Identification of underlying genetic mutation in a case of familial epilepsy

A Thesis Submitted to

NIRMA UNIVERSITY

In partial fulfillment of the award of the Degree of

MASTERS OF SCIENCE

IN

BIOTECHNOLOGY

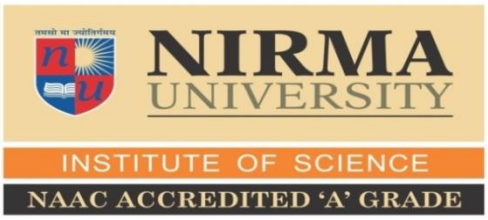
BY

AAKANSHA SHAH 15MBT025

Under the guidance of

Dr. Sonal Bakshi

(Asst. Professor, ISNU)



ACKNOWLEDGMENT

At the end of my thesis I would like to thank all those who have helped me through this journey. First and foremost, I would like to thank Almighty God; it is he who blessed me with the people whose names I feel privileged to mention here.

Words cannot express the feelings and heartfelt gratitude towards my family who have always been the rock strong support through this journey.

Secondly, I would like to thank my guide, Dr.Sonal Bakshi for always supporting me, motivating me and moulding me to become a better researcher.

I am grateful to Prof. Sarat Dalai, Director of, Institute of Science, Nirma University for providing all the facilities to complete my dissertation work.

My wholehearted thanks to the family that, I selected for my case-study. My research would have been incomplete without their technical support.

I am deeply thankful to Dr. Jayshankar Das, Sector Specialist of GSBTM and his team Dr. Shivarudrappa Bhairappanavar, Scientist B for his scholarly guidance, Mr. Inayat Shaikh, Research Associate, for providing his insight and expertise which greatly helped in completion of my thesis.

I am also thankful to all the teaching and non-teaching staff for their constant support and help.

A special thanks to all Ph.D. Scholars Fulesh Kunwar, Shikha Tiwari, and Nazia Saiyed for always sharing their valuable inputs.

I would also like to thank my amazing dissertation group members Shivani Sheth, Kiran Lalwani, Chakshu Rathi, Shweta Patel, Silkey Mehta, Krupa Desai, Rupal Jatvada, and Prapti Purohit for always providing me moral support.

ABBREVIATIONS

ACMG- American College of Medical Genetics and Genomics

ASD- Autism Spectrum Disease

BAM- Binary Alignment/Map

BWA- Burrows Wheeler Aligner

CARS- Childhood Autism Rating Scale

CLNSIG- clinical significance

CNV- Copy Number Variant

CSF- Cerebrospinal fluid examination

DALY- Disability-adjusted life years

DST- Developmental Screening Test

EDTA- Ethylenediaminetetraacetic acid

EEG- Electroencephelography

GBD- Global Burden of Disease

GTG- G banding by Trypsin with Giemsa

ID/DD- Intellectual Disability/ Developmental Disorder

IGV- Integrated Genome Viewer

ILAE- International League against Epilepsy

INDEL- Insertion deletion

ISCN- International System for Human Cytogenomic Nomenclature

MDCT- Multidetector Computed Tomography

MSCT- multi-slice computed tomography

NGS- Next generation sequencing

OMIM- Online Mendelian Inheritance in Man

PWE- Persons with epilepsy

QC- Quality Control

RPMI- Roswell Park Memorial Institute

RT- Room temperature

SAM- Sequence Alignment Map

SEARO- South-East Asia Regional Office

SNP- Single Nucleotide Polymorphism

TES- Targeted Exome Sequencing

TP- Terminated pregnancy

UCSC- University of California, Santa Cruz

VCF- Variant call format

LIST OF FIGURES AND TABLES

Contents

ABSTRACT:

Epilepsy is a chronic neurological condition characterised by recurrent unprovoked seizures and often associated with cognitive dysfunction. The major underlying factors include genetic factors, head trauma, infections, and brain disorders. The present case-study involves a family in which multiple members are affected in different generations suggesting a strong inherited genetic component. The proband's electroencephelography demonstrated frequent epileptiform activity and so this study aims to identify the underlying genetic variants in case of familial epilepsy using targeted exome sequencing which consists of a gene panel covering 5649 genes including genes associated with epileptic encephalopathy genes as well. Total 27 pathogenic mutations were identified in 26 genes. Out of these 27 mutations, seven pathogenic mutations DPYD (c.85T>C), BBS2 (c.209A>G), GLI2 (c.4558G>A), CLN6 (c.214G>T), PRODH (c.1562A>G), HEXB (c.185C>T), and MAG (c.399C>G) are associated with neurological disorders as per OMIM database. Remaining mutations are reported to be associated with congenital, muscular, metabolic, immunodeficiency disorders, etc. It may suggest that the proband might develop other complications as well. The analysis of a proband with unidentified genetic anomaly provided list of sequence variants for further functional investigation to understand the complex molecular mechanism underlying this familial occurrence epilepsy.

1. INTRODUCTION:

As per International League against Epilepsy (ILAE; 1993) epilepsy is a condition characterized by recurrent (two or more) epileptic seizures, unprovoked by any immediate identified cause (Santosh et al, 2014). Epilepsy is a chronic neurological condition and has often been associated with cognitive dysfunction. The seizures precipitate when clusters of nerve cells or neurons generate electrical and chemical signals that act on other neurons which results in strong depolarization mainly due to imbalance between excitatory and inhibitory signals at the synapse. Patient may have strange sensation and emotions or behave strangely. They may have violent spasms or lose consciousness. At the time of seizure, many neurons fire (signal) at the same time – as many as 500 times a second, this is much faster than normal (figure1). This sudden rush of excessive electrical activity happening at the same time causes various involuntary movements. The disturbance of this normal neuronal activity may result in a loss of awareness. (https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Epilepsies-and-Seizures-Hope-Through).

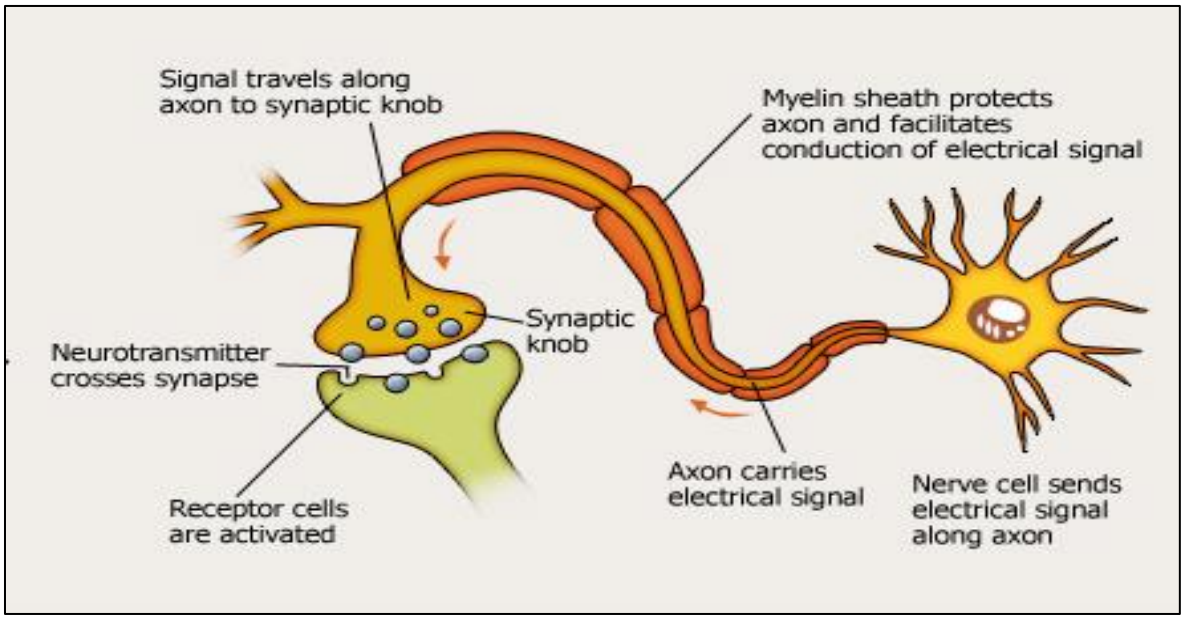


Figure 1: The figure describes the pathway of electrical signals

Epilepsy is also known as spectrum disorder because of its different causes, its ability to vary in severity, different seizure types and impact from person to person, and its range of co-existing conditions (https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Epilepsies-and-Seizures-Hope-Through). In general, a person is considered epileptic only if he or she has had two or more unprovoked seizures separated by at least 24 hours. A provoked seizure is one caused by a known precipitating factor such as high fever, nervous system infections, acute traumatic brain injury, or fluctuations in blood sugar or electrolyte levels. Two major categories of seizure are focal seizures and generalized seizures. Focal seizure occurs only in one part of the brain while generalized affects both the parts of brain. However, there are many different types of seizures in each of these categories. In fact, neurosurgeons have described more than 30 different types of seizures. Generalized seizure is further divided into absence seizures, tonic seizures, clonic seizures, myoclonic seizures, atonic seizures, tonic-clonic seizures, secondary generalized seizures (Amudhan et al, 2015).

1.1 CAUSES:

Anything that disturbs the normal pattern of neuronal activity viz, illness, brain damage, abnormal brain development etc can lead to seizures (https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Epilepsies-and-Seizures-Hope-Through). Primary etiology of epilepsy involves a complex interaction between genes and environment. Primary and secondary epilepsy varies significantly because the cause for primary epilepsy remains unclear and, many conditions being causative for secondary epilepsy. Primary epilepsy has many possible causes including genetic factors, illness, brain injury, and abnormal brain development, genetics, infections like meningitis and encephalitis, medical conditions like brain tumor or stroke or Alzheimer's, problems during pregnancy or injury to infant's brain. Clinically it can be treated by its symptoms or etiological treatment can be provided or through surgery. In half of the cases, the cause is unknown and this type of epilepsy is called secondary epilepsy (https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Epilepsies-and-Seizures-Hope-Through).

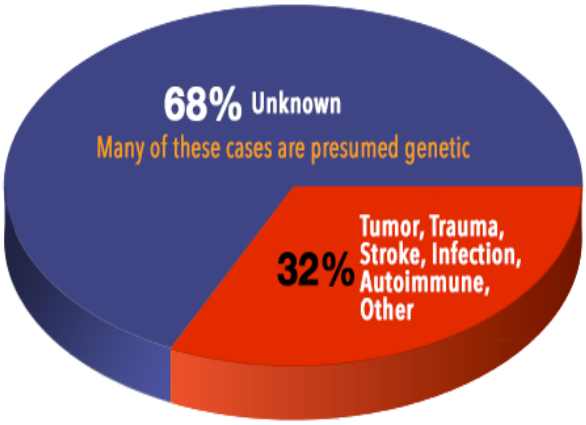


Figure 2: Pie chart depicting the major causes of epilepsy (http://www.cureepilepsy.org)

1.1.1 GENETIC FACTOR:

Genetic mutations may play a key role in the development of certain epilepsies. Many types of epilepsy suggest strong inherited genetic component while in some cases defects is caused by new mutations or changes to the DNA with no family history of the disorder (called "denovo" mutations). In such cases, the mutation will only be passed on if it occurs in the germ line. Epilepsy due to genetic factors can be due to single gene or it can be polygenic. Overall, it is estimated by the researchers that hundreds of genes could play a role in the epilepsy (Liu et al, 2012). In the given case study, it is believed that epilepsy is caused due to genetic disorder because it has affected multiple blood-relatives in extended family members, pointing to a strong inherited genetic component. Gene mutations can also occur spontaneously and contribute to development of epilepsy in people with no family history.

1.1.2 OTHER FACTORS:

Epilepsy can be traced through factors in some of the cases. Other factors are:

- Head trauma due to traumatic injury like accident can cause epilepsy
- Abnormal blood vessel formation (arteriovenous malformations) or bleeding in the brain (hemorrhage)
- Inflammation of the brain

- Infectious diseases, such as meningitis, AIDS and viral encephalitis, can cause epilepsy
- Increased consumption of alcohol, Alzheimer's, multiple sclerosis are some of the other reasons causing epilepsy
- Prenatal exposure history such as infection in the mother, poor nutrition or oxygen deficiency can cause epilepsy. It can also be caused due to some genetic disorders including constitutional chromosomal aberrations

1.1 PREVELANCE:

The prevalence of epilepsy is a strong indicator reflecting the presence of epilepsy (in the past or at a given point of time) indicating the number of people requiring clinical management. In recent times, epilepsy registries have been established in India for studies focusing exclusively on epilepsy. These registries have helped in better understanding of the epidemiology of epilepsy in the development of programs and services (Amudhan et al, 2015). Mental retardation is a condition, which leads to congenital lifelong impairment of cognitive and adaptive functioning and daily live skills. DALYs (Daily Adjusted Life Years) across the globe has varied from 6.2 million in Africa to as low as 1.6 million in the European region with Southeast Asia contributing for 3.2 million DALYs; an increase from 0.3 to 0.5% during 1990-2010. Within the Southeast Asia region, the prevalence of epilepsy varied from 2 to 10 per 1,000 and more than half of the total DALYs due to epilepsy (as estimated from GBD (Global Burden Disease 2010) were accounted from India. Of the 70 million persons with epilepsy (PWE) worldwide, nearly 12 million PWE are expected to reside in India; which contributes to nearly one-sixth of the global burden and is estimated to be 5-9 per 1,000 (Amudhan et al, 2015). This contribution of India towards global epilepsy incidence is due to large population, lower income and education, inadequate resources, sociocultural prejudices, competing infectious and noncommunicable diseases, and the low importance given for public health aspects of epilepsy. Epilepsy affects both males and females of all races, ethnic backgrounds, and ages. The onset of epilepsy starts more often in children and older age groups (http://www.epilepsy.com/learn/epilepsy-statistics). A changing pattern observed towards older age group is due to sociodemographic and epidemiological transition.

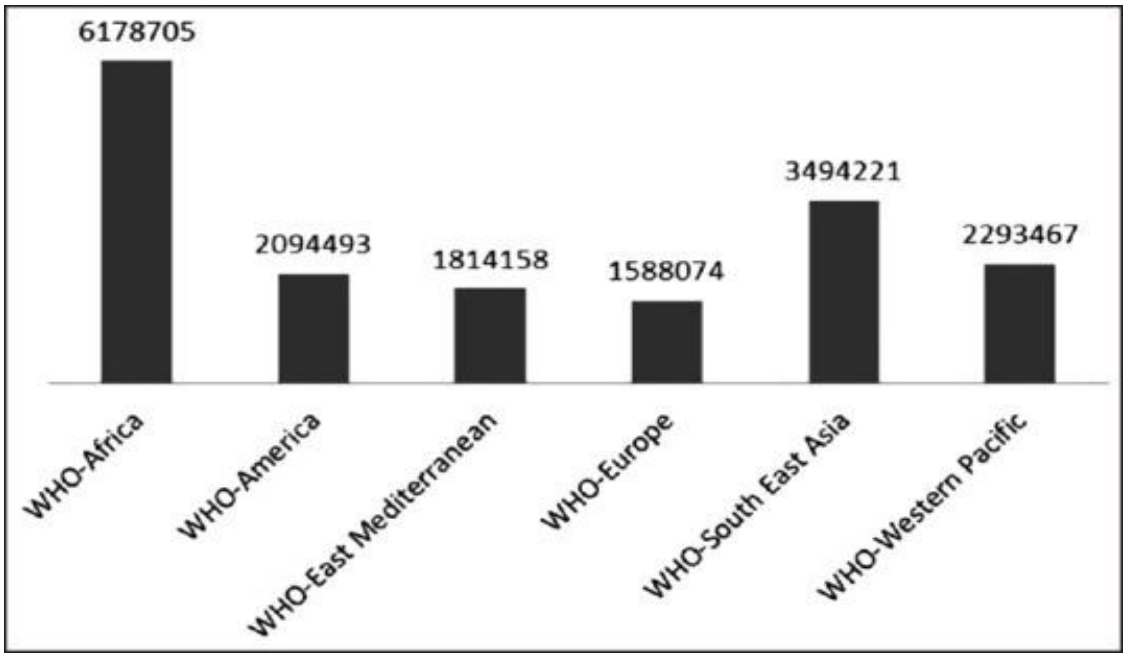


Figure 3: Disability-adjusted life years (DALYs) due to epilepsy among six WHO regions (GBD 2010). WHO = World Health Organization, GBD = Global Burden of Disease

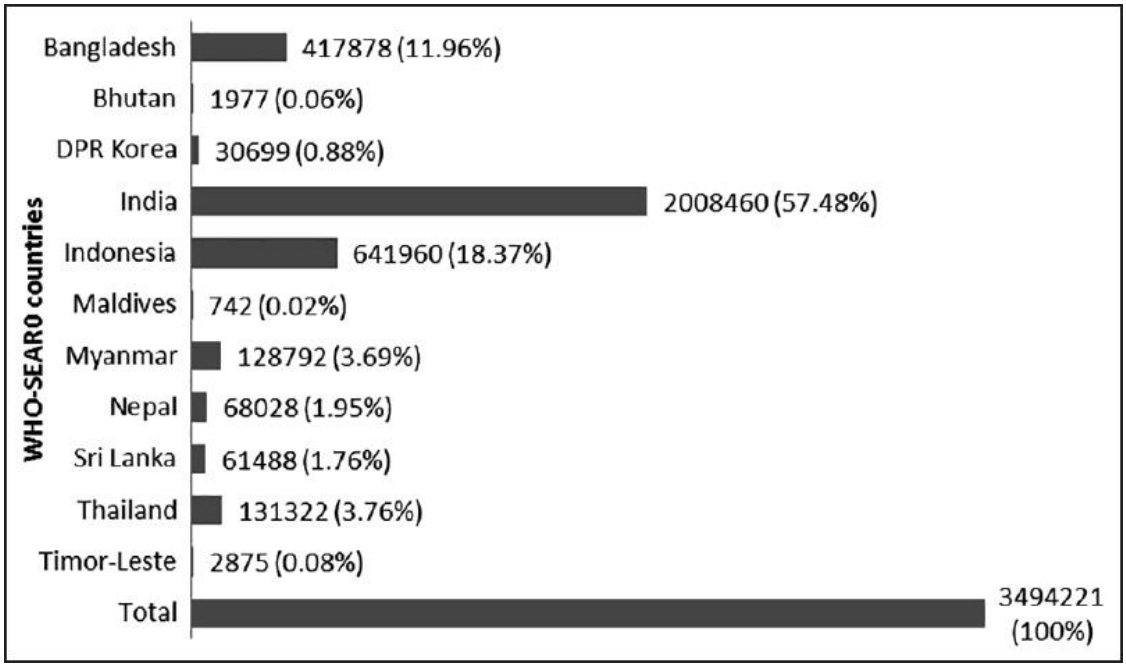


Figure 4: DALYs (Disability-adjusted life years) due to epilepsy in Southeast Asia (GBD-2010) (SEARO = South-East Asia Regional Office)

1.2 TREATMENT:

Some people with epilepsy can lead productive lives while some get severely impacted. In many cases, lifelong treatment is required to control the seizures, while in some cases the seizures eventually stop. Epilepsy can be controlled with drug therapies and surgery. However, as much as 30 to 40 percent of people with epilepsy continue to have seizures because the treatments that they get do not completely control their seizures (called intractable or medication resistant epilepsy). The odds of becoming seizure-free are not as good with severe epilepsy syndromes, but it is possible that seizures will decrease or even stop over time. Seizures can be controlled if the epilepsy begins during childhood, has been well controlled by medication, or if the person has had surgery to remove the brain focus of the abnormal cell firing. There has been medical and research advances where more than 20 different medications and a variety of dietary treatments and surgical techniques are now available and may provide good control over seizures (Amudhan et al, 2015). Many devices are also developed that can modulate brain activity to decrease seizure frequency. Advance neuroimaging can also identify brain abnormalities that give rise to seizures, which can be cured by neurosurgery. Research on the underlying causes of the epilepsies has led to a great improvement in understanding of these disorders that may lead to treatments that are more effective or even to new ways of predicting and preventing epilepsy in the future (Amudhan et al, 2015).

It is very difficult to determine a person's risk of inheriting or passing on this complex disorders because they often cluster in families, and may not have a clear pattern of inheritance. Chromosome study has been widely used for diagnosis of numerous genetic disorders. Currently it provides the only direct view of the genome as a whole and the entire genome can be viewed only during mitosis in an orderly and systemic manner. Development of many technologies like Next generation sequencing (NGS), whole exome sequencing has led to an explosion of gene identification in unknown disorders. NGS has become the premier tool in genetic and genomic analysis.

1.3 ROLE OF NGS PLATFORM:

The production of large numbers of low-cost reads makes the NGS platforms described are useful for many applications. These include discovery of a variant by resequencing targeted regions of interest (targeted exome sequencing) or whole genomes, de novo assemblies of bacterial and lower eukaryotic genomes, cataloguing the transcriptomes of cells, tissues, and organisms (RNA–seq), genome-wide profiling of epigenetic marks and chromatin structure using other seq-based methods (ChIP–seq, methyl– seq and DNase–seq), and species classification and/ or gene discovery by metagenomics studies.There are various platforms for NGS suited for different applications. SOLiD/Ion Torrent PGM typically represents NGS systems from Life Sciences, Genome Analyzer/HiSeq 2000/MiSeq from Illumina, and GS FLX Titanium/GS Junior from Roche. Beijing Genomics Institute (BGI), which has the world's biggest sequencing capacity, has multiple NGS systems including 137 HiSeq 2000, 27 SOLiD, one Ion Torrent PGM, one MiSeq, and one 454 sequencer. It helps in sample handling, sequencing, and bioinformatics analysis (Santosh et al.2014). Applications of NGS includes using Illumina/ Solexa and Life/APG platforms for variant discovery by resequencing human genomes because gigantic volumes of high-quality variants are produced per run. Furthermore, the Helicos BioSciences platform is well suited for applications that demand quantitative information in RNA–seq or direct RNA sequencing, as it sequences RNA templates directly without the need to convert them into cDNAs (http://www.cureepilepsy.org).

Particularly targeted exome sequencing (TES) has been used in the present case study because whole-genome sequencing is an expensive endeavour. TES uses target specific regions of interest. This strategy can be used to examine all of the exons in the genome, specific gene families that constitute known drug targets or megabase size regions that are associated with disease or pharmacogenetic effects through genome-wide association studies (Santosh et al, 2014). Examples: Mutations of MTOR genes were identified by TES and WES together, which was associated with spectrum of brain overgrowth phenotypes extending from FCD (Focal cortical dysplasia) type 2a to diffuse

megalencephaly, distinguished by different mutations and levels of mosaicism (Mirza et al, 2016). Using whole-exome sequencing, rare homozygous missense variants (c.526C>T [p.Arg176Trp] and c.629C>T in SLC45A1 were identified, encoding cerebral glucose transporter, in two consanguineous families with moderate to severe ID (Intellectual Disability), epilepsy, and variable neuropsychiatric features (Srour, 2017). In another case, WES followed by targeted analysis of 64 epilepsy genes, on 40 consecutive children and adults who had MRI-negative focal epilepsy and a family history of febrile seizures or any type of epilepsy in at least one first- or second-degree relative was performed. A pathogenic SCN1A variant in a patient with drug-resistant epilepsy was identified. It led to discontinue long-standing carbamazepine therapy in the patient, resulting in complete seizure control. The data demonstrated that WES with targeted gene analysis is an effective diagnostic tool for patients with common focal epilepsies in whom a genetic etiology is suspected. It can also influence clinical decision-making, including antiepileptic drug selection and consideration of epilepsy surgery, hence supporting its incorporation in the routine clinical care of patients (Perucca et al, 2017). Despite so many applications, there are no reported mutations identified by TES from India on genetics of epilepsy.

2 OBJECTIVE:

- To identify the genetic mutation possibly responsible for the high incidence of epilepsy in the family.
- To help with the genetic counselling to predict and prevent the future progeny from being affected by the same.
- The increased knowledge about the causative genetic variants of epilepsies has had a major impact on its diagnosis. These have already been translated into treatment recommendations like genetic counselling for a few genes.

3 STUDY DESIGN:

- Patient selection
- Chromosomal analysis
- Targeted Exome Sequencing

4 CASE STUDY:

A 17-year-old female, born as a second child with one normal elder sibling to the non-consanguineous couple was delivered normally after the completion of 280 days of gestation. The proband in the given case study was normal until the age of 1 year 6 months. There was no significant history of prenatal or postnatal exposures. They had three TP (terminated pregnancy) in between the first and the second child. There were six affected family members from the paternal side of the proband suffering from similar symptoms. Out of the six affected family members, five are females while one is male. Only two of the affected members are alive at present, one is a 17-year-old female and another is 49-year-old male (figure 5). They all had the similar neurodevelopmental symptoms.

The patient presented here had seizures in the form of head banging without fever, clinching teeth, which recurred in a particular period. Developmental delay was seen alongwith the loss of speech and stereotypes. She could sit and crawl, but could not walk. She had disturbed gait movements. At the age of 2 years 2 months psychological test report showed that, she needed assessment in motor functioning (walking) and speech. Electroencephalography (EEG) done at the age of 2 years revealed developmental retardation, mental retardation, primary generalized epilepsy and diffuse brain dysfunction. Proband was also assessed for Rett syndrome because she had purposeless hand movements, hand wringing, episodes of laughing, drooling but was tested negative. Despite multiple antiepileptic therapies, she continued to have intractable seizures.

With time she showed abnormal physical growth and weight gain alongwith neuro-regression. Limb development stopped and she had repetitive act stereotypes. A few subtle dysmorphic features were evident, such as, her head size, which was stable increased dynamically. She stopped standing. She could make some vocalizations and was fed orally without any swallowing difficulties. She had dental caries at the age of 15 but it was cured by the surgery. EEG report at the age of 12 years was abnormal with frequent interictal epileptiform discharges from right hemisphere suggestive of right sided epileptiform activity with diffuse right sided brain damage. She is presently on multiple antileptic medications. Extensive neurometabolic investigations were essentially unremarkable, including 2D-Echo, normal brain MRI, DST (Developmental Screening Test), CARS (childhood Autism Rating Scale), MDCT (Multidetector Computed Tomography), CSF (cerebrospinal fluid) examination, video EEG, MSCT (multi-slice computed tomography) and other biochemical reports.

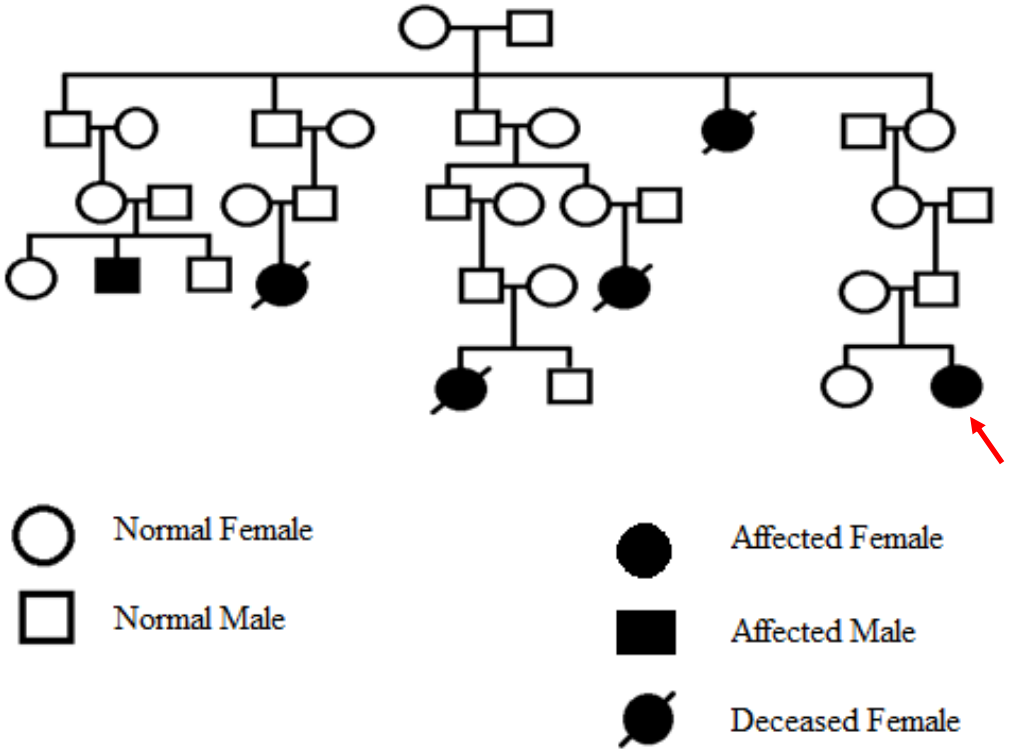


Figure 5: Pedigree chart depicting multiple cases of epilepsy

5 METHODS AND MATERIALS:

5.1 PREPARATION OF REAGENTS:

5.1.1 Stock Giemsa:

1 gm Giemsa was added to 54 ml glycerol. It was kept in waterbath at 60˚C for 8hours. It was filtered with funnel in dry blue-cap bottle. In addition, the funnel should be covered so that it does not evaporate.

5.1.2 Working Giemsa :

2 ml of Giemsa was added to 60 ml Sorenson's buffer.

5.1.3 PBS :

0.1 gm KCl (potassium chloride), 4 gm NaCl (sodium chloride), 0.72 gm NaH_2PO_4 (sodium dihydrogen phosphate), and 0.12 gm KH_2PO_4 (potassium dihydrogen phosphate) was added in 500 ml distilled water (Freshly prepared).

5.1.4 Stock trypsin :

0.01 gm EDTA (Ethylenediaminetetraacetic acid) and 0.02 gm Trypsin was added in 50 ml PBS (Phosphate buffered saline).

5.1.5 Working trypsin:

25 ml stock trypsin in 25 ml PBS was added.

5.1.6 Sorenson's buffer:

0.345 gm NaH_2PO_4 and 0.45 gm NA_2HPO_4 (Disodium hydrogen phosphate) in 250 ml distilled water (pH=7) was added.

5.1.7 Fixative:

The ratio of acetic acid: methanol should be 3:1. It should be freshly prepared and chilled (90 ml acetic acid + 30 ml methanol).

5.1.8 Hypotonic solution:

0.56 gm KCl added in 100 ml distilled water and kept in waterbath at 37˚ C.

5.2 METHODS:

5.2.1 BLOOD COLLECTION:

A 17-year-old female patient was examined for her phenotypic characteristics. Written informed consent was obtained from her parents. Patient's past and present clinical history was verified by neurosurgeon. Detailed family history was obtained, and peripheral blood samples of the 2 affected and 2 unaffected family members were collected. Venous blood was collected aseptically into sodium heparin and EDTA vacuatte with the help of multi-sample needle. The blood collected in the EDTA vacuatte was used for DNA isolation for targeted exome sequencing. The blood collected in sodium heparin tubes was used for cell culture and karyotyping.

5.2.2 DNA ISOLATION:

DNA isolation was done from Whole Blood (EDTA vacuatte) using the QIAGEN kit. This protocol is for purification of genomic DNA from up to 2 ml of whole blood.

i) 100µl QIAGEN Protease was pipetted in the bottom of the 15 ml centrifuge tube
ii) 1ml blood was added and mixed
iii) 1.2ml of Buffer AL was added and was mixed vigorously
iv) It was incubated at 70˚C
v) 1ml ethanol was added and was transferred to the QIAamp Midi column and was placed in a15ml centrifuge tube and was centrifuged for 3min at 3000rpm
vi) After discarding the filtrate, it was centrifuged again for 1 min after addition of Buffer AW1 and the same thing was repeated for Buffer AW2 but it was centrifuged for 15min instead of 1min at 5000rpm
vii) 500µl of Buffer AE was added to the QIAamp Midi column and was incubated for 5minutes at RT (Room temperature) and centrifuged at 5000rpm for 15min

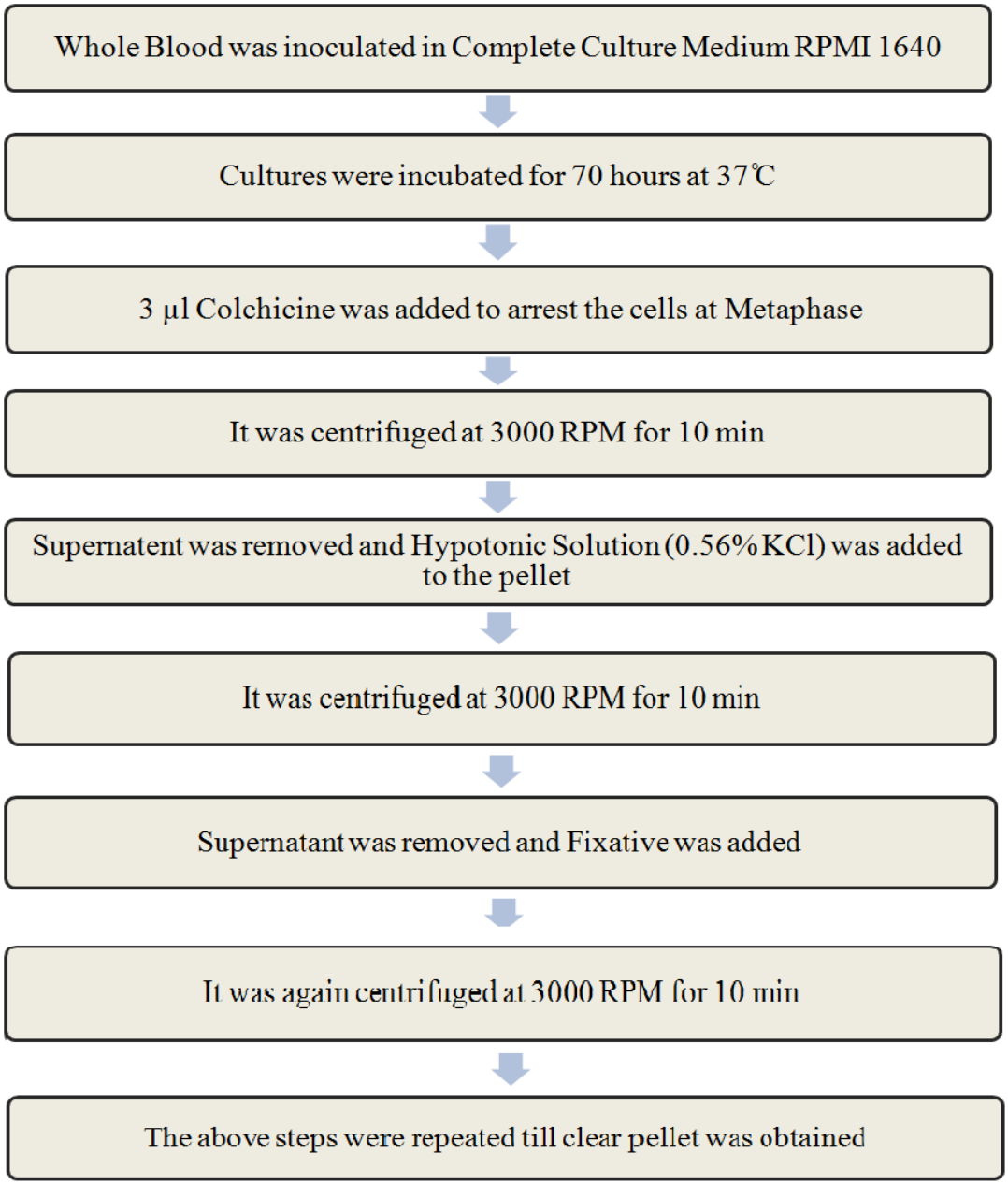
viii) Isolated DNA was stored at -20˚C

ix) DNA concentration was measured by nanodrop technique and it was found to be 104ng/µl

(260nm:280nm) Purity of DNA – 1.8

5.2.3 SHORT TERM CULTURE:

Short-term culture of whole blood was done for metaphase preparation.



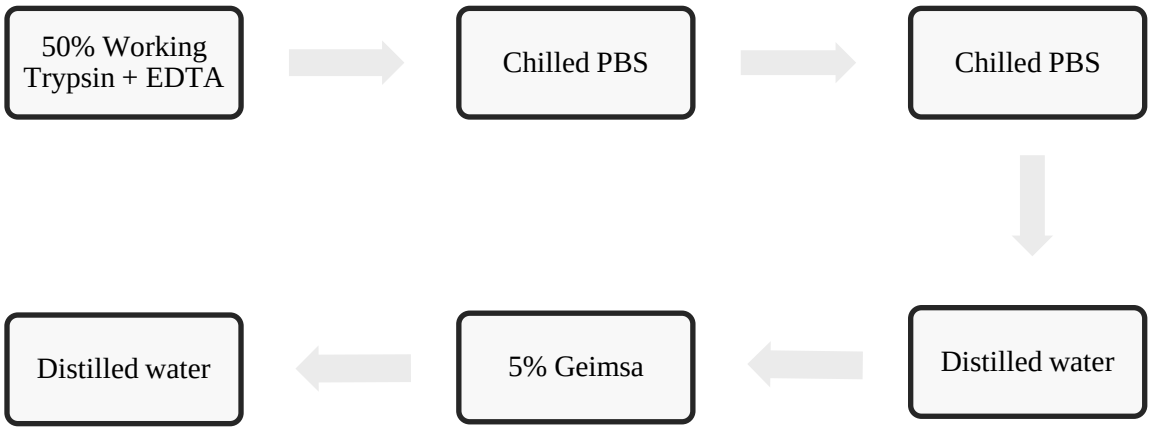
5.2.4 SLIDE MAKING:

i) Slides were kept in concentrated chromic acid for 24 hours and then were rinsed in running tap water for 30minute and then with distilled water for 30 minutes. Now the slides were immersed in methanol for 30minutes. And finally, before using the slides, they were immersed in distilled water.
ii) The cell pellets were added with fresh fixative again and resuspended until turbid.
iii) 2-3 drops of this cell suspension were dropped on a clean slide.
iv) Slides were air dried and was observed under the microscope for frquency and quality of metaphase cells.

5.2.5 GTG BANDING:

For GTG banding, slides were aged for 10 days. Slides can also be baked at 90˚C for 2 hours before banding. The procedure for banding is:

i) Slide was treated for 30-60 seconds with trypsin and EDTA.
ii) After removing the slide from working trypsin solution, it was given two chilled PBS rinse and one distilled water rinse.
iii) Then the slide was stained with 5% Giemsa for 8-15 minutes and then it was rinsed with distilled water.
iv) Slide was observed under the light microscope.
v) 20 metaphases plate were scored per person.



vi) Metaphase plates was scored and chromosomes were grouped according to their size,shape and banding (patterns as per ISCN, 2016)

- Group A – large metacentric – 3 pairs of chromosomes
- Group B – large submetacentric – 2 pairs of chromosomes
- Group C- medium submetacentric - 7 pairs of chromosomes
- Group D – medium acrocentric – 3 pairs of chromosomes
- Group E – small submetacentric – 3 pairs of chromosomes
- Group F – small metacentric – 2 pairs of chromosomes
- Group G – small acrocentric – 2 pairs of chromosomes
- XX / XY

5.2.6 KARYOTYPE:

Banded slides were used for karyotyping. The slides were analyzed using microscope with image analysis system metasystem Ikaros software.

5.2.7 TARGETED EXOME SEQUENCING:

Clinical or Targeted Exome sequencing is a very useful tool for analyzing specific mutations in a given case of genetic condition. A focused gene panel contains a selected set of genes or gene regions that have known or suspected associations with the disease or phenotype under study. Here, a gene panel comprising of 5469 genes was used for TES using Illumina platform. Analysis of fastqc file (raw data) was done to obtain VCF file, which was used for identifying mutations. The pipeline used for analysis is shown in figure 6. After the causal variant is identified, it is visualized by IGV (Integrated Genomic Viewer).

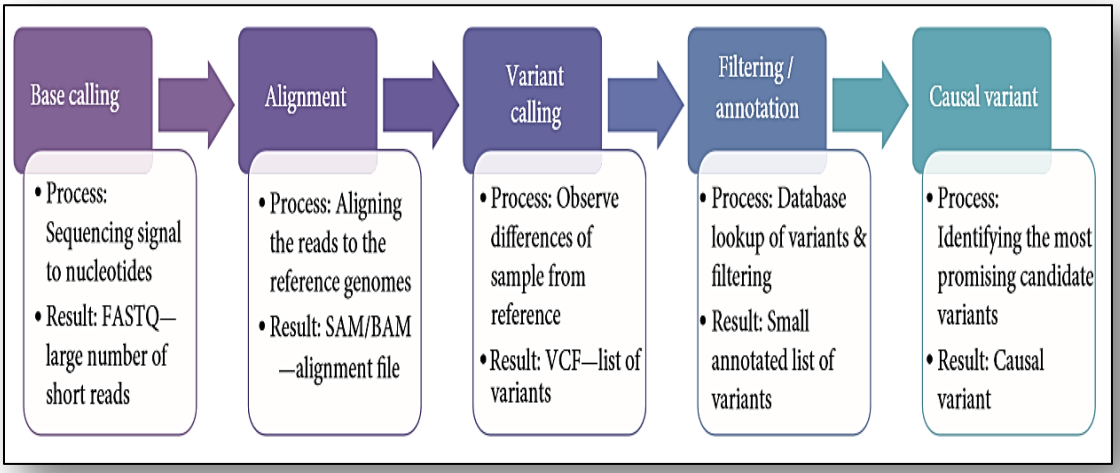


Figure 6: NGS pipeline (Dolled-Filhart et al, 2012)

1. QUALITY ASSESMENT:

Modern high throughput sequencers can read tens of millions of sequences in a single run. Some simple quality control checks should always be performed to ensure that the quality of raw data is good and there are no problems or biases in the data before analysing this sequence to draw biological conclusions (https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual). FastQC is a popular tool for analysing FASTQ files and it will provide a wide range of information related to the quality profile of the reads (Ramirez et al, 2014). A QC report is generated as part of the analysis pipeline, but this is usually only focused on identifying artifacts, which were generated by the sequencer itself. The main aim of FastQC is to provide a QC report, which can spot quality, which originate either in the sequencer or in the starting library material (https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual).

2. ALIGNMENT OF SEQUENCES:

The first step in the pipeline involves the alignment of reads in the .fastq file against a reference genome, for which the reference genome has to be obtained. Various sources are available to download reference genome. The UCSC genome browser contains publicly available reference genome assemblies for a large variety of organisms (The Hitchhiker's Guide to Whole Exome Analysis by Shrey Gandhi, Vinod Scaria). For this study, hg19 reference genome has been used which was obtained from UCSC Genome bioinformatics browser.

2.1 Map to reference: Post-processing is an important part of the pipeline and helps polish the alignment output results. After mapping reads to the reference genome, these post-alignment processing steps are necessary to minimize the artifacts that may affect the quality of downstream variant calling (The Hitchhiker's Guide to Whole Exome Analysis by Shrey Gandhi, Vinod Scaria). Here, indexing of reference sequence has been done for further mapping steps using **BWA** and **SAM Tools** and then both paired end sequences

was mapped with most recent version of human genome (GRCh37/hg19) sequence and produced a file in SAM format using BWA v0.7.12 software.

BWA (Burrows Wheeler Aligner)

BWA is software for mapping low-divergent sequences against a large reference genome, such as the human genome. It consists of three algorithms: BWA-backtrack, BWA-SW and BWA-MEM.

BWA-MEM, which is the latest, fastest and accurate is generally recommended for high-quality queries (http://bio-bwa.sourceforge.net/).

SAM Tools

The Sequence Alignment Map (SAM) format is a generic alignment format for storing read alignments against reference sequences, supporting short and long reads (up to 128 Mbp) produced by different sequencing platforms. Various utilities for post-processing alignments in the SAM format are implemented such as indexing, variant caller and alignment viewer, and thus provide universal tools for processing read alignments (Li et al, 2007).

2.2 SAM to BAM conversion: To improve the performance, a companion format Binary Alignment Map (BAM) was designed, which is the binary representation of SAM and keeps exactly the same information as SAM. The advantage of BAM file is that it is compact in size and supports fast retrieval of alignments in specified regions (Li et al, 2007). Conversion of SAM files into BAM was done using Picard-tools-1.119

Picard Tools

Picard is a collection of Java-based command-line utilities that manipulate SAM files, and creates a new program that reads and write SAM files. Both SAM text format and SAM binary (BAM) format are supported.

3. VARIANT CALLING:

Variant call format (VCF) is the standardized generic format for storing sequence variation including SNPs, Indels, larger structural variants and annotations. The output of the variant calling process is a VCF format file. This text based format file gives details about the variants detected by the variant detection software (The Hitchhiker's Guide to Whole Exome Analysis by Shrey Gandhi, Vinod Scaria). Variant Calling of SNPs/INDELs was done using SAMtools/BCFtools v1.3.1.

SAM Tools

It collects summary information in the input BAMs, computes the likelihood of data given each possible genotype and stores in the format of BCF. It does not call variants.

BCF Tools

It is utilized for variant calling and manipulating VCFs and BCFs. BCF tools is a set of utilities that manipulate variant calls in the VCF and its binary counterpart BCF (https://samtools.github.io/bcftools/bcftools.html).

4. VARIANT FILTERING AND ANNOTATION:

After variant detection, a large number of variant calls are generated. Manually inspecting such a large number of variants is practically impossible. Variant annotation tools perform the function of combining the raw putative variant calls with auxilary data to annotate variants. It is important to annotate the variants wih attributes such as genomic feature, exonic function and amino acid changes. These annotation tools draw in information from various sources, and an integrated summary is presented. It is possible to predict the function and effect of the individual SNP. Gene annotation allows for codon analysis to determine whether variants are synonymous, non-synonymous, non-sense, missense, or create early stop codons. (The Hitchhiker's Guide to Whole Exome Analysis by Shrey Gandhi, Vinod Scaria). For the variant filtration and annotation step, SnpEff and SnpSift v4.3k was used.

SnpEff

It is a genetic variant annotation and effect prediction toolbox. It annotates and predicts the effects of variants on genes (such as amino acid changes).

Input: The inputs are predicted variants (SNPs, insertions, deletions and MNPs). The input file is usually obtained as a result of a sequencing experiment, and it is usually in variant call format (VCF).

Output: SnpEff analyzes the input variants. It annotates the variants and calculates the effects they produce on known genes (e.g. amino acid changes). (http://snpeff.sourceforge.net/SnpSift.html).

SnpSift

SnpSift is used for filtering and manipulating genomic annotated files (VCF). Once the file is annotated using SnpEff, SnpSift can be used to filter large genomic datasets in order to find the most significant variants. Given the large data files, this is not a trivial task (e.g., All the variants cannot be loaded into XLS spreadsheet). SnpSift helps to perform this VCF file manipulation and filtering required at this stage in data processing pipelines (http://snpeff.sourceforge.net/SnpSift.html). Arbitrary expressions were used for instance, "(QUAL≥200)". In the Annotation step using SnpEff 3 types of files are generated HTML, .txt file and .vcf files.

Annotation with DATABASE

After making .vcf file, data annotation was done with ClinVar database. ClinVar provides a freely available archive of reports of relationships among medically important variants and phenotypes. Building from the foundation of the variants submitted with minimal phenotypic descriptions to dbSNP and dbVar, ClinVar now accepts direct submissions with rich, structured details of phenotype, interpretation of functional and clinical significance, methodology used to capture variant calls and supporting evidence. ClinVar thus provides access to a broader set of clinical interpretations collected on their own and the promise of a comprehensive site for obtaining current and historical data (Melissa J. Landrum et al, 2014)

5. CLINICAL VARIANT ANALYSIS:

Both annotated VCF files were filtered i.e. CinVar and dbSNP manually in Excel sheet. ClinVar database that contains information about **CLNSIG** is a string that describes the variants' clinical significance (Table no.1). Reference SNP ID (rsID) was checked in ClinVar database and the disease information were noted that were reported in MedGene, OMIM, and ClinGen. Variants were classified accoording to the clinical sgnificance provided by ACMG guidelines (Table 1). For example, command CLNSIG=5 will filter out all the pathogenic variants.

CLNSIG#	**ClinVar Definition**
0	**Uncertain**
1	**Not Provided**
2	**Benign**
3	**Likely Benign**
4	**Likely Pathogenic**
5	**Pathogenic**
6	**Drug-response related**
7	**Histocompatibility-related**
255	**Other (Conflicts, flips, etc)**

Table 1: Variant Clinical Significance
(https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/)

6. VISUALIZATION:

Data visualization is an essential component of genomic data analysis. Aligned reads were visualized using IGV v2.3.x.

Integrated Genomics Viewer (IGV):

Analysis of large, diverse datasets holds the promise of a more comprehensive understanding of the genome and its relation to human disease. Human review is an essential component of the process, complementing computational approaches. This call for efficient and intuitive visualization tools able to scale to very large datasets and to flexibly integrate multiple data types, including clinical data IGV is such a lightweight visualization tool that enables intuitive real-time exploration of diverse, large-scale genomic datasets on standard desktop computers. It is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next generation sequence data, and genomic annotations. It supports flexible integration of a wide range of genomic data types including aligned sequence reads, mutations, copy number, RNAi screens, gene expression, methylation, and genomic annotations (James T. Robinson, et al, 2011).

6. RESULTS AND DISCUSSION:

For this study, two patients and two unaffected family members were selected. DNA samples from all four members were obtained. The proband is a 17-year-old female. Clinical history has been described and general characteristics have been mentioned in the table given below (Table2). The results obtained showed that EEG report was abnormal with frequent interictal epileptiform discharges from right hemisphere suggestive of right sided epileptiform activity with diffuse right sided brain damage (figure: 1).

GENERAL CHARACTERISTICS	AGE	17-year-old
	SEX	Female
	ORIGIN	Ahmedabad
DEVELOPMENTAL HISTORY		Patient was normal for one and half year, then neuro-regression begin
GENERAL OBSERVATION	LIMBS	Abnormal (small)
		Difficulty in standing and walking, no response to external stimuli, involuntary hand movements and clinching of teeth
FAMILY HISTORY		3TP (Terminated Pregnancy), one healthy sister of proband Total 6 patients in the whole family

Table 2: General characteristics of the proband

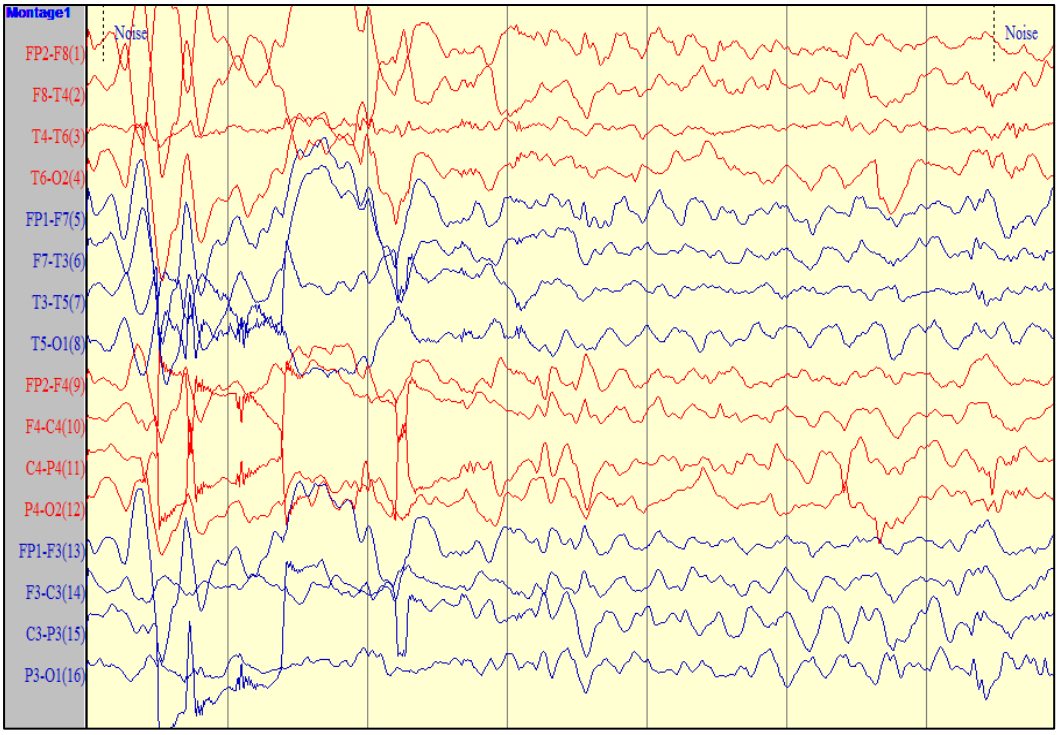
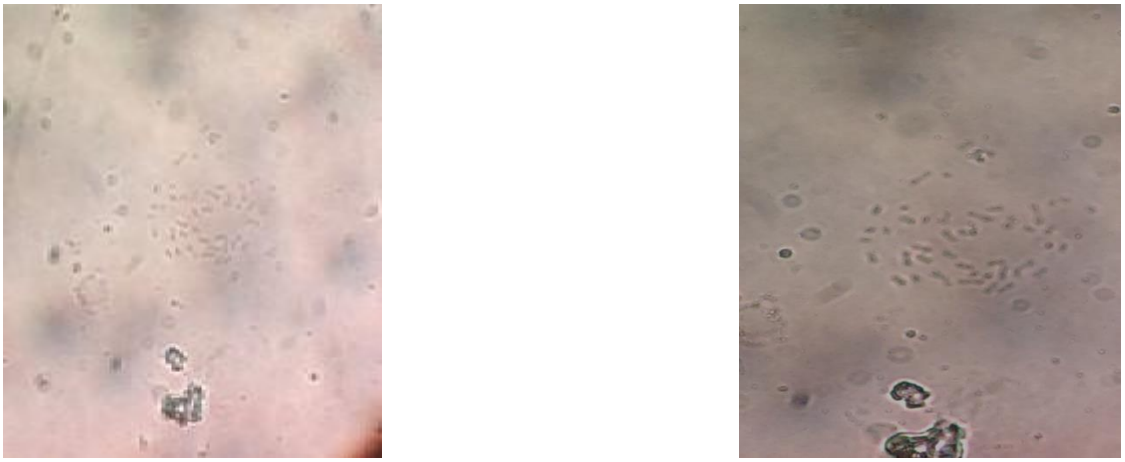


Figure 7: Electroencephalogram of a patient was abnormal with frequent interictal epileptiform discharges from right hemisphere suggestive of right sided epileptiform activity with diffuse right sided brain damage.

6.1 KARYOTYPE RESULTS:

Karyotype was normal at microscopic level (Figure 10) but any changes less than 10 mb cannot be detected by karyotype. Next Generation Sequencing techniques can detect such changes. Targeted exome sequencing detected 27 such pathogenic mutations (Table 4).



(A) (B)

Figure 8: Photomicrography of the image shows unstained metaphase plate used for GTG banding as the chromosomes are well spread (Magnification: 40X + Digital zoom)

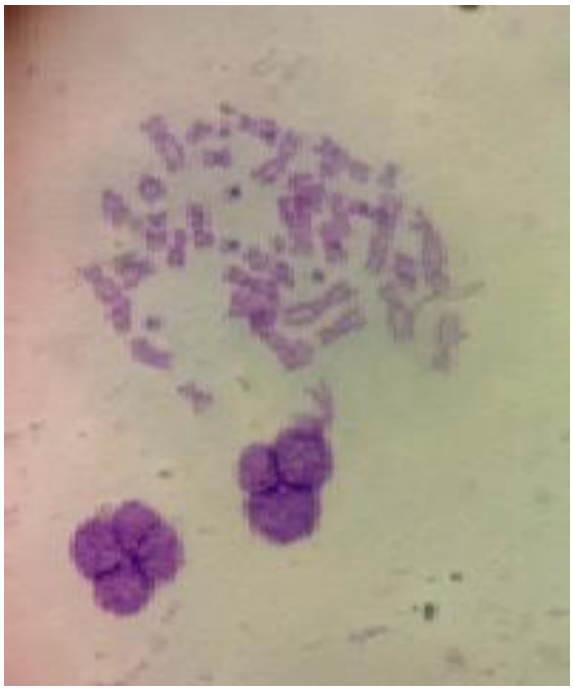


Figure 9: Photomicrography of the banded chromosomes after giving trypsin-PBS treatment (Magnification: 40X + Digital zoom)

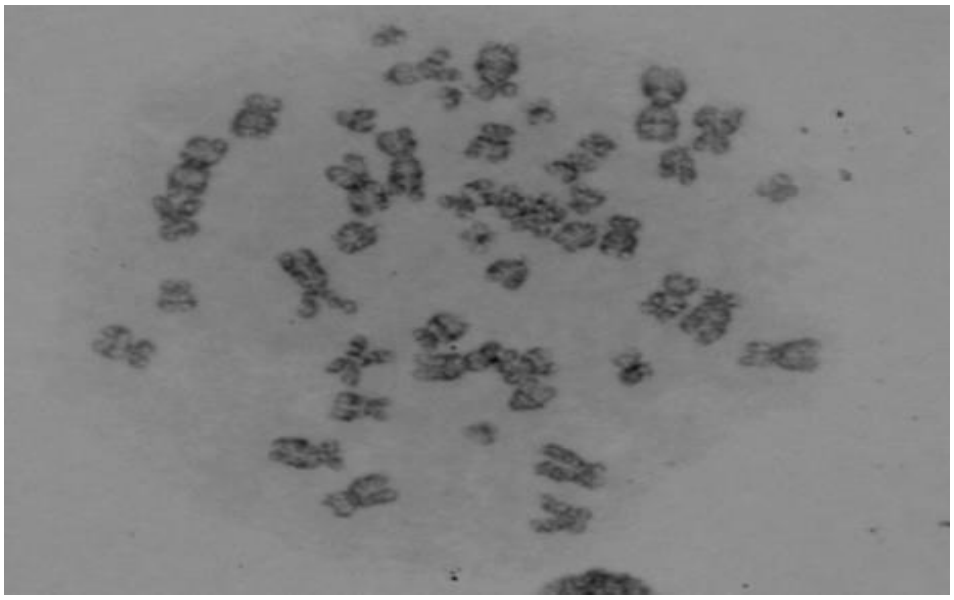


Figure: 10A

Figure: 10B

Figure 10(A) and (B): GTG banded metaphases plates of 10(A) and 10(B) of a normal family member (Magnification: 100X + digital zoom)

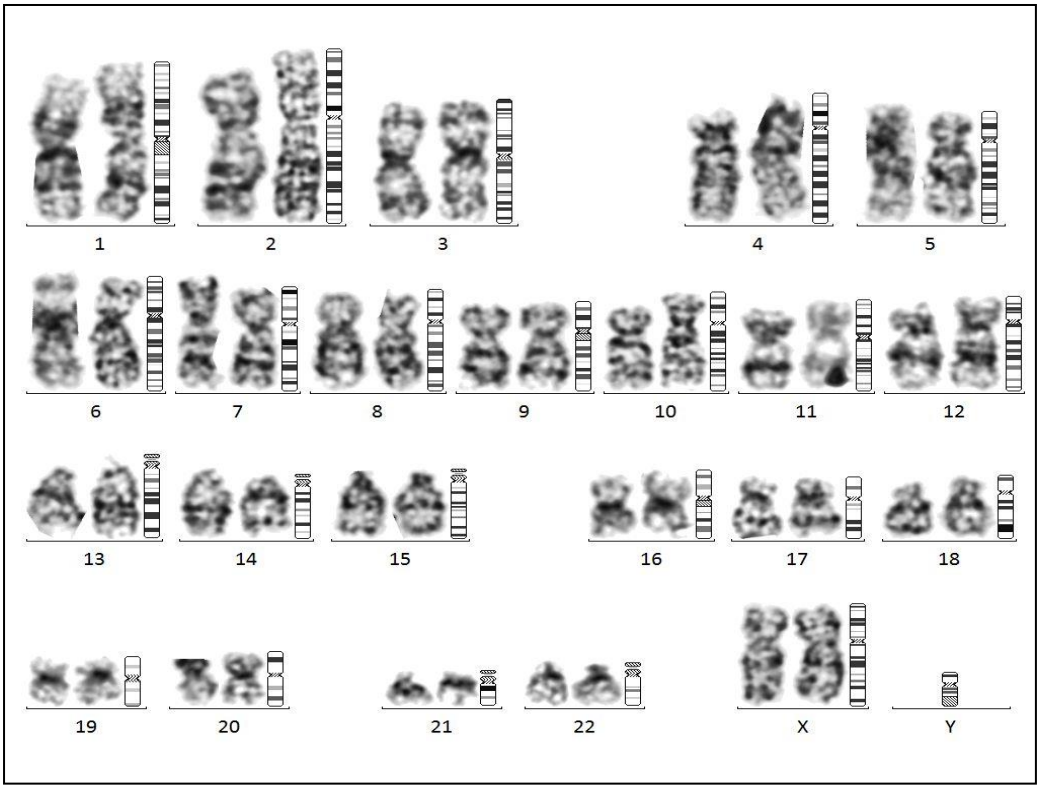


Figure 11: A normal karyotype of an unaffected family member

6.2 TARGETED EXOME SEQUENCING RESULTS:

TES was performed with a gene panel consisting of 5469 genes, which gave 27 pathogenic mutations after filtration from .fastq files. These genes are listed below in Table 4.

Out of these 27 pathogenic mutations, seven are related to neurological disorders. Among them are DPYD, BBS2, GLI2, CLN6, PRODH, HEXB, and MAG shown in Table 3 alongwith the gene alteration, aminoacid change, related disorders, position and rsID.

6.2.1 QUALITY ASSESSMENT:

a) Quality Control: FastQC has been used for quality assessment. TES generates millions of sequences in a single run so it is good to check the quality of the data as it affects the result.Total sequences generated were 27446521.

b) Per base sequence quality: An overview of the range of quality values at each position of the bases in fastq files have been described here. The y-axis represents the quality scores of the bases. The background of y-axis is divided into three regions which consists of green region (very good quality scores), orange (reasonable quality scores), and red region (poor quality scores). All the base calls lies in the green region (Figure 12) hence it is a good quality sequence. Central red line shows the median value.Yellow box shows inter-quartile region. Blue line shows mean quality. Y-axis indicates the Quality score.

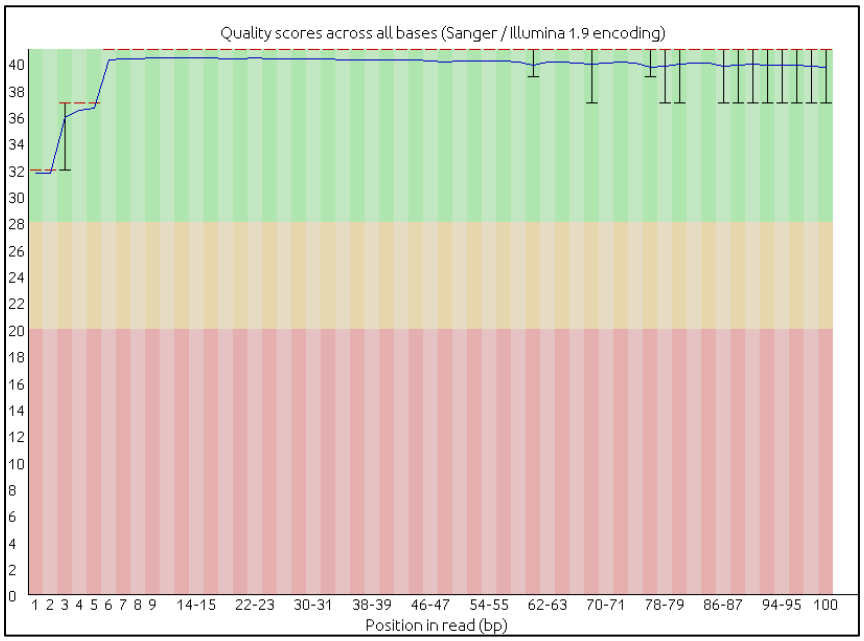


Figure 12: Per base quality

c) Per base sequence content: The graph plots out the total proportion of position of base that has been called for four normal DNA bases. Here, less overrepresented sequences can be seen as the difference between bases is less (Figure 13).

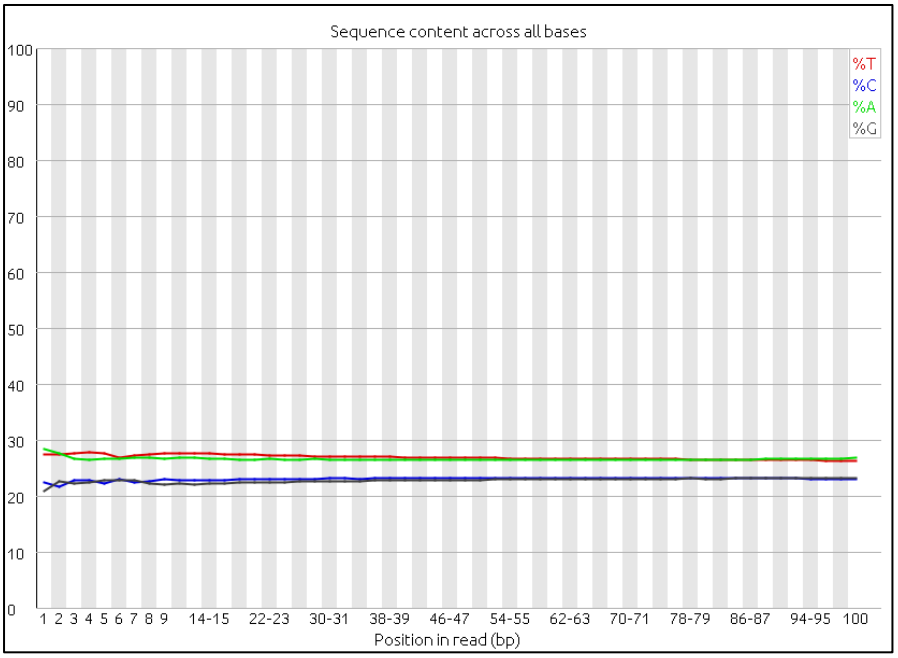


Figure 13: Per base sequence content

d) Per sequence GC content: The GC content across the whole length of each sequence was measured. The overlapping of two peaks confirms about less contamination while preparing the library (Figure 14).

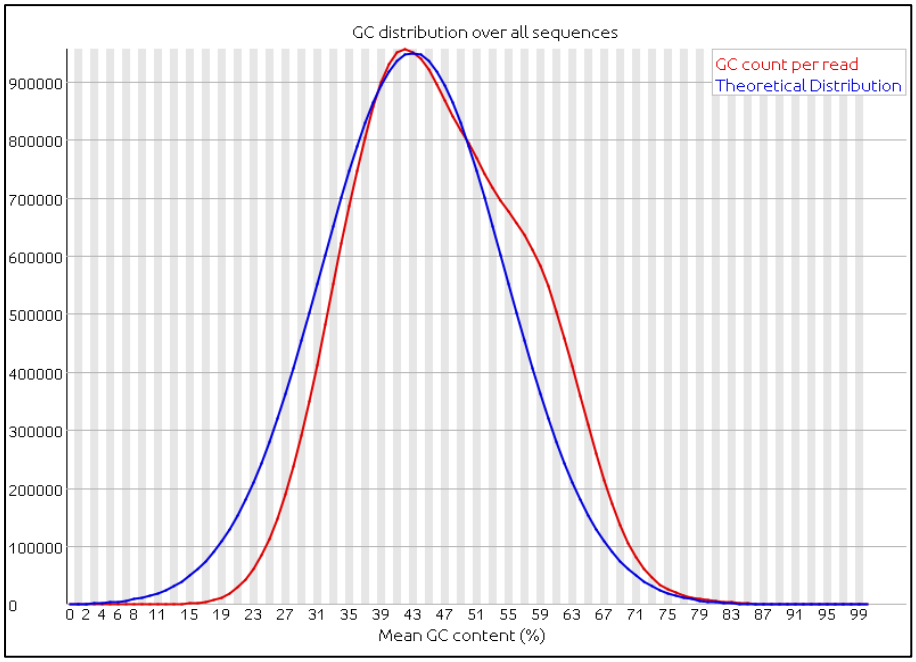


Figure 14: Per seuence GC content

6.2.2 Variant identification and filteration:

The DNA was sequenced keeping the average depth of 200X coverage. It was then compared with GRCh37.75 human genome reference sequencing using SnEff tool. Many filtering tools were used in the processing steps, for example Snpsift, which gave 36,802 variants as an output (Figure 15).

Genome	GRCh37.75
Date	2017-03-28 14:31
SnpEff version	SnpEff 4.3i (build 2016-12-15 22:33), by Pablo Cingolani
Command line arguments	SnpEff -stats aligned025.html GRCh37.75 aligned025.filter.vcf
Warnings	37,502
Errors	160
Number of lines (input file)	36,632
Number of variants (before filter)	36,804
Number of not variants (i.e. reference equals alternative)	0
Number of variants processed (i.e. after filter and non-variants)	36,802
Number of known variants (i.e. non-empty ID)	0 (0%)
Number of multi-allelic VCF entries (i.e. more than two alleles)	128
Number of effects	254,155
Genome total length	32,036,512,403
Genome effective length	3,095,677,432
Variant rate	1 variant every 84,117 bases

Figure 15: Summary of the report of targeted exome sequencing

a) SNPs and INDELs: SNPs can be in the form of transition and transversions. However, the frequency of SNP occurrence varies from person to person and not all of them are pathogenic. They are classified on the basis of their clinical significance by ACMG Guidelines. INDELs are genomic insertion and deletions. They can vary from single nucleotide change to fewer large regions. Number of SNPs are higher than number of insertions and deletions (Table 4)

TYPE	TOTAL
SNP	34,146
INS	1329
DEL	1327
TOTAL	36,802

Table 3: Number of variants by type

b) Transition and Transversion: Transition is the change between purine-purine or pyrimidine-pyrimidine nucleotides. Transversion is the change between purine-pyrimidine nucleotides and vice versa. The numbers of transversions were found to be higher than transitions. The ideal range of transition and transversion ratio should be in-between 2-4. In this case, the ratio is 2.4 which fall in the normal range (Figure 16).

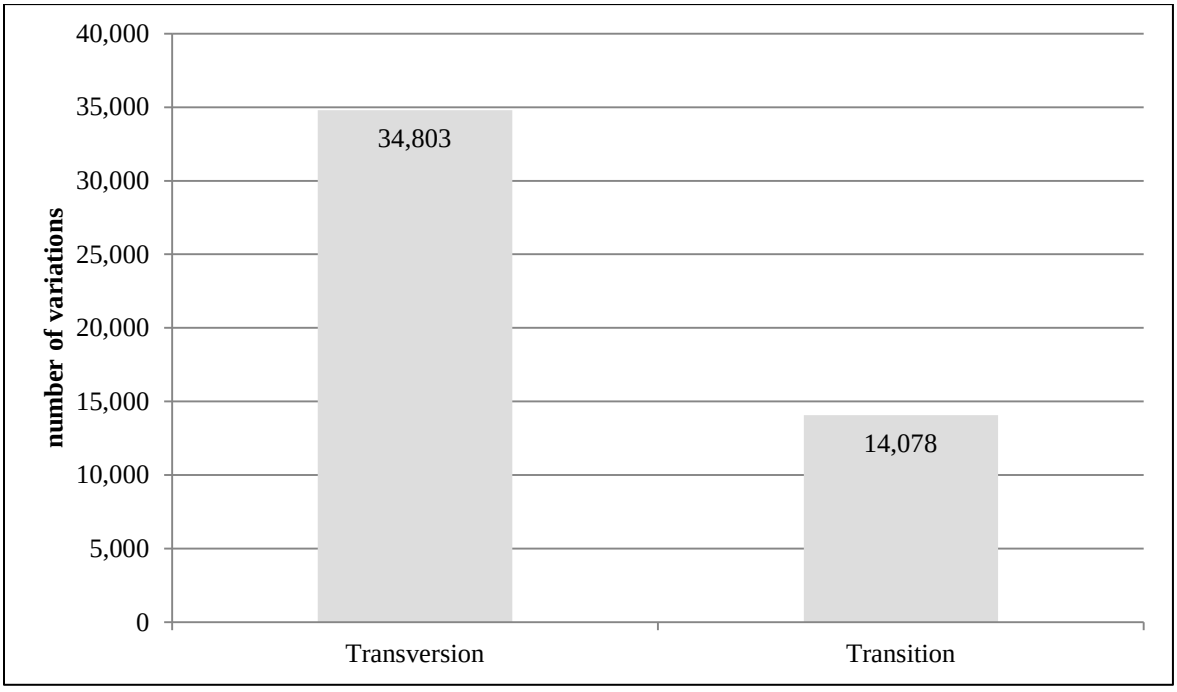


Figure 16 – Ratio of transition/transversion

c) Missense and Non-sense variations:

A non-sense mutation occurs when a codon gets converted into stop codon. Missense mutations are those mutations in which a change in a single nucleotide in a codon results in change in codon that code for a new aminoacid. Silent mutations are less significant hence they are not taken into consideration here (Figure 17).

Number of variations found in intronnic regions was found to be higher than variations in exons. However, exonic variations are more responsible for phenotypic and genotypic change (Figure 18).

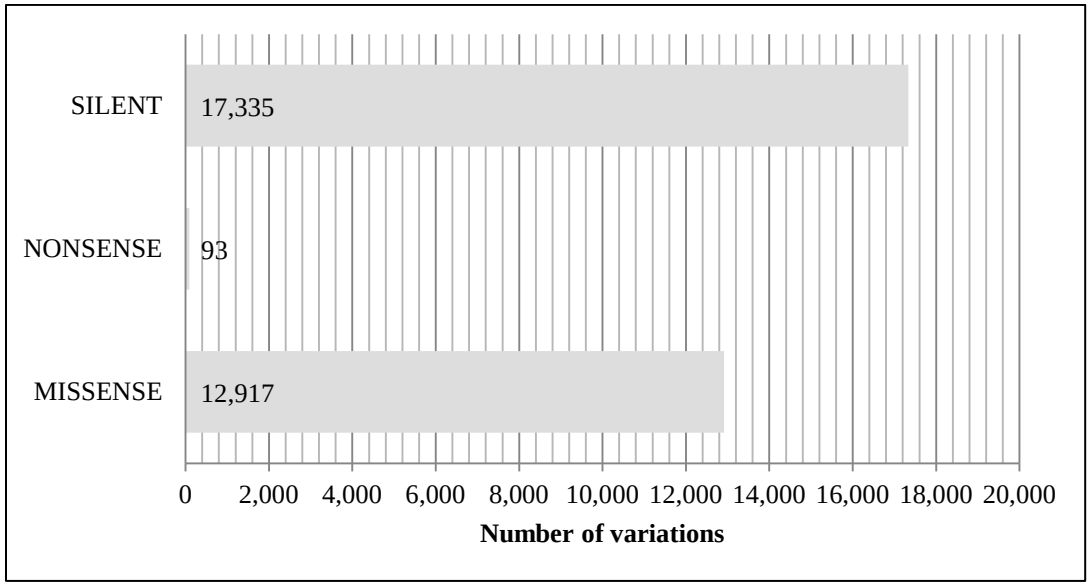


Figure 17: Number of variants by functional class

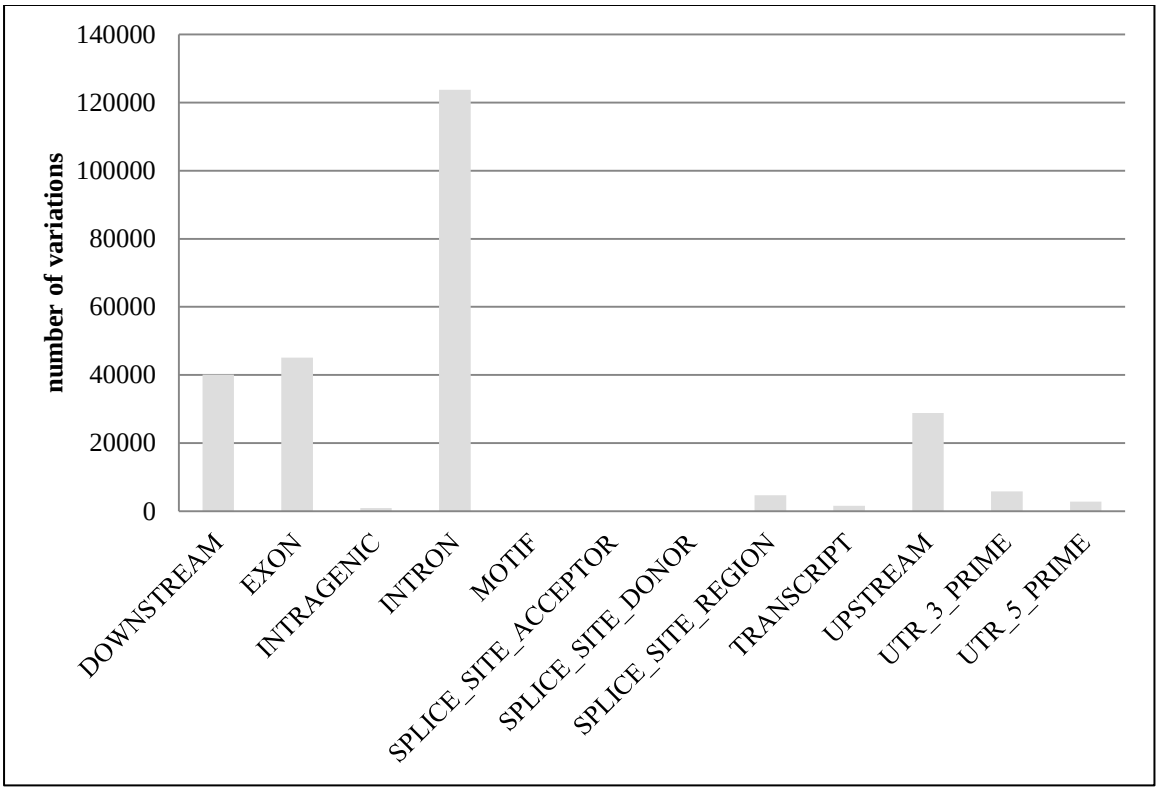


Figure 18: Number of Effects by type

d) Clinical mutation identification:

CHR. NO.	GENES	DISORDERS	PROTIEN	ALT	rsID
1	DPYD	NEUROLOGICAL DISORDER	CYS29=,CYS29 ARG	c.85T>C	rs1801265
16	BBS2	NEUROLOGICAL DISORDER	ASN70=,ASN70 SER	c.209A>G	rs4784677
2	GLI2	NEUROLOGICAL DISORDER	ASP1520ASN	c.4558G>A	rs114814747
15	CLN6	NEUROLOGICAL DISORDER	GLU72TER,GLU72GLN	c.214G>T	rs104894483
22	PRODH	NEUROLOGICAL DISORDER	GLN521ARG	c.1562A>G	rs450046
5	HEXB	NEUROLOGICAL DISORDER	SER62,SER62LEU	c.185C>T	rs820878
19	MAG	NEUROLOGICAL DISORDER	SER133ARG	c.399C>G	rs2301600
6	RHAG	HEMOLYTIC ANEMIA	VAL270ILE	c.808G>A	rs16879498
6	F13A1	HEMATOLOGIC DISORDER		c.-19+12A=	rs2815822
9	ADAMTS13	HEMOLYTIC ANEMIA	GLN448GLU	c.1342C>G	rs2301612
5	IL7R	IMMUNODEFICIENCY SYNDROME	VAL138ILE	c.197T>C	rs1494558
11	RAG2	IMMUNODEFICIENCY SYNDROME	THR215ILE	c.644C>T	rs35691292

5	1L7R	IMMUNODEFICIENCY SYNDROME	ILE66THR	c.197T>C	rs1494558
1	DBT	METABOLIC DISORDER		c.1150G>A	rs12021720
12	HPD	METABOLIC DISORDER	ALA33THR	c.-21G>A	rs1154510
18	FECH	METABOLIC DISORDER		c.68-23C>T	rs2269219
5	IRGM	BOWEL DISEASE	LEU105=	c.313C>T	rs10065172
14	RPGRIP1	EYE DISORDER	ALN547SER	c.1639G>T	rs10151259
7	ATP6V0A4	KIDNEY AND URINARY TRACT DISORDER	MET580THR	c.1739T>C	rs3807153
12	HNF1A	DIABETES	GLY574=,GLY574SER	c.1720G>A	rs1169305
X	SERPINA7		LEU303PHE	c.909G>T	rs1804495
1	F5	BLOOD COAGULATION	ARG534=,ARG534GLN	c.1601G>A	rs6025
3	CP	IRON DEFICIENCY	HIS117=,HIS117GLN	c.2991T>G	rs34394958
12	KRT74	CONGENITAL DISORDER	PHE274SER	c.821T>C	rs147962513
12	KRT85	CONGENITAL DISORDER	ARG78HIS	c.233G>A	rs61630004

11	ANKK1		GLU713LYS	c.2137G>A	rs1800497
11	RAG1		ARG249HIS,ARG249GLnfs	c.746delGinsAA	rs3740955;rs886041745

Table 4 –List of pathogenic variants obtained from targeted exome sequencing

The clinically relevant mutations gained from annotation by SnpEff were then mapped by variants which are already published in literature and other databases. Variants which had rsID (Reference SNP cluster ID) were filtered out from the raw vcf file. Clinical significance was classified by ACMG guidelines while other functional aspects were provided by Clinvar. CLNSG=5 is the most important terms so we filtered other parameters manually.

CLNSIG = 5 showed 27 pathogenic genes which are classified in Table 1.

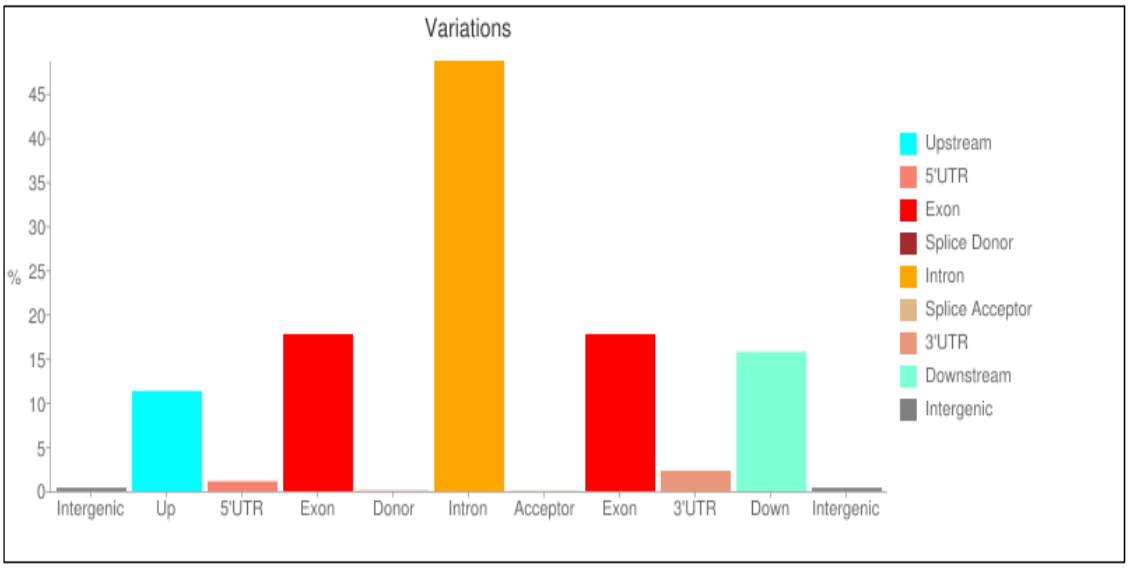


Figure 19: Graph depicting the number of variation according to the region and it seems that the inton region has the highest number of variations

Our study generated a list of sequence variants that will enrich the spectrum of genetic variants implicated in epilepsy. Even, computational filtering is not an ideal solution, as it still misses errors and reduces the sensitivity to detect the variants (DNA damage or DNA variant?)

The obtained results were confirmed by IGV (Integrated Genomics Viewer)

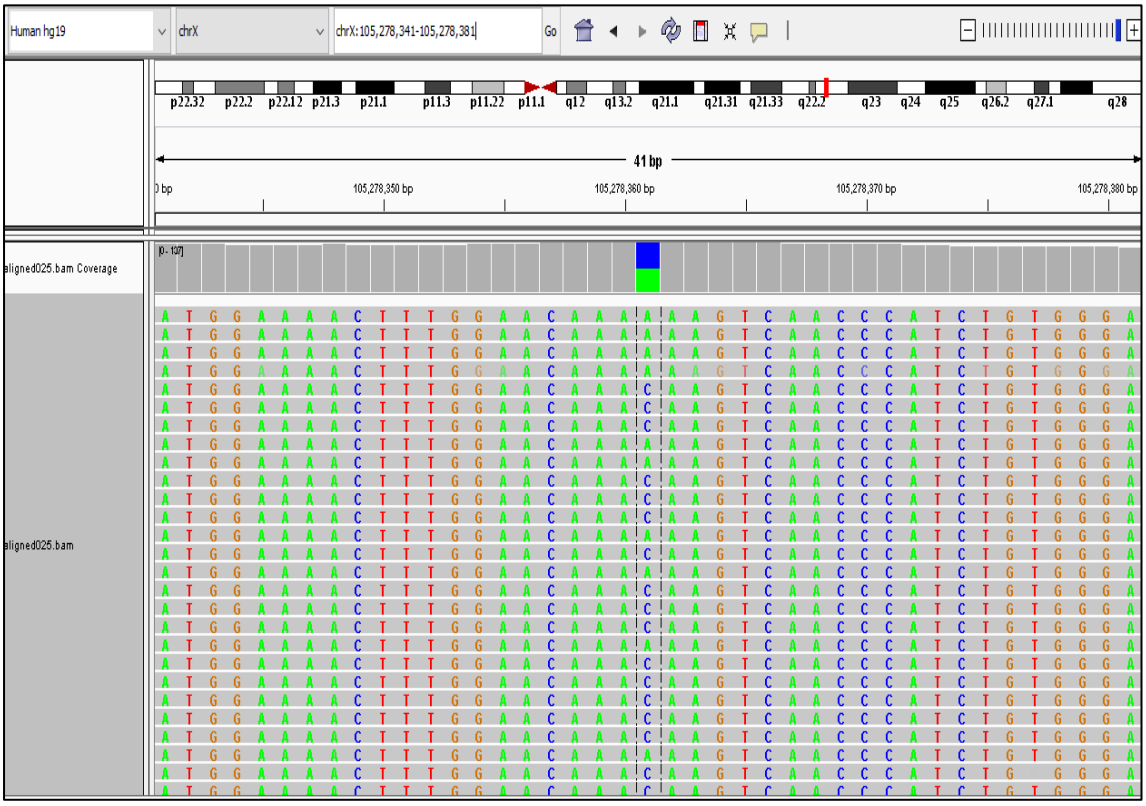


Figure 20: IGV RESULTS- visualization of the mutations for confirmation of mutation. This has been done with all 27 mutations.

Visualization of genes confirmed that the analysis has been done proper.

Epilepsies present with a broad range of clinical features and their genetic causes remain unknown in the vast majority of cases, although several genes have been identified in rare Mendelian forms, either heritable or sporadic. Finding the disease genes can be

challenging, as the same epileptic phenotype may be associated with several genes. Presently, high-throughput sequencing is becoming the most promising approach to improve molecular diagnosis of this condition, although the interpretation of the results is far from being a standardized process. Recent genomics studies using next generation sequencing (NGS) technique have identified a large number of genetic variants including copy number variant (CNV) and single nucleotide variant (SNV) in a small set of genes from individuals with epilepsy. These discoveries have contributed significantly to evaluate the etiology of epilepsy in clinic and lay the foundation to develop molecular specific treatment. Recently, targeted capture using NGS panels have been used in both research and clinical molecular genetics laboratories to understand the molecular etiology of epilepsy disorders and gene identifications in monogenic epilepsies including both familial epilepsies and severe epilepsies. The NGS panels used in research studies vary in gene composition, sample size, inclusion criteria, and sequence analytic protocol (The Hitchhiker's Guide to Whole Exome Analysis by Shrey Gandhi, Vinod Scaria). These technologies have also been successfully used in the research setting for gene discovery in large cohorts of individuals with Autism Spectrum Disorder (ASD), ID/DD, epilepsy, and schizophrenia, generating vast amounts of publicly available genomic data and some has also been translated into treatment recommendations (Li et al, 2009). Indeed, next-generation sequencing (NGS) does not magically make diagnoses but typically provides a handful of possibilities requiring further studies on the function of each candidate gene.

To overcome these problems, we composed a panel containing most epilepsy genes, covering several relevant phenotypes (Landrum et al, 2009). These findings support clinical usage of the NGS panel in epilepsy clinic to determine the molecular bases underlying epilepsy. It also develops a comprehensive knowledge base of candidate genes related to developmental brain disorders. Genes were prioritized based on the inheritance pattern. In these study we obtained 27 pathogenic mutations with the help of targeted exome sequencing which were prioritize based on the strength of evidence and their level of clinical significance. We excluded benign, likely benign, VUS and likely pathogenic mutations to avoid ambiguity in the interpretation of their functional consequences.

This approach results in higher specificity and reduced sensitivity (Landrum et al, 2014).

According to ACMG Guidelines, variants found through genetic testing are currently classified and reported as follows:

- Pathogenic Mutation: Alterations with sufficient evidence to classify as pathogenic (capable of causing disease). Targeted testing of at-risk family members and appropriate changes in medical management (*i.e.* high-risk surveillance) for pathogenic mutation carriers recommended. A pathogenic mutation is always included in results reports.

- Likely Pathogenic: alterations with strong evidence in favor of pathogenicity. Targeted testing of at-risk family members and appropriate changes in medical management (*i.e.* high-risk surveillance) for likely pathogenic carriers recommended. A likely pathogenic is always included in results reports.

- Variant, Unknown Significance (VUS): Alterations with limited and/or conflicting evidence regarding pathogenicity. Targeted testing of informative family members to collect cosegregation data via our Family Studies Program recommended. Medical management based on personal and family histories, not VUS carrier status. A VUS is always included in results reports.

- Likely Benign: alterations with strong evidence against pathogenicity Targeted testing of at-risk family members not recommended. Medical management based on personal and family histories. A benign is not routinely included in results reports.

- Benign: alterations with very strong evidence against pathogenicity. Targeted testing of at-risk family members not recommended. Medical management based on personal and family histories. Benign alterations are not routinely included in results reports (ACMG GUIDELINES)

After identifying mutations, all the 27 mutations were categorized according to their clinical significance, level of pathogenicity and their functional aspect.

1) DPYD: Cytogenetic location: 1p21.3
   Phenotype: Dihydropyrimidine dehydrogenase deficiency
   Clinical significance: Pathogenic, not provided
   Inheritance: Autosomal recessive
   Description: The DPYD gene encodes dihydropyrimidine dehydrogenase, the initial and rate-limiting enzyme in the catabolism of the pyrimidine bases uracil and thymine. Mutations in this gene result in dihydropyrimidine dehydrogenase deficiency, an error in pyrimidine metabolism associated with thymine-uraciluria and an increased risk of toxicity in cancer patients receiving 5-fluorouracil chemotherapy (OMIM). Dihydropyrimidine dehydrogenase may have important roles for neurodevelopment as this pathway is the only endogenous source of the neurotransmitter β-alanine. Phenotypic variation is large, with convulsive disorders, microcephaly, motor retardation, and mental retardation the most frequent manifestations (Au Km et al, 2003).
   DPYD Deficiency can result in premature termination, substitution and the common splice donor site mutation which can lead to convulsive disorders (genecard).

2) BBS2: Cytogenetic location: 16q13
   Phenotype: Bardet-Biedl syndrome 2
   Clinical significance: Benign, pathogenic
   Inheritance: Autosomal recessive
   Description: BBS2 is 1 of 7 BBS proteins that form the stable core of a protein complex required for ciliogenesis. The gene is evolutionarily conserved and displays a wide pattern of tissue expression, including brain, kidney, adrenal gland, and thyroid gland.
   Clinical features like mild developmental delay, speech delay, congenital heart disease, poor coordination, and an increased incidence of diabetes mellitus and hypertension

have been reported in BBS patients (Dong jun Xing,2014). Bardet-Biedl syndrome (BBS) is an autosomal recessive disorder characterized by central obesity, mental impairment, rod-cone dystrophy, polydactyly, hypogonadism in males, and renal abnormalities(Hirano M et al,2015) To date, mutations in 21 different genes have been described as being responsible for BBS (Ece Solmaz A et al,2015)

3) GLI2: Cytogenetic location: 2q14.2
   Phenotype: Culler-Jones syndrome, Holoprosencephaly 9
   Clinical significance: Conflicting interpretation
   Inheritance: Autosomal dominant
   Description: The GLI2 gene encodes a vertebral transcription factor involved in SHH signal transduction and are also associated with Hedge-hog signalling. On the basis of the functional domains affected by known human GLI2 mutations, mutant GLI2 proteins exhibited either loss-of-function or dominant-negative activity (OMIM). A data from a paper implicate perturbation of the Shh pathway in at least 37% of individuals with the HH (Hypothalamic hamartoma) epilepsy syndrome, consistent with the concept of a developmental pathway brain disease (Michael S, 2016). Children with HPE have many medical problems: developmental delay and feeding difficulties, epilepsy, instability of temperature, heart rate and respiration (Dubourg C, 2003).

4) CLN6: Cytogenetic location: 15q23
   Phenotype: Ceroid lipofuscinosis, neuronal, 6
   Clinical significance: Likely benign, pathogenic
   Inheritance: Autosomal recessive
   Description: CLN6 encodes a deduced 311-amino acid protein with 7 predicted transmembrane domains and a predicted molecular mass of 36 KD. Northern blot analysis detected expression of a major 2.4-kb CLN6 mRNA in adult and embryonic brain and in peripheral tissues of mouse and human. Recent data show that CLN6 mutation may cause severe neurologic disorders beginning in both

childhood and adult life (Laura Canafoglia, 2015). The neuronal ceroid-lipofuscinoses (NCL) is a group of neurodegenerative disorders characterized by epilepsy, visual failure, progressive mental and motor deterioration, myoclonus, dementia and reduced life expectancy (Patino LC et al, 2014).

5) PRODH: Cytogenetic location: 22q11.21

   Phenotype: Hyperprolinemia, type I, Schizophrenia, susceptibility to, 4

   Clinical significance: Pathogenic

   Inheritance: Autosomal recessive - Hyperprolinemia, type I

   Autosomal dominant - Schizophrenia, susceptibility to, 4

   Description: This gene encodes a mitochondrial protein that catalyzes the first step in proline degradation. Mutations in this gene are associated with hyperprolinemia type 1 and susceptibility to schizophrenia 4 (SCZD4) (Genecard). PRODH is one of the candidate genes for susceptibility to schizophrenia and other neurological disorders. It codes for a proline dehydrogenase enzyme, which catalyses the first step of proline catabolism and most likely is involved in neuromediator synthesis in the CNS (Maria Santsova, 2013). PRODH was identified as a schizophrenia-associated gene in some studies. PRODH is a mitochondrial enzyme catalyzing the first step of proline catabolism, and PRODH malfunctions cause hyperprolinemia type I. This condition is known to cause neurological abnormalities. PRODH also may play a role in the synthesis of neuromediators and PRODH malfunction may cause a neuromediator misbalance.

6) HEXB: Cytogenetic location: 5q13.3

   Phenotype: Sandhoff disease, infantile, juvenile, and adult forms

   Clinical significance: Benign, Pathogenic

   Inheritance: Autosomal recessive

   Description: Beta-hexosaminidase is composed of two subunits, alpha and beta, which are encoded by separate genes.The HEXB gene encodes the beta subunit of the enzyme hexosaminidase, which is involved in the breakdown of gangliosides. Mutations in the

alpha or beta subunit genes lead to an accumulation of GM2 ganglioside in neurons and neurodegenerative disorders termed the GM2 gangliosidoses. Beta subunit gene mutations lead to Sandhoff disease (GM2-gangliosidosis type II). (Genecard). Sandhoff disease is a rare, genetic, lipid storage disorder characterized by progressive degeneration of the nerve cells (neurons) in the brain and spinal cord (Masri A et al, 2014).

7) MAG: Cytogenetic location: 19q13.12

   Phenotype: Spastic paraplegia 75, autosomal recessive

   Clinical significance: Pathogenic

   Inheritance: Autosomal recessive

   Description: Myelin-associated glycoprotein (MAG) is a cell adhesion molecule involved in myelin maintenance and glia-axon interaction (OMIM). It is thought to be involved in the process of myelination. A lectin binds to sialylated glycoconjugates and mediates certain myelin-neuron cell-cell interactions (genecard). MAG interacts directly with axons of nerve cells.

   These are the seven genes and their clinical significance and epilepsy has been described as the primary cause in these mutations. Other 20 genes have been listed below which are not directly responsible for causing epilepsy but it is one of their secondary effects.

8) RHAG: Cytogenetic location: 6p12.3

   Phenotype: Anemia, hemolytic, Rh-null, regulator type (Autosomal recessive)

   Overhydrated hereditary stomatocytosis (Autosomal dominant)

   Clinical significance: Pathogenic

   Description: The protein encoded by this gene is erythrocyte-specific and is thought to be part of a membrane channel that transports ammonium and carbon dioxide across the blood cell membrane. The encoded protein appears to interact with Rh blood group antigens and Rh30 polypeptides. Defects in this gene are a cause of regulator type Rh-null hemolytic anemia (RHN), or Rh-deficiency syndrome (genecard)

9) F13A1: Cytogenetic location: 6p25.1

Phenotype: Factor XIIIA deficiency -Autosomal recessive

Venous thrombosis, protection against - Autosomal dominant

Clinical significance: Benign, Pathogenic

Description: The F13A1 gene encodes the A subunit of factor XIII, the last enzyme generated in the blood coagulation cascade. It is the zymogen for fibrinoligase, a transglutaminase that forms intramolecular gamma-glutamyl-epsilon-lysine crosslinking between fibrin molecules and thus stabilizes blood clots (OMIM).

10) ADAMTS13: Cytogenetic location: 9q34.2

Phenotype: Thrombotic thrombocytopenic purpura, familial

Inheritance: Autosomal recessive

Clinical significance: Benign, Pathogenic

Description: ADAMTS13 is a multidomain protease that cleaves VWF (Von Willebrand factor) in circulating blood and thereby limits platelet thrombosis. Diseases associated with ADAMTS13 include Thrombotic Thrombocytopenic Purpura, Familial and Familial Thrombotic Thrombocytopenia Purpura (genecard)

11) IL7R: Cytogenetic location: 5p13.2

Phenotype: Severe combined immunodeficiency, T-cell negative, B-cell/natural killer cell-positive type

Inheritance: Autosomal recessive

Clinical significance: conflicting interpretation

2 mutations in the same gene has been observed

Description: The IL7R gene encodes a receptor for interleukin-7 (IL7; 146660), a 25-kD glycoprotein involved in the regulation of lymphopoiesis. Defects in this gene may be associated with severe combined immunodeficiency (SCID). Diseases associated with IL7R include Severe Combined Immunodeficiency, T Cell-Negative, B-Cell/Natural Killer-Cell Positive and Severe Combined Immunodeficiency (genecard)

12) RAG2: Cytogenetic location: 11p12

Phenotype: Severe combined immunodeficiency, B cell-negative, Omenn syndrome, Combined cellular and humoral immune defects with granulomas (OMIM)

Inheritance: Autosomal recessive

Clinical significance: conflicting interpretation

Description: This gene encodes a protein that is involved in the initiation of V (D) J recombination during B and T cell development. Mutations in this gene cause Omenn syndrome, a form of severe combined immunodeficiency associated with autoimmune-like symptoms (genecard)

13) DBT: Cytogenetic location: 1p21.2

Phenotype: Maple syrup urine disease, type II (OMIM)

Inheritance: Autosomal recessive

Clinical significance: benign/likely benign, pathogenic

Description: DBT (Dihydrolipoamide Branched Chain Transacylase E2) is a Protein Coding gene. Diseases associated with DBT include Maple Syrup Urine Disease, Type Ii and Intermediate Maple Syrup Urine Disease (genecard)

14) HPD: Cytogenetic location: 12q24.31

Phenotype: Hawkinsinuria (Autosomal Dominant)

Tyrosinemia, type III (Autosomal recessive) (OMIM)

Clinical significance: Pathogenic

Description: The protein encoded by this gene is a key enzyme in the degradation of tyrosine. The encoded protein catalyzes the conversion of 4-hydroxyphenylpyruvate to homogentisate. Defects in this gene are a cause of tyrosinemia type 3 (TYRO3) and hawkinsinuria (HAWK).

15) FECH: Cytogenetic location: 18q21.31

Phenotype: Protoporphyria, erythropoietic (OMIM)

Inheritance: Autosomal recesive

Clinical significance: Conflicting interpretation

Description: The protein encoded by this gene is localized to the mitochondrion, where it catalyzes the insertion of the ferrous form of iron into protoporphyrin IX in the heme synthesis pathway. Mutations in this gene are associated with erythropoietic protoporphyria (genecard)

16) IRGM: Cytogenetic location: 5q33.1

Phenotype: Inflammatory bowel disease (Crohn disease)

Clinical significance: Pathogenic

Description: This gene encodes a member of the p47 immunity-related GTPase family. The encoded protein may play a role in the innate immune response by regulating autophagy formation in response to intracellular pathogens. Polymorphisms that affect the normal expression of this gene are associated with a susceptibility to Crohn's disease and tuberculosis

17) RPGRIP1: Cytogenetic location: 14q11.2

Phenotype: Cone-rod dystrophy 13, Leber congenital amaurosis 6

Clinical significance: Conflicting interpretation

Desscription: Mutations in the retinitis pigmentosa GTPase regulator gene cause X-linked retinitis pigmentosa-3 (RP3), a severe, progressive, and degenerative retinal dystrophy that eventually leads to complete blindness.

18) ATP6V0A4: Cytogenetic location: 7q34

Phenotype: Renal tubular acidosis

Inheritance: Autosomal recessive

Clinical significance: Conflicting interpretation

Description: This gene encodes a component of vacuolar ATPase (V-ATPase), a multisubunit enzyme that mediates acidification of intracellular compartments of

eukaryotic cells. Mutations in this gene are associated with renal tubular acidosis associated with preserved hearing

19) HNF1A: Cytogenetic location: 12q24.31

Phenotype: MODY, type III (Autosomal Dominant), Renal cell carcinoma, Diabetes mellitus (Autosomal recessive)

Clinical significance: Conflicting interpretation

Description: The protein encoded by this gene is a transcription factor required for the expression of several liver-specific genes. Defects in this gene are a cause of maturity onset diabetes of the young type 3 (MODY3) and also can result in the appearance of hepatic adenomas.

20) SERPINA7: Cytogenetic location: Xq22.3

Phenotype: Thyroxine-binding globulin QTL

Clinical significance: Pathogenic

Description: This gene encodes the major thyroid hormone transport protein, TBG, in serum. Mutations in this gene result in TBG deficiency, which has been classified as partial deficiency, complete deficiency, and excess, based on the level of serum TBG.

21) F5: Cytogenetic location: 1q24.2

Phenotype: Factor V deficiency (Autosomal recessive), Thrombophilia due to activated protein C resistance (Autosomal dominant)

Clinical significance: Pathogenic, Drug response

Description: This gene encodes an essential cofactor of the blood coagulation cascade. This factor circulates in plasma, and is converted to the active form by the release of the activation peptide by thrombin during coagulation. Defects in this gene result in either an autosomal recessive hemorrhagic diathesis or an autosomal dominant form of thrombophilia, which is known as activated protein C resistance.

22) CP: Cytogenetic location: 3q24-q25

Phenotype: Cerebellar ataxia (Autosomal recessive), Hemosiderosis (Autosomal recessive)

Clinical significance: BENIGN/LIKELY BENIGN, PATHOGENIC

Description: The protein encoded by this gene is a metalloprotein that binds most of the copper in plasma and is involved in the peroxidation of Fe (II) transferrin to Fe (III) transferrin. Mutations in this gene cause aceruloplasminemia, which results in iron accumulation and tissue damage, and is associated with diabetes and neurologic abnormalities.

23) KRT74: Cytogenetic location: 12q13.13

Phenotype: Ectodermal dysplasia 7 (Autosomal recessive), Woolly hair (autosomal dominant)

Clinical significance: CONFLICTING INTERPRETATION

Description: Keratins are intermediate filament proteins responsible for the structural integrity of epithelial cells and are subdivided into epithelial keratins and hair keratins. Diseases associated with KRT74 include Ectodermal Dysplasia 7, Hair/Nail Type and Woolly Hair.

24) KRT85: Cytogenetic location: 12q13.13

Phenotype: Ectodermal dysplasia 4

Inheritance: Autosomal recessive

Clinical significance: Pathogenic

Description: The protein encoded by this gene is a member of the keratin gene family. As a type II hair keratin, it is a basic protein which heterodimerizes with type I keratins to form hair and nails. It is related to congenital disorder.

25) ANKK1: Cytogenetic location: 11q23.2

Phenotype: Dopamine receptor D2 (reduced brain density)

Clinical significance: DRUG RESPONSE

Description: Diseases associated with ANKK1 include Alexithymia and Impulse Control disorder. It is not clear if this gene plays any role in neuropsychiatric disorders.

26) RAG1: Cytogenetic location: 11p12

Phenotype: Omenn syndrome, Severe combined immunodeficiency, granulomas

Inheritance: Autosomal recessive

Clinical significance: BENIGN, PATHOGENIC

Description: The protein encoded by this gene is involved in activation of immunoglobulin V-D-J recombination. Defects in this gene can be the cause of several diseases.

6.2.3 LIMITATIONS:

Despite these significant progresses, the molecular etiologies for the majority of epilepsy patients remain elusive. For the most of epilepsy patients with known genetic causes the genotype phenotype correlation has not yet been fully delineated. Diagnosing the patient can be challenging as there is both genetic heterogeneity for a given epilepsy syndrome and phenotypic heterogeneity for a specific gene (https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual). VUS and likely pathogenic variants have also been excluded from the study which can be the cause of mutation. However, we noted that the positive detection rate for causal variants is not significantly increased because our finding may suggest a challenge of choosing candidate genes genome wide on the basis of functional prediction and in silico analysis. Very little is known about the underlying molecular mechanism contributing to the variable expressivity associated with these mutations.

In known epileptic encephalopathy genes, the full phenotypic spectrum associated with mutations is not known. Very few studies have investigated the role of any given gene across a wide spectrum of epileptic encephalopathy syndromes. This makes the majority of

cases remain unexplained because serial gene testing in the clinical setting is an inefficient and expensive process. Furthermore, it is necessary to faciliate diagnosis of the additional gene that causes epileptic encephalopathies. Moreover, functional studies performed with experimental models have not yet provided results that have been directly translated to the clinical setting. This is a disappointing lack of progress, reflecting as it does on the complex nature of epilepsies and their treatment (Landrum et al, 2014).

In summary, our study supported the use of NGS panel as an effective tool to detect the genetic cause for epilepsy of unknown etiology. Our study generated a list of novel sequence variants that will enrich the spectrum of genetic variants implicated in epilepsy. The analysis of a proband with unidentified genetic anomaly provided list of sequence variants for further functional investigation to understand the complex molecular mechanism underlying this familial occurrence epilepsy.

.

The roles of these genes in epilepsy have not been reported in literature and additional studies are clearly warranted to support the pathogenicity. It would be interesting in future study to compare them systematically in parallel with larger sample size and determine whether a significant modifier effect may be present in different genetic background

These results suggest that pathogenic mutations in these genes, while important, can maybe the causes of epileptic encephalopathies, or cause only very distinct syndromes that were not prevalent in our cohort. These findings support a clinical approach to genetic diagnosis that employs large gene panels or whole exome sequencing (https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual). Applying this approach across various neurodevelopmental disorders will identify additional mutation positive patients for a specific gene and enhance our understanding of disease mechanisms (https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual). We emphasize that sequencing and careful interpretation could result in a correct genetic

diagnosis for such patients and ultimately facilitate changes in patient management (The Hitchhiker's Guide to Whole Exome Analysis By Shrey Gandhi, Vinod Scaria)

Obviously novel mutations require that their causative role is further confirmed by segregation or functional analyses (https://samtools.github.io/bcftools/bcftools.html). The correlations between phenotype and genotype in genetic epilepsies is rapidly changing in relation to new findings emerging from exome sequencing and the use of diagnostic panels as unexpected phenotypes become associated with mutations of specific genes and vice versa. A constantly updated database will be essential to establish all the known gene mutations and polymorphisms, and their clinical correlates, so that genotype–phenotype correlations can be determined (Lanthrum et al, 2014). It is hoped that a constantly updated database will be established for all the known gene mutations and polymorphisms, and their clinical correlates, so that genotype–phenotype correlations can be determined. (Lanthrum et al, 2014)

REFERENCE:

1. Ramirez-Gonzalez, Ricardo H., et al. "StatsDB: platform-agnostic storage and understanding of next generation sequencing run metrics." F1000 Research 2 (2014).
2. https://biof-edu.colorado.edu/videos/dowell-short-read-class/day-4/fastqc-manual
3. The Hitchhiker's Guide to Whole Exome Analysis By Shrey Gandhi, Vinod Scaria
4. http://bio-bwa.sourceforge.net/
5. Li, Heng, et al. "The sequence alignment/map format and SAMtools." Bioinformatics 25.16 (2009): 2078-2079.
6. https://samtools.github.io/bcftools/bcftools.html
7. Landrum, Melissa J., et al. "ClinVar: public archive of relationships among sequence variation and human phenotype." Nucleic acids research 42.D1 (2014): D980-D985.
8. Robinson, James T., et al. "Integrative genomics viewer." Nature biotechnology 29.1 (2011): 24-26.
9. NGS pipeline Dolled-Filhart et al, 2012
10. Nomenclature, Human Cytogenomic. "ISCN 2016." Fetal Diagn Ther 40.1 (2016): 1-80.
11. DNA variants or DNA damage?
12. Richards, Sue, et al. "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 17.5 (2015): 405-423.
13. https://www.omim.org/entry/612779?search=rs1801265&highlight=rs1801265
14. http://www.genecards.org/cgi-bin/carddisp.pl?gene=DPYD&keywords=rs1801265
15. Au, K. M., et al. "Diagnosis of dihydropyrimidine dehydrogenase deficiency in a neonate with thymine-uraciluria." Hong Kong Medical Journal 9.2 (2003): 130-133.
16. Hirano, Makito, et al. "The First Nationwide Survey and Genetic Analyses of Bardet-Biedl Syndrome in Japan." PloS one 10.9 (2015): e0136317.
17. Solmaz, Asli Ece, et al. "Targeted multi-gene panel testing for the diagnosis of Bardet Biedl syndrome: identification of nine novel mutations across BBS1, BBS2, BBS4, BBS7, BBS9, BBS10 genes." European journal of medical genetics 58.12 (2015): 689-694.

18. Xing, Dong-Jun, et al. "Comprehensive molecular diagnosis of Bardet-Biedl syndrome by high-throughput targeted exome sequencing." PloS one 9.3 (2014): e90599.
19. https://www.omim.org/entry/606151?search=rs4784677&highlight=rs4784677
20. https://www.omim.org/entry/165230?search=rs114814747&highlight=rs114814747
21. Hildebrand, Michael S., et al. "Mutations of the sonic hedgehog pathway underlie hypothalamic hamartoma with gelastic epilepsy." The American Journal of Human Genetics 99.2 (2016): 423-429.
22. Dubourg, Christèle et al. "Holoprosencephaly." Orphanet Journal of Rare Diseases 2 (2007): 8. PMC. Web. 30 Apr. 2017.
23. http://www.genecards.org/cgi-bin/carddisp.pl?gene=PRODH&keywords=rs450046
24. Suntsova, Maria, et al. "Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH." Proceedings of the National Academy of Sciences 110.48 (2013): 19472-19477.
25. https://www.omim.org/entry/606873?search=rs820878&highlight=rs820878
26. http://www.genecards.org/cgi-bin/carddisp.pl?gene=HEXB&keywords=rs820878
27. Masri, Amira, et al. "Homozygous p. R284* mutation in HEXB gene causing Sandhoff disease with nystagmus." European Journal of Paediatric Neurology 18.3 (2014): 399-403.
28. http://www.genecards.org/cgi-bin/carddisp.pl?gene=MAG&keywords=rs2301600
29. https://www.omim.org/entry/159460?search=rs2301600&highlight=rs2301600
30. http://www.genecards.org/cgi-bin/carddisp.pl?gene=RHAG&keywords=rs16879498
31. http://www.genecards.org/cgi-bin/carddisp.pl?gene=IL7R&keywords=rs1494558
32. http://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAMTS13&keywords=rs2301612
33. http://www.genecards.org/cgi-bin/carddisp.pl?gene=RAG2&keywords=rs35691292
34. https://www.omim.org/entry/179616?search=rs35691292&highlight=rs35691292
35. http://www.genecards.org/cgi-bin/carddisp.pl?gene=DBT&keywords=rs12021720
36. http://www.genecards.org/cgi-bin/carddisp.pl?gene=HPD&keywords=rs1154510
37. https://www.omim.org/entry/609695?search=rs1154510&highlight=rs1154510
38. https://www.omim.org/entry/612386?search=rs2269219&highlight=rs2269219
39. http://www.genecards.org/cgi-bin/carddisp.pl?gene=FECH&keywords=rs2269219
40. https://www.omim.org/entry/608212?search=rs10065172&highlight=rs10065172

41. http://www.genecards.org/cgi-bin/carddisp.pl?gene=IRGM&keywords=rs10065172
42. https://www.omim.org/entry/605446?search=rs10151259&highlight=rs10151259
43. http://www.genecards.org/cgi-bin/carddisp.pl?gene=SERPINA7&keywords=rs1804495
44. https://www.omim.org/entry/314200?search=rs1804495&highlight=rs1804495
45. http://www.genecards.org/cgi-bin/carddisp.pl?gene=HNF1A&keywords=rs1169305
46. https://www.omim.org/entry/142410?search=rs1169305&highlight=rs1169305
47. https://www.omim.org/entry/605239?search=rs3807153&highlight=rs3807153
48. http://www.genecards.org/cgi-bin/carddisp.pl?gene=F5&keywords=rs6025
49. https://www.omim.org/entry/612309?search=rs6025&highlight=rs6025
50. http://www.genecards.org/cgi-bin/carddisp.pl?gene=CP&keywords=rs34394958
51. http://omim.org/entry/117700?search=cp&highlight=cp
52. http://www.genecards.org/cgi-bin/carddisp.pl?gene=KRT74&keywords=rs147962513
53. http://omim.org/entry/608248?search=rs147962513&highlight=rs147962513
54. http://www.genecards.org/cgi-bin/carddisp.pl?gene=KRT85&keywords=rs61630004
55. http://omim.org/entry/602767?search=rs61630004&highlight=rs61630004
56. http://omim.org/entry/608774?search=rs1800497&highlight=rs1800497
57. http://www.genecards.org/cgi-bin/carddisp.pl?gene=ANKK1&keywords=rs1800497
58. http://omim.org/entry/179615
59. http://www.genecards.org/cgi-bin/carddisp.pl?gene=RAG1&keywords=rs3740955
60. Wang, Yimin, et al. "Genetic Variants Identified from Epilepsy of Unknown Etiology in Chinese Children by Targeted Exome Sequencing." Scientific Reports 7 (2017): 40319.
61. Carvill, Gemma L., et al. "Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1." Nature genetics 45.7 (2013): 825-830.
62. Patel, Jaina, and Saadet Mercimek-Mahmutoglu. "Epileptic encephalopathy in childhood: a stepwise approach for identification of underlying genetic causes." The Indian Journal of Pediatrics 83.10 (2016): 1164-1174.
63. Gonzalez-Mantilla, Andrea J., et al. "A cross-disorder method to identify novel candidate genes for developmental brain disorders." JAMA psychiatry 73.3 (2016): 275-283.

64. Møller, Rikke S., Hans A. Dahl, and Ingo Helbig. "The contribution of next generation sequencing to epilepsy genetics." Expert review of molecular diagnostics 15.12 (2015): 1531-1538.
65. Della Mina, Erika, et al. "Improving molecular diagnosis in epilepsy by a dedicated high-throughput sequencing platform." European Journal of Human Genetics 23.3 (2015): 354-362.
66. Guerrini, Renzo, Carla Marini, and Massimo Mantegazza. "Genetic epilepsy syndromes without structural brain abnormalities: clinical features and experimental models." Neurotherapeutics 11.2 (2014): 269-285.
67. Amudhan, Senthil, Gopalkrishna Gururaj, and Parthasarathy Satishchandra. "Epilepsy in India I: Epidemiology and public health." Annals of Indian Academy of Neurology 18.3 (2015): 263.
68. http://www.epilepsy.com/learn/epilepsy-statistics
69. https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Epilepsies-and-Seizures-Hope-Through
70. Liu, Lin, et al. "Comparison of next-generation sequencing systems." BioMed Research International 2012 (2012).
71. Santhosh, Nandanavana Subbareddy, Sanjib Sinha, and Parthasarathy Satishchandra. "Epilepsy: Indian perspective." Annals of Indian Academy of Neurology 17.5 (2014): 3.
72. Metzker, Michael L. "Sequencing technologies—the next generation." Nature reviews genetics 11.1 (2010): 31-46.
73. http://www.cureepilepsy.org
74. Mirzaa, Ghayda M. et al. "Wide Spectrum of Developmental Brain Disorders from Megalencephaly to Focal Cortical Dysplasia and Pigmentary Mosaicism Caused by Mutations of MTOR." JAMA neurology 73.7 (2016): 836–845. PMC. Web. 1 May 2017.
75. Srour, Myriam, et al. "Dysfunction of the Cerebral Glucose Transporter SLC45A1 in Individuals with Intellectual Disability and Epilepsy." *The American Journal of Human Genetics* (2017).
76. Perucca, Piero, et al. "Real-world utility of whole exome sequencing with targeted gene analysis for focal epilepsy." *Epilepsy Research* 131 (2017): 1-8.