WHOLE EXOME SEQUENCING OF EPILEPTIC GENE FOR DETERMINATION OF SEQUENCE VARIATION AND MUTATIONAL ANALYSIS

M.Sc. Thesis 2016

Submitted to

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

By

Sunita Koirala

Registration No: 5-2-48-2593-2005

Supervisors

Dr. Tilak R. Shrestha
Professor
Central Department of Biotechnology,
Tribhuvan University,
Kirtipur, Nepal

Dr. Jozef Gecz
Professor
Senior Principal Research Fellow,
NH&MRC, University of Adelaide,
Australia

ACKNOWLEGMENT

I express deepest gratitude to my supervisor Prof. Tilak R. Shrestha, for guiding me through my dissertation and encouraging me to work at the Neurogenetics laboratory, University of Adelaide, Adelaide, Australia as a part of my research work. It was through his guidance and persistence that I got an opportunity to work at such a prestigious laboratory, which not only helped me in expanding the horizon of my knowledge but also no doubt shall act as a major stepping stone in all my future endeavors.

I am particularly thankful to Prof. Jozef Gecz for his generosity to work within his group leading investigations in the field of genetic and molecular biology of neurodevelopmental disorders.

I am grateful to the entire Neurogenetic Group at the Women's and Children's Hospital, an outstanding team required for the completion of this project.

I am highly indebted to Dr. Raman Sharma for his patience and guidance in teaching me the valuable lessons through informative discussions and advice. He was a constant source of inspirations that guided me through the tough tasks and help me accomplish them.

I am thankful to Prof. Krishna Das Manandhar, Head of Department of Central Department of Biotechnology, Tribhuvan University.

I am also grateful to members of the Neurogenetics Group, specially Marie Shaw, Mark Corbett, Alison Gardner, Duyen Pham, Stanley Tan, Thuong Ha, Sayakya Kayumi and Anthony Correll without whom the completion of my work would not have been possible. Through their constant support, I was able to learn and build up confidence to carry out many of the experiments on my own and learnt basic concepts of doing real science. I am especially grateful to them for showing utmost level of patience while I was in my initial learning stage in genetics and they gracefully helped me to overcome my deficiencies. I am thankful to all the wonderful members of Evan Eichler Collaboration.

I am thankful to wonderful people of Central Department of Biotechnology all the faculty members and teaching and nonteaching staffs of Central Department of Biotechnology, seniors all the people involved directly and indirectly and my loving friends.

At last but not the least, I am emphatically thankful to my parents, whose support and love helped me get through some of the toughest moments and always encouraged me to perform better and never give up the hope.

ABBREVIATIONS

A	-	Adenine
AA	-	Amino Acid
ADNFLE	-	Autosomal dominant nocturnal frontal lobe epilepsy
AP1B1	-	Adaptor Related Protein Complex 1 Beta 1 Subunit Glyoxalase Domain
ATP6V1B2	-	ATPASE H+ Transporting V1 Subunit B2
ANK3	-	Ankyrin 3
ATP7B	-	ATPase Copper Transporting Beta
ATXN3	-	Ataxia Type 3
C	-	Cytosine
CNV	-	Copy Number Variant
CADD	-	Combined Annotation Dependent Depletion
cDNA	-	Complementary DeoxyriboNucleic Acid
dATP	-	2ˈ- Deoxyadenosine 5ˈ-Triphosphate
dCTP	-	2ˈ- Deoxycytidine 5ˈ-Triphosphate
dGTP	-	2ˈ- Deoxyguanosine 5ˈ-Triphosphate
DLG2	-	Disks large homolog 2
DPYSL3	-	Dihydropyrimidinase like 3
DNA	-	Deoxy-ribose Nucleic Acid
dNTP	-	2ˈ- Deoxynucleotide 5ˈ-Triphosphate
dTTP	-	2ˈ- Deoxythymidine 5ˈ-Triphosphate
EDTA	-	Ethylene-diamine-tetra-acetic acid
EE	-	Epileptic Encephalopathy
EEG	-	Electroencephalography
EtBr	-	Ethidium Bromide
G	-	Guanine
GGE	-	Genetic generalized epilepsy
GLOD5	-	Glyoxalase Domain Containing 5
HGNC	-	Human Genome Nomenclature Committee

ID	-	Intellectual disability
IGE	-	Idiopathic generalized epilepsy
Kb	-	Kilobases
MAP7D3	-	Microtubule-associated protein 7 Domain Containing 3
MEGF8	-	Multiple Epidermal Growth Factor-like Domains 8
MLPA	-	Multiplex Ligation-dependent Probe Amplification
MAOA	-	Monoamine oxidase a
NCOA1	-	Nuclear Receptor Coactivator 1
NFW	-	Nuclease Free Water
NALCN	-	Sodium Leak Channel
NGS	-	Next Generation Sequencing
NCOR1	-	Nuclear receptor co-repressor gene
ORF	-	Open Reading Frame
OTOG	-	Otogelin
PCR	-	Polymerase Chain Reaction
PIK3R4	-	Phosphoinositide-3-Kinase Regulatory Subunit 4
POGZ	-	Pogo transposable element with Zinc Finger domain
PPP2R1A	-	Protein phosphatase 2 (formerly 2a), regulatory subunit a (pr 65), alpha isoform
RNA	-	Ribonucleic Acid
rpm	-	Revolutions per Minute
RT PCR	-	Reverse Transcription Polymerase Chain Reaction
SNP	-	Single-Nucleotide Polymorphism
SNV	-	Single-Nucleotide Variant
T	-	Thymine
TE/TAE	-	Tris-EDTA/Tris-Acetate-EDTA
TBE	-	Tris-Borate-EDTA
WES	-	Whole Exome Sequencing
WTG	-	Wild Type Genotype

Table of content

List of Figure

List of Table

ABSTRACT

Epilepsy is a group of disorders followed by frequent seizures. The genetic bases of epilepsy are clear from epidemiological studies and from rare gene findings. Advances in genomic technologies have facilitated genome-wide discovery of both common and rare variants which have led to a rapid increase in our understanding of epilepsy genetics. Genetic insights into the epilepsies have come primarily from Next generation sequencing (Whole exome sequencing and Whole genome sequencing) which has emerged as a building block for the identification of disease-causing mutations. In our study whole exome sequencing has been used for the screening of mutations in patients after the validation of variants to identify novel genes. Sequencing was performed on Illumina platform. After prioritization of variants, validation of variants was done by different bioinformatics analysis tools and one putative variant was screened. The causative mutation was validated by capillary sequencing and also the carrier status among the family members was checked. We ended with the prediction of the mutation in the gene to be real with the screening of eleven variants with the prediction of gene *PPP2R1A* to be real in family 1, the variant c.G2833A:p.E945K present in exon 17 in family 2, *MAOA* leading affected males for three generations with Intellectual disability in family 3, c588A>T:p.R196S and c.587G>A:p.R196K in exon 8 with mutation in gene *DLG2* in family 4,c.G953A:p.R318H with mutation in *ALG13* of fifth family, p.1673T with mutation of *GRIA1* in family 6, c.C1058T:p.A353V as the causing factor of mutation in *DPYSL3* in family 9 and , a novel gene *ATP6V1B2* in family 11 and the cause of the disorders. Most of these genes are expressed in the brain and instruct subunits of ion channels that play vital roles in stabilizing and propagating neuronal activity. Disruption of these genes in general induces neuronal hyper excitability, thus causing seizures. Epilepsy syndromes are found to be the result of severe mutations in single genes whereas more commonly epilepsy is likely to be caused by the combined effect of variants in number of genes.

Key words: Epilepsy, Mutation, whole exome sequencing, novel genes, seizures.

CHAPTER 1

INTRODUCTION

1.1 Background

Epilepsy is a disorder of the central nervous system characterized by recurrent seizures unprovoked by acute systemic or neurologic insults (Bromfield, 2006). It causes nerve cells in the brain to signal abnormally causing strange sensations, emotions, behavior, muscle spasms and loss of consciousness. People with epilepsy have recurring seizures that often occur spontaneously and without warning. The official definition of a seizure is a transient occurrence of signs and symptoms due to an abnormal excessive or synchronous neuronal activity in the brain (Fisher, 2014).

During a seizure, large numbers of brain cells are activated abnormally at the same time. It is like an "electrical storm" in the brain.

The nature of the seizures depends on many factors, such as the person's age, the sleep-wake cycle, prior injuries to the brain, genetic tendencies, medications, which circuits in the brain are involved, and many others (Fisher *et al.,* 2016).

Separating seizures into different types helps guide further testing, treatment, and prognosis or outlook. Using a common language for seizure classification also makes it easier to communicate among clinicians caring for people with epilepsy and doing research on epilepsy. The classification also provides common words for people with epilepsy and the general public to describe their seizures.

1.2 Classification of epileptic seizures

Classification of epileptic seizures proposed by the Commission on Classification and Terminology of the International League against Epilepsy (ILAE) and approved in September 1981. This classification is based on the clinical expression of the seizure and the electro-encephalographic picture during and between the seizures (Fisher, 2016).

1.3 International classification of epileptic seizures

The most common type (60%) of seizures is sudden. Of these, one third begin as generalized seizures from the start, affecting both hemispheres of the brain. Two-thirds begin as partial seizures (which affect one hemisphere of the brain) which may then progress to generalized seizures. The remaining 40% of seizures are non-convulsive (Sharma S. *et al*., 2013).

Partial seizures begin in one side of the brain. Partial seizures are the most common type of seizure experienced by people with epilepsy. Partial seizures consist visual, hearing, smell, psychic, autonomic, and motor phenomena. Jerking activity may start in a specific muscle group and spread to surrounding muscle. Automatisms may occur which are non-consciously generated activities and mostly simple repetitive movements like smacking of the lips or more complex activities such as attempts to pick up something.

Generalized seizures include tonic-clonic, tonic, clonic, myoclonic, absence and atonic seizure. They all involve loss of consciousness and typically happen without warning (Sharma *et al.*, 2013). Tonic-clonic seizures occur with a contraction of the limbs followed by their extension along with arching of the back which lasts 10–30 seconds (the tonic phase).Tonic seizures result in continuous muscle contraction and rigidity, while tonic-clonic seizures involve alternating tonic activity with rhythmic jerking of muscle groups.

A cry may be heard due to contraction of the chest muscles, followed by a shaking of the limbs in unison (clonic phase). Tonic seizures produce constant contractions of the muscles. Myoclonic seizures involve spasms of muscles in either a few areas or all over. Absence seizures are most common in children and naturally don't cause any long-term problems. The person does not fall over and returns to normal right after it ends. Atonic seizures involve loss of muscle activity for greater than one second. This typically occurs on both sides of the body sensory symptoms (Gashaut *et al.*, 1964).

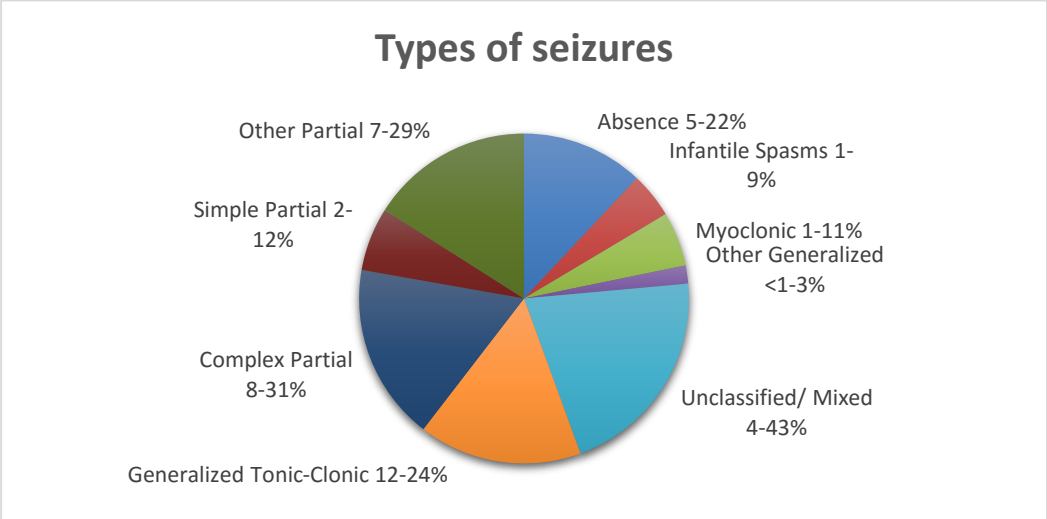

Fig 1.1: Types of Seizures. Pie chart includes ranges of seizures. Unclassified seizures range from 4 to 43%, seizures with inadequate and incomplete data whereas myoclonic and infantile spasms comes under similar range (Cowan, 2002).

1.4 Etiology

Epilepsy may be an isolated neurological symptom, or it may occur as part of a more complex syndrome. There are many different causes of epilepsy, including genetic disorders, metabolic diseases, and structural brain abnormalities. However, in some cases, the cause of epilepsy is not known. The International League against Epilepsy (ILAE) has described the underlying causes of epilepsy as the following (Fisher, 2016).

Genetic: "The concept of genetic epilepsy is that the epilepsy is, as best as understood, the direct result of a known or presumed genetic defect(s) in which seizures are the core symptom of the disorder". The genetic contribution must be evident by extensive and replicable molecular studies or familial studies.

Structural/metabolic**:** If there is a specific structural or metabolic disorder that has been proven in various studies to be associated with increased risk of epilepsy, the cause of epilepsy can be defined as structural or metabolic. The cause for such a disorder can be acquired (e.g. trauma, stroke, infection), genetic (malformations of cortical development) or both (e.g. West syndrome).

Unknown cause: This means that the underlying cause is yet unknown. Epilepsy might be presenting because of either some genetic defect or as a result of disorder that is not yet recognized.

1.5 Types of Epilepsy

Although the definition of epilepsy suggests that it is a single disorder, it is more accurate to describe epilepsy as a group of disorders with diverse etiologies and outcomes. There are many different epilepsy syndromes that can be characterized by the seizure type(s), age of onset, developmental status, co-morbid features and etiology. There are then several specific syndromes within each class defined by differences in specific seizure types, electroencephalogram (EEG) patterns and disease progression (Stafstrom and Carmant, 2015). Epilepsy can be broadly grouped into three classes which are Genetic generalized epilepsy (GGE) formerly idiopathic generalized epilepsy, Focal epilepsy and Epileptic encephalopathy (EE).

The GGE syndromes, characterized by generalized seizures that involve both sides of the brain, include juvenile myoclonic epilepsy and childhood absence epilepsy among others. The GGEs tend to start in childhood or adolescence and are usually associated with normal development and intellect (Pella *et al*., 2014).

Focal seizures originate in one hemisphere of the brain. Examples of focal epilepsy syndromes are temporal lobe epilepsy, autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE), and autosomal dominant epilepsy with auditory features.

The EEs are severe, early onset condition characterized by refractory seizures, developmental delay or regression associated with ongoing epileptic activity, and generally poor prognosis. Dravet, Ohtahara and West syndromes are some of the well-studied EEs. Importantly, epilepsy is often a co-morbid condition in individuals with intellectual disability (ID), autism or schizophrenia and may be a feature of many metabolic conditions and genetic syndromes (Myers and Mefford, 2015).

1.6 Diagnosis and clinical presentation

The types of seizures determines to the presentation and characteristics of syndromes related to epilepsy. In-case of clonic seizure include jerking (clonic), muscles becoming limp or weak (atonic), tense or rigid muscles (tonic), brief muscle twitching (myoclonus), or epileptic spasms (Conway *et al*.,2015). There may also be automatisms or repeated automatic movements, like clapping or rubbing of hands, lip smacking or chewing, or running. Some of the clinical presentations are as follows:

- Changes in sensation, emotions, thinking or cognition, autonomic functions (such as gastrointestinal sensations, waves of heat or cold, goose bumps, heart racing etc).
- Lack of movement (behavior arrest).
- Epileptic spasms.
- Movement stops, the person may just stare and not make any other movements.

1.6.1 Laboratory diagnosis

Epilepsy is usually difficult to diagnose quickly. In most cases, it cannot be confirmed until you have had more than one seizure. Laboratory studies done for the diagnosis of individuals with seizures include the follow complete neurological consultation for epilepsy and related conditions. Neurophysiological tests including routine EEGs and outpatient and inpatient video-EEG monitoring, Long-term inpatient video-EEG monitoring with scalp or intracranial electrodes and neuroimaging in general (Andrade *et al*., 2017) are some common ways to diagnose it.

a. Electroencephalogram (EEG) test can detect unusual brain activity associated with epilepsy by measuring the electrical activity of your brain through electrodes. The purpose of the EEG recording is to detect interictal activity and localize the region of

interictal activity. In the presence of epileptiform activity, the EEG recording will also help to determine the type of seizure or epilepsy syndrome (Flink *et al*., 2002).

b. Magnetic resonance imaging (MRI) scan uses strong magnetic fields and radio waves to produce detailed images of the inside of body. It can be useful in cases of suspected epilepsy because it can often detect possible causes of the condition, such as defects in the structure of brain or the presence of a brain tumor.

1.6.2 Molecular diagnosis

Molecular genetic tests includes detection of specific pathogenic variants-genotyping sequencing of a gene to detect pathogenic variants, amplification or hybridization methods. To detect these potentially pathogenic alterations, known to be involved in several epilepsies, panels should be coupled with additional analysis methods such as multiplex amplicon quantification (MAQ), multiplex ligation-dependent probe amplification (MLPA) and array comparative genomic hybridization (CGH) (Ceulemans *et al.*, 2012) .These are analyzed with an NGS structural variations detection method (qPCR, array CGH, MLPA) to detect copy number variants involving one or more genes. Methylation targets to specific techniques to detect epigenetic changes that influence gene expression. Exome sequencing targets on sequencing of the complete exons set of human.

a. Array comparative genomic hybridization (CGH)

Discovery in epilepsy genetics, inevitably, has followed technological advances (Scheffer IE, Mefford HC, 2014). In the current era, the first new genetic technology that became widely available was array comparative genomic hybridization (aCGH), which permits comparison of segments of patient's DNA with, typically, pooled DNA from a group of controls without the same condition, usually healthy individuals. The technique highlights segments where the number of copies of that segment is different to that seen in controls (copy number variation), down to a certain size resolution, usually of order of few kilobases but occasionally with higher resolution (Mefford, 2011). CGH is now offered by many clinical genetics laboratories and because of its higher resolution and reliability, has replaced karyotyping as the first-line testing complex epilepsy phenotypes that do not implicate an obvious candidate gene. CGH is indicated when the presenting epilepsy is syndromic, being associated with other features, such as facial or somatic dysmorphism, intellectual disability, autism spectrum disorder or multiple co-morbidities. Microdeletions and microduplications together falling within the category of copy number variants, CNVs have been increasingly reported in association with complex epilepsies (Olson *et al.*, 2014; Helbig *et al.*, 2014), and have sometimes pointed to novel candidate epilepsy genes (Saitsu *et al,* 2008).

b. Whole Exome Sequencing

Whole Exome sequencing of parent-offspring trios offers a cost-effective method for screening coding regions for mutations and has been successful in identifying candidate de novo variants in sporadic cases of Intellectual disability (Vissers *et al*., 2010). However, the limitations of current exome capture and sequencing methodologies include incomplete or variable coverage of exons and the inability to study regulatory variation. Whole-genome sequencing (WGS) studies are not limited by these aspects, and when they are implemented in a framework, they have methodical advantages. The basic steps in WES involves lysing cells to extract DNA, library preparation, cluster generation, high through-put massively parallel sequencing and computational data analysis (Metzker, 2010). Library preparation involves optimal fragmentation of the DNA, end repair of the fragments, adapter ligation, PCR enrichment, probes hybridization and capture and finally PCR enrichment and size selection (Bamshad *et al*., 2011; Metzker, 2010). Those captured exome are then clonally amplified to generate millions of clusters which are then subjected for sequencing. The raw sequence is then processed through various quality check softwares such as FASTQC and Trimmomatic (Rimmer *et al.*, 2014). The quality passed data are then aligned against reference genome and variants are called and annotated. Finally the causal mutation is identified among the numerous variations by adopting different approaches such as removing common variations using public databases.

c. Next Generation Sequencing

Next Generation Sequencing has contributed to the identification of many monogenic epilepsy syndromes and is favoring earlier and more accurate diagnosis in a subset of pediatric patients with epilepsy. Massively-parallel-sequencing technologies are revolutionizing the process of discovering genetic variants that cause disease (Bamshad *et al*., 2011). Although ion channel genes represent the gene family most frequently causally related to epilepsy, other genes have gradually been associated with complex developmental epilepsy conditions, revealing the pathogenic role of mutations affecting diverse molecular pathways that regulate membrane excitability, synaptic plasticity, neurotransmitter release, postsynaptic receptors, transporters, cell metabolism, and many formative steps in early brain development. No specific treatment is available for most of the monogenic disorders that can now be diagnosed early using NGS, and the main benefits of knowing the specific cause include etiological diagnosis, better prognostication and genetic counseling however, for a limited number of disorders, timely treatment based on their known molecular pathology is already possible and sometimes decisive. Discovery of a causative gene defect associated with a non-progressive course may reduce the need for further diagnostic investigations in the search for a progressive disorder at the biochemical and imaging level (Mei *et al*., 2017).

1.7 Global outlook

1.7.1 Salient findings

A total of about 43,704,000 people with epilepsy are reported from 108 countries covering 85.4% of the world's population. The mean number of people with epilepsy per 1000 population is 8.93 (SD 8.14, median 7.59) from 105 responding countries.

While it is 12.59 and 11.29 in the Americas and Africa, respectively, it is 9.97 in South-East Asia, 9.4 in the Eastern Mediterranean, 8.23 in Europe, and 3.66 in the Western Pacific. The mean number of people with epilepsy per 1000 population ranges from 7.99 in the high-income countries to 9.50 in the low-income countries (Reynolds, 2005).

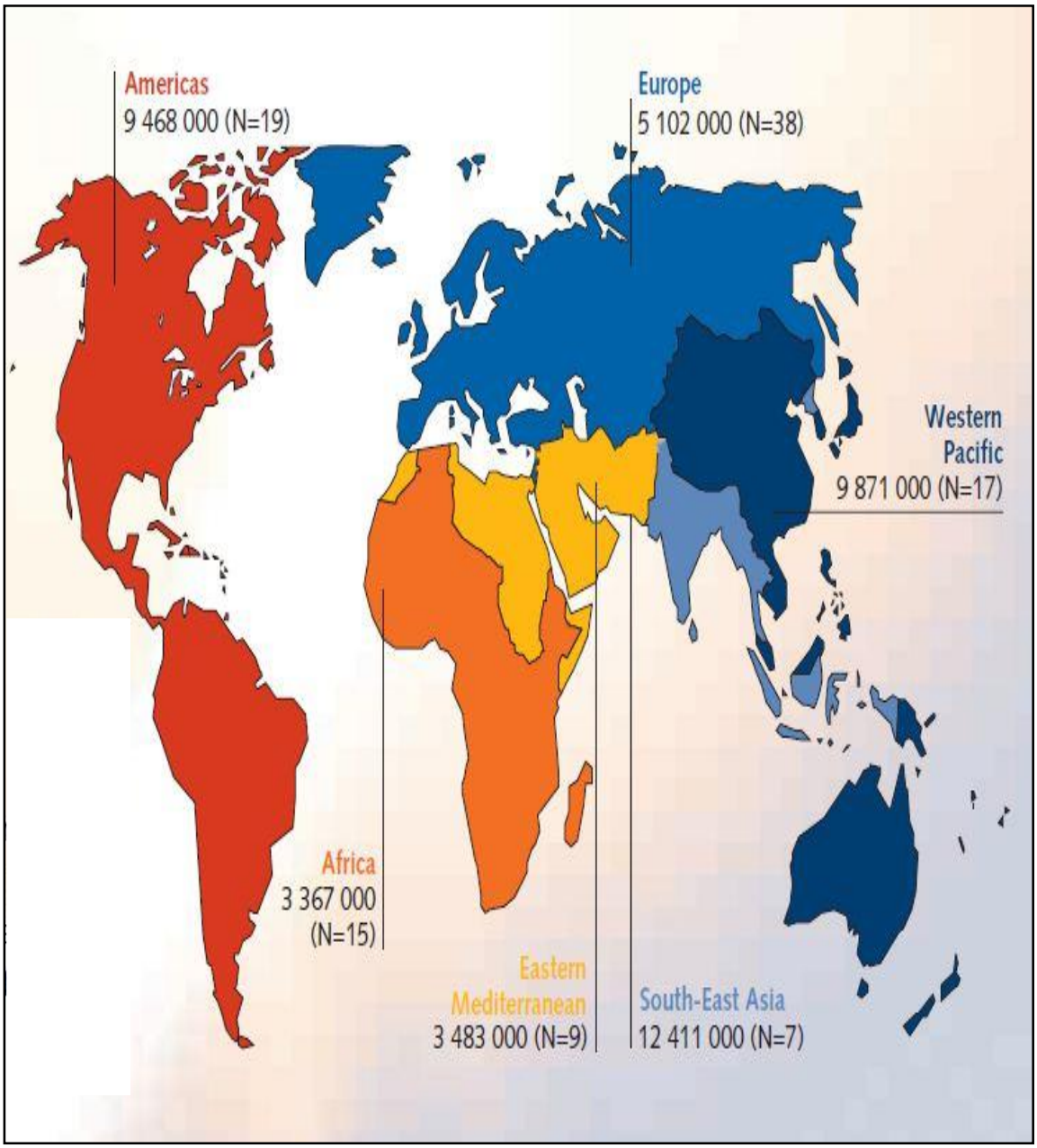

Fig 1.2: **Global outlook of people with epilepsy**. (Epilepsy Atlas WHO, 2005).These numbers are only indicative based on the information provided by Atlas respondents. The numbers have not been corrected for non-responding countries. The number of people with epilepsy in WHO regions is estimated to N=105*.The mean number of people with epilepsy per 1000 population varies across region.

1.7.2 Limitations

The data regarding the number of people with epilepsy were not collected using stringent research methods as for epidemiological studies; such methods are costly and are not easy to carry out. The sources of information vary across responding countries, limiting the interpretation of the data set. For example, some respondents provided figures based on generic prevalence or findings from one particular area of the country or on the number of people eligible for anti-epileptic drugs. One of the reasons for low prevalence reported from the Western Pacific could be the lower prevalence rates reported from Pacific Islands; it also puts a bias on the global outcome, as the Western Pacific comprises 27% of the population in all WHO regions. Information regarding the number of people with active epilepsy was not obtained regarding the number of people with epilepsy among special groups, e.g. children, was not obtained.

1.7.3 Conclusion

The number of people with epilepsy is high in most regions of the world, thus constituting epilepsy as a major public health concern. There is a need to carry out multinational epidemiological studies using standardized definitions and case ascertainment methods.

1.8 Genes observed under research study corresponding to putative variants

The genes which were found to be associated in epileptic diseases were identified from eleven families. Out of these eleven genes thus analyzed, the *ATP6V1B2* gene was found to be novel.

1.8.1. *PPP2R1A* (Protein phosphatase 2 scaffold subunit A alpha)

This gene encodes a constant regulatory subunit of protein phosphatase 2 with Size, 589 amino acids and molecular mass 65309 Da with genomic size 26613bp, Exon and coding exon count 9. It is one of the four major Ser/Thr phosphatases, and it is implicated in the negative control of cell growth and division.

It consists of a common heteromeric core enzyme, which is composed of a catalytic subunit and a constant regulatory subunit that associates with a variety of regulatory subunits. The constant regulatory subunit A serves as a scaffolding molecule to coordinate the assembly of the catalytic subunit and a variable regulatory B subunit. (RefSeq, Apr 2010). This is a Protein Coding gene. Diseases associated with PPP2R1A include mental retardation, Autosomal dominant 26 and endometrioid ovary carcinoma. Among its related pathways are P70S6K/ Ribosomal protein S6 kinase beta-1 (S6K1) signaling and Reelin Pathway (cajal-Retzius cells). This gene involves protein heterodimerization activity and antigen binding. Interacts with FOXO1 (Fork head box protein 01) the interaction dephosphorylates FOXO1 on AKT-mediated

phosphorylation sites. PP2A consists of a common hetero dimeric core enzyme, composed of PPP2CA a 36 kDa catalytic subunit (subunit C) and *PPP2R1A* a regulatory subunit (PR65 or subunit A), that associates with a variety of regulatory subunits. This gene has 12 transcripts (splice variants), 59 orthologues, 1 paralogue, is a member of 1 Ensembl protein family and is associated with 31 phenotypes (Source: HGNC Symbol;Acc:HGNC:9302).

1.8.2. *POGZ* (Pogo transposable element with Zinc Finger domain)

This gene plays a role in mitotic cell cycle progression and is involved in Kinetochore assembly and mitotic sister chromatid cohesion. This protein was found to interact with the transcription factor SP1 in a yeast two-hybrid system. Alternatively spliced transcript variants encoding distinct isoforms have been observed (provided by Ref Seq, Aug 2010).Protein Coding gene. Diseases associated with POGZ include mental retardation, autosomal dominant 37 and Syndromic Intellectual Disability (RefSeq: NM_207171.2). This gene has 21 transcripts (splice variants), 65 orthologues, 4 paralogues, is a member of 2 Ensembl protein families and is associated with phenotypes in exon 17.

Probably through its association with CBX5 plays a role in mitotic chromosome segregation by regulating aurora kinase B/AURKB activation and AURKB dissociation from chromosome arms (POGZ_HUMAN, Q7Z3K3).The protein encoded by this gene appears to be a zinc finger protein containing a transposase domain at the C-terminus (Source: HGNC Symbol; Acc:HGNC:18801).

1.8.3. *MAOA* (Monoamine oxidase A)

This gene is one of two neighboring gene family members that encode mitochondrial enzymes which catalyze the oxidative deamination of amines, such as dopamine, norepinephrine, and serotonin. Mutation of this gene results in Brunner syndrome. This gene has also been associated with a variety of other psychiatric disorders, including antisocial behavior. Alternatively spliced transcript variants encoding multiple isoforms have been observed (RefSeq, Jul 2012). This gene has 4 transcripts (splice variants), 66 orthologues, 1 paralogue, is a member of 1 Ensembl protein family and is associated with 3 phenotypes (Source: HGNC Symbol;Acc:HGNC:6833)

1.8.4. *DLG2* (Discs large scaffold protein 2)

This gene is one of two neighboring gene family members that encode mitochondrial enzymes which catalyze the oxidative deamination of amines, such as dopamine, norepinephrine, and serotonin. This gene has also been associated with a variety of other psychiatric disorders, including antisocial behavior. Alternatively spliced transcript variants encoding multiple isoforms have been observed (RefSeq, Jul 2012).This gene has 30 transcripts (splice variants), 69 orthologues, 12 paralogues, is a member of 1 Ensembl protein family and is associated with 59 phenotypes. (Source: HGNC Symbol; Acc: HGNC: 2901)

1.8.5. *ALG13* (UDP-N-acetylglucosaminyltransferase subunit)

The protein encoded by this gene is a subunit of a bipartite UDP-N-acetylglucosamine transferase. It heterodimerizes with asparagine-linked glycosylation 14 homolog to form a functional UDP-GlcNAc glycosyltransferase that catalyzes the second sugar in addition of the highly conserved oligosaccharide precursor in endoplasmic reticulum N-linked glycosylation (RefSeq, Dec 2009). This gene has 41 transcripts (splice variants), 60 orthologues, 2 paralogues, is a member of 2 Ensembl protein families and is associated with 5 phenotypes in exon8. Multiple transcript variants encoding different isoforms have been found for this gene (Source: HGNC Symbol; Acc:HGNC:30881) .

1.8.6. *GRIA1* (Glutamate ionotropic receptor AMPA type subunit 1)

Glutamate receptors are the predominant excitatory neurotransmitter receptors in the mammalian brain and are activated in a variety of normal neurophysiologic processes. These receptors are heteromeric protein complexes with multiple subunits, each possessing transmembrane regions, and all arranged to form a ligand-gated ion channel. The classification of glutamate receptors is based on their activation by different pharmacologic agonists. This gene belongs to a family of alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) receptors. Alternatively spliced transcript variants encoding different isoforms have been found for this gene (RefSeq, Jul 2008). This gene has 13 transcripts (splice variants), 75 orthologues, 10 paralogues, is a member of 1 Ensembl protein family and is associated with 17 phenotypes (Source: HGNC Symbol;Acc:HGNC:4571).

1.8.7. NCOR1 (Nuclear receptor corepressor 1)

This gene encodes a protein that mediates ligand-independent transcription repression of thyroid-hormone and retinoic-acid receptors by promoting chromatin condensation and preventing access of the transcription machinery. It is part of a complex which also includes histone deacetylases and transcriptional regulators similar to the yeast protein Sin3p. This gene is located between the Charcot-Marie-Tooth and Smith-Magenis syndrome critical regions on chromosome 17. Alternate splicing results in multiple transcript variants. Pseudogenes of this gene are found on chromosomes 17 and 20 (RefSeq, Jun 2010).This gene has 27 transcripts (splice variants), 72 orthologues, 4 paralogues, is a member of 1 Ensembl protein family and is associated with 37 phenotypes (Source: HGNC Symbol; Acc: HGNC: 7672).

1.8.8. NALCN (Sodium leak channel)

This gene has 5 transcripts (splice variants), 72 orthologues, 8 paralogues, is a member of 1 Ensembl protein family and is associated with 9 phenotypes. NALCN forms a voltage-independent, nonselective, noninactivating cation channel permeable to Na+, K+, and Ca (2+). It is responsible for the neuronal background sodium leak conductance (Lu *et al*., 2007).NALCN is a Protein Coding gene. Diseases associated with NALCN include hypotonia, infatile with psychomotor retardation and

contractures of the limbs (OMIM, Mar 2008). Among its related pathways are ion channel pathway and transport of glucose and other sugars (Source: HGNC Symbol; Acc:HGNC:19082)

1.8.9. DPYSL3 (Dihydropyrimidinase like 3)

It is a Protein Coding gene. Diseases associated with DPYSL3 include hypertensive encephalopathy. Among its related pathways are semaphorin interactions and developmental biology. Observations related to this gene include hydrolase activity and phosphoprotein binding. This gene has 9 transcripts (splice variants), 74 orthologues, 5 paralogues, is a member of 1 Ensembl protein family and is associated with 1 phenotype Source : HGNC Symbol;Acc:HGNC:3015).

1.8.10. ANK3 (Ankyrin 3)

Ankyrins are a family of proteins that are considered to bond the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play key roles in activities such as cell motility, activation, proliferation, contact, and the conservation of specialized membrane domains. Multiple isoforms of ankyrin with different affinities for various target proteins are expressed in a tissue-specific, developmentally regulated manner.

Most ankyrins are typically combined of three structural domains: an amino-terminal domain containing several ankyrin repeats; a central region with a highly conserved spectrin binding domain; and a carboxy-terminal regulatory domain which is the least conserved and subject to variation. Ankyrin 3 is an immunologically distinct gene product from ankyrins 1 and 2, and was initially found at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Multiple transcript variants encoding different isoforms have been found for this gene (provided by RefSeq, Feb 2011).This gene has 32 transcripts (splice variants), 77 orthologues, 11 paralogues, is a member of 1 Ensembl protein family and is associated with 2 phenotypes(Source: HGNC Symbol;Acc:HGNC:494).

1.8.11. ATP6V1B2 (ATPase H+ transporting V1 subunit B2)

This gene encodes a component of vacuolar ATPase (V-ATPase), a multi-subunit enzyme that mediates acidification of eukaryotic intracellular organelles. V-ATPase dependent organelle acidification is necessary for such intracellular processes as protein sorting, zymogen activation, receptor-mediated endocytosis, and synaptic vesicle proton gradient generation.

This gene has 6 transcripts (splice variants), 70 orthologues, 1 paralogue, is a member of 1 Ensembl protein family and is associated with 5 phenotypes (RefSeq, Feb 2011).V-ATPase is composed of a cytosolic V1 domain and a transmembrane V0 domain. The V1 domain consists of three A, three B, and two G subunits, as well as a C, D, E, F, and H subunit. The V1 domain contains the ATP catalytic site. The protein encoded by this gene is one of two V1 domain B subunit isoforms and is the only B isoform highly expressed in osteoclasts (Source: HGNC Symbol; Acc: HGNC: 854).

1.9 Objectives

1.9.1 General Objective

i. Identifying novel variants associated with epilepsy.
ii. Identification of novel genes with follow up screening of additional patient cohorts to find additional causes with mutation in those genes.

1.9.2 Specific Objectives

a. To perform whole Exome sequencing of epileptic genes.
b. Analysis and interpretation of WES data.
c. Selection of the possible putative gene on the basis of functional features and known gene.
d. Segregation of variants.
e. To find out the novel mutation if any, present in the epileptic gene.

1.10 Hypothesis

The study will lead to identification of novel genes involved in epilepsy and mutation in those genes and finding new genetic causes of neuro-developmental disorders. It may also lead in finding out relationship with protein domain, identification of new pathways and contribute in molecular works.

1.11 Rationale of the study

The study of the mutation in genes related with epilepsy will allow better understanding of epilepsy and its diagnosis thus lead to the development of recombinant therapeutic products as well as gene therapy. In addition, the detection of carrier status, prediction of development of inhibitors and genetic counseling can be done. Such type of study will also provide deeper insights for the disease pattern which is shared by the patients with similar phenotypic complications thus can support the patients to deal with the upcoming generations too.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

Within 17th and 18th centuries epilepsy was considered as a brain disorder. During these two centuries epilepsy was one of several key areas of debate in the gradual identification and separation of nervous disorders from mental disorders, which led to the beginnings of modern neurology in the 19th century. Many treatises on convulsive diseases appeared which included hysteria, tetanus, tremors, rigors and other paroxysmal movement disorders. The latter were gradually separated off from epilepsy in the 19th century, as illustrated in the distinguished lectures on convulsive diseases by Robert Bentley Todd in 1849 and Jackson in 1890. With the development of neuropathology as a new discipline in the 19th century, there also began a great debate, which is still with us to some extent, as to the distinction between pure primary idiopathic epilepsy, in which the brain is macroscopically normal, from secondary symptomatic epilepsy, associated with many different brain pathologies (Reynolds, 2005).

Several attempts have been made recently to reduce the difficulties in the classification of epilepsy (Gastaut *et al*., 1964). These classifications were based not only on detailed clinical observation of the seizures, but also on EEG recordings. Gastaut's classification, as presented at the 11th International Congress of the International League against Epilepsy, and published later is the latest and most comprehensive attempt of this kind. However, the data on which Gastaut established his classification are not always fully available to the neurologist and electroencephalographer. Reports of seizures are sometimes scant, details are overlooked by the patient and relatives, and there is little opportunity to observe a seizure during the short session when the EEG is being taken. With the development of the concept of functional localization in the brain and the discovery, for example, of the motor cortex, the concept of epileptiform or partial seizures arose as models for the study of generalized seizures. By studying the clinical features, seizures were more difficult to classify and localize; it was not until the discovery of human electroencephalography (EEG) in the 20th century that the concepts of temporal lobe or frontal epilepsy were gradually clarified. In the second half of the 20th century remarkable progress was made in diagnostic facilities and possibilities through structural and functional neuroimaging.

2.2 Pathophysiology of seizures

Seizure is the clinical manifestation of epilepsy. This occurs basically due to excessive firing of the neurons and fast spread of these impulses over the brain. There are two phenomenons in pathophysiology of a seizure hyper-excitability of a neuron and hyper synchronization. Hyper synchronization happens when a large electrical impulse is generated in one part of the brain from a focus of tissues millions of neurons in the brain fire excessively in addition bringing on a seizure. Neuronal messages are transmitted by electrical impulses (Benarroch *et al*., 2011). This is actually a net positive inward ion flux that leads to depolarization or voltage change in the neuronal membrane. The ions involved include sodium, potassium, calcium and chloride. Normally brain tissues prevent hyper excitability by several inhibitory mechanisms involving negative ions like chloride ions.

Disturbance in this normal excitability leads to hyper-excitability. In this state there is increases excitatory transmission of impulses and decreases inhibitory transmission. In addition there is alteration in the voltage gated ionic channels. These ion channels normally open when the voltage difference across the neuronal membrane is changed favorably (Siegelbaum & Hudspeth, 2000). Once activated the impulses flow via the neuronal circuits along the axons of the nerves. An action potential travels down the axon to the terminal buttons and then releases neurotransmitters in the synaptic cleft. This carries the action potential from one nerve to another. Excitatory transmission involves Glutamate that is the principal excitatory neurotransmitter in the brain. GABA or Gamma amino butyric acid is the principal inhibitory neurotransmitter in the brain as shown in Fig 2.1.

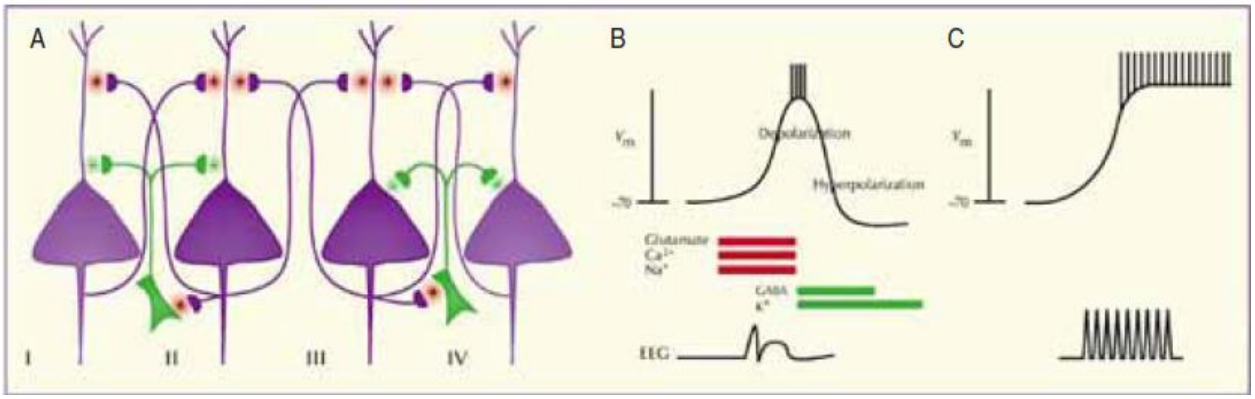

Fig 2.1: Mechanism of Paroxysmal depolarization shift (PDS). GABA is mediated via Chloride and Potassium channel. Excitation of a group of nerves is caused by inward currents of Na, Ca and involvement of excitatory neurotransmitters like Glutamate and Aspartate (GL *et al*., 2003).

When a PDS occurs as an abnormally prolonged run of action potentials during sustained membrane depolarization in a single neuron, as shown in the upper trace in B, the event is

detectable only with microelectrodes, increased glutamate concentration is associated with influx of cations initially, followed by increased GABA concentration with efflux of potassium. When PDSs in a large number of neurons are synchronized for less than 200 ms, as shown in A, these electrical potentials may summate as a spike-wave complex that is recorded with macro-electrodes, as shown in the lower trace in B. When sustained repetitive firing of PDSs in a large number of neurons becomes synchronized for many seconds or longer, an electrographic seizure occurs, as shown in C (Henry *et al*., 2012).

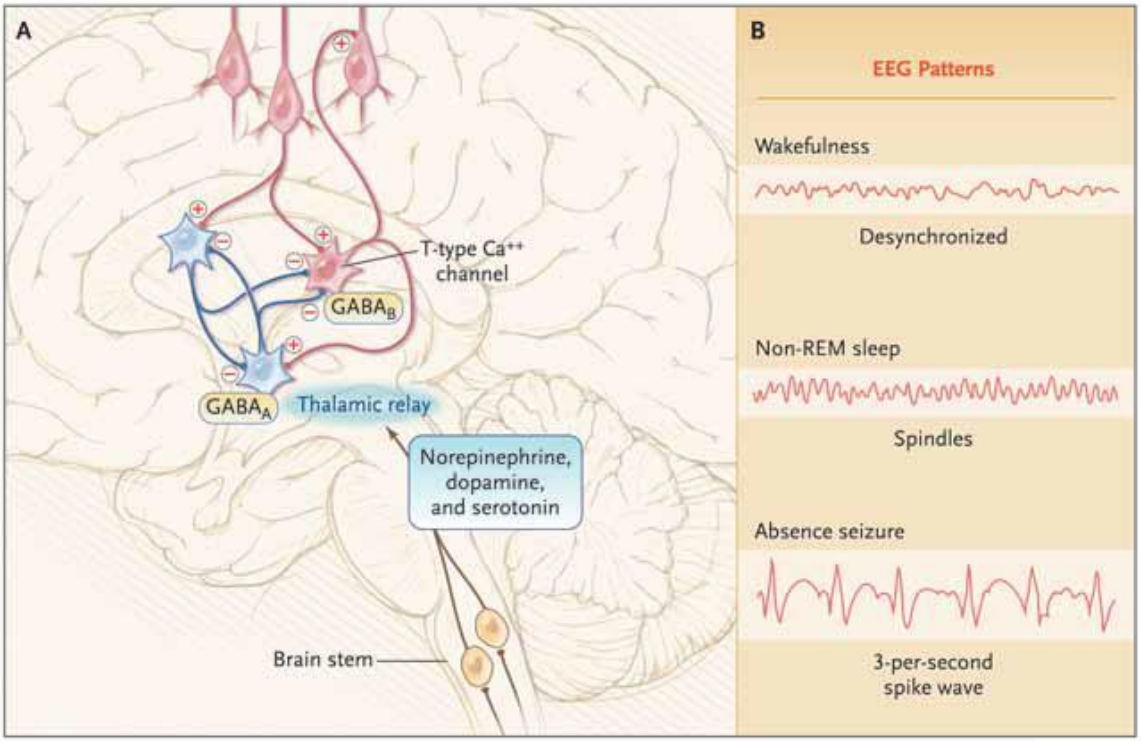

Fig 2.2 : Thalamocortical circuits in generalized epilepsies (Chang *et al*., 2003).

Under normal conditions, brainstem mono aminergic projections synchronize the thalamic reticular neurons into cycles of slow waves and GABA projections of these neurons synchronize thalamocortical relay neurons into cycles of slow waves. Thus resulting in synchronized glutamate thalamocortical projections to widespread cortical areas that generate the electroencephalography recorded slow waves.

In generalized epilepsies, repetitive paroxysmal depolarization shifts (PDSs) in thalamic reticular neurons synchronize PDSs in thalamocortical relay neurons, which in turn synchronize PDSs in cortical neurons, thus generating EEG-recorded spike-wave discharges and sometimes absence seizures (Henry *et al*.,2012). A structurally normal complement of thalamic and cortical neurons and their pathways are able to generate spike-wave discharges and absence seizures due to genetically based dysfunction of channels, receptors, or neurochemical elements. A minobutyric acid (GABA) functions initially as an excitatory

neurotransmitter in the premature infant and a developmental switch changes its role to inhibition closer to term. This switch occurs as a result of the maturation of the cation-chloride co transporter. Subsequently, there is a surge of glutamatergic-related excitatory connections, which in turn lead to a developmental phase during which the immature brain favors excitability (Sharma *et al*., 2017).

2.3 Genetics of epilepsy

A major source of complexity in the study of the genetic epilepsies resides in the frequently reported observation that a single gene can be associated with different syndromes, ranging from very benign epilepsies to severe encephalopathies. Various mechanisms have been blamed for this wide range of severity, including somatic mosaicism, intragenic genomic rearrangements, protein localization of the mutations and loss- versus gain-of-function mutations. For example, somatic mosaicism for SCN1Amutations has often been found in mildly affected parents of children exhibiting Dravet syndrome (Xu *et al*., 2015). Dravet syndrome is a rare genetic epileptic encephalopathy (dysfunction of the brain). It begins in the first year of life in an otherwise healthy infant.

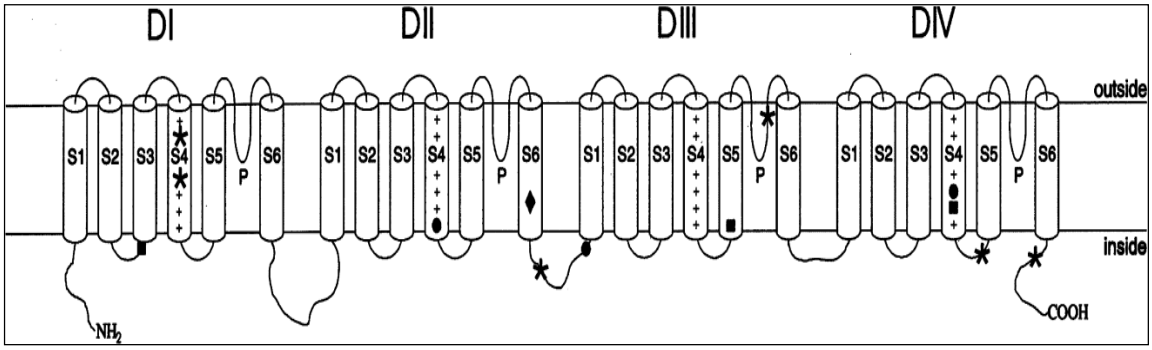

Fig 2.3 : Organization of SCN1A. The neuronal voltage-gated sodium-channel α-subunit *SCN1A* is a monomer and consists of four homologous domains (DI–DIV). Each domain has six transmembrane segments (S1–S6). S4 has several positively charged amino acids and represents the voltage sensor. P = the pore loop, which delineates the pore of the channel. Mutations identified denoted as : asterisks (*) = deletion, insertion and nonsense mutations; diamond (♦) = missense mutation; circles (○) = GEFS+ missense mutations reported by Escayg *et al*., 2000, 2001; and squares (▪) = GEFS+ missense mutations reported by Wallace *et al*., 2001.

The Aristaless-related homeobox gene, *ARX,* is an important transcription factor with a crucial role in forebrain, pancreas and testes development. At least fifty-nine mutations have been described in the ARX gene in seven X-chromosome linked disorders involving mental retardation. Recent studies with *ARX* screening suggest that the gene is mutated in 9.5% of X-linked families with these disorders (Gecz *et al*., 2006).

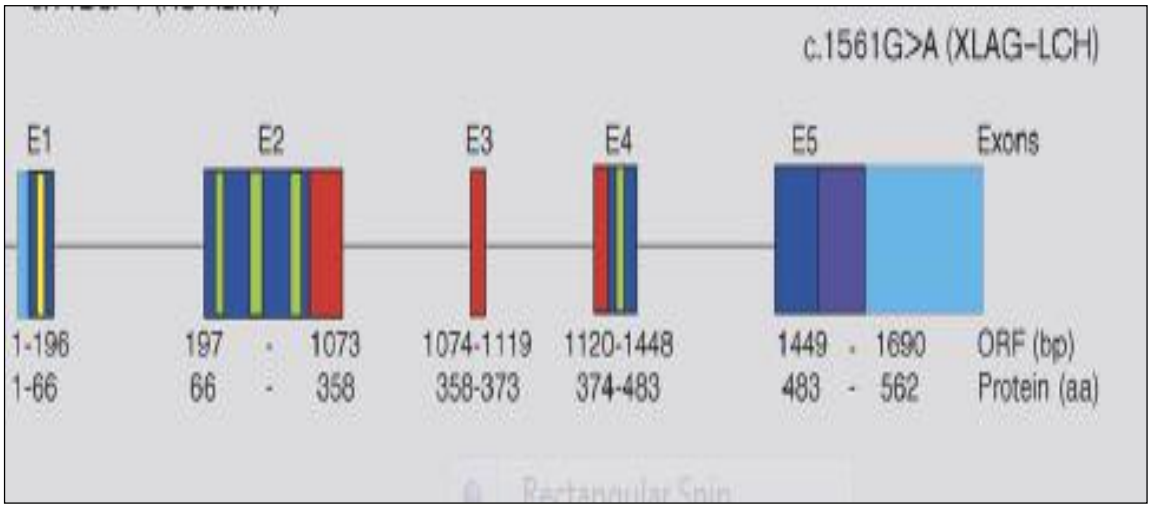

Fig 2.4: Mutations of the ARX gene (Adopted from Current Opinion in Genetics & Development 2006, 16:308–316). Five exons of the ARX gene are boxed together with the open reading frame in dark blue, 50 and 30 untranslated regions in light blue, and ATG start and STOP codon positions. Base pair and corresponding amino acid sequence-spans of individual exons are shown below each exon. Known functional domains are highlighted : octapeptide is shown as a yellow rectangle in exon 1,polyalanine tracts are shown as four green rectangles in exons 2 and 4: homeodomain sequences spanning exons 2, 3 and 4 are shown as red boxes and Aristaless domain is shown in exon 5 as a purple box.

Different insertion and deletion mutations were found which are of yet unknown functional significance to the ARX protein. The mutated variant found to be c.1561, G>A (X-linked lissencephaly with abnormal genitalia (XLAG) -Langerhans Cell Histiocytosis).

2.4 Impact of Next Generation Sequencing (NGS) and Whole Exome Sequencing (WES) on diagnosis and mutation of Epileptic gene

Next generation sequencing, also known as deep sequencing (Behjati and Tarpey, 2013), in simple, is a catch-all term for all those sequencing platforms that are able to perform massively parallel sequencing (up to millions and even billions reads at a time) with an immensely high throughput in gigabase (Gb) or even terabase (Tb) scale (Rizzo and Buck, 2012). Although common epilepsy syndromes, such as genetic generalized epilepsy (GGE), are only rarely monogenic, the overall diagnostic yield of NGS in epilepsy reaches up to 20–30% of all cases (Moller *et al*.,2015). Although the advent of NGS has open the door for sequencing the whole genome of organisms, whole genome sequencing (WGS) is still not feasible due to

the time, cost (Biesecker *et al.*, 2011) and data (Boycott *et al*., 2013) constraints. WGS entails the sequencing of the entire genome with uniform coverage and therefore has a higher chance to detect structural rearrangements and copy number variants. WGS data analysis is usually limited to the 1–2% of the genome that encodes genes. WES is less complex and expensive than WGS since sequencing the coding regions reduces wet-lab, computational and data storage costs. WES is not limited to selected genes, as for gene panels, but includes the coding of almost all the known 20,000 human genes (Mie *et al*., 2017).

The increased knowledge about causative genetic variants has had a major impact on diagnosis of genetic epilepsies and has already been translated into treatment recommendations for a few genes (Moller *et al*., 2015). Next generation sequencing technologies such as targeted gene panels, whole exome sequencing and whole genome sequencing have led to an explosion of gene identifications in monogenic epilepsies including both familial epilepsies and severe epilepsies, often referred to as epileptic encephalopathies (Moller *et al*., 2015). The widening of the phenotypic spectrum associated with a single gene has been remarkable for some genes, including sodium voltage-gated channel alpha subunit 2 (*SCN2A*), potassium voltage-gated channel subfamily Q member 2 *(KCNQ2*), sodium voltage gated channel alpha subunit 8 (*SCN8A*) and DEP domain containing 5 (*DEPDC5*), and has largely been achieved using WES or targeted resequencing of gene panels. For instance the *SCN2A* gene was initially associated with benign familial neonatal-infantile seizures (BFNIS) (Heron *et al*., 2002).

The variant interpretation process is a key step in the NGS analysis indeed, guidelines for interpreting sequence variants have been published (Richard et al.,2015) that provide uniform criteria for classifying a variant as pathogenic, likely pathogenic, of unknown significance (VUS), likely benign or benign. In this perspective, segregation analysis performed in parents or other relatives to determine the carrier status or de novo occurrence of a variant is often helpful to provide further evidence for or against pathogenicity.

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1 Study of the cases and spot of research work

The family which meets the criteria in favor of the terms genetic epilepsy and includes frequent seizures and have abnormal characters especially phenotypic characters and clinical reports were priotised to identify the most comprehensive and effective method for retrieving studies specific to the study objective.

Some of the families were from Evan Eichler Collaboration which created an Expanded CNV morbidity map from 29,085 children with developmental delay in comparison to 19,584 healthy controls, identifying 70 significant CNVs and resequenced 26 candidate genes in 4,716 additional cases with developmental delay and 2,193 controls. An integrated analysis of CNV and single-nucleotide variant data pinpointed 10 familial cases enriched for putative loss of function. Follow-up of a subset of affected individuals identified new clinical subtypes of pediatric disease and the genes responsible for disease-associated CNVs (Eichler, 2014). Over the course of five years, large cohorts were sent to Evan Eichler lab in Seattle in collaboration for membrane interface probe (MIP) screening ID and autism genes.

The genes were Follow-up of these individuals, singleton set. Blood Samples from the members of family 1 to family 10 were collected from a family with at least one member with epilepsy and were from South Australia and samples of family 11 were from Poland, a new and large family-Polish family. The identification of genes was performed in Women's and Children Hospital, Neurogenetics Department, Adelaide.

3.2 Controls

Control were recruited among unrelated visitors and the clinical assessment of controls was limited to answers to general medical questionnaire with specific on any neurological symptoms individuals with a family history positive for any neurological disorder were excluded.

3.3 DNA Quality check and Quantification

Quality of the DNA extracted from the samples was checked by agarose gel electrophoresis. 1 % agarose gel was prepared by dissolving 1 gm of agarose in 100 mL of 1X TBE buffer and heated till boiling. 3µl of Ethidium Bromide (EtBr) from stock solution of 15mg/mL was added to the solution and mixed uniformly. Gel was left to cool down, poured onto a gel tray and was allowed

to set. DNA samples were diluted with nuclease free water (NFW) and stained with 4X gel loading dye in the ratio of 1:8:1, to final volume of 10µl. Samples were loaded into their respective wells and 2µl of 1kb DNA ladder was used as a marker to compare the size of the DNA. Electrophoresis was carried out at a constant voltage of 120V for 30 minutes.

DNA Quantification was done by using Nanodrop ND-1000 (Thermo Scientific, USA). With the sampling arm opened, 1µl of NFW was pipetted onto the lower measurement pedestal. The sampling arm was closed both the lower and upper measurement pedestals were wiped properly by piece of lint free paper. Again 1µl of NFW was loaded and was set as blank. Both the measurement pedestals were wiped out again and 1µl DNA samples were loaded and concentrations of DNA samples were noted down along with the A260/280 ratio.

3.4 Library Preparation

The genomic DNA were made ready for whole exome sequencing. Approximately 100ul of 5ng/µl final concentration of all the samples was prepared with Tris-EDTA buffer. The library preparation of the samples for whole exome sequencing was performed by using manufacturer provided protocol of TruSeq Exome Library Prep Illumina, USA. The brief work flow for whole exome library preparation is shown in below in Fig 3.1.

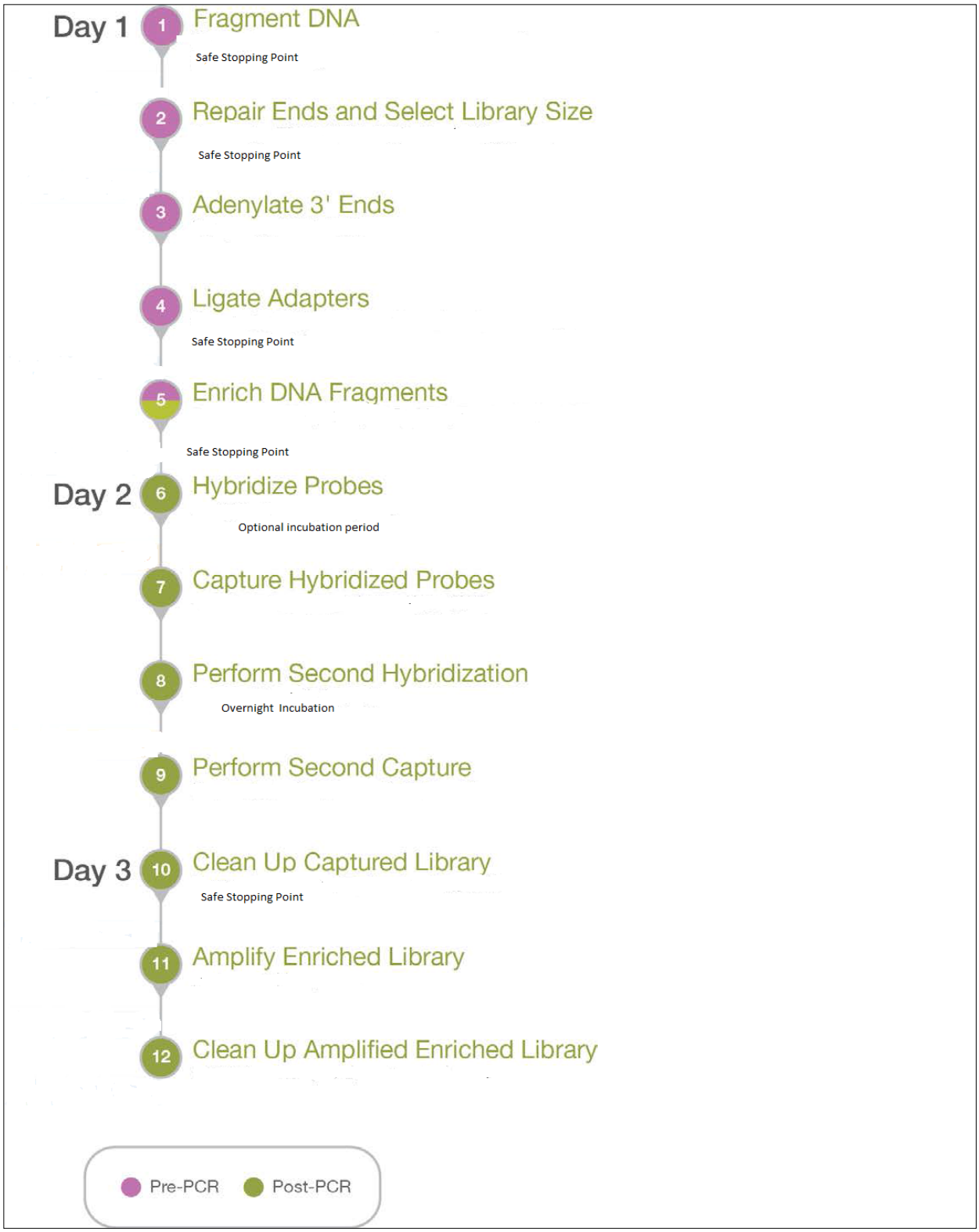

Fig 3.1: Workflow for the preparation of Exome Library.

3.4.1 DNA Fragmentation

RSB (Resuspension Buffer) was thawed at room temperature and stored at 8°C after the initial thaw. SPB (Sample Purification Beads) were brought to room temperature letting to stand for 30 minutes and 80% ethanol was freshly prepared. This process leads to optimally fragment gDNA to a 150 bp insert size. Covaris shearing generates dsDNA fragments with 3' or 5' overhangs and the following steps were followed.

Normalization of gDNA was done taking 100ng of each gDNA samples which were normalized with RSB as shearing buffer premix in a 15 ml conical tube for fragmentation, by mixing 20ul each of 5ng/µl samples with 40ul of RSB. Then the process was followed by fragmentation .The mixture was pipetted to the covaris tubes. Shearing buffer premix was created. Tube was centrifuged briefly. 50µl supernatant was transferred from each Covaris tube to 8-tube strip. Cleanup of fragmented DNA was performed. SPB was vortexed until well-dispersed. 100 µl SPB was added to each well or to the tube, and then mixed thoroughly by pipetting up and down. Incubated at room temperature for 5 minutes and centrifuged briefly. It was placed on a magnetic stand and wait until the liquid is clear (~8 minutes).All supernatant was removed and discarded from the tube. Washed 2 times, 200 µl freshly prepared 80% EtOH was added to each well or to the tube. It was then incubated on the magnetic stand for 30 seconds. Supernatant was removed and discarded from the tube. Again centrifuged briefly and Incubated on the magnetic stand for 30 seconds. Pipette was used to remove residual EtOH from each well or from the tube, pellet was air-dried on the magnetic stand (~5 minutes). RSB was added to the tube. Removed from the magnetic stand, and then mix thoroughly by Pipetting up and down. Then it was incubated at room temperature for 2 minutes. Placed on a magnetic stand and waited until the liquid is clear (2–5 minutes). 60µl of the supernatant was transferred to the new 8-tube strip.

3.4.2 End Repair and Library Size Selection

This process converts the overhangs resulting from fragmentation into blunt ends using ERP3 (End Repair Mix). The 3' to 5' exonuclease activity of this mix removes the 3'overhangs and the 5' to 3' polymerase activity fills in the 5' overhangs. Following end repair, the library size is selected using SPB (Sample Purification Beads). 40 µl ERP3 was added to the tube, and then mix thoroughly by Pipetting up and down. Incubated by Placing on the thermal cycler and run the ERP program (30°C for 45 minutes and hold 4°C). Optimization of Fragment Length of the library was performed as; Sample Purification Beads were vortexed until well-dispersed. 90 µl SPB was added to each well or to the tube, and then mixed thoroughly. It was incubated at room temperature for 5 minutes and centrifuged briefly. Placed on a magnetic stand and wait until the

liquid is clear (2–5 minutes). 185 µl supernatant was transferred to the new 1.5 ml microcentrifuge tube. Sample Purification Beads were vortexed until well-dispersed. 125 µl SPB was added to each well or to the tube, and then mix thoroughly .Incubated at room temperature for 5 minutes and centrifuged briefly. Placed on a magnetic stand and wait until the liquid is clear (2–5 minutes). Supernatant was removed and discarded from the tube. Washed for 2 times as 200 µl freshly prepared 80% EtOH was added to each well or to the tube and incubated on the magnetic stand for 30 seconds. Supernatant was removed and discarded from the tube and centrifuged briefly. Incubated on the magnetic stand for 30 seconds.20 µl pipette was used to remove residual EtOH from the tube. It was air-dried on the magnetic stand until dry about 5 minutes. 20 µl RSB was added to the tube. Removed from the magnetic stand, and then mix thoroughly by Pipetting up and down. Incubated at room temperature for 2 minutes and centrifuged briefly. Placed on a magnetic stand and wait until the liquid is clear (2–5 minutes). 17.5 µl supernatant was transferred to a new 8-tube strip.

3.4.3 Adenylation of 3' ends

A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. After end repair and optimization of library size, a single 'A' nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. 12.5µl of A Tailing Mix (ATL2) to each tube and was mixed thoroughly. The tube was placed on thermal cycler and following program was set and run. The preheat lid option set to 100°C, 37°C for 30 minutes, 70°C for 5 minutes and hold at 4°C. Each tube contains 30µl.

3.4.4 Adapter Ligation

This process ligates multiple indexing adapters to the ends of the DNA fragments, which prepares them for hybridization onto a flow cell. To the tube containing DNA fragments 2.5µl of LIG2 (Ligation Mix 2) and 2.5µl of DNA adapters (DAP) were added and mixed thoroughly. The tube was placed on thermal cycler and following program was set and run. The preheat lid option set to 100°C, 37°C for 30 minutes, 70°C for 5 minutes and hold at 4°C. Each tube contains 37.5 µl. 5µl STL (Stop Ligation Buffer) was added to the tube and mixed thoroughly. After the indexing adapter ligation the ligated fragments are cleaned by using SPB.

42.5µl SPB was added to the tube and mixed thoroughly by pipetting. Further processing was done following the step 2 to 7 of clean up fragmented DNA. 52.5µl RSB was added to the tube

and the tube was removed from the magnetic stand and then mixed thoroughly by pipetting up and down. It was then incubated at room temperature for 2 minutes. Again it was place on a magnetic stand and waited until the liquid was clear (~5 minutes).50µl of the supernatant was transferred to a new 8-tube strip. Again this whole process from 1 to 5 was repeated by using 50µl of SPB and 27.5µl of RSB. Finally 25µl of the supernatant was transferred to new 8-tube strip.

3.4.5 Enrichment of DNA fragments by PCR

This process used PCR to selectively enrich those DNA fragments that had adapter molecules on both ends and to amplify the amount of DNA in the library. PCR was performed with PPC (PCR Primer Cocktail) that anneals to the ends of the adapters.

The tube was placed on ice and 5µl PPC firstly thawed at room temperature. Inverted to mix, then centrifuged at 600 × g for 1 minute. 20 µl EPM (Enhanced PCR Mix) was added tube and mixed thoroughly. The tube was placed on thermal cycler and following program was set and run. The preheat lid option set to 100°C, 98°C for 20 minutes, 60°C for 15 seconds, 72°C for 30 seconds, 72°C for 5 minutes and hold at 4°C. Each tube contains 50µl.

Amplified DNA was then cleaned by using SPB. 35µl SPB was added to the tube and was mixed thoroughly. It was incubated at RT for 5 minutes. The mixture was then placed on a magnetic stand and waited until the liquid was clear (~8 minutes). 82µl supernatant was transferred to the corresponding tube and 82µl SPB was added to each tube. They were mixed thoroughly by pipetting. Further processing was done following the step 2 to 7 of clean up fragmented DNA.

Then, 17.5µl RSB was added to the tube. The tube was removed from the magnetic stand and then mixed thoroughly by pipetting. The tube was incubated at RT for 2 minutes. Again it was place on a magnetic stand and waited until the liquid was clear (~5 minutes).Finally 15µl of the supernatant was transferred to a new 8-tube strip.

3.4.6 Quantify Libraries

The libraries are quantified using the Qubit dsDNA HS Assay Kit and the quality of library is checked by diluting the DNA library 1:10 with RSB and run 1 µl diluted DNA library. **T**he library size distribution was observed on an Agilent Technologies 2100 Bioanalyzer.

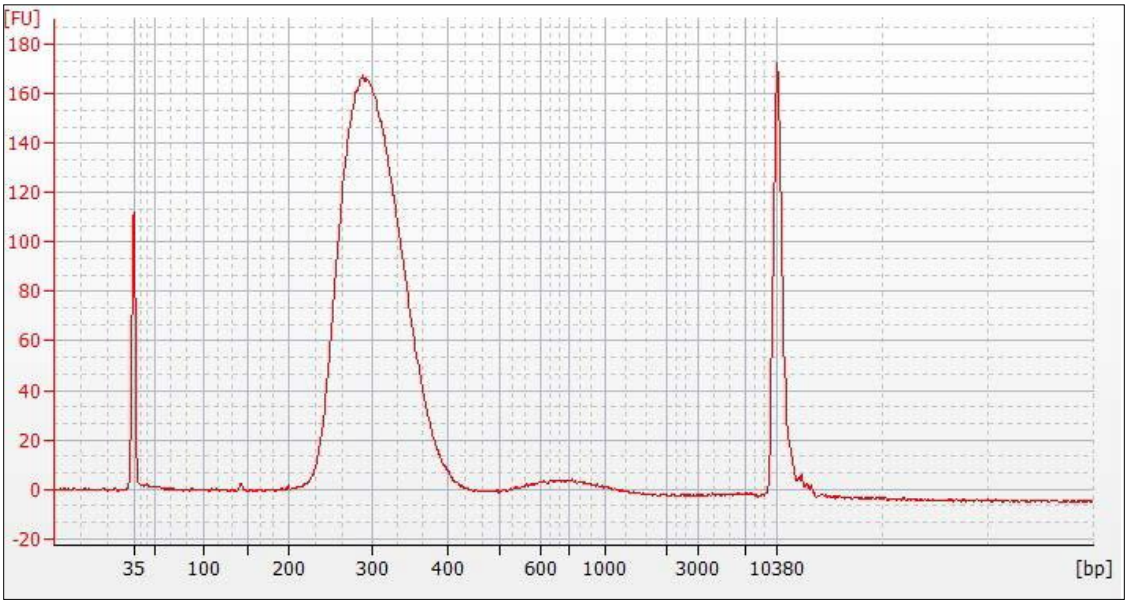

Fig 3.2: 150 bp Insert library distributions on a high sensitivity DNA chip (1:10 Dilution). Some remaining adapter dimer (~150 bp) and a secondary peak, as a concatemer of the library (~550 bp–1000 bp) were observed which seems normal.

3.4.7 Hybridization of Probes

This step combined DNA libraries containing unique indexes into a single pool, and then bond targeted regions of the DNA with capture probes. The libraries were pooled into one pool in such a way that the pool contains 100ng of each library. Final volume of the pool was adjusted to 40µl with Resuspension buffer. The probes were in turn hybridized to the pooled libraries. The following reagents were added in the order listed to a new 8-tube strip and were mixed by pipetting.

- DNA library pool- 40µl
- CT3 (Capture Target Buffer 3) - 50µl
- CEX (Coding Exome Oligos) - 10µl

The tube was placed on thermal cycler and following program was set and run. The preheat lid option set to 100°C, 95°C for 10 minutes. Amplification step of 18 cycles of 1 minute each, starting at 94°C, then decreasing 2°C per cycle and 58°C for 90 minutes. Hold at 58°C for approximately 20 hours. Each tube contains 100µl.

3.4.8 Capture of Hybridized Probes

This step uses SMB (Streptavidin Magnetic Beads) to capture probes hybridized to the targeted regions of interest. Two heated washes remove nonspecific binding from the beads. The enriched library is then eluted from the beads and prepared for a second round of hybridization. The probes hybridized to the targeted regions of the DNA libraries were in turn captured using SMB. The SMB was incubated at RT for approximately 30 minutes before use. 250µl SMB was added to a new 1.5ml micro centrifuge tube (MCT). The total sample volume (~100µl) from the thermal cycler was directly transferred to the MCT containing SMB with the 8-tube strip still on thermal cycler. It was pipetted to mix. The MCT was incubated at RT for 25 minutes with gentle tapping done at an interval of around 7 minutes to ensure that the beads did not get settled at the bottom of the tube. It was centrifuged briefly and was incubated on a magnetic stand until the liquid was clear (2–5 minutes). Supernatant was discarded into new MCT. The tube was removed from the magnetic stand. Nonspecific binding from the beads was removed by two heated washes with SWS (Streptavidin Wash Solution). 200µl SWS was added to the tube and tube was tapped to mix the beads and SWS properly. The tube was placed on the 50°C heat block for 30 minutes with 800 rpm shaking at an interval of 7 minutes. After the heat incubation, the tube was immediately placed on a magnetic stand until the liquid was clear (~2 minutes). Supernatant was discarded. The tube was removed from the magnetic stand. These steps were repeated for a total of 2 washes**.**

The enriched library was then eluted from the beads and prepared for a second round of hybridization. The elution premix was prepared by mixing 28.5µl EE1 (Enrichment Elution Buffer 1) and 1.5µl HP3 (2N NaOH) in a 1.5 ml MCT and then was tapped and vortexed. 23µl elution premix was added to the tube that contains the beads and was tapped to mix. The MCT was incubated at RT for 2 minutes and was given a short spin. Placed on a magnetic stand until the liquid was clear (~5 minutes).21µl supernatant was transferred to a new 8-tube strip. To the tube 4µl ET2 (Elute Target Buffer 2) was added and was pipetted to mix. Short spin was given.

3.4.9 Second Hybridization

This step bound targeted regions of the enriched DNA with capture probes to ensure the high specificity of the captured regions a second time. The 8-tube strip followings were added and mixed by pipetting.

- RSB- 15µl
- CT3 (Capture Target Buffer 3)- 50µl
- CEX (Coding Exome Oligos)- 10µl

Short spin was given to the tube and was place on the thermal cycler with the same PCR program used for first. The tube was kept at the 58°C holding temperature for approximately 24 hours.

3.4.10 Second Capture

The same procedure of the first capture was repeated to ensure the high specificity of the captured regions.

3.4.11 Cleanup of Captured Library

The captured library was again purified using SPB. 45µl of the well-dispersed SPB was added to the tube and pipetted to mix. Further processing was done following the step 2 to 7 of clean up fragmented DNA. The library was then eluted on 27.5µl of RSB. The tube was incubated at RT for 2 minutes. Again it was place on a magnetic stand and waited until the liquid was clear (~5 minutes). Finally 25µl of the supernatant was transferred to a new 8-tube strip.

3.4.12 Amplification of Enriched Library

Enriched library was then amplified by an 8-cycle PCR program using PPC. To the tube 5µl PPC and 20µl NEM (Enrichment Amp Mix) were added and pipetted to mix. A short spin was given to the tube and it was placed on thermal cycler and following program was set and run. The preheat lid option set to 100°C,98°C for 10 seconds, 8 cycles of 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds, 72°C for 5 minutes and hold at 4°C. Each well or tube contains 50µl.

3.4.13 Cleanup of Amplified Enriched Library

Following the amplification using PCR program, the amplified enriched library was purified by using SPB removing the unwanted products. 45µl SPB was added to the tube and pipetted to mix. Further processing was done following the step 2 to 7 of clean up fragmented DNA. The library was then eluted on 22µl of RSB. The tube was incubated at RT for 2 minutes. Again it was place on a magnetic stand and waited until the liquid was clear (~5 minutes). Finally 20µl of the supernatant was transferred to a new 8-tube strip.

3.4.14 Validate enriched Libraries

The libraries were quantified using the Qubit dsDNA HS Assay Kit and the quality of library is checked by diluting the DNA library 1:10 with RSB and run 1 µl diluted DNA library. The following procedures were performed to quantify enriched libraries and check enriched library quality. The post enriched library was quantified using the Qubit dsDNA HS Assay Kit.

3.5 Whole Exome sequencing and analysis of variants

DNA extraction was done from blood using QIAGEN maxi kit as per manufacturers' protocol. DNA was sent away for sequencing. TruSeq Exome Library Prep protocol was followed when preparing libraries for Illumina sequencing systems that require balanced index combinations, followed by enrichment reagents provided in TruSeq Exome Library Prep kits and sequencing was performed with either Illumina, HiSeq machines. Each read was then aligned to the reference genome Ensembl core database and single-nucleotide variants (SNVs) and small indels were identified using Platypus. The variants were analysed using ANNOVAR and variants were uploaded to CADD directly with a request of full annotation data .The variants were priotised on the basis of following points;

a. Focusing on deleterious variants i.e., nonsynonymous SNVs, coding indels, frameshifts.
b. Mode of inheritance- since most of disorders have autosomal dominant as well as autosomal recessive mode of inheritance, homozygous and heterozygous variants as well as hemizygous variants were focused.
c. Rare variants with ExAC allele frequency <0.01 taken from Exome Aggregation Consortium.
d. Predicting and retaining variants with functional effects using in-silico tools like SIFT, Polyphen, CADD score and on the basis of known genes, epilepsy, Intellectual disability or X linked genes.
e. Assessing biological effect of the variant and functional characteristics based on literature reviews and online tools.

3.6 PCR Primer Designing

Amplicons selected for PCR amplification and sequencing were single exons with smaller than 800bp in size. Primers were designed for different genes automatically with prime 3 and primer sequence were available on request as listed in Table 3.1.

Based on clinically diagnosed epileptic patients, primers were designed for 11 genes. After careful analysis of different variant in different genes of eleven families, such mutation inference can be used for diagnosis of epileptic patients related to intellectual disability and X-linked genes simply by the PCR primer design for those eleven epileptic genes. The detail of epileptic gene sequences and PCR primers has been documented in Appendix 2.

Table 3.1: PCR Primers

Gene	Acc. No. NCBI	HGNC	Primer	Sequence	Tm (°C)
PPP2R1A	010320.10	9302	Forward	5’-GCAATCTGCGAAGTGTCTCA-3’	52
			Reverse	5’-ACCGAGTCCTAGGGAGAGGA-3’	52.3
POGZ	046601.1	18801	Forward	5'-TGATCAGGATGAAGGGAGCC-3'	60
			Reverse	5'-TTCGATCATCACTGCTCAGC-3'	61
MAOA	008957.2	6833	Forward	5'-GGCCTTGTGTCTTTTGCAGTT-3'	56.9
			Reverse	5'-GGTCATTATGTGTGGCCAAGG-3'	62.4
DLG2	021375.1	2901	Forward	5'-TGGGGCCTCAGATAGTGAGT-3'	59.66
			Reverse	5'-GGGTTCTGACTCAACTCCCG-3'	60.04
ALG13	016238.1	30881	Forward	5'-TCTTTACCCCTTCTCCTCCCAA-3'	56.9
			Reverse	5'-TAGTCATTGAGAAGCCAATTAGGG-3'	58.3
GRIA1	047078.1	4571	Forward	5'-CAGTTGATGTCCATGCTGCT-3'	56.9
			Reverse	5'-AGCCCTCAGAGCATTAGCAC-3'	55.8
NCOR1	047111.1	7672	Forward	5'- TCAGCCAGGACAAAACTCAG -3'	55.5
			Reverse	5'- CCATGGATTACAAATACTCAGC -3'	56.3
NALCN	053176.1	19082	Forward	5'-TCTAGGTGTGAATGCTTGGGG-3'	66.1
			Reverse	5'-TCAAGACTGTTCTTGCCCTGAT-3'	64.0
DPYSL3	011373.4	3015	Forward	5'-CCTAGGGAGCAAGGCAATGAG-3'	58.6
			Reverse	5'-CCCTCTGCCAGACATTGCTC-3'	58.3
ANK3	020987.4	494	Forward	5'- CGTGACAAAACTGAAAAGCTAA -3'	59.2
			Reverse	5'- TGGGCATGCTCATCTTTGGT -3'	61.9
ATP6V1B2	025853.17	854	Forward	5'-TGCCAGGAAGAGACAGTAGGA -3'	60
			Reverse	5'-GATGGGGAAGAGAGAAGTGG -3'	58.7

3.7 PCR amplification of the target sequence

Reagents of PCR for eleven families were prepared with different ratio for different samples. PCR was performed for the target sequence of different DNA samples. The primer for PCR was designed using PRIMER 3 input online tool. PCR mixture was prepared for all the DNA samples of eleven families and followed to amplify with different PCR conditions listed in Table 3.2.

Table 3.2: PCR conditions for different genes.

Gene	Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	Hold
PPP2R1A	95 °C	95 °C	60°C	72 °C	72 °C	4 °C
	3 min	30 sec	30 sec	30 sec	5 min	
	1 cycle	35 cycles			1 cycle	
POGZ	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
MAOA	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
DLG2	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
ALG13	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
GRIA1	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
NCOR1	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	
NALCN	95 °C	98°C	60 °C	72 °C	72 °C	4 °C
	3 min	20 sec	15 sec	15 sec	1 min	
	1 cycle	35 cycles			1 cycle	
DPYSL3	95 °C	98 °C	60 °C	72 °C	72 °C	4 °C
	3 min	20 sec	15 sec	15 sec	1 min	
	1 cycle	35 cycles			1 cycle	
ANK3	95 °C	98 °C	60 °C	72 °C	72 °C	4 °C
	3 min	20 sec	15 sec	15 sec	1 min	
	1 cycle	35 cycles			1 cycle	
ATP6V1B2	95 °C	95 °C	55 °C	72 °C	72 °C	4 °C
	3 min	45 sec	30 sec	45 sec	5 min	
	1 cycle	35 cycles			1 cycle	

3.8 Preparation of Elution Solution

The reagents were thoroughly mixed and were examined for precipitation. If any reagent has formed a precipitate, was warmed at 55-65^{o}C until the precipitate dissolves and allowed to cool to room temperature before use.

3.9 Preparation of Wash Solution

Wash Solution Concentrate was diluted with 48ml of 100% ethanol. After each use, tightly cap the diluted Wash Solution to prevent the evaporation of ethanol and centrifuged at 12,000 rpm for 30 seconds.

3.10 Observation of PCR products on agarose gel and purification

The PCR products were loaded on a gel and when the gel run was complete, the gel was observed under UV transilluminator for bright bands. The PCR product was cleaned using PCR Clean-up kit as per the manufacturer's instruction with slight modification. A Gen Elute plasmid mini spin column was inserted into a provided collection tube. 0.5 ml of the Column Preparation Solution was added to each mini spin column and centrifuged at 12,000 rpm for 1 minute. The elute was discarded. 5 volumes of Binding Solution to 1 volume of the PCR reaction was added and mixed. The solution was transferred into the binding column. The Column was centrifuged at maximum speed 12,000 rpm for 1minute.The elutes was discarded, but the collection tube was retained back. The binding column was replaced into the collection tube. 0.5 ml of diluted Wash Solution was used to the column and centrifuged at maximum speed for 1 minute. The elute was discarded but retained in the collection tube. The column was replaced into the collection tube. The column was centrifuged at maximum speed for 2minutes, without any additional wash solution, to remove excess ethanol. Any residual elute was removed as well as the collection tube.

The column was transferred to a fresh 2ml collection tube. 50µL of Elution Solution was put to the center of each column. Incubate at room temperature for 1 minute. To elute the DNA, the column was centrifuged at maximum speed for 1 minute. The PCR amplification product was present in elute and ready for immediate use or storage at -20^{o}c.

3.11 Sanger sequencing

DNA sequencing was carried out by the di-deoxy chain termination method of Sanger (Sanger *et al*., 1977) modified by Chen and Seeburg (1985). A high-throughput, parallel Sanger sequencing was designed to move multiple samples through the amplification, sequencing and subsequent mutational analysis. Sequencing was performed with Big Dye terminator cycle sequencing kit (Applied Bio System) provided the reaction mixture containing all the dNTP's and fluorescently

labeled four dideoxynucleotides along with Taq DNA polymerase in the sequencing buffer. Primers were designed for different genes automatically with prime 3 and primer sequences were listed in Table 3.3. The condition for sequencing the genes is listed in Table 3.4.

Table 3.3 : Primers used for sequencing.

Gene	Primer	Sequence	Tm (°C)
PPP2R1A	Reverse primer	5'-ACCGAGTCCTAGGGAGAGGA-3'	52
POGZ	Forward primer	5'-TTCGATCATCACTGCTCAGC-3'	60
MAOA	Forward primer	5'-GGCCTTGTGTCTTTTGCAGTT-3'	59.86
DLG2	Forward primer	5'-TGGGGCCTCAGATAGTGAGT-3'	59.66
ALG13	Reverse primer	5'-TAGTCATTGAGAAGCCAATTAGGG-3'	58.3
GRIA1	Reverse primer	5'-AGCCCTCAGAGCATTAGCAC-3'	55.8
NCOR1	Reverse primer	5'-TCAGCCAGGGACAAAACTCAG-3'	56.3
NALCN	Reverse primer	5'- TCAAGACTGTTCTTGCCCTGAT -3'	64
DPYSL3	Reverse primer	5'-CCCTCTGCCAGACATTGCTC-3'	58.3
ANK3	Reverse primer	5'-TGGGCATGCTCATCTTTGGT-3'	61.9
ATP6V1B2	Reverse primer	5'- GATGGGGAAGAGAGAAGTGG -3'	58.7

Cycle sequencing was carried out by first giving a rapid ramp to 95^{o}C and holding at this temperature for 10 seconds. This step was followed by rapid thermal ramp to 50^{o}C and holding for 5 seconds. Last step was rapid thermal ramp to 60^{o}C for 3 minutes and 30 seconds. The cycle consisting of the above three steps was repeated 35 times. At last final extension was done at 72^{o}C for 5 minutes followed by rapid temperature drop to 4^{o}C. The unincorporated dye terminators were removed by ethanol precipitation.

Table 3.4: Conditions for cycle sequencing for different genes.

Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	Hold
95 °C, 3 min	95 °C, 10 sec	50 °C, 5 sec	60 °C, 3min 30 sec	72 °C, 5 min	4 °C
	35 cycles				

3.12 Sequencing reaction and clean up

75% of isopropanol was added with 80 µl volume to each tube and vortexed for few seconds. The tubes were kept in dark for 15 minutes and centrifuged in 1300 rpm for 25minutes. The Supernatant was discarded. 200µl of 75% isopropanol was added and centrifuged in 1300rpm for 15 minutes. The same process was followed again. Supernatant was discarded and rotated in speedy vacuum for 10 min to obtain dry pellet. Dry tubes were kept ready for sequencing analysis.

3.13 Sequence analysis

After dry pellet is obtained, 10 µl of formaldehyde is mixed to each Eppendorf tube and vortexed for few seconds. The tubes were kept in dry bath incubator for 5 minutes in 98□C and quickly kept in ice bath for 5 minutes. 15 µl of sample from each Eppendorf tube is then transferred to 96 well plates and covered. The plate is kept in contact of 3130 and 3100 Avant Capillary Array 50 cm in 3130 genetic analyzer where samples are picked up by electric current by four needles laser produces florescent and emits different colored bands. First five bands were picked for sequencing analysis and then the remaining bands were picked continuously. The in-silico tools were used to assemble the nucleotide sequences obtained so far.

3.14 Permutation Bioinformatic Analysis

To study rare variants in families, mutational changes were observed and the segregation was analyzed using bioinformatics tools and literatures. The Map files were sorted using Samtools and duplicates were removed to avoid false variants marked by Picard. The variants were called using Platypus to give a cohort of variants in variant call format and annotated using ANNOVAR. The output file of the sequencer was in .fastq format and contained the raw sequencing reads. The trimmed sequences reads were aligned to human reference genome (hg19/GRCh37). The output SAM (Sequence Alignment/Map) file was converted to a compressed BAM (Binary Alignment/Map) file format and its sorting was done according to its position in the reference genome using Samtools.

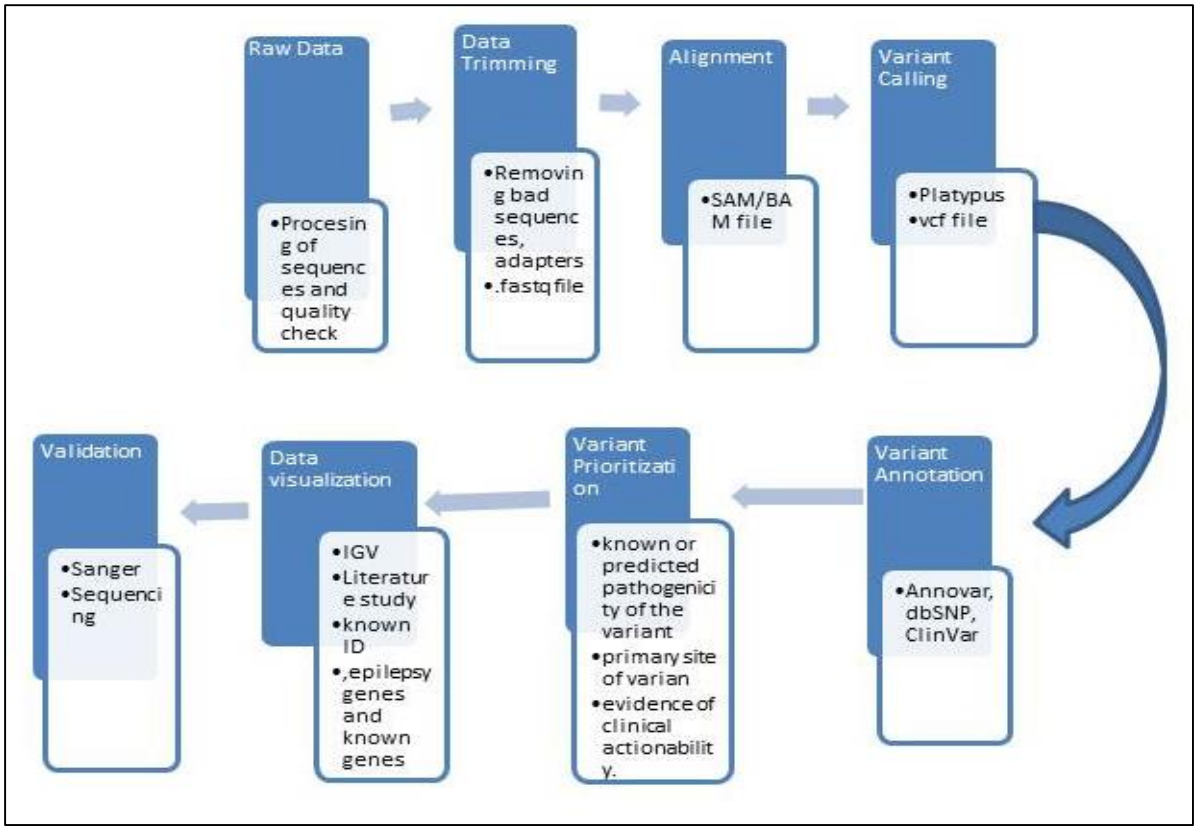

Fig 3.4: Bioinformatic analysis of sequenced data showing a high-throughput variant prioritization pipeline to annotate and prioritize variants, particularly focusing mutations.

The PCR duplicates were marked by Picard. Variants were called using Platypus to give an output variant file in variant call format, prioritized and selected. The putative variants were screened, visualized in IGV and capillary sequenced to verify if the mutation is real and predict genes for mutational analysis.

3.15 Variant Prioritization

Variants mostly likely to be protein-disrupting were Sanger validated nonsense, frameshift, or splice and missense variants with a CADD score > 30, read depth > 8X, a quality score > 20 and an allele count of < 3 to get ultra-rare and private deleterious events. Among the variants obtained deleterious variants affecting protein sequences were prioritized.

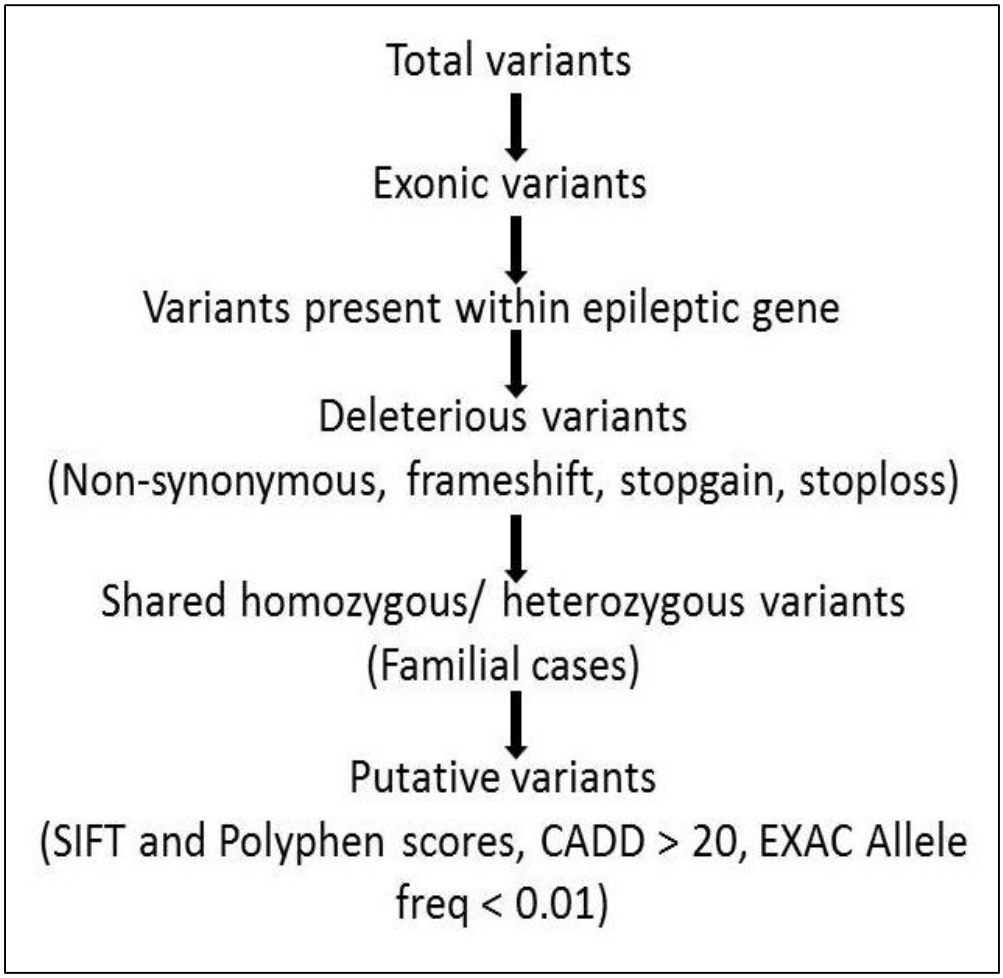

Fig 3.5: Variant prioritization strategy. Exonic variants were targeted having high chances of mutations priotising deleterious variants to validate putative variant for the identification of most probable gene with mutation.

CHAPTER 4

RESULTS

Among the different 11 families, after WES prioritized variants were Sanger sequenced and the segregation of variants were analyzed to determine novel gene and mutation analysis.

4.1 Analysis of segregation of variants in Family 1

4.1.1 Clinical presentation

This family consists of a male patient with complications regarding mental retardation and a disorder characterized by mental retardation, a disorder characterized by significantly below average general intellectual functioning associated with impairments in adaptive behavior.The sample of family members were encoded with certain DNA numbers as 21298(daughter), 22556(mother)and 33170(father).

4.1.2 Quantification of DNA and PCR amplification

Family consisted of three DNA samples, out of which one sample could not be amplified. The concentration of the extracted DNA samples 21298(daughter), 22556(mother) and 33170(father) was obtained 15ng/µl, 9.44ng/µl and 1.65ng/ µl respectively which was measured by using Nanodrop. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR was 5'-GCAATCTGCGAAGTGTCTCA-3'as forward primer & 5'-ACCGAGTCCTAGGGAGAGGA-3' as reverse primer.

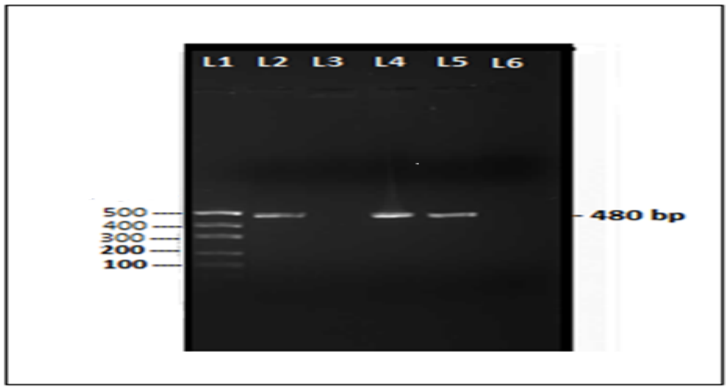

Fig 4.1: PCR amplification of genetic region including the variant of family 1. L1– 500 bp DNA ladder (puc19). L2 to L5– PCR amplicon where DNA extracted from all samples were with good yields except sample in L3 which suggests it is a contaminated sample, L5-Blood bank-positive control and L6 as negative control. Product size of the amplified region was 480 bp.

4.1.3 Prioritization of variants and validation by capillary sequencing

After the targeted capillary sequencing, it was clear that the variant was in heterozygous form in 33170(father) and was found homozygous in rest. Only 33170 had the G>A variant and found heterozygous with both WTG and mutant allele. This proves our exome sequencing results with the identification of gene which tend to be real.

From the findings of WES;

Chromosome number	Position	Gene list	Id	Ref
19	52715956	PPP2R1A	33170	G>A

After the validation of variant, it was clear that complications regarding mental retardation and mental retardation, a disorder characterized by significantly below average general intellectual functioning is caused due to the mutation in the gene Protein Phosphatase (formerly 2A), Regulatory (Subunit PR 65) Alpha isomer (PPP2R1A). . This gene encodes an alpha isoform of the constant regulatory subunit A. Alternatively spliced transcript variants have been described (Provided by RefSeq, Apr 2010). This is a Protein Coding gene. Diseases associated with *PPP2R1A* include mental retardation, Autosomal dominant 26. The variant is found to be heterozygous in father who resulted with mental retardation and seizures. Daughter is not affected whereas sample of mother was found to be contaminated so couldnot be sequenced.

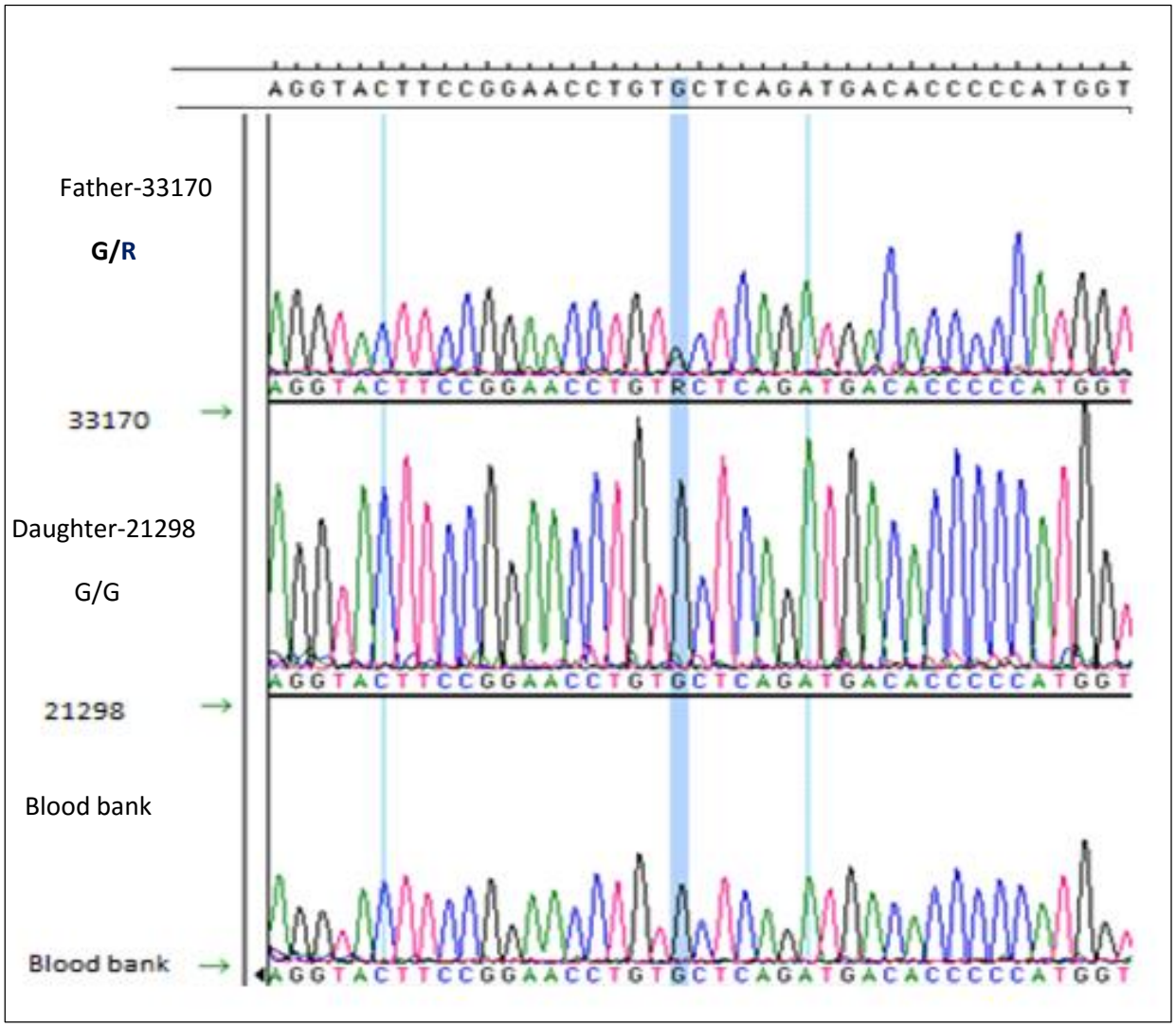

Fig 4.2: Chromatogram derived from targeted capillary sequencing for family 1 testing segregation of the G>A variant. The variant is found to be heterozygous in patient with DNA number 33170(father). The shown fragments were amplified by using reverse primer. The nucleotide code R in DNA number 33170 encodes the base change G> A(R-A or G).

4.2 Analysis of segregation of variants in Family 2

4.2.1 Clinical presentation

The family members were found likely with affected multiple transcripts and lead to mild learning problems and microcephaly. The family in our WES sample set consists of two siblings, brother and sister and their parents. This is a family with what looks like two separate phenotypes segregating separately but in dominant way. This family was analyzed in two separate ways, to elucidate the inheritance coming from mom and the predicted different inheritance coming from

dad. The predicted disorders were especially, Nooonas which is characterized by mildly unusual facial features, short stature, heart defects, bleeding and affect different parts of body and intellectual disability with microcephaly. This was the family with four members- mother (45269), father (45294) and their sibling's daughter (45295) and son (45268) with their respective DNA numbers.

4.2.2 PCR amplification and quantification of DNA

Family consisted of four DNA samples, results in good yield and quality. The concentration of the extracted DNA samples was measured by using Nanodrop and were DNA quantified as 45268(son)-312.85ng/ µl, 45269(mother)-378.32ng/ µl, 45294 (father)-254.56 ng/ µl and 45295(daughter)-140.92ng/µl. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region encompassing the variant c.G2833A:p.E945K present in exon 17. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR were 5'-TGATCAGGATGAAGGGAGCC-3' as forward primer and 5'-TTCGATCATCACTGCTCAGC-3' as reverse primer.

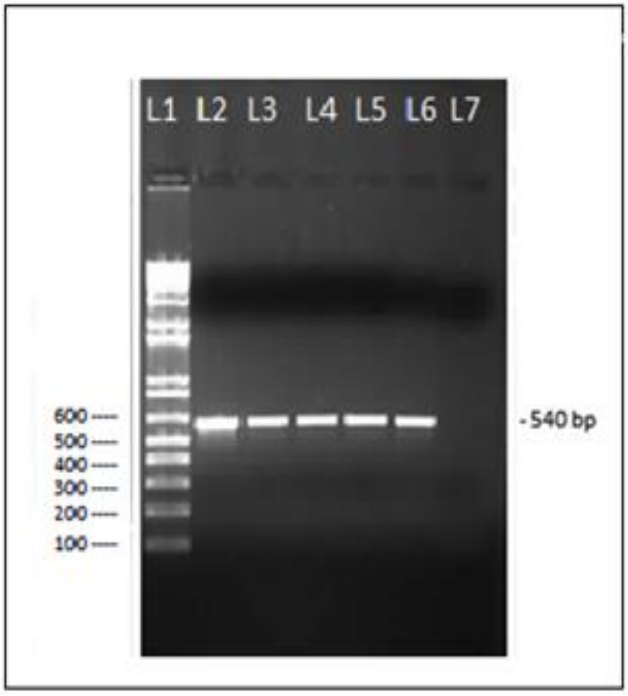

Fig 4.3: Family 2, PCR amplification including c.C2833T:p.E945K present on exon 17 of POGZ. L1– 1000 bp DNA ladder, L2 to L6– PCR amplicon where DNA extracted from all samples were with good yield obtained by Phusion High Fidelity Taq polymerase, L6-Blood bank-positive control and L7-negative control. Product size of the amplified region was 540bp.

4.2.3 Screening of variants and analysis of variants on the basis of functional candidates, CADD, Intellectual disability(ID), X-linked and known gene.

The sorting of variants were done deleterious variations, Known ID genes, high CADD score, SIFT and Polyphen scores as well as EXAC allele frequency(<0.01). Moreover, variants were prioritized based on pathogenicity, clinical phenotype, familial segregation and rare allele frequency. The variants shared with father are listed in Table 4.1.

Table 4.1: Variants shared with father (45294) based on functional character, high CADD and ID gene. After the validation of variant, the listed variants were priotised mainly on the basis of functional basis but found only shared with father.

Chr	Start	End	Ref	Obs	Gene.gene	Exonic Functional gene
chr1	43738471	43738471	G	A	*TMEM125*	Stopgain
chr1	44804892	44804892	T	C	*ERI3*	nonsynonymous SNV
chr1	78429963	78429963	C	T	*FUBP1*	nonsynonymous SNV
chr2	27447260	27447260	C	T	*CAD*	nonsynonymous SNV
chr2	43458467	43458467	G	A	*THADA*	synonymous SNV
chr2	74166070	74166070	C	G	*DGUOK*	nonsynonymous SNV
chr2	131897817	131897817	G	A	*PLEKHB2*	nonsynonymous SNV
chr3	25672317	25672317	G	A	*TOP2B*	synonymous SNV
chr3	38316563	38316563	G	A	*SLC22A13*	nonsynonymous SNV
chr3	100515254	100515254	A	G	*ABI3BP*	synonymous SNV
chr3	113376117	113376119	TGC	-	*USF3*	nonframeshift deletion
chr3	127298941	127298941	G	A	*TPRA1*	nonsynonymous SNV
chr6	135611646	135611646	T	C	*AHI1*	nonsynonymous SNV
chr6	139237099	139237099	G	A	*REPS1*	nonsynonymous SNV
chr7	57528609	57528613	TGTTT	-	*ZNF716*	frameshift deletion
chr7	57528613	57528613	TA	AAC	*ZNF716*	nonframeshift insertion
chr7	103003814	103003814	A	G	*PSMC2*	synonymous SNV
chr7	151097241	151097241	C	T	*WDR86*	nonsynonymous SNV
chr8	110432793	110432793	C	G	*PKHD1L1*	nonsynonymous SNV
chr9	116131994	116131994	G	C	*BSPRY*	nonsynonymous SNV
chr9	117188529	117188529	G	A	*DFNB31*	synonymous SNV
chr9	135205647	135205647	T	C	*SETX*	synonymous SNV
chr10	395358	395358	G	A	*DIP2C*	nonsynonymous SNV
chr10	47668689	47668689	G	A	*ANTXRL*	synonymous SNV
chr10	86008759	86008759	T	C	*RGR*	synonymous SNV

chr10	97919617	97919617	A	T	*ZNF518A*	Unknown
chr11	4388695	4388695	C	T	*OR52B4*	synonymous SNV
chr11	28045363	28045363	T	G	*KIF18A*	nonsynonymous SNV
chr11	55798102	55798102	G	A	*OR5AS1*	nonsynonymous SNV
chr11	77618847	77618847	T	C	*INTS4*	synonymous SNV
chr12	104152932	104152932	G	A	*STAB2*	nonsynonymous SNV
chr14	58915164	58915164	T	C	*KIAA0586*	nonsynonymous SNV
chr14	92537353	92537353	A	G	*ATXN3*	frameshift insertion
chr14	104559037	104559037	C	T	*ASPG*	synonymous SNV
chr17	18144178	18144178	G	A	*LLGL1*	synonymous SNV
chr17	46052996	46052996	G	A	*CDK5RAP3*	synonymous SNV
chr17	61558936	61558936	G	T	*ACE*	nonsynonymous SNV
chr17	61611276	61611276	G	A	*KCNH6*	synonymous SNV
chr19	7963770	7963770	C	T	*LRRC8E*	synonymous SNV
chr19	15784436	15784436	C	T	*CYP4F12*	nonsynonymous SNV
chr19	16910775	16910775	G	T	*NWD1*	nonsynonymous SNV
chr19	19377305	19377305	G	A	*TM6SF2*	synonymous SNV
chr19	36002393	36002393	TGCTG	GCCACTGCTGCT	*DMKN*	nonframeshift insertion
chr19	41822446	41822446	C	T	*CCDC97*	synonymous SNV
chr19	41916721	41916721	C	T	*BCKDHA*	synonymous SNV
chr19	44778605	44778605	A	T	*ZNF233*	nonsynonymous SNV
chr19	49107042	49107042	G	C	*FAM83E*	synonymous SNV
chr20	61468494	61468494	C	T	*COL9A3*	nonsynonymous SNV
chr22	19344558	19344558	C	T	*HIRA*	nonsynonymous SNV
chr22	31658200	31658200	C	T	*LIMK2*	nonsynonymous SNV

4.2.3.1 Validation of variants on the basis of known ID genes and on the basis of high CADD and functional candidates shared with father.

Out of 51 variants detected where variants were shared with father, some of the variants were likely to be putative variants on the basis of known genes and high CADD and functional candidates which are listed below.

On the basis of known ID genes (variants shared with father-45294)

After the information were obtained from screening of variants, variants shared with father (45294) was prioritized on the basis of ID genes. Focusing on deleterious variants i.e.,

nonsynonymous SNVs, coding indels, frameshifts. Moreover Rare variants with ExAC allele frequency <0.01 taken from Exome Aggregation Consortium were preferred.

We found AHI1 (Abelson Helper Integration Site 1with Nonsynonymous SNV, low frequency SNP (rs199578341),CADD score of19.1. The mutation was found heterozygous in father and the rest affected with homozygous mutation.*AHI1* was found associated with Joubert syndrome-3 (*JBTS3*) which didnot matched with phenotypic characters of affected members. In case of *ATXN3*(Ataxin-3),exonic function was Frameshift insertion and followed by Progressive neurologic disorder characterized principally by ataxia, spasticity, and ocular movement abnormalities. *MYH3* (Myosin Heavy Chain 3) had Slice variant, NM_002470:exon19:c.2048-2A>G with CADD score 25.4 and no polyphen. Heterozygous mutation in MYH3 seems associated with distal arthrogryposis type 2A, 2B and 8 – arthrogryposis, a joint contracture disease, with these subtypes having various other associated phenotypes. *ACE* (Angiotensin-converting enzyme) was found to be Nonsynonymous SNV, variant in p.A319S, low frequency SNP (rs34126458) with CADD score 23.4. Various inheritance patterns were associated with quite different phenotypes.

On the basis of high CADD and functional candidates (variants shared with father-45294)

After the Visualization of the variants in Integrative Genomics Viewer (IGV),some of the putative variants were shared with father with DNA number 45294and sorted out if those are real and putative on the basis of ID genes, Focusing on deleterious variants i.e., nonsynonymous SNVs, coding indels and frameshifts. Moreover Rare variants with ExAC allele frequency <0.01 taken from Exome Aggregation Consortium were preferred.*COL9A3* (Collagen Type IX Alpha 3 Chain) was found to be Nonsynonymous SNV, p.R555W characterised with low frequency SNP (rs201783992) and CADD score 31. Variants were found associated with epiphyseal dysplasia. *FUBP1* (Far upstream element-binding protein 1) was found to be Nonsynonymous SNV characterized with high CADD score 34 and heterozygous in father and affected members. It was revealed, Far upstream element binding protein 1 aberrant expression of this gene has been found in malignant tissues, and this gene seemed to be important to neural system and lung development, quite broad expression moreover predominant in bone marrow. *TMEM125* (Transmembrane Protein 125) was found to be stopgain encompassing p.W26X in genetic region with high CADD score 37 .It seems to have quite broad expression, not represented in brain atlas but highly expressed in brain. In *REPS1* (RALBP1 Associated Eps Domain Containing 1), Nonsynonymous SNV with low frequency SNP (rs373034841) and CADD 24.4 was found.RALBP1 associated Eps domain containing 1 encoded a signaling adaptor interacts with proteins that participate in signaling, endocytosis and cytoskeletal changes. *DIP2C* (Disco Interacting Protein 2

Homolog C) was characterised with Nonsynonymous SNV, p.H1008Y, low frequency SNP (EXAC) and CADD 26.5.

4.2.3.2 Validation of variants on the basis of x linked genes high CADD and functional candidates shared with mother (45269)

On the basis of X Linked Genes

Variants were prioritized based on pathogenicity, clinical phenotype, familial segregation and rare allele frequency. The Clinical phenotype shared with mother includes Microcephaly and Learning Problems. Clinical phenotype which were inheritated with mother were mild learning problems and microcephaly as well as with sister were mild learning problems, microcephaly, mild Noonan's and brother was inheritated with severe learning problems, microcephaly, severe Noonan's.

The clinical phenotype resembeled to mother and predicted the variant may be putative within the shared variants from mother which happened to be true. The variants shared with mother are listed in Table 4.2.

Table 4.2: Variants shared with mother (45269) based on high CADD and X-linked functional candidate. After the validation of variant, the listed variants were priotised mainly on the basis of functional basis but only shared with mother in which *POGZ* was found to be real candidate.

Chr	Start	End	Ref	Obs	Gene.gene	Exonic Functional gene
chr1	1025751	1025751	C	T	*C1orf159*	nonsynonymous SNV
chr1	9416073	9416073	G	A	*SPSB1*	synonymous SNV
chr1	**151378393**	**151378393**	**C**	**T**	***POGZ***	**nonsynonymous SNV**
chr1	159778905	159778905	C	T	*FCRL6*	synonymous SNV
chr1	204412618	204412618	C	T	*PIK3C2B*	nonsynonymous SNV
chr1	207642235	207642235	G	A	*CR2*	nonsynonymous SNV
chr1	224872519	224872519	G	C	*CNIH3*	nonsynonymous SNV
chr1	228589812	228589812	C	T	*TRIM11*	synonymous SNV
chr2	21236337	21236337	C	A	*APOB*	nonsynonymous SNV
chr2	26533673	26533673	G	T	*ADGRF3*	nonsynonymous SNV
chr2	73659410	73659410	T	C	*ALMS1*	synonymous SNV
chr2	148731033	148731033	C	T	*ORC4*	nonsynonymous SNV
chr2	172823428	172823428	C	T	*HAT1*	nonsynonymous SNV

chr2	236626225	236626225	G	A	*AGAP1*	nonsynonymous SNV
chr2	237395520	237395520	A	T	*IQCA1*	synonymous SNV
chr2	241066272	241066272	A	G	*MYEOV2*	nonsynonymous SNV
chr2	242380856	242380856	C	T	*FARP2*	synonymous SNV
chr3	3178946	3178946	T	A	*TRNT1*	nonsynonymous SNV
chr3	77542460	77542460	C	T	*ROBO2*	nonsynonymous SNV
chr3	112546521	112546521	T	C	*CD200R1L*	synonymous SNV
chr3	113172517	113172517	C	T	*SPICE1*	synonymous SNV
chr3	113376117	113376119	TGC	-	*USF3*	nonframeshift deletion
chr3	113376119	113376119	TAG	TGC	*USF3*	nonframeshift insertion
chr4	2214853	2214853	G	A	*POLN*	nonsynonymous SNV
chr4	5664870	5664870	G	A	*EVC2*	nonsynonymous SNV
chr5	74916221	74916221	G	A	*ANKDD1B*	nonsynonymous SNV
chr5	140782951	140782951	A	C	*PCDHGA9*	nonsynonymous SNV
chr5	140797930	140797930	C	T	*PCDHGB7*	synonymous SNV
chr6	87966140	87966140	A	G	*ZNF292*	synonymous SNV
chr6	146755269	146755269	T	C	*GRM1*	synonymous SNV
chr6	151790102	151790102	G	T	*ARMT1*	nonsynonymous SNV
chr6	154439836	154439836	G	A	*OPRM1*	nonsynonymous SNV
chr7	100334608	100334608	C	T	*ZAN*	Unknown
chr7	120979156	120979156	C	T	*WNT16*	synonymous SNV
chr7	150839330	150839330	C	T	*AGAP3*	synonymous SNV
chr8	24775639	24775639	G	A	*NEFM*	synonymous SNV
chr9	116060043	116060043	G	A	*RNF183*	nonsynonymous SNV
chr10	27687437	27687437	C	A	*PTCHD3*	nonsynonymous SNV
chr10	102003456	102003456	T	C	*CWF19L1*	nonsynonymous SNV
chr10	112572446	112572446	C	A	*RBM20*	nonsynonymous SNV
chr11	47829973	47829973	C	T	*NUP160*	nonsynonymous SNV
chr11	62677200	62677200	C	T	*CHRM1*	nonsynonymous SNV
chr11	93463070	93463070	A	T	*CEP295*	nonsynonymous SNV
chr11	94224032	94224032	G	A	*MRE11A*	synonymous SNV
chr11	123624766	123624766	A	T	*OR6X1*	nonsynonymous SNV
chr12	52755268	52755268	A	G	*KRT85*	nonsynonymous SNV
chr12	70723216	70723216	T	A	*CNOT2*	synonymous SNV

chr12	80632673	80632673	C	T	*OTOGL*	nonsynonymous SNV
chr12	98925548	98925548	C	A	*TMPO*	nonsynonymous SNV
chr12	122729198	122729198	T	C	*VPS33A*	nonsynonymous SNV
chr12	124330324	124330324	G	A	*DNAH10*	nonsynonymous SNV
chr12	124840081	124840081	C	T	*NCOR2*	nonsynonymous SNV
chr12	131616321	131616321	G	A	*ADGRD1*	nonsynonymous SNV
chr13	72440182	72440182	C	G	*DACH1*	synonymous SNV
chr14	24878988	24878988	C	T	*NYNRIN*	nonsynonymous SNV
chr14	81610687	81610687	C	T	*TSHR*	nonsynonymous SNV
chr14	92537353	92537353	A	G	*ATXN3*	frameshift insertion
chr14	105693053	105693053	T	C	*BRF1*	nonsynonymous SNV
chr15	58254315	58254315	G	A	*ALDH1A2*	synonymous SNV
chr15	81172070	81172070	T	C	*CEMIP*	synonymous SNV
chr16	2051096	2051096	G	A	*ZNF598*	synonymous SNV
chr16	67354737	67354737	C	T	*KCTD19*	nonsynonymous SNV
chr17	1538499	1538499	C	A	*SCARF1*	nonsynonymous SNV
chr17	40342881	40342881	T	G	*GHDC*	nonsynonymous SNV
chr17	72764631	72764631	A	C	*SLC9A3R1*	nonsynonymous SNV
chr17	76446773	76446773	G	C	*DNAH17*	nonsynonymous SNV
chr17	76450691	76450691	G	A	*DNAH17*	nonsynonymous SNV
chr17	78210809	78210810	AG	-	*SLC26A11*	frameshift deletion
chr19	9066029	9066029	G	A	*MUC16*	synonymous SNV
chr20	56964537	56964537	C	T	*VAPB*	synonymous SNV
chr20	61513496	61513496	G	A	*DIDO1*	nonsynonymous SNV
chr22	50957657	50957657	G	A	*NCAPH2*	nonsynonymous SNV
chrX	150817170	150817170	C	T	*PASD1*	nonsynonymous SNV

Out of 76 variants detected where variants were shared with mother, some variants were likely to be putative variants on the basis of known genes, X linked genes and functional candidates which are listed below.

On the basis of X LINKED GENES (variants shared with mother-45269)

After the screening of the variants some of the putative variants were shared with mother and sorted out if those are real and putative on the basis of X linked genes. *PASD1 (*PAS Domain Containing 1) was found to be Nonsynonymous SNV, p.S238L characterised by low frequency

SNP (rs146310228) and CADD score 11.41. PAS domain containing 1 which is thought to function as a transcription factor and looks to be expressed in bone and testis only (unigene) so did not matched with the expression of related disorders.

On the basis of Known ID Genes (variants shared with mother-45269)

Some of the putative variants were shared with mother and sorted out if those are real and putative on the basis of ID genes**.** *TSHR* (Thyroid Stimulating Hormone Receptor)was found to be Nonsynonymous SNV encompassing the variant p.T762M within genetic region with CADD score 25.3.There were a number of mutations identified and a variety of clinical phenotypes associated with variants in *TSHR*, all associated with hypothyroidism. Homozygous mutation in a female related with hypothyroidism and ID. *ATXN3* (Ataxin 3) was found to be with Frameshift insertion and repeated expansion was seen. *EVC2* (EvC Ciliary Complex Subunit 2) was found to be Nonsynonymous SNV with low frequency SNP (EXAC)and CADD score 23.EVC associated with Ellis-van Creveld syndrome was with heterozygous mutation associated with Weyers acrofacial dysostosis, overlapping skeletal dysplasia disorders with other associated features characterised by Autosomal dominant condition with dental anomalies, nail dystrophy, postaxial polydactyly and mild short stature. *APOB* (Apolipoprotein B) was found to be Nonsynonymous SNV, p.G1304V, low frequency SNP (rs142798172) with high CADD score 31.Heterozygous mutation in *APOB* was found to be associated with autosomal dominant type B hypercholesterolemia, and homozygous mutation.

On the basis of high CADD and functional candidates (variants shared with mother-45269)

Some of the putative variants were shared, *POGZ* (Pogo Transposable Element Derived with ZNF Domain) was found more putative and real with Nonsynonymous SNV and CADD = 33 (no polyphen). Heterozygous mutation in *POGZ* associated with White-Sutton syndrome (Ye *et al*., 2015) reported LOF mutations. Stopgains and nonsynonymous SNV were found associated with Intellectual disability and microcephaly.

4.2.4 Variant annotation and variations

Large cohort of variants were detected and annotated to obtain the putative variant only. The variants need to be functionally annotated which was done by various annotation tools. Following annotation variants were prioritised and mainly targeted with deleterious variations. 51 variants were shared with father and 76 variant were shared with mother which was prioritised in this family where the number of exonic function was predicted. This results with the number of exonic mutation regarding deleterious variation.

Table 4.3: Prediction of various exonic mutation (variant shared with father)-45294

Mutation Type	Number of Variants with mutation
Stopgain	1
Non synonymous SNVs	24
Synonymous SNVs	19
Non frameshift deletion	1
Frameshift deletion	1
Frameshift insertion	1
Non frameshift insertion	1
Unknown	2
Others	1

Table 4.4: Prediction of various exonic mutation (variant shared with mother-45269)

Mutation Type	Number of Variants with mutation
Non synonymous SNVs	47
Synonymous SNVs	22
Non frameshift deletion	1
Frameshift deletion	2
Frameshift insertion	1
Non frameshift insertion	2
Unknown	1

The siblings from this family bear different variations including mostly deleterious variations. Deleterious variations which affect protein sequences (nonsynonymous and frameshift, stopgain) were selected. Out of 51 variants shared with father, nonsynonymous SNVs cover highest percentage (47%) and out of 76 variants shared with mother, nonsynonymous SNVs cover highest percentage (62%) where frameshift deletion, frameshift insertion, nonframeshift deletion bear almost same proportion which is represented in Fig:4.4 and Fig:4.5. The variants shared with mother and father was found to have nonsynonymous SNVs with the highest proportion among other mutations.

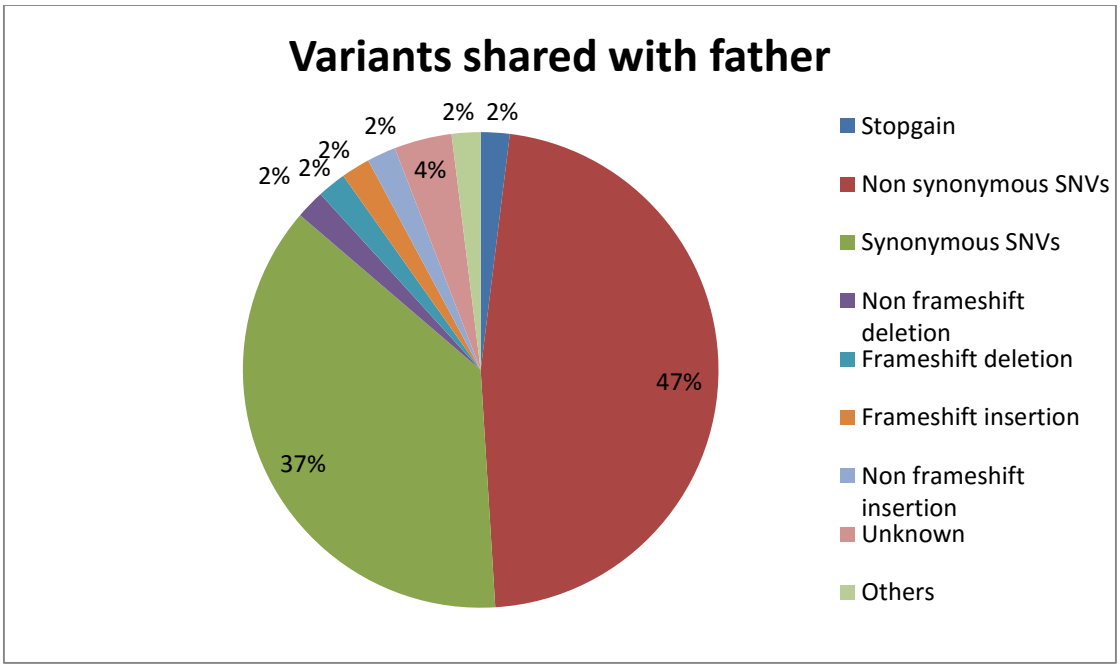

Fig 4.4: Pie chart representing the relative percentage of various types of variations. Out of 51 variants shared with father non synonymous SNVs cover highest percentage (47%) followed by synonymous SNVs (37%) where frameshift deletion, frameshift insertion, nonframeshift deletion bear almost same proportion.

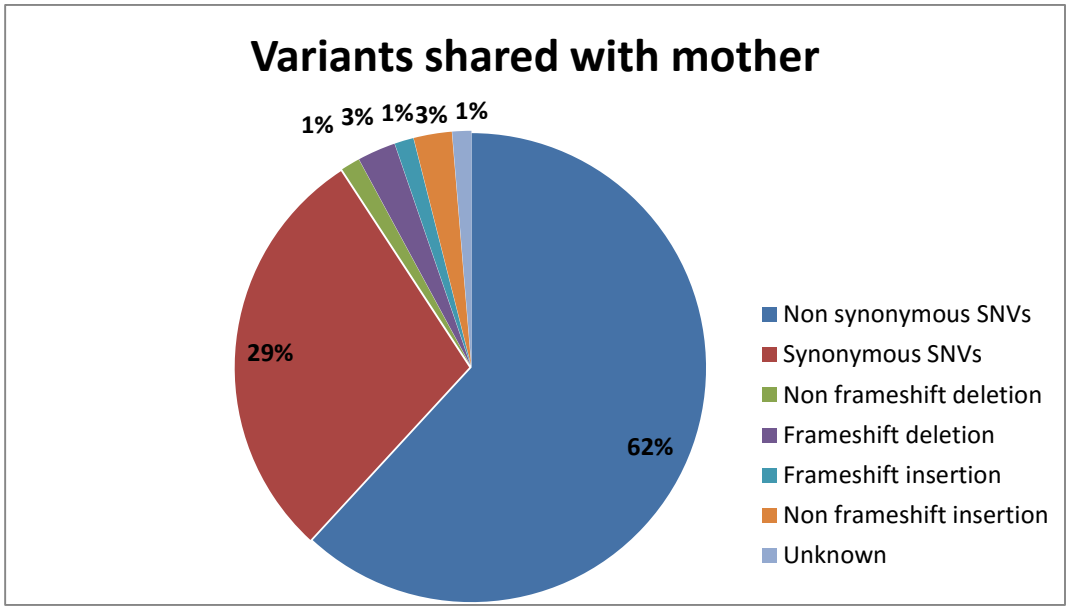

Fig 4.5: Pie chart representing the relative percentage of various types of variations. Out of 76 variants shared with mother, nonsynonymous SNVs cover highest percentage (62%) followed by synonymous SNVs (22%) where frameshift deletion, frameshift insertion, nonframeshift deletion bear almost same proportion.

4.2.5 Prioritization of variants and mutational analysis

The second families consisted of 13 variants and were followed but after filtering *POGZ* was found likely which affected multiple transcripts and lead to mild learning problems and microcephaly. Sanger follow up showed segregation in mother (45269) and 2 kids (daughter-45295 and son-45298) but not present in father (45294). Sample 45269, 45295 and 45298 were affected, heterozygous with C>T variant and 45294 was found homozygous WTG.

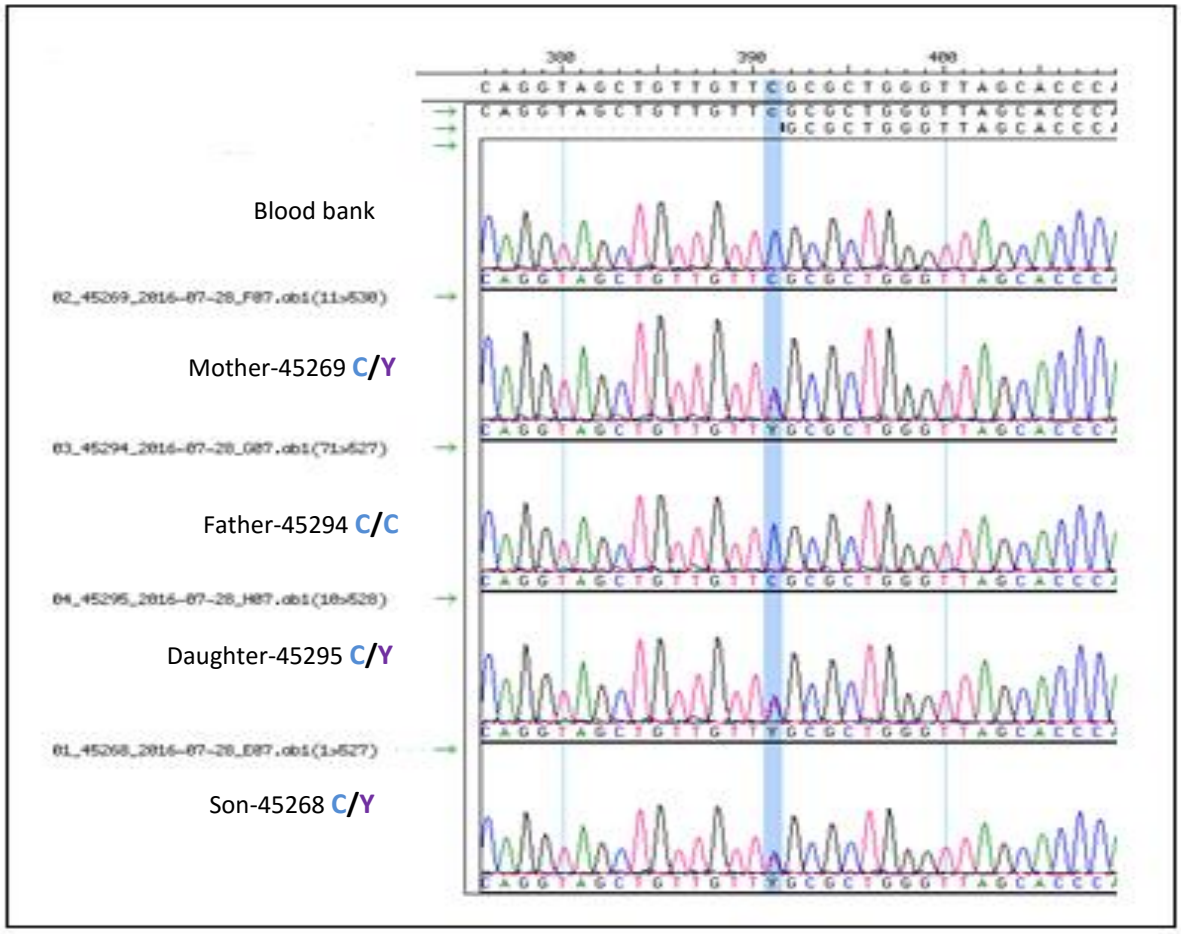

Fig 4.6 : Chromatogram derived from targeted capillary sequencing of family 2, testing segregation of the C>T variant, c.C2833T:p.E945K p Presents in exon 17 of *POGZ*. Segregation found to be in mother (45269) and 2 kids (45295 and 45298) but not present in father (45294). 45269, 45295 and 45298 were affected, heterozygous with C>T variant. The nucleotide code Y in DNA number 45269, 45295 and 45268 encodes the base change C>T (Y- C or T).

Variant being tracked was POGZ, chr1:151378393C>T in this family.

DNA	Putative variants	Genotype
45269	POGZ Chr:151378393c/t	C/T
45294	POGZ Chr:151378393c/t	C
45295	POGZ Chr:151378393c/t	C/T
45268	POGZ Chr:151378393c/t	C/T

The results of capillary sequencing of family members, the putative variant c.C2833T:p.E945K lies in exon 17 of *POGZ* and validates the phenotypic features were due to mutation in *POGZ*.

4.2.6 Visualization of the putative variant c.G2833A:p.E945K in exon 17 of *POGZ* in Integrative Genomics Viewer (IGV)

Large quantities of diverse types of genomic data were viewed with the Integrative Genomics Viewer (IGV) for interactive visual exploration of diverse genomic data. After the prediction of variant, IGV was used to foresee the mutational analysis of the gene. In family 2, the mutational change was clearly figured out through IGV.

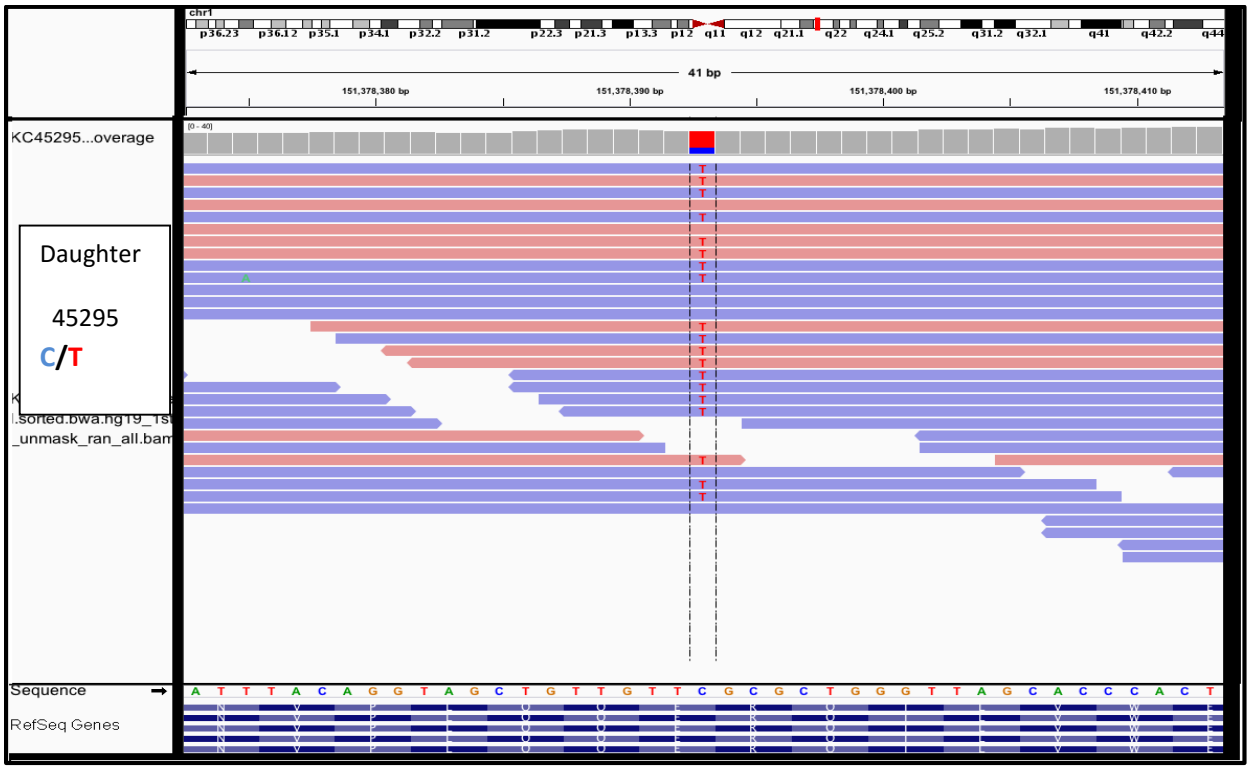

Fig 4.7: Visualization of variant c.C2833T:p.E945K in exon 17 of *POGZ*, chr1:151378393C>T in daughter using IGV. The IGV shows 45295(daughter) with heterogeneous mutation C/T.

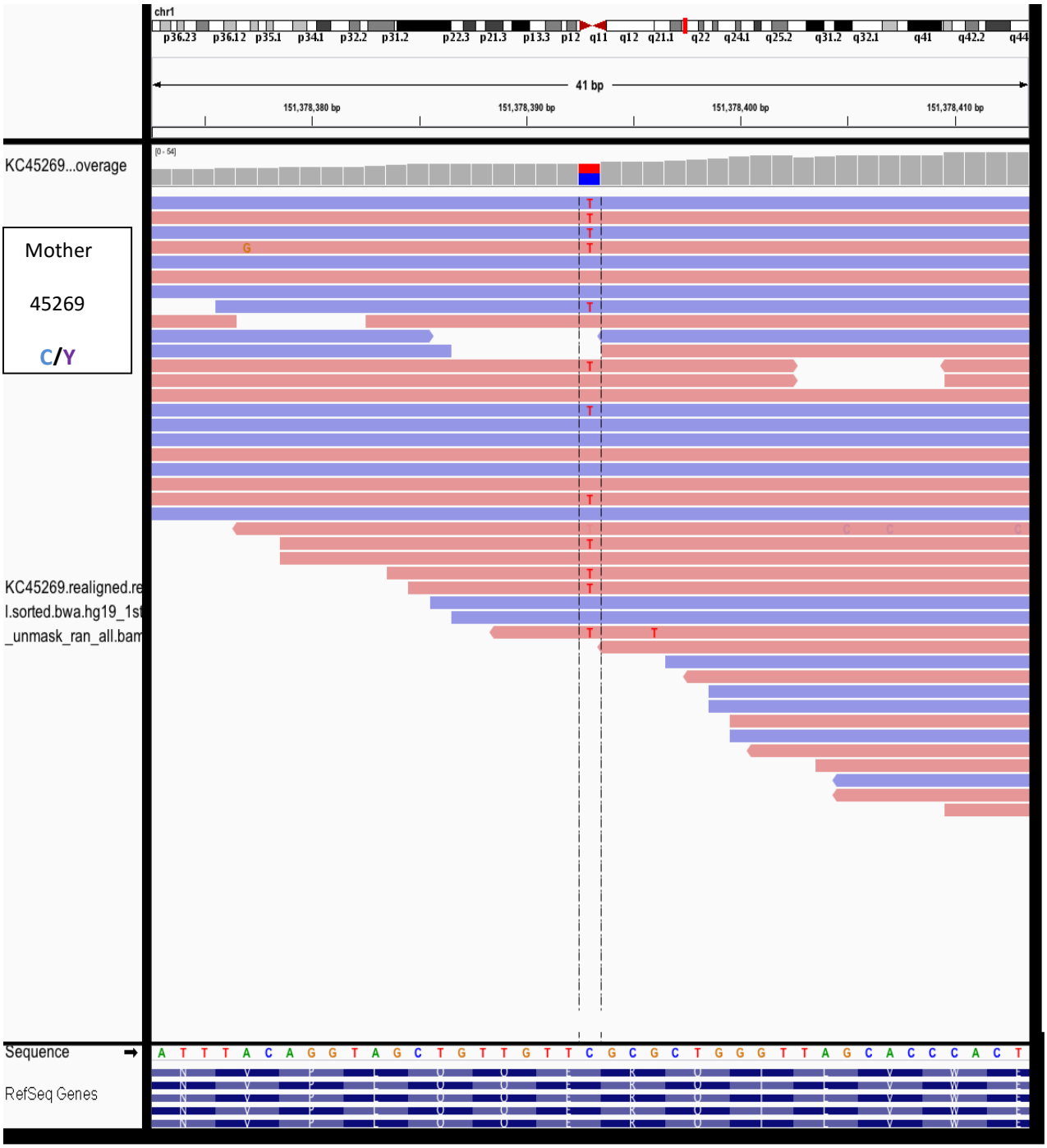

Fig 4.8: Visualization of variant c.C2833T:p.E945K in exon 17 of *POGZ*, chr1:151378393C>T in mother using IGV. The IGV shows 45269(mother) with heterogeneous mutation C/Y. The nucleotide code Y in DNA number 45269 encodes the base change C>T(Y- C or T).

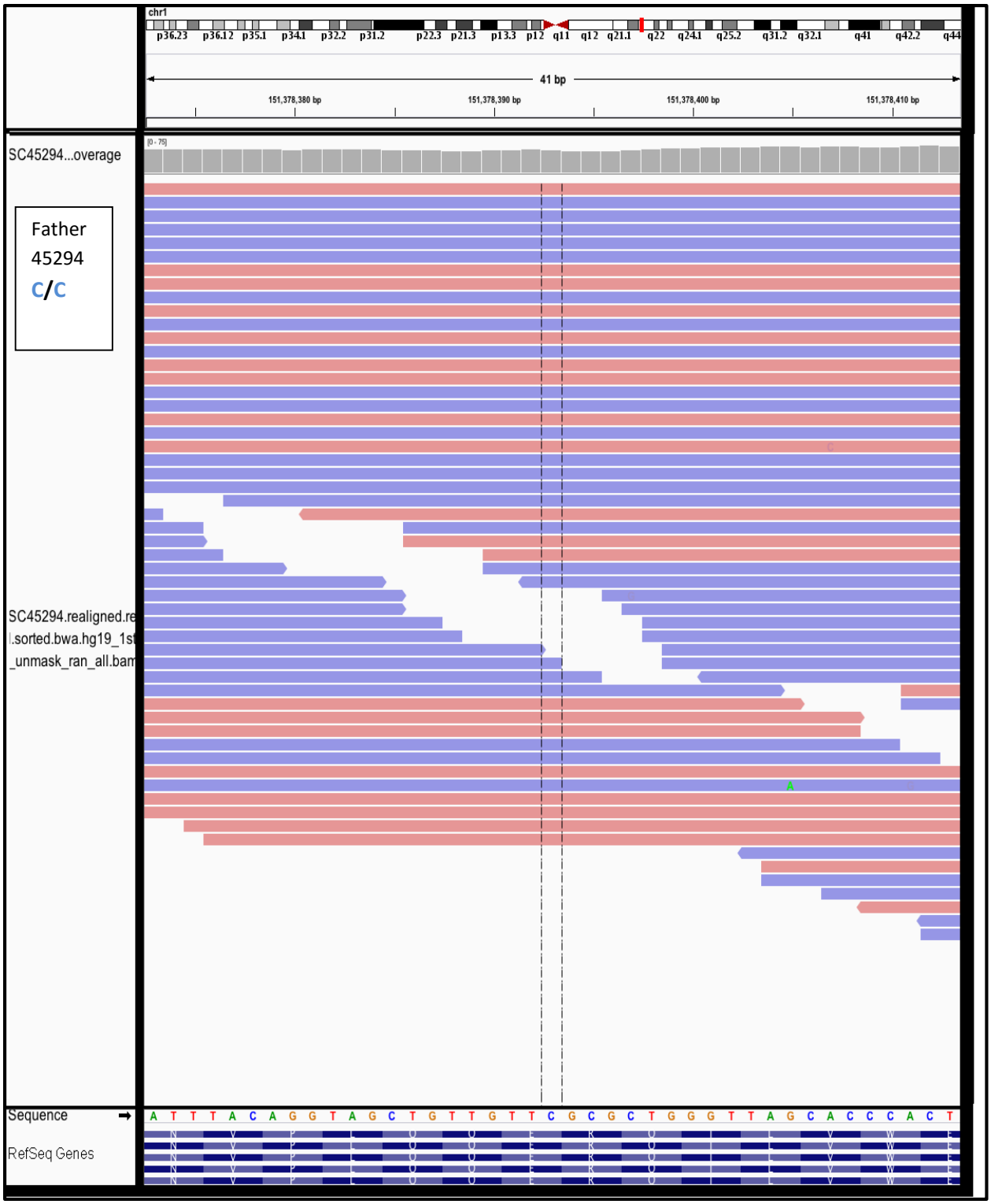

Fig 4.9: Visualization of variant c.C2833T:p.E945K in exon 17 of *POGZ*, chr1:151378393C>T in father using IGV. The IGV shows 45294 (father) was not affected.

Fig 4.10: Visualization of variant c.C2833T:p.E945K in exon 17 of *POGZ*, chr1:151378393 C>T using IGV in son. The IGV shows 45268(son) with heterogeneous mutation C/T. After the visualization of variant using IGV, it was clear that mother, son and daughter were with heterogeneous mutation C>T whereas father was not affected.

4.2.7 Pathogenicity prediction and conformation of variant

The variant was inherited in heterozygous form (45269, 45295-0/1) with CADD score 33 Nonsynonymous SNV and polyphen score of 0.995 implies that variant has probably damaging effect.

Table 4.5: Mutation analysis of *POGZ*

phastConsElements 46way	SIFT_score	Polyphen2_HVAR_score	MutationTaster_score	CADD_phred	Interpro_domain	KC45269	KC45295
Score=441,LOD=83	0.003	0.995	1	31	HTH CenpB-type DNA-binding domain	0/1	0/1

LOD (logarithm of the odds) score is found to be 83 with 441 score of conserved element refers to statistical estimate of whether two genes, or a gene and a disease gene, are likely to be located near each other on a chromosome and are therefore likely to be inherited.

4.3 Analysis of segregation of variants in Family 3

4.3.1 Clinical presentation

The third family was a family of affected males for three generations with Intellectual disability, psychosis, hypotonia feeding issues and club foot. The members with their respective DNA numbers were son (33881), daughter (33882) and mother (33883).

4.3.2 Quantification of DNA and PCR amplification

Family consisted of three DNA samples results in good yield. Extracted DNA samples were run on 1% agarose gel and quantified by Nanodrop. The concentration of the extracted DNA samples were 33881 (son)- 142.72ng/µl, 33882(daughter)- 9.44ng/µl and 33883(mother)- 1.65ng/µl was measured by using Nanodrop. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region containing the variant p.1673Tpresent in *MAOA***.** All the DNA samples were PCR amplified by using primers designed to amplify the genetic region.

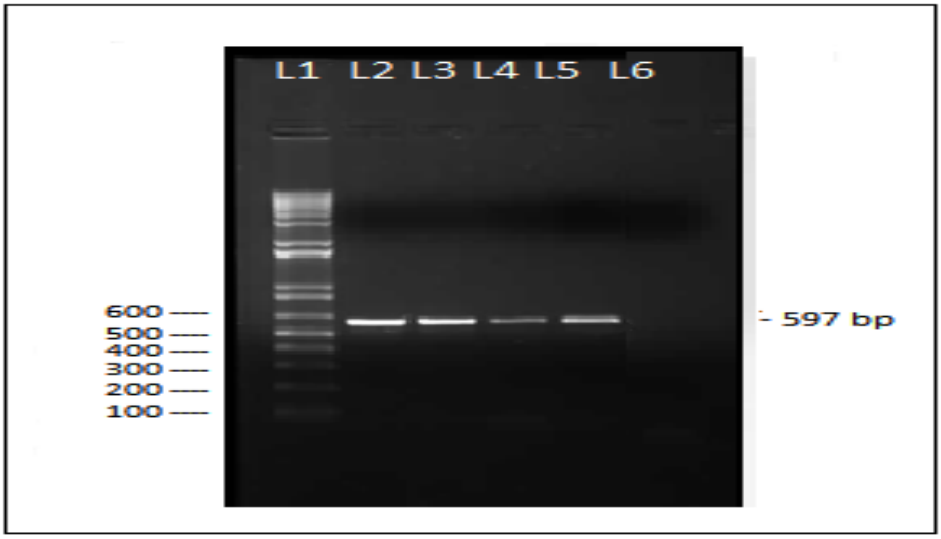

Fig 4.11: Family 3, Agarose gel (1%) electrophoresis including the variant p.1673Tof *MAOA*. L1– 1000 bp DNA ladder, L2 to L6– PCR amplicon, all samples were with good yield obtained by Phusion High Fidelity Taq polymerase, L5-Blood bank-positive control and L6-negative control.

4.3.3 Prioritization of variants and mutational analysis

After filtering 11 variants from this family, MAOA was found to be more likely affecting multiple transcripts. As expected affected male were hemizygous and mutation lead to impulsive aggressiveness and mild mental retardation, Variant was found de-novo in pro-band, however this family have two duplications (Xq and 18p segregating with various phenotypes). This means the amino acid change was actually (A or G) R>H (A or C or T) due to the adjacent base change as tabled below. From the findings of WES: *MAOA*, chrX: 43591035 G>A/G>T.

Chromosome number	Position	Gene list	Id	Ref
X	43591035	*MAOA*	33883	G>A/G>T

Table 4.6: Amino acid change due to the adjacent base change.

DNA	Genotype	AA at this position
Son-33881	A, G>T	Histidine
Daughter-33882	G, G>T	Arginine
Mother-33883	G/A, G>T	Histidine/Arginine

After the result obtained from targeted capillary sequencing, it is clear that the mutation in *MAOA* leads to of affected males for three generations with Intellectual disability, psychosis, hypotonia, feeding issues and club foot.

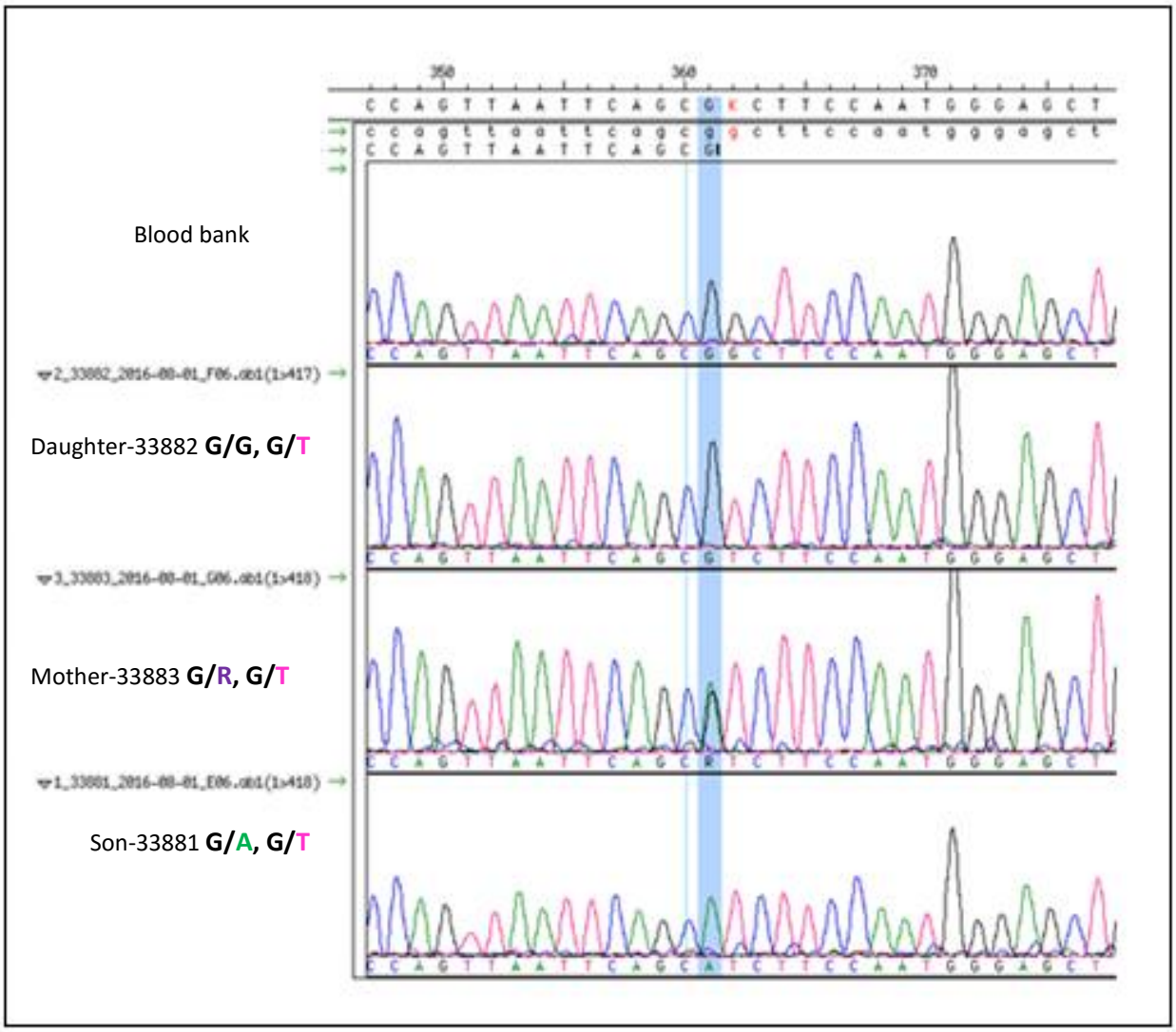

Fig 4.12: Chromatogram derived from targeted capillary sequencing of family3 including the variant p.1673T of *MAOA*. Testing segregation of the G>A/G>T variant, p.1673T present in *MAOA*. Affected male (son-33881) was with hemizygous mutation. The nucleotide code R in DNA number (mother) 33883 encodes the base change G>A (R- A or G).

Primer used was forward primer in all DNA samples. Among the variants, *MAOA* was identified real and its mutation is the causing factor which affects the protein sequences leading to the numerous disorders. This gene has also been associated with a variety of other psychiatric

disorders, including antisocial behavior. Alternatively spliced transcript variants encoding multiple isoforms have been observed.

4.3.4 Visualization of the putative variant p.1673T of *MAOA* in Integrative Genomics Viewer (IGV)

After the prediction of variant p.1673T of *MAOA*, IGV was used to predict the mutational analysis of the gene. Visualization of the G>A variant, p.1673T present in *MAOA*, affected male (son-33881) was found with hemizygous mutation to be as shown in Fig: 4.15.

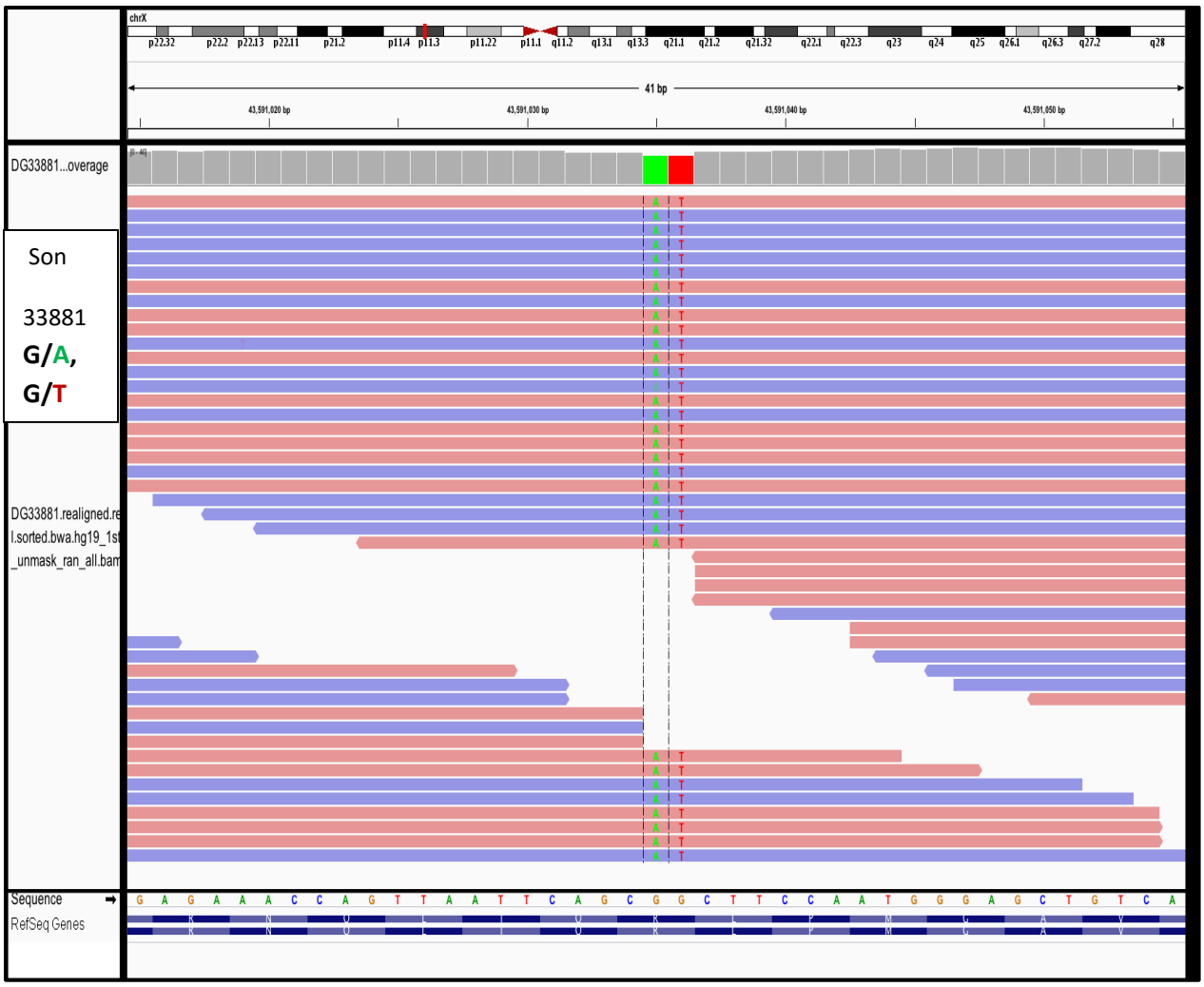

Fig 4.13: Visualization of variant p.1673T present in *MAOA*, chrX: 43591035 using IGV. The IGV shows son with hemizygous mutation **(G/A) and G/T.**

After the visualization of variant using IGV, it was clear that son was with hemizygous mutation G>A and G>T in *MAOA*. This mutation in *MAOA* was the main factor to cause Intellectual disability, feeding issues and club foot which were present in son.

4.4 Analysis of segregation of variants in Family 4

4.4.1 Clinical presentation

The family members were found to be associated with a variety of psychiatric disorders, including antisocial behavior in the affected samples.The members with their respective DNA numbers were 30096(mother), 41112(daughter), 44093 (son) and 48899 (father).

4.4.2 Quantification of DNA and PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were run on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured and resulted 30096(mother)-1057.15ng/µl, 41112(daughter)-214ng/µl, 44093(son)-538.62ng/µl and 48899(father)-226.12ng/µl by using Nanodrop.

All the DNA samples were PCR amplified by using primers designed to amplify the genetic region encompassing the variant c.588A>T: p.R196Sandc.587G>A: p.R196K present in exon 8. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR was 5'-TGGGGCCTCAGATAGTGAGT-3' as forward primer and 5'-GGGTTCTGACTCAACTCCCG-3' as reverse primer.

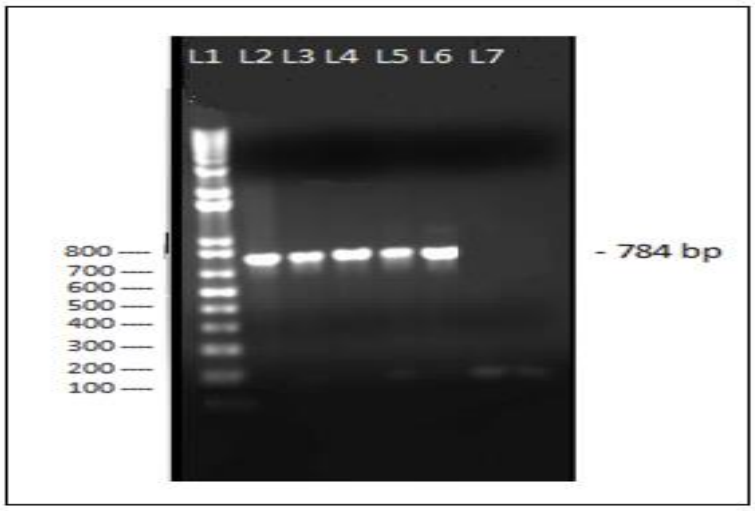

Fig 4.14: Family 4, PCR amplification of genomic region including the variant c.588C>T:p. R196S and c.587T>A:p.R196K in exon 8 of DLG2. 1000 bp DNA ladder. L2 to L5– PCR amplicon all samples were with good yield obtained by Phusion High Fidelity Taq-polymerase. L6-Blood bank-positive control and L7-negative control. Product size of the amplified region was 784bp.

4.4.3 Screening of basis of variants and analysis on the basis of heterogeneous character.

Variants were prioritised based on clinical phenotype, familial segregation, rare allele frequency and other criteria. Among these variants deleterious variations were targeted which are listed below in Table 4.7.

Table 4.7: Based on heterogeneous variants. These were the variants shared with mother (30096) based on functional character, high CADD and ID gene with heterogeneous mutations. After the validation of variant, the listed variants were priotised mainly on the basis of functional candidates shared with mother.

Chr	Start	End	Ref	Obs	Gene.gene	Exonic Functional gene
chr4	74453634	74453634	T	A	*RASSF6*	nonsynonymous SNV
chr4	74453635	74453635	C	T	*RASSF6*	nonsynonymous SNV
chr5	56234763	56234763	T	C	*MIER3*	nonsynonymous SNV
chr5	82785955	82785955	T	G	*VCAN*	nonsynonymous SNV
chr5	140784362	140784362	G	A	*PCDHGA9*	nonsynonymous SNV
chr5	150892052	150892052	A	G	*FAT2*	nonsynonymous SNV
chr5	178410177	178410177	C	T	*GRM6*	nonsynonymous SNV
chr6	170176444	170176444	C	T	*ERMARD*	nonsynonymous SNV
chr7	29134718	29134718	C	T	*CPVL*	nonsynonymous SNV
chr11	47362773	47362773	C	T	*MYBPC3*	nonsynonymous SNV
chr11	63974858	63974858	T	A	*FERMT3*	nonsynonymous SNV
chr11	64116879	64116879	C	T	*CCDC88B*	nonsynonymous SNV
chr11	67222904	67222904	G	C	*CABP4*	nonsynonymous SNV
chr11	**83180352**	**83180352**	**T**	**A**	***DLG2***	**nonsynonymous SNV**
chr11	**83180353**	**83180353**	**C**	**T**	***DLG2***	**nonsynonymous SNV**
chr11	88911575	88911575	C	T	*TYR*	nonsynonymous SNV
chr11	93839196	93839196	T	C	*HEPHL1*	nonsynonymous SNV
chr11	108128246	108128246	T	A	*ATM*	nonsynonymous SNV
chr11	129743891	129743891	G	A	*NFRKB*	nonsynonymous SNV
chr12	1036390	1036390	C	T	*RAD52*	nonsynonymous SNV
chr12	50358929	50358929	G	A	*AQP5*	nonsynonymous SNV

chr13	114107590	114107590	C	T	*ADPRHL1*	nonsynonymous SNV
chr14	73959574	73959574	A	G	*C14orf169*	Unknown
chr14	105413203	105413203	C	G	*AHNAK2*	nonsynonymous SNV
chr14	105418811	105418811	C	T	*AHNAK2*	nonsynonymous SNV
chr16	4934403	4934403	C	T	*PPL*	nonsynonymous SNV
chr16	16205325	16205325	G	A	*ABCC1*	nonsynonymous SNV
chr16	21739665	21739665	C	T	*OTOA*	nonsynonymous SNV
chr16	48122566	48122566	A	G	*ABCC12*	nonsynonymous SNV
chr16	70187381	70187381	C	T	*PDPR*	Stopgain
chr16	77328872	77328872	G	C	*ADAMTS18*	nonsynonymous SNV
chr17	15973554	15973554	T	C	*NCOR1*	nonsynonymous SNV
chr17	26678753	26678753	A	G	*POLDIP2*	Unknown
chr17	36487042	36487042	G	A	*GPR179*	nonsynonymous SNV
chr17	43227090	43227090	T	G	*HEXIM1*	nonsynonymous SNV
chr17	58123532	58123532	T	A	*HEATR6*	nonsynonymous SNV
chr17	73498328	73498328	G	A	*CASKIN2*	nonsynonymous SNV
chr17	73758832	73758832	G	A	*GALK1*	nonsynonymous SNV
chr17	76134472	76134472	C	T	*TMC8*	nonsynonymous SNV
chr17	76459054	76459054	T	C	*DNAH17*	nonsynonymous SNV
chr19	815771	815771	G	C	*LPPR3*	nonsynonymous SNV
chr20	44587908	44587908	T	A	*ZNF335*	nonsynonymous SNV
chr21	44151969	44151969	C	T	*PDE9A*	nonsynonymous SNV
chr22	19839703	19839703	G	C	*C22orf29*	nonsynonymous SNV
chr22	20920816	20920816	TA	CAG	*MED15*	nonframeshift insertion
chr22	50682306	50682306	G	A	*TUBGCP6*	nonsynonymous SNV

Among the 47 variants DLG2 seems to be real after viewing the mode of inheritance, intersection of cohort and assessing biological effect of variants based on literature reviews and online tools. Moreover the chromatogram obtained from the sequencing validated it to be real.

4.4.4 Variant annotation and variations

Large cohort of variants were detected and annotated to obtain the putative variant only based on heterogeneous mutation. The variants need to be functionally annotated which was done by

various annotation tools. Following annotation variants were prioritised and mainly targeted with deleterious variations.47 variants were prioritised in this family where the number of exonic function was predicted. This results with the number of exonic mutation regarding Non synonymous SNVs mutation to be the highest which is shown in Table 4.8.

Table 4.8: Summary of heterogeneous mutation shared with mother.

Mutation Type	Number of Variants with mutation
Non synonymous SNVs	43
Frameshift insertion	1
Stopgain	1
Unknown	2

The siblings from this family bear heterogeneous mutation including mostly deleterious variations. Out of 47 variants non synonymous SNVs cover highest percentage (92%) where non frameshift insertion and stopgain bear same proportion (2%) which is represented in Fig: 4.15.

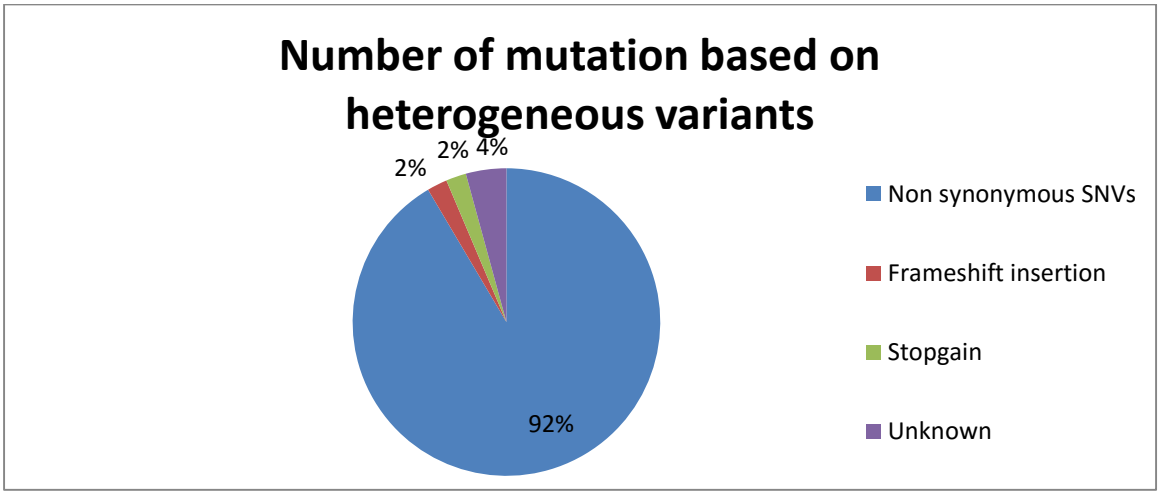

Fig 4.15: Pie chart representing the relative percentage of various types of variations. Out of 47 variants based on heterogeneous mutation, nonsynonymous SNVs cover highest percentage (92%) followed by non frameshift insertion and stopgain which bear same proportion (2%).

4.4.5 Prioritization of variants and mutational analysis

In fourth family, out of four samples, three of them (30096-mother-, 41112-daughter and 44093-son) have adjacent heterozygous changes with wild type and mutant allele C>T /T>A. From the findings of WES, variants were tracked; *DLG2*, chr11:83180352T>A and *DLG2*, chr11:83180353C>T.

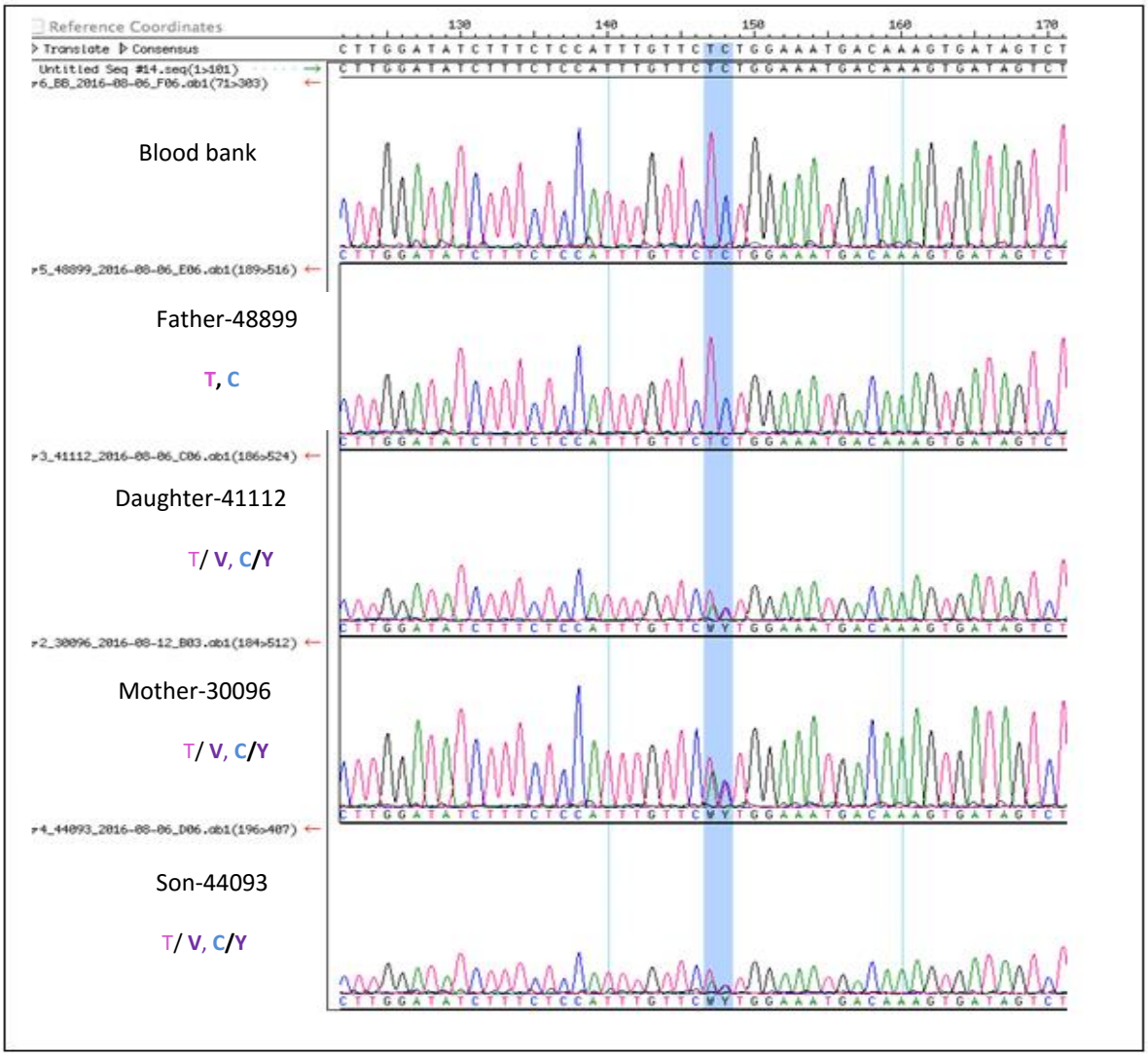

Fig 4.16: Chromatogram derived from targeted capillary sequencing of family 4 including the variantc.588C>T: p. R196S and c.587T>A: p.R196K in exon 8 of *DLG2* testing segregation of the C>T /T>A variant. Three of the family members (30097, 41112 and 44093) have adjacent heterozygous changes with wild type and mutant allele C>T /T>A. Primer used was reverse primer in all DNA samples. The nucleotide code V and Y in DNA number 41112, 30096 and 44093 which encodes the base change C>T/T>A(V- A or G or C and Y- C or T).

4.4.6 Visualization of the putative variantc.588A>T:p.R196S and c.587G>A:p.R196K in exon 8 of *DLG2* in Integrative Genomics Viewer (IGV)

After the prediction of the putative variantc.588A>T:p.R196S and c.587G>A:p.R196K in exon 8 of *DLG2,* IGV was used to predict the mutational analysis of the gene. After the visualization of

the variant C>T /T>C in IGV, aaffected male (son-44093) was found with adjacent heterozygous mutation as shown in Fig: 4.17.

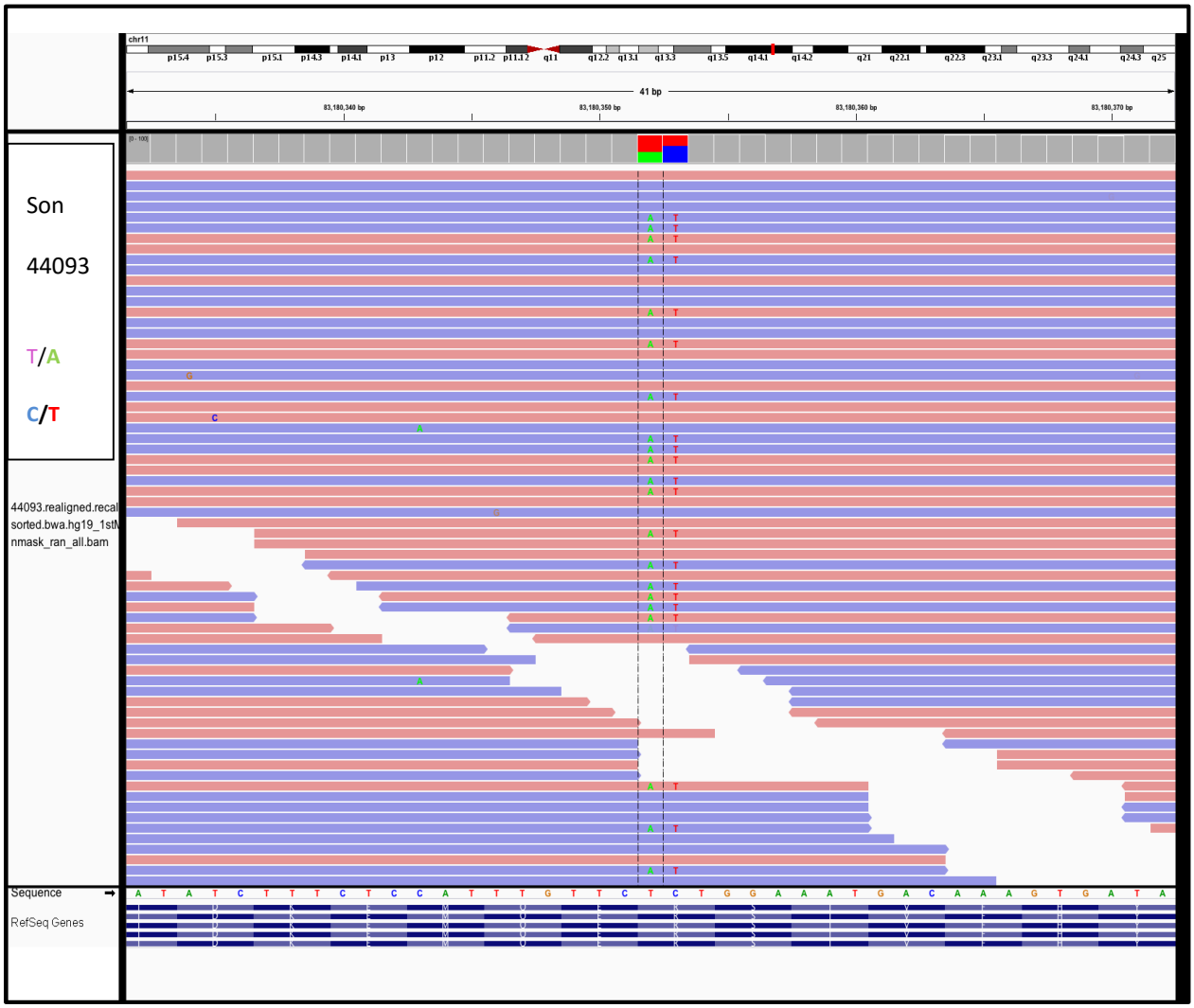

Fig 4.17: Visualization of putative variantc.588C>T:p.R196S and c.587T>A:p.R196K in exon 8 of *DLG2* USING IGV. The IGV snapshot shows son with adjacent heterozygous mutation.

4.5 Analysis of segregation of variants in Family 5

4.5.1 Clinical presentation

The family members were found to be associated with a variety of, autistic features, speech delay / nonverbal, seizures, craniofacial features, history of cerebellar ataxia in distant relatives (Parkinsonian-like) in both males and females. The members with their respective DNA numbers were 39696 (father), 39697 (mother) and 45416 (son).

4.5.2 Quantification of DNA and PCR amplification

Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. All the samples were of high yield and good quality. The concentration of the extracted DNA samples 39696- 220.27ng/µl, 45416- 335.19ng/µl and 39697- 170.1ng/µl was measured by using Nanodrop. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region encompassing the variant c.G953A:p.R318H present in exon-8**.** All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR was 5'-TCTTTACCCCTTCTCCTCCCAA-3' Forward primer and 5'-TAGTCATTGAGAAGCCAATTAGGG-3'as reverse primer.

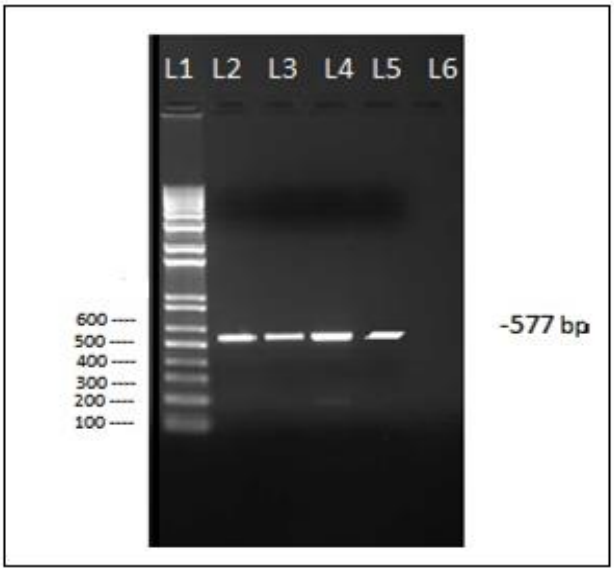

Fig 4.18: Family 5, PCR amplification of genetic region encompassing the variant c.C953T:p. R318H present in exon 8. L1- 1000 bp DNA ladder. L2 to L6, PCR amplicon with high yield and good quality obtained by Phusion High Fidelity. L5-Blood bank-positive control and L6-negative control. Product size of the amplified region was 577bp.

4.5.3 Screening of variants on the basis of deleterious variations

Variants were prioritised based on clinical phenotype, functional characteristics and high CADD score. Among these variants with deleterious variations were targeted which are listed below in Table 4.9.

Table 4.9: Variants with deleterious variations. These were the variants shared with Son (45416) based on functional character, high CADD and ID gene with heterogeneous mutations and deleterious variations. After the validation of variant, the listed variants were priotised mainly on the basis of functional basis in which *ALG13* was found to be shared and found real.

chr	Start	End	ref	obs	Gene.gene	Exonic Functional .gene
chr1	15844606	15844606	G	A	*CASP9*	synonymous SNV
chr1	39322575	39322575	T	C	*RRAGC*	synonymous SNV
chr1	43002200	43002201	AG	-	*CCDC30*	frameshift deletion
chr1	61547635	61547637	CCT	-	*NFIA*	nonframeshift deletion
chr1	85624690	85624690	C	T	*SYDE2*	nonsynonymous SNV
chr1	89848246	89848246	G	A	*GBP6*	synonymous SNV
chr1	150416763	150416763	C	T	*RPRD2*	nonsynonymous SNV
chr2	96616501	96616501	G	AA	*ANKRD36C*	frameshift insertion
chr2	96616506	96616507	GC	-	*ANKRD36C*	frameshift deletion
chr2	97818261	97818262	GC	-	*ANKRD36*	frameshift deletion
chr2	97818264	97818264	C	TT	*ANKRD36*	frameshift insertion
chr2	97851074	97851074	C	-	*ANKRD36*	frameshift deletion
chr2	97851078	97851078	A	G	*ANKRD36*	frameshift insertion
chr2	97851083	97851086	TAAT	-	*ANKRD36*	Unclassified
chr2	133075479	133075479	C	-	*ZNF806*	frameshift deletion
chr2	133076118	133076118	A	-	*ZNF806*	frameshift deletion
chr4	39878643	39878643	T	C	*PDS5A*	nonsynonymous SNV
chr4	81122525	81122525	C	A	*PRDM8*	nonsynonymous SNV
chr4	155244406	155244409	TTTG	-	*DCHS2*	frameshift deletion
chr5	56177852	56177854	CAA	-	*MAP3K1*	nonframeshift deletion
chr6	57512788	57512788	C	TA	*PRIM2*	Unclassified
chr6	57512794	57512796	ATT	-	*PRIM2*	Unclassified
chr7	100550473	100550473	G	A	*MUC3A*	nonsynonymous SNV
chr7	100551376	100551376	C	A	*MUC3A*	synonymous SNV
chr7	100551390	100551390	T	A	*MUC3A*	synonymous SNV
chr7	100551392	100551392	C	T	*MUC3A*	nonsynonymous SNV
chr7	100551393	100551393	G	A	*MUC3A*	synonymous SNV
chr7	100551397	100551397	G	T	*MUC3A*	nonsynonymous SNV

chr7	100551400	100551400	G	A	*MUC3A*	nonsynonymous SNV
chr9	95237028	95237030	TCA	-	*ASPN*	nonframeshift deletion
chr10	28024153	28024153	C	G	*MKX*	nonsynonymous SNV
chr11	56143125	56143125	C	T	*OR8U1,OR8U8*	nonsynonymous SNV
chr11	56143260	56143261	GT	-	*OR8U1,OR8U8*	frameshift deletion
chr11	56143375	56143375	A	T	*OR8U1,OR8U8*	synonymous SNV
chr11	56143490	56143490	T	A	*OR8U1,OR8U8*	nonsynonymous SNV
chr11	56143729	56143729	C	T	*OR8U1,OR8U8*	synonymous SNV
chr11	56143785	56143786	AT	-	*OR8U1,OR8U8*	frameshift deletion
chr11	104878040	104878040	A	T	*CASP5*	frameshift insertion
chr12	40907337	40907337	C	T	*MUC19*	Unknown
chr12	45429832	45429832	G	A	*DBX2*	nonsynonymous SNV
chr13	39608335	39608335	T	A	*PROSER1*	Unclassified
chr13	48664556	48664556	T	A	*MED4*	Unclassified
chr14	105404421	105404421	G	C	*AHNAK2*	synonymous SNV
chr16	58577316	58577316	T	A	*CNOT1*	frameshift insertion
chr16	88780635	88780635	G	0	*CTU2*	Unclassified
chr17	28505136	28505136	T	C	*NSRP1*	synonymous SNV
chr19	30500122	30500124	TGA	-	*URI1*	nonframeshift deletion
chrX	38020200	38020200	C	G	*SRPX*	nonsynonymous SNV
chrX	105159737	105159737	C	T	*NRK*	nonsynonymous SNV
chrX	**110956473**	**110956473**	**C**	**T**	***ALG13***	**nonsynonymous SNV**
chrY	21154466	21154466	T	A	*CD24*	nonsynonymous SNV

Out of 52 variants which were prioritised, mutation in *ALG13* seems to be real, after viewing functional characterstics, intersection of cohort and assessing biological effect of variants based on literature reviews and online tools. Moreover the chromatogram obtained from the sequencing validated it to be real.

4.5.4 Variant annotation and variations

Large cohort of variants were detected and annotated to obtain the putative variant only. The variants need to be functionally annotated which was done by various annotation tools. Following annotation variants were prioritised and mainly targeted with deleterious variations. 52 variants were prioritised in this family where the number of exonic function was predicted

and shows the number of exonic mutation regarding deleterious variation –Non synonymous SNVs is high.

Table 4.10: Prediction of various exonic mutations.

Mutation Type	Number of Variants with mutation
Frameshift deletion	9
Frameshift insertion	5
Non frameshift deletion	3
Non synonymous SNVs	16
Synonymous SNVs	10
Unknown	1
Others	8

The siblings from this family bear different variations including mostly deleterious variations. Deleterious variations which affect protein sequences (nonsynonymous and frameshift) were selected. Out of 52 variants non synonymous SNVs cover highest percentage (31%) where frameshift deletion, frameshift insertion, nonframeshift deletion bear almost same proportion.

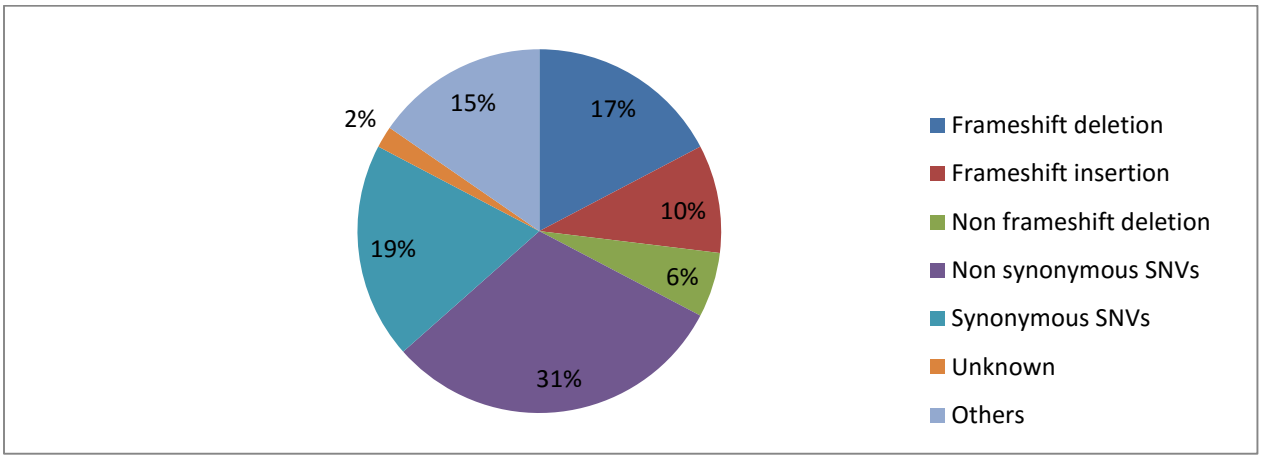

Fig 4.19: Pie chart representing the relative percentage of various types of variations. Out of 52 variants non synonymous SNVs cover highest percentage (31%) followed by synonymous SNVs (19%) where frameshift deletion, frameshift insertion, nonframeshift deletion bear almost same proportion (17%, 10% and 6% respectively).

4.5.5 Prioritization of variants and mutational analysis

After the Sequencing analysis in fifth family, revealed 4 variants. After filtering *ALG13* seemed more likely with nonsynonymous SNV, affects multiple transcripts, low frequency SNP (EXAC) and CADD value was 33. Among three members of this family, both the males (39696 and 45416)

found to be hemizygous with C>T variant and mother (39697) found to be heterozygous with WTG as well as mutant allele. Variant was tracked as *ALG13*, chrX: 110956473C >T.

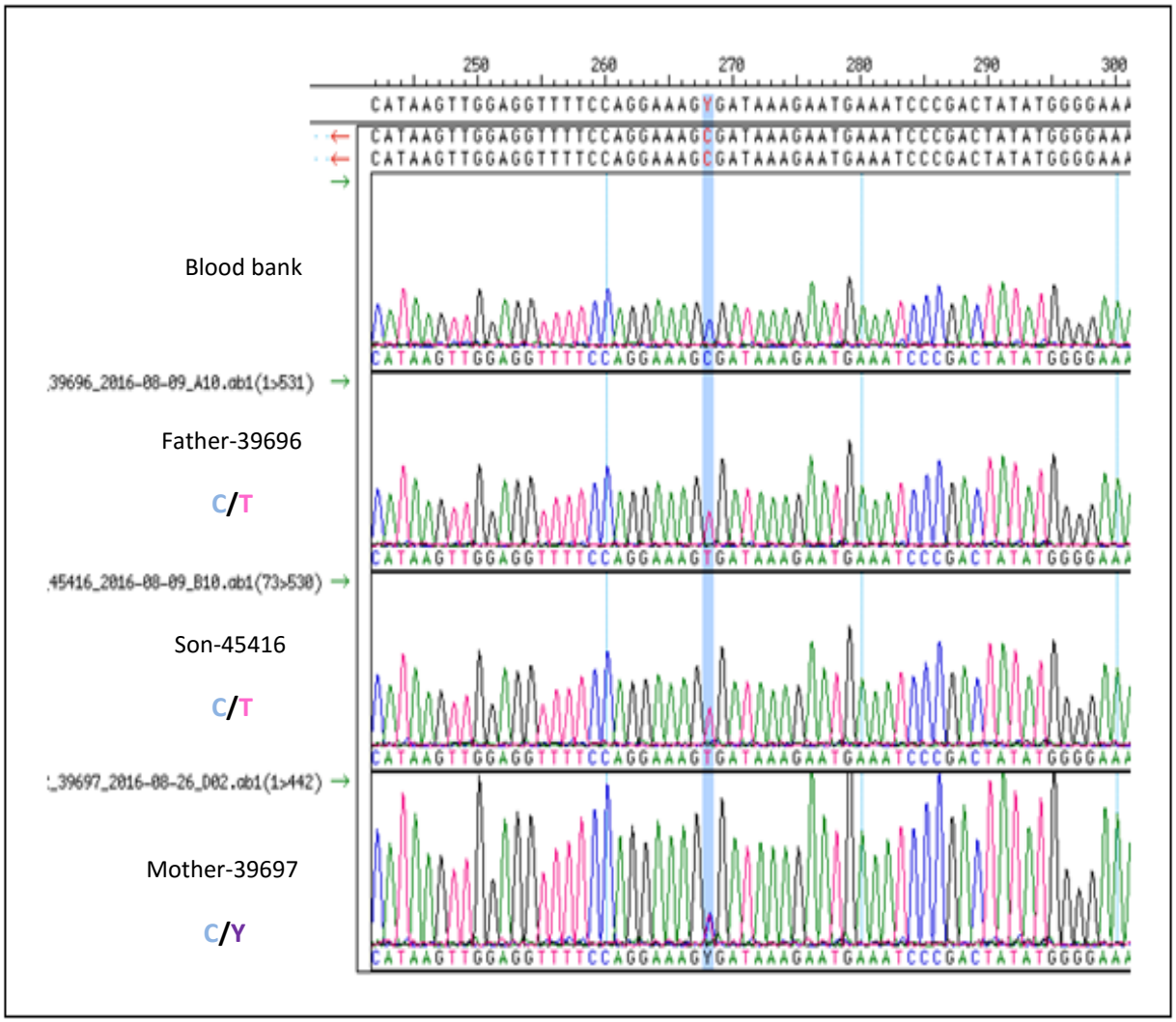

Fig 4.20: Chromatogram derived from targeted capillary sequencing of family 5. Sequencing image for family 5 shows the segregation of the C>T variant. Both the males (39696 and 45416 found to be hemizygous with C>T variant and mother (39697) found to be heterozygous with WTG as well as mutant allele. Primer used was reverse primer in all DNA samples. The nucleotide code Y in DNA number which encodes the base change C>T(Y-C or T).

It was found segregating in both affected males hemizygous mutation in *ALG13* associated with early infantile epileptic encephalopathy-36 with definite variations in the phenotypes reported. Both males were found with autistic features, speech delay, seizures and craniofacial feature

4.5.6 Pathogenicity prediction and conformation of variant

The variant was inherited in hemizygous form in males with CADD score 33 Nonsynonymous SNV and polyphen score of 0.993 implies that variant has probably damaging effect.

Table 4.11: Mutation analysis of *ALG13*

AAChange.gene	phastConsElements46way	ExAC.r0.1.filtered	EpilepsyGene	IDGene	LoFToolScores	SIFT_score	Polyphen2_HVAR_score	CADD_phred	phastCons20way_mammalian
ALG13:NM_001099922:exon8:c.C953T:p.R318H	Score=437;Name=lod=80	0.00006456	Name=NA	Name=NA	Name=0.273	0.02	0.993	33	0.999

The SIFT score 0.02 means it has deleterious effect in *ALG13* and Polyphen2_HVAR_score 0.993 implies that the variant has probably damaging effect in ALG 13. LOD 80 implies that variant found is likely to be inherited.

4.6 Analysis of segregation of variants in Family 6

4.6.1Clinical presentation

The mutation in this family was probably contributing to learning difficulties.The members with their respective DNA numbers were 25431(father), 25618(mother) and their siblings 43356(male), 25619 (male), 43358(female) and 43359(female).

4.6.2 Quantification of DNA and PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were run on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured by using nanodrop were 25431(father)-67.73ng/µl, 25618 (mother) -21.74ng/µl and their siblings 25619-61.29ng/µL, 43356-69.29ng/µl, 43358- 326.74ng/µl and 43359- 359.56ng/µl.

All the DNA samples were PCR amplified by using primers designed to amplify the genetic region encompassing the variant p.T1673C.All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR were 5’-CAGTTGATGTCCATGCTGCT3’ as forward primer and 5’-AGCCCTCAGAGCATTAGCAC-3’ as reverse primer.

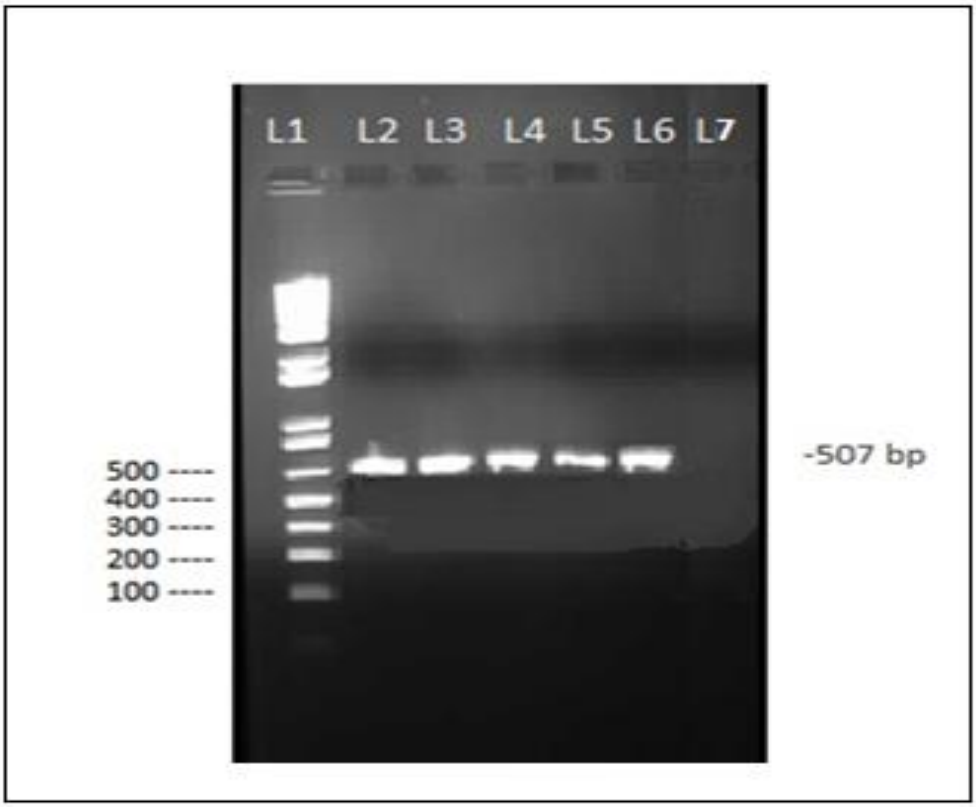

Fig 4.21: Family 6, PCR amplification of genetic region encompassing the variant p.1673T. L1– 1000 bp DNA ladder. L2 to L5– PCR amplicon obtained by Phusion High Fidelity Taq polymerase. L5-Blood bank-positive control and L6-negative control. Product size of the amplified region was 507 bp.

4.6.3 Prioritization of variants and mutational analysis

In sixth family, *GRIA1* was found likely was not a known ID gene among the sets of variants obtained after WES. 25618(mother) was found not to be affected whereas 25431(father) and 43356(male) have adjacent heterozygous changes with wild type and mutant allele T>C concluded that was probably contributing to learning difficulties. Variant being tracked was *GRIA1*, chr5:153144050T>C in this family.

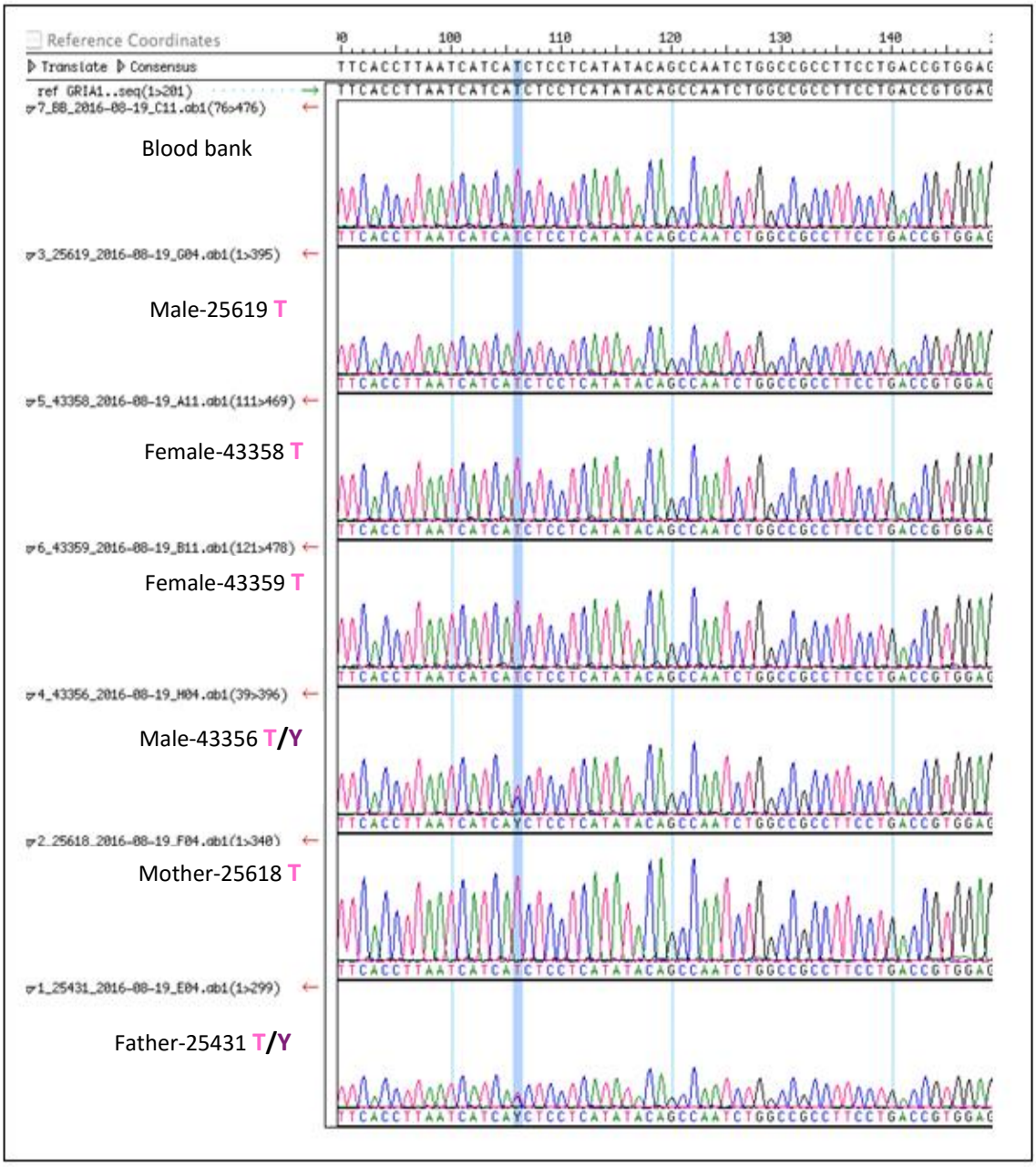

Fig 4.22: Chromatogram derived from targeted capillary sequencing of family 6 of the variant p.T1673C of GRIA1. Sequencing image for family 6 shows the segregation of the T>C variant. Primer used was reverse primer in all DNA samples. The nucleotide code Y in DNA number 25431(father) and 43356(male) with heterozygous mutation which encodes the base change T>C (Y-C or T).

4.7 Analysis of segregation of variants in Family 7

4.7.1 Quantification of DNA and PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured by using Nanodrop 34764- 47.13ng/µl, 50878- 66.29ng/µl and 50792- 154.52ng/µl. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region.

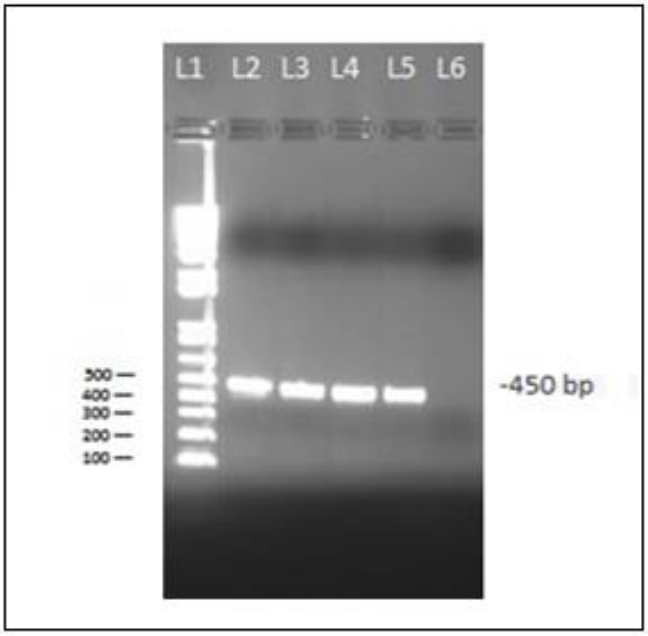

Fig 4.23 : Family 7, PCR amplification of genetic region. L1–1000 bp DNA ladder. L2 to L5– PCR amplicon obtained by Phusion High Fidelity Taq polymerase. L5-Blood bank-positive control and L6-negative control. Product size of the amplified region was 450 bp.

4.7.2 Prioritization of variants and mutational analysis

From the seventh family, three members were tested for segregation analysis among them mother (50878) and son (34764) were found to be heterozygous with WTG as well as mutant allele and father (50792) was not affected. Variant being tracked was *NCOR1* chr17:15968867C>T.

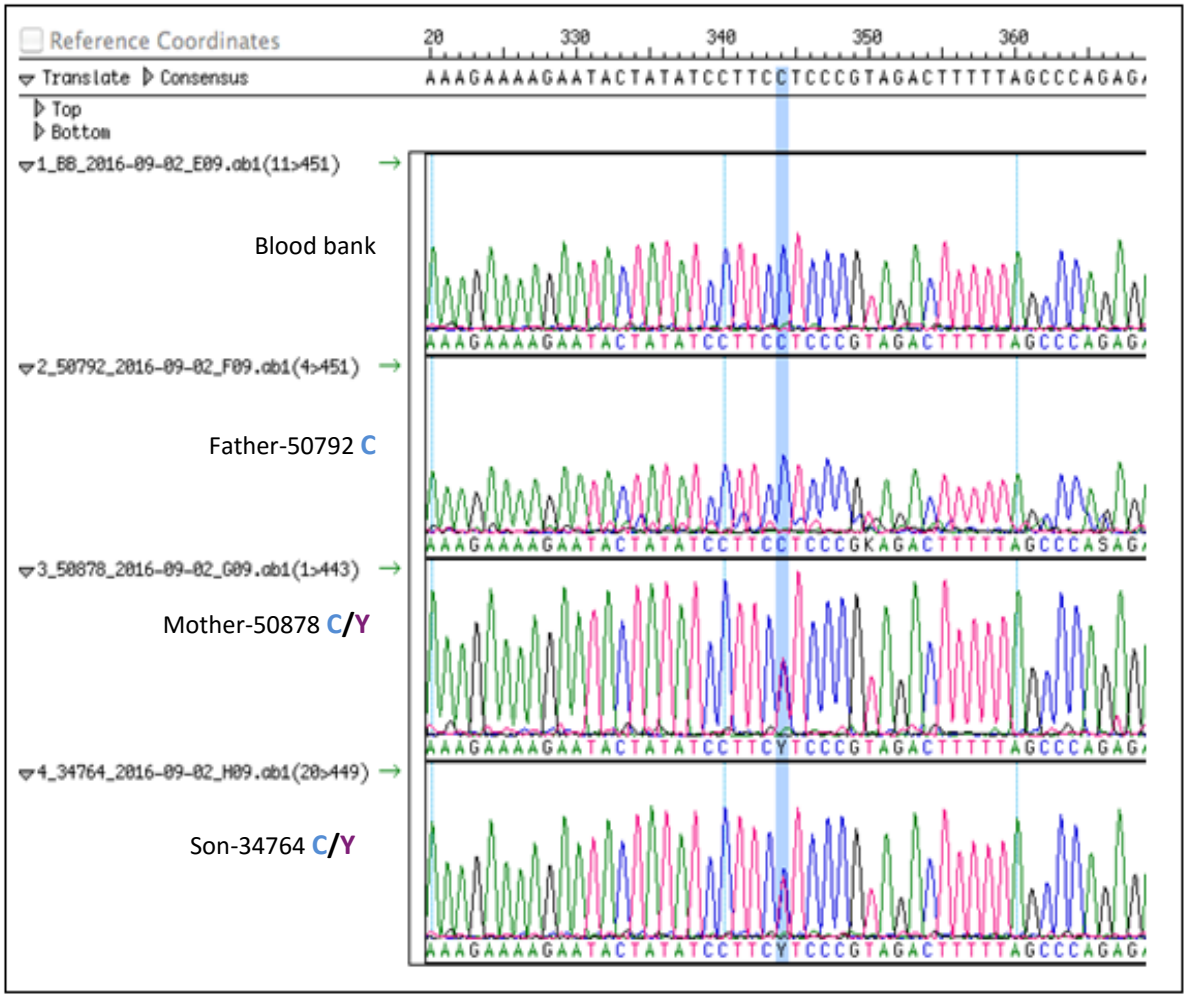

Fig 4.24: Chromatogram derived from targeted capillary sequencing of family 7. Testing segregation of the C>T variant. Mother (50878) and son (34764) were found to be heterozygous with WTG as well as mutant allele and father (50792). Primer used was reverse primer in all DNA samples. The nucleotide code Y in DNA number 50878 and 34764 which were with heterozygous mutation which encodes the base change C>T (Y- C or T).

4. 8 Analysis of segregation of variants in Family 8

4.8.1 Clinical presentation

The affected members were having learning difficulties and mild seizures with psychomotor retardation. Family members tested with their respective DNA numbers were 25064(mother), 43094(father) and their siblings 25094(female), 25545(female), 30702(male), 43093(female), 43095(male) and 45068(male).

4.8.2 Quantification of DNA and PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured by using Nanodrop which was 25064- 305.01ng/µl, 25094- 318.65ng/µl, 25545-200.86ng/µl, 30702- 277.15ng/µl, 43093- 284.89ng/µl, 43094- 228.13ng/µl, 43095- 225.05ng/ul and 45068-179.79ng/µl. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primer used in PCR were 5'-TCTAGGTGTGAATGCTTGGGG-3'as forward and 5'-TCAAGACTGTTCTTGCCCTGAT-3'as reverse primer.

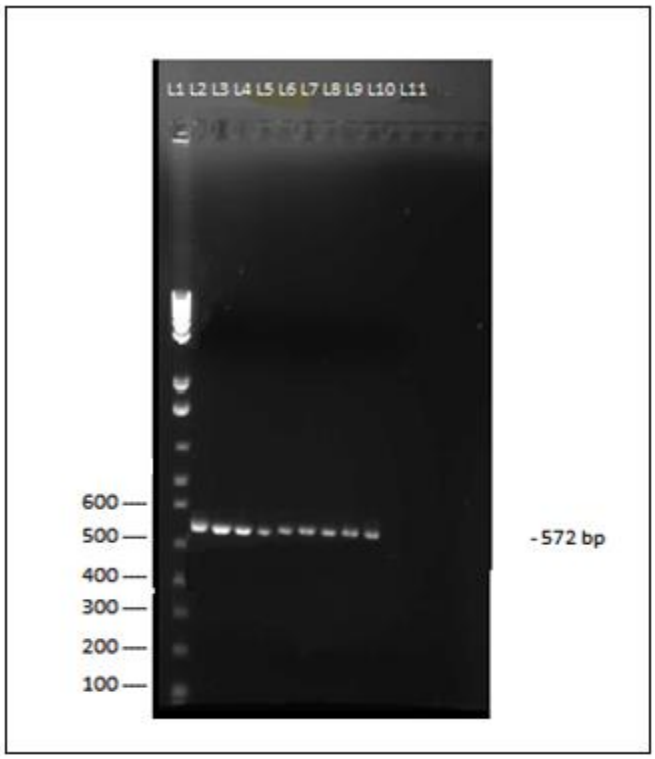

Fig 4.25: Family 8, PCR amplification of genetic region. L1– 1000 bp DNA ladder. L2 to L10– PCR amplicon obtained by Phusion High Fidelity Taq polymerase. L10-Blood bank-positive control and L11-negative control. Product size of the amplified region was 572bp.

4.8.3 Prioritization of variants and mutational analysis

From eighth family, Samples with DNA numbers numbers were 43094(father) and sibling 30702 (male) and 43095(male) found to be heterozygous with WTG as well as mutant allele whereas 45068(male) as homozygous with WTG and 25064(mother) was found to be not affected. From the findings of WES the variant tracked was *NALCN*, chr13:101844315 G>T.

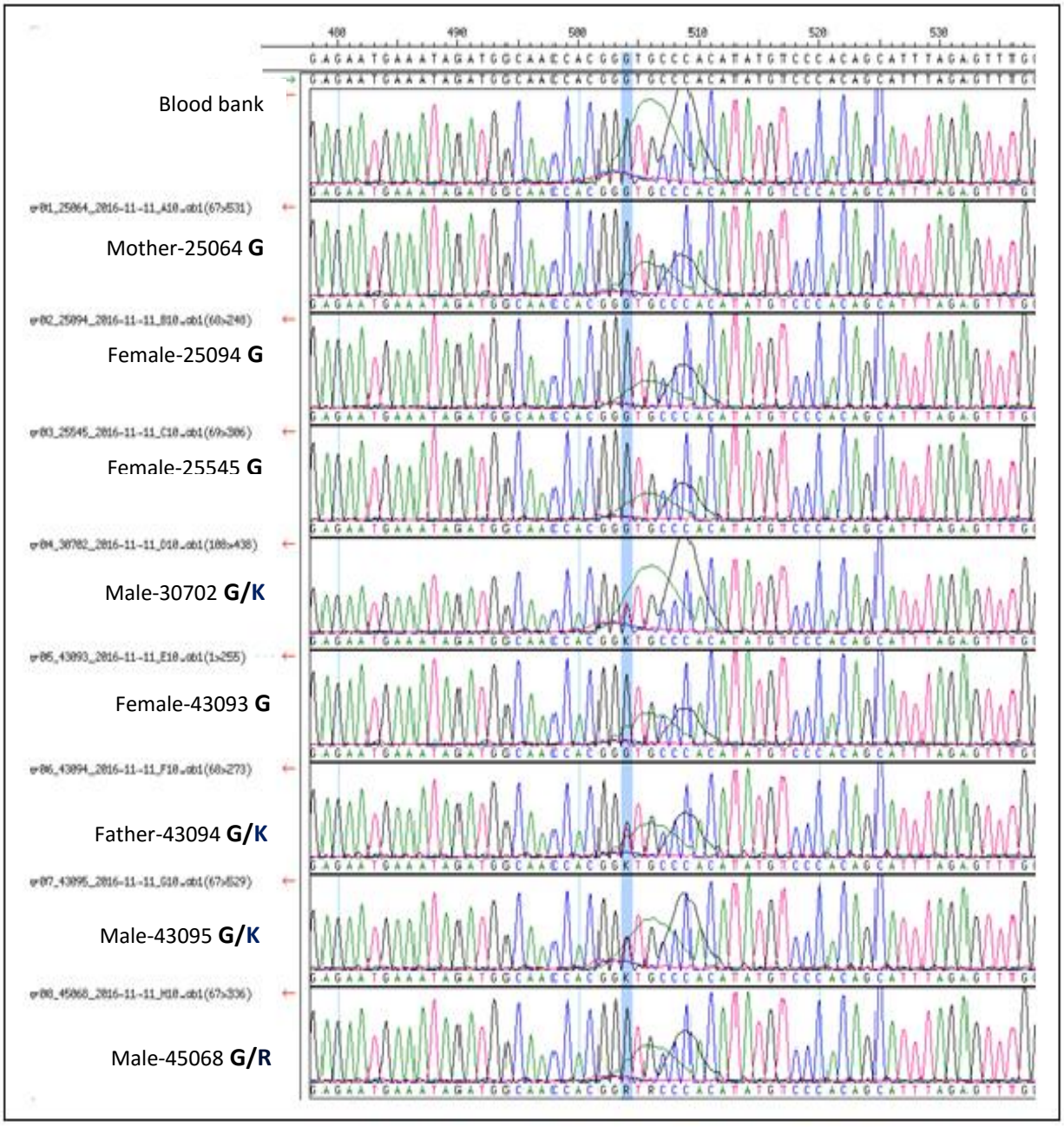

Fig 4.26: Chromatogram derived from targeted capillary sequencing of family 8 testing segregation of the G>T variant. Samples with DNA numbers 43094 (father) and siblings 30702(male), 45068(male) found to be heterozygous with WTG as well as mutant allele. Primer used was reverse primer in all DNA samples.The nucleotide code K in DNA number 30702, 43094 and 43095 which were sample with heterozygous mutation which encodes the base change G>T (K-G or T) and G>G (R-A or G).

Due to screening and analysis of variant, Sodium leak channel *(NALCN*) was confirmed real. Diseases associated with *NALCN* included hypotonia, infatile with psychomotor retardation and contractures of the limbs. The Clinical features which resembled with father predicted it was due to mutation in *NALCN*.

4.8.4 Visualization of the putative variant in Integrative Genomics Viewer (IGV) of *NALCN*,chr13:101844315 G>T.

After the prediction of the putative variant, IGV was used to predict the mutational analysis of gene. After the visualization of the variant of *NALCN*, chr13:101844315, aaffected male (30702) was found with heterozygous mutation G>T.

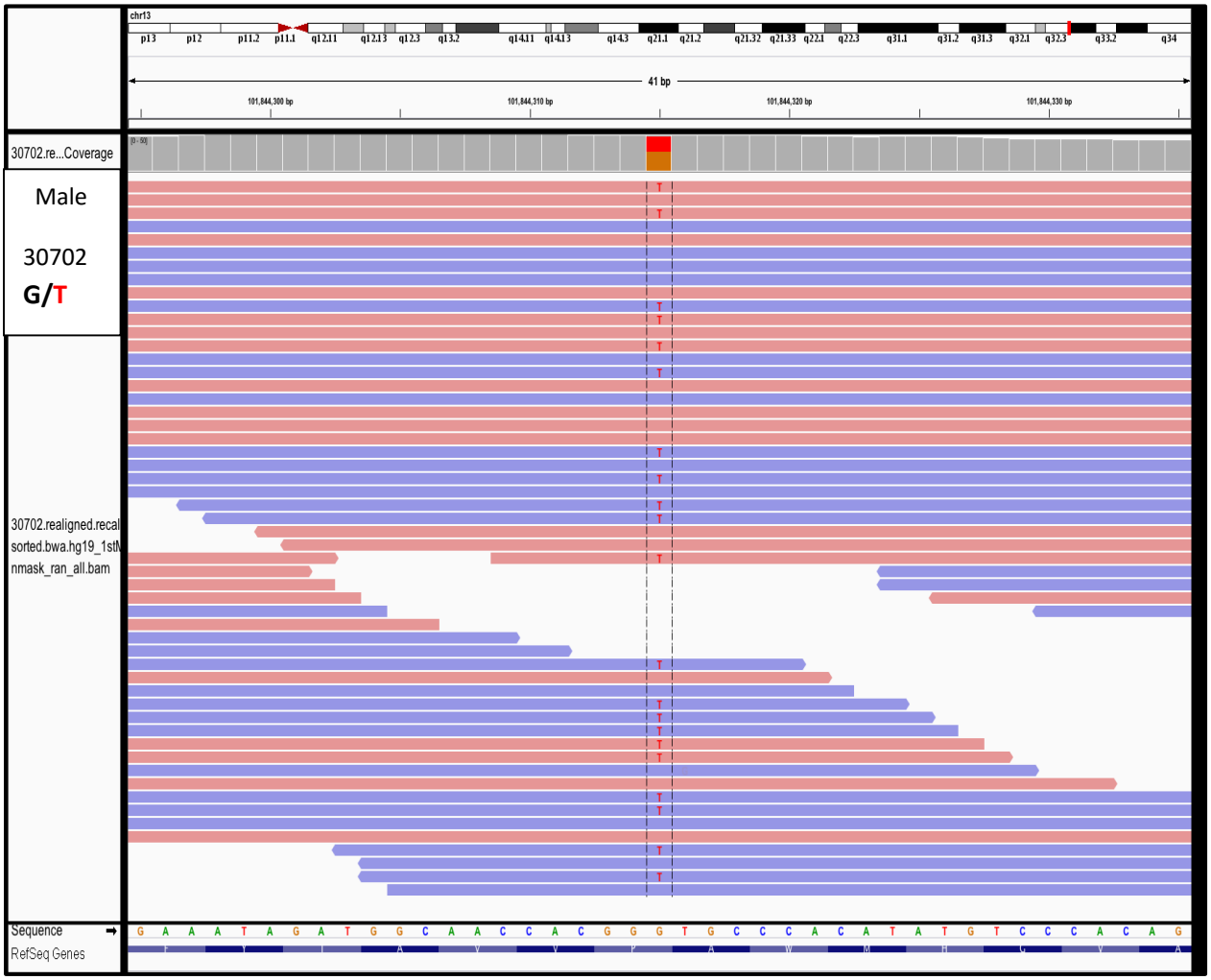

Fig 4.27: Visualization of putative variant of *NALCN*, chr13:101844315 G>T using IGV. The IGV snapshot shows 30702 being heterozygous with WTG as well as mutant allele for the mutation chr13:101844315 G>T.

4. 9 Analysis of segregation of variants in Family 9

4.9.1Clinical presentation

This is a family where brothers (19610 and 24176) and 1 nephew (19608) affected with nsID (no dysmorphic features). The members with their respective DNA numbers were 19608(nephew-male), 19609(father), 19610(brother), 19614(mother), 24176(brother) and their siblings 19611(male), 19612(female), 19613(female), 25087(female), 26434(male) and 26435 (male).

4.9.2 Quantification of DNA and PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured by using Nanodrop. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region. Primers used in PCR were CCTAGGGAGCAAGGCAATGAG-F & CCCTCTGCCAGACATTGCTC-R.

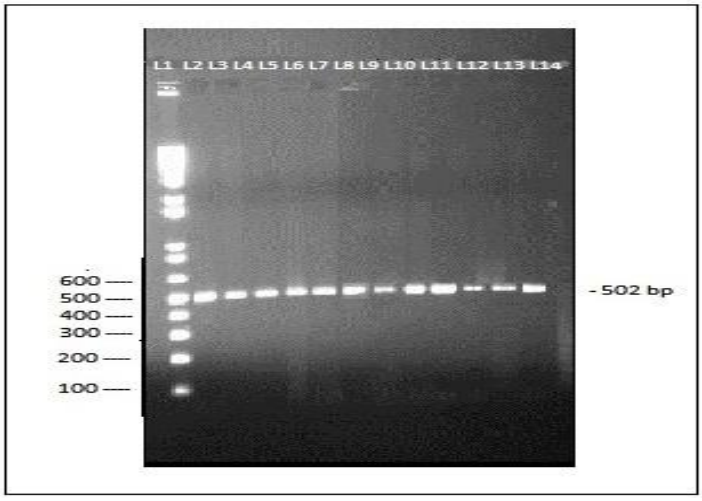

Fig 4.28: Family 9, PCR amplification of genetic region encompassing the variant c.G1058A:p.A353V of *DPYSL3*. L1– 1000 bp DNA ladder. L2 to L13 – PCR amplicon with high yield obtained by Phusion High Fidelity Taq polymerase. L13-Blood bank-positive control and L14-negative control. Product size of the amplified region was 502bp.

4.9.3 Screening of variants on the basis of functional candidates

Sequencing revealed 2 possible X-Chromosome variants and Autosomal Coding variant to follow up both from noncoding region. Validations of variants from noncoding region were as;

chr	start	end	ref	obs	Func	Gene
chrX	40962796	40962796	G	A	intronic	USP9X
chrX	79975267	79975267	G	A	intronic	BRWD3

After the shorting of the variants some of the putative variants were found shared with mother. *USP9X* (Ubiquitin Specific Peptidase 9, X-Linked) was found to be in noncoding region and base not to be conserved. It isn't found in regulatory region, sits in intron1-2. The phenotype linked with USP9x related to developmental delay, aggression, phenotype could fit possible to the change sites in a very large intron and in a not particularly conserved domain. *BRWD3* (Bromodomain and WD Repeat Domain Containing 3) found in noncoding region and base not particularly conserved. It is located in intron 17-18, not convincing as phenotype linked with BRWD3 shows macrocephaly. Sequencing revealed 6 Autosomal Coding variants to follow up:

Table 4.12: Possible Autosomal Coding variant.

Chr	**Start**	**End**	**Ref**	**obs**	**Gene**	**Exonic Function**	**AAChange**
chr2	136668972	136668972	C	G	DARS	nonsynonymous SNV	c.G1322C:p.R441T
chr3	42251580	42251580	C	GGA	TRAK1	nonframeshift insertion	c.1844_1845insGGA:p.E615delinsEE
chr5	**146785268**	**146785268**	**G**	**A**	**DPYSL3**	**nonsynonymous SNV**	**c.C1058T:p.A353V**
chr6	30954375	30954378	CACA	-	MUC21	frameshift deletion	c.423_426del:p.141_142del
chr6	30954378	30954378	GCA	TGTG	MUC21	frameshift insertion	c.426_427insTGTG:p.T142fs
chr16	88798129	88798129	G	C	PIEZO1	nonsynonymous SNV	c.C3181G:p.P1061A

DARS (Aspartyl-tRNA synthetase) and *PIEZO1* (Piezo Type Mechanosensitive Ion Channel Component 1) didn't fit unique patient phenotype. *TRAK1* (Trafficking Kinesin Protein 1) found to have non-frameshift insertion, located in exon contained in 1 human transcript not conserved in other species. Moreover this transcript found to be not likely to be expressed in the brain. *DPYSL3* (Dihydropyrimidinase like 3) found to be deleterious, base and AA conserved, sits in metallo-dependent hydrolase domain.

4.9.4 Prioritization of variants and mutational analysis

Variants which were prioritised revealed mutation in *DPYSL3* to be real. After viewing functional characteristics, intersection of cohort and assessing biological effect of variants based on literature reviews and online tools, variants were sorted. Moreover the chromatogram obtained from the sequencing validated it putative.

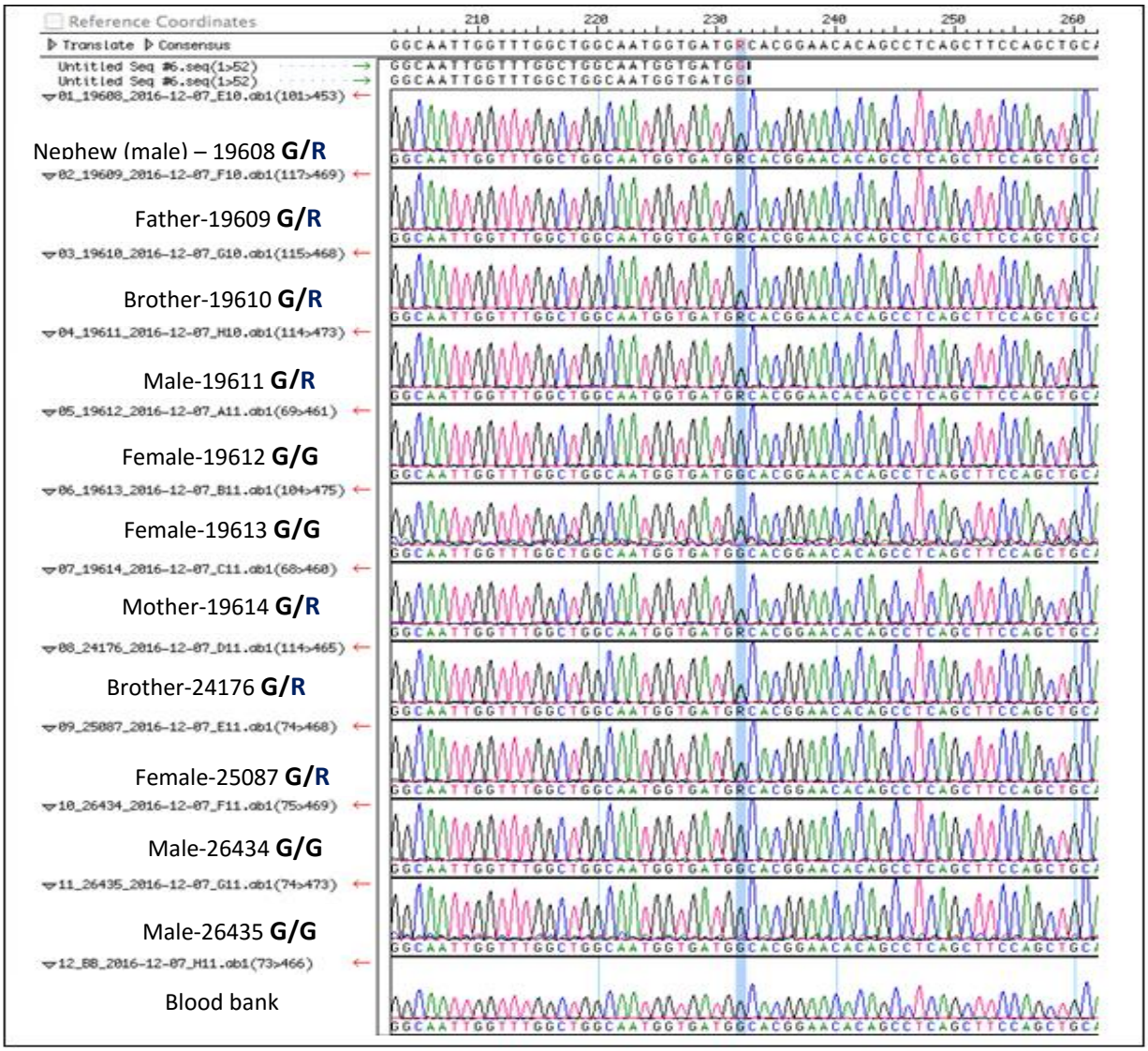

Fig 4.29: Chromatogram derived from targeted capillary sequencing of family 9 in *DPYSL3*. The nucleotide code R in DNA number 19608, 19609, 19610, 19611, 19614, 24176, 25087 found to be heterozygous mutation encodes the base change G>A (R-A or G). Samples with DNA numbers 19608 (nephew), 19609(father), 19610(brother), 19614(mother), 24176(brother) and siblings 25087(female) and 19611(male) found to be heterozygous. Siblings 19612 & 19613(female), 26434 & 26435(male) found not affected. Primer used was reverse primer in all DNA samples.

4.9.5 Visualization of the putative variant in Integrative Genomics Viewer (IGV) of *DPYSL3*.

After the prediction of the putative variant, IGV was used to predict the mutational analysis of gene. After the visualization of the variant of *DPYSL3*, chr5:146785268 G>A, affected father (19609) and brother (24176) were found with heterozygous mutation (G>A).

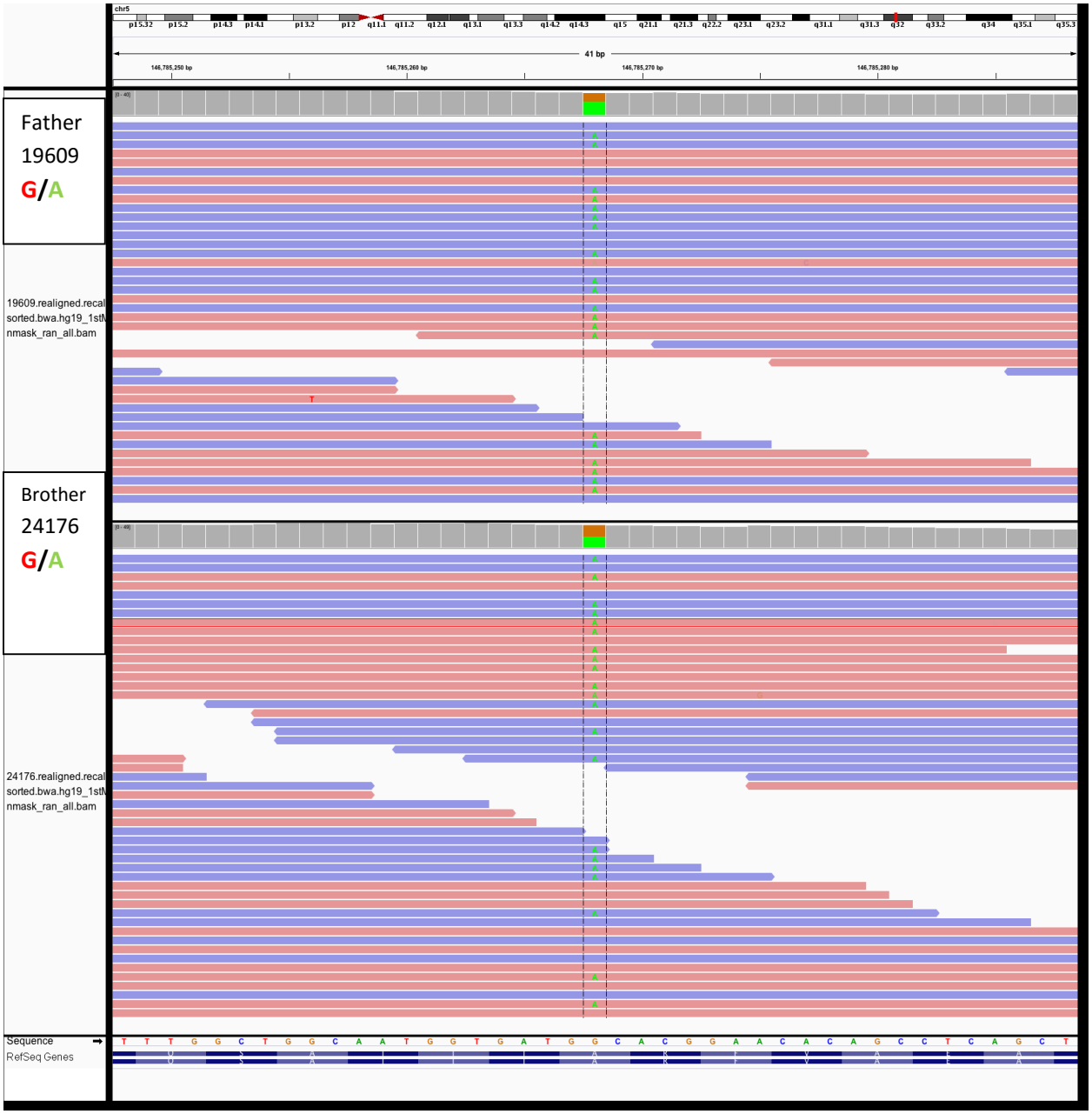

Fig 4.30: Visualization of putative variant of *DPYSL3*, chr5:146785268 G>A using IGV. The IGV snapshot shows father and brother being heterozygous with WTG as well as mutant allele for that mutation (chr5:146785268 G>A).

4.10 Analysis of segregation of variants in Family 10

4.10.1Clinical presentation

The affected members were having learning difficulties and mild seizures with psychomotor retardation. The members with their respective DNA numbers were 41987(mother), 41988(son), 45299(son), 51439(father) and 51440(daughter).

4.10.2 PCR amplification

All the samples were of high yield and good quality. Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region.

Fig 4.31: Family 10, PCR amplification of genetic region in *ANK3*. L1– 1000 bp DNA ladder. L2 to L7– PCR amplicon obtained with high yield by Phusion High Fidelity Taq polymerase. L7-Blood bank as positive control and L8- negative control. Product size of the amplified region was 600bp.

4.10.3 Prioritization of variants and mutational analysis

Variants were prioritised and shorted which revealed mutation in *ANK3*. After viewing functional characteristics and assessing biological effect of variants based on literature reviews and online tools and the chromatogram obtained from the sequencing validated it.

Fig 4.32: Chromatogram derived from targeted capillary sequencing for family 10 in *ANK3*. Testing segregation of the T>G variant, all the members were found heterozygous with WTG as well as mutant allele except 51439(father) who is not affected. Reverse primer was used in all DNA samples. The nucleotide code K(K-G or T) in DNA number 41967, 41988,45299 and 51440 found to be heterozygous mutation encodes the base change T>G.

4.10.4 Visualization of the putative variant in IGV OF FAMILY 10

After the prediction of the putative variant, IGV was used to predict the mutational analysis of gene. After the visualization of the variant of *ANK3*, chr10:61832591T>G, affected mother and both the son were observed with heterozygous mutation (T>G).

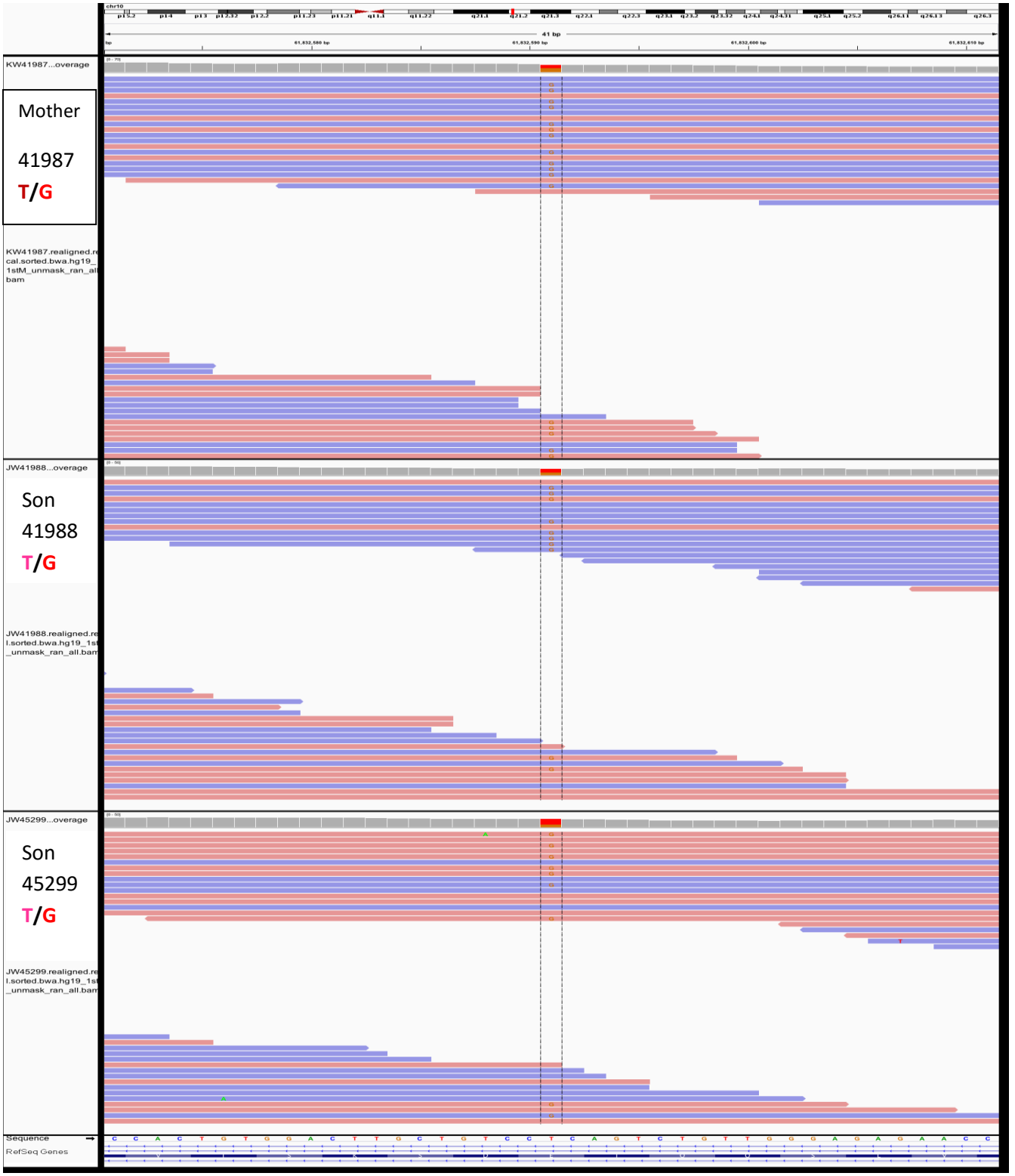

Fig 4.33: Visualization of putative variant of *ANK3*, chr10:61832591T>G using IGV. The IGV snapshot shows mother (41987) and her sons (41988 &45299) being heterozygous with WTG as well as mutant allele for the mutation (chr10:61832591T>G) of *ANK3*.

4.11 Analysis of segregation of variants in Family 11

4.11.1 Clinical presentation

The eleventh family from Poland was included in WES epilepsy sample set. This family consisted of affected males and females across multiple generations. The affected members were having learning difficulties and mild seizures with psychomotor retardation. The initial thoughts regarding the affected member were co-segregation of Intellectual disability & dysmorphic features suggesting multi-gene deletion. Members from this family with their respective notification and DNA numbers as II:3-45661,III:1-45662,III:2-45663,II:1-45664,II:4-45665,III:3 -45666 and II:5 -47904.

4.11.2 PCR amplification and quantification of DNA

All the samples were of high yield and good quality. Extracted DNA samples were checked qualitatively on 1% agarose gel electrophoresis and quantitatively by Nanodrop as described in materials and methodology section. The concentration of the extracted DNA samples was measured by using Nanodrop as, 45661-147.52ng/µl, 45662-625.70ng/µl, 45663-161.82ng/µl, 45664-221.89ng/µl, 45665-104.93ng/µl, 45666-228.13ng/µl and 47904-317.2ng/µl. All the DNA samples were PCR amplified by using primers designed to amplify the genetic region.

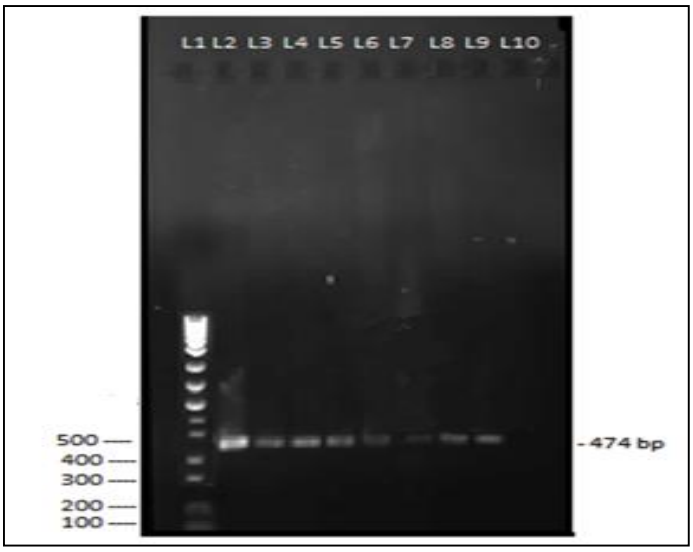

Fig 4.34 : Family11, PCR amplification of genetic region encompassing the variantc.C1192 G:p L398V on exon12 of ATP6V1B2. L1– 1000 bp DNA ladder. L2 to L9 – PCR amplicon, L9- Blood bank-positive control and L10- negative control. Product size of the amplified region was 474 bp.

4.11.3 Screening of variants on the basis of functional candidates, CADD score and known ID genes

Human genome comprises of large cohort of variations, where lies the disease causing the mutation. Deleterious variations affecting the protein sequences (frameshift, nonsynonymous, stopgain, stoploss) were targeted which revealed 95 variants which tend to be putative shown in Appendix 4 in brief.

Table 4.13: Variants based on functional character, high CADD and ID gene. These were the variants shared with son (45661-II: 3) second generation, based on functional character, high CADD and ID gene. After the validation of variant, the listed variants were priotised mainly on the basis of functional basis in which variant of *ATP6V1B2* was found to be real and putative.

Chr	Start	End	Ref	Obs	Gene.gene	Exonic Functional gene
chr1	17949604	17949604	C	G	*ARHGEF10L*	nonsynonymous SNV
chr1	19487529	19487529	G	A	*UBR4*	nonsynonymous SNV
chr1	22142473	22142473	G	A	*LDLRAD2*	synonymous SNV
chr1	29064190	29064190	C	G	*YTHDF2*	nonsynonymous SNV
chr1	32140895	32140895	G	A	*COL16A1*	synonymous SNV
chr1	33478899	33478899	G	A	*AK2*	synonymous SNV
chr1	44684084	44684084	C	T	*DMAP1*	synonymous SNV
chr1	45288989	45288989	G	A	*PTCH2*	synonymous SNV
chr1	53986422	53986422	G	A	*GLIS1*	synonymous SNV
chr1	67206976	67206976	G	C	*SGIP1*	nonsynonymous SNV
chr1	161696588	161696588	C	T	*FCRLB*	nonsynonymous SNV
chr2	24916179	24916179	G	A	*NCOA1*	synonymous SNV
chr2	26689619	26689619	T	A	*OTOF*	nonsynonymous SNV
chr2	68385266	68385266	G	C	*PNO1*	nonsynonymous SNV
chr2	75094895	75094895	G	A	*HK2*	nonsynonymous SNV
chr2	103334987	103334987	G	A	*MFSD9*	synonymous SNV
chr2	179474006	179474006	T	C	*TTN*	nonsynonymous SNV
chr2	187370519	187370519	A	G	*ZC3H15*	nonsynonymous SNV
chr3	4718380	4718380	G	T	*ITPR1*	nonsynonymous SNV
chr3	100170600	100170600	AG	TCC	*LNP1*	nonframeshift insertion
chr3	113374248	113374248	C	A	*USF3*	nonsynonymous SNV
chr3	113376111	113376119	TGCT	-	*USF3*	nonframeshift deletion
chr3	124281737	124281737	G	A	*KALRN*	synonymous SNV
chr3	128853715	128853715	A	G	*ISY1*	synonymous SNV

chr3	130452817	130452817	C	T	*PIK3R4*	nonsynonymous SNV
chr3	139173613	139173613	G	T	*RBP2*	nonsynonymous SNV
chr3	148596462	148596462	A	C	*CPA3*	nonsynonymous SNV
chr6	42932606	42932606	T	C	*PEX6*	nonsynonymous SNV
chr6	43970503	43970503	AG	GCG	*C6orf223*	nonframeshift insertion
chr6	84894982	84894982	T	A	*CEP162*	nonsynonymous SNV
chr6	165693571	165693571	C	T	*C6orf118*	nonsynonymous SNV
chr7	6449512	6449512	C	G	*DAGLB*	nonsynonymous SNV
chr11	55135807	55135807	C	T	*OR4A15*	Stopgain
chr11	62290653	62290653	A	G	*AHNAK*	nonsynonymous SNV
chr11	95724684	95724684	C	T	*MAML2*	synonymous SNV
chr11	101981600	101981600	G	A	*YAP1*	synonymous SNV
chr12	2622093	2622093	G	A	*CACNA1C*	nonsynonymous SNV
chr12	4655494	4655494	C	G	*RAD51AP1*	nonsynonymous SNV
chr12	4737620	4737620	A	G	*AKAP3*	nonsynonymous SNV
chr13	52518281	52518281	G	T	*ATP7B*	nonsynonymous SNV
chr13	114751194	114751194	T	C	*RASA3*	nonsynonymous SNV
chr14	32561112	32561112	G	C	*ARHGAP5*	nonsynonymous SNV
chr14	92537353	92537353	A	G	*ATXN3*	frameshift insertion
chr15	101606295	101606295	G	A	*LRRK1*	nonsynonymous SNV
chr16	12009530	12009530	GA	CCG	*GSPT1*	nonframeshift insertion
chr7	48314109	48314109	C	T	*ABCA13*	nonsynonymous SNV
chr7	89894669	89894669	T	G	*CFAP69*	synonymous SNV
chr7	149490487	149490487	G	A	*SSPO*	nonsynonymous SNV
chr7	151136984	151136984	T	A	*CRYGN*	nonsynonymous SNV
chr8	11666231	11666236	TCC	-	*FDFT1*	nonframeshift deletion
chr8	17613005	17613005	C	A	*MTUS1*	nonsynonymous SNV
chr8	**20074761**	**20074761**	**C**	**G**	***ATP6V1B2***	**nonsynonymous SNV**
chr8	22168751	22168751	G	A	*PIWIL2*	nonsynonymous SNV
chr8	24181402	24181403	AC	-	*ADAM28*	frameshift deletion
chr8	27509122	27509122	C	T	*SCARA3*	synonymous SNV
chr9	33466372	33466372	G	A	*NOL6*	nonsynonymous SNV
chr9	34996478	34996478	G	A	*DNAJB5*	nonsynonymous SNV
chr9	95237030	95237030	GA	TCA	*ASPN*	nonframeshift insertion
chr11	608597	608597	C	A	*PHRF1*	synonymous SNV
chr11	6411936	6411941	GCT	-	*SMPD1*	nonframeshift deletion
chr11	8132302	8132302	G	A	*RIC3*	synonymous SNV

chr11	17590724	17590724	C	T	*OTOG*	Stopgain
chr11	46905499	46905499	T	C	*LRP4*	synonymous SNV
chr16	88497443	88497443	C	A	*ZNF469*	nonsynonymous SNV
chr17	40027943	40027943	G	A	*ACLY*	nonsynonymous SNV
chr17	40270030	40270068	CTG	-	*KAT2A*	nonframeshift deletion
chr17	57272206	57272206	A	G	*PRR11*	nonsynonymous SNV
chr18	32826129	32826129	G	A	*ZNF397*	nonsynonymous SNV
chr19	40023438	40023438	G	A	*EID2B*	nonsynonymous SNV
chr19	41086245	41086245	G	A	*SHKBP1*	nonsynonymous SNV
chr19	41383114	41383114	C	T	*CYP2A7*	nonsynonymous SNV
chr19	42219023	42219023	G	A	*CEACAM5*	synonymous SNV
chr19	42880522	42880522	G	A	*MEGF8*	synonymous SNV
chr20	48504455	48504455	C	T	*SLC9A8*	synonymous SNV
chr20	49458319	49458319	T	-	*BCAS4*	frameshift deletion
chr20	60937477	60937477	G	A	*LAMA5*	synonymous SNV
chr21	43963651	43963651	C	T	*SLC37A1*	synonymous SNV
chr22	21288316	21288316	C	T	*CRKL*	synonymous SNV
chr22	29737500	29737500	G	A	*AP1B1*	nonsynonymous SNV
chr22	37906314	37906314	GA	CTC	*CARD10*	nonframeshift insertion
chr22	41747615	41747615	A	G	*ZC3H7B*	nonsynonymous SNV
chr22	47095277	47095277	TA	AAG	*CERK*	nonframeshift insertion
chrX	47272852	47272852	T	C	*ZNF157*	synonymous SNV
chrX	48620240	48620240	A	-	*GLOD5*	frameshift deletion
chrX	55779851	55779851	G	A	*RRAGB*	synonymous SNV
chrX	69459686	69459686	G	A	*AWAT1*	nonsynonymous SNV
chrX	117526804	117526804	G	C	*WDR44*	nonsynonymous SNV
chrX	135333478	135333478	G	A	*MAP7D3*	nonsynonymous SNV
chrX	135333480	135333480	C	-	*MAP7D3*	frameshift deletion
chrX	135333481	135333481	T	A	*MAP7D3*	nonsynonymous SNV
chrX	135333483	135333483	C	A	*MAP7D3*	nonsynonymous SNV
chrX	135333486	135333488	CCA	-	*MAP7D3*	nonframeshift deletion
chrX	141290724	141290724	G	A	*MAGEC2*	synonymous SNV
chrX	149638719	149638719	G	A	*MAMLD1*	nonsynonymous SNV

Out of 95 variants, five of the variants were targeted to be real on the basis of having epilepsy gene and ID genes and remaining five were targeted to be real on the basis of functional features and high CADD.

4.11.4 Variant annotation and variations

Following annotation, variants were prioritised and mainly targeted with deleterious variations shared with II:3 (45661). 95 variants were prioritised in this family where the number of exonic function was predicted. This results with the number of exonic mutation regarding deleterious variation. Non synonymous SNVs was found to be highest in this which is shown in Table 4.14.

Table 4.14: Summary of various exonic mutation with deleterious variation.

Mutation Type	Number of Variants with mutation
Non synonymous SNVs	49
Synonymous SNVs	28
Non frameshift deletion	5
Non frameshift insertion	7
Frameshift deletion	4
Frameshift insertion	1
Stopgain	2

The siblings from this family bear different variations including mostly deleterious variations. Deleterious variations which affect protein sequences (nonsynonymous and frameshift, stopgain) were selected. Out of 95 variants non synonymous SNVs cover highest percentage (51%) followed by synonymous SNVs (29%). Frameshift deletion, frameshift insertion, non-frameshift deletion bear almost same proportion which is represented in Fig: 4.35.

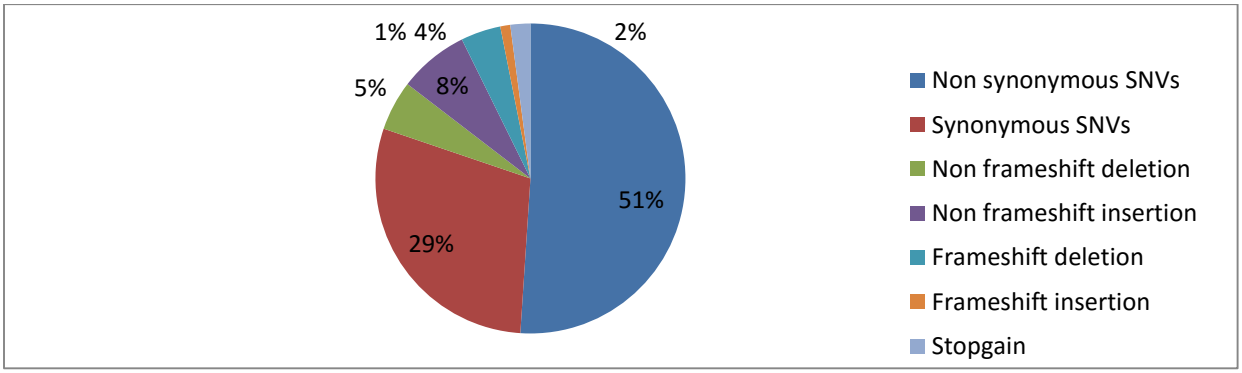

Fig 4.35: Pie chart representing the relative percentage of various types of variations. Out of 95 variants non synonymous SNVs cover highest percentage (51%) followed by synonymous SNVs (29%) where frameshift deletion, frameshift insertion, nonframeshift deletion, non-frameshift insertion and stopgain bear almost same proportion (4%, 1%, 5%, 8% and 2%respectively).

4.11.5 Prioritization of variants and mutational analysis

Variants which were prioritised revealed mutation in *ATP6V1B2* (ATPase H+ Transporting V1 Subunit B2) to be real, after viewing functional characterstics and assessing biological effect of variants based on literature reviews and online tools. Moreover, the chromatogram obtained from the sequencing validated it to be real. From the findings of WES, *ATP6V1B2*, chr8:20074761 C>G. The shorting of variants based on the pipeline followed for putative variant associated with epilepsy is shown in Fig 4.36.

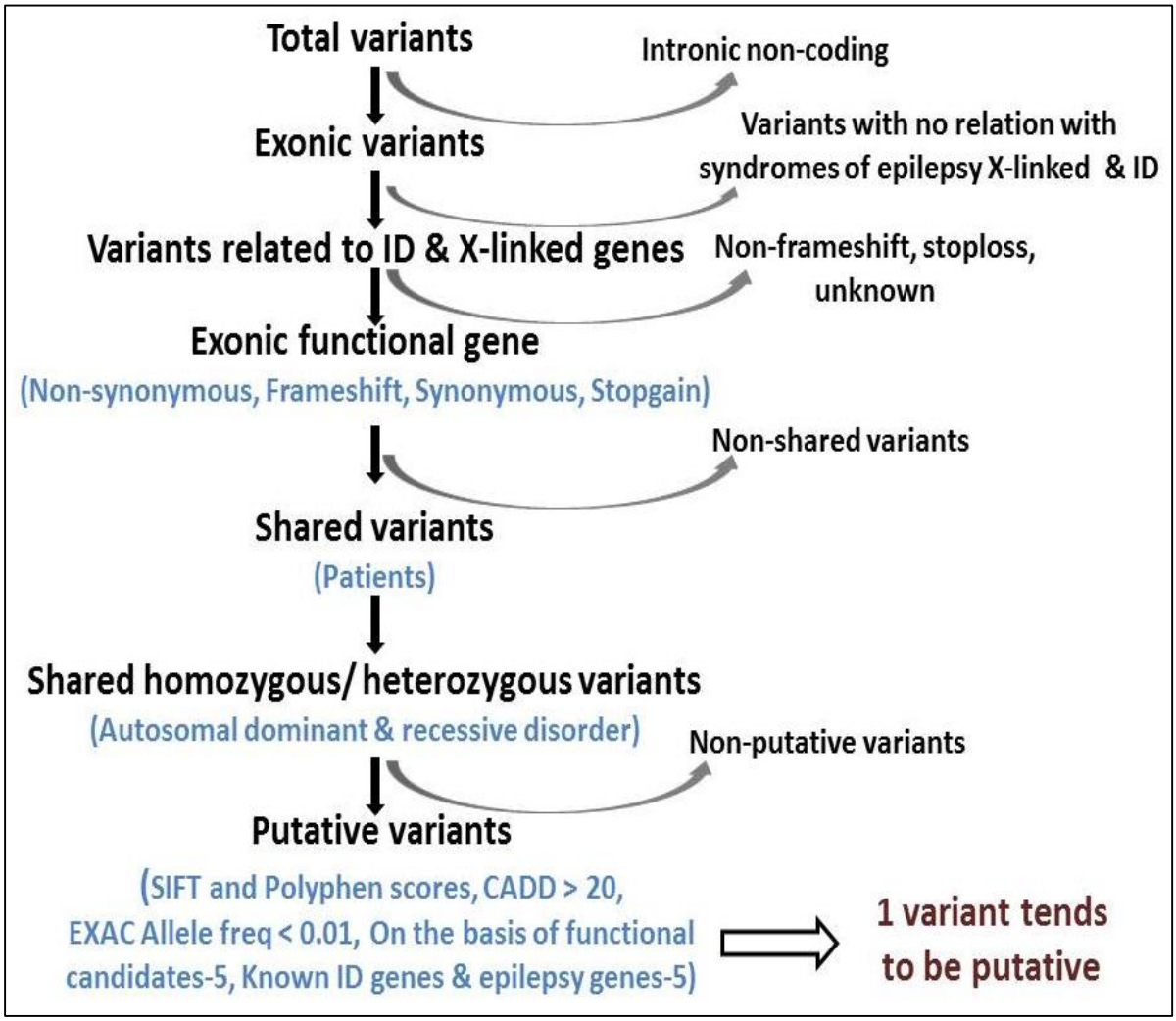

Fig 4.36: Pipeline used for prioritization of variants. Out of 95 variants obtained after annotation of variant, variants were sorted out based on clinical phenotype, familial segregation and functional basis to validate one possible variant. This result the putative variants revealing the mutation of *ATP6V1B2*.

4.11.6 Pathogenecity prediction of putative variant

The variant was inheritated heterozygous with WTG as well as mutant allelein all the Family members except II: 5 (47904) with CADD score 25.2, Nonsynonymous SNV and polyphen score of 0.964 implies that variant has probably damaging effect. The SIFT score 0.001 means it has deleterious effect.

Table 4.15: Prediction of putative variant.

chr	Ref	Obs	AAChange.gene	IDGene	SIFT_score	Polyphen2_HVAR_score	CADD_phred
chr8	C	G	ATP6V1B2:NM_001693:exon12:c.C1192G:p.L398V	Name=NA	0.001	0.964	25.2

4.11.7 Presentation of Family Pedigree with phenotypic features and segregation of variant

The individuals were from Poland and clinically diagnosed with seizures mild mental retardation, posteriorly rotated ears, a pit on ear helix, soft hands, MRI- hyper intense lesions and other phenotypic character. The members of second generations were denoted II with their respective DNA numbers II: 3 (45661), II: 1 (45664), II: 4 (45665) and II: 5 (47904) with Intellectual disability and unclassified epilepsy. The offspring's of third generations were III:3 (45666), III:1(45662) and III:2(45663) were with dysmorphic features, unclassified epileptic epilepsy, GGE and Intellectual disability.

After segregation of variant and the analysis predicted, a novel gene *ATP6V1B2* which is the cause for these clinical features in the affected members was found. All the Family members tested were found to be heterozygous with WTG as well as mutant allele except II: 5 (47904).

Mutation lead to Epilepsy, Intellectual disability and dysmorphic features in the affected family members across multiple generations. After the mutational analysis *ATP6V1B2* is real and relevant gene. The segregation of variants clearly showed the mutation in all the Family members tested, which were found to be heterozygous with WTG as well as mutant allele except II: 5 (J)-47904. Mutation lead to Epilepsy, Intellectual disability and dysmorphic features in the affected family members across multiple generations.

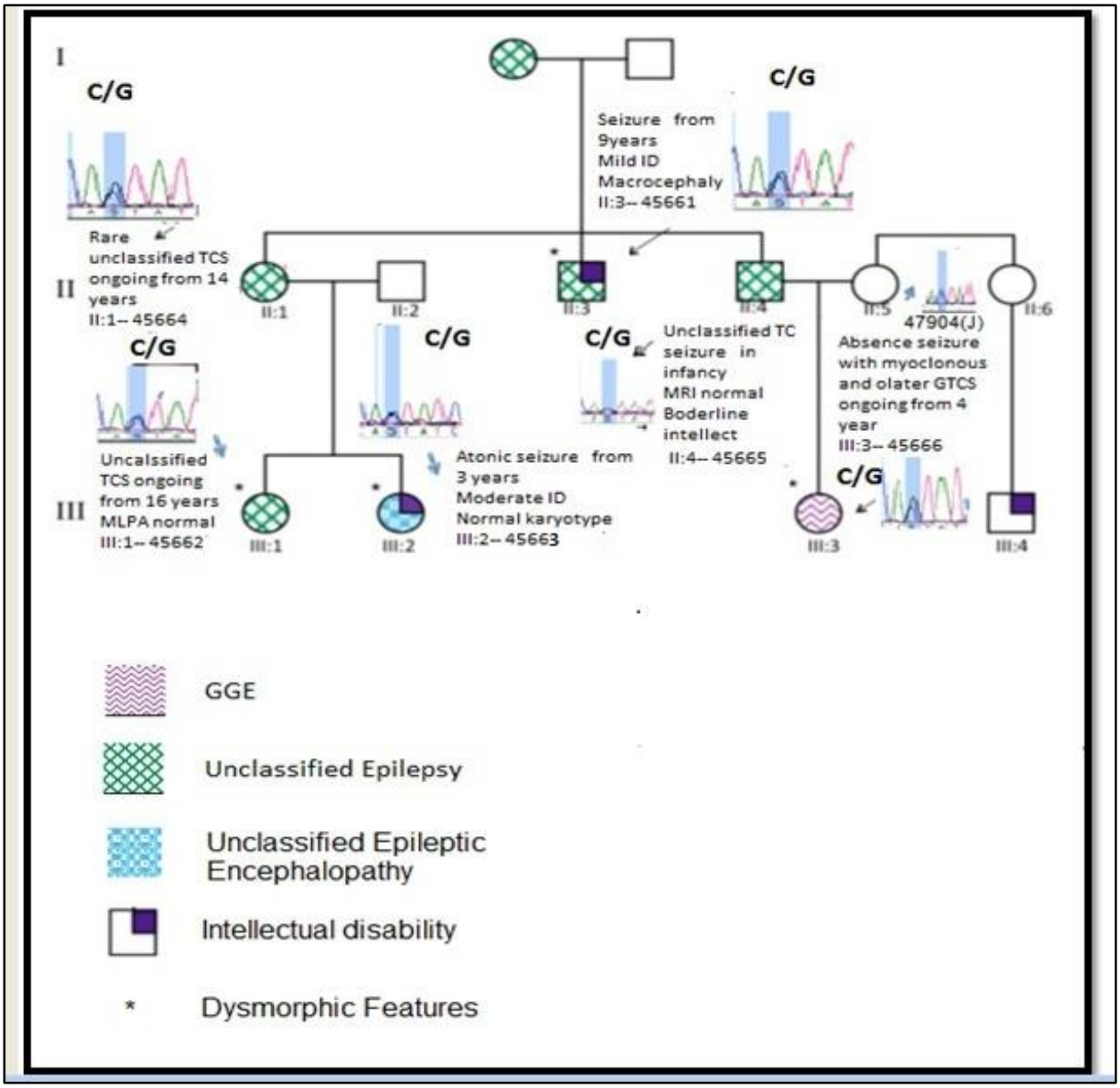

Fig 4.37: Showing segregation of variants in ATP6V1B2. The segregation of variants clearly showed the mutation in all the Family members tested was found to be heterozygous with Wild type genotype (WTG) as well as mutant allele except(II: 5)-47904.

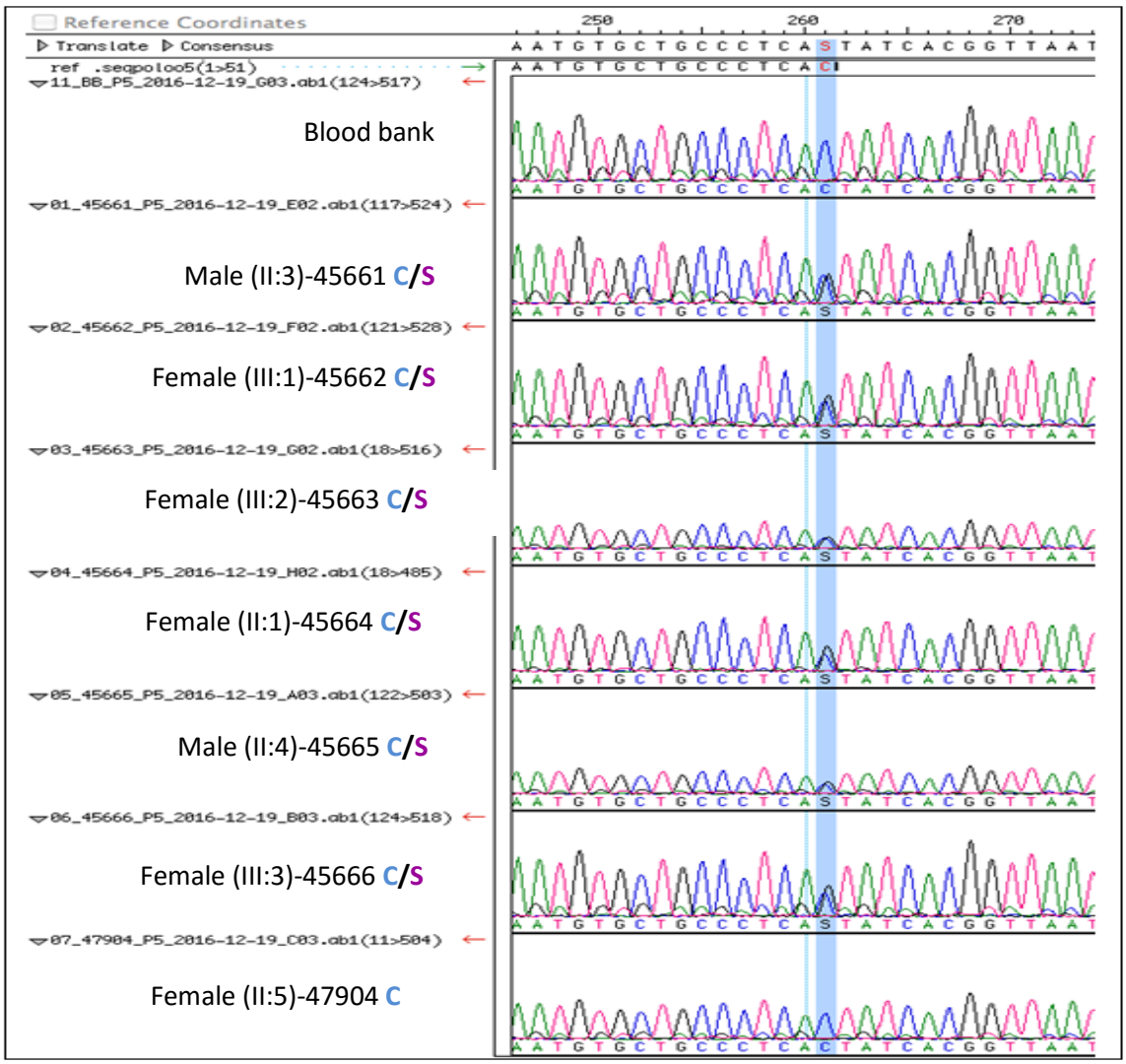

Fig 4.38: Chromatogram derived from targeted capillary sequencing for family 11 testing segregation of the C>G variant, c.C1192G:p.L398V of ATP6V1B2 on exon12. All the Family members tested were found to be heterozygous with WTG as well as mutant allele except II:5 47904.Reverse primer was used in sequencing for all the samples. The nucleotide code S(S-G or C) in DNA number 45661,45662,45663,45664,45665 and 45666 found to be heterozygous mutation encodes the base change C>G.

4.11.8 Visualization of the putative variant in Integrative Genomics Viewer (IGV) of *ATP6V1B2*.

After the prediction of the putative variant, IGV was used to predict the mutational analysis of gene. After the visualization of the variant c.C1192G:p.L398V of ATP6V1B2 on exon12, affected members with DNA number 45661 and 45662 were found with heterozygous mutation (C>G).

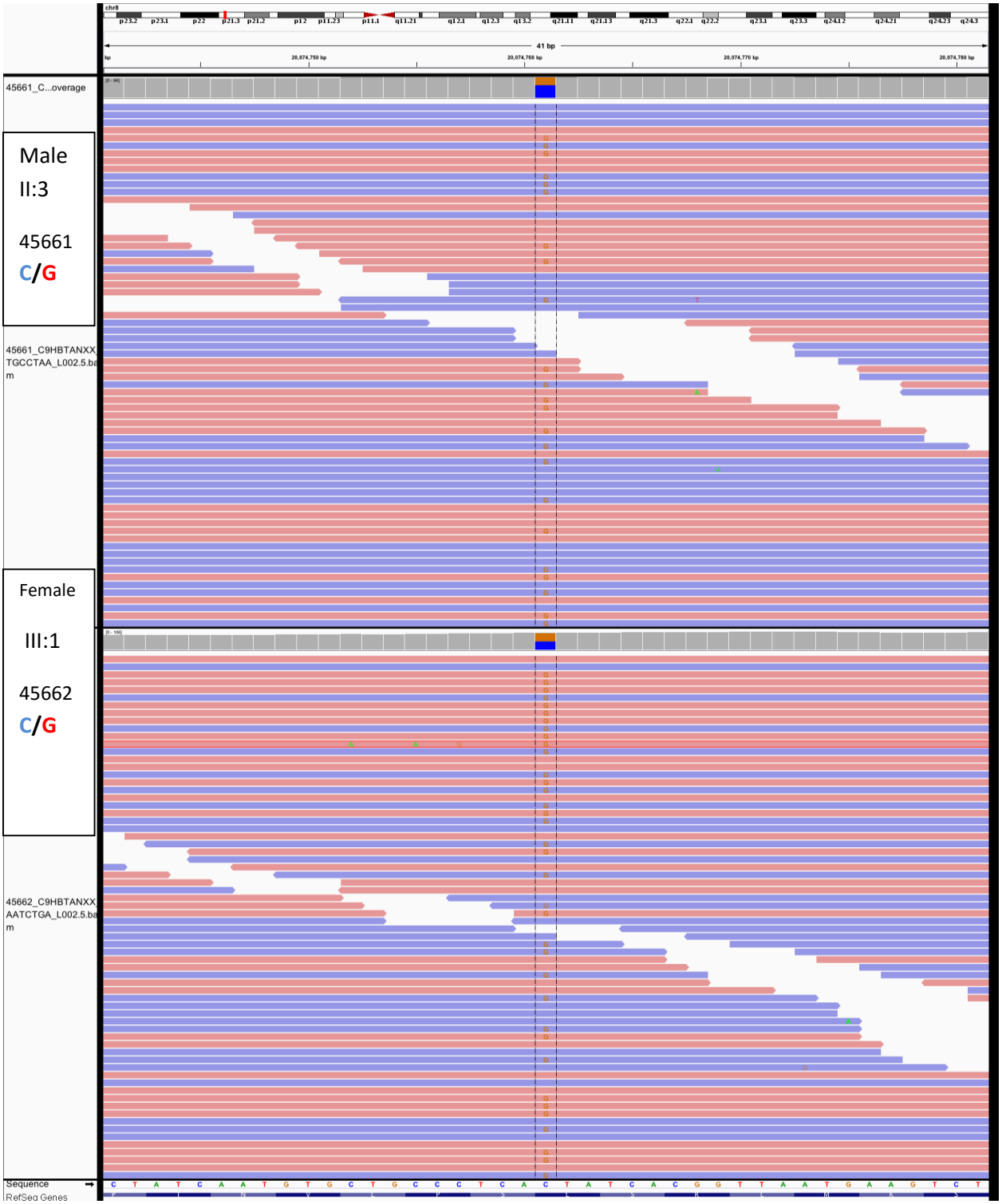

Fig 4.39: Visualisation of putative variant c.C1192G:p.L398V of ATP6V1B2 on exon12 in IGV. The IGV snapshot shows affected members with DNA number 45661 and 45662 being heterozygous with WTG as well as mutant allele for the mutation(chr8:20074761C>G).

CHAPTER 5

DISCUSSION

The causes of epilepsy are multiple, and may be influenced by several factors including the epilepsy syndrome. Most adult patients share different problems such as mental slowness, memory difficulties and attention deficits. In children, mental problems are more diffuse, responsible for language troubles, learning difficulties, poor academic outcome, behavior problems and finally unfortunate socio-professional prognosis. Generally, traditional methods have been used for screening and diagnosis of epileptic disorders. Due to advancement in science and technology, Next Generation Sequencing have been emerged as a cost effective and convenient tool in the field of genetic engineering. The fall in sequencing costs has exposed the major deficits in our understanding of the pathogenic significance of most variants Next-generation sequencing methods produce detailed information, including large variation datasets.

Every human genome contains about 3 million single nucleotide substitutions in comparison with a reference genome, but only a few of those are disease-related. 1000 Genomes project estimated that each genome codes for about 11,000 amino acid substitutions, approximately 12,500 synonymous substitutions, and hundreds of insertions and deletions (Abecasis *et al*., 2010). In addition, there are other more complex variations, fewer in number but challenging to interpret. As more personal genomes are sequenced and made publicly available, we will uncover more genomic variation. This will likely further illuminate how the totality of genomic variation accounts for polymorphic traits and both complex and Mendelian diseases. The situation reduces clinical utility and greatly affects the potential for disease treatment and prevention.

This work was mainly targeted to screen variants and identify gene causing the disease and showed that mutational changes in several genes leads to epilepsy which is possible to be transferred from generation to generation. Exome sequencing has been used extensively to diagnose novel diseases and find novel causative mutations for known disease phenotypes.

5.1 Whole Exome Sequencing

Exome sequencing revealed the availability of large databases of known SNPs of the affected members from eleven families with the informations of pathogenic variants. During the overall process of analysis of variants, nonputative variants found in these databases can generally be excluded. Some of nonputative variants found to be Nonsynonymous SNV characterised with low

frequency SNP and low CADD score. Variants were found associated with epiphyseal dysplasia. In family 2, *FUBP1* was found to be Nonsynonymous SNV characterized with high CADD score 34 and heterozygous in father and affected members. It was revealed, Far upstream element binding protein 1 aberrant expression of this gene has been found in malignant tissues, and this gene seemed to be important to neural system and lung development, quite broad expression moreover predominant in bone marrow .*TMEM125* was found to be stopgain encompassing p.W26X in genetic region with high CADD score 37 .It seems to have quite broad expression, not represented in brain atlas but highly expressed in brain in family 2 and only *POGZ* can be targeted, significantly reducing the variant pool. Excluding those variants, we obtained various novel genes after the validation of NGS results by capillary sequencing. The Exome Aggregation Consortium (ExAC) has created a user-friendly database containing the exome sequences of more than 60,000 unrelated individuals, which is freely available (Exome Aggregation Consortium (ExAC) 2014).

5.2 Bioinformatic Analysis of WES Data

The exomes have been analyzed using a uniform bioinformatic pipeline and were from individuals with Intellectual disability and X linked epileptic disease. The database provides information such as depth of coverage, genotype quality, allele frequency, and variant consequences. In family 2 *EVC2* (EvC Ciliary Complex Subunit 2) was found to be Nonsynonymous SNV with low frequency SNP (EXAC) and CADD score 23. *POGZ* was found more putative and real with Nonsynonymous SNV and CADD = 33 (no polyphen). *ATP6V1B2* was found to have CADD score 25.2, Nonsynonymous SNV and polyphen score of 0.964 implies that variant has probably damaging effect. The SIFT score 0.001 means it has deleterious effect.Filtering capabilities and large number of exomes will simplify the process of prioritizing variants when using exome sequencing as a diagnostic tool.

5.3 variation databases

When investigating the effects of variations, one has to find out whether the identified alteration has been previously found in patients or in healthy individuals. If a variant is frequent in healthy individuals in populations, possibly in several ethnic groups, it is likely

not harmful, although there are some exceptions. Central variation databases, such as the Human Gene Mutation Database (Stenson *et al*., 2014), Online Mendelian Inheritance in Man(OMIM) include variants for many human genes The UniProtKB/SwissProt database contains manually annotated protein entries with some variants (Yip *et al*., 2008). Variations can have their effects on many levels. The most studied are amino acid substitutions however, also other

types of variations as insertions and deletions are important. In family 3, variant was found de-novo in pro-band, however this family have two duplications (Xq and 18p segregating with various phenotypes). This means the amino acid change was actually (A or G) R>H (A or C or T) due to the adjacent base change Amino acid substitutions can have numerous effects and the mechanisms behind them are diverse (Steward *et al*., 2003).

5.4 Variant prioritization

Variations can have their effects on many levels. In our cases, amino acid substitutions and also other types of variations such as insertions and deletions were targeted where many unrelated, affected individuals were available which can be used to simplify analysis and search for common variants likely to affect gene function.

In family 1 the G>A variant was found putative leading to the mutation, mutation of the variant c.C2833T:p.E945K present in exon 17 was found affecting the gene function in family 2. In family 3,mutation was found in chromosome X with prioritization of variant G>A/G>T. In fourth family variant C>T /T>A was found putative. In fifth family variant was found in exon8, c.C953T:p.R318H which led to mutation in *ALG13*. Variant T>C was prioritized in sixth family and Sequencing of family 7 and 8 revealed segregation of the C>T variant and G>T respectively. The variant c.G1058A:p.A353V was found putative as the causing factor of mutation in *DPYSL3* in family 9. After prioritizing variant T>G revealed mutation in *ANK3* in family 10. After the segregation analysis of variant a novel gene *ATP6V1B2* was identified, this was not notified in relation with any epileptic gene.

We found variants relating to mutation which lead to identification of novel gene. Analysis examining how conserved an amino acid sequence is through evolution and between genes in a family can help to increase the confidence of a mutation having a deleterious effect (Wang *et al*., 2011).

5.5 Mutation discovery

We selected (Intellectual disability and X linked) disease-risk genes, on the basis of sequencing studies. Genes were selected and ranked according to functional category, *de novo* recurrences, overlap with a CNV morbidity maps, pathway connectivity and absence of *de novo* variants. We designed bioinformatic distributed across pools to cover all annotated RefSeq coding exons These genes were targeted for sequencing in large cohorts of cases (some including unaffected siblings) with a primary ascertainment diagnosis of the disorders with the identification of variants. In case of family 1, *PPP2R1A* was found putative relating to unusual facial features and intellectual disability. In case of family 2, affect different parts of body, Intellectual disability with

microcephaly which was proved to be the result of mutation of the variant c.C2833T:p.E945K present in exon 17.In family 3, two duplications Xq and 18p were observed segregating with various phenotypes revealing the mutation in *MAOA*. In fourth family the affected members have adjacent heterozygous changes with mutation of *DLG2*. After the validation of variant in exon 8, c.C953T:p.R318H, mutation in *ALG13* of fifth family was identified. The affected members of sixth family have adjacent heterozygous changes were found with mutation in *ALG13*. In case of family 7and 8 segregation of the C>T, G>T variant lead to the identification of mutation in *NCOR1* and *NALCN* respectively. The variant c.G1058A:p.A353V was found putative as the causing factor of mutation in *DPYSL3* in family 9. testing segregation of variant T>G revealed mutation in *ANK3*.Segregation of variant and the analysis predicted, a novel gene *ATP6V1B2* which is the cause for these clinical features in the affected members in family 11.

The mutational analysis was done excluding common variant and targeting potentially deleterious (nonsense, stop-gain, start-loss, frameshift, or disruptive splicing mutations) or missense events with a combined annotation-dependent depletion (CADD) score >20 in all the eleven families. The number of private, high-impact events identified in probands was significantly greater than that in unaffected siblings in the study.

5.5.1 WES and Bioinformatics Analysis of putative variants

Variants mostly likely to be protein-disrupting were Sanger validated nonsense, frameshift, or splice and missense variants with a CADD score > 20, read depth > 8X, a quality score > 20 and an allele count of < 3 to get ultra-rare and private deleterious events. Variations at functionally and structurally important sites have a more severe impact than those at nonfunctional sites.

5.5.2 Prediction of insertion and deletion

Insertions and deletions in coding regions can lead to three types of variations in a protein sequence which amphigoric alteration of the amino acid sequence due to RNA frameshift, insertion or deletion of one or more amino acid without changing the reading frame and termination of amino acid sequence by introducing an mRNA stop codon at the variation site or soon after it. SIFT Indel can predict the impact of non frameshifting insertions and deletions (Hu and Ng, 2013). MutationTaster2 can predict impacts of insertions, deletions, as well as amino acid substitutions (Schwarz *et al*., 2014).

CHAPTER 6

SUMMARY

Whole exome sequencing has very high potential on finding the genetic cause of various disorders and leads to the identification of novel genes through validation of variants. The affected members from eleven families were with recurrent seizures, mild retardation problems, learning difficulties and dysmorphic features. It is important to have a comprehensive and critical analysis of laboratory results, clinical phenotype and genotype in patients and disorders related to epilepsy. After the set of WES performed in South Australian patients from family 1 to 10 and polish family, family 11 based in intellectual disability result various variants and mutations. To validate the variants we performed capillary sequencing after analyzing those variants with different tools.

In case of family 1, 33170 had the G>A variant and found heterozygous with both WTG and mutant allele. 21298 was found to be homozygous WTG leading to the prediction of gene Protein Phosphatase (formerly 2A), Regulatory (Subunit PR 65) Alpha isomer *(PPP2R1A*) to be putative. In family 1, members were found to be associated with a variety of psychiatric disorders, mental retardation including antisocial behavior in the affected samples which intersect with diseases associated with *PPP2R1A.*

In case of family 2, the affected member were characterized by mildly unusual facial features, short stature, heart defects, bleeding and affect different parts of body, Intellectual disability with microcephaly which was proved to be the result of mutation of the variant c.C2833T:p.E945K present in exon 17 with mutation of *POGZ*.

In case of family 3, Variant was found de-novo in pro-band; however this family has two duplications, Xq and 18p segregating with various phenotypes. Male patients are affected by borderline mental retardation and exhibit abnormal behavior, including disturbed regulation of impulsive aggression. Obligate female carriers have normal intelligence and behavior (ECO: 0000269 PubMed: 8211186).This phenotypic character resemble to affected members in family 3 with the mutation of MAOA. .In fourth family the affected members have adjacent heterozygous changes with wild type and mutant allele C>T /T>A leading to mutation of *DLG2.* Mutation of *DLG2* results in Brunner syndrome which is not associated with this family. This gene has also been associated with a variety of other psychiatric disorders which resemble to that of affected members. After the validation of variant in exon 8, c.C953T:p.R318H it was clear about the disorder created due to mutation in *ALG13* of fifth family. The affected members of sixth

family have adjacent heterozygous changes with wild type and mutant allele T>C concluded that was probably contributing to learning difficulties with mutation in *GRIA1*. Disturbing neurophysiologic processes, results in neurological disorder. After testing segregation of the T>C variant confirmed mutation in *GRIA1*.Sequencing of family 7and 8 revealed segregation of the C>T variant and *NALCN*, chr13:101844315 G>T respectively. The variant c.G1058A:p.A353V was found putative as the causing factor of mutation in *DPYSL3* in family 9. Testing segregation of variant T>G revealed mutation in *ANK3* (Ankyrin 3) after viewing functional characteristics and assessing biological effect of variants based on literature reviews and online tools in family 10.

The variants from these 10 families were validated after variant calling and prioritization as preliminary works were already done, sequencing and validation was done which revealed those genes to be real. In family 11, segregation of variant and the analysis predicted, a novel gene *ATP6V1B2* which is the cause for these clinical features in the affected members. All the Family members tested were found to be heterozygous with WTG as well as mutant allele except 47904.This gene has never been listed till date in relation to epilepsy and was found real.

Molecular analysis of mutation of epileptic gene can help in confirming uncertain phenotypical diagnosis or solving the dilemma of differential diagnosis and can provide supplementary genotype-phenotype information. Furthermore, WES can also serve as the efficient technique for carrier detection and prenatal diagnosis.

CHAPTER 7

CONCLUSION

This work shows that different phenotypical characteristics among epilepsy patients are due to an almost breath taking diversity of mutations that have been characterized in the genes. This work was mainly focused on finding out of novel mutations. After the segregation of variants in all the families, the variants were found to be real causing mutations.

After finding out different mutations and sequence variation in the patients with similar phenotypic data, this suggests useful information in pharmacogenomics. Insights gained from mutations and their effects from the molecular pathology of disease will reveal both refined and major complexities at the genetic, protein levels and ion channels. Further, helps to functional, cell and molecular biology studies of some of these new genes or variants using various types of cells including neurons or stem cells. Similarly, insights gained from polymorphisms and their ethnic variations will hint at a complex evolution awaiting further exploration. Finally, insights gained from the investigation of patient genetics and an etiology of the inhibitor complication will serve as paving the path for gene therapy in previously untreated patients.

The overall process leads to the finding of novel genes. I resolved one large polish family, with epilepsy through the identification of a novel variant in a gene, which was previously not involved in epilepsy.

CHAPTER 8

RECOMMENDATIONS AND LIMITATIONS

At present, genetic testing is indicated for specific disorders, as outlined above, where the finding of a genetic mutation may save further potentially invasive investigations into the underlying cause and may influence patient management and genetic counseling. In our study the identification of deleterious molecular changes with validation and screening of variants lead to mutation analysis and identification of novel genes. The work which identified mutation in epileptic patients may help to:

- Unravel the underlying genetic lesions in epileptic patients.
- After the verification of mutation of genes, the study of molecular domains and the relationship with pathways becomes easier leading to easy means of molecular analysis.
- Interpretation of protein domain and their linkage associated with disorders.
- Furthermore, carrier status among the family members can be checked by WES in an efficient way both on economic aspect as well as on time factor.

Still their interpretation remains complex. It is not worth investing in whole exome or whole genome sequencing of patients ($1500- $1800per sample) despite the relatively low cost ($2,000–$5,000) often delivered with a basic bioinformatic analysis thrown in) as we will be faced with an interpretable set of genetic variants and have considerable difficulty explaining the findings. Success has been achieved in identifying mutations in all of the family but dissecting the genetic basis of common epilepsy syndromes has proven more difficult. Moreover common epilepsies are genetically complex disorders believed to be influenced by variation in several susceptibility genes. It was difficult to put all information together due to the family members belonging to Australia and had to follow terms and conditions of not showing pedigree of families and pictures of mutation in genes. Some of the major limitations are given below:

- Difficulty in dealing with the findings and following the molecular level of work.
- Understanding the genetic basis of common epilepsy syndromes has proven more difficult.
- Working under terms and conditions.
- Time limit
- Studies of molecular level work of mutation in novel genes still left.

REFERENCES

Andrade, D. M., Bassett, A. S., Bercovici, E., Borlot, F., Bui, E., Camfield, P., ... & Greenaway, J. (2017). Epilepsy: Transition from pediatric to adult care. Recommendations of the Ontario epilepsy implementation task force. Epilepsia.

Ali, A., Akram, F., Khan, G., & Hussain, S. (2017). Paediatrics brain imaging in epilepsy: common presenting symptoms and spectrum of abnormalities detected on MRI. Journal of Ayub Medical College Abbottabad, 29(2), 215-218.

Al-Sayed, M. D., Al-Zaidan, H., Albakheet, A., Hakami, H., Kenana, R., Al-Yafee, Y., ... & Al-Qubbaj, W. (2013). Mutations in NALCN cause an autosomal-recessive syndrome with severe hypotonia, speech impairment, and cognitive delay. The American Journal of Human Genetics, 93(4), 721-726.

Benarroch, E. E. (2011). Na+, K+-ATPase Functions in the nervous system and involvement in neurologic disease. Neurology, 76(3), 287-293.

Beth, S. L., Underhill, D. M., Cran, K. M., & Gluck, S. L. (1995). Transcriptional Regulation of the Vacuolar H+-ATPase B2 Subunit Gene in Differentiating THP-1 Cells. The American Society for Biochemistry and Molecular Biology, Inc.

Bernasconi, P.,Rausch, T., Struve, I., Morgan, L., & Taiz, L. (1990). An mRNA from human brain encodes an isoform of the B subunit of the vacuolar H (+)-ATPase. Journal of Biological Chemistry, 265(29), 17428-17431.

Bissar-Tadmouri, N., Donahue, W. L., Al-Gazali, L., Nelson, S. F., Bayrak-Toydemir, P., & Kantarci, S. (2014). X chromosome exome sequencing reveals a novel ALG13 mutation in a nonsyndromic intellectual disability family with multiple affected male siblings. American Journal of Medical Genetics Part A, 164(1), 164-169.

Basel-Vanagaite, L., Hershkovitz, T., Heyman, E., Raspall-Chaure, M., Kakar, N., Smirin-Yosef, P., ... & Lagovsky, I. (2013). Biallelic SZT2 mutations cause infantile encephalopathy with epilepsy and dysmorphic corpus callosum. The American Journal of Human Genetics, 93(3), 524-529.

Bamshad, M. J., Shendure, J. A., Valle, D., Hamosh, A., Lupski, J. R., Gibbs, R. A., ... & Mane, S. (2012). The Centers for Mendelian Genomics: A new large-scale initiative to identify the genes underlying rare Mendelian conditions. American journal of medical genetics Part A, 158(7), 1523-1525.

Benbadis, S.R.(2006).The EEG in nonepileptic seizures. Journal of clinical neurophysiology, 23(4), 340-352.

Carton, S., Thompson, P. J., & Duncan, J. S. (2003). Non-epileptic seizures: patients' understanding and reaction to the diagnosis and impact on outcome. Seizure, 12(5), 287-294.

Chen, D. K., Sharma, E., & LaFrance, W. C. (2017). Psychogenic Non-Epileptic Seizures. Current Neurology and Neuroscience Reports, 17(9), 71.

Candace T. Myers, H. C. (2015). Advancing epilepsy genetics in the genomic era.

Cowan, L. D. (2002). The epidemiology of the epilepsies in children. Developmental Disabilities Research Reviews, 8(3), 171-181.

Conway, E., Devinsky, O., & Glick, C. S. (2015). Epilepsy in Children: What Every Parent Needs to Know. Springer Publishing Company.

Chong, J. X., McMillin, M. J., Shively, K. M., Beck, A. E., Marvin, C. T., Armenteros, J. R., ... & Bocian, M. (2015). De novo mutations in NALCN cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. The American Journal of Human Genetics, 96(3), 462-473.

Chen, D. T., Jiang, X., Akula, N., Shugart, Y. Y., Wendland, J. R., Steele, C. J. M., ... & Cheng, A. (2013). Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. Molecular psychiatry, 18(2), 195.

Ceulemans, S., van der Ven, K., & Del-Favero, J. (2012). Targeted screening and validation of copy number variations. Genomic Structural Variants: Methods and Protocols, 311-328.

Coe, B. P., Witherspoon, K., Rosenfeld, J. A., Van Bon, B. W., Vulto-van Silfhout, A. T., Bosco, P., ... & Schuurs-Hoeijmakers, J. H. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nature genetics, 46(10), 1063-1071.

Carvill, G. L., Heavin, S. B., Yendle, S. C., McMahon, J. M., O'Roak, B. J., Cook, J., ... & Malone, S. (2013). Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nature genetics, 45(7), 825-830.

Chien, C. C., Lin, C. H., Chang, Y. Y., & Lung, F. W. (2010). Association of VNTR polymorphisms in the MAOA promoter and DRD4 exon 3 with heroin dependence in male Chinese addicts. The World Journal of Biological Psychiatry, 11(2-2), 409-416.

Caveness, W. F., AM LORENTZ, H. J. M., & RADERMECKER, J. (1964). A proposed international classification of epileptic seizures. Epilepsia, 5, 297-306.

Dannlowski, U., Ohrmann, P., Konrad, C., Domschke, K., Bauer, J., Kugel, H., ... & Mortensen, L. S. (2009). Reduced amygdala–prefrontal coupling in major depression: association with MAOA genotype and illness severity. International Journal of Neuropsychopharmacology, 12(1), 11-22.

Demirkan, A., van Duijn, C. M., Ugocsai, P., Isaacs, A., Pramstaller, P. P., Liebisch, G., ... & Kirichenko, A. V. (2012). Genome-wide association study identifies novel loci associated with circulating phospho-and sphingolipid concentrations. PLoS genetics, 8(2), e1002490.

Edward B Bromfield, J. E. (2006). An Introduction to Epilepsy. American Epilepsy Society.

Eichler, E. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay.

Fan, M., Liu, B., Jiang, T., Jiang, X., Zhao, H., & Zhang, J. (2010). Meta-analysis of the association between the monoamine oxidase-A gene and mood disorders. Psychiatric genetics, 20(1), 1-7.

Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., ... & Smoller, J. W. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nature genetics, 40(9), 1056-1058.

Flink, R., Pedersen, B., Guekht, A. B., Malmgren, K., Michelucci, R., Neville, B., ... & Özkara, C. (2002). Guidelines for the use of EEG methodology in the diagnosis of epilepsy. Acta Neurologica Scandinavica, 106(1), 1-7.

Gan, S., Ni, W., Dong, Y., Wang, N., & Wu, Z. (2015). Population genetics and new insight into range of CAG repeats of spinocerebellar ataxia type 3 in the Han Chinese population. Us National Library of Medicine National Institute of Health .

Global Campaign against Epilepsy, Programme for Neurological Diseases, Neuroscience (World Health Organization), International Bureau of Epilepsy, & International League against Epilepsy. (2005). Atlas: Epilepsy Care in the World. World Health Organization.

Gécz, J.,Cloosterman, D., & Partington, M. (2006). ARX: a gene for all seasons. Current opinion in genetics & development 16(3), 308-316.

Gan, S. R., Ni, W., Dong, Y., Wang, N., & Wu, Z. Y. (2015). Population genetics and new insight into range of CAG repeats of spinocerebellar ataxia type 3 in the Han Chinese population. PLoS one, 10(8), e0134405.

Hirata, T.,Iwamoto-Kihara, A., Sun-Wada, G. H., Okajima, T., Wada, Y., & Futai, M. (2003). Subunit rotation of vacuolar-type proton pumping ATPase relative rotation of the G and c subunits. Journal of Biological Chemistry, 278(26), 23714-23719.

Heron, S. E., Crossland, K. M., Andermann, E., Phillips, H. A., Hall, A. J., Bleasel, A., ... & Mulley, J. C. (2002). Sodium-channel defects in benign familial neonatal-infantile seizures. The Lancet, 360(9336), 851-852.

Iqbal, Z., Vandeweyer, G., van der Voet, M., Waryah, A. M., Zahoor, M. Y., Besseling, J. A., ... & Van der Aa, N. (2013). Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. Human molecular genetics, 22(10), 1960-1970.

Kim, E., Cho, K. O., Rothschild, A., & Sheng, M. (1996). Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins. Neuron, 17(1), 103-113.

Kuzniecky, R. I. (2005). Neuroimaging of epilepsy: therapeutic implications. NeuroRx, 2(2), 384-393.

Kerner, B., Jasinska, A. J., DeYoung, J., Almonte, M., Choi, O. W., & Freimer, N. B. (2009). Polymorphisms in the GRIA1 gene region in psychotic bipolar disorder. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150(1), 24-32.

Köroğlu, Ç., Seven, M., & Tolun, A. (2013). Recessive truncating NALCN mutation in infantile neuroaxonal dystrophy with facial dysmorphism. Journal of medical genetics, jmedgenet-2013.

Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., ... & NHLBI Exome Sequencing Project. (2012). Copy number variation detection and genotyping from exome sequence data. Genome research, 22(8), 1525-1532.

Kortüm, F., Caputo, V., Bauer, C. K., Stella, L., Ciolfi, A., Alawi, M., ... & Grammatico, P. (2015). Mutations in KCNH1 and ATP6V1B2 cause Zimmermann-Laband syndrome. Nature genetics, 47(6), 661-667.

Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., ... & Morgan, M. (2011). Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell, 145(7), 1036-1048.

Leyland, M. L., & Dart, C. (2004). An alternatively spliced isoform of PSD-93/chapsyn 110 binds to the inwardly rectifying potassium channel, Kir2. 1. Journal of Biological Chemistry, 279(42), 43427-43436.

Lavy,S.,Carmon, A., & Yahr, I. (1972). Assessment of a clinical and electroencephalographic classification of epileptic patients in everyday neurological practice A survey of 450 cases. Epilepsia, 13(4), 498-505.

Lupton, C. J., Steer, D. L., Wintrode, P. L., Bottomley, S. P., Hughes, V. A., & Ellisdon, A. M. (2015). Enhanced molecular mobility of ordinarily structured regions drives polyglutamine disease. Journal of Biological Chemistry, 290(40), 24190-24200.

Lu, B., Su, Y., Das, S., Liu, J., Xia, J., & Ren, D. (2007). The neuronal channel NALCN contributes resting sodium permeability and is required for normal respiratory rhythm. Cell, 129(2), 371-383.

Liu, Y., Zhang, Y., & Wang, J. H. (2014). Crystal structure of human Ankyrin G death domain. Proteins: Structure, Function, and Bioinformatics, 82(12), 3476-3482.

LaFrance, W. C., & Benbadis, S. R. (2006). Avoiding the costs of unrecognized psychological nonepileptic seizures. Neurology, 66(11), 1620-1621.

Martin, R. C., Gilliam, F. G., Kilgore, M., Faught, E., & Kuzniecky, R. (1998). Improved health care resource utilization following video-EEG-confirmed diagnosis of nonepileptic psychogenic seizures. Seizure, 7(5), 385-390.

Mulley, J. C., & Mefford, H. C. (2011). Epilepsy and the new cytogenetics. Epilepsia, 52(3), 423-432.

Myers , C. T., & Mefford, H. C. (2015). Advancing epilepsy genetics in the genomic era. Genome Medicine .

Møller, R. S., Dahl, H. A., & Helbig, I. (2015). The contribution of next generation sequencing to epilepsy genetics. Expert review of molecular diagnostics, 15(12), 1531-1538.

Mevissen, T. E., Hospenthal, M. K., Geurink, P. P., Elliott, P. R., Akutsu, M., Arnaudo, N., ... & Freund, S. M. (2013). OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. Cell, 154(1), 169-184.

Martin, M. S., Tang, B., Papale, L. A., Yu, F. H., Catterall, W. A., & Escayg, A. (2007). The voltage-gated sodium channel Scn8a is a genetic modifier of severe myoclonic epilepsy of infancy. Human molecular genetics, 16(23), 2892-2899.

Myers, C. T., & Mefford, H. C. (2015). Advancing epilepsy genetics in the genomic era. Genome medicine, 7(1), 91.

Magri, C., Gardella, R., Barlati, S. D., Podavini, D., Iatropoulos, P., Bonomi, S., ... & Barlati, S. (2006). Glutamate AMPA receptor subunit 1 gene (GRIA1) and DSM-IV-TR schizophrenia: A pilot case-control association study in an Italian sample. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 141(3), 287-293.

Nozawa, R. S., Nagao, K., Masuda, H. T., Iwasaki, O., Hirota, T., Nozaki, N., ... & Obuse, C. (2010). Human POGZ modulates dissociation of HP1 [alpha] from mitotic chromosome arms through Aurora B activation. Nature cell biology, 12(7), 719.

Pella, R. D., Mukundan, L., & Szabo, C. A. (2014). Juvenile Myoclonic Epilepsy—A Maturation Syndrome Coming of Age. In Epilepsy Topics.

Palmer, N. D., Mychaleckyj, J. C., Langefeld, C. D., Ziegler, J. T., Williams, A. H., Bryer-Ash, M., & Bowden, D. W. (2010). Evaluation of DLG2 as a positional candidate for disposition index in African-Americans from the IRAS Family Study. Diabetes research and clinical practice, 87(1), 69-76.

Qian, Q. J., Yang, L., Wang, Y. F., Zhang, H. B., Guan, L. L., Chen, Y., ... & Faraone, S. V. (2010). Gene–gene interaction between COMT and MAOA potentially predicts the intelligence of attention-deficit hyperactivity disorder boys in China. Behavior genetics, 40(3), 357-365.

Robert S. Fisher, J. H. (2016). Operational Classification of Seizure Types by the International League Against Epilepsy.

Ruggeri, F., Longo , G., Faggiano, S., Lipiec, E., Pastore , A., & Dietler, G. (2015). Infrared nanospectroscopy characterization of oligomeric and fibrillar aggregates during amyloid formation. US National Library of Medicine National Institutes of Health .

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., ... & Frankish, A. (2005). The DNA sequence of the human X chromosome. Nature, 434(7031), 325.

Ruggeri, F. S., Longo, G., Faggiano, S., Lipiec, E., Pastore, A., & Dietler, G. (2015). Infrared nanospectroscopy characterization of oligomeric and fibrillar aggregates during amyloid formation. Nature communications, 6, 7831.

Reuber, M., Fernandez, G., Bauer, J., Helmstaedter, C., & Elger, C. E. (2002). Diagnostic delay in psychogenic nonepileptic seizures. Neurology, 58(3), 493-495.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... & Voelkerding, K. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine: official journal of the American College of Medical Genetics, 17(5), 405-424.

Silva, U. C. A., Marques, W., Lourenço, C. M., Hallak, J. E. C., & Osório, F. L. (2015). Psychiatric disorders, spinocerebellar ataxia type 3 and CAG expansion. Journal of neurology, 262(7), 1777-1779.

Smith, D., Defalla, B. A., & Chadwick, D. W. (1999). The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. Qjm, 92(1), 15-23.

Siesjö, B. K. (1992). Pathophysiology and treatment of focal cerebral ischemia: Part I: Pathophysiology. Journal of neurosurgery, 77(2), 169-184.

Schulze, T. G., Detera-Wadleigh, S. D., Akula, N., Gupta, A., Kassem, L., Steele, J., ... & Propping, P. (2009). Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. Molecular psychiatry, 14(5), 487.

Sharma, S., & Prasad, A. N. (2017). Inborn Errors of Metabolism and Epilepsy: Current Understanding, Diagnosis, and Treatment Approaches. International Journal of Molecular Sciences, 18(7), 1384.

Sharma, S., & Dixit, V. (2013). Epilepsy–a comprehensive review. Int J Pharma Res. Rev., 2(12), 61-80.

Siegelbaum, S. A., & Hudspeth, A. J. (2000). Principles of neural science(Vol. 4, pp. 1227-1246). E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.). New York: McGraw-hill.

Singh, N. A., Pappas, C., Dahle, E. J., Claes, L. R., Pruess, T. H., De Jonghe, P., ... & White, H. S. (2009). A role of SCN9A in human epilepsies, as a cause of febrile seizures and as a potential modifier of Dravet syndrome. PLoS genetics, 5(9), e1000649.

Stafstrom, C. E., & Carmant, L. (2015). Seizures and epilepsy: an overview for neuroscientists. Cold Spring Harbor perspectives in medicine, 5(6), a022426.

Timal, S., Hoischen, A., Lehle, L., Adamowicz, M., Huijben, K., Sykut-Cegielska, J., ... & Gilissen, C. (2012). Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. Human molecular genetics, 21(19), 4151-4161.

Uanda , S. C., Wilson , M., Charles , L. M., Jaime , E. C., & Flávia , O. L. (2015). Psychiatric disorders, spinocerebellar ataxia type 3 and CAG expansion. Journal of Neurology , 1777.

Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., ... & van Bon, B. W. (2010). A de novo paradigm for mental retardation. Nature genetics, 42(12), 1109-1112.

Wang, J., Hoshino, T., Redner, R. L., Kajigaya, S., & Liu, J. M. (1998). ETO, fusion partner in t (8; 21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proceedings of the National Academy of Sciences, 95(18), 10860-10865.

Xu, X., Yang, X., Wu, Q., Liu, A., Yang, X., Ye, A. Y., ... & Wang, S. (2015). Amplicon resequencing identified parental mosaicism for approximately 10% of “de novo” SCN1A mutations in children with Dravet syndrome. Human mutation, 36(9), 861-872.

Yuan, Y., Zhang, J., Chang, Q., Zeng, J., & Wu, J. (2014). De novo mutation in ATP6V1B2 impairs lysosome acidification and causes dominant deafness-onychodystrophy syndrome. National Library of Medicine National Institutes of Health .

www.epilepsy.com/article/2016/12/2017-revised-classification-seizures

www.bbc.co.uk/.../epilepsy1.shtml
www.lucigen.com/illumina-dnalibraryprep
https://support.illumina.com/.../samplepreps_truseq/truseqexome/truseq-exome-library.
https://support.illumina.com/sequencing/sequencing_kits/truseq-exome-kit.htm
www.qiagen.com/NGS/Library-Prep
www.openbioinformatics.org/annovar/annovar_filter.htm
annovar.openbioinformatics.org/

APPENDICES

Appendix 1: Reagents and Composition

1.1 Gel electrophoresis

Quality of the DNA extracted from the samples was checked by agarose gel electrophoresis. 1 % agarose gel was prepared by dissolving 1 gm of agarose in 100 mL of 1X TBE buffer.

- Agarose gel -1%
- Buffer-TBE Tris/Borate/ *EDTA* (1%)
- Volt -110v

❖ Composition of 1x TBE buffer

■ 89 mM Tris (pH 7.6) ■ 89 mM boric acid ■ 2 mM EDTA

1.2 PCR Master MiX

The PCR components were mixed together according to manufacturer provided protocol (KAPPA) in appropriate volume to finalize the volume of 50µl in all of the eleven families. Sample (Blood bank) was used as positive control and nuclease free water was used as negative control.

1.3 Master Mix for sequencing reaction

Composition of sequencing reagents were mixed together according to manufacturer provided protocol in appropriate volume to finalize the volume of 10µl in all of the eleven families sequencing was performed with Big Dye Terminator cycle sequencing Kit.

Appendix 2: Design of primer after validation of putative variant and gene

Primers were designed for different genes by primer Blast and primer sequences were available on request. The mutations of genes were verified by capillary sequencing after screening and validation of variants. The segregation of variant was check among family members with epileptic disorders.

2.1 Primer design for *PPP2R1A (*010320.10) of family 1

GAATGATGTAATCGTCAGCACTTAGCATTGACTAGATTTATTATGC**GCAATCTGCGAAGTGTCTCA**CACACACTATTTTCATTTA
AACCTCATGCGGACCTGTGGGGTAGGTACTGTTACTATCAGCTCCGTTTCATAGGGCTGGGAAGACAGAGAGGGGGTCATCA
CTTGCCCAAGGTCATTCAGCTAAAACCTGGACCCACACAACTGCAGAGTCTGTGCTTGCTCCTCTCTGCCATACTGCCTGCTGCC
TCAGGATCCCCGTCCCCGACTCCCAGGTACTTCCGGAACCTGT***A***CTCAGATGACACCCCCATGGTGCGGCGGGCCGCAGCCTC
CAAGCTGGGGGAGTTTGCCAAGGTGCTGGAGCTGGACAACGTCAAGAGTGAGATCATCCCCATGTTCTCCAACCTGGCCTCTG
ACGAGCAGGTGAGTTTTGCTTCCTGGCCCTCTGCTCTCCCGTCCTTCTGGTGGTTCCTGCCCATGAAAGAGAATCCCAGAGCTC
AGCAAGGCCTCTGCTGCCCTCCCACTGTTCCTC**TCCTCTCCCTAGGACTCGGT**GCGGCTGCTGGCGGTGGAGGCGTGCGTGAA
CATCGCCCAGCTTCTGCCCCAGGAGGATCTGGAGGCCCTGGT

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	GCAATCTGCGAAGTGTCTCA	20	52
Reverse	ACCGAGTCCTAGGGAGAGGA	20	52.3

Fig 1: Primer designed for *PPP2R1A*. Primer was designed against segregation of the G>A variant of *PPP2R1A* of family 1 of Father (33170). The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.2 Primer design for *POGZ* (046601.1) of family 2

CCGAATGTCTGAATGTTGA**TGATCAGGATGAAGGGAGCC**CAGTCACCCAAGAACCTGAGCTAGCATCAGGTGGTGGTGGTA
GTGGTGGAGTTGGCAAAAAGGAGCAGCTGTCTGTGAAGAAGCTTCGAGTAGTACTGTTTGCTCTATGCTGCAATACAGAACA
GGCAGCTGAACACTTCCGAAATCCCCAGCGACGTATTCGCCGTTGGCTTCGACGTTTCCAGGCCTCCCAGGGGGAGAATCTAG
AGGGCAAATATCTGAGCTTTGAGGCAGAAGAGAAACTGGCTGAGTGGGTGCTAACCCAGCGCGAACAACAGCTACCTGTAAA
TGAGGAGACCTTGTTCCAGAAGGCCACCAAAATAGGACGTTCTTTGGAAGGGGGGTTTAAGATCTCCTATGAGTGGGCTGTG
CGTTTCATGC***T***GCGGCACCACCTGACTCCCCATGCCCGGCGAGCTGTGGCCCACACCCTACCTAAGGATGTAGCAGAGAATGC
AGGACTCTTCATTGATTTTGTACAACGGCAGATTCACAACCAGGACTTACCCTTGTCTATGATTGTGGCTATTGATGAGATCTCT
TTGTTCCTGGATACAGAGGT**GCTGAGCAGTGATGATCGAA**AGGAGAATGCC

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TGATCAGGATGAAGGGAGCC	20	61
Reverse	TTCGATCATCACTGCTCAGC	20	60

Fig 2: Primer designed for *POGZ*. Primer was designed against including c.C2833T:p.E945K , from mother (45269) C>T of family 2 present on exon 17 of *POGZ*. The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.3 Primer design for *MAOA* (008957.2) of family 3

ATA**GGCCTTGTGTCTTTTGCAGTT**TCTAACCAGATATAATTCAAGCAGTAGCATAATTAATGCTGACATGTTTCCGC
CTCAGTCTTCCAAATTACCAGCCTTGGTTGACCTACCTTGATCTGTTTTGTTGCCTCACAGTTGCCTCATTTTCTCATT
TTGTATGTTTTTACTGCTCAATGTTGTTTAGAAGTGATAAAGATGATTTTTCACGCCTGCCACAAAGACTGCAGCTC
ACATTTGAGGTAATATTTTGCTTGTGTGTGTTTTAGTGCAAATACGTAATTAATGCGATCCCTCCGACCTTGACTGC
CAAGATTCACTTCAGACCAGAGCTTCCAGCAGAGAGAAACCAGTTAATTCAGC***AT***CTTCCAATGGGAGCTGTCATT
AAGTGCATGATGTATTACAAGGAGGCCTTCTGGAAGAAGAAGGGTAGGCTGCTATTATTCATGTTTAAACTGTATT
ATATGAAGAAATCACAGTCTTTTAGGATTATTGAACAGAAGGTATTGAACAGAAACAAAGTATTTTCAAACTGGTA
GAAAAAGGATTAGAAATCTTGGTCTGTCAATTTCCTCATAT**CCTTGGCCACACATAATGACC**CCAAGAGCACTTGT
TGGCAATGGGAGGGAAGAAGGAGATCACATCAGTCATAAGGCCACCATTGCCCTGACTCCTGGCATCTGTCC

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	GGCCTTGTGTCTTTTGCAGTT	21	56.9
Reverse	GGTCATTATGTGTGGCCAAGG	19	62.4

Fig 3: Primer designed for *MAOA*. Primer was designed against including the variant p.1673Tof *MAOA,* segregation of the G>A/G>T variant from Son (33881) of family 3. The variant is highlighted green. Forward primer is highlighted yellow and reverse is highlighted with blue color and shown by using arrow.

2.4 Primer design for *DLG2* (021375.1) of family 4

CTCTTCTACCATGAGGATGTCAAGTTTTCAGAGAGAGAGAAAAAAAATAACTTCCTCTAACAAAATAATTCAGGAATGCTATCC
CCTTGGTTTACAACTAAAGTAGTAAACCTGCCTTAATAAAATCAAGGTTGTGTAT**TGGGGCCTCAGATAGTGAGTC**AAAGACA
AAGCCCTCTTCCTACTTTTTCCAAGTCAGTGAGAATACTTTATTGAAATTAGAACAAAGATTAATGACATTGCTCTAAGGTGATT
CTCTGCTTGGATGTTTTAAATTATTAAATCAAAATTTACAATTGCAATCAAGAAAGAGTAATGACCAAAAAAAAAAAAAAAATGC
AGTTTGCAACCTAGAAACCTTCTTCAGGGGAGGTGTAAGAAGATTGGTGAACATACTCTTTCTGCTACAAATCTCACAGACTGC
ACACTGGTTCCATATAAATTGTCATTGTACTGGCCGGCTTCTATAAACTTGTGCTCTTGGA***TA***TCTTTCTCCATTTGTTCTCTGGA
AATGACAAAGTGATAGTCTCTGCCATCCACCTCGTAGTCTCGCTTTGGCCTCGTAGTATCTTTATAAAACAAAAAATGATAAAAT
TAAGGTGCATTTATACCCATAAACTTGCTTTACACCTATATTCTCCACCACATCCTCAACATTGATATGTAGAAATATGTAGTATG
AAGAAATATGACATAGTAACAGGTACTAGTCTTACCCTCCCACCATAAGCAACTAAAAAACTGGACAAAATATATGAAACAACT
GTTTTCAGATATCAGACAACAGGAAGTACAGGGCTGTAATCTTTGAGAAAAGAGAAACATATGAGAGAGCATTCACCCCACCT
CTCTACCTAGAGGCCATTTCCAAATGGAGAGTTTAGACATCACATGGTGCTAGAAGACA**CGGGAGTTGAGTCAGAACCCA**TG
GAGATCCCAGAAAGGCCACATTTTAAAAGGTTACTCTAGACTTACTCTACACCCATCTTCACAAAGTCTAAAACCATGTCACA

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TGGGGCCTCAGATAGTGAGT	18	59.66
Reverse	GGGTTCTGACTCAACTCCCG	19	60.04

Fig 4: Primer designed for *DLG2*. Primer was designed against c.588C>T: p.R196S and c.587T>A: p.R196K from Son (44093) in exon 8 of *DLG2*. The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.5 Primer design for *ALG13* (016238.1) of family 5

ATAATTGGTTG**TCTTTACCCCTTCTCCTCCCAA**ATTAAGATACATAAAAAAAAATCCATAGGACTTTAGGGGCAATTAGCTGTCT
ACCTTGAATTAACATCATTTAGCCCTTTTGGGTTATTTTAAAGCCAAACATTAAACCGGATATTTATATTACTTTATATGAATAAG
GTAGGCAACTTAAAAGCACAAATAGTTGTTCCTATAACTATCTCTTACGTATGCCAAATTATGTCTTCCCACTTACCTTTGTTTTC
CCCATATAGTCGGGATTTCATTCTTTATCGCTTTCCTGGAAAACCTCCAACTTATGTCACAGATAATGGCTATGAAGACAAGGTA
AGAAGATGAGTGAATGTTGACTTATATAAAAGAAG***T***TGAATGATGCTTCTGGCTTGCCTGGATATTAAATCATTTTTAACCTAA
TTAATGAAAAGCATAGTCTGAATGACAGGTTTCTCTGTTACTAGCATTTTCCTCTCTAGTCAGTTAATCTCAGTGTTGTTAGCCCT
CTAAACAGACCTGATCGTCATTTCTCTGTGATAGGAAAAAAGCCACCCCAAA**CCCTAATTGGCTTCTCAATGACTA**AAAAGGAT

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TCTTTACCCCTTCTCCTCCCAA	22	56.9
Reverse	TAGTCATTGAGAAGCCAATTAGGG	24	58.3

Fig 5: Primer designed for *ALG13*. Primer was designed against the variant c.C953T:p. R318H of mother (39697), C>T present in exon 8 of family 5. The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.6 Primer design for *GRIA1* (047078.1) of family 6

CCAAGCCCCGGACCCGTGCTCAGTCCCTCCCCAGAGCTTTCCA**CAGTTGATGTCCATGCTGCT**GATGCCTCTTCCCTTTGGCCTG
CAGGTCCCTGTCTGGTCGCATCGTTGGTGGCGTCTGGTGGTTCTTCACCTTAATCATCA***C***CTCCTCATATACAGCCAATCTGGCC
GCCTTCCTGACCGTGGAGAGGATGGTGTCTCCCATTGAGAGTGCAGAGGACCTAGCGAAGCAGACAGAAATTGCCTACGGGA
CGCTGGAAGCAGGATCTACTAAGGAGTTCTTCAGGGTAGGGACATCCTTTGTCCCAAGACACTTTCACAAAAGGGAACCACAA
CTGTCATTAAAATGTCCTTCTCATTATATAGCCTGAGATAACAGGCAGCCAGAGCAAGTGCTTAACTGACTGTAAGTCAGGGAA
ACTCAGGACATTTCTAAGCTTTCTACCACCCCCCTGAGAAATCTACATTTCTACATTTGTAGAAGCTTCAAAGCTCCTGCATTGAT
TTGGATCCCTTGCCTGTATCAGTCT**GTGCTAATGCTCTGAGGGCT**GCAAAAGACA

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	CAGTTGATGTCCATGCTGCT	20	56.9
Reverse	AGCCCTCAGAGCATTAGCAC	20	55.8

Fig 6: Primer designed for *GRIA1*. Primer was designed against variant p.1673T with T>C from Father (25431) of family 6. The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.7 Primer design for *NCOR1* (047111.1) of family 7

CAGATGCCAGGTAAAAATATGATAAAGTTACATGTCAAAGATTGAAAGTAAAACTACTTGAGAAGTTTTTAGCTACTGTTATCCTTCCCCTTTT
CCATTACATCTTAGACATTAGAATTGCTATTGCTAATACATCAAGGTCTAAAATAAATCCACCTTAATCCTTGTAGCTAAGTAAAAGTTCCATGT
GCACTAGCTCAGCATGTGATTTTTCTCTACTCTTTTGTACGTGGTGTTTAGCAAATGTATT**TCAGCCAGGACAAAACTCAG**TGCCTTTCTTTCAT
TCTCTCAATCAGCAGCGGCTGCTTACCTGTTTCAGAGACAGCTTTCACCAACTCCAGGTTACCCAAGTCAGTATCAGCTTTACGCAATGGAGAA
CACAAGACAGACAATCTTAAATGATTACATTACCTCACAACAGATGCAAGTGAACTTGCGTCCAGATGTGGCCAGAGGACTCTCCCCAAGAG
AGCAGCCACTGGGTCTCCCATACCCAGCAACGAGAGGTGTGTGTCTTTGCATTTCAGATTTCGTAGAATTTATTATGTTTACCTAGTGTTTCTT
AGGTACTCAGTAAATACTTGTTGAATGAATGATGGAAAGAAAGAAAGAAAAGAATACTATATCCTTCC***T***CCCGTAGACTTTTTAGCCCAGAGA
CATACATTGAGACAAGGTGTTTTGGTGAATACTGAAATTAACTAAACCAAATCCTTGT**GCTGAGTATTTGTAATCCATGG**CTTACAAATTTGCT
CTCTTTAATCTGCCCATACTTTGAGGTTACTGATGTGAAGTACAGTCTTTAATATCCTTCTTTAAAGAATGTCTTAAAAAGTAATGATGATAGTA
AACATTTTTATTTATATTATGTACAAATTACCCTTGATTCACAGAGATAGACAAAATCCAAAGTCATGTATGGTTCTGTACCTTTAGAATAACAC
TTAACATAAATTGTATTGTCTTTTGGATAGGAATCATTGACCTGACCAATATGCCTCCAACAATTTTAGTGCCTCA

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TCAGCCAGGACAAAACTCAG	20	55.5
Reverse	CCATGGATTACAAATACTCAGC	23	56.3

Fig 2: Primer designed for *NCOR1* of family 7. Primer was designed for the validation of segregation of the C>T variant from son (34764). The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.8 Primer design for NALCN (053176.1) of family 8

AAAATCTGCTTGTGTGACAGAAGTCGGTTCTTACTGGACTTTGGTATTCAGTTTTCTTGTGCTTTTACAATACTTGCAGCAATGTGTTTT
ATTTTTTTCCATGTGTTTTTTTCCAGTCGCACATAGCACAAAACCACTGTGCTGATTTCTAATTAACTGACTGCACCCTGAGGTCTATAG
CTTCTGTCTTCAGTTTAATTACAGCTGTGAAGAGGATGCAGTGAAACTGAAGACAGTGTCTCTACATGAGAAAAAAAATAATAATGGA
GTTCCGGTTTGATCCTGAGATCA**TCTAGGTGTGAATGCTTGGGG**ATACTTAAATTGGTATGTAAATGTGAGATTTTCATTTTGTTCTGT
CTTGTTTTCCCTCTTACAAACACTGGTACCTGATGGAGAAAATATCTAACTTAACAAATCAAGCAGTTCTTTGATGTGTAGAAACTTAG
AATGGAAGGAGGGATCTCATCCAACCAGATTAGTCTGCATTAGTCCTTTGGCTCCCATTTTGAAATTGACAATAGGCCAAATATCTGTT
GATGTTCATCCCATTATAAATTCCAAGATATAATTGAAAAAAAAATGCTGAACTCACCAGAGTGGCAAAAAGATGATAGAGAATGAA
ATAGATGGCAACCACGGG***T***GCCCACATATGTCCCACAGCATTTAGAGTTTGGTCCATTACGTCCACCCATCCTTCCTGGGTGAGGATC
TGGAACATGGACATAAATGCCTAAGAGGAAAAGAACAAAGGTATTACACACTAGCTTTAGGCTTGCATACAACTTCAAATTAGCATGT
GATGTTTGAAGACTTTGATCTCAATATTATTGATTTTTATCTG**ATCAGGGCAAGAACAGTCTTGA**TTCAACACATAGGAAAGATATAA
GAATAAAGATCTTAAAAATGTGATATTTAACTATTTGATGTGGAATAAGAAAATAATTACCCTGAATTAAAATTAGTTTTGTTCGCAGT
AAATGCTGTTGAGCAAATTAGAAACAAATTCAAGATCT

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TCTAGGTGTGAATGCTTGGGG	20	66.1
Reverse	TCAAGACTGTTCTTGCCCTGAT	22	64.0

Fig 8: Primer designed for *NALCN* of family 8. Primer was designed s mutation which encodes the base change Primer was designed for the validation of segregation of G>T variant of Father (43094). The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.9 Primer design for *DPYSL3* (011373.4) of family 9

ACCTACAATCGGTTATTCACTTGTGTTTGTTTAAGCCAGTTTGAGCAGGATCTTTGAATTTTGCACCCAATTTAATCCTGACTGATATGTCCGTT
TATAAAACTCAGACAAGGTGAACTGAAAATCAAAGATCAAGGGGAGAAGTAAGATGCACAAAGTCTCAGAAGAACTGGTTCAAGCTGTGTC
ATCTTGGACAACTCATCTAACCTTCCTGAGCCTCAGTTTACTTAAATGCCAAGTGGGAGAAATCCCAGGATCCTCCCATCCACACAGGATGTG
GAGAGCGGCTT**CCTAGGGAGCAAGGCAATGAG**CAAAGTAGGCACCATTTATATTACACAGGACAACCAGAGAGGGAAGGAGCCACACAGT
GCCATCAGGGGCTCCGAGATCTTGGAAACACAAATCACCTTTTTTCCTGGCTTGTGAGATGAGGTCAGCTGCACTCTTGCTCATGACCTTTGTG
ACGTAGAGAGGGCAATTGGTTTGGCTGGCAATGGTGATG***A***CACGGAACACAGCCTCAGCTTCCAGCTGCAAGAAAAAGTCTCCAGGAGTCA
ATAGAAACATGAACCTCTCAAACTGGTGGCTCCCACCTCCACGAACTGTGAGGATGCAGTGCTGCACAAAGGTTATGGATAATTTTGGCAGG
ATCTGGAATTAGATAGCCTATATCCACATCTCAGATCTACCACTTCCTAACTGTGTCACTTTACTCAGCTTCCCAGTGCCTCTATCCCTCACTAGT
AAAATAGTATAGCACAGCATTGTCATCTGA**GAGCAATGTCTGGCAGAGGG**TAAGTACTTTTTCCTTAGCCTCTGTGGTTATCACAAAGGTGAC
ATCTGAATGAAAGCCTTTGCTCCTCAATCTTCTCAGGACAGAATCTCTGGGCAAACACAATAGCAAACCATTATGTCTAAAATAACCAGCCAA
GGCAGTTAGGCATCATGGTTCCTCAAATCCTTTCCTGCAGCAGTAAGTGGTCAGCAAAG

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	CCTAGGGAGCAAGGCAATGAG	21	58.6
Reverse	CCCTCTGCCAGACATTGCTC	20	58.3

Fig 9: Primer designed for *DPYSL3* of family 9. Primer was designed for the validation of segregation of G>A variant from Father (19609). The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.10 Primer design for ANK3 (020987.4) of family 10

ATGAGCATTTTGATAAAGTGACAGTGTTGCACTATTCTGGCAATGTTAGTAGTCCAAAACATGCCATGTGGATGCGCTTTACTGAGGACAGAT
TAGACAGAGGTAGAGAGAAGTTGATATATGAAGATAGGGTGGACAGGACTGTGAAGGAGGCTGAAGAAAAACTGACTGAAGTGTCACAGT
TTTTT**CGTGACAAAACTGAAAAGCTAA**ATGATGAACTGCAGTCCCCAGAGAAAAAGGCACGCCCTAAAAATGGCAAAGAATATTCTTCTCAA
AGCCCTACCAGTAGCAGCCCTGAGAAAGTGCTACTGACAGAACTGCTGGCATCCAATGATGAGTGGGTTAAGGCAAGACAGCATGGCCCTG
ATGGACAAGGCTTCCCCAAGGCCGAGGAGAAGGCACCCAGTCTGCCCAGCAGCCCAGAGAAGATGGTTCTCTCCCAACAGACTGA***G***GACAG
CAAGTCCACAGTGGAAGCCAAAGGAAGTATTTCACAGAGCAAAGCACCAGATGGGCCCCAGTCTGGATTCCAGCTCAAACAATCTAAACTCA
GTTCCATTAGATTAAAATTTGAACAAGGCACACACGCAAAAAGTAAGGACATGTCTCAAGAAGACAGAAAGTCAGATGGCCAGTCCAGAATC
CCAGTTAAAAAAATACAGGAGAGCAAGCTACCCGTCTACCAAGTTTTTGCTAGAGAAAAACAGCAGAAGGCCATAGACCTCCCAGATGAAAG
TGTATCTGTGCAAAAAGATTTTATGGTATTAAAA**ACCAAAGATGAGCATGCCCA**AAGCAACGAAATTGTTGTAAATGATTCTGGCTCTGATAA
TGTGAAAAAACAGAGAACTGAAATGTCAAGTAAAGCAATGCCTGACTCTTTTTCTGAGCAGCAGGCTAAAGACTTGGCATGTCATATAACCTC

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	CGTGACAAAACTGAAAAGCTAA	22	59.2
Reverse	TGGGCATGCTCATCTTTGGT	20	61.9

Fig 10: Primer designed for *ANK3* of family 10. Primer was designed for the validation of segregation of variant T>G from mother (41987). The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

2.11 Primer design for *ATP6V1B2* (025853.17) of family 11

TTCCTTTATTGTGTCTTAAAACAGAGTCACTTGTCTCCAAGAACTTTTTTGGGAATTCTGTTCCATAATGGTATCATCTACCAATCTTGTTATTGATAGAAAAAAAT
GACTTTTCCATGCCTCTCTAGTAGATGCTGACTTCTAAAATCAGCTTTTAGGAATTATTGTTAATTTTTTTTATTTTGTGTATAATTTTTTGAGTCAGTTCAGAAGAC
CATTTTGACTCGATAATGTGGTAGTACTTATCAAACAGATGTTGAAAGTTGAGTTTTTTAAAATAGTCTTGTACAGTGTTATAGATGACACTTTTTTGGATTTGATT
CTCGAAAGAAATGCTATGCTATAAAGTGTTTCTTACTACCTTTTCT**TGCCAGGAAGAGACAGTAGGA**TTCATCTAGGAGACAAATGCAGCGGTTGTCTTTGATTT
AAGTTGTAATAATGAGTTTCAAATGTTTTTTTATTACCAGTAAAAGTTGAGTTTTGTACTTAAATATAGTATTTTTTCCCTCTCATTCTACCCAAGATTTATCCACCT
ATCAATGTGCTGCCCTCA***G***TATCACGGTTAATGAAGTCTGCTATTGGAGAAGGGATGACCAGGAAGGATCATGCCGATGTATCTAACCAGCTAGTATGTACATT
CTTCTAAGAATGGTGTTTGAAAATATGGATACGTTTCTGCTGTCTTACACCTCTTGTCTTTCACCCTTA**CCACTTCTCTCTTCCCCATC**CACATGGAATTTTAGTGTA
GGCTTGATGTGGAATACATAATCATTCATTAGAACCAAAGCAGCCTTTTCCCATAAGGACAGATGCAA**ATGATCAAGATTAGCTTG**AAACCCATGTCTAATGCTA
TCAAGTTTCCAAATAAAAAGCCAAGCAATGAGGTTGTTTAAATTTTTTATTAGCAATTTTAATGTTAATTATAGAGTATTCCTGAAAGGCTACTCTGACCTATTTCT
TGCTTAACATATTAAGCCGTCACCAGAACACTAAAGTATACAGTTTTATGTAGGCTCTGAAATCCTTGATTTTTATACTCATATAAGAT

Primer	Sequence 5'-3'	Length bp	Tm ^{o}C
Forward	TGCCAGGAAGAGACAGTAGGA	21	60
Reverse	GATGGGGAAGAGAGAAGTGG	20	58.7

Fig 11: Primer designed for *ATPV1B2* of family 11.Primer was designed against the variant c. C1192G: p L398V, C>G on exon12 of ATP6V1B2 from Female (III:3)-45666. The variant is underlined, bold and italicized. Forward and reverse primer are indicated by bold letters and shown by using arrow.

Appendix3: Presentation of Polish family-family 11

3.1 Initial thoughts of family 11

The initial thoughts of members which were predicted suggested mutation of *ATP6V1B2* was real and causing factors of phenotypic characters, clinical features and epileptic disorders.

This family consisted of three generation where mother (first generation-I) had unclassified epilepsy and father was not affected with epilepsy. In second generation daughter with DNA number 45664 (II: 1) and son 45665 (II: 4) were carrier of unclassified epilepsy whereas II: 3, Son (45661) had ID with unclassified seizures and dysmorphic features. The siblings of third generation wit numbers III: 1 (45662), III: 2(45663), III: 3 (45666) and III: 4 were affected with unclassified epilepsy, ID with unclassified epileptic encephalopathy, GGE and ID respectively.

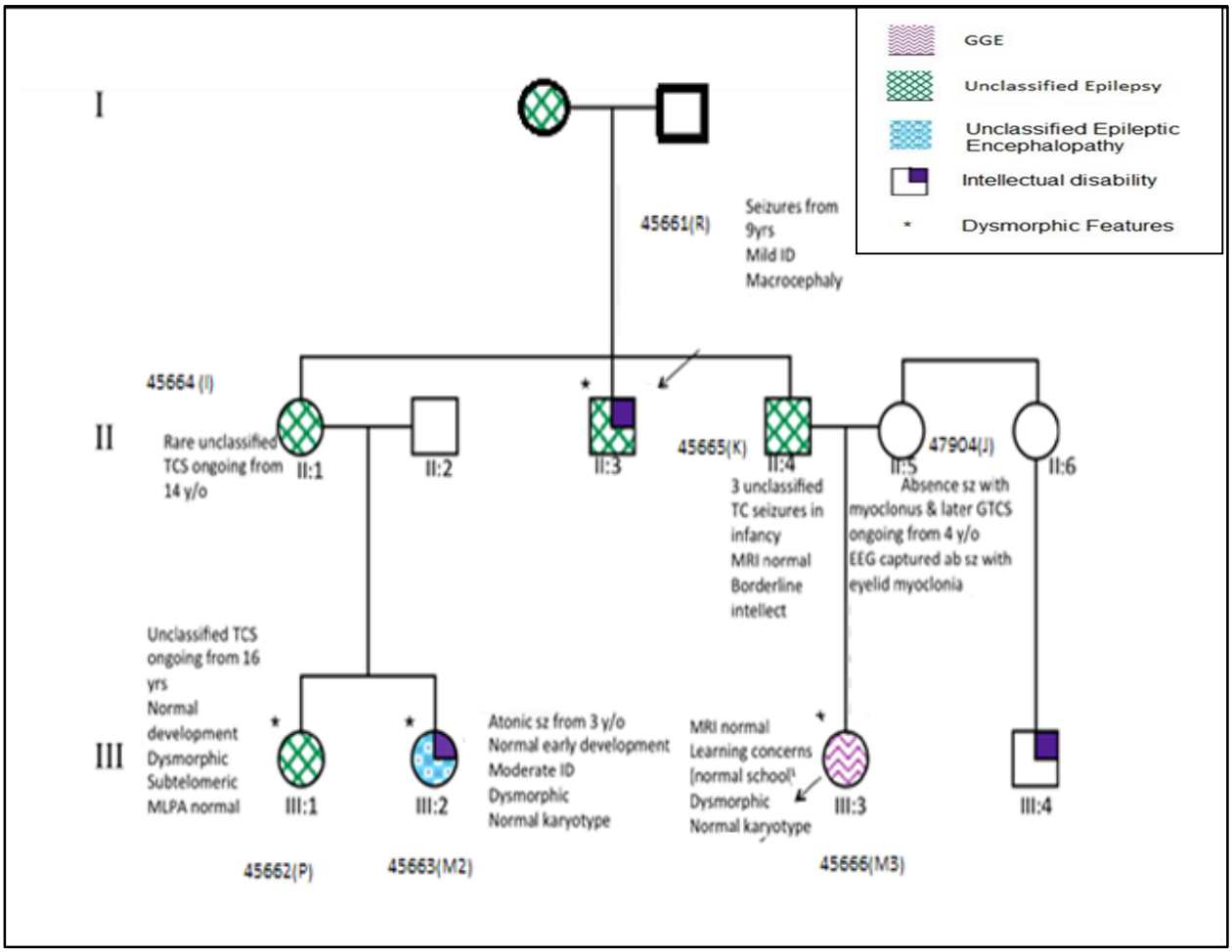

Fig 12: Representation of initial thoughts of members of family 11.

Appendix 4 : Presentation of 95 variants of Polish family-family 11

These were the variants based on functional character shared with son (45661-II: 3) of second generation from family 11, based on high CADD and ID gene. After the validation of variant, the listed variants were priotised mainly on the basis of functional basis in which mutation leads to identification of *ATP6V1B2* and validated the novel variant.

Table 1 : Variants with deleterious variations. These were the variants shared with son (45661-II: 3) based on functional character, high CADD and ID gene with heterogeneous mutations and deleterious variations.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	chr	start	end	ref	obs	Gene.gene	ExonicFunc.gene	AAChange.gene	phastCons	ExAC	SIFTscore	Polyphen	CADD
2	chr1	17949604	17949604	C	G	ARHGEF1	nonsynonymous SNV	ARHGEF10L:NM_001011722:exon10:c.C1017G:p.I339M	Score=625	0	0.018	1	25.7
3	chr1	19487529	19487529	G	A	UBR4	nonsynonymous SNV	UBR4:NM_020765:exon38:c.C5288T:p.S1763L	Score=485	0.000308	0.303	0.001	13.16
4	chr1	22142473	22142473	G	A	LDLRAD2	synonymous SNV	LDLRAD2:NM_001013693:exon3:c.G549A:p.R183R	Score=395	0.000276	0	0	0
5	chr1	29064190	29064190	C	G	YTHDF2	nonsynonymous SNV	YTHDF2:NM_016258:exon2:c.C48G:p.N16K,YTHDF2:NM	Score=558	0.000703	0.002	0.262	23.1
6	chr1	32140895	32140895	G	A	COL16A1	synonymous SNV	COL16A1:NM_001856:exon43:c.C2790T:p.N930N	Score=435	0.000909	0	0	0
7	chr1	33478899	33478899	G	A	AK2	synonymous SNV	AK2:NM_001199199:exon6:c.C579T:p.Y193Y,AK2:NM_0	Score=425	0.000363	0	0	0
8	chr1	44684084	44684084	C	T	DMAP1	synonymous SNV	DMAP1:NM_019100:exon4:c.C495T:p.R165R,DMAP1:NI	Score=560	0.000947	0	0	0
9	chr1	45288989	45288989	G	A	PTCH2	synonymous SNV	PTCH2:NM_001166292:exon20:c.C3183T:p.P1061P,PT	0	0.000111	0	0	0
10	chr1	53986422	53986422	G	A	GLIS1	synonymous SNV	GLIS1:NM_147193:exon6:c.C1086T:p.D362D	0	0.000959	0	0	0
11	chr1	67206976	67206976	G	C	SGIP1	nonsynonymous SNV	SGIP1:NM_001308203:exon21:c.G1730C:p.R577T,SGIP	Score=535	3.95E-05	0.016	0.997	28
12	chr1	161696588	161696588	C	T	FCRLB	nonsynonymous SNV	FCRLB:NM_001002901:exon7:c.C670T:p.H224Y	Score=521	0.000224	1	0.981	25.2
13	chr2	24916179	24916179	G	A	NCOA1	synonymous SNV	NCOA1:NM_003743:exon8:c.G771A:p.T257T,NCOA1:NM	Score=596	0	0	0	0
14	chr2	26689619	26689619	T	A	OTOF	nonsynonymous SNV	OTOF:NM_194322:exon18:c.A2393T:p.D798V,OTOF:NM	Score=577	0.000671	0.256	0.999	29.6
15	chr2	68385266	68385266	G	C	PNO1	nonsynonymous SNV	PNO1:NM_020143:exon1:c.G200C:p.G67A	0	0.000729	0.647	0	0.033
16	chr2	75094895	75094895	G	A	HK2	nonsynonymous SNV	HK2:NM_000189:exon3:c.G359A:p.R120Q	0	0	0.604	0.209	21.8
17	chr2	103334987	103334987	G	A	MFSD9	synonymous SNV	MFSD9:NM_032718:exon6:c.C1317T:p.C439C	Score=325	0.000197	0	0	0
18	chr2	179474006	179474006	T	C	TTN	nonsynonymous SNV	TTN:NM_003319:exon101:c.A24836G:p.H8279R,TTN:NM	Score=435	2.41E-05	0.204	0.984	13.25
19	chr2	187370519	187370519	A	G	ZC3H15	nonsynonymous SNV	ZC3H15:NM_018471:exon8:c.A917G:p.D306G	Score=620	0	0.143	0.473	23.4
20	chr3	4718380	4718380	G	T	ITPR1	nonsynonymous SNV	ITPR1:NM_001168272:exon23:c.G2817T:p.M939I,ITPR1:	Score=393	0	0.33	0.017	17.92
21	chr3	100170600	100170600	AT	TC	LNP1	nonframeshift insertion	LNP1:NM_001085451:exon3:c.194_195insTCCTAGAAG	0	0	0	0	0
22	chr3	113374248	113374248	C	A	USF3	nonsynonymous SNV	USF3:NM_001009899:exon7:c.G6281T:p.R2094L	Score=626	0	0	1	32
23	chr3	113376111	113376119	TGC	-	USF3	nonframeshift deletion	USF3:NM_001009899:exon7:c.4410_4418del:p.1470_14	Score=569	0	0	0	0
24	chr3	124281737	124281737	G	A	KALRN	synonymous SNV	KALRN:NM_001024660:exon34:c.G4977A:p.A1659A	0	0.000112	0	0	0
25	chr3	128853715	128853715	A	G	ISY1,ISY1-I	synonymous SNV	ISY1-RAB43:NM_001204890:exon8:c.T501C:p.D167D,IS	Score=531	0	0	0	0

26	chr3	130452817	130452817	C	T	PIK3R4	nonsynonymous SNV	PIK3R4:NM_014602:exon4:c.G1025A:p.R342H	Score=612	0.000371	0.004	0.998	34
27	chr3	139173613	139173613	G	T	RBP2	nonsynonymous SNV	RBP2:NM_004164:exon3:c.C312A:p.N104K	Score=563	1.58E-05	0.414	0.859	18.5
28	chr3	148596462	148596462	A	C	CPA3	nonsynonymous SNV	CPA3:NM_001870:exon5:c.A401C:p.D134A	0	0.000158	0.572	0	0.394
29	chr6	42932606	42932606	T	C	PEX6	nonsynonymous SNV	PEX6:NM_000287:exon16:c.A2728G:p.T910A,PEX6:NM	Score=393	0	0.002	0.999	26.3
30	chr6	43970503	43970503	-	GC	C6orf223	nonframeshift insertion	C6orf223:NM_153246:exon4:c.369_370insGCGGCG:p.R	0	0	0	0	0
31	chr6	84894982	84894982	T	A	CEP162	nonsynonymous SNV	CEP162:NM_001286206:exon13:c.A1358T:p.E453V,CE	Score=425	0	0.003	1	25.2
32	chr6	165693571	165693571	C	T	C6orf118	nonsynonymous SNV	C6orf118:NM_144980:exon9:c.G1385A:p.C462Y	0	0.000458	0	0.025	0.022
33	chr7	6449512	6449512	C	G	DAGLB	nonsynonymous SNV	DAGLB:NM_001142936:exon13:c.G1588C:p.V530L,DA(	0	3.95E-05	1	0.008	0.487
34	chr1	55135807	55135807	C	T	OR4A15	stopgain	OR4A15:NM_001005275:exon1:c.C448T:p.R150X	0	7.1E-05	0	0	32
35	chr1	62290653	62290653	A	G	AHNAK	nonsynonymous SNV	AHNAK:NM_001620:exon5:c.T11236C:p.S3746P	0	3.16E-05	0.041	0.822	23.7
36	chr1	95724684	95724684	C	T	MAML2	synonymous SNV	MAML2:NM_032427:exon3:c.G2343A:p.A781A	Score=325	0.000104	0	0	0
37	chr1	101981600	101981600	G	A	YAP1	synonymous SNV	YAP1:NM_001130145:exon1:c.G21A:p.P7P,YAP1:NM_0	Score=343	0.000434	0	0	0
38	chr1	2622093	2622093	G	A	CACNA1C	nonsynonymous SNV	CACNA1C:NM_000719:exon9:c.G1333A:p.E445K,CACI	Score=454	0	0.001	1	34
39	chr1	4655494	4655494	C	G	RAD51AP	nonsynonymous SNV	RAD51AP1:NM_006479:exon4:c.C229G:p.L77V,RAD51,	Score=425	7.89E-06	0.317	0.141	18.3
40	chr1	4737620	4737620	A	G	AKAP3	nonsynonymous SNV	AKAP3:NM_001278309:exon5:c.T448C:p.S150P,AKAP3	Score=272	5.53E-05	0.084	0.01	14.81
41	chr1	52518281	52518281	G	T	ATP7B	nonsynonymous SNV	ATP7B:NM_001005918:exon10:c.C2586A:p.H862Q,ATP	Score=443	0.000893	0.019	1	25.1
42	chr1	114751194	114751194	T	C	RASA3	nonsynonymous SNV	RASA3:NM_007368:exon23:c.A2321G:p.Q774R	Score=549	0.000142	0.679	0	9.234
43	chr1	32561112	32561112	G	C	ARHGAP5	nonsynonymous SNV	ARHGAP5:NM_001030055:exon2:c.G1237C:p.V413L,AF	Score=798	0	0.41	0.005	9.811
44	chr1	92537353	92537353	-	G	ATXN3	frameshift insertion	ATXN3:NM_001164782:exon2:c.68_69insC:p.G23fs,ATX	0	0	0	0	0
45	chr1	101606295	101606295	G	A	LRRK1	nonsynonymous SNV	LRRK1:NM_024652:exon32:c.G5653A:p.G1885S	0	5.56E-05	0.848	0	0.001
46	chr1	12009530	12009530	-	CC	GSPT1	nonframeshift insertion	GSPT1:NM_001130006:exon1:c.47_48insCGG:p.G16del	Score=566	0	0	0	0
47	chr7	48314109	48314109	C	T	ABCA13	nonsynonymous SNV	ABCA13:NM_152701:exon17:c.C4846T:p.P1616S	0	0.000352	0.38	0.087	0.021
48	chr7	89894669	89894669	T	G	CFAP69	synonymous SNV	CFAP69:NM_001039706:exon5:c.T411G:p.T137T,CFAP	Score=503	4E-05	0	0	0
49	chr7	149490487	149490487	G	A	SSPO	nonsynonymous SNV	SSPO:NM_198455:exon39:c.G5963A:p.C1988Y	Score=466	0	0	0.999	17.17
50	chr7	151136984	151136984	T	A	CRYGN	nonsynonymous SNV	CRYGN:NM_001308292:exon1:c.A20T:p.K7M,CRYGN:NI	Score=321	7.4E-05	0	0.562	22.8
51	chr8	11666231	11666236	TCC	-	FDFT1	nonframeshift deletion	FDFT1:NM_001287750:exon1:c.205_210del:p.69_70del	Score=259	0	0	0	0
52	chr8	17613005	17613005	C	A	MTUS1	nonsynonymous SNV	MTUS1:NM_001001924:exon2:c.G312T:p.Q104H,MTUS1	0	0.000351	0.001	0.997	12.97
53	**chr8**	**20074761**	**20074761**	**C**	**G**	**ATP6V1B2**	**nonsynonymous SNV**	**ATP6V1B2:NM_001693:exon12:c.C1192G:p.L398V**	**0**	**0**	**0.001**	**0.999**	**25.2**
54	chr8	22168751	22168751	G	A	PIWIL2	nonsynonymous SNV	PIWIL2:NM_001135721:exon16:c.G1927A:p.E643K,PIWI	Score=506	0.000134	0.008	0.227	26.2
55	chr8	24181402	24181403	AC	-	ADAM28	frameshift deletion	ADAM28:NM_001304351:exon9:c.776_777del:p.D259fs,	Score=376	7.9E-06	0	0	0
56	chr8	27509122	27509122	C	T	SCARA3	synonymous SNV	SCARA3:NM_016240:exon3:c.C204T:p.A68A,SCARA3:N	0	4.74E-05	0	0	0
57	chr9	33466372	33466372	G	A	NOL6	nonsynonymous SNV	NOL6:NM_022917:exon17:c.C2143T:p.R715W	0	0.000347	0.008	0.992	25.8
58	chr9	34996478	34996478	G	A	DNAJB5	nonsynonymous SNV	DNAJB5:NM_001135005:exon3:c.G644A:p.R215H,DNA	Score=520	2.37E-05	0.475	0.017	23.3
59	chr9	95237030	95237030	-	TC	ASPN	nonframeshift insertion	ASPN:NM_001193335:exon2:c.149_150insTGA:p.D50de	0	0	0	0	0
60	chr1	608597	608597	C	A	PHRF1	synonymous SNV	PHRF1:NM_001286581:exon14:c.C3141A:p.S1047S,PH	0	0.000746	0	0	0
61	chr1	6411936	6411941	GCT	-	SMPD1	nonframeshift deletion	SMPD1:NM_000543:exon1:c.108_113del:p.36_38del,SM	0	0	0	0	0
62	chr1	8132302	8132302	G	A	RIC3	synonymous SNV	RIC3:NM_001135109:exon2:c.C507T:p.T169T,RIC3:NM_	0	0.000458	0	0	0
63	chr1	17590724	17590724	C	T	OTOG	stopgain	OTOG:NM_001277269:exon15:c.C1702T:p.Q568X,OTO(	Score=644	0	0	0	39
64	chr1	46905499	46905499	T	C	LRP4	synonymous SNV	LRP4:NM_002334:exon19:c.A2535G:p.T845T	Score=635	1.58E-05	0	0	0
65	chr1	88497443	88497443	C	A	ZNF469	nonsynonymous SNV	ZNF469:NM_001127464:exon2:c.C3481A:p.P1161T	0	0.000166	0.76	0.982	22.6
66	chr1	40027943	40027943	G	A	ACLY	nonsynonymous SNV	ACLY:NM_001303275:exon24:c.C3068T:p.S1023L,ACL	Score=435	0	0	0.998	35
67	chr1	40270030	40270068	CTG	-	KAT2A	nonframeshift deletion	KAT2A:NM_021078:exon8:c.1181_1186del:p.394_396de	Score=317	0	0	0	0
68	chr1	57272206	57272206	A	G	PRR11	nonsynonymous SNV	PRR11:NM_018304:exon6:c.A734G:p.E245G	0	0.00015	0.225	0.996	23.5
69	chr1	32826129	32826129	G	A	ZNF397	nonsynonymous SNV	ZNF397:NM_001135178:exon4:c.G1460A:p.G487D	Score=615	0.000164	0.012	1	25.4
70	chr1	40023438	40023438	G	A	EID2B	nonsynonymous SNV	EID2B:NM_152361:exon1:c.C5T:p.A2V	0	0	0	0.993	24
71	chr1	41086245	41086245	G	A	SHKBP1	nonsynonymous SNV	SHKBP1:NM_138392:exon7:c.G416A:p.R139Q	Score=539	0.00075	0	0.998	24
72	chr1	41383114	41383114	C	T	CYP2A7	nonsynonymous SNV	CYP2A7:NM_030589:exon6:c.G989A:p.R330Q,CYP2A7:	0	3.16E-05	0.074	0.956	18.96
73	chr1	42219023	42219023	G	A	CEACAM5	synonymous SNV	CEACAM5:NM_001291484:exon3:c.G558A:p.P186P,CE	0	0.000663	0	0	0
74	chr1	42880522	42880522	G	A	MEGF8	synonymous SNV	MEGF8:NM_001410:exon41:c.G7932A:p.P2644P,MEGF	0	0.000277	0	0	0
75	chr2	48504455	48504455	C	T	SLC9A8	synonymous SNV	SLC9A8:NM_001260491:exon16:c.C1776T:p.D592D,SL	0	0.000482	0	0	0

76	chr2	49458319	49458319	T	-	BCAS4	frameshift deletion	BCAS4:NM_017843:exon4:c.371delT:p.V124fs,BCAS4:N	Score=391	0.000158	0	0	0
77	chr2	60937477	60937477	G	A	LAMA5	synonymous SNV	LAMA5:NM_005560:exon2:c.C429T:p.N143N	Score=569	0.000304	0	0	0
78	chr2	43963651	43963651	C	T	SLC37A1	synonymous SNV	SLC37A1:NM_018964:exon9:c.C669T:p.F223F	Score=485	2.37E-05	0	0	0
79	chr2	21288316	21288316	C	T	CRKL	synonymous SNV	CRKL:NM_005207:exon2:c.C561T:p.H187H	0	8.68E-05	0	0	0
80	chr2	29737500	29737500	G	A	AP1B1	nonsynonymous SNV	AP1B1:NM_001127:exon13:c.C1786T:p.R596C,AP1B1:I	Score=500	1.58E-05	0.002	0.998	34
81	chr2	37906314	37906314	-	CT	CARD10	nonframeshift insertion	CARD10:NM_014550:exon4:c.813_814insAAGGAG:p.K	Score=391	0	0	0	0
82	chr2	41747615	41747615	A	G	ZC3H7B	nonsynonymous SNV	ZC3H7B:NM_017590:exon17:c.A1999G:p.M667V	Score=554	7.89E-06	0.536	0.232	10.94
83	chr2	47095277	47095277	-	AA	CERK	nonframeshift insertion	CERK:NM_022766:exon8:c.875_876insGGGCTACGGC	Score=593	0	0	0	0
84	chrX	47272852	47272852	T	C	ZNF157	synonymous SNV	ZNF157:NM_003446:exon4:c.T1380C:p.H460H	Score=455	0.000852	0	0	0
85	chrX	48620240	48620240	A	-	GLOD5	frameshift deletion	GLOD5:NM_001080489:exon1:c.46delA:p.R16fs	0	0	0	0	0
86	chrX	55779851	55779851	G	A	RRAGB	synonymous SNV	RRAGB:NM_006064:exon7:c.G639A:p.L213L,RRAGB:NI	Score=659	0	0	0	0
87	chrX	69459686	69459686	G	A	AWAT1	nonsynonymous SNV	AWAT1:NM_001013579:exon6:c.G734A:p.R245H	0	9.47E-05	0.089	0.031	19.53
88	chrX	117526804	117526804	G	C	WDR44	nonsynonymous SNV	WDR44:NM_001184966:exon3:c.G321C:p.E107D,WDR4	Score=343	1.58E-05	0.425	0.012	0.091
89	chrX	135333478	135333478	G	A	MAP7D3	nonsynonymous SNV	MAP7D3:NM_001173517:exon1:c.C40T:p.P14S,MAP7D	0	0	0.785	0	0.312
90	chrX	135333480	135333480	C	-	MAP7D3	frameshift deletion	MAP7D3:NM_001173517:exon1:c.38delG:p.S13fs,MAP7	0	0	0	0	0
91	chrX	135333481	135333481	T	A	MAP7D3	nonsynonymous SNV	MAP7D3:NM_001173517:exon1:c.A37T:p.S13C,MAP7D	0	0	0.005	0.998	17.89
92	chrX	135333483	135333483	C	A	MAP7D3	nonsynonymous SNV	MAP7D3:NM_001173517:exon1:c.G35T:p.G12V,MAP7D	0	0	0.002	0.962	22.7
93	chrX	135333486	135333488	CCA	-	MAP7D3	nonframeshift deletion	MAP7D3:NM_001173517:exon1:c.30_32del:p.10_11del,	0	0	0	0	0
94	chrX	141290724	141290724	G	A	MAGEC2	synonymous SNV	MAGEC2:NM_016249:exon3:c.C1050T:p.T350T	0	0.000363	0	0	0
95	chrX	149638719	149638719	G	A	MAMLD1	nonsynonymous SNV	MAMLD1:NM_001177466:exon2:c.G799A:p.A267T,MAM	0	5.53E-05	0.928	0.003	0.05