11-2020

IDENTIFICATION OF INHERITED AND DE NOVO EXOMIC VARIATIONS IN AN EMIRATI FAMILY WITH NEURODEVELOPMENTAL DISORDERS

Asmaa Samir Abdelaziz Refaey

United Arab Emirates University

College of Science

Department of Biology

IDENTIFICATION OF INHERITED AND *DE NOVO* EXOMIC VARIATIONS IN AN EMIRATI FAMILY WITH NEURODEVELOPMENTAL DISORDERS

Asmaa Samir Abdelaziz Refaey

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Science in Molecular Biology and Biotechnology

Under the Supervision of Dr. Ranjit Vijayan

November 2020

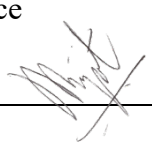
Declaration of Original Work

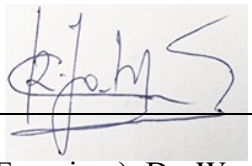
I, Asmaa Samir Abdelaziz Refaey, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled *"Identification of Inherited and De Novo Exomic Variations in An Emirati Family with Neurodevelopmental Disorders"*, hereby, solemnly declare that this thesis is my own original research work that has been done and prepared by me under the supervision of Dr. Ranjit Vijayan in the College of Science at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: _____ Date: 30/12/2020

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Dr. Ranjit Vijayan
   Title: Associate Professor
   Department of Biology
   College of Science
   Signature  Date 22/11/2020

2) Member: Professor Rabah Iratni
   Title: Professor
   Department of Biology
   College of Science
   Signature  Date 22/11/2020

3) Member (External Examiner): Dr. Wael Mohammed Osman
   Title: Assistant Professor
   Department of Chemistry
   Institution: Khalifa University, Abu Dhabi, UAE
   Signature  Date 22/11/2020

This Master Thesis is accepted by:

Dean of the College of science: Professor Maamar Benkraouda

Signature Maamar Benkraouda Date December 31, 2020

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature Ali Hassan Date December 31, 2020

Copy ____ of ____

Abstract

Neurodevelopmental disorders (NDDs) are a heterogenous group of disorders that affect children at any point of development and lead to mental and motor function deficits. Often, the underlying cause could be genetic and inherited. This study investigated possible genetic variations that could have led to these neurological abnormalities and other genetic disorders in an Emirati family. Whole exome sequencing (WES) was used to sequence the protein-coding regions of the genome to identify potential *de novo* and inherited variants that are associated with disorders in this family.

WES of DNA from the parents and ten children were performed. Several variants were identified in high-risk genes associated with autism and epilepsy. However, most of these have previously been classified as benign. Several potentially pathogenic inherited and *de novo* variants were also identified These include a homozygous deletion of *HBA2* gene in some of the family members indicating potential thalassemia, the Protein S (*PROS1)* variant rs146366248 (AF= 0.0007675) associated with protein S deficiency and thrombophilia, Fc fragment of IgG receptor IA (*FCGR1A)* variant rs74315310 (AF= 0.004104) associated with familial deficiency of IGG receptor I and the *de novo* rs132630331 variant in Glycerol Kinase (GK) associated with glycerol kinase deficiency. NDDs have very complex aetiology and could also have been caused by chromosomal abnormalities and copy number variations, which cannot be detected with WES, as well as environmental factors. Hence, further study is required to confirm these findings and to extend it to genomic regions not covered in this study.

Keywords: Neurodevelopmental disorders, autism spectrum disorder, epilepsy, autoimmune disorders, single nucleotide polymorphisms, whole exome sequencing.

Title and Abstract (in Arabic)

تحديد الاختلافات الجينية الموروثة والمستحدثة في عائلة إماراتية لديها اضطرابات النمو العصبي

الملخص

اضطرابات النمو العصبية هي مجموعة من الاضطرابات المتنوعة التي تؤثر على الأطفال في أي نقطة من النمو وتؤدي إلى عجز الوظائف العقلية والحركية. عادة ما تكون الأسباب جينية وموروثه. تتحرّى الدراسة الاختلافات الجينية المحتملة التي ربما قد أدت إلى اضطرابات عصبية وغيرها من الاعتلالات المشتركة في عائلة إماراتية. تم استخدام تقنية التسلسل الكامل للإكسوم لتسلسل مناطق ترميز البروتين من الجينوم وذلك لتحديد الاختلافات الجينية المحتملة المستحدثة والموروثة التي قد ترتبط باضطرابات النمو العصبي في هذه العائلة.

تم إجراء التسلسل الكامل للإكسوم للحمض النووي للوالدين وعشرة من الأبناء. معظم الاختلافات التي حددتها الدراسة كان مرتبطًا بال الجينات الموروثة عالية المخاطر لكل من أطياف التوحد والصرع. وتمكنت الدراسة من تحديد عدد قليل من الطفرات الموروثة والمستحدثة المسببة للأمراض ومنها اكتشاف حذف الجين HBA2 في بعض أفراد الأسرة وهذا قد يحتمل وجود التلاسيميا، وأيضا الاختلاف في بروتين S (PROS1) rs146366248 المرتبط بلجلومبفيليا الناتجة عن نقص البروتين S والمتغير في الجين (FCGR1A) rs74315310 المرتبطة بنقص مستقبلات IGG I. وتم تعريف الاختلاف المستحث في أحد أفراد العائلة في جين GK rs132630331 المرتبطة بنقص كيناز الجليسيرول. الأسباب التي ينتج عنها اضطرابات النمو العصبي معقدة جداً ويمكن أيضاً أن يكون سببها شذوذ الكروموسومات واختلافات رقم النسخ غير المرتبطة بمنطقة الترميز. أو قد تكون عوامل بيئية ومن ثم، يلزم إجراء مزيد من الدراسة لتأكيد هذه النتيجة وتوسيع نطاقها لتشمل المناطق التي لم تشملها هذه الدراسة.

مفاهيم البحث الرئيسية: اضطرابات النمو العصبي، التوحد، الصرع، التسلسل الكامل للأكسوم، الأمراض المناعية، المعلوماتية الحيوية.

Acknowledgements

I would like to express my utmost feeling of gratitude to my supervisor Dr. Ranjit Vijayan for his support, patient guidance and understanding during this journey. This study would not have taken place without his endless effort in humbling the obstacles.

I would like to thank my committee for their guidance, support, and assistance during the preparation of this thesis. I would also like to thank the Chairman and all Professors of the Department of Biology at the United Arab Emirates University for assisting and inspiring me throughout my studies and research.

I would like to extend my deepest gratitude to my colleagues; Ms. Amna Al-Dhmanie who supported and cheered me from the first moment to the end. Special thanks to Mr. Amanat Ali for his dedicated help and guidance through the lab work. I would like to extend my thanks to Ms. Bincy Baby for her encouragement and support.

At the end all thanks to my family who were always here for me. Thank you for the overwhelming encouragement and support through all the study years.

Dedication

To my family and to those who see the world from a different perspective.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACMG	The American College of Medical Genetics and Genomics
ADHD	Attention Deficit Hyperactivity Disorder
AD	Autoimmune Disorder
AF	Allele Frequency
ASD	Autism Spectrum Disorder
CNV	Copy Number Variant
DM	Diabetes Mellitus
DSM-5	Diagnostic and Statistical Manual of Mental Disorders-Fifth edition
EE	Epileptic Encephalopathy
FXS	Fragile X Syndrome
ID	Intellectual Disability
MCCD2	3-methylcrotonoyl-CoA carboxylase 2 deficiency
NCBI	National Center for Biotechnology Information
NDD	Neurodevelopmental Disorders
NGS	Next Generation Sequencing
OMIM	Online Mendelian Inheritance in Man
SIFT	Sort Intolerant From Tolerant
SNP	Single Nucleotide Polymorphism
WES	Whole Exome Sequencing

Chapter 1: Introduction

1.1 Neurodevelopmental disorders

Neurodevelopmental disorders (NDDs) are a group of complex heterogenous disorders that results from pathologies in the nervous system. It could impact children at any point of development. NDDs manifest as cognitive impairment or delay and motor abnormalities. The effect on brain function is variable and often affects memory, intellectual abilities and social behaviour. The symptoms of NDDs vary from mild to profound. NDDs can be spotted in early childhood by monitoring childhood milestones if the pathology is severe enough to be detectable. NDDs encompass a wide range of disorders including intellectual disabilities (IDs), autism spectrum disorder (ASD), attention deficit hyperactive disorder (ADHD), motor coordination disorders, and Tic disorders such as Tourette's disorder[1]. Often NDDs have common neuro-abnormalities and motor deficits. They are recognized as a spectrum of disorders rather than specifically distinguished disorders, suggesting a shared underlying genetic background for these disorders[2,3]. The prevalence of NDDs is increasing, along with the morbidity as well as the economic and social burden associated with them. Six in every one thousand children are thought to have NDDs[1]. The aetiology of NDDs can be environmental, genetic or both. Prenatal exposure to toxins, radiations or some medications during pregnancy or postnatal causes as hypoxia, brain trauma or infections during labour are environmental risk factors linked to NDDs[4]. The identification of the role of the *FMR1* gene in Fragile X syndrome (FXS) established the way to investigate the genetic causes of NDDs[5]. Genetic and chromosomal abnormalities and the cellular and biological mechanisms affected by those variations

are still major causes of NDDs. Genetic variations of NDDs are diverse and hence complex. The underlying cause of most NDDs are polygenic rather than incited by a single gene. The onset of NDDs symptoms usually happen by the contribution of multiple genetic variations rather than monogenic pathology[2]. Mutation in a single gene could have a broad spectrum of clinical abnormalities, e.g., *SCN2A*, gene encodes a brain sodium channel that is the most mutated channel in epilepsy. It was identified as a source of *de novo* pathogenic mutations in autism, epileptic encephalopathy (EE), intellectual dysfunction (ID) and schizophrenia[6]. Some genetic syndromes such as myotonic dystrophy, Angelman and Prader-Willi syndromes, neurofibromatosis, fragile X syndrome, tuberous sclerosis and Joubert syndrome have autism as a recognized feature[20]. This emphasizes the complicated and shared genetic architecture of NDDs.

1.2 Autism spectrum disorders

Autism was first defined by Leo Kanner in 1943 as a specific syndrome noticed in early childhood and manifested by abnormal social and emotional relationships[7]. In 2013, a new definition and classification was introduced in the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5). DSM-5 specifies autism spectrum disorder (ASD) as diverse neurological and behavioural abnormalities with two main criteria: intellectual and social functions deficit that significantly affect verbal and nonverbal communication along with repetitive stereotypic behaviour. ASD manifests at an early age and can be observed sometimes at the age of three years or younger. Each child under the spectrum would show unique symptoms according to functions that are affected[1]. The ASD phenotypic scale vary from mild to severe neurobehavioral abnormalities. Although there is no cure or treatment for the core

symptoms of ASD, early diagnosis along with applying assistive behavioural therapies would improve the prognosis. Only the antipsychotics risperidone and aripiprazole drugs are licensed and approved by the US Food and Drug Administration to alleviate ASD associated hyperactivity, inattention, irritability, aggression, self-injurious behaviour and stereotypies[7]. The first step to find a treatment to ASD is to have more understanding to the causes and factors contributing to ASD.

1.2.1 Autism spectrum disorder etiology

The prevalence of ASD is increasing. Worldwide, one in every 160 children is believed to have ASD. ASD has a higher incidence in males than in females. In spite of this, the potential causes of ASD are still not fully understood[8]. The leading causes of ASD could be genetic, environmental or a combination of both. Advanced parental age, very low birth weight, or exposure to medication during pregnancy such as valproic acid (treatment of epilepsy) brain infections and trauma are thought to be factors that contributes to ASD[9]. ASD prevalence has also been reported to be higher in preterm infants[10]. Children with other NDDs or children whose siblings with ASD have a higher chance of displaying ASD[11]. The recurrence risk of autism in siblings is approximately 45 times higher than in the general population. Twin studies have also pointed to a higher concurrence rate in monozygotic (60%–91%) than in dizygotic twins (0%–6%)[12].

1.2.2 ASD associated genetic variations

The genetic landscape of ASD is complex and not fully understood. Genetic abnormalities and variations count for 90% of the ASD aetiology[13]. Chromosomal structural abnormalities including single nucleotide polymorphisms (SNPs),

insertions, deletions, translocations, and copy number variations (CNVs) could also often play a role. Over 600 genes have been associated with ASD. The genes affected are diverse and generally involved in protein coding, chromosomal splicing, neuronal signalling, and metabolism. The interplay between those gene functions which lead to ASD neurobehavioral abnormalities remains unclear[14]. Sometimes, ASD is a possible manifestation of single-gene variations, such as mutations in *FMR1, TSC1, TSC2, MECP2,* and *PTEN*[15]. Mutations in synaptic genes such as *NLGN3, NLGN4X* and *SHANK3,* and rare deletion CNVs of *SHANK3* are also associated with ASD[16,17]. A large study based on thousands of ASD subjects shortlisted the following 33 genes with very significant association with ASD (*ADNP, ANK2, ARID1B, CHD8, CUL3, DYRK1A, GRIN2B, KATNAL2, POGZ, SCN2A, SUV420H1, SYNGAP1, TBR1, ASXL3, BCL11A, CACNA2D3, MLL3, ASH1L, CTTNBP2, GABRB3, PTEN, RELN, APH1A, CD42BPB, ETFB, NAA15, MYO9B, MYT1L, NR3C2, SETD5, TRIO, MIB1, VIL1*). The same study was able to identify 107 high risk genes that showed unusual evolutionary constraint against mutations, to have *de novo* loss-of-function mutations in over 5% of autistic subjects[18]. The role of those genes involved encoding proteins for synaptic formation (voltage-gated ion channels regulating the propagation of action potentials pathways), transcriptional regulation and chromatin-remodelling, notably the pathways that mediate post-translational lysine methylation/demethylation and modifications of histones[18]. Studies have also investigated the role of epigenetic regulatory mechanisms in ASD. Over 600 confirmed and putative human epigenes have been identified many of which are associated with ID and ASD[19,20]. This stresses the multifactorial aetiology of ASD and that ASD could arise as an interaction between genetic and environmental confounding factors mediated by epigenetic mechanisms[21].

1.2.3 ASD and comorbidity

Autistic behaviours can often associate with ASD comorbidities such as epilepsy and autoimmune abnormalities. The incidence of these are four times higher in NDD affected children than in general population[1]. Metabolic disorders such as mitochondrial dysfunctions are also seen in 10 to 20% of patients with ASD[22].

1.2.4 ASD and autoimmune abnormalities

Studies have associated immune abnormalities and pathologies to ASD[23]. Several epidemiological studies highlighted the higher prevalence of ASD in children with a familial history of certain autoimmune disorders as eczema, psoriasis, type I diabetes mellitus and rheumatoid arthritis (RA)[24,25]. Other studies have investigated the role of human leukocyte antigen (HLA) gene (gene that plays important role in ADs) variations in ASD children. Autistic children had signiifcantly higher frequency. of HLA-DRB1*11 and HLA-B*07 alleles than controls[26]. *NLRP2* and *MOGS* are two genes with prominent roles in immunity that have been linked to ASD. The study identified rare variants associated with the genes in 90% of ASD study subjects[27]. Overall, the altered immune function in ASD children is suggestive of a contribution or involvement of the autoimmune system in the pathology of ASD.

1.3 Epilepsy

Epilepsy is a neurological disorder that manifests as recurrent, unprovoked seizures. It affects approximately 50 million people worldwide. The underlying cause of epilepsy varies. Epilepsy could be a result of brain damage due to trauma or infection, or it could also be other metabolic, immune and genetic causes[28]. More than 900 genes have been linked to epilepsy[29]. Epilepsy phenotypes can result from single

gene or multigene variations[30]. Several genes linked to epilepsy are associated with neuronal transmissions and the variations lead to a set of channelopathies e.g. acetylcholine receptor α4 subunit (*CHRNA4*), potassium channels like (*KCNQ2, KCNQ3*) and the voltage-gated sodium channel β1 subunit gene (*SCN1B*)[31–33]. Epilepsy and ASD coexist in up to 20% of children with either disorder[34]. Several studies have suggested an overlap between epilepsy and ASD associated genes[35,36].

1.4 Whole exome sequencing

Advanced massively parallel and rapid DNA sequencing techniques like next generation sequencing (NGS) have enabled scientists to gain insights into genetic variations and inherited disorders. Whole exome sequencing (WES) is one such technique that can be used to identify genetic variations within exomes (the coding sequence in DNA) and to identify if these variations are inherited or *de novo* by downstream analysis. Applying clinical exome sequencing approaches and informatics have identified over 5000 genes with a role in neuronal development and function in affected individuals[22,37]. Large cohort studies which utilized WES to investigate NDDs have achieved a high success in identifying ASD associated *de novo* mutations[18,38]. The basic strategies in using WES in such studies is either to identify pathogenic *de novo* variants in small sized sample using trios-based analysis, or to find the rare inherited variants in a large sample[37].

However, sequencing techniques identify a huge number of variations which are of uncertain clinical significance, most probably because the genomic datasets are lacking, limitations of bioinformatics and computational predictions, and relevance in relationship to the normal population and a lack of biological correlation[6].

Understanding the genetic causes underlying NDDs is the first step in gaining a deeper insight into the biological and cellular pathways of such abnormalities. This could be of significance if applied to develop candidate therapies which can improve the quality of life, cure the affected population or perhaps used in genetic screening programs.

1.5 Hypothesis

The ten children studied here have various disorders that could have genetic origins. Some of these conditions include NDDs, epilepsy, and metabolic disorders. NDDs have multiple genetic aetiology. NDDs could be caused by *de novo* and inherited single nucleotide variations, other chromosomal abnormalities as well as environmental factors. Several lines of evidence suggest that genetic variations in the exome could be associated with NDDs. Hence, the hypothesis of this study was that NDDs and the other disorders in the chosen Emirati family could have been caused by either *de novo* or inherited genetic variations in the exome. This was investigated in this study using whole exome sequencing.

1.6 Objective of the study

The objective of this study was to sequence the whole exome of an Emirati family consisting of the parents and ten children and to use trio-based analysis to identify potential inherited and *de novo* variations in the exome that could explain the range of possibly related genetic disorders in this family including NDDs, epilepsy, metabolic disorders and diabetes.

Chapter 2: Methods

2.1 Ethical approval

Ethical approval was obtained from the Centre of Research and Statistics, Ministry of Health and Prevention (MOHAP) with approval no: MOHAP /DXB-REC-50/2018 granted on 19 March 2019. Blood samples were collected at Fujairah Hospital from the participating members of the family (parents and ten children) after obtaining informed consent and assent of minors. Subsequent laboratory work and data analysis were done at United Arab Emirates University (UAEU).

2.2 Participants

The participating family has multiplex ASD children. The healthy parents, designated as S1 and S2, were subsequently used for trio-based analysis to identify *de novo* and inherited variations. The extended family has a history of epilepsy, diabetes, leukaemia and heart abnormalities. The samples of the ten children were designated with identifiers S3-12. Of the ten children, three were diagnosed with autistic behaviours (S3, S5 and S7). The remaining seven children S4, S6, S8, S9, S10, and S11 were considered as either controls or to evaluate potentially related disorders such as epilepsy (samples S1, S3, S9 and S11), metabolic abnormalities (samples S7, S8, S10, and S11) and autoimmune disorders (samples S4, S9, and S12). It is worth noting that all the study probands were preterm. Figure 1 provides an illustrative pedigree of the family with NDDs and other conditions. In this family, variations in reported genes associated with NDDs, epilepsy, ASD and metabolic abnormalities that may associate with NDDs were investigated along with other pathogenic variants associated with unrelated disorders.

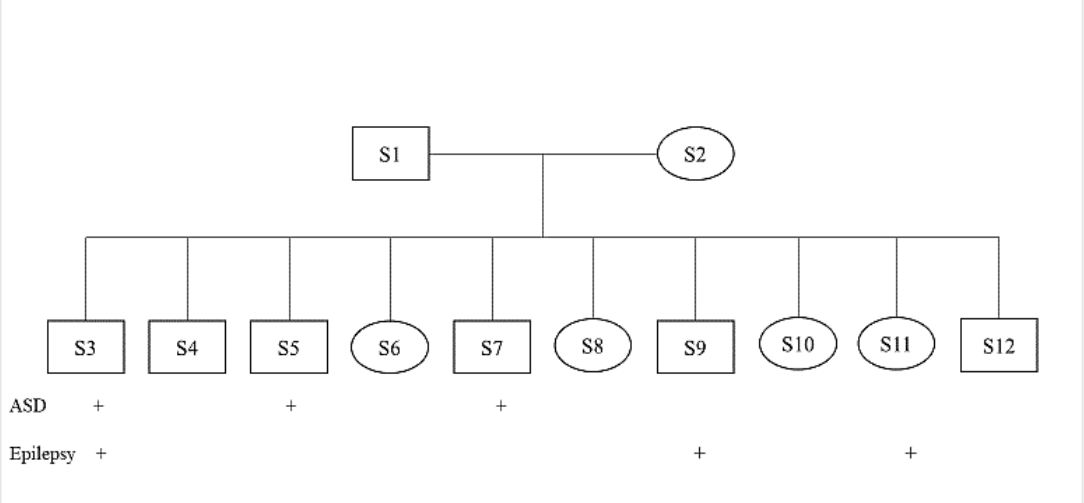


Figure 1: Pedigree of the family studied indicating probands with ASD and epilepsy

2.3 Sample collection

5 ml of whole blood was collected in ethylene diamine tetra-acetic acid (EDTA) tubes from the participants and labelled by a registered phlebotomist at Fujairah Hospital. Samples were transferred to UAEU in a blood transport box and stored at -20℃ in UAEU's Department of Biology.

2.4 DNA isolation and purification

DNA isolation and purification were performed to obtain a minimum concentration of 50 ng/µl DNA. QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) was used for this purpose.

2.4.1 DNA extraction

DNA extraction was done using QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) based on manufacturer's recommended protocol. Briefly, Protease K was resuspended, buffers AW1 &AW2 were dissolved in 30 ml 100% ethanol as labelled. 20 µl of QIAGEN Protease (proteinase K) was pipetted into twelve 1.5 ml

microcentrifuge labelled tubes. 200 µl of each of the twelve samples were added to the microcentrifuge tubes with proteinase K followed by the addition of 200 µl Buffer AL to each sample. Samples were pulse-vortexed for 15 s. and then incubated at 56°C for 10 min. Following incubation, a volume of 200 µl ethanol (96%) was added to each sample, and mixed using pulse-vortexing for 15 s, followed by brief centrifugation. The mixed samples were transferred to the QIAamp Mini spin column in a 2 ml collection tubes, capped and centrifuged at 8000 rpm for 1 min. The columns were moved to a clean 2 ml collection tubes and collection tubes with the filtrates were discarded.

2.4.2 DNA purification

DNA purification was performed using QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) based on manufacturer's recommended protocol. This involved two washing steps. For washing step one, 500 µl buffer AW1 was added to each column and the columns were centrifuged at 8000 rpm for 1 min. The columns were then placed in new 2 ml collection tubes, and tubes with filtrates were discarded.

In washing step two, 500 µl buffer AW2 was added to each column and the mix was centrifuged at 14,000 rpm for 3 min. The step was repeated for 1 minute and the tubes with the filtrates were discarded.

For elution, the QIAamp Mini spin columns were transferred to clean 1.5 ml microcentrifuge tubes. A volume of 50 µl buffer AE was added to each column as divided portions; 30 µl AE buffer were added first and pulse-vortexed then the remaining 30 µl was added and incubated at room temperature for 5 min. The samples were centrifuged at 8000 rpm for 1 min. Using the microvolume spectrophotometer

Nanodrop 2000/c (Thermo Fisher Scientific, USA), 1µl of each purified DNA isolate was loaded and the quality and quantity were assessed.

2.5 Whole exome sequencing

Twelve DNA isolates were sent to Macrogen Korea for WES using Illumina HiSeq platform. The Illumina NGS workflow includes 4 basic steps shown in the below Figure 2. The Agilent SureSelect v6 kit was used for exome capture.

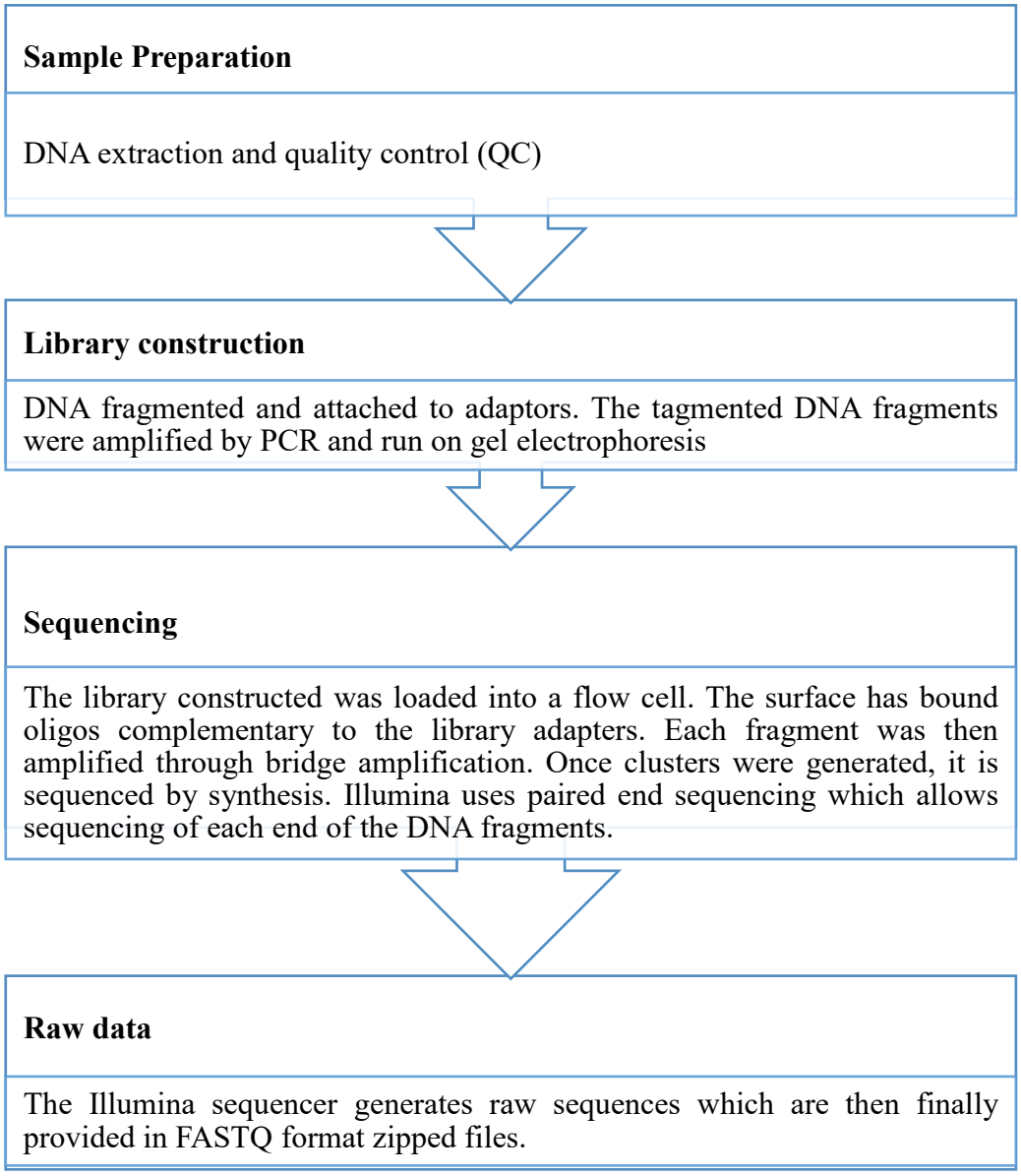


Figure 2: The Illumina NGS workflow

2.6 Data analysis pipeline

Raw WES data in FASTQ format obtained from Macrogen was analysed at the Computational Biology & Bioinformatics Lab in the Department of Biology, UAEU. The *SeqMule* pipeline was used to analysed the sequenced data[39]. The analysis employed the series of steps shown in Figure 3.

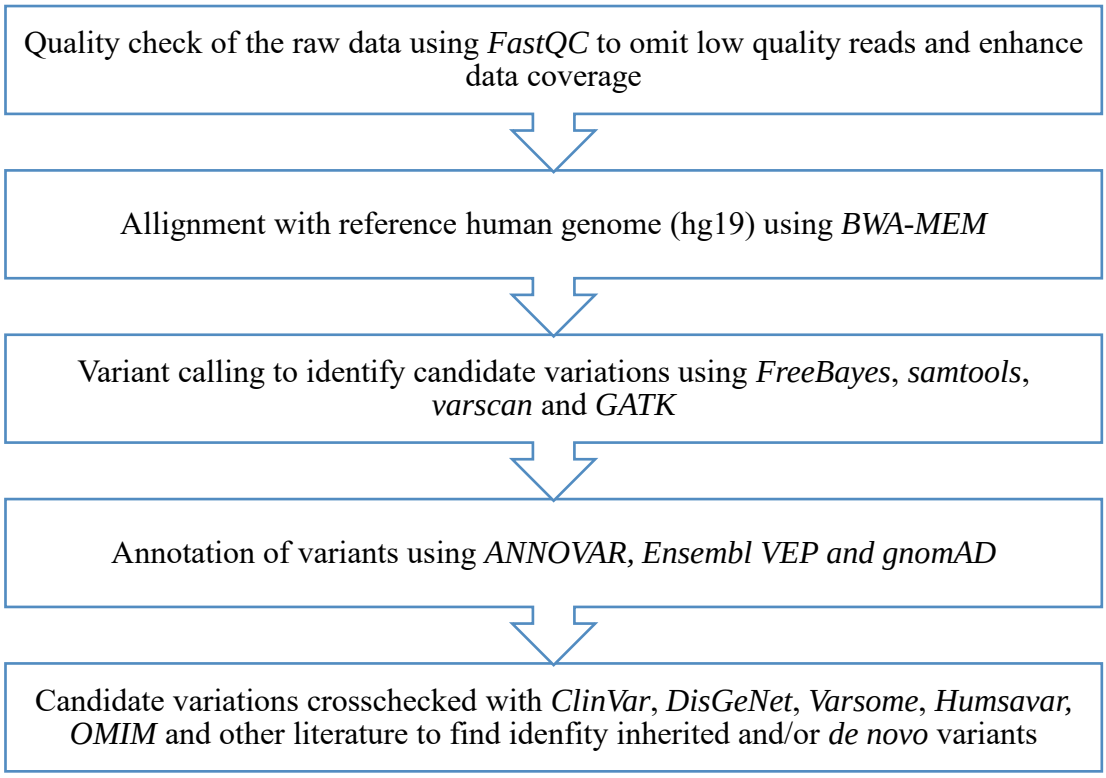


Figure 3: Bioinformatics data analysis pipeline

SeqMule 1.2.6 is an automated pipeline that allows the integration of multiple tools and algorithms for the analysis of sequencing data. *FastQC* 0.11.8 was used for quality check; the low-quality reads were filtered or trimmed. Quality check passed reads were aligned to the human reference hg19 sequence using *bwa-mem (BWA* 0.7.17). Deduplication of reads was done by *samtools* using *picard*. Variants were called using multiple callers: *FreeBayes* 1.0.2, *samtools* 0.1.19, *GATK* 2.8-1, and *varscan* 2.4.3. VCF files with the variants were annotated using *ANNOVAR*, Ensembl

Variant Effect Predictor (VEP) and exome allele frequency was obtained from gnomAD. The aligned reads in BAM format were visualized in IGV 2.8.10.

Chapter 3: Results

3.1 Quantity and quality of the DNA isolates

The quantity and quality of the isolated DNA from the twelve samples were evaluated using a Nanodrop 2000/c since the minimum quantity/quality required by Macrogen for sequencing was a concentration of 50 ng/µl and a purity of A260/280 >1.7. As shown in Table 1, all samples met these requirements and were sent to Macrogen (Korea) for WES.

Table 1: Quantity and quality of the DNA isolates

Sample	**DNA concentration (ng/µl)**	**260/280**
S1	146	1.86
S2	83.5	1.8
S3	87	1.82
S4	70	1.85
S5	50.4	1.86
S6	113.6	1.8
S7	99	1.87
S8	55.4	1.93
S9	130	1.91
S10	60	1.9
S11	57.8	1.9
S12	51.5	1.8

3.2 Statistics of sequenced and assembled data

The sequenced and assembled reads were assessed based on statistics provide by *SeqMule* and a summary of this is provided in Table 2. An average of 57627696 reads passed QC of which over 95% were mapped to the human reference genome (hg19) with an average coverage of 83X and quality score of 36 (Figure 4). The GC content of all samples was 51%, which in in the expected range for exome sequencing.

Table 2: Statistics of sequenced and assembled data per each sample

Sample ID	QC passed reads	Mapped reads	% reads mapped to targeted region	Average coverage	GC %
S1	79772073	75948415 (95.21%)	62.61%	116.93	51
S2	57982674	55246560 (95.28%)	61.13%	83.11	51
S3	74813209	71237450 (95.22%)	62.68%	109.71	51
S4	74250876	70724526 (95.25%)	62.91%	109.26	51
S5	50963733	48569215 (95.30%)	61.95%	74.15	51
S6	51922709	49513727 (95.36%)	62.27%	76.17	51
S7	40541404	38648743 (95.33%)	61.47%	58.65	51
S8	51469745	49006370 (95.21%)	62.11%	75.03	51
S9	52275935	49816033 (95.29%)	62.36%	76.69	51
S10	49865353	47483913 (95.22%)	62.44%	73.07	51
S11	50623239	48193489 (95.20%)	60.72%	72.11	51
S12	57406380	54630172 (95.16%)	61.67%	82.86	51
Average	57627696	54918217.75 (95%)	62.03%	83.97	51

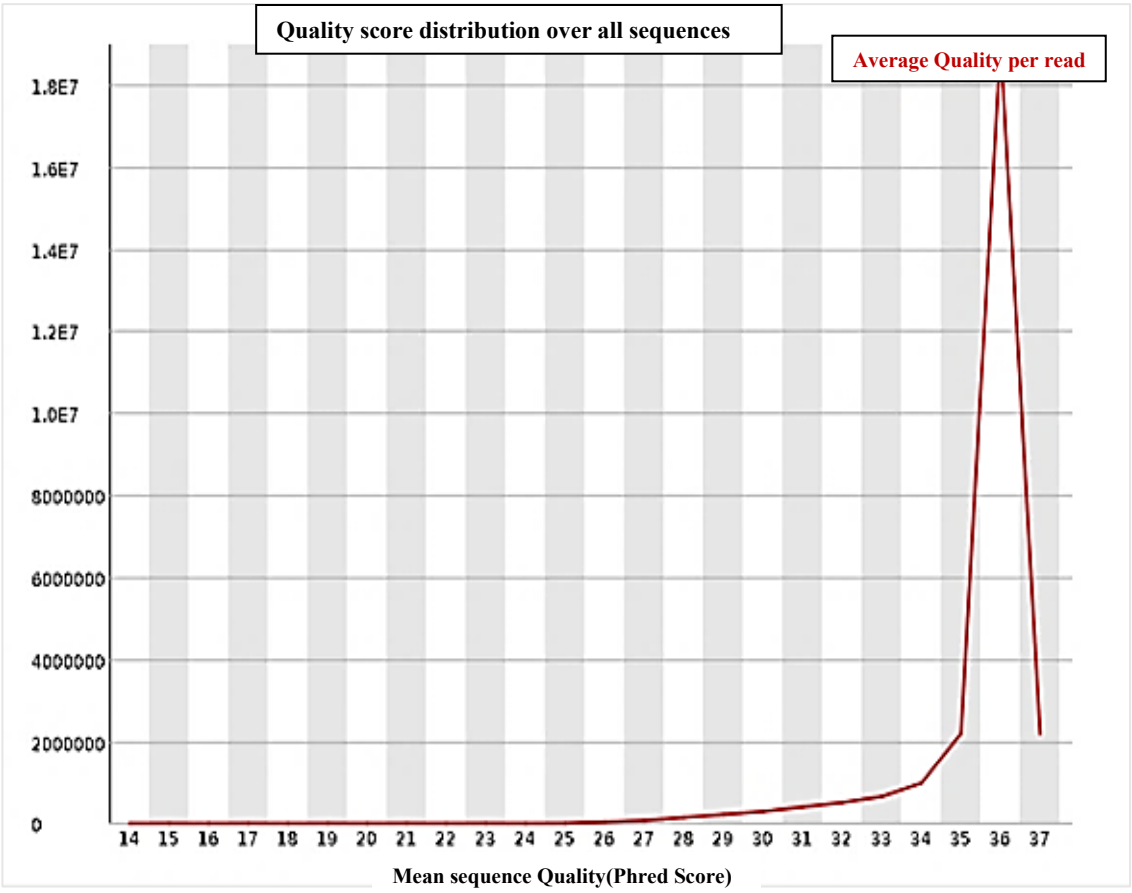


Figure 4: Average quality score

3.3 Summary of detected variants

Variant calling using FreeBayes, based on WES data aligned to the human reference hg19 reference genome, identified an average of 56928 variants per sample (Figure 5). Since WES also normally covers short non-exonic areas that flank exons, variants were also identified in this region. In this study, an average of 35166 intronic and 21761 exonic variations were detected per sample. Exonic variations include synonymous, nonsynonymous, and missense variations as well as frameshift and non-frameshift indels and SNPs. About half of the WES detected variations were nonsynonymous and less than 1% were frameshift and non-frameshift (Figure 6). The parents shared around half of all detected exonic variants. 50% of the exonic variants were synonymous. Around one third of the shared nonsynonymous variants were homozygous in both parents (Figure 7). 57% of the exonic nonsynonymous variants the probands were inherited from both parents and homozygosity was slightly higher in the inherited variants (Table 3).

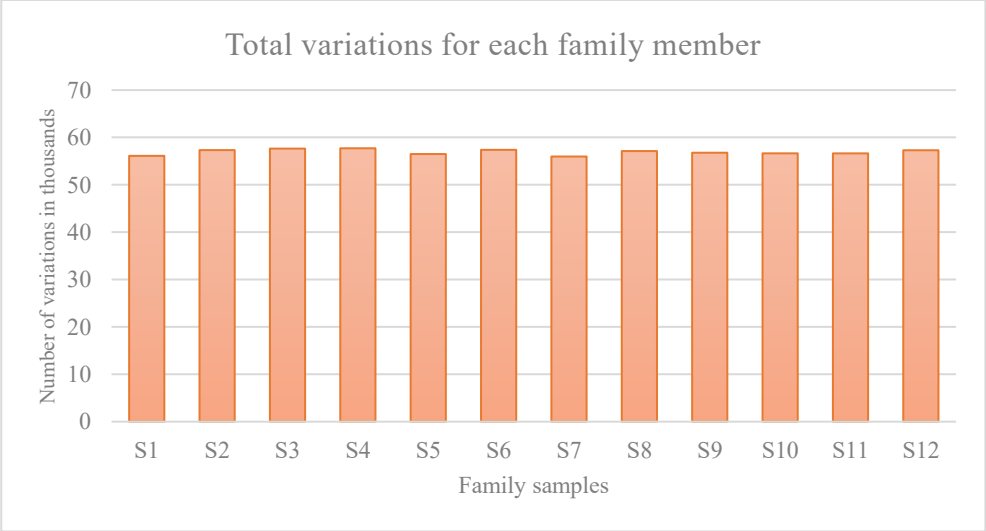


Figure 5: Total number of variations detected by WES in each sample

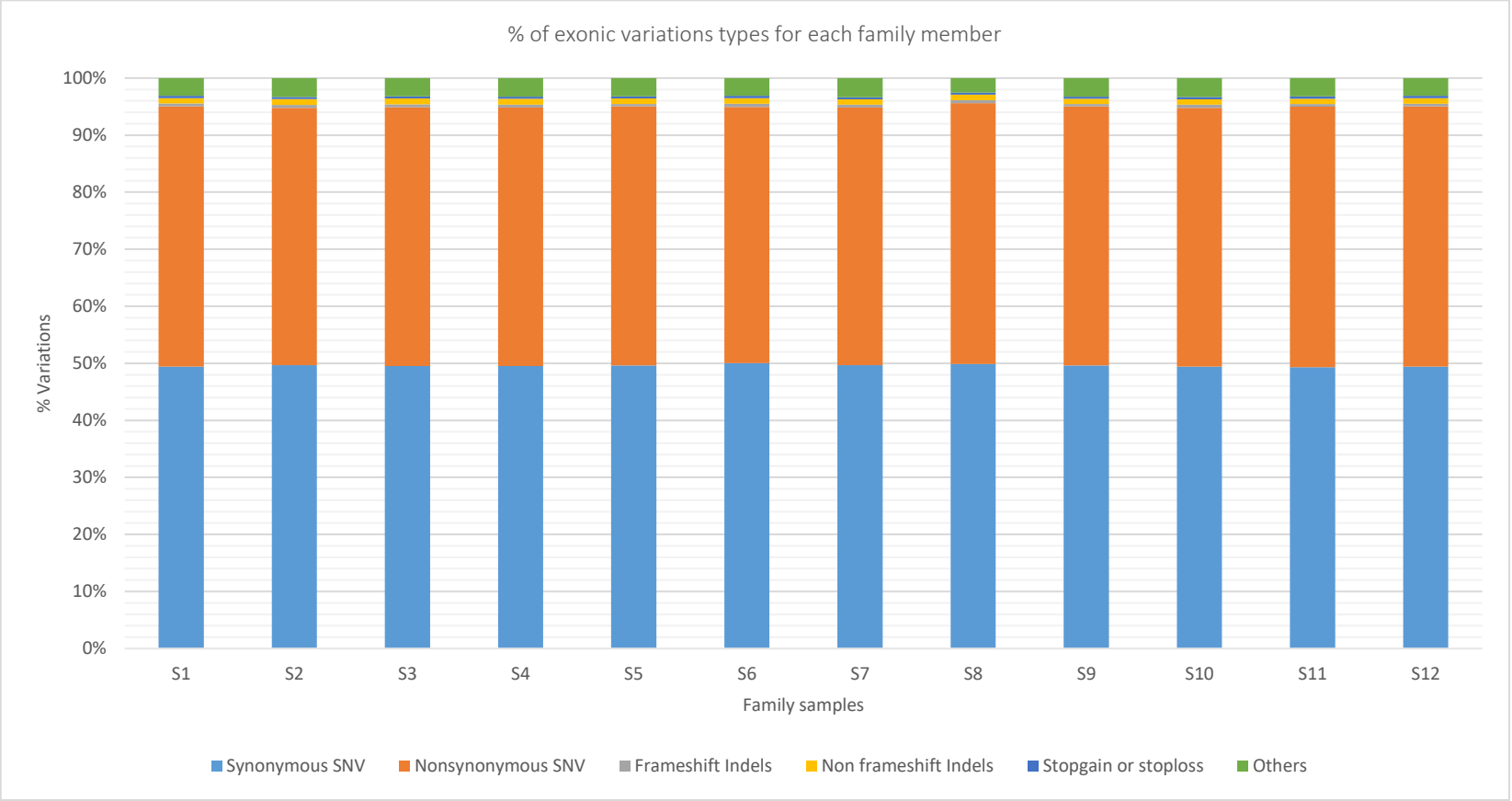


Figure 6: Variation type per sample

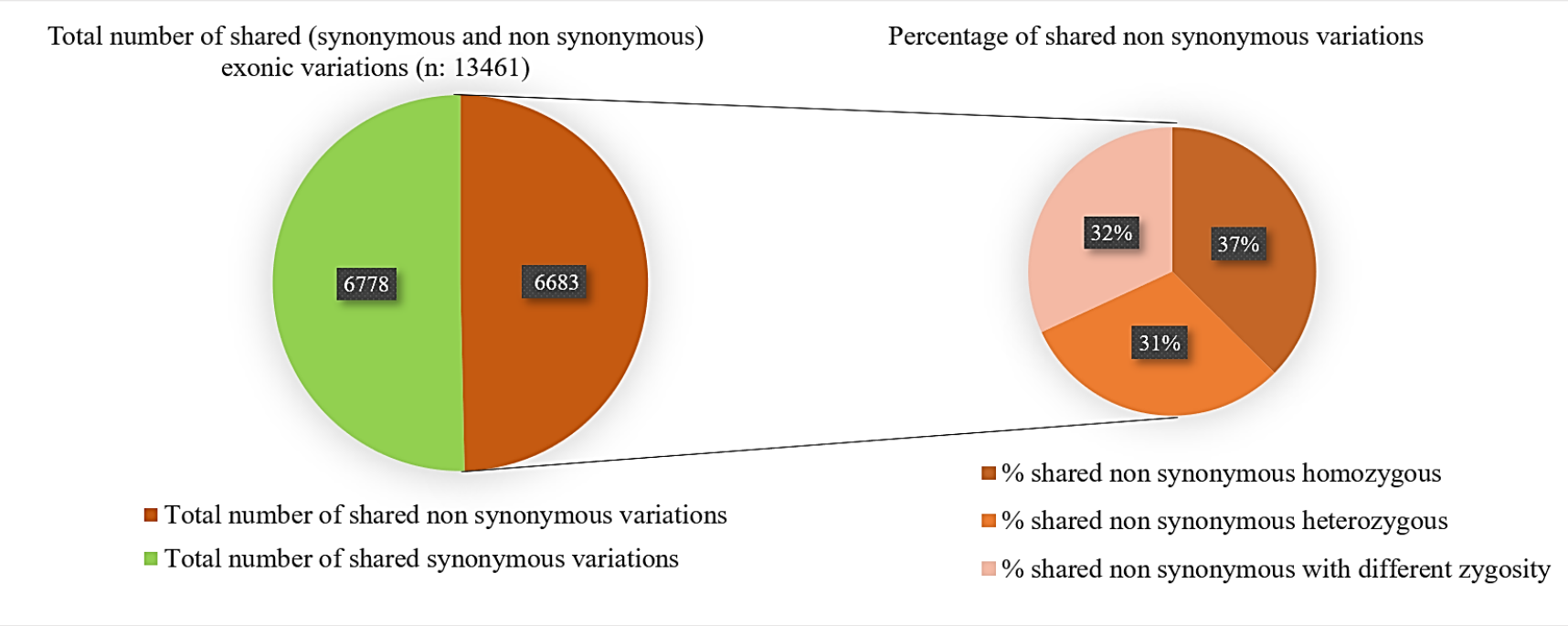


Figure 7: Total exonic variants (synonymous and nonsynonymous) along with percentage of shared homo or heterozygosity of the nonsynonymous variants in the parents (S1 and S2)

Table 3: Nonsynonymous exonic variations inherited and *de novo* along with zygosity in the probands. HE: heterozygous, HO: homozygous

Variation source	**Inherited and in both parents**		***De novo* reported and not reported in dbSNP**	
Sample ID / Zygosity	HE	HO	HE	HO
S3	2222	3972	162	26
S4	2199	4002	122	14
S5	2172	3971	101	15
S6	2355	3376	1472	20
S7	2180	3948	141	16
S8	2254	3807	141	22
S9	2184	3922	127	18
S10	2215	3958	142	7
S11	2256	3867	128	7
S12	2192	3978	145	15

3.4 Variations associated with ASD or NDDs in the family

Analysis of the variants showed that almost all of the genetic variations identified in ASD related genes are benign and most of them exist in both ASD and non-ASD probands (Table 4). WES data results indicated the presence of a high number of benign ASD associatied missense variants in the ASD-linked gene *ANKRD11,* which is involved in transcription regulation[40]. Nearly all of these missense variants in *ANKRD11* exists in at least one of the two parents except for the *de novo* variants rs767958 (AF= 0.03310) and rs60520302 (AF= 0.00001961)which exists only in S6, a child with no clinical ASD phenotype. The benign variation rs11669628 (AF= 0.1332) was also identified in *CC2D1A* (serotonin receptor transcription regulator gene), a gene associated with NDDs. This variation was present in heterozygous mode in four of the probands S3, S5, S9, and S11 affected with ASD, epilepsy, or behavioural abnormality and one of the parents is also a carrier of this variant. In the *CHD* gene family, only the benign rs10467770 (AF= 0.2458) variant in *CHD8* was present in S7, ASD affected child. Multiple gene variants that correlated

to susceptibility to schizophrenia were noticed in the family members including the rs6280 (AF= 0.6275) in *DRD3* which is reported to have a high-risk for schizophrenia. This variant was heterozygous in most of the probands except for S7, ASD reported child who was homozygous. The rs1051061 variant (AF= 0.3588), which is also related to schizophrenia, is homozygous in two probands (S3 reported ASD traits and S11 reported behavioural problems). Some additional variants associated with other NDDs, which may have autistic behavioral traits, were also identified. The rs86312 variant (AF= 0.01837) in *NAGLU* and rs7503034 (AF= 0.3686) in *SGSH* are associated with Sanfilippo syndrome and mucopolysaccharidosis, both of which have been reported to have ASD traits[41,42]. Benign variants rs611326 and rs17522826 in *TCF4* gene, associated with Pitt-Hoppkins syndrome, was present in most of the probands. Two variants, one high risk and the other pathogenic, with association to Bardet-Biedl syndrome 2 were identified. Bardet-Biedl syndrome 2 is an autosomal recessive disorder characterized by clinical features such as pigmentary retinal dystrophy, obesity, developmental deficits, and renal disorders and requires three mutant alleles to manifest[43,44]. The two variants were detected with similar zygosity in all ASD affected male probands and one unaffected female proband. The heterozygous *CCDC28B* rs41263993 (AF= 0.01158) variation was identified in S3, S5, S7 and S8 and the high risk homozygous *BBS2* rs4784677 was identified in all probands.

Table 4: Variations in genes associated with ASD and/or other abnormalities along with inheritance pattern in the family members. M: Male, F: Female, Y: yes, F*: female with reported behavioral abnormalities, AF: Exome allele frequency from gnomAD.

Designated sample Id		**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	-	-
Gender		M	F	M	M	M	F	M	F	M	F	F*	M		
ASD				Y		Y		Y							
Epilepsy		Y		Y						Y		Y			
Gene	**rsID**													**AF**	**ClinVar significance**
ANKRD11	rs2279348	HO	HOs	HO	HO	HO	HE	HO	HE	HO	HO	HO	HO	0.6354	Benign, Association with ASD
ANKRD11	rs145906515	HE	.	.	.	HE	.	HE	.	HE	.	.	HE	0.004140	Benign, Association with ASD
ANKRD11	rs143743958	.	HE	HE	HE	HE	.	.	HE	.	HE	.	HE	0.0005340	Benign, Association with ASD & encephalopathy, progressive, early-onset, with episodic rhabdomyolysis
ANKRD11	rs76793093	.	.	.	.	.	HE	.	.	.	.	.	.	0.03310	Benign, Association with ASD
ANKRD11	rs60520302	.	.	.	.	.	HE	.	.	.	.	.	.	0.00001961	Benign, Association with ASD
CC2D1A	rs11669628	HE	.	HE	.	.	.	HE	.	HE	.	HE	HE	0.1332	Benign, History of NDDs
CHD8	rs10467770	.	HE	.	.	.	.	HE	.	.	.	.	.	0.2458	Benign, History of NDDs
DAOA	rs2391191	.	.	.	.	.	HE	.	.	.	.	.	.	0.3951	Related to early onset Schizophrenia
DRD3	rs6280	HE	HE	HE	.	HE	HE	HO	HE	HE	.	HE	.	0.6275	Risk factor & susceptibility to. Schizophrenia and tremor. (controversial)
VRK2	rs1051061	HE	HE	HO		HE	HE		HE	HE	HE	HO	HE	0.3588	Associated schizophrenia
GABRG2	rs211035	HO	.	.	HO	.	.	.	.	.	.	.	.	0.8233	Interactions of the GABRG2 polymorphisms and childhood trauma on suicide attempt. Also related traits in depressed patients.

Table 4: Variations in genes associated with ASD and/or other abnormalities along with inheritance pattern in the family members. M: male, F: female, Y: yes, F*: female with reported behavioural abnormalities, AF: Exome allele frequency from gnomAD. (Continued)

Designated sample Id		**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	-	-
Gender		M	F	M	M	M	F	M	F	M	F	F*	M		
ASD				Y		Y		Y							
Epilepsy		Y		Y						Y		Y			
Gene	**rsID**													**AF**	**ClinVar significance**
GLO1	rs4746	.	.	.	.	.	HE	.	.	.	.	.	.	0.3588	Uncertain significance to autism
NRXN1	rs9636391	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	0.8518	Benign, Pitt-Hopkins-like syndrome 4
TCF4	rs611326	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	.	HO	0.9996	Benign or likely benign, Pitt Hopkins syndrome
TCF4	rs17522826	HO	HE	HO	HE	HO	.	HE	HO	HE	HO	HO	HE	0.2130	Benign or likely benign, Pitt Hopkins syndrome.
PIGN	rs3862712	HE	HE	.	HE	HE	HO	.	.	HE	HE	.	HE	0.1106	Benign; History of neurodevelopmental disorder
CCDC28B	rs41263993	HE	.	HE	.	HE	.	HE	HE	.	.	.	.	0.01158	Risk factor, Bardet-Biedl Syndrome
BBS2	rs4784677	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	0.9941	Bardet-Biedl syndrome 2 (disorder)
SGSH	rs7503034	.	HE	.	.	.	HO	.	HE	HE	HE	HE	HE	0.3686	Benign, Sanfilippo syndrome and Mucopolysaccharidosis, MPS-III-A
SHANK3	rs9616915	HE	.	HE	.	HE	HE	HE	HE	.	HE	HE	.	0.4268	Benign History of NDDs

3.5 Variations in genes associated with metabolism

Several variants identified were related to metabolic abnormalities, some of which may be related to developmental delay (Table 5). Variations associated with genes *MTHFR* (methylenetetrahydrofolate reductase), involved in folate metabolism pathway, and *MTR* (5-methyltetrahydrofolate-homocysteine methyltransferase), *MTRR* (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), associated with cobalamin metabolism were observed in all family members. Two *MTHFR* variants rs1801133 and rs1801131 (AF= 0.3149 and 0.2890 respectively) are benignly associated with neural tube deficites and susceptibility to schizopherania. A similar combination of multiple heterozygous nonsynonymous variants in the three genes was observered in two of the ASD reported probands (S3 and S7). The *AGA* gene variant rs74626221 (AF= 0.01437) was reported to have a mild association with aspartylglucosaminuria (AGU). AGU is a autosomal recessive disorder associated with a high level of abnormal metabolites in urine[45]. This, along with *MTHFR* and *MTR* variants may be associated with the urological abnormalities reported in most of the probands in this family especially S5, S7, and S11. The benign rs2476601 was identified in *PTPN22*, the rs1052553 variant in *MAPT* and rs237025 in *SUMO4* (AF= 0.1446, 0.6155 and 0.7718 consecutively) were reported to confer DM risk and these were observed in the probands. Also the *GALT* rs2070074 variant (AF= 0.09) associated with galactosemia was identified in some of the probands. Multiple variants with potential DM risk have been identified in the family members correlating with reported prediabetic trait and glucose intolerance. Among the findings were variants associated with cortisone metabolism abnormalities and consequently electrolye imbalance which were also reported in the probands, especially hyperkalemia.

Corticosterone methyl oxidase 2 deficiency (CMO-2 deficiency), an autosomal recessive condition, associated with the variant rs61757294 (AF= 0.08452) in *CYP11B2* leads to hypoaldosteronism and hyperkalemia[46].

Table 5: Variations in genes associated with metabolic abnormalities in the family members

Genes	**rsID**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	**AF**	**Varsome significance**	**ClinVar significance**
MTHFR	rs1801133	.	HE	.	HE	.	.	.	HE	HE	HE	HE	.	0.3149	Benign	Likely benign, Neural tube defects, folate-sensitive
MTHFR	rs1801131	.	HE	HE	.	HE	HE	HE	.	.	.	.	HE	0.2890	Benign	Likely benign, Neural tube defects, folate-sensitive
MTR	rs1805087	HE	HE	HE	HE	.	.	HE	HE	.	.	.	.	0.2030	Benign	Disorders of Intracellular Cobalamin metabolism
MTR	rs1131449	HO	HE	HE	HE	HE	.	HE	.	HO	HE	HO	.	0.5762	Benign	Disorders of intracellular Cobalamin metabolism
MTRR	rs1532268	HE	HE	HE	.	.	HE	HE	HE	.	.	.	.	0.3108	Benign	Disorders of Intracellular Cobalamin metabolism
COMT	rs4680	HO	HE	HE	HO	HE	HE	HE	HO	HO	HO	HE	HO	0.4611	Benign	drug response; Tramadol response, nicotine response - Efficacy
NAGLU	rs86312	HO	HO	HO	HO	HO	HO	HO	HO	HO.	HO	HO	HO	0.01837	Benign	Benign, Mucopolysaccharidosis, MPS-III-B
AGA	rs74626221		HE	HE			HE	HE	HE		HE		HE	0.01437	Benign	Benign, Aspartylglucosaminuria (AGU)
BCHE	rs1803274	HE	HE	HO	HE	HE		HE	HE		HE	HO		0.1759	Benign	Butyrylcholinesterase deficiency (BCHED)
CYP11B2	rs61757294		HE		HE		HE			HE		HE		0.08452	Benign	Corticosterone methyl oxidase 2 deficiency (CMO-2 deficiency)
H6PD	rs6688832		HE	HE	HE					HE		HE		0.2845	Benign	Cortisone reductase deficiency 1 (CORTRD1)
PTPN22	rs2476601	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	0.9289	Benign	Diabetes mellitus, insulin dependent (IDDM)
MAPT	rs1052553	HE				HE							HE	0.1446	Benign	Diabetes Mellitus, Insulin-Dependent
SUMO4	rs237025		HE				HE			HE	HE		HE	0.6155	Benign	Diabetes Mellitus, Insulin-Dependent
DPYD	rs1801265	HO	HE	HE	HE	HO	HE	HO		HE	HO	HE	HE	0.7718	Benign	Dihydropyrimidine dehydrogenase deficiency (DPYDD)
GALT	rs2070074		HE	HE	HE		HE		HE	HE		HE	HE	0.0991	Benign	Galactosemia (GALCT)
PRODH	rs450046	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	0.9283	Benign	Hyperprolinemia 1 (HYRPRO1)
SLCO1B1	rs4149056	HE					HE			HE	HE	HE	HE	0.1329	Benign	Pathogenic, Gilberts syndrome (Autosomal recessive)

3.6 Variations associated with blood disorders

Some of the probands were reported to have blood related conditions such as sickle cell anaemia (S1 and S9) and thalassemia carriers (S2, S6, S7, and S11). Analysis of the aligned sequence data indicated a homozygous deletion of *HBA2* gene in S1 and a few probands (S3, S4, S9 and S12) suggesting potential thalassemia (Figure 8), while the other probands had carrier status. However, the data could not confirm the sickle cell anaemia reported in both S1 and S9 since the rs334 (*HBB*: c.20A>T [p. Glu7Val]) which is responsible for the sickling form of haemoglobin, HbS formation was not identified despite the high coverage of the *HBB* gene (Figure 9). Instead, the data indicated genetic variations associated with other hematopathologies (Table 6). The study identified rs17261572 (AF= 0.1994) in *C1GALT1C1* (Xq24) that has a corelation to Tn Syndrome. Tn Syndrome (Polyagglutinability of the red blood cell to all human sera except autologous serum or the sera of new-borns) is an acquired disorder that is corelated to any abnormality that affect red blood cell membrane such as thalassemia, sickle cell anaemia, leukaemia and autoimmune abnormalities[47]. The homozygous rs17261572 associated with Tn syndrome was identified in six male probands. Hereditary hemochromatosis, an autosomal recessive inherited disorder of iron metabolism manifests by slow accumulation of iron in body tissues that results in many problems such as liver cirrhosis, arthritis, DM and cardiac events in later stage of life. The *HFE* gene variants were mainly implicated in the mechanism of hereditary hemochromatosis[48,49] The *HFE* gene variant rs1799945 (AF= 0.1092), which is corelated to the disorder, is homozygous in one parent and five probands. The *PROS1* variant rs146366248 (AF= 0.0007675) associated with protein S deficiency and

thrombophilia in autosomal dominant mode of transmission. This variant was also observed in one parent and seven probands.

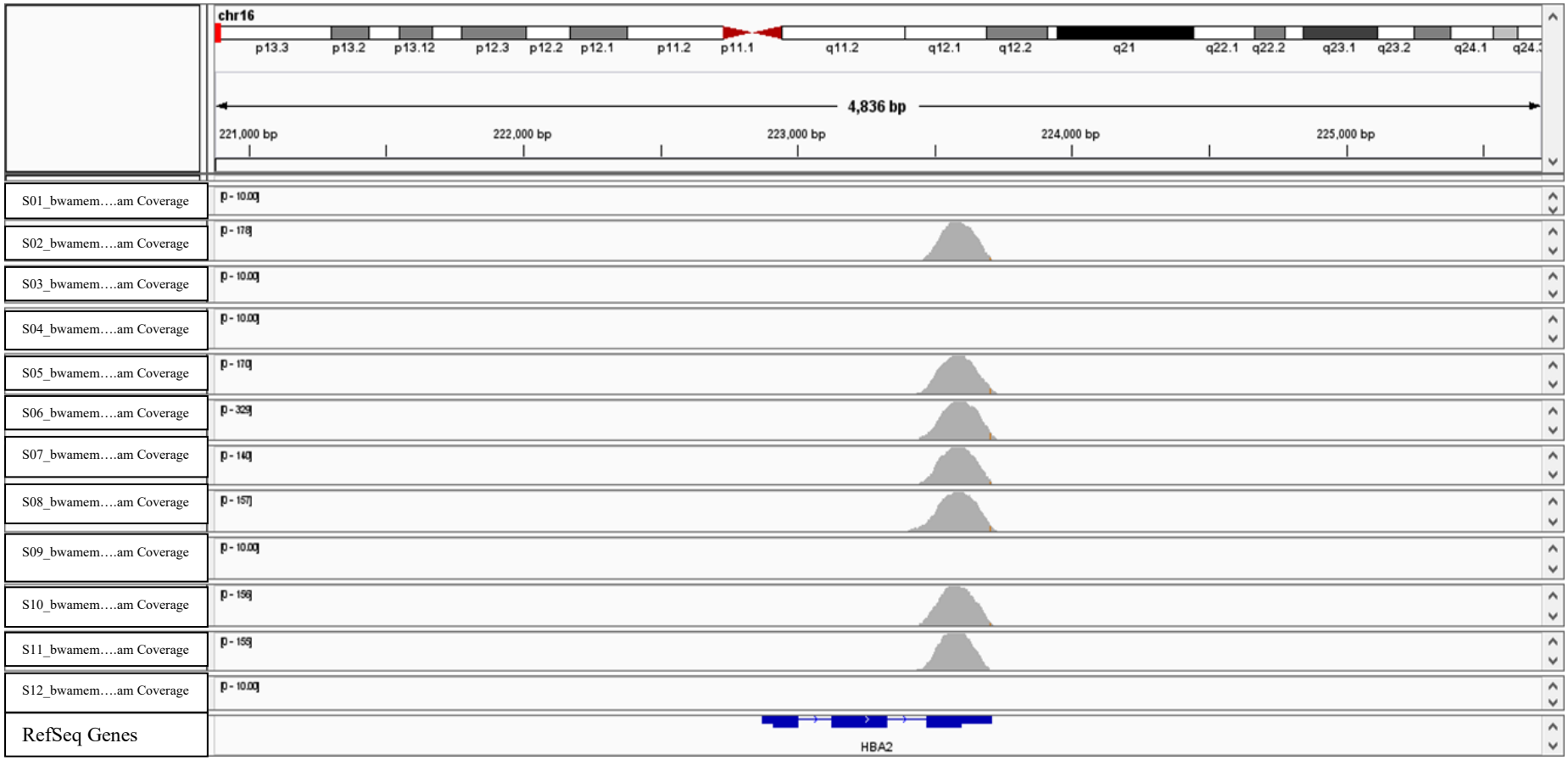


Figure 8: Aligned reads of HBA2 gene visualized using IGV showing double deletion of HBA2 gene in S1, S3, S4, S9 & S12

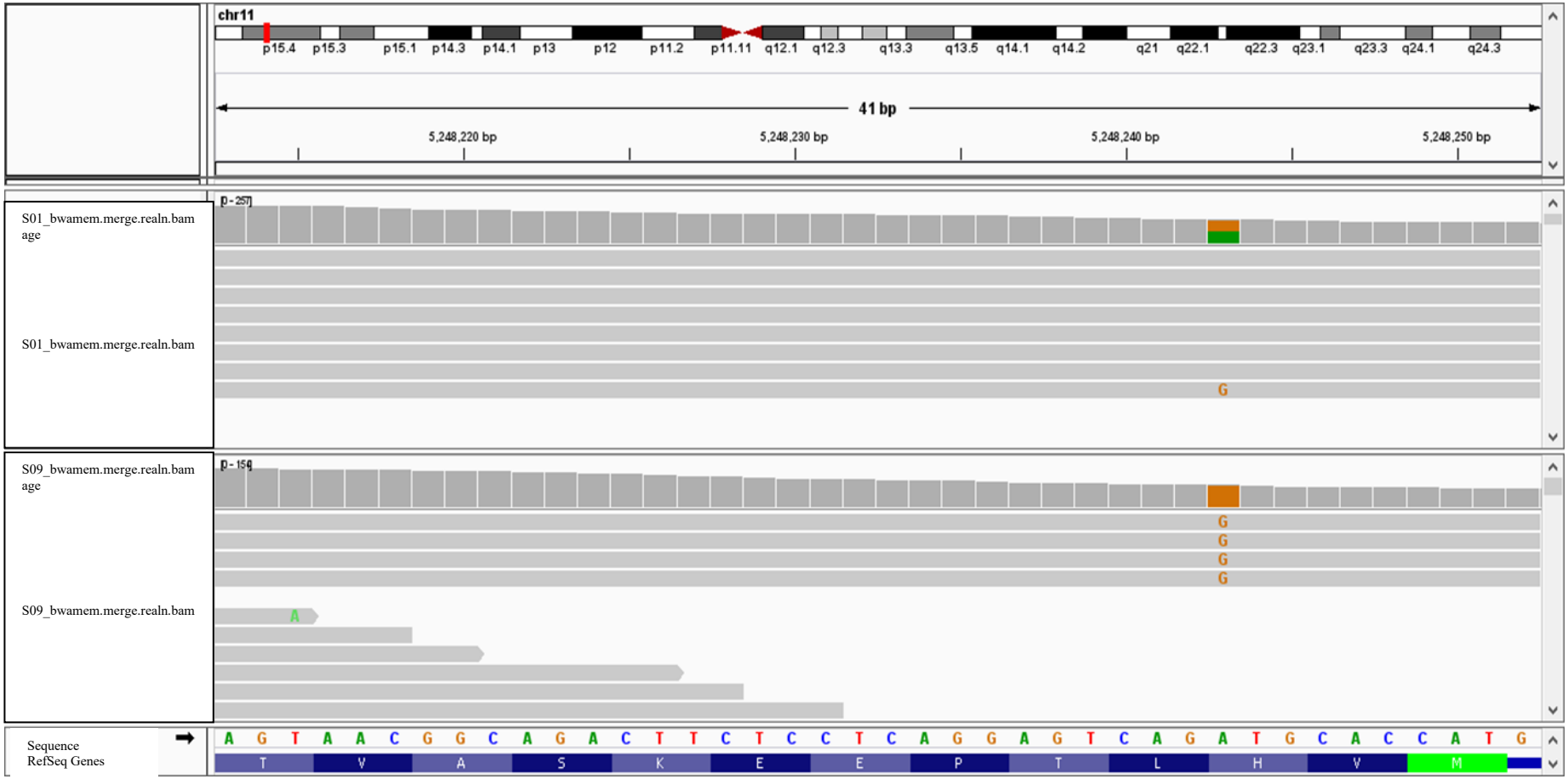


Figure 9: Aligned reads of HBB gene visualized using IGV showing no amino acid change in S1 & S9 at amino acid position 7 related to sickle cell anemia

Table 6: Variations associated with blood disorders

Genes	**Rs I.D.**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	**AF**	**Varsome significance**	**Clinvar significance**
PROS1	rs146366248	HE	.	HE	HE	.	.	HE	.	HE	HE	HE	HE	0.0007675	Likely Pathogenic	Uncertain. Thrombophilia due to protein S deficiency, autosomal dominant
C1GALT1C1	rs17261572	.	HE	HO	HO	HO	.	HO	HE	HO	HE	.	HO	0.1994	Benign	Tn Syndrome X-linked
HFE	rs1799945	HO	HE	HO	HE	HO	HE	HO	HE	HE	HO	HE	HO	0.1092	Likely Pathogenic	Pathogenic. Hereditary hemochromatosis (controversial)

3.7 Variations associated with nephropathies

Abnormalities in kidney function has also been reported in the family. Several variants associated with kidney and metabolic abnormality were identified in the probands (Table 7). The S*LC6A19* variant rs35329108 (AF= 0.2161) was observed in four probands and leads to hyperglycinuria (autosomal dominant). Hyperglycinuria results from defect in renal tubular transport of glycine and amino acids, a condition which is associated with some clinical features such as hypertension and nephrolithiasis and even mental retardation[50–52]. The *DBT* variant rs12021720, associated for maple syrup disease (autosomal recessive) represents the most common abnormal organic acidurias in infants. The disorder is a result of an abnormal activity of branched-chain alpha-keto acid dehydrogenase (BCKDH)[53,54]. An associated variant was observed in four probands.

3.8 Variations associated with other malignancies

The data also suggested the presence of variants potentially associated with cancer, a condition reported in members of the extended family. The *FGFR4* (fibroblast growth factor receptor) variant rs351855 (AF= 0.3271) was reported to be associated with metastasis and poor prognosis in breast cancer (Table 8).

Table 7: Some identified variations that associate with kidney abnormalities

Gene	rsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
SLC6A19	rs35329108	HE	.	HE				HE	HE		HE			0.2161	Benign	Pathogenic, Hyperglycinuria
DBT	rs12021720	HE	HO	HO	HO	HE	HE	HE	HE	HO	HE	HO	HO	0.9174	Benign	Benign, Maple Syrup Urine Disease
AGX T	rs34116584	HO		HE	HE	HE		HE	HE	HE	HE	HE	HE	0.1470	Benign	Conflicting interpretation of pathogenicity, Primary hyperoxaluria, type I
ATP6V0A4	rs3807153	HO		HE	HE	HE		HE	HE	HE	HE	HE	HE	0.06583	Benign	Benign, Renal tubular acidosis, distal, autosomal recessive (RTADR)]
CYP21A2	rs6467	HO	HE	HO	HE	HO	HE	HO	HE	HE	HO	HE	HO	0.6369	Benign	Conflicting interpretation of pathogenicity ,21-hydroxylase deficiency

Table 8: Some identified variations potentially associated with cancer

Gene	RsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
PTPRJ	rs1566734		HE	HE				HE	HE		HE	HE	HE	0.1686	Benign	Pathogenic, Malignant tumour of colon
FGFR4	rs351855	HO	.	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	0.3271	Benign	Pathogenic, Malignant neoplasm of prostate / Cancer progression and tumour cell motility
ELAC2	rs4792311	HE						HE	HE	HE	HE			0.2736	Benign	Pathogenic, Prostate cancer, hereditary 2

3.9 Variations associated with eye disorders

The family also has a history of eye problems (glaucoma and macular dystrophy). The study identified a few variants associated with eye problem (Table 9). Amaurosis congenita of Leber type 1 is one of retinal dystrophies which has autosomal recessive mode of transmission[55]. A reported variant in *GUCY2D* gene rs61749665 (AF= 0.4159 was inherited in some of the probands. A glaucoma/exfoliation syndrome high risk variant was also identified in the *LOXL1* gene. Also, the reported *ABCA4* s1801466 (AF= 0.04255) associated with Stargardt disease. The study was also able to identify a pathogenic *de novo* variant associated with Con Rod Dystrophy in one of the probands (Table 13).

3.10 Variations associated with autoimmune abnormalities

A family also has a history of potentially autoimmune associated disorders such as eczema, asthma and RA. Eczema was reported in S1 and in all probands, asthma in S1 and S9 and juvenile RA in S9. Pathogenic or high-risk variants related to these conditions are listed in Table 10. Most of the autoimmune disorders results from multifactorial causes and are of polygenic origin. WES identified several variants in high-risk genes related autoimmune disorders. However, most of these were not reported before to have any pathogenic significance.

3.11 Variations associated with chronic disorders

The data identified variants that pose a high risk to cardiac problems. *HNF1A* rs1169288 and *AGT* rs699 (AF= 0.3549 and 0.5481 consecutively) were identified to confer high risk to coronary artery disease (CAD). Furthermore, hypercholesterolemia is a predisposing factor to CAD. *PCSK9* rs11583680 (AF= 0.1140) variant which is

associated with familial hypercholesterolemia was identified in S1, S5 and S7 (Table 11).

Table 9: Variations associated with eye problems

Gene	rsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
GUCY2D	rs61749665	HE	.	HE	HE	HE	HE	.	.	.	.	HE	HE	0.4159	Benign	Amaurosis congenita of Leber, type 1
LOXL1	rs1048661	HO	HE	HE	HO	HO	HO	HO	HE	HO	HE	HO	HO	0.3279	Benign	High risk Glaucoma / Exfoliation syndrome (autosomal dominant)
ABCA4	rs1801466	HE	HO	HO	HO	HE	HE	HO	HE	HO	HE	HE	HO	0.04255	Benign	Likely pathogenic or uncertain significance, Stargardt disease

Table 10: Variations associated with autoimmune diseases

Gene	RsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
IRGM	rs10065172	.	HE	.	.	.	HO	HE	HE	HE	.	.	.	0.1650	Benign	Inflammatory Bowel Disease 19
ATG16L1	rs2241880	HO	HO	HO	HO	HO	HE	HO	HO	HO	HO	HO	HO	0.4532	Benign	Risk factor, Inflammatory bowel disease 10 (IBD10)
IL13	rs20541	HO	.	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	0.7206	Benign	High risk, Asthma
CTLA4	rs231775		HE	HE	HE	HE	HE				HE			0.4151	Benign	Risk factor DM, SLE & Hashimoto thyroiditis
KRT74	rs147962513	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	HO	0.01016	Benign	Conflicting interpretation of Pathogenicity, ectodermal dysplasia 7, hair/nail type
FCGR1A	rs74315310	HE	.	HE	HE					HE		HE	HE	0.004104	Pathogenic	Pathogenic, Familia deficiency of IGG receptor I phagocytic

Table 11: Variations associated with heart problems

Gene	RsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
PCSK9	rs11583680	HE				HE		HE				HE		0.1140	Benign	Hypercholesterolemia, Familial
HNF1A	rs1169288	HO	HE	HO	HE	HO	HO	HO	HO	HE	HE	HE	HO	0.3549	Benign	High risk, Coronary Artery Disease
AGT	rs699	HE	HO	HE	HE	HO	HE	HE	HO	HE	HE	HE	HO	0.5481	Benign	High risk, Coronary Artery Disease
SCN1B	rs55742440	HE	HE	HH	HE	.	HH	HH	HE	HH	HE	HE	HH	0.3684	Benign	Benign; Brugada syndrome 5.

3.12 Variations associated with epilepsy in the family

Around 400 nonsynonymous variations in epilepsy associated genes were identified in the family. Venn diagrams (Figures 10-13) represent the number of shared mutations in parents and in probands with epilepsy (S3, S9 and S11). Most of the shared variants are benign and following similar zygosity in all epilepsy and non-epilepsy reported probands. 38% of the inherited shared variants in S3, S9, S11 were same homozygosity as in the parents and similar to other probands as well, only variants *OPRM1* rs677830 and *SLC22A1* rs628031 (AF= 0.1905 and 0.6405 consecutively) existed as homozygous in epilepsy reported probands and heterozygous in both parents and other probands, however they pose only benign significance. Further analysis to identify the relevant epilepsy associated variations are provided in Table 12.

WES indicated many variations in genes associated with epilepsy. Most of those variants were reported with benign significance for different types of epilepsy and inherited in most of the probands (Table 12). The *EFHC1* gene variant rs3804506 (AF= 0.1016) is reported to have a benign association with Juvenile myoclonic epilepsy and typical absence seizures. This variant is heterozygous in both parents but homozygous in one of the epilepsies/ASD reported child (S3) and ASD reported child (S7) who were also reported to have seizers in early childhood, and heterozygous on some other probands. The rs782304760 (AF= 0.00002842) variant associated with *KDM2B* gene was reported to be associated with infantile seizers and developmental delay in Saudi Arabia. However, the variant is inherited from the father in two of ASD and seizure reported probands (S3 and S7) and one normal child. In general, several variants in genes associated with epilepsy and involved in synaptic transmission

(*CACNA1B, CACNA1H, CHRNA2, CACNA1B, RELN, SCN9A, CACNA1A*) with reported benign significance exists in all family members.

Table 12: Variations associated with epilepsy and neurological abnormalities

Gene	rsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
KDM2B	rs782304760	HE	.	HE	HE	.	.	HE	.	.	.	.	HE	0.00002842	Likely Benign	Likely benign; Global developmental delay. Microcephaly Muscular hypotonia Infantile spasms
CACNA1H	rs1054645	HO	HE	HE	HO	HE	HE	HE	HE	HO	HE	HE	HO	0.6567	Benign	Benign; Epilepsy, childhood absence 6
CPA6	rs10957393	.	HE	HE	.	HE	.	HE	HE	HE	HE	HE	.	0.2289	Benign	Epilepsy, familial temporal lobe, 5, likely/Benign; All Highly Penetrant
RELN	rs115913736	.	HE	HE	HE	.	HE	HE	HE	.	HE	HE	.	0.006603	Benign	Uncertain significance; Norman-Roberts syndrome, Benign; Epilepsy, familial temporal lobe, 7
CPA6	rs17343819	.	HE	HE	.	HE	.	HE	HE	HE	HE	HE	.	0.1123	Benign	Benign; Epilepsy, familial temporal lobe, 5.
KCNQ2	rs1801475	HE	HE	.	HE	HO	HO	.	HO	HE	.	HE	HE	0.6092	Benign	Benign; Benign familial neonatal seizures 1, Benign; Early infantile epileptic encephalopathy with suppression bursts
GOSR2	rs197922	.	HE	HE	.	HE	HE	HE	HE	HE	HE	HE	.	0.3459	Benign	Benign; Progressive Myoclonic Epilepsy, Benign; Progressive myoclonic epilepsy, X-linked, Benign; Seizures.
PRDM8	rs200010979	HE	.	HE	.	HE	.	HE	HE	HE	HE	.	.	0.001204	Benign	Benign; Epilepsy, progressive myoclonic, 10
EFHC1	rs3804506	HE	HE	HO	HE	HE	HE	HO	HE	.	.	.	.	0.1016	Benign	Benign/Likely benign; Juvenile myoclonic epilepsy, Typical absence seizures
CACNA1B	rs4422842	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	0.5000	Benign	benign; neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements
PRDM8	rs544862921	HE	.	HE	.	HE	.	HE	HE	HE	HE	.	.	0.002125	Benign	Benign; Epilepsy, progressive myoclonic, 10
CACNA1H	rs61056448	.	HE	.	HE	.	.	.	.	HE	.	.	HE	0.001938	Benign	Benign; not provided, Benign; Benign; Idiopathic generalized epilepsy Hyperaldosteronism, familial, type IV, not provided; Epilepsy, childhood absence 6

Table 12: Variations associated with epilepsy and neurological abnormalities (Continued)

Gene	rsID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	AF	Varsome significance	ClinVar significance
CACNA1H	rs61734410	HE	HE	.	HO	HE	.	.	.	HO	HE	.	HO	0.3856	Benign	Benign; Epilepsy, childhood absence 6 Generalized epilepsy with febrile seizures plus, type 7
CHRNA2	rs891398	.	HE	HE	.	.	HO	HE	.	.	HE	HE	.	0.5292	Benign	Benign; Epilepsy, nocturnal frontal lobe, type 4, Benign; Seizures.
RBFOX1	rs113298071	HE	.	.	.	HE	.	.	HE	HE	.	.	HE	0.0008156	Benign	Benign; Idiopathic generalized epilepsy
CACNA1B	rs145816559	.	HE	HE	.	HE	.	HE	HE	HE	HE	.	.	0.2155	Benign	Benign; neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements
RELN	rs150850005	HE	.	.	.	HE	.	.	.	HE	.	.	HE	0.0001877	Uncertain significance	Uncertain significance; Norman-Roberts syndrome. Familial temporal lobe epilepsy 1, Epilepsy, familial temporal lobe, 7
DEPDC5	rs16989528	.	HE	.	HE	HE	HE	.	HE	HE	.	.	.	0.03479	Benign	Benign; Familial focal epilepsy with variable foci, Seizures
CACNA1A	rs16027	.	HE	.	.	.	.	.	HE	HE	.	HE	HE	0.1088	Benign	Benign; Episodic ataxia type 2. Familial hemiplegic migraine type 1, Benign; History of neurodevelopmental disorder
WWOX	rs11545029	HE	.	HE	.	.	.	HE	.	HE	HE	HE	HE	0.4511	Benign	Uncertain significance; Spinocerebellar ataxia, autosomal recessive 12, Epileptic encephalopathy, early infantile,
HEPACAM	rs10790715	HO	.	HE	HE	HE	.	HE	HE	HE	HE	HE	HE	0.7221	Benign	Benign; not specified, Benign; Megalencephalic leukoencephalopathy with subcortical cysts Lissencephaly 2
MUSK	rs375737188	HE		HE		HE			HE				HE	0.00007599	Likely pathogenic	Myasthenic syndrome, congenital, 9, associated with acetylcholine receptor deficiency, Pena-Shokeir syndrome type I(AD)

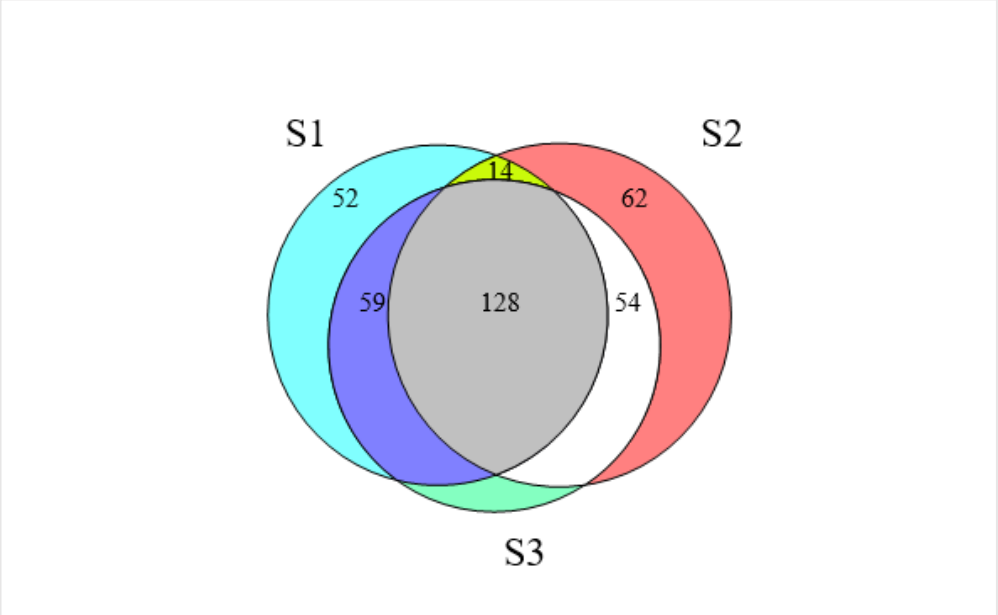


Figure 10: Number of variations identified in genes linked to epilepsy in S3 and parents

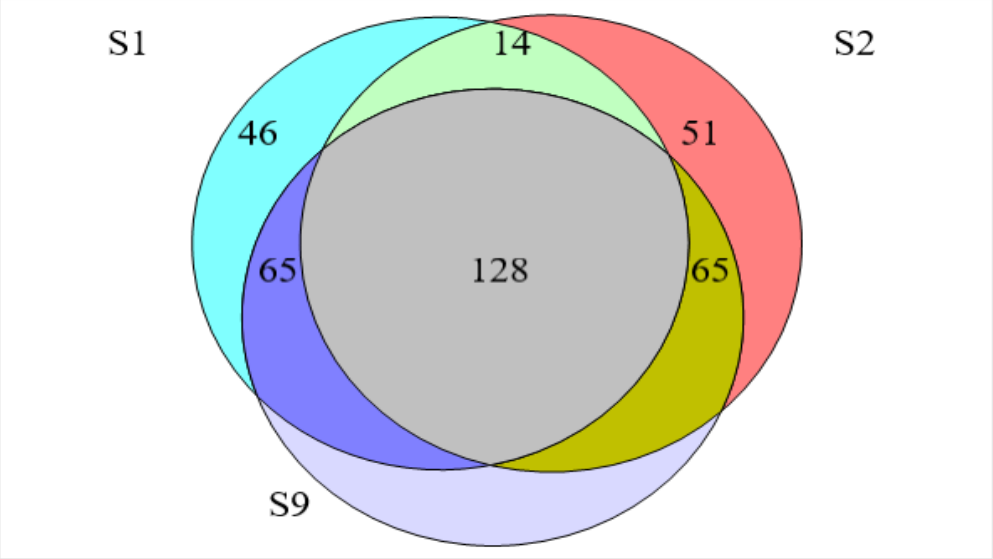


Figure 11: Number of variations identified in genes linked to epilepsy in S9 and parents

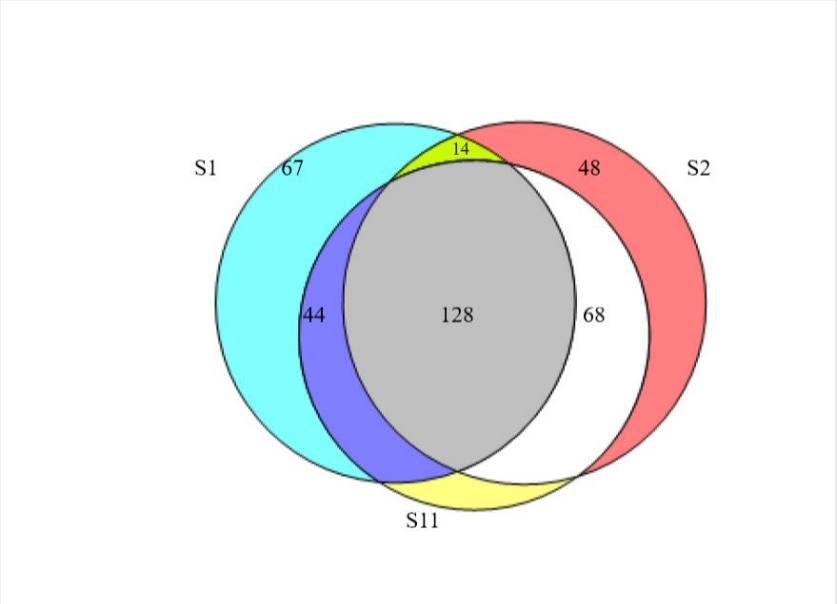


Figure 12: Number of variations identified in genes linked to epilepsy in S11 and parents

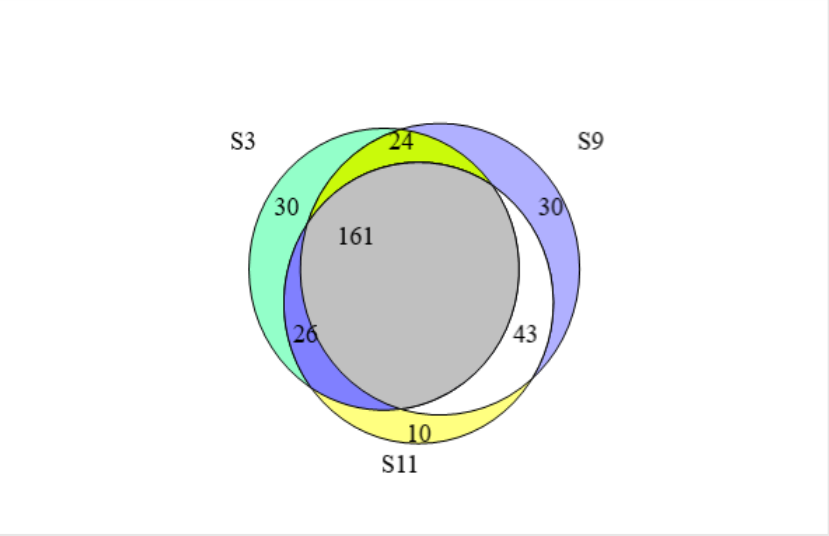


Figure 13: Number of variations identified in genes linked to epilepsy in confirmed epilepsy probands (S3, S9 and S11)

3.13 Summary of significant *de novo* and novel variants

De novo (variations that were found in probands but not inherited from the parents) nonsynonymous variants represented 2.6% of the total exonic nonsynonymous variants. Most of the *de novo* variants were heterozygous (Table 2). One of the female proband who has no ASD traits (S6) showed a high number of *de novo* variants compared to other probands (Figure 14). A high number of variations was also reported by variant callers FreeBayes, SAMtools and GATK for this sample. Visualization of the aligned BAM files also confirmed the *de novo* variations in S6. The calculated average coverage of the *de novo* nonsynonymous variants in S6 was 76X. 83% of *de novo* variants had a coverage above 30X. These variants are distributed in genes on different chromosomes (Figure 15) However, it is not clear how such a disproportionate number of variants were present only in S6.

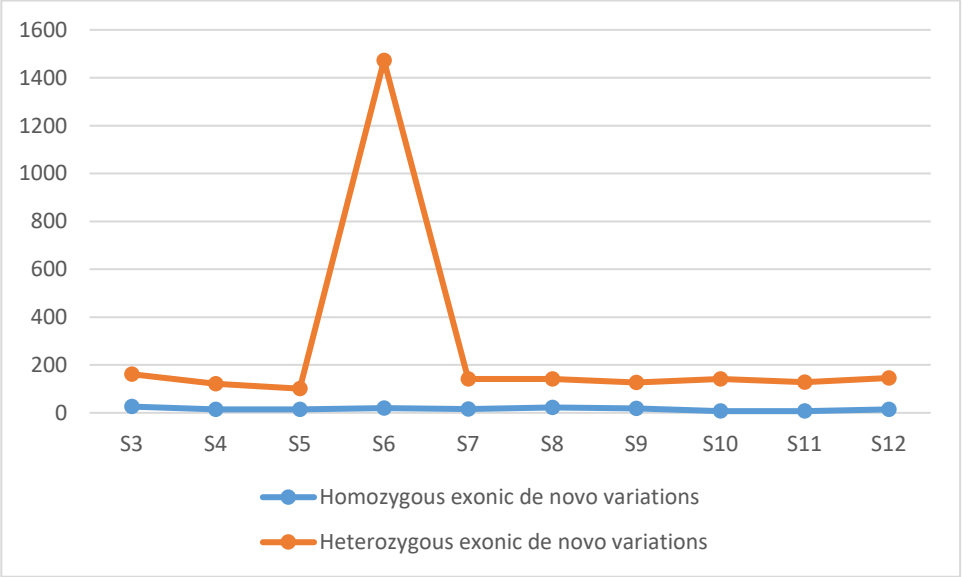


Figure 14: Number of exonic de novo mutations in the probands based on zygosity

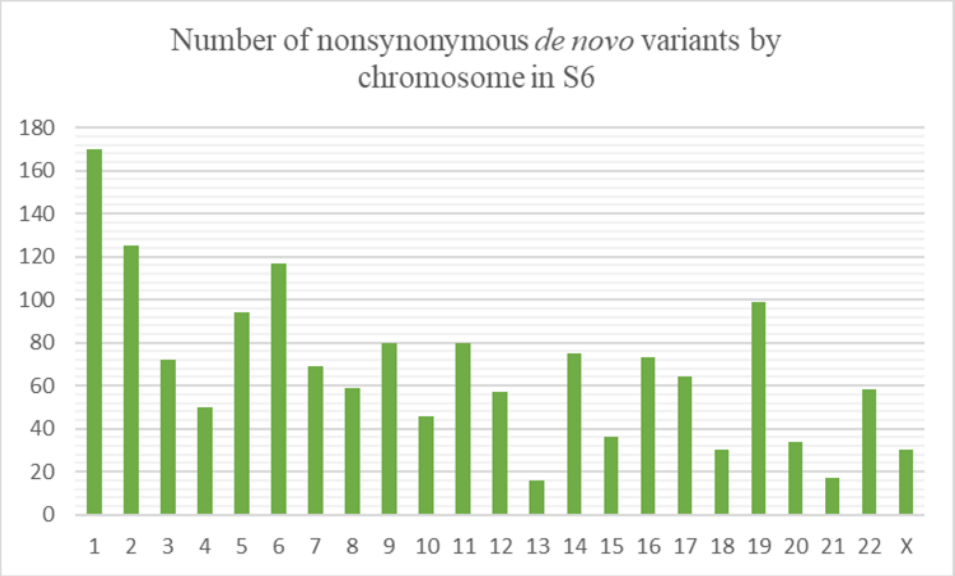


Figure 15: Number of *de novo* variants in S6 in each chromosome

Around 14% of the *de novo* variations identified were predicted to be deleterious using the computational pathogenicity predictor SIFT[56]. Novel variations, not currently listed in dbSNP and ClinVar, were detected in genes associated with both epilepsy and NDDs. *De novo* variations that were common in ASD or epilepsy affected children are highlighted here. Variants reported to be associated with some disorders are provided in Table 13. The heterozygous rs132630331 variant *GK* gene was observed in a proband with behavioural abnormality. This variant leads to glycerol kinase deficiency (GKD), an X-linked recessive disorder associated with neurodevelopmental delay[57]. Also, the rs10151259 (AF= 0.2007) in *RPGRIP1* is a variant linked to Cone-rod dystrophy 13 (CORD13). Tables 13 and 14 summarize these findings.

Table 13: Pathogenic *de novo* variations with established associations to disorders

Sample	Rs ID	Gene	Associated disorder	Zygosity	AF	Varsome significance	OMIM ID
S11(F)	rs132630331	*GK*	Glycerol kinase deficiency (GKD) X-linked recessive disorder	HE	-	Uncertain Significance	307030
S6(F)	rs150591260	*MCCC2*	3-methylcrotonoyl-CoA carboxylase 2 deficiency (MCC2D) AR	HE	0.0007837	Pathogenic	210210
S6(F)	rs10151259	*RPGRIP1*	Cone-rod dystrophy 13 (CORD13) AR	HE	0.2007	Benign	608194

Table 14: Benign and high risk *de novo* variations associated with disorders

Sample	Rs ID	Gene	Associated disorder	Zygosity	AF	Varsome significance	OMIM ID
S3(M)	rs2549677	*PKD1*	Polycystic kidney disease 1 with or without polycystic liver disease (PKD1)	HE	0.07212	Benign	173900
S6(M)	rs28564871	*ABHD12B*	A breast cancer sample	HE	0.2462	Benign	
S6(F)	rs35077384	*ZFYVE27*	Spastic paraplegia 33, autosomal dominant (SPG33)	HE	0.009874	Benign	610244
S9 (M), S10(F)& S11(F)	rs1136743	*SAA1*	Reactive amyloid systemic amyloidosis, also called AA-amyloidosis. autosomal dominant (controversial)	HE	0.4790	Benign	104750

Finally, variants shared by probands affected by ASD (S3, S5 and S7) were evaluated. Five variations were observed to be shared by all three probands (Figure 16 and Table 15). However, all of these were benign or without evidence of pathogenicity and were also found in other unaffected children with similar zygosity. Some *de novo* variants were identified to be present in epilepsy reported probands (Figure 17 and Table 16). Variants were investigated for any reported clinical significance. However, the prediction was benign for all of these according to Varsome's ACMG classification. A few novel variants not reported in dbSNP were detected with WES (Table 17). Two of these - *LMTK3* (lemur tyrosine kinase 2*)* and *MCTP2 (*Multiple C2 And Transmembrane Domain Containing 2) exist only in two ASD and neurobehavioral abnormality reported probands. Both variants are heterozygous and both genes likely follow recessive mode of transmission according to Varsome.

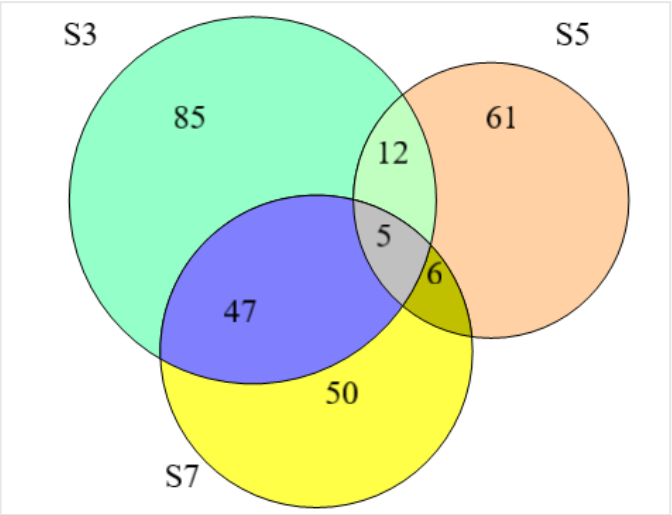


Figure 16: Venn diagram representing the total number of exonic *de novo* variants identified in ASD reported probands

Table 15: Five shared *de novo* exonic variants identified in ASD probands. AR: Autosomal recessive, AD: Autosomal dominant, AF: Allele frequency

Gene	**Rs ID**	**Significance (Varsome)**	**AF**	**Zygosity**
RBMXL2 (AR or AD)	rs11041171	Benign	0.9949	HO
OR2T27(AR)	rs28533004	Uncertain significance	0.5470	HE
FNDC1 (AR)	rs295332	Benign	0.5270	HE
SPEN (AD)	rs776474446	Uncertain significance	-	HE
HRNR (AR)	rs80018286	Uncertain significance	0.3343	HE

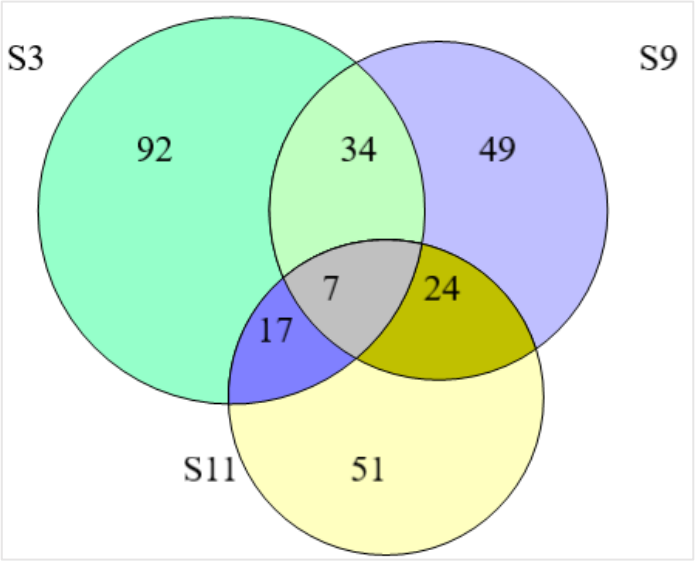


Figure 17: Venn diagram representing the total number of exonic *de novo* variants identified in epilepsy reported probands

Table 16: Seven exonic *de novo* variations shared by children with epilepsy traits

Gene	**rsID**	**Varsome significance**	**AF**	**Zygosity**
PDE4DIP	rs138083036	Benign	0.002036	HE
RASA4	rs144395384	Benign	0.1240	HE
OR2T27	rs28533004	Benign	0.5470	HE
FNDC1	rs295332	Benign	0.5270	HE
GOLGA8F	rs566731488	Benign	0.001861	HE
NUTM2F	rs75315722	Benign	0.3134	HE
SPEN	rs776474446	Benign	0.03952	HE

Table 17: Variations not reported in dbSNP but detected in ASD reported children. HE: Heterozygous, D: Deleterious, T: Tolerant, B: Benign, P: Probably damaging

Gene and cytogenic location	**Chromosomal and protein change**	**SIFT Pathogenicity Prediction**	**PolyPhen Pathogenicity Prediction**	**Gene Inheritance pattern/ Varsome**	**S3**	**S5**	**S7**
FGD1 Xp11.22	*FGD1*: NM_004463: exon3:c.C520G: p. P174A	T	B	X-linked recessive	HE	HE	HE
LMTK3 19q13.33	*LMTK3*: NM_001080434: exon15: c.: p. L1451R	D	D	Autosomal recessive	HE	HE	.
MCTP2 15q26.2	*MCTP2*: NM_001159643: exon18:c.C2221G: p.L741V, *MCTP2*: NM_018349: exon20:c.C2386G: p. L796V	D	D	Autosomal recessive	HE	HE	.
SIX2 2p21	SIX2: NM_016932: exon2: c. A853C: p. N285H	D	P	Autosomal dominant	HE	HE	.
SMPD1 11p15.4	*SMPD1*: NM_000543: exon1: c.102_103insCTGGCGCTGGCG: p.L35_V36insALAL, SMPD1: NM_001007593: exon1: c.102_103insCTGGCGCTGGCG: p.L35_V36insALAL, *SMPD1*: NM_001318087: exon1: c.102_103insCTGGCGCTGGCG: p.L35_V36insALAL, SMPD1: NM_001365135: exon1: c.102_103insCTGGCGCTGGCG: p.L35_V36insALAL	.	.	Autosomal recessive	HE	HE	HE

3.14 Variations identified in ACMG's incidental findings gene list

ACMG recommends reporting incidental findings in the exons of 59 genes[58]. WES identified an average of 230 variations in this list of genes, most of the were either benign or likely benign. Some of these variants had conflicting interpretation of pathogenicity or uncertain significance as shown in Table 18.

Table 18: Variants identified in ACMG recommended incidental findings gene list

Genes	rsID	Significance and related disorders	Varsome significance	Parental source	AF value	Probands
PMS2	rs576055272	Conflicting interpretation of pathogenicity, Hereditary cancer-predisposing syndrome. Uncertain significance, Hereditary nonpolyposis colorectal cancer type 4	Benign	S1(HE)	0.0013	S4, S5, S7, S10 & S11
TSC2	rs45517144	Uncertain significance, Hereditary cancer-predisposing syndrome & Tuberous sclerosis	Uncertain Significance	S1(HE)	0.000008	S4, S5, S9 & S12
APC	rs459552	Uncertain-Significance, Hereditary cancer-predisposing syndrome	Benign	S1(HE) & S2(HO)	0.7981	S3, S4, S7, S8, S9, S10, S11& S12(HO), S5, S6(HE)
RET	rs1799939	Uncertain-Significance, Multiple endocrine neoplasia, type 2a	Benign	S1 (HO) & S2(HE)	0.2033	S3, S7, S6, S9, S10, S11(HO), S4, S5(HE)
BRCA2	rs144848	Uncertain-Significance, Ductal breast carcinoma	Benign	S1 (HO) & S2(HE)	0.2779	S3, S5, S7, S9, S10, S11 & S12(HE) S4, S6 & S8(HO)
PCSK9	rs509504	Conflicting-Interpretations-Of-Pathogenicity, Familial hypercholesterolemia 1. Likely-Benign, Familial hypobetalipoproteinaemia	Benign	S1 & S2(HO)	0.9945	All
MSH2	rs17217723	Conflicting-Interpretations-Of-Pathogenicity, Hereditary cancer-predisposing syndrome. Uncertain-Significance, Lynch syndrome	Uncertain Significance	S2(HE)	2.78E-05	S3, S5, S6, S10, S11 &S12

3.15 Summary of variations identified

A list of all the variations included here is summarized in Table 19.

Table 19: Summary of variations reported in this study

Gene	rsID	Parental source	Probands	AF	ClinVar significance
PROS1	rs146366248	S1	S3, S4, S7, S9, S10, S11 & S11	0.0007675	Uncertain. Thrombophilia due to protein S deficiency, autosomal dominant
HFE	rs1799945	Both	All	0.1092	Pathogenic. Hereditary hemochromatosis (controversial)
SLC6A19	rs35329108	S1	S3, S7, S8 & S10	0.2161	Pathogenic, Hyperglycinuria
PTPRJ	rs1566734	S2	S3, S7, S8, S10, S11 & S12	0.1686	Pathogenic, Malignant tumour of colon
ELAC2	rs4792311	S1	S7-S10	0.2736	Pathogenic, Prostate cancer, hereditary 2
LOXL1	rs1048661	Both	All	0.3279	High risk Glaucoma / Exfoliation syndrome (autosomal dominant)
CTLA4	rs231775	S2	S3, S4, S5, S6 & S10	0.4151	Risk factor DM, SLE & Hashimoto thyroiditis
KRT74	rs147962513	Both	All	0.01016	Conflicting interpretation of Pathogenicity, ectodermal dysplasia 7, hair/nail type
FCGR1A	rs74315310	S1	S3, S4, S9, S11 & S12	0.004104	Pathogenic, Familia deficiency of IGG receptor I phagocytic
PCSK9	rs11583680	S1	S5, S7 & S11	0.114	Hypercholesterolemia, Familial
HNF1A	rs1169288	Both	All	0.3549	High risk, Coronary Artery Disease
MTHFR	rs1801133	S2	S4, S8, S9, S10 & S11	0.3149	Likely benign, Neural tube defects, folate-sensitive
	rs1801131	S2	S3, S5, S6, S7& S12	0.289	
MTR	rs1805087	Both	S3, S4, S7& S8	0.203	Disorders of Intracellular Cobalamin metabolism
	rs1131449	Both	S3, S4, S5, S7, S9, S10 & S11	0.5762	
GALT	rs2070074	S2	S3, S4, S6, S8, S9, S11 & S12	0.0991	Galactosemia (GALCT)
KDM2B	rs782304760	S1	S3, S4, S7, & S12	0.00002842	Likely benign; Global developmental delay. Microcephaly Muscular hypotonia Infantile spasms
RELN	rs150850005	S1	S5, S9 & S12	0.0001877	Uncertain significance; Norman-Roberts syndrome. Familial temporal lobe epilepsy 1, Epilepsy, familial temporal lobe, 7
	rs115913736	S2	S3, S4, S6, S7, S8, S10 & S11	0.006603	
CPA6	rs10957393	S2	S3, S5, S7, S8, S9, S10 & S11	0.2289	Epilepsy, familial temporal lobe, 5, likely/Benign; All Highly Penetrant
	rs17343819	S2	S3, S5, S7, S8, S9, S10 & S12	0.1123	
PRDM8	rs200010979	Both	S3, S5, S7, S8, S9, & S10	0.001204	Benign; Epilepsy, progressive myoclonic, 10
	rs3804506	Both	S3-S8	0.1016	Benign/Likely benign; Juvenile myoclonic epilepsy, Typical absence seizures
	rs544862921	S1	S3, S5, S7, S8, S9 & S10	0.002125	Benign; Epilepsy, progressive myoclonic, 10

Table 19: Summary of variations reported in this study (Continued)

Gene	rsID	Parental source	Probands	AF	ClinVar significance
CACNA1H	rs61056448	S2	S4, S9 & S12	0.001938	Benign; Idiopathic generalized epilepsy, Hyperaldosteronism, familial, type IV; Epilepsy, childhood absence 6
DEPDC5	rs61734410	Both	S4, S5, S9, S10 & S12	0.3856	Benign; Epilepsy, childhood absence 6, Generalized epilepsy with febrile seizures plus, type 7
HEPACAM	rs10790715	S1	All	0.7221	Benign; not specified, Benign; Megalencephalic leukoencephalopathy with subcortical cysts Lissencephaly 2
MUSK	rs375737188	S1	S3, S5, S8 & S12	0.00007599	Myasthenic syndrome, congenital, 9, associated with acetylcholine receptor deficiency, Pena-Shokeir syndrome type I(AD)
ANKRD11	rs2279348	Both	All	0.6354	Benign, Association with ASD
	rs145906515	S1	S5, S7, S9 & S12	0.00414	
	rs143743958	S2	S3, S4, S5, S8, S10 & S12	0.000534	Benign, Association with ASD & encephalopathy, progressive, early-onset, with episodic rhabdomyolysis
CC2D1A	rs11669628	S1	S3, S7, S9, S11 & S12	0.1332	Benign, History of NDDs
CHD8	rs10467770	S2	S7	0.2458	Benign, History of NDDs
VRK2	rs1051061	Both	S3, S5, S6, S8, S9, S10, S11 & S12	0.3588	Associated schizophrenia
GK	rs132630331	De novo	S11	.	Glycerol kinase deficiency (GKD) X-linked recessive disorder
KCNMA1	rs747029218	S1	All except S6	0.000007988	Uncertain significance: Paroxysmal non kinesigenic dyskinesia, 3, with or without generalized epilepsy
RBFOX1	rs113298071	S1	S5, S8, S9 & S12	0.0008156	Benign; Idiopathic generalized epilepsy
CACNA1B	rs4422842	Both	All	0.5	benign; neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements
	rs145816559	S2	S3, S5, S7, S8, S9 & S10	0.2155	

Chapter 4: Discussion

Consanguinity is a significant contributing factor for the transmission of inherited disorders in the Arabian region[59]. However, the parents in this family are not known to be related and there is no previous history of ASD or other NDDs in closely related members of the family. However, epilepsy, metabolic disorders such as diabetes mellitus (DM), autoimmune disorders (eczema and juvenile arthritis), blood disorders (sickle cell anaemia and β-thalassemia), and cancer have been reported in members of the extended family. Many genes have been linked to ASD. One of the tasks of this study was to identify why non-ASD reported parents of non-consanguineous marriage has ASD in the probands and what are the potential genetic contributing factor, if any, that could be detected in the exome. Could this be a result of a single homozygous inherited pathogenic gene variant that was inherited in recessive mode from carrier parents or perhaps a complex interplay of multiple genes that contributed to the emergence of ASD traits in some probands?

Variants associated with neurobehavioral conditions were observed in multiple ASD linked genes. *ANKRD11* mutations are associated with seizures and intellectual disability and *ANKRD11* is considered a key candidate gene for autosomal dominant intellectual function deficits[40]. Microdeletions in *ANKDR11* and *ZNF778* have also been correlated with ASD[40,60]. Variations associated with both genes exist in the study family members rs10625512 and rs9921361 (AF= 0.8138 and 0.8071 consecutively) in *ZNF778*, a homozygous SNV and a frameshift insertion, respectively, were present in ASD and non-ASD probands along with the *ANKDR11* missense variations. However, both *ZNF778* variants are not reported to be clinically significant according to ClinVar. *CC2D1A* gene encodes a transcriptional repressor protein that regulates

expression of the 5-hydroxytryptamine (serotonin) receptor (*5HTR*). Mutations associated with *CC2D1A* are likely recessive and protein-truncating mutation in *CC2D1A* have established association with ID, ASD, seizures and non-syndromic mental retardation[61–63]. *CHD8* is a member of the cadherin superfamily. It is an integral membrane protein that mediates calcium-dependent cell-cell adhesion. Altered *CHD8* expression has an effect on a wide range of genetic pathways involved in brain development such as synapse formation, cell adhesion, and neuronal differentiation[64,65]. Truncated variants and rare familial *CDH8* 16q21 microdeletions pose a high risk for autism and intellectual difficulties[64,66].

Several studies have emphasized the overlap between ASD and schizophrenia associated genetic and phenotypic traits[19,64,65]. In this study, variations were identified in several schizophrenia associated genes such as *DRD3, VRK2, LAVL4, DECR1, MTHFR, ADGRV1, RENBP, GPX3: TNIP1, TRIM31, CYP2D7, ITIH3, TAOK2, ADAMTSL3*. While most of the identified study variants related to schizophrenia had similar zygosity in the parents and probands, two variants *DRD3* rs6280 and *VRK2* rs1051061 were homozygous in only ASD or neurobehavioral abnormality reported probands. The role of rs6280 in schizophrenia has been the subject of some debate[67,68]. In Han Chinese population, the variant rs1051061 has been reported to be involved in schizophrenia pathways[69]. Whether or not those variants contribute to ASD traits needs further investigation.

Metabolic abnormalities and increased oxidative stress that results from impaired or declined capacity of both methylation and transulfuration pathways are prevalent in ASD children[70]. *MTHFR* polymorphism has also been associated with leukaemia, neural tube deficits, abortion, still birth and other NDDs[71–74].

It is possible that a combination of *MTHFR, MTR* and *MTRR* variants could have played a role in contributing to ASD in this family. In this study, WES identified an average of ten variants in *MTR* gene and five variants in *MTHFR* gene. The combination of heterozygous *MTHFR* rs1801133, rs1801131 and *MTR* rs1805087 polymorphism were highly prevalent in children with Down syndrome's and ASD in some populations[75] and the existence of both *MTHFR* rs1801133 and rs1801131 polymorphism bestow an increased ASD risk in Saudi population[76]. Complicated mechanisms underlie immune disorder pathologies and the high risk *CTLA4* rs231775 was linked to many ADs such as DM and RA[77,78]. The family history indicates the prevelance of both disorders. Some of the probands have clinical traits related to DM including polyuria and polyphagia, obesity, metabolic syndromes as well as juvenile arthritis which is suggestive of the role of these variants in autoinmmune abnormalities reported in the family.

Multiple genetic variants identified were related to haematological abnormalities reported in the family. A Tn Syndrome associated variant was identified in many probands. Clinical features associated or reported with Tn syndrome are autoimmune or haematological[47,79,80], some of which exist in the family. Another variant, *HFE* rs1799945, has been reported with conflicting reports of pathogenicity for hereditary hemochromatosis[81–83]. Good prognosis of this disorder is associated with early diagnosis and harm can be minimized by early intervention[49]. Some other identified variants confer risk to lung and heart problems, e.g. Protein S deficiency cascade results in thrombophilia, that may lead to recurrent venous thrombosis and pulmonary embolism (PE)[84]. However, a definitive association between the variants and the reported abnormalities cannot be established with just theoretical corelation discussed here. Additional functional and association studies need to be done to fully

ascertain the role of these variants and how it might contribute to the conditions in the family.

Epilepsies have variable phenotypes and those phenotypes also differ in mode of transmission and genetic contributions. Some epilepsies are thought to result from a single pathogenic mutation and proven later to be a contribution of multiple genes, e.g. nocturnal frontal lobe epilepsy was associated only with *CHRNA4* and later shown to involve other genes like *CHRNB2, CHRNA2,* and *KCNT1*[29]. In this study, multiple genes associated with different types of epilepsy were detected. Nocturnal frontal lobe epilepsy associated benign variants in this family exist on *CHRNA2*. Familial temporal lobe epilepsy associated benign variants in *RELN* and *CPA6* were also identified. Benign variants rs1801475 in *KCNQ2* and rs11545029 in *WWOX* were associated with early infantile epileptic encephalopathy which has autosomal dominant mode of transmission. However, all the WES identified variants are benign and they had a similar zygosity in all probands.

The *UBA1* gene variant rs2070169 was found only in probands who have ASD and/or epilepsy. However, this variant has only been reported to have a benign association with spinal muscular atrophy X-linked 2, a rare X-linked recessive disorder which results from neuronal degeneration and muscular hypotonia[85]. The probands have not been reported to have these traits.

Glycerol kinase deficiency (GKD) is associated with severe neurodevelopmental delay in infants along with the glucose intolerance and hyperglycerolemia[57]. GKD could manifest in phenotypes of behavioural abnormalities along with GIT and kidney abnormalities. *GK* mutations have X-linked recessive pattern of inheritance[86].

3-methylcrotonoyl-CoA carboxylase 2 (MCC2) enzyme plays an important role in both leucine and biotin metabolic process and protein hetero-oligomerization. 3-methylcrotonoyl-CoA carboxylase 2 deficiency (MCC2D) has a large number of phenotypes including seborrheic dermatitis, intellectual disability, seizure, muscular hypotonia, lethargy, global developmental delay, generalized hypotonia, hyperreflexia, metabolic acidosis, organic aciduria, ketoacidosis, vomiting, ketonuria, hyperglycinuria, feeding difficulties[87]. Several conditions including eczema, GIT abnormalities and nephropathies were observed in the children with different degrees. However, the proband who carries a heterozygous pathogenic *de novo MCCC2* rs150591260 variant has no other significant clinical abnormalities. This variant was reported in a study to diagnose children with inborn errors of metabolism in UAE where this pathogenic variant was falsely detected by WES[88].

The genetic picture that underlies NDDs and epilepsies in the family cannot be precisely pinned down using the WES data. The nature of the complicated genetic background, as well as contributing environmental factors, of both disorders have been under investigation in large cohort studies. These studies have emphasized the heterogenicity of the genetic abnormalities including structural abnormalities and copy number variations (CNVs), which could have been the case in this family. There could have been other contributing factors to ASD in the family. For example, there is a higher incidence of ASD in preterm infants and infections. While all the children in this family were born preterm, not all of them developed ASD traits. However, this, along with genetic predisposition could have resulted in ASD in some children.

A few metabolic pathways associated genes have been variations in this family. *MTHFR, MTR* and *MTRR* variants are widely reported in NDDs and abortion, conditions that have been reported in this family. Autoimmune abnormality has also

been reported in the family that may be associated with identified variations in many genes. The contribution of *de novo* variants to NDDs is not significant in this study. Most of the *de novo* identified variants in ASD reported probands are benign, have not been reported before or not restricted to ASD reported probands.

One of the limitations of this study was the use of WES to identify variations. Though inexpensive, this technique only introspects around 1-2% of the genome that covers the exome. This cannot be effectively used to identify variations in the whole genome or chromosomal abnormalities. Other advanced and expensive techniques such as chromosomal microarrays or whole genome sequencing may be necessary to get the full range of genomic variations. In this study, the human hg19 reference, which is one of the most widely used reference genomes, was used. However, this is not truly representative of the Arabian population. Hence, several variants identified may be polymorphisms that exist in this population. Additionally, another limitation was the inability to verify the effect of some of these variations on the individuals. However, the insights gained here provides several directions that could be pursued to understand the genetic underpinnings of various disorders that run in this family.

Chapter 5: Conclusion

Several genes have been directly or indirectly implicated in the pathophysiology of NDDs. Many pathogenic variants have also been reported in the coding region of these genes. Hence, WES was employed as a rapid and cost-effective approach to study these genes and its coding region. However, in this instance, the exomic region does not provide a clear indication of any underlying genetic reason that could have led to NDDs, specifically ASD, in this family. Most likely the cause may be chromosomal structural abnormalities, in other non-coding regions that make up more than 98% of the genome or copy number variations. These are nearly impossible to identify from exome sequencing techniques. More expensive and detailed analysis using whole genome sequencing and chromosomal microarrays are recommended for such studies. Most of the detected pathogenic or high-risk variants are in genes involved in the metabolism and accounts for some of the conditions reported in the family. Several genes that are related to NDDs have several inherited benign or *de novo* variations. Compound heterozygous variations and expressivity of these in the family members may also account for why some probands have ASD phenotype while others do not. Clearly, male children are more affected with ASD and epilepsy. The reason for that would also require further work to confirm.

References

1. American Psychiatric Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub. Accessed September 22, 2020. https://www.psychiatry.org/psychiatrists/practice/dsm

2. Cardoso AR, Lopes-Marques M, Silva RM, et al. Essential genetic findings in neurodevelopmental disorders. *Hum Genomics*. 2019;13(1):31. doi:10.1186/s40246-019-0216-4

3. Rizzo R, Gulisano M, Domini CN, Ferro MC, Curatolo P. The Relationship between Autism Spectrum Disorder and Tourette Syndrome in Childhood: An Overview of Shared Characteristics. *J Pediatr Neurol*. 2017;15(03):115-122. doi:10.1055/s-0037-1602821

4. Nelson KB, Chang T. Is cerebral palsy preventable? *Curr Opin Neurol*. 2008;21(2):129-135. doi:10.1097/WCO.0b013e3282f4958b

5. Kremer EJ, Yu S, Pritchard M, et al. Isolation of a human DNA sequence which spans the fragile X. *Am J Hum Genet*. 1991;49(3):656-661.

6. Sherr EH. Chapter 36 - Neurodevelopmental Disorders, Causes, and Consequences. In: Lehner T, Miller BL, State MW, eds. *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*. Academic Press; 2016:587-599. doi:10.1016/B978-0-12-800105-9.00036-6

7. Lamy M, Erickson CA. Pharmacological management of behavioral disturbances in children and adolescents with autism spectrum disorders. *Curr Probl Pediatr Adolesc Health Care*. 2018;48(10):250-264. doi:10.1016/j.cppeds.2018.08.015

8. Elsabbagh M, Divan G, Koh Y-J, et al. Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Res*. 2012;5(3):160-179. doi:10.1002/aur.239

9. Autism spectrum disorders. Accessed September 22, 2020. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

10. Agrawal S, Rao SC, Bulsara MK, Patole SK. Prevalence of Autism Spectrum Disorder in Preterm Infants: A Meta-analysis. *Pediatrics*. 2018;142(3): e20180134. doi:10.1542/peds.2018-0134

11. NIMH » Autism Spectrum Disorder. Accessed September 22, 2020. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml

12. Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet*. 2001;2(12):943-955. doi:10.1038/35103559

13. Schuch JB, Mariath LM, Roman T, Schuler-Faccini L. The Genetic Basis of Autism Spectrum Disorder. In: Robinson-Agramonte M de los A, ed. *Translational Approaches to Autism Spectrum Disorder*. Springer International Publishing; 2015:39-63. doi:10.1007/978-3-319-16321-5_3

14. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015;87(6):1215-1233. doi:10.1016/j.neuron.2015.09.016

15. Jamain S, Quach H, Betancur C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 2003;34(1):27-29. doi:10.1038/ng1136

16. Pinto D, Pagnamenta AT, Klei L, et al. Functional Impact of Global Rare Copy Number Variation in Autism Spectrum Disorder. *Nature*. 2010;466(7304):368-372. doi:10.1038/nature09146

17. Sebat J, Lakshmi B, Malhotra D, et al. Strong Association of De Novo Copy Number Mutations with Autism. *Science*. 2007;316(5823):445-449. doi:10.1126/science.1138659

18. De Rubeis S, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209-215. doi:10.1038/nature13772

19. Turinsky AL, Turner B, Borja RC, et al. DAnCER: disease-annotated chromatin epigenetics resource. *Nucleic Acids Res*. 2011;39 (Database issue):D889-894. doi:10.1093/nar/gkq857

20. McCarthy SE, Gillis J, Kramer M, et al. De novo Mutations in Schizophrenia Implicate Chromatin Remodeling and Support a Genetic Overlap with Autism and Intellectual Disability. *Mol Psychiatry*. 2014;19(6):652-658. doi:10.1038/mp.2014.29

21. Siu MT, Weksberg R. Epigenetics of Autism Spectrum Disorder. *Adv Exp Med Biol*. 2017;978:63-90. doi:10.1007/978-3-319-53889-1_4

22. Genovese A, Butler MG. Clinical Assessment, Genetics, and Treatment Approaches in Autism Spectrum Disorder (ASD). *Int J Mol Sci*. 2020;21(13):4726. doi:10.3390/ijms21134726

23. Onore C, Careaga M, Ashwood P. The role of immune dysfunction in the pathophysiology of autism. *Brain Behav Immun*. 2012;26(3):383-392. doi:10.1016/j.bbi.2011.08.007

24. Croen LA, Qian Y, Ashwood P, et al. Family history of immune conditions and autism spectrum and developmental disorders: Findings from the study to explore early development. *Autism Res*. 2019;12(1):123-135. doi:10.1002/aur.1979

25. Atladóttir HÓ, Pedersen MG, Thorsen P, et al. Association of family history of autoimmune diseases and autism spectrum disorders. *Pediatrics*. 2009;124(2):687-694. doi:10.1542/peds.2008-2445

26. Al-Hakbany M, Awadallah S, AL-Ayadhi L. The Relationship of HLA Class I and II Alleles and Haplotypes with Autism: A Case Control Study. Autism Research and Treatment. doi:https://doi.org/10.1155/2014/242048

27. Al-Mubarak B, Abouelhoda M, Omar A, et al. Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from Saudi families. *Sci Rep*. 2017;7(1):5679. doi:10.1038/s41598-017-06033-1

28. Epilepsy. Accessed September 27, 2020. https://www.who.int/news-room/fact-sheets/detail/epilepsy

29. Wang J, Lin Z-J, Liu L, et al. Epilepsy-associated genes. *Seizure*. 2017;44:11-20. doi:10.1016/j.seizure.2016.11.030

30. Ellis CA, Petrovski S, Berkovic SF. Epilepsy genetics: clinical impacts and biological insights. *Lancet Neurol*. 2020;19(1):93-100. doi:10.1016/S1474-4422(19)30269-8

31. Biervert C, Schroeder BC, Kubisch C, et al. A Potassium Channel Mutation in Neonatal Human Epilepsy. *Science*. 1998;279(5349):403-406. doi:10.1126/science.279.5349.403

32. Charlier C, Singh NA, Ryan SG, et al. A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. *Nat Genet*. 1998;18(1):53-55. doi:10.1038/ng0198-53

33. Wallace RH, Wang DW, Singh R, et al. Febrile seizures and generalized epilepsy associated with a mutation in the Na + -channel ß1 subunit gene SCN1B. *Nat Genet*. 1998;19(4):366-370. doi:10.1038/1252

34. Tuchman R, Cuccaro M. Epilepsy and Autism: Neurodevelopmental Perspective. *Curr Neurol Neurosci Rep*. 2011;11(4):428-434. doi:10.1007/s11910-011-0195-x

35. Long S, Zhou H, Li S, et al. The Clinical and Genetic Features of Co-occurring Epilepsy and Autism Spectrum Disorder in Chinese Children. *Front Neurol*. 2019;10. doi:10.3389/fneur.2019.00505

36. McTague A, Howell KB, Cross JH, Kurian MA, Scheffer IE. The genetic landscape of the epileptic encephalopathies of infancy and childhood. *Lancet Neurol*. 2016;15(3):304-316. doi:10.1016/S1474-4422(15)00250-1

37. Sener EF, Canatan H, Ozkul Y. Recent Advances in Autism Spectrum Disorders: Applications of Whole Exome Sequencing Technology. *Psychiatry Investig*. 2016;13(3):255-264. doi:10.4306/pi.2016.13.3.255

38. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515(7526):216-221. doi:10.1038/nature13908

39. Guo Y, Ding X, Shen Y, Lyon GJ, Wang K. SeqMule: automated pipeline for analysis of human exome/genome sequencing data. *Sci Rep*. 2015;5(1):14283. doi:10.1038/srep14283

40. Isrie M, Hendriks Y, Gielissen N, et al. Haploinsufficiency of ANKRD11 causes mild cognitive impairment, short stature and minor dysmorphisms. *Eur J Hum Genet*. 2012;20(2):131-133. doi:10.1038/ejhg.2011.105

41. Andrade F, Aldámiz-Echevarría L, Llarena M, Couce ML. Sanfilippo syndrome: Overall review. *Pediatr Int Off J Jpn Pediatr Soc*. 2015;57(3):331-338. doi:10.1111/ped.12636

42. Wijburg FA, Węgrzyn G, Burton BK, Tylki-Szymańska A. Mucopolysaccharidosis type III (Sanfilippo syndrome) and misdiagnosis of idiopathic developmental delay, attention deficit/hyperactivity disorder or autism spectrum disorder. *Acta Paediatr*. 2013;102(5):462-470. doi:10.1111/apa.12169

43. Katsanis N, Ansley SJ, Badano JL, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science*. 2001;293(5538):2256-2259. doi:10.1126/science.1063525

44. Mykytyn K, Nishimura DY, Searby CC, et al. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat Genet*. 2002;31(4):435-438. doi:10.1038/ng935

45. Saarela J, Laine M, Oinonen C, et al. Molecular pathogenesis of a disease: structural consequences of aspartylglucosaminuria mutations. *Hum Mol Genet*. 2001;10(9):983-995. doi:10.1093/hmg/10.9.983

46. Picco P, Garibaldi L, Cotellessa M, Di Rocco M, Borrone C. Corticosterone methyl oxidase type II deficiency: a cause of failure to thrive and recurrent dehydration in early infancy. *Eur J Pediatr*. 1992;151(3):170-173. doi:10.1007/BF01954376

47. Berger EG. Tn-syndrome. *Biochim Biophys Acta BBA - Mol Basis Dis*. 1999;1455(2):255-268. doi:10.1016/S0925-4439(99)00069-1

48. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet*. 1996;13(4):399-408. doi:10.1038/ng0896-399

49. Griffiths W, Cox T. Haemochromatosis: novel gene discovery and the molecular pathophysiology of iron metabolism. *Hum Mol Genet*. 2000;9(16):2377-2382. doi:10.1093/hmg/9.16.2377

50. Bröer S, Bailey CG, Kowalczuk S, et al. Iminoglycinuria and hyperglycinuria are discrete human phenotypes resulting from complex mutations in proline and glycine transporters. *J Clin Invest*. 2008;118(12):3881-3892. doi:10.1172/JCI36625

51. Käser H, Cottier P, Antener I. Glucoglycinuria, a new familial syndrome. *J Pediatr*. 1962;61(3):386-394. doi:10.1016/S0022-3476(62)80369-2

52. Vries A de, Kochwa S, Lazebnik J, Frank M, Djaldetti M. Glycinuria, a hereditary disorder associated with nephrolithiasis. *Am J Med*. 1957;23(3):408-415. doi:10.1016/0002-9343(57)90320-0

53. Baulny HO de, Saudubray JM. Branched-chain organic acidurias. *Semin Neonatol*. 2002;7(1):65-74. doi:10.1053/siny.2001.0087

54. Tsuruta M, Mitsubuchi H, Mardy S, et al. Molecular basis of intermittent maple syrup urine disease: novel mutations in the E2 gene of the branched-chain alpha-keto acid dehydrogenase complex. *J Hum Genet*. 1998;43(2):91-100. doi:10.1007/s100380050047

55. Chung DC, Traboulsi EI. Leber congenital amaurosis: clinical correlations with genotypes, gene therapy trials update, and future directions. *J AAPOS Off Publ Am Assoc Pediatr Ophthalmol Strabismus*. 2009;13(6):587-592. doi:10.1016/j.jaapos.2009.10.004

56. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814.

57. Gaudet D, Arsenault S, Pérusse L, et al. Glycerol as a correlate of impaired glucose tolerance: dissection of a complex system by use of a simple genetic trait. *Am J Hum Genet*. 2000;66(5):1558-1568. doi:10.1086/302903

58. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):249-255. doi:10.1038/gim.2016.190

59. Bener A, Dafeeah EE, Samson N. Does consanguinity increase the risk of schizophrenia? Study based on primary health care centre visits. *Ment Health Fam Med*. 2012;9(4):241-248.

60. Willemsen MH, Fernandez BA, Bacino CA, et al. Identification of ANKRD11 and ZNF778 as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome. *Eur J Hum Genet*. 2010;18(4):429-435. doi:10.1038/ejhg.2009.192

61. Basel-Vanagaite L, Taub E, Halpern GJ, et al. Genetic screening for autosomal recessive nonsyndromic mental retardation in an isolated population in Israel. *Eur J Hum Genet EJHG*. 2007;15(2):250-253. doi:10.1038/sj.ejhg.5201750

62. Manzini MC, Xiong L, Shaheen R, et al. CC2D1A Regulates Human Intellectual and Social Function as well as NF-κB Signaling Homeostasis. *Cell Rep*. 2014;8(3):647-655. doi:10.1016/j.celrep.2014.06.039

63. Zamarbide M, Oaks AW, Pond HL, Adelman JS, Manzini MC. Loss of the Intellectual Disability and Autism Gene Cc2d1a and Its Homolog Cc2d1b Differentially Affect Spatial Memory, Anxiety, and Hyperactivity. *Front Genet*. 2018;9. doi:10.3389/fgene.2018.00065

64. Sugathan A, Biagioli M, Golzio C, et al. CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci U S A*. 2014;111(42):E4468-E4477. doi:10.1073/pnas.1405266111

65. Wilkinson B, Grepo N, Thompson BL, et al. The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Transl Psychiatry*. 2015;5:e568. doi:10.1038/tp.2015.62

66. Pagnamenta AT, Khan H, Walker S, et al. Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. *J Med Genet*. 2011;48(1):48-54. doi:10.1136/jmg.2010.079426

67. Crocq MA, Mant R, Asherson P, et al. Association between schizophrenia and homozygosity at the dopamine D3 receptor gene. *J Med Genet*. 1992;29(12):858-860.

68. Spurlock G, Williams J, McGuffin P, et al. European multicentre association study of schizophrenia: a study of the DRD2 Ser311Cys and DRD3 Ser9Gly polymorphisms. *Am J Med Genet*. 1998;81(1):24-28. doi:10.1002/(SICI)1096-8628(19980207)81:1<24::AID-AJMG5>3.0.CO;2-N

69. Yu HG, Yan HK, Li JM, et al. Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. Published online 2017. doi:10.1038/mp.2016.212

70. James SJ, Melnyk S, Jernigan S, et al. Metabolic endophenotype and related genotypes are associated with oxidative stress in children with autism. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet*. 2006;141B(8):947-956. doi:10.1002/ajmg.b.30366

71. Ah A, Na E. Methylenetetrahydrofolate Reductase Gene Variants Confer Potential Vulnerability to Autism Spectrum Disorder in a Saudi Community. Neuropsychiatric disease and treatment. doi:10.2147/NDT.S230348

72. Chung-Filho AA, Brisson GD, Vieira TMF, et al. MTHFR rs1801133 polymorphism is associated with increased risk of B-cell precursor lymphoblastic leukaemia with recurrent genetic aberrations of fetal origin. *Cancer Epidemiol*. 2020;65:101693. doi:10.1016/j.canep.2020.101693

73. H Z, B R, M P, et al. Increased frequency of combined methylenetetrahydrofolate reductase C677T and A1298C mutated alleles in spontaneously aborted embryos. European journal of human genetics : EJHG. doi:10.1038/sj.ejhg.5200767

74. Ma L, Jiang Y, Kong X, et al. Interaction of MTHFR C677T polymorphism with smoking in susceptibility to diabetic nephropathy in Chinese men with type 2 diabetes. *J Hum Genet*. 2019;64(1):23-28. doi:10.1038/s10038-018-0531-y

75. Jiajin L, Shuyan C, Ying W, Junxiao C, Xiudi W. Genetic polymorphisms in folate metabolism as risk for Down syndrome in the southern China. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet*. 2019;32(12):2030-2035. doi:10.1080/14767058.2018.1424818

76. Parmeggiani F, Gallenga CE, Costagliola C, et al. Impact of methylenetetrahydrofolate reductase C677T polymorphism on the efficacy of photodynamic therapy in patients with neovascular age-related macular degeneration. *Sci Rep*. 2019;9(1):2614. doi:10.1038/s41598-019-38919-7

77. Marron MP, Raffel LJ, Garchon HJ, et al. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet*. 1997;6(8):1275-1282. doi:10.1093/hmg/6.8.1275

78. Nisticò L, Buzzetti R, Pritchard LE, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. *Hum Mol Genet*. 1996;5(7):1075-1080. doi:10.1093/hmg/5.7.1075

79. Beck ML. Red blood cell polyagglutination: clinical aspects. *Semin Hematol*. 2000;37(2):186-196. doi:10.1016/s0037-1963(00)90043-x

80. Vainchenker W, Vinci G, Testa U, et al. Presence of the Tn antigen on hematopoietic progenitors from patients with the Tn syndrome. *J Clin Invest*. 1985;75(2):541-546. doi:10.1172/JCI111730

81. Aguilar-Martinez P, Bismuth M, Picot MC, et al. Variable phenotypic presentation of iron overload in H63D homozygotes: are genetic modifiers the cause? *Gut*. 2001;48(6):836-842. doi:10.1136/gut.48.6.836

82. Beutler E. The significance of the 187G (H63D) mutation in hemochromatosis. *Am J Hum Genet*. 1997;61(3):762-764.

83. Tomatsu S, Orii KO, Fleming RE, et al. Contribution of the H63D mutation in HFE to murine hereditary hemochromatosis. *Proc Natl Acad Sci U S A*. 2003;100(26):15788-15793. doi:10.1073/pnas.2237037100

84. Pintao MC, Garcia AA, Borgel D, et al. Gross deletions/duplications in PROS1 are relatively common in point mutation-negative hereditary protein S deficiency. *Hum Genet*. 2009;126(3):449-456. doi:10.1007/s00439-009-0687-9

85. Kobayashi H, Baumbach L, Matise TC, Schiavi A, Greenberg F, Hoffman EP. A gene for a severe lethal form of X-linked arthrogryposis (X-linked infantile spinal muscular atrophy) maps to human chromosome Xp11.3-q11.2. *Hum Mol Genet*. 1995;4(7):1213-1216. doi:10.1093/hmg/4.7.1213

86. Dipple KM, Zhang YH, Huang BL, et al. Glycerol kinase deficiency: evidence for complexity in a single gene disorder. *Hum Genet*. 2001;109(1):55-62. doi:10.1007/s004390100545

87. Shepard PJ, Barshop BA, Baumgartner MR, et al. Consanguinity and rare mutations outside of MCCC genes underlie nonspecific phenotypes of MCCD. *Genet Med Off J Am Coll Med Genet*. 2015;17(8):660-667. doi:10.1038/gim.2014.157

88. Al-Shamsi A, Hertecant JL, Souid A-K, Al-Jasmi FA. Whole exome sequencing diagnosis of inborn errors of metabolism and other disorders in United Arab Emirates. *Orphanet J Rare Dis*. 2016;11(1):94. doi:10.1186/s13023-016-0474-3

Appendix

List of genes referenced in the thesis

Abbreviation	Gene name
ABHD12B	Abhydrolase Domain Containing 12B
ADNP	Activity Dependent Neuroprotector Homeobox
AKT	AKT Serine/Threonine Kinase 1
ANK2	Ankyrin 2
ANKRD11	Ankyrin Repeat Domain 11
APC	APC Regulator Of WNT Signalling Pathway
APH1A	Aph-1 Homolog A, Gamma-Secretase Subunit
APOE	Apolipoprotein E
ARID1B	AT-Rich Interaction Domain 1B
ASH1L	ASH1 Like Histone Lysine Methyltransferase
ASXL3	ASXL Transcriptional Regulator 3
ATRX	ATRX Chromatin Remodel
BCL11A	BAF Chromatin Remodelling Complex Subunit BCL11A
CACNA2D3	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 3
CC2D1A	Coiled-Coil And C2 Domain Containing 1A
CHD7	Chromodomain Helicase DNA Binding Protein 7
CHD8	Chromodomain Helicase DNA Binding Protein 8
COMT	Catechol-O-Methyltransferase
CTTNBP2	Cortactin Binding Protein 2
CUL3	Cullin 3
DAOA	D-Amino Acid Oxidase Activator
DEPDC5	DEP Domain Containing 5, GATOR1 Subcomplex Subunit
DISC1	DISC1 Scaffold Protein
DNMT3A	DNA Methyltransferase 3 Alpha
DRD3	Dopamine Receptor D3
DYRK1A	Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A
ETFB	Electron Transfer Flavoprotein Subunit Beta
FKTN	Fukutin
FMR1	FMRP Translational Regulator 1
FNDC1 (AR)	Fibronectin Type III Domain Containing 1
GABRB3	Gamma-Aminobutyric Acid Type A Receptor Subunit Beta 3
GABRG2	Gamma-Aminobutyric Acid Type A Receptor Subunit Gamma 2
GAD1	Glutamate Decarboxylase 1
GK	Glycerol Kinase
GLO1	Glyoxalase I

List of genes referenced in the thesis (Continued)

Abbreviation	**Gene name**
GPCR	G Protein Coupled Receptor
GRIN2B	Glutamate Ionotropic Receptor NMDA Type Subunit 2B
HOXA1	Homeobox A1
HRNR/EGFR	Hornerin/Epidermal Growth Factor Receptor
IRS1	Insulin Receptor Substrate 1
KATNAL2	Katanin Catalytic Subunit A1 Like 2
KCNJ12	Potassium Inwardly Rectifying Channel Subfamily J Member 12
KCNJ18	Potassium Inwardly Rectifying Channel Subfamily J Member 18
KCNQ2	Potassium Voltage-Gated Channel Subfamily Q Member 2
KCNQ3	Potassium Voltage-Gated Channel Subfamily Q Member 3
MAP1B	Microtubule Associated Protein 1B
MCCC1	Methylcrotonoyl-Coa Carboxylase 1
MCCC2	Methylcrotonoyl-Coa Carboxylase 2
MECP2	Methyl-Cpg Binding Protein 2
MIB1	Mindbomb E3 Ubiquitin Protein Ligase 1
MOGS	Mannosyl-Oligosaccharide Glucosidase
MTHFR	Methylenetetrahydrofolate Reductase
MTR	5-Methyltetrahydrofolate-Homocysteine Methyltransferase
MTRR	5-Methyltetrahydrofolate-Homocysteine Methyltransferase Reductase
MYH9	Myosin Heavy Chain 9
MYO9B	Myosin IXB
MYT1L	Myelin Transcription Factor 1 Like
NAA15	N-Alpha-Acetyltransferase 15, Nata Auxiliary Subunit
NAGLU	N-Acetyl-Alpha-Glucosaminidase
NGF	Nerve Growth Factor
NLGN3	Neuroligin 3
NLGN4X	Neuroligin 4 X-Linked
NLRP2	NLR Family Pyrin Domain Containing 2
NR3C2	Nuclear Receptor Subfamily 3 Group C Member 2
NRXN1	Neurexin 1
OR2T27	Olfactory Receptor Family 2 Subfamily T Member 27
PAX5	Paired Box 5
PHF2	PHD Finger Protein 2
PKD1	Polycystin 1, Transient Receptor Potential Channel Interacting
POGZ	Pogo Transposable Element Derived with ZNF Domain
PTEN	Phosphatase and Tensin Homolog

List of genes referenced in the thesis (Continued)

Abbreviation	Gene name
RBMXL2	RBMX like 2
RELN	Reelin
RET	Ret Proto-Oncogene
RPGRIP1	RPGR Interacting Protein 1
SCN1A	Sodium Voltage-Gated Channel Alpha Subunit 1
SCN1B	Sodium Voltage-Gated Channel Beta Subunit 1
SCN2A	Sodium Voltage-Gated Channel Alpha Subunit 2
SETD5	SET Domain Containing 5
SGSH	N-Sulfoglucosamine Sulfohydrolase
SHANK3	SH3 And Multiple Ankyrin Repeat Domains 3
SPEN (AD)	Spen Family Transcriptional Repressor
SYNGAP1	Synaptic Ras Gtpase Activating Protein 1
TBR1	T-Box Brain Transcription Factor 1
TCF4	Transcription Factor 4
TGFB1	Transforming Growth Factor Beta 1
TRIO	Trio Rho Guanine Nucleotide Exchange Factor
TSC1	TSC Complex Subunit 1
TSC2	TSC Complex Subunit 2
TTC21B	Tetratricopeptide Repeat Domain 21B
UBE3A	Ubiquitin Protein Ligase E3A
VIL1	Villin 1
ZFYVE27	Zinc Finger FYVE-Type Containing 27