Identification of Putative Pathogenic Variants in Genes Responsible for Intellectual disability

By

Ghazala Shaheen

Department of Biochemistry,

Quaid-i-Azam University

Islamabad, Pakistan

2018

Acknowledgement

"Certainly Allah has (complete) knowledge of that which is hidden in the Heavens and the Earth and Allah is the One who Sees what all of you (His creations) do." (Surah al-Hujurat, verse: 20). All praise is due to Allah, Who is the most beneficent and merciful. I am extremely thankful to my Almighty Allah for His countless blessings on me and enabled me to accomplish my quest of knowledge. I offer my countless salutations to the Holy Prophet Muhammad (blessings and Peace be upon Him), who is an excellent model of humanity and enabled us to recognize our creator and enlightened us with knowledge and showed us a right path.

I would like to pay my huge gratitude to the families for their cooperation and donation of blood samples for this study. I express my deepest gratitude to my spiritual father an honourable personality **Dr. Muhammad Ansar** for his patience, support and constant guidance throughout my study. I am blessed to have a mentor like him, who has encouraging and inspiring personality that helped me to complete my study with a great learning experience. It has been a great honour to work under his kind supervision. I must say he is one of the cherished personalities of our department.

I am extremely grateful to my respectable seniors, Dr. Ehsan Ullah, Sundus Sajid, Memoona Rasheed, Madiha Amin and Rabbia Basharat for their guidance and constant support and positive criticism during my research work which increased my knowledge and learning experience. I am also grateful to my lab fellows; Muhammad Zahid, Zainab Khan, Sana Iqbal, Zahid Ullah, Irfan Ullah, Sumaira Aslam, Maria, Maryum, Fizza, Umayyah for their assistance in various ways during my research work. I am also thankful to Muhammad Qadeer for their assistance and cooperation during my lab work. I am also grateful to the office staff members of the Biochemistry Department for their cooperation in all my official work.

I would also like to pay my gratitude to my lovely friends, Gulfareen Hafeez, Sehrish Jabeen, Saira Bibi and all my classmates who support me directly or indirectly throughout my research work. I would like to thank my beloved family especially my lovely sisters and brothers. Without their support and encouragement nothing would have been possible for me. I cannot forget to pay my gratitude to my beloved brothers Khalid Mehmood and Tariq Mehmood for their unconditional love and support. I am so blessed to have family like you. Thank you so much for all your support and guidance. I truly don't have words

to express my love and respect to all of you who support me, encouraged me and helped me to accomplish this task and achieve my goal.

Ghazala Shaheen

Abstract

Intellectual disability (ID) is a neurodevelopmental disorder that seriously affects cognition, adaptive and intellectual functions of the individuals. It is a heterogeneous disorder with prevalence rate of 1-3%. There are ~700 known ID genes but the underlying genes are unknown in large number of ID patients.

For this study two families (A & B) were recruited from the remote areas of Sindh and Punjab, provinces of Pakistan. Both families exhibit non-syndromic ID. Whole genome single nucleotide polymorphism (SNP) genotyping was performed on both families which results in the identification of multiple homozygous regions in family B only. For family A, the genotype data was manually accessed to find the presence of genomic deletion, which identified a probable deletion on chromosome 2. This region was flanked by rs6706472 and rs12713108 and spans a genomic region of ~559kb (chr2:50,228,332-50,786,770), but PCR based assays did not confirm the presence of deletion in the affected members of this family.

In case of family B, Homozygosity Mapper identified 4 homozygous regions. Exome sequencing of IV-4 individual of this family could not identify the ID causing variant. As a consequence, BAM file of this individual was evaluated to check the coverage of genes present in the HBD regions. For chromosome 1: 38401933-64109264 HBD region two genes *POMGNT1* (exon-23) and *SZT2* (exon-5-15) were partially uncovered while on chromosome 14:56307478-75947712 homozygous by descent (HBD) region *KIAA0586* (exon-6 & 33) gene was uncovered. Sanger sequencing of these uncovered regions, could not identify any pathogenic variant. Only one polymorphism was found in exon-6 of *KIAA0586* gene which was present in the intronic region, but has no impact on splicing. Though no pathogenic variant was identified in these families by using exome sequencing, the disease causing mutations can be identified by using genome sequencing.

Abbreviations	Full Name
ABI	Applied bio-system
ADID	Autosomal dominant Intellectual Disability
ARID	Autosomal Recessive Intellectual Disability
ASDs	Autism Spectrum Disorders
BAM	Binary Alignment/Map
BEC	Bioethics Committee
BLAT	Blast like alignment tool
bp	Base pair
◦C	Degree centigrade
CNS	Central Nervous System
CNVs	Copy Number Variations
CSF	Cerebrospinal Fluid
dbSNP	Single Nucleotide Polymorphism Database
EDTA	Ethylene di-amine tetra acetate
EtBr	Ethidium bromide
ExAC	Exome Aggregation Consor
ExoSAP	Exonuclease1-Shrimp Alkaline Phosphatase
F.P	Forward Primer
g	Gram
gDNA	Genomic DNA
HBD	Homozygous by descent region
HC	Head Circumference
HSF	Human splice site Finder
ID	Intellectual Disability
IQ	Intelligence Quotient
JS	Joubert syndrome

LDs	Learning Disabilities
MEB	Muscle-Eye-Brain
Mb	Mega base
Mins	Minutes
Ml	Milli litre
mM	Milli mole
NGS	Next Generation Sequencing
NS-ARID	Non- Syndromic Autosomal Recessive Intellectual Disability
NS-ID	Non- Syndromic Intellectual Disability
NVLD	Non-verbal Learning Disability
OMIM	Online Mendelian inheritance in Man
PNS	Peripheral Nervous System
S-ARID	Syndromic Autosomal Recessive Intellectual Disability
SDS	Sodium dodecyl- Sulphate
Sec	Second
SGS	Second Generation Sequencing
S-ID	Syndromic Intellectual Disability
SIFT	Sorting Intolerant From Tolerant
SNP	Single Nucleotide Polymorphism
ROH	Runs of Homozygosity
Rpm	Revolution per minute
R.P	Reverse Primer
TBE	Tris Borate EDTA
TE	Tris-EDTA
UCSC	University of California Santa Cruz
µl	Micro litre

V	Voltage
VCF	Variant Call File
WES	Whole exome sequencing

Table of Contents

Acknowledgement……..... i
List of Tables……..... iii
List of Figures……..... iv
List of Abbreviations…….. vi
Abstract……... ix
1. Introduction…….. 1
1.1 Development of Human Brain……...….. 1
1.1.1 Neurogenesis……..……... 2
1.1.2 Brain Parts and their Functions……. 2
1.1.2.1 Cerebrum……..……..... 2
1.1.2.2 Cerebral Cortex……. 2
1.1.2.3 Prefrontal Cortex……..… 3
1.1.2.4 Basal Nuclei…….. 3
1.1.2.5 Hippocampus…….... 3
1.1.2.6 Cerebellum……... 3
1.1.3 Neurodevelopmental Disabilities……..…… 6
1.1.3.1 Autism Spectrum Disorder (ASDs)…….…. 6
1.1.3.2 Learning Disabilities (LDs)……..……. 7
1.1.3.3 Intellectual Disabilities (ID)……..…… 7
1.1.3.3.1 Epidemiology of ID……..…….. 7
1.1.3.3.2 Disease Etiology…….……... 7
1.1.3.3.2.1 Genetic Factors…….……... 8
1.1.3.2.1.1 Chromosomal Abnormalities…….…….. 8
1.1.3.2.1.2 Monogenic Forms of ID…….…. 8
1.1.3.3.2.2 Non-genetic Factors……..……... 9
1.1.4 Classification……..……... 9
1.1.4.1 ID types based on IQ level and severity……..…….. 9
1.1.4.1.1 Mild ID……..……... 9
1.1.4.1.2 Moderate ID……..……... 10
1.1.4.1.3 Severe ID……..…… 10
1.1.4.1.4 Profound ID……..…… 10
1.1.4.2 ID types based on mode of inheritance…….. 10

1.1.4.2.1 X-linked Forms of ID……. 10
1.1.4.2.2 Autosomal Dominant ID…….…… 11
1.1.4.2.3 Autosomal Recessive Intellectual Disability (ARID)…… 11
1.1.4.2.3.1 Syndromic Autosomal Recessive Intellectual Disability (S-ARID)…… 11
1.1.4.2.3.1.1 SZT2……. 11
1.1.4.2.3.1.2 POMGNT1…….……... 12
1.1.4.2.3.1.3 KIAA0586…….…… 12
1.1.4.2.3.1.3.1 Joubert Syndrome (JS)……. 13
1.1.4.2.3.1.4 NRXN1……..…… 14
1.1.4.2.3.2 Non-Syndromic Autosomal Recessive Intellectual Disability (NS-ARID)……..……. 14
1.1.5 Next generation Sequencing (NGS)……..……... 15
1.1.5.1 SNP Genotyping……..…….. 16
1.1.5.1.1 Advantages and Disadvantages of SNP Genotyping..…… 17
1.1.5.2 Whole Exome Sequencing (WES)…….……... 18
1.1.5.2.1 Benefits and Challenges of Exome Sequencing…….…… 20
1.1.6 Aims and Objectives……..……... 22
2. Materials and Methods……..... 23
2.1 Ethical Approval and Subjects……..…... 23
2.2 Clinical Evaluation of Affected Patients……..……... 23
2.3 Sample Collection……..…….. 23
2.4 Genomic DNA Extraction……..……..... 24
2.4.1 DNA Quantification……..……..... 25
2.5 Genetic Analysis of ID Families……..…… 27
2.5.1 Genome Wide Genotyping……..……. 27
2.5.1.1 Family A……..…….. 27
2.5.2 Whole Exome Sequencing (WES)…….…... 30
2.6 PCR Amplification……..……. 30
2.6.1 PCR Amplification Confirmation……..…… 31
2.7 Purification of Amplified Products by ExoSAP-IT……..……... 31
2.8 Sequencing Reaction……..…… 32
2.8.1 Sequenced Data Analysis……..…..... 33

3. Results…….. 36
3.1 Family Description…….…..... 36
3.1.1 Family A……..……. 36
3.1.2 Family B……..…….. 36
3.2 Genetic Analysis……...…… 41
3.2.1 Whole genome genotyping and Homozygosity mapping…….... 41
3.2.1.1 Family A……..…….. 41
3.2.1.2 Family B..... 44
3.2.1.2.1 Exome Sequencing in Family B……..……. 44
3.2.1.2.1.1 Uncovered Genes during Exome Sequencing of Family B…… 49
4. Discussion……..…… 59
5. References……..…… 63
Electronic References……..……. 82

List of Figures

Figure No.	Title	Page No.
1.1	Embryonic human brain development……	5
2.1	Graphical presentation of primer designed to test the probable deletion detected in the family A ……...	29
3.1	Pedigree of family (A) with five generations and eight affected individuals……...	37
3.2	Pedigree of four generation family (B) with three affected individual…….	39
3.3	Schematic presentation of SNP no call region on chr2: p16.3 of family A and its ruling out through PCR amplification ……	42
3.4	Gel image of PCR products amplified from upstream and downstream of probable deleted region……....	43
3.5	Homozygosity Mapper image of HBD regions identified in family B	45
3.6	Screen shot of exon-23 of *POMGNT1* gene which is partially covered in the exome sequence obtained from affected individual IV-4…….	47
3.7	Screen shot of exon 5-15 of *SZT2* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4…….	47
3.8	Screen shot of exon-6 of *KIAA0586* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4…….	48
3.9	Screen shot of exon-33 of *KIAA0586* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4…….	48
3.10	Chromatogram of exon 23 of *POMGNT1* gene sequenced in affected individual IV-4 of family B……..	50

3.11	Chromatogram of exon 5 of *SZT2* gene sequenced in affected individual IV-4 of family B……	51
3.12	Chromatogram of exon 6 & 7 of *SZT2* gene sequenced in affected individual IV-4 of family B……	52
3.13	Chromatogram of exon 8-9 of *SZT2* gene sequenced in affected individual IV-4 of family B……	53
3.14	Chromatogram of exon 10-11 of *SZT2* gene sequenced in affected individual IV-4 of family B…….	54
3.15	Chromatogram of exon 12-13 of *SZT2* gene sequenced in affected individual IV-4 of family B……	55
3.16	Chromatogram of exon 14-15 of *SZT2* gene sequenced in affected individual IV-4 of family B……	56
3.17	Chromatogram of exon 6 & 33 of *KIAA0586* gene sequenced in affected individual IV-4 of family B……..	57

List of Tables

Table No.	Title	Page No.
2.1	List of Solutions used for DNA extraction along with their composition...	**26**
2.2	Sequence of primers used to amplify and sequence a genomic region with probable deletion in family A…….	**34**
2.3	Sequence of primers used to amplify and sequence the uncovered exons of genes present in the HBD region identified in family B…….	**35**
3.1	Clinical features of affected individuals of family A……..	**38**
3.2	Clinical features of affected individuals of family B…….....	**40**
3.3	List of homozygous by descent (HBD) regions identified by Homozygosity Mapper…….....	**45**
3.4	Genomics coordinates of uncovered regions identified in individual IV-4……	**46**
3.5	Variants identified during the sequencing of *POMGNT1*, *SZT2* and *KIAA0586* genes in family B……...	**58**

1. INTRODUCTION

Two major parts of the *Homo sapiens* nervous system include central nervous system (CNS) and the peripheral nervous system (PNS). The CNS comprises of the spinal cord and the brain while the PNS involves the nerve fibres that originate from spinal cord and spread to the body parts like neck, skeletal muscles, arms, internal organs, legs and torso. The brain is a soft, wrinkled organ having approximately 100 billion neurons located inside the cranial cavity and it weighs about 3 pounds. The approximate storage capacity of human brain is $1.25x10^{12}$ bytes (Cherniak, 1990; Hofman, 2012).

1.1 Development of Human Brain

The development of human brain is a chain of processes (which are dynamic and adaptive) that promotes the differentiation and development of neural structures. These processes work within extremely controlled and genetically organized, but continuously changing environments, that support the development of dynamic and complex structures (Stiles, 2008). The first neural tissue that develops in human brain is neural plate which is also called primitive neural tissue and this neural tissue is part of the outermost embryonic tissue layer known as ectoderm. The ectoderm is present on the surface of early embryo, and later forms the epidermis of the skin. In the early stages of development a groove appears in the ectoderm, which deepens and fused to form a neural tube. The fused part of ectoderm develops into a separate structure called neural crest which is present in the middle of the neural tube and the ectoderm. Eventually, the neural crest will develop in to the ganglia of the PNS while neural tube will become the CNS. On the neural tube's anterior end, three distinct swellings are formed which later form the forebrain (Prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon). In the developed human brain, forebrain has two parts diencephalon and telencephalon whereas the hindbrain has metencephalon and myelencephalon (Fig 1.1). The telencephalon disproportionately grows into enormous cerebral hemispheres that shield the midbrain, the diencephalon, and a portion of the hindbrain (Fig 1.1). The CNS originates as a hollow tube and remains hollow until the brain regions are formed. The cavities of the brain (ventricles) are filled with cerebrospinal fluid (CSF).

In an adult brain 10^{11} (100 billion) neurons are present which gets about 15% of the total blood flow per minute, and thus indicates the high metabolic needs of brain (Fox, 2016).

1.1.1 Neurogenesis

The brain of adult human contains neural stem cells which can develop into neurons and glial cells. There are two locations in the human brain that are involved in the development of new neurons from neural stem cells. One of the locations is the subventricular zone; new neurons are generated in this area and migrate into striatum. The other location is the subgranular zone of the hippocampus in which neurogenesis results in newly formed interneurons that functions within the hippocampus (Fox, 2016).

1.1.2 Brain Parts and their Functions

1.1.2.1 Cerebrum

The only structure of telencephalon is the cerebrum which is the major portion of brain and accounts for almost 80% of its mass. Within two convoluted hemispheres it contains five paired lobes, encloses gray matter in its cortex and in deeper cerebral nuclei. The cerebral hemispheres are interconnected internally by a large fiber tract called the corpus callosum (Fox, 2016). Cerebrum is primarily involved in determining intelligence, thinking, reasoning, personality, motor functions, producing and understanding language, interpretation of sensory impulses and processing of sensory information (Bailey, 2013).

1.1.2.2 Cerebral Cortex

The outer folds and grooves (convolutions) of cerebrum are called cerebral cortex which is comprised of underlying white matter and 2 to 4 mm of gray matter. The depressed grooves of the convolutions are called sulci while the elevated folds are called gyri. Each cerebral hemisphere is further divided into five lobes by fissures or deep sulci. These lobes are the frontal, parietal, temporal and occipital while the fifth lobe is deep insula which is enclosed by the portions of the parietal, frontal and temporal lobes. Frontal lobe is involved in voluntary motor control of skeletal muscles, higher intellectual processing (e.g., planning and decision making); personality and verbal communication (Fox, 2016). Parietal lobe is involved in speech, muscular and cutaneous sensations and expression of emotions and thoughts (Fox, 2016). Temporal lobe has role in understanding of acoustic sensations and storage (memory) of visual and auditory experiences (Fox, 2016). Occipital lobe is primarily involved in vision and coordination of eye movements (Fox, 2016). Insula has role in sensory (principally pain), memory and visceral integration (Fox, 2016).

1.1.2.3 Prefrontal Cortex

In the frontal lobes, the anterior most portions are occupied by prefrontal cortex. Prefrontal cortex occupies one-third portion of entire cerebral cortex. Prefrontal cortex has ability to carry out and initiate patterns of behaviour, constant attention (Luria, 1966), short term memory tasks (Chao, 1998), working memory (Fuster *et al*., 2000), stimulus detection and sequencing tasks (Lepage *et al*., 1996), delayed responding, set shifting, flexibility and effective problem solving (Romine *et al*., 2004). Prefrontal cortex is also involved in other higher cognitive functions like planning, memory and judgement (Fox, 2016).

1.1.2.4 Basal Nuclei

Basal nuclei are composed of masses of gray matter that contain neuron cell bodies located deep in the cerebral white matter. Basal nuclei and related nuclei engaged primarily in motor control, motor learning, emotions, executive functions and behaviour (Lanciego *et al*., 2012). Basal ganglia refers to nuclei rooted deep in the brain hemispheres (globus pallidus, striatum or caudate-putamen) while related nuclei include structures localized in the diencephalon (subthalamic nucleus), mesencephalon (substantia nigra) and pedunculopontine nucleus in pons (Lanciego *et al*., 2012).

1.1.2.5 Hippocampus

The hippocampus is like a seahorse-shaped and is a part of limbic system. The limbic system is present just below the cortex and above the brain stem. The hippocampus is found in the temporal lobe and has two parts present on each hemisphere. The hippocampus is a $1/100^{th}$ part of the cerebral cortex and is mainly involved in learning and memory. It plays important role in two types of memories: declarative and spatial relationship memories. Declarative memories include memories associated with facts and events while spatial relationship memories include routes or pathways. The spatial relationship memories are stored in the right hippocampus, and this function is compromised in epilepsy, Alzheimer's and depression (Anand & Dhikav, 2012).

1.1.2.6 Cerebellum

The 2^{nd} largest structure of the brain is the cerebellum, having about 50 billion neurons. It also contains outer gray matter and inner white matter like cerebrum. It contributes in the coordinating movements, by receiving input from proprioceptors (tendon, joint, and muscle receptors) and mutually working with the motor areas of the cerebral cortex and

the basal nuclei. It is also required for coordinating movements of different joints, proper timing and force required for limb movement e.g., cerebellum is necessary for different actions i.e. to take a fork of food to your mouth, or touch your nose with your finger etc. It is suggested that the cerebellum have different functions beyond motor coordination including attainment of sensory data, emotion, memory, and other higher functions (Fox, 2016). It is also believed that cerebellum may also be involved in autism and schizophrenia (Fox, 2016).

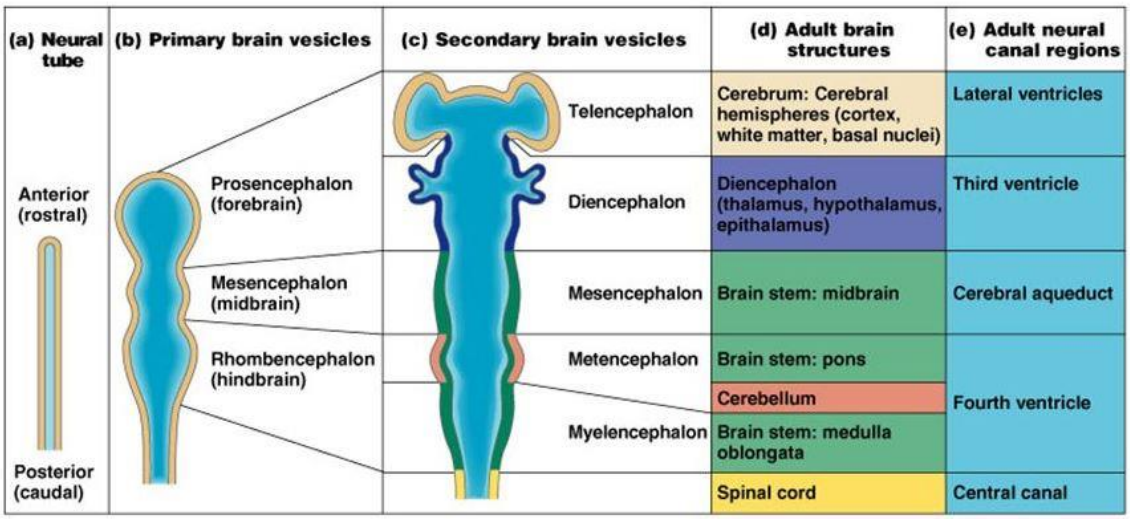

Figure 1.1: Embryonic human brain development (Adopted from Marieb & Hoehn, 2010).

1.1.3 Neurodevelopmental Disabilities

Neurodevelopmental disabilities are a large group of chronic conditions that occur during early developmental processes (including conception, birth and growth) up to the age of 22 years and last during the lifetime of an individual (Accardo, 2008). As a result of prenatal, metabolic, perinatal, and other factors, fundamental deficits occur in the developing brain which gives rise to these disorders (WHO, 2001). In the United States, about 17% of individuals (below18 years of age) are affected by developmental disabilities (Accardo, 2008). Major categories of developmental disabilities include: autism spectrum disorder, learning disabilities, intellectual disabilities, neurogenetic disorders, communication disorders, neurobehavioral disorders, neurometabolic disorders, neuromuscular disorders, sensory impairments, neuromotor disorders, traumatic brain injuries and disabilities linked with chronic diseases and spinal cord injuries (Patel *et al*., 2011a).

1.1.3.1 Autism Spectrum Disorder (ASDs)

ASDs are a group of neurodevelopmental disorders having deficits in communication, social and behavioural domains (Centres for Disease Control and Prevention, 2009). The three major types of ASDs include asperger syndrome, autistic disorder and pervasive developmental disorder. In asperger syndrome, individuals have unusual behaviours and interests, and deficits in social domain but they do not have intellectual disability and deficits in language (Patel *et al*., 2011a). The individuals with ASD have significant language delays and unusual behaviours and they face social and communication challenges (Matson & Shoemaker, 2009). In comparison to asperger syndrome, individuals with autistic disorder have significant percentage of intellectual disability (Patel *et al*., 2011a)**.** In pervasive developmental disorder, individuals have some features of asperger syndrome and some features of autistic disorder (http://www.cdc.gov/ncbddd/autism, 2009).

In United States the overall occurrence of ASDs is 9/1000 in children below 8 years age (Centres for Disease Control and Prevention, 2009), but male incidence rate is four times greater than females (http://www.cdc.gov/ncbddd/autism, 2009). There is a high degree of covariance between ASDs and IDs, about 50-70% of ASD individuals will have some type of ID while 40% of those with IDs will also have an ASD (Matson & Shoemaker, 2009).

1.1.3.2 Learning Disabilities (LDs)

Learning disability is defined as the difficulty in particular areas of learning and it directly effect the academic performance of an individual. The different types of learning disabilities are mathematics disability (dyscalculia), reading disability (dyslexia), non-verbal learning disability (NVLD) and disorder of written expression (dysgraphia) (Lagae, 2008; Patel & Merrick, 2011b). Early identification of LDs is essential to prevent behavioural disturbances, low early school dropout, self-confidence, academic failure, socioeconomic welfare, occupation and subsequently health (Patel *et al.*, 2011a). LD can be identified when a child or teenager's scores are significantly below in a separately administered standardized attainment test (in mathematics, interpretation, or written expression) than expected on his/her age level of IQ and education (Lagae, 2008; Patel & Merrick, 2011b).

1.1.3.3 Intellectual Disability (ID)

Intellectual disability (ID) also known as general learning disability is a neurodevelopmental disorder having impaired intellectual and adaptive functions. ID is caused due to chromosomal abnormalities, genetic mutations and environmental factors. ID is characterised by intellectual quotient (IQ) below 70 having significant impairment in cognitive and adaptive behaviour (American Psychiatric Association, 2013).

1.1.3.3.1 Epidemiology of ID

ID prevalence rate in general population is 1-3%. The most predominant ID risk factors in developing countries are cultural deprivation, genetic causes (due to consanguineous marriages) and malnutrition (Khan *et al*., 2016). In western countries the approximate incidence of ID is around 1.5-2% while in the developing countries it is about twice as that of the developed countries (Leonard and Wen, 2002; Maulik *et al*., 2011). ID is more common in males than in females with the percentage of 30% but this difference goes down in case of severe ID where females have higher ratio than males (Katusic *et al*., 1996; Bradley *et al*., 2002).

1.1.3.3.2 Disease Etiology

ID development and its prevalence involve genetic and non-genetic factors. The non-genetic factors include environmental and socio-economic factors (Iqbal *et al*., 2014).

1.1.3.3.2.1 Genetic Factors

Genetic factors that contribute in ID include chromosomal abnormalities and monogenic/multigenic disorders like metabolic defects.

1.1.3.3.2.1.1 Chromosomal Abnormalities

Changes in chromosomal structure include aneuploidies, sub- microscopic deletions, copy number variations, duplications, and balanced/unbalanced translocations. It is estimated that about 22% of severe ID patients have chromosomal defects (Stromme, 2000). Chromosomal abnormalities mostly lead to the syndromic forms of ID and this probably results due to the deletion of adjacent genes (Karnebeek *et al.,* 2005). Among the chromosomal aneuploidies; trisomy 21 is the most common cause of ID, with the prevalence rate of 1/1000 live births (Ballif *et al.,* 2007; Ledbetter *et al*., 2007). Down syndrome can also occur due to the translocation between chromosome 21 and chromosome 15. The analysis of large patient cohorts revealed that depending on patient selection, copy number variations (CNVs) may contribute about 15-20% of all ID individuals (Zahir *et al*., 2007; Cooper *et al.,* 2011). CNVs cause periodic conditions of ID along with additional features such as in Prader-Willi syndrome, Smith-Magenis syndrome and Williamms-Beuren syndrome (Ravnan *et al*., 2006). Several kb to Mb region of DNA can be affected by these CNVs that causes contiguous gene syndrome containing altered expression of several genes (Iqbal *et al*., 2014). On the other hand, in some cases haploinsufficiency can describe ID, such as *EHMT1* gene in Kleefstra syndrome (Kleefstra *et al*., 2006), *KANSL1* gene in Koolen - de Varies syndrome (Koolen *et al*., 2012), *LIS1* in the 17p13.3 micro deletion syndrome (Dobyns *et al*., 1993) and *MECP2* gene in the MECP2 duplication syndrome (Bijlsma *et al*., 2012).

1.1.3.3.2.1.2 Monogenic Forms of ID

A gene having a mutation can arrest the function of a particular gene product and this defect can be subcategorized on the basis of type of defect and the involvement of particular pathway. Examples of monogenic defects include metabolic disorders (Karnebeek & Stockler, 2012), mitochondrial defects (Valenti *et al*., 2014) and repeat extension defects (Budworth & McMurray, 2013). These genetic defects cause different disorders like Phenylketonuria, which is a metabolic disorder (Zurfluh *et al*., 2008), Alpers syndrome which is a mitochondrial defect (Naviaux & Nguyen, 2004; Nguyen *et al.,* 2005) and Huntington's disease which is a repeat extension disorder.

1.1.3.3.2.2 Non-genetic Factors

Acquired factors or non-genetic factors can cause ID since birth or in some instances phenotype becomes obvious later in life. These non-genetic factors include noxious vulnerability during pregnancy and obstetric-related problems, nutritional factors and environmental insults (Iqbal *et al*., 2014). The prevalence of ID in the poor and underdeveloped world is two to three folds greater than in Western world (Ropers, 2010). One of the highest risk factors in poor countries is mal-nutrition.

During pregnancy maternal exposure to toxic substances (e.g. use of drugs, environmental chemicals and alcohol abuse), maternal conditions such as phenylketonuria or diabetes and premature birth results in variable effects. Due to the maternal diffusion of infectious diseases like toxoplasmosis, rubella and cytomegalovirus infection causes prenatal and perinatal difficulties such as fetal brain damage, trauma and hypoxia which are the important contributors of ID. The frequency of these various factors depends on maternal life style along with quality which varies greatly among different countries (Chiurazzi *et al*., 2016).

1.1.4 Classification

ID is generally classified either into four grades of severity (mild, moderate, severe and profound) or by mode of inheritance, which can follow autosomal recessive, autosomal dominant and X-linked inheritance patterns.

1.1.4.1 ID Types Based on IQ Level and Severity

ID is classified on the basis of severity into four grades depending on the degree to which an individual is incompetent to meet the requirements of his age group (Matson *et al*., 2005).

1.1.4.1.1 Mild ID

The children with mild ID have the ability to improve communication and social skills, but their motor skills are slightly compromised. In the age of 6-20 years, they can usually read at the level of $4^{th}/5^{th}$ grade. At the age of 21 years and older, they can efficiently develop social skills. Majority of these individuals can work and support themselves but some of them may need assistance in social or financial stress (Katz *et al*., 2008). The IQ level of these individuals is usually in the range of 50-55 to ~70 (American Psychiatric Association, 1987; WHO, 2001).

1.1.4.1.2 Moderate ID

Moderate ID individuals can speak or learn to communicate, but their motor coordination is fair and can learn some social and professional skills. They may be able to support themselves economically under protected conditions but they usually need assistance in stressful periods. The prevalence of moderate ID is 10% and has IQ level in the range of 35-49 (Iqbal *et al*., 2014).

1.1.4.1.3 Severe ID

Severe ID patients have minimal language ability with significant limitations in motor skills. They can have ability to learn or speak to communicate. They can learn health habits and elemental self-care. They characteristically need lifelong guidance and aid in daily activities. They support themselves economically under full supervision. About 5% of ID patients are classified under severe ID with IQ level in the range of 20-34% (Iqbal *et al.,* 2014).

1.1.4.1.4 Profound ID

Patients with profound ID show significant developmental delay with minimum sensorimotor abilities. They want essential care that can last for life span. Teenagers have partial motor and language development. They can acquire poor personal care skills. The 5% of ID patients have profound form of cognitive impairment, having IQ level <20% (Iqbal *et al*., 2014).

1.1.4.2 ID Types Based on Mode of Inheritance

On the basis of physical location of a gene on the particular chromosome (X-chromosome or autosomes) and mode of inheritance, ID is categorize into X-linked ID, autosomal dominant and autosomal recessive forms.

1.1.4.2.1 X-linked Forms of ID

ID is assumed to be most common in males as 8-12% of ID is present in males than in females due to mutations on X-chromosome (Raymond, 2006; Ropers, 2008). There are about 100 X-linked genes that are involved in ID (Ross *et al.*, 2005; Piton *et al*., 2013). Due to typical inheritance pattern X-linked ID can easily be recognized. The X-linked ID probably contributes about 20-25% of ID (Turner, 1996) and is subdivided into syndromic and non-syndromic forms (Roper & Hamel, 2005).

In case of syndromic X-linked ID, patients have one or more defects or clinical features along with ID (Ropers, 2006). About 140 X-linked ID genes have been identified so far that contains 20 non-syndromic X-linked ID genes (Ropers & Hamel, 2005).

1.1.4.2.2 Autosomal Dominant ID (ADID)

ADID is defined as the ID genes located on non-sex chromosome in which disease associated mutation in a single gene copy is sufficient to cause a disease. ADID generally occurs due to *de novo* mutations. The identification of doubted genes involved in ADID is usually carried out through mapping of chromosomal breakpoints in ID patients having inversions or duplications and balanced translocations (Vanagaite, 2007). According to Online Mendelian inheritance in Man (OMIM), 40 gene loci (MRD1-MRD40) are listed that are involved in the development of ADID. ADID includes both syndromic as well as non-syndomic forms depending on the presence or absence of additional clinical features.

1.1.4.2.3 Autosomal Recessive Intellectual Disability (ARID)

ARID is defined as the genes located on non-sex or autosomal chromosomes in which mutation in both gene copies are needed to cause the disease. ARID can be divided into syndromic ID (S-ID), in which more than one physical features or comorbid illness are present and non- syndromic (NS-ID) in which there is no obvious comorbidities. Approximately 700 ID genes are known, which includes both syndromic and non-syndromic genes (Harripaul *et al*., 2017). According to OMIM data, 53 loci and 33 genes are reported for non-syndromic and syndromic ARID. Out of 33 genes only 14 are involved in NS-ARID (Saleha *et al*., 2017). It is estimated that >2500 genes are autosomal ID genes and the majority of them are recessive. In populations having consanguineous marriages, most ID cases are recessive. Even in outbred population, 13-24% of ID is caused by autosomal recessive genes (Musante & Ropers, 2014).

1.1.4.2.3.1 Syndromic Autosomal Recessive Intellectual Disability (S-ARID)

S-ARID is associated with the impairment in nervous system, hearing, vision and skeletal muscles abnormalities (Bokhoven, 2011). Some of these genes are summarized in the subsequent section.

1.1.4.2.3.1.1 *SZT2*

SZT2 is a gene that encodes a protein called "seizure threshold II homolog". This gene is expressed in the brain mainly in the parietal and frontal cortex as well as in dorsal root

ganglia. It is confined to the peroxisome and involved in oxidative stress resistance (https://www.ncbi.nlm.nih.gov/gene/23334). *SZT2* performs its function by increasing the activity of SOD (superoxide dismutase) but itself has no superoxide dismutase activity. Mice studies showed that *SZT2* convenes low seizure threshold and may also has role to enhance epileptogenesis (Frankel *et al*., 2009). Vanagaite *et al*. (2013), found a homozygous transition (c.73C>T) in exon-2 of *SZT2* gene that results in the formation of premature stop codon (p.Arg25*). The other mutation is compound heterozygous for c.2092C>T nonsense mutation and c.1496G>T is a substitution mutation. These mutations cause infantile epileptic syndrome, distinct neuroradiological anomalies, severe developmental delay, intractable seizures and epilepsy (Vanagaite *et al*., 2013). Falcone *et al*. (2013) found 3bp homozygous deletion c.4202_4204delTTC in three children while their parents had heterozygous mutation. Authors concluded that this deletion causes disruption of *SZT2* protein function which might lead to moderate or mild ID without seizures.

1.1.4.2.3.1.2 *POMGNT1*

POMGNT1 is a gene which encodes a protein (enzyme) called "Protein O-linked mannose beta-1, 2-N- acetylglucosaminyltransferase-1, that catalyzes glycoproteins and dystroglycan by transferring N-acetylglucosamine to O-mannose. A mutation in *POMGNT1* gene causes muscle-eye-brain (MEB) disease (Yoshida *et al*., 2001). MEB is an autosomal recessive hereditary disorder which includes muscular dystrophy, optical anomalies and brain deformity. In MEB patients, 13 autonomous disease causing mutations were found in *POMGNT1* (Yoshida *et al*., 2001; Taniguchi *et al*., 2003). These mutations are simple point mutations that cause missense, nonsense, frame-shift, splice-site and pre-mature termination mutation that are dispersed throughout the gene. Mutations near 5' end of coding region causes severe brain symptoms (e.g. hydrocephalus) while in 3' end causes milder phenotypes (Manya *et al*., 2003).

1.1.4.2.3.1.3 *KIAA0586*

KIAA0586 is an autosomal recessive gene and it has five isoforms. The isoform (transcript NM_001244189.1) translates into a protein having 1644 a.a (amino-acids) with one C-terminal protein rich region and four coiled coil domains (predicted by N-coil; Lupas *et al*., 1991). Bachmann-Gagescu *et al*. (2015a), found a compound heterozygous mutations in *KIAA0586* gene; a G to C transversion (c.1413-1G-C) resulting in splice site

mutation and 1bp deletion (c.428delG, NM_001244189.1) that results in frame shift and premature termination. Biallelic mutations of *KIAA0586* are correlated with mild JS (Joubert syndrome) in ~2.5% of families with Joubert syndrome (Bachmann-Gagescu *et al*., 2015a). The other JS patient have frame shift (p.E477Gfs*7) mutation besides synonymous substitution mutation near splice site (last bp of exon 23). This splice site substitution mutation directs cryptic splice site in exon 23 and causes 56bp deletion which leads to frame shift and premature truncation of protein. Another individual with JS contains a homozygous missense mutation (p. D566V) which support pathogenicity with the CADD score of 27.4 (Kircher *et al*., 2014). *KIAA0586* (*TALPID3*) gene codes a centrosomal protein which is involved in ciliogenesis and hedgehog signalling (Yin *et al*., 2009; Bangs *et al*., 2011; Ben *et al*., 2011).

1.1.4.2.3.1.3.1 Joubert Syndrome (JS)

Joubert syndrome (JS) is a neurodevelopmental disorder. It is a rare autosomal recessive disorder that mainly affects the brain and causes mental, physical and visual impairments. In individuals having Joubert syndrome two parts of brain i.e. the brainstem and the cerebellar vermis do not develop properly (Roosing *et al*., 2015). JS is thought to be linked with defects in the maintenance and development of cilia. Cilia are tiny hair like structures found on the surface of cell (Roosing *et al*., 2015). JS (OMIM 213300) was first identified in 1969 by pediatric neurologist Marie Joubert. The neurological features of JS include abnormal eye movements, episodic hyperpnea, ataxia, agenesis of the cerebellar vermis and cognitive impairment (Joubert *et al*., 1968). Within first months of life the typical features of JS includes abnormal respiration pattern, abnormal eye movements, hypotonia and developmental delay (Vilboux *et al*., 2017). Although some JS patients have multisystem organ involvement including liver diseases, fibrocystic kidney retinal dystrophy, polydactyly, chorioretinal colobomas, and occipital encephalocele but others have neurological abnormalities (Parisi *et al*., 2009; Aygun, 2009). Recessive mutations in >28 genes causes Joubert syndrome and all of them encode proteins that are localized to the basal body and primary cilium (Romani *et al*., 2013). The causal mutations in these genes are identified in about 62% of individuals with Joubert syndrome which suggests that other genes may also be involved (Bachmann-Gagescu *et al*., 2015b).

1.1.4.2.3.1.4 *NRXN1*

NRXN1 gene is located on chromosome 2p16.3 and is a component of small group of proteins which was initially known as synaptic transmembrane receptors for α-latrotoxin (the black widow spider toxin) (Ushkaryov *et al*., 1992). *NRXN1* play an important role in synaptic maturation by modifying synaptic properties (Ushkaryov *et al*., 1992), facilitate trans-synaptic interactions (Ichtchenko *et al*., 1995) and helps in regulating synaptic transmission by interacting with neuroligins (Ushkaryov *et al*., 1992). *NRXN1* gene produces two isoforms α- and β- which differs in their extracellular domains but similar in their intracellular and transmembrane structures (Sudhof, 2008). α-*NRXN1* has six LNS (Laminin –G, NRXN, Sex hormone binding globulin) domains, three intercalated EGF (epidermal growth factor) like domains that interact with neurexophilins (Missler *et al*., 1998), LRRTM proteins (Siddiqui *et al*., 2010) and also regulate some calcium channels (Dudanova *et al*., 2006). Whereas β-*NRXN1* lacks EGF-like domains and contains single LNS and fewer laminin G domains (Reissner *et al*., 2008). *NRXN1* go through extensive splicing which is spatially and temporally controlled via calcium/calmodulin- dependent kinase-4 signalling by neuronal activity (Iijima *et al*., 2011).

Intra *NRXN1* deletions have been observed in individuals having wide range of neurodevelopmental disorders (Kim *et al*., 2008; Marshall *et al*., 2008; Rujescu *et al*., 2008; Vrijenhoek *et al*., 2008; Zahir *et al*., 2008; Bucan *et al*., 2009; Glessner *et al*., 2009; Kirov *et al*., 2009; Zweier *et al*., 2009; Ching *et al*., 2010; Kowalnik *et al*., 2010; Pinto *et al*., 2010; Gregor *et al*., 2011; Harrison *et al*., 2011). Schizophrenia and type-1 diabetes has also been identified in family with *NRXN1* deletions (Suckow *et al*., 2008) and this is possible because neurexin-1 is also expressed in pancreatic β-cells (Duong *et al*., 2012).

1.1.4.2.3.2 Non-Syndromic Autosomal Recessive Intellectual Disability (NS-ARID)

The mode of inheritance of autosomal recessive ID is ~1/4th of the total ID genetic cases (Higgins *et al*., 2004). About 34 genes and 51 loci of NS-ARID are identified so far and most of these genes are localized on chromosome 6 or 19. Missense or small indels are the most common genetic defects causing NS-ARID (Khan *et al*., 2016). High ratio of consanguineous marriages in a population have high rate of ARID (Higgins *et al*., 2004; Basel –Vanagaite *et al*., 2007; Molinari *et al*., 2008). Najmabadi *et al.* (2011), has identified 23 known and 50 novel candidate genes causing ARID. The NS-ARID genes

are found to be involved in different downstream processes such as methylation (*NSUN2* and *METTL23*); ion homeostasis (*TUSC3*); protease activity (*PRSS12* and *CRBN*); signal transduction pathway (*LINS, GRIK2, PGAP2, CRADD, ANK3, PGAP1, TRAPPC9* and *TNIK*); tRNA modification enzymes (*ADAT3, NSUN2* and *PUSC3*); transcription factors (*CC2D1A, MED23* and *TAF2*); Ubiquitination (*HERC2* and *FBXO31*); DNA damage response (*TTI2*); mRNA degradation (*EDC3* and *DCPS*); neuro-morphogenesis (*KPTN* and *KIAA1033*)*;* cargo protein (*LMAN2L*); lipid metabolism (*TECR*) and carbohydrate metabolism (*ST3GAL3, MAN1B1, NDST1 and PIGG*) (Khan *et al*., 2016).

PRSS12 (Protease serine 12) also known as neurotrypsin was the first gene reported in NS-ARID. It shows predominant expression in the hippocampus, the cortex, the amygdala, the olfactory bulb, the brainstem and the spinal cord (Yamamura, 1997). This neurotrypsin possibly takes part in neuron plasticity and synapse maturation (Gschwend *et al*., 1997; Wolfer *et al*., 2001; Molinari *et al*., 2002). *PRSS12* gene mutations are involved in moderate and severe ID phenotype (Molinari *et al*., 2002).

Another gene *TRAPPC9* (trafficking protein particle complex 9) (MIM: 611966) is involved in NS-ARID phenotype and protein produced by this gene is NIBP (NIK- and IKK-β-binding protein). This protein directly correlates with NIK (NF-κB-inducing kinase) and IKKβ (IkappaB kinase beta) and involved in the activation of NF-κB pathway. It plays an important role in the nerve cell survival and axonal outgrowth (Hu *et al*., 2005). *CRBN* gene (MIM: 609262) translates into a protein called ATP-dependent Lon protease cereblon that is involved in neuronal surface expression of calcium (Ca^{2+}) regulated potassium (K^{+}) channels and has a role in release of neurotransmitter and neuronal- excitability (Higgins *et al*., 2004; Jo *et al*., 2005). Nonsense mutations in this gene lead to mild form of intellectual disability. *GRIK2* (MIM: 138244) encodes GLuR6 protein which is a part of kainate receptor (KAR), which are iono-tropic glutamate receptors that binds with excitatory neurotransmitter l glutamate. These receptors show high expression in the brain predominantly in the hippocampal mossy fibers (Bortolotto *et al*., 1999; Contractor *et al*., 2001; Motazacker *et al*., 2007).

1.1.5 Next Generation Sequencing (NGS)

The DNA sequencing technique is a broad way of attaining genome information of any living organism. Sanger sequencing has been the universal sequencing technology for decades, in which the fluorescently labelled terminating nucleotides and electrophoresis

was used. The bacterial genome (Haemophilus influenza) was first time completely sequenced in 1995 with the help of Sanger sequencing and formulated a preliminary effect in the arena of microbial genomics (Fleischmann *et al*., 1995). For whole human genome sequencing project, Sanger sequencing was made possible by using various sequencing instruments and automated sample preparation with the cost of US $2.7 billion and more than ten years to compete (Lander *et al*., 2001; Venter *et al*., 2001). Even though Sanger sequencing offered vital improvements in the growth of genomics, but it is too costly and time consuming to sequence entire human genome. This reality provoke the pioneers to develop cheaper, faster and high throughput sequencing techniques that enable to sequence billions of bases in only a matter of hours (Collins *et al*., 2003). The massively parallel high-throughput technologies are also known as next generation sequencing (NGS). The main advantage of a new "massively parallel" or "second generation" sequencing technology is high throughput and low cost per sequenced base. Numerous human genomes can be sequenced in a single run on a second-generation sequencing (SGS) machine within a couple of days (Berglund *et al*., 2011). In fact today, a single NGS system (Illumina's X10; Illumina Inc., San Diego, California) at a cost of ~ $1,000 per genome can sequence 18000 human genomes every year (Gilissen *et al*., 2012).

1.1.5.1 SNP Genotyping

Before single nucleotide polymorphism (SNP) genotyping, microsatellite markers were used as a standard process to genotype genomes. Microsatellite markers are highly polymorphic and used in the form of panels of 300-400 microsatellite markers that are equally distributed throughout the genome (Kennedy *et al*., 2003). In large consanguineous families, the microsatellite marker genotyping has been used to identify runs of homozygosity (ROH) that contained the particular genetic defect, but at the same time the resolution of this method was not enough to study small non-consanguineous families (Collin *et al*., 2011). About ten years ago, a new robust microarray based SNP genotyping and high throughput system was introduced which is highly potent in the detection of CNVs (Rauch *et al.*, 2004) but the present SNP arrays can detect duplications and deletions up to 20kb and 10kb resolution and can also detect ROHs (Iqbal *et al*., 2014).

SNP is a variation at a particular position in the genome and is present in >1% of the population (Brookes, 1999). Each SNP has two different alleles at a genomic position in

such a way that each appears to be present in a substantial portion of the human population and comprises major part of these DNA variants. Generally SNPs do not include deletion or insertion polymorphism but in practice sometimes biallelic variations include deletions, insertions and variations with allelic frequency less than 1%. The two alleles of SNP are often arbitrarily labelled as A and B and hence each individual usually gets single copy of each SNP from each parent and therefore the genotype of an individual is either AA, AB or BB at a SNP site. It is estimated that in human genome there are some 10 million SNPs present (Kruglyak & Nickerson, 2001). The central aim of human genetics is to identify DNA variants that cause disease. A few years ago, scientists have focused on copy number variants (CNVs) which are important contributors to genetic variations in humans (Iafrate *et al*., 2004; Sebat *et al*., 2004). CNVs are chromosomal segments which are at least 1000 base pairs in length and differ only in number of copies from one individual to the other (Feuk *et al*., 2006). Soon after their discovery, numerous high profile studies have associated CNVs with a number of common diseases which includes Alzheimer disease (Lecrux *et al*., 2006), autism (Sebat *et al*., 2007; Weiss *et al.,* 2008), Crohn's disease (Fellermann *et al*., 2006), Parkinson's disease (Sanchez *et al*., 2008), psoriasis (Hollox *et al*., 2008) and schizophrenia (Walsh *et al.,* 2008).

Iqbal *et al*. (2012) come across a case involving patient from a consanguineous family from Pakistan which was sequenced with the help of targeted panel designed for these homozygous regions and found that there was a 5 exon deletion in the *TPO* gene. The deletion could not be detected with SNP genotyping because it is away from the limits of CNV detection algorithms. More than 300 genes of autosomal recessive intellectual disability (ARID) have been identified, mostly by homozygosity mapping using SNP microarray while Sanger sequencing was used for subsequent follow- up of candidate genes. About 97% of these ARID genes are involved in recessive disorders that comprise ID as a major feature. Actually very few recessive genes that cause isolated ID have been identified (Hoeijmakers *et al*., 2011; Musante & Ropers, 2014).

1.1.5.1.1 Advantages and Disadvantages of SNP Genotyping

SNP genotyping is helpful in generating several genotypes easily and rapidly for the assessment of individual ancestry and is also predominantly significant in disease association studies comprising individuals with diverse ancestries (LaFramboise, 2009). SNP genotyping signals may also be used to measure SNP allele frequencies which are

continuously changeable in DNA pooled from hundreds of individuals. This approach is helpful in which a study may investigate data from pooled controls and pooled disease cases in batches to evaluate changes in allele frequencies which may indicate association between disease susceptibility and genetic variants (Butcher *et al.,* 2004). Another unexpected application of genotyping is to assess allele specific expression by genotyping RNA rather than DNA. Recently, both Affymetrix (Gimelbrant *et al*., 2007) and Illumina (Milani *et al*., 2009) platforms have been used to identify genes for which parental specific allelic expression was measured which showed high expression of one parental allele as compared to other. This is done by employing the arrays to genotype transcribed SNPs which are present in heterozygous form in the individual's genomic DNA, afterwards the comparative abundance of each of two SNP alleles are algorithmically assessed in the RNA. These studies disclose large number of genes that indicates difference in allelic expression (LaFramboise, 2009).

The pitfall of SNP genotyping is that genomic regions may contain causal mutations which are poorly encompassed by the SNP array. A family having 4 affected individuals that shared autozygous interval which was unclear and missed by both direct review and by linkage analysis (Zahrani *et al*., 2013). Currently, the molecular basis of (heterogeneous) monogenic disorders is widely explained by exome sequencing. Approximately 1.5% of the genome contains all protein coding genes and within the protein coding regions around 85% of all Medelian disease causing mutations are present which are covered by exome sequencing (Cooper *et al*., 1995; Gregory, 2005; Benitez *et al*., 2011).

1.1.5.2 Whole Exome Sequencing (WES)

WES is a technique in which only 1% part of the genome i.e. the coding parts (exons) or the genetic regions which are decoded into structural or functional proteins are sequenced (Teer *et al*., 2010). In WES technique each exon is sequenced many times to ensure sequence accuracy, therefore this technique is also known as massively parallel sequencing. WES can be used to sequence all exons, or segment of chromosome or a set of genes which may or may not be genomic neighbours but it is also used to discover rare alleles that are involved in complex traits and Mendelian phenotypes but does not assess non-coding alleles. WES has more advantages over Sanger sequencing and can identify mutations that had been skipped by Sanger sequencing. WES is cheaper than Sanger sequencing and at the similar cost of Sanger sequencing of one or two genes, WES may

be performed (Harding & Robertson, 2014). In case of target gene sequencing, only a group of candidate genes responsible for a clinical condition are sequenced which may or may not identify the molecular cause behind the disease (Zhang *et al*., 2012).

Exome sequencing has a great effect on the discovery of ID genes and tackling some of the major issues of gene discovery in this group of diseases. In case of Doors syndrome which is a rare syndrome that includes ID, deafness, seizures, osteodystrophy and onychodystrophy (Campeau *et al.,* 2014). It was difficult to identify the molecular cause of this syndrome by using previous genetic techniques because of its small number and phenotypic variability. International multicentre study recruited 30 families and performed whole exome sequencing on 17 and found recessive mutation in *TBC1D24* gene in only 7 families. This gene was then sequenced in 9 additional families as a validation cohort by using Sanger sequencing and found *TBC1D24* mutations in two more families (Campeau *et al.,* 2014).

The other case solved by exome sequencing is the sporadic and non-syndromic cases of ID. The genetic basis of these cases is extremely challenging because it is difficult to presume that a specific group of patients having NS-ID and may have mutations in the same or different genes. It is possible that *de novo* mutations are also involved in many of the ID cases, as humans have a quite high mutation rate per generation (Roach *et al*., 2010; Lynch, 2010) and ~15% of ID cases are CNVs (Varies *et al*., 2005; Cook & Scherer, 2008). Vissers *et al*. (2010) recruited 10 affected children to perform exome sequencing and their normal parents, to find splice site or exonic variants present in child but absent in both parent. They found <2 *de novo* exonic variants in each individual and these variants were further filtered on the basis of their expected effect on protein function and evolutionary constraint that results in identification of causative mutations in 6 different genes in 6 patients. *RAB39B* and *SYNGAP1* are the two of those genes implicated in ID (Hamdan *et al*., 2009; Giannandrea *et al*., 2010) and other four genes (*YY1*, *DYNC1H1*, *CIC* and *DEAF1*) may have an important role in proper development of nervous system evidenced through network studies and mouse models (Topper *et al.*, 2011).

In familial cases of ID, many families have multiple affected children in which inheritance pattern of disease is both X-linked and autosomal recessive. In exome sequencing, exome sequences in combination with pedigree data are utilized to detect mutations in autosomal forms of ID (Topper *et al.,* 2011). Caliskan *et al*. (2011)

examined large consanguineous family in which 5 out of 13 children had a NS-ID which includes poor language, developmental delay and fine motor skills. Linkage studies were conducted and identified gene rich region on chromosome 19p13. Afterwards, high resolution SNP genotyping was used to identify 2 Mb (mega base) region in all affected individuals having over 30 homozygous genes. As only affected individuals shared homozygous region, so it is definitely a fully penetrant recessive condition. Investigators chose parents and performed exome sequencing for the analysis of disruptive, novel, heterozygous variants in the 2 Mb region that were absent in Single Nucleotide Polymorphism database (dbSNP). These filters promptly results in the identification of single variant (missense mutation) in a highly conserved residue of the *TECR* gene. *TECR* is a synaptic glycoprotein highly expressed in the brain and is involved in biosynthesis of long chain fatty acids (Topper *et al.*, 2011).

1.1.5.2.1 Benefits and Challenges of Exome Sequencing

The major benefit of exome sequencing is to sequence almost all the coding content of the genome which is expected to have most of the disease-causing mutations, in a single experiment. Exome sequencing also play an important role in identifying new genes that are involved in unpredicted pathways or cellular processes. Most significantly, the sample size needed for exome sequencing is reduced to a few phenotypically associated individuals, a single extended family or a single child and parent. This efficiency of the exome sequencing approach will have the most thoughtful impact on the human genome sequencing field (Topper *et al*., 2011).

Despite of many benefits of exome sequencing for detecting mutations for rare diseases, there are certain challenges and flaws in this sequencing platform. There are several reasons behind the incomplete exome data generation by the current technologies. The first reason is the coverage depth across the exomes is not uniform and some regions are usually neglected, this occurs due to biases in the capture, sequencing and alignment processes. The capture process is biased due to variations in secondary structure and GC content. Normally, exact matches to a probe sequences are more expected to be captured, therefore heterozygous variants, particularly true for genomic DNA fragments having indels, can represent significantly less than half of the captured DNA from a locus. The other reason is, sequence alignment can introduce biases and it can operate poorly in disambiguating sequences amongst regions with repeated sequences or highly homologous sequences. In addition to this, variant calling can further bias by mapping

exome sequences to a single reference sequence since reads with variants are often assigned lower mapping quality scores, which can eventually ignore those variants from the following analysis. Most of these complications are solved by using longer capture probes and longer, paired- end sequence reads. Therefore, many of these shortcomings will be less challenging and less problematic as the technology improves. The interpretation of exome sequencing is the bigger challenge. It produces a huge amount of data which includes sequence information of millions of bases and tens of thousands of variants per individual. As the technology become fast, the prompt data generation aspects become routine, the analytical strategies will be streamlined for quick classification of pertinent pathogenic versus benign variants. This should be possible only if more and more exome sequences of diseased and healthy individuals become available (Topper *et al*., 2011).

1.1.6 Aims and Objectives

- To recruit the consanguineous families affected with intellectual disability.
- To perform the targeted Sanger sequencing and data analysis.
- To identify disease causing mutations in the patients with ID.

2. Materials and Methods

2.1 Ethical Approval and Subjects

The current research work was approved by the Institutional review board and bioethics committee (BEC) of Quaid-i-Azam University Islamabad. For this study, two consanguineous families with Intellectual Disability (ID) were recruited from the remote areas of Pakistan.

Each family was collected by visiting the respective area and the detailed information was collected from the healthy individuals. The collected information was partly represented in the form of pedigrees, which show graphical presentation of family's genetic relationships (Resta, 1993; Bennett, 1999). Both pedigrees were drawn by using the standard instructions of Bennett *et al.* (2008), by using the Haplopainter 1.043 software (http://www.mybiosoftware.com/haplopainter-1-043-pedigree-haploytype-drawingtool.html). In pedigree, square indicates male individual whereas circle indicates female. The filled circles and squares denote the ID patients whereas healthy individuals are denoted by hollow circles and squares. The diagonal line over circles and squares showed deceased individuals and consanguinity was represented by double line between spouses. Generations are denoted with Roman numerals while individuals with in generations are denoted with Arabic numerals.

2.2 Clinical Evaluation of Affected Patients

The detailed evaluation of each ID individual was done by visiting their home residences in the Punjab and Sindh provinces of Pakistan. Each affected individual was assessed for the presence of neurological, dermatological, morphological, behavioural and skeletal abnormalities. The severity of ID in affected individuals was assessed by asking questions in their local language and the result was reflected in the form of IQ score. Prenatal, perinatal and postnatal history of each affected individual was collected, along with weight, height and head circumference (HC) measurement to determine their development trends.

2.3 Sample Collection

Peripheral blood samples were collected from affected and normal individuals of each family by using 10mL sterile syringes (Becton Dickinson, UK) with written consent. The blood samples were then preserved in ethylene di-amine tetra acetate (EDTA) vacutainer tubes (red top) (Becton, Dickson-Plymouth, UK) and were stored at $4^{\circ}C$ in Genomics lab at Quaid-i-Azam University, Islamabad for genomic DNA isolation.

2.4 Genomic DNA Extraction

The gDNA (Genomic DNA) extraction from peripheral blood was carried out by using phenol-chloroform (organic method) method (Sambrook & Russell, 2006). The brief description of the protocol is as follows but the composition of different solutions used for DNA extraction are mentioned in the table 2.1.

- The Vacutainers® (BD-plymouth, UK) containing blood samples were incubated at 25 ˚C for about 30 minutes.
- After mixing, 750 µl blood was taken in each properly labelled autoclaved Eppendorf tube (Extra Gene, USA) and equal volume (750 µl) of Solution A was added (Table 2.1).
- Tubes were inverted 4-7 times to facilitate thorough mixing and then kept at room temperature for approximately 15-20 minutes.
- Then centrifugation was performed at 13,000 revolutions per minute (rpm) for 1 minute in a microcentrifuge (Hettich Mikro120, Germany).
- After centrifugation, the supernatant was discarded and 400µl of solution A was added to dissolve the nuclear pellet.
- Once again, centrifugation was performed at 12,000 rpm for 1 minute.
- The supernatant obtained as a result of centrifugation was discarded and the nuclear pellet was re-dissolved in 400µl solution B, 12µl of 20% SDS (Invitrogen, Japan) and 8µl of proteinase K (Thermo Fisher Scientific, USA).
- The pellet was dissolved and samples were incubated at 37˚C for about 12 hours in an incubator (Binder B28, Germany).
- On next day, 500µl of freshly prepared mixture containing equal volume of Solution C+ Solution D (Table 2.1) was added to the samples and tubes were inverted several times to ensure mixing and then centrifuged at 12,000 rpm for 10 minutes to separate the aqueous and organic layer.
- After centrifugation, the aqueous (upper) layer was collected in a new autoclaved Eppendrof tube and the rest of fraction was discarded.
- In the aqueous layer, 500µl of Solution D was added and centrifuged at 12,000 rpm for 10 minutes.
- After centrifugation, the upper (aqueous) layer was again transferred to new autoclaved Eppendrof tube and the DNA was precipitated by the addition of 55µl sodium acetate (3M, pH 6) and 500µl chilled Iso-propanol.

- To precipitate DNA, tubes were inverted several times and then centrifuged at 12,000 rpm for 10 minutes to precipitate DNA at the bottom of the tube.
- The supernatant was discarded carefully and 200µl of 70% ethanol was added to wash the DNA pellet and then centrifuged at 8,000 rpm for 7 minutes.
- After centrifugation, the ethanol was discarded and the pellet was dried in the incubator at 37°C for 20-30 minutes.
- Then the dried gDNA pellet was dissolved in a suitable volume (80-150µl) of T.E (Tris EDTA) buffer (Table 2.1) and incubated at 37°C for overnight. Afterwards, samples were stored at 4°C.
- The quality of DNA band was checked on 1% agarose gel before checking the DNA concentration.

2.4.1 DNA Quantification

For DNA quantification, Colibri Microvolume Spectrometer (Titertek Berthold, Germany) was used. Initially, 1-2µl T.E was used as a blank and then pipetted an aliquot of l-2 µl DNA sample to measure the DNA concentration of each sample. After quantification, DNA dilutions were prepared to get the concentration of 20ng/µl for each sample.

Table 2.1: List of Solutions used for DNA extraction along with their composition

SR No.	Solution	Composition
1.	**Solution A**	5mM MgCl2
		0.32M Sucrose
		10mM Tris HCl (pH 7.5)
		1% Volume/Volume Triton X-100
2.	**Solution B**	10mM Tris-HCl (pH 7.5)
		400mM NaCl
		2mM EDTA (pH 8.0)
3.	**Solution C**	Saturated Phenol
4.	**Solution D**	Chloroform (24ml): Iso-amyl alcohol (1ml)
5.	**10X TBE**	0.025M EDTA (pH 8.3)
		0.89M Tris
		0.89M Boric acid
6.	**20% SDS**	5g Sodium dodecyl sulphate in 25ml H_2O
7.	**T.E (Tris EDTA Buffer)**	1mM EDTA
		10 mM Tris (pH 8.0)
		0.5M EDTA (pH 8.0)
		1M Tris (pH 7.5)

2.5 Genetic Analysis of ID Families

In order to find out the ID causing gene in both families (A & B) whole genome genotyping was performed through international collaborations. The details are given below;

2.5.1 Genome Wide Genotyping

Three members of family A and all available members of family B were genotyped through Infinium Human Core Exome Beadchip (Illumina, USA). This chip contains ~550,000 single nucleotide polymorphism (SNPs). Genotyping was peformed in the University of Washington, USA, but the genotyping data was analysed in Pakistan by using Homozygosity Mapper (Seelow *et al*., 2012) (http://www.homozygositymapper.org/). As the inheritance pattern of both families (A, B) is recessive, so the homozygous by descent (HBD) regions shared by the affected members of each family were considered as a potential regions which may contain disease causing genes. The flanking SNPs of each identified HBD region were carefully noted and the regions were further analyzed through UCSC genome browser to find genes present in the region.

2.5.1.1 Family A

The genotype data of this family was also analysed to detect copy number variations (CNVs). As a simple procedure, we looked for the presence of continuous regions with the SNP no calls in the affected individuals. These regions were carefully identified and subjected to further validations when two or more SNPs were not called in the affected individuals.

When a SNP no call region was found and shared by affected members, it was further tested through PCR amplification and sequencing. For PCR amplification primers were designed by importing the deleted sequences along with flanking 1000bp from Ensemble Genome Browser (http://asia.ensemble.org/index.html). Three primer sets (Table 2.2) were designed; first set was designed from the upstream region of the probable deleted region, second set was designed within deleted region and third primer set was designed from the downstream of the deleted region (Fig 2.1). The primers were designed with the online primer designing tool Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/). To confirm single hit and primer specificity BLAT (Blast like alignment tool) (https://genome.ucsc.edu/cgi-bin/hgBlat) was used which is available on UCSC genome browser. Product size and melting temperature of primer was confirmed through *Insilico*

PCR (https://genome.ucsc.edu/cgi-bin/hgPcr). Primers were received in the lyophilised form from Gene Link ^{TM}, USA, and were re-suspended in T.E buffer, before making dilutions of 10µM concentration.

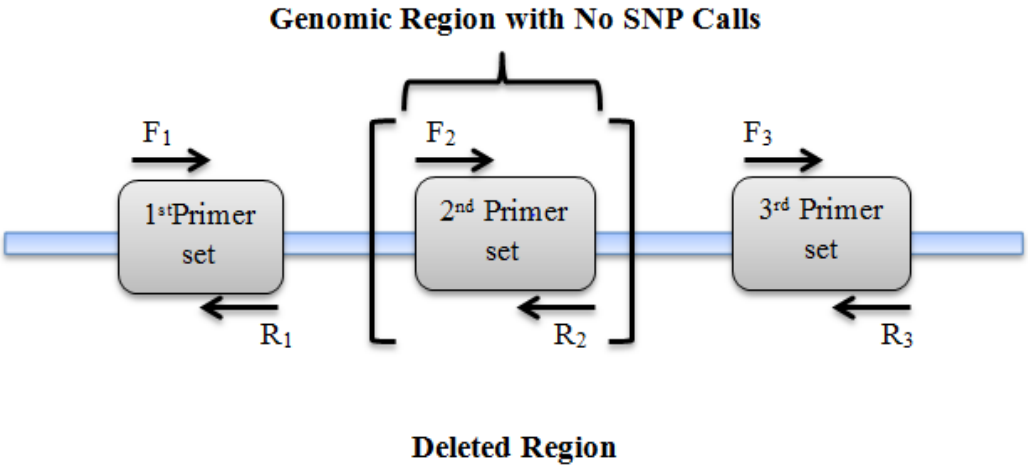

Figure 2.1 Graphical presentation of primer designed to test the probable deletion detected in the family A. First primer set (F_1R_1) and third primer set (F_3R_3) were designed before and after probable deletion while 2nd primer set was designed within the deleted region.

2.5.2 Whole Exome Sequencing (WES)

WES was performed on one affected member (IV-4) of family B on Illumina HiSeq in the University of Washington, USA. Exome data was generated in the form of Variant Call File (VCF) and Binary Alignment/Map (BAM) file. Later various parameters were applied on VCF file to find out the ID causing variants.

Initially, the BAM file of individual IV-4 was analysed to check the coverage and read depth by using Golden Helix Genome Browse 2.1.2 (http://goldenhelix.com/products/GenomeBrowse/). In this browser, sequencing reads obtained from ID patient were piled up against GRCH37 reference genome to check the read depth and coverage of the targeted genomic region, especially the homozygous regions identified in the family B. Uncovered regions and their genomic coordinates were identified (Table 2.3), and were used to design primers to Sanger sequence the uncovered regions detected in individual IV-4.

To rule out the involvement of uncovered regions, the uncovered exons of three genes present in these regions were amplified and sequenced. Primers for *KIAA0586* (NM_001244189.1), *POMGNT1* (NM_001243766.1) and *SZT2* (NM_015284.3) genes were designed by importing the sequences of uncovered regions from Ensemble Genome Browser (http://asia.ensemble.org/index.html) and rest of primer designing procedure was same as described in family A. Each uncovered exon present within the HBD region was amplified by Sanger sequencing.

2.6 PCR Amplification

PCR amplification was carried out by using primers listed in table 2.3 and the DNA sample from the affected individual (IV-4) of family B. These reactions were carried out in PCR tubes (200µL, Axygen, USA) after adding the following reagents:

- 10X Taq Buffer with ammonium sulphate 2.5µL
- (25mM) $MgCl_2$ 1.5µL
- (10mM) dNTPs 0.5µL
- DNA sample (20ng/µl) 1µL
- (0.4µM) Forward Primer 1µL
- (0.4µM) Reverse Primer 1µL
- (5U/µl) Taq DNA polymerase 0.2µL
- Ultrapure DNase free PCR water 17.3µL

The tubes containing reaction mixture were then vortexed briefly and centrifuged for 30 seconds at 8000 rpm. The PCR tubes were placed in a T1 Thermo-cycler (Biometra, Germany) and thermal cycling conditions were set as follows:

Steps	**Temperature & Time**	
First denaturation of template DNA	95°C for 10 minutes	
2nd denaturation	95°C for 1 minute	cycles (2:39)
Primer annealing	58-60 °C for 55 seconds	
Primer extension	72°C for 50 seconds	
Final extension	72°C for 10 minutes	
Paused	4°C	

2.6.1 PCR Amplification Confirmation

The amplified PCR product was visualized on 2% agarose gel. For this purpose, 2% agarose gel was prepared by adding 1.6g of agarose powder in 80ml of 1X Tris Boric EDTA (TBE) buffer (Table 2.1) in a 100ml conical flask. The mixture (1X TBE and agarose) in a conical flask was heated in a microwave for 2 minutes and in the meantime the gel casting tray was set with comb and spacers. The heated mixture was allowed to cool and then 8µL ethidium bromide (EtBr) dye (4µL EtBr in case of 40ml of 1% gel) was added and mixed. The mixture was then poured in the gel casting tray and left to solidify at room temperature for about 20-25 minutes. Once the gel solidified, the comb and spacers were removed carefully without distorting wells. The gel was then placed in a gel running tank (Biometra, Germany) which also contains 1xTBE running buffer. After that, 3µL of amplified PCR product along with 3µL of loading dye was loaded in each well. In order to confirm the size of PCR amplified product, 3µL of Gene Ruler 100 bp (ThermoFisher Scientific, USA) was also loaded in the well. The voltage (110 V) for about 15-25 minutes was provided to resolve the amplified product and gel image was then visualized by using gel documentation system (SYNGENE, UK). The specific PCR products were selected for purification and Sanger sequencing.

2.7 Purification of Amplified Products by ExoSAP-IT

ExoSAP-IT (Affymetrix, USA) is an exonuclease-1(removes left over primers) shrimp alkaline phosphatase (removes any remaining dNTPs) which was used to purify PCR products before proceeding for DNA sequencing. This was done by taking 1µl ExoSAP reagent and 5µl PCR product in a PCR tube followed by brief vortex and short spin.

These tubes were given incubation cycles, first at 37°C for 15 minutes and second at 80°C for 15 minutes in T1 Thermocycler to inactivate the ExoSAP reagent.

2.8 Sequencing Reaction

After purification of PCR products, Sequencing reaction was performed, with the Big Dye Terminator Cycle Sequencing kit (ThermoFisher Scientific, USA). The Big Dye terminator cycle sequencing kit utilizes four fluorescently labelled di-deoxynucleotides (ddNTPs) which results in the single stranded products of various lengths that are separated on the basis of their size. The first step includes cycle sequencing of purified PCR product which involves the following;

- PCR amplified DNA template 2.5-95 ng/µL
- Primer (F or R) 3.2µM
- Big dye terminator ready reaction mix 4µL
- ddH_2O (DNase & RNase free distilled H_2O) was added to make final volume up to10 or 20µL depending on volume of other reagents.

Reagents in the microamp tubes were mixed by brief vortexing and tubes were placed in the thermo-cycler by setting the thermal cycling conditions as:

Initial denaturation	96°C for 1min	
Denaturation	96°C for 10 sec	25cycles
Annealing	50°C for 5 sec	
Final Extension	60°C for 4 mins	
Hold	4°C	

After thermal cycling, tubes were centrifuged and the sequencing reaction was then purified with ethanol/EDTA (60µl of absolute ethanol and 5µl of 125mM EDTA). The mixture was thoroughly mixed and centrifuged at 1000 x g for 10 seconds and then incubated for 15 minutes at room temperature. After incubation, samples were centrifuged (4°C) at 1870 x g for 45 minutes. The supernatant produced after centrifugation was discarded and in the pellet 60µl of 70% ethanol was added. Again centrifugation was performed at 4°C for 15 minutes and supernatant was discarded while samples were left at room temperature for 15 minutes. After purification, sequencing was performed on ABI 310 sequencer (Applied Bio-systems, USA).

2.8.1 Sequenced Data Analysis

The bioinformatics tool CLC Genomics Workbench v3.6.5 was used to analyse the sequence data and the reference sequence was taken from the Ensemble genome browser (http://www.ensemble.org/index.html). The sequenced data of each exon was then aligned to its respective reference sequence to identifying the potential variant. The observed variants were then checked on SIFT (http://sift.jcvi.org/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), Huaman Splicing Finder (http://www.umd.be/HSF3/HSF.html), and Mutation Taster (http://www.mutationtaster.org/) to predict their pathogenic nature.

The 1000 Genomes Browser (http://browser.1000genomes.org/index.html), ExAC (Exome aggregation Consortium) Browser (http://exac.broadinstitute.org/) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) were used to explore the nature of variant identified in this study.

Table 2.2: Sequence of primers used to amplify and sequence a genomic region with probable deletion in family A.

Sr. No.	Gene-Exon		Primer Sequence	Annealing Temp.	Product Size
1	**Chr2:50227704_ 50228124**	F	CTCTGATTGGTGGGTCCTCTA	59.15	421bp
		R	GCAAGTGTTCAGAGGTGGAC	58.28	
2	**NRXN1 Exon16**	F	CTACCACATTAAAGGGGAGCAG	60.01	676bp
		R	GAGGGAATTGTGGAAACAGG	59.38	
3	**Chr2: 50786947-50787490**	F	TGCCTGTTCTGAGTTTGAGAAG	59.66	544bp
		R	CAGATGACCCTTGCCTTGTAG	59.74	

Table 2.3: Sequence of primers used to amplify and sequence the uncovered exons of genes present in the HBD region identified in family B.

Sr.#	Gene-Exon		Primer Sequence	Annealing Temp.	Product size
1	**KIAA0586 Exon-6**	F	TGTGAAGTGATGGGTATGCTAA	58.4	556bp
		R	AGACAAGCTGCTGTGGTTTTT	57.4	
2	**KIAA0586 Exon-33**	F	TGCCTAGTCAGTCAACCAAATC	60.3	500bp
		R	ACTGGAAGCTATCATCCTTTGC	60.3	
3	**POMGNT1 Exon-23**	F	GAGTTTGAGACTCAGGCATCC	61.3	584bp
		R	AACACCCCGTTCCTGAGTTA	58.4	
4	**SZT2 Exon-5**	F	GCTGAGGACTGATGGAGATTG	56.3	491bp
		R	GGGGAGTATATCTGAGTCCTGCT	58.6	
5	**SZT2 Exon-6**	F	AGGGAGGAAGCACAGTCAAG	58.2	488bp
		R	GGTTGTCAGAGTCGAGGGTTC	58.9	
6	**SZT2 Exon-7**	F	GGTAAGGGAGAAGGAAAGGTG	56	443bp
		R	AGTCCTCTCCCACGATGCTAC	59.8	
7	**SZT2 Exon-8,9**	F	GAAGGGCTTAAGGAAGGAAAGA	55.1	901bp
		R	CAATTCAAAACCCACTGTCTCA	53.7	
8	**SZT2 Exon-10,11**	F	CCTGAAACCTGTGGAACCAT	58.4	829bp
		R	GGTTAAAGCTTGGGCTTCTG	58.4	
9	**SZT2 Exon-12-13**	F	CAGTTTTTGCCTGTGCTTCA	54.7	839bp
		R	GGGGAGAAATGGGAGTCAGT	57.8	
10	**SZT2 Exon-14,15**	F	ACTCCTCCCTGACCTCTCTGCT	63.2	938bp
		R	CCACACCTCACCCCGCTAAG	62.4	

3. Results

3.1 Family Description

Two consanguineous families (A& B) were collected from the remote areas of Pakistan.

3.1.1 Family A

The Family A consists of five generation, but eight affected individuals (II-5, IV-3, IV-5, IV-8, V-1, V-2, V-4 & V-6) are present in generation II, IV, and V. Seven affected individuals are male which probably indicates the X-linked inheritance of ID in this family (Fig 3.1). This was also supported by the presence of an affected female (V-6) whom father was intellectually disabled. However, detailed analysis of the pedigree did not reveal any relationship between female individuals II-2 and II-4, which may indicate the autosomal recessive inheritance of ID in this family.

Affected individuals of this family have non-syndromic, mild ID without any history of fits. They have short term memory and cannot recognize home and other relatives. They do not fight and have non-aggressive behaviour. They are unable to speak, unable to take names of simple things and have mild IQ level. Other details are summarized in table 3.1.

3.1.2 Family B

The pedigree of family B shows four generations, and the three affected individuals (IV-1, IV-2 & IV-4) were present in 4^{th} generation. The mode of inheritance of ID in this family is autosomal recessive as unaffected parents have affected children (Fig 3.2).

The affected members of this family have normal hearing and vision, can recognise home and relatives, and have moderate speech development. All three ID patients show aggressive behaviour and were capable to speak, say few words and short sentences. Additional details are summarized in table 3.2.

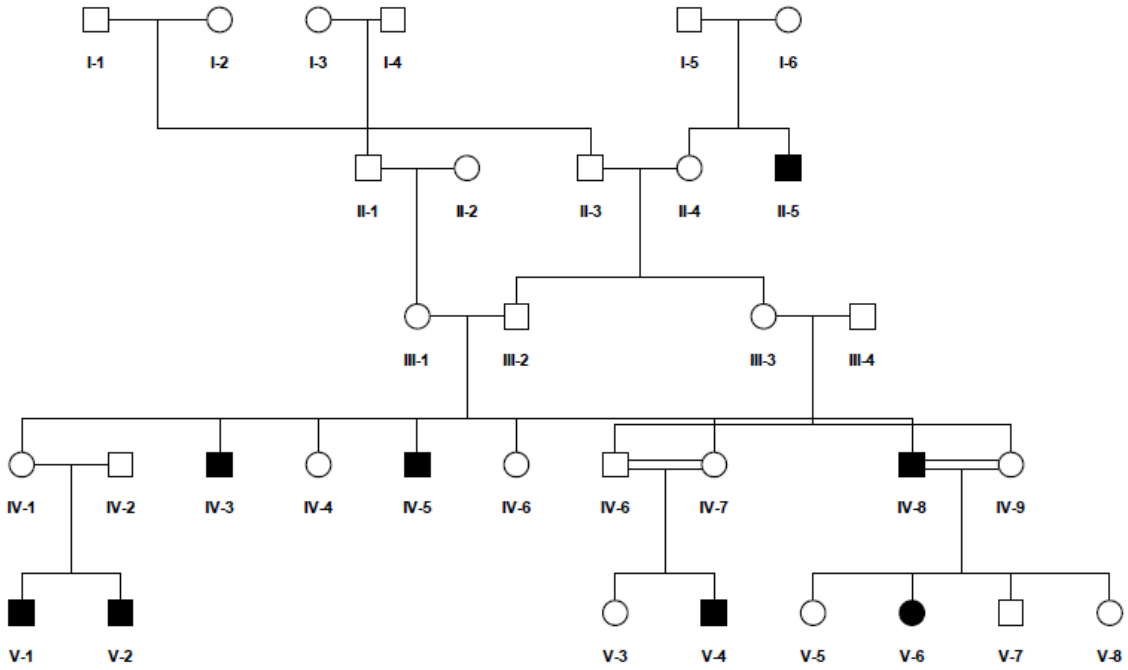

Figure 3.1: Pedigree of family (A) with five generations and eight affected individuals.

Table 3.1: Clinical features of affected individuals of family A.

	Affected Individuals				
Features	IV-3	IV-5	IV-8	V-4	V-6
Intellectual Disability	Mild	Mild	Mild	Mild	Mild
Aggressive Behaviour	No	No	No	No	No
Language	Can't Speak	Can't Speak	Can't Speak	Can't Speak	Can't Speak
Reproductive organs development	Large testis	Large testis	Large testis	Large testis	No
Fits History	No	No	No	No	No
Concept of many	No	No	No	No	No
Self-Care Concept	No	No	No	No	No
Home recognition	No	No	No	No	No
Learning Disability	Severe	Severe	Severe	Severe	Severe
Movement Coordination	Normal	Normal	Problem with movement Coordination	Normal	Normal
Memory	Short term memory	Short term memory	Short term memory	Short term memory	Short term memory
Head Circumference	52cm	53cm	NA	54cm	47cm
Gender	Male	Male	Male	Male	Female

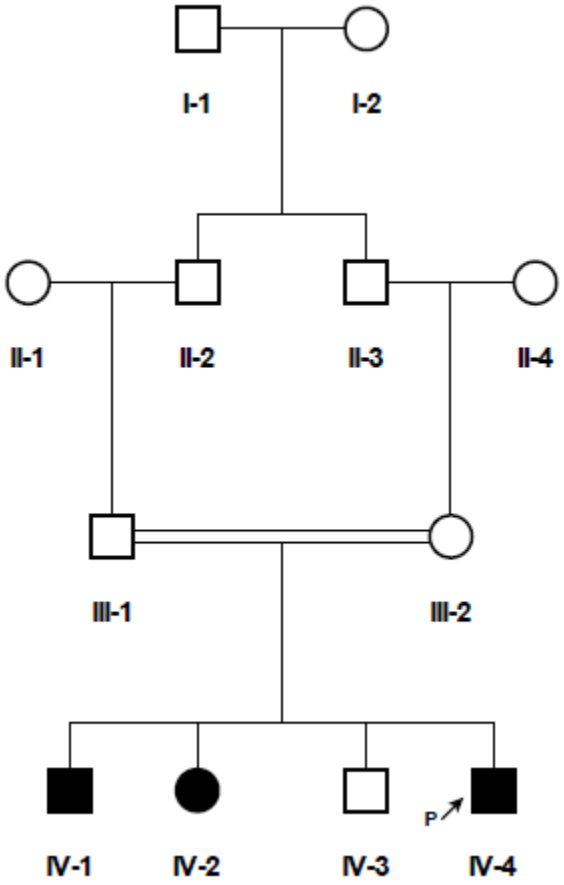

Figure 3.2: Pedigree of four generation family (B) with three affected individual. The arrow under the square indicates proband (P).

Table 3.2: Clinical features of affected individuals of family B.

	Affected Individuals		
Features	IV-1	IV-2	IV-4
Age	16 years	21 years	19 years
Head Circumference	54cm	53cm	51.5cm
Self-care concept	No	No	No
Fits History	No	No	No
Aggressive behaviour	Yes	Yes	Yes
Quantity Concept	No	No	No
Home recognition	Yes	Yes	Yes
Relatives recognition	Yes	Yes	Yes
Speech Development	To some extent	To some extent	To some extent
Hearing & Vision	Normal	Normal	Normal
Ambulation delay	Yes	Yes	Yes
Repetitive behaviour	No	Yes	No
Gender	Male	Female	Male

3.2 Genetic Analysis

The genomic applications like whole genome genotyping & exome sequencing were used to identify the ID causing variants in the affected individuals of families A and B.

3.2.1 Whole genome genotyping and Homozygosity mapping

3.2.1.1 Family A

The mode of inheritance of ID in the affected members of family A is not clear and may indicates the possibilities of either autosomal recessive or X-linked ID. As a result, three affected members of this family were subjected for genome wide scan rather than X-chromosome analysis. In family A, genome wide genotyping was performed by using DNA samples of three affected members (IV-8, V-4 &V-6). Analysis of the genotype data with the homozygosity mapper tool did not identify a homozygous region. Subsequently, the genotype data was manually accessed in excel to find continuous SNP no call regions, which can be indicative of the genomic deletions. These analyses identified a genomic region with multiple SNP no calls on chromosome 2. This region was flanked by rs6706472 and rs12713108 and spans a genomic region of ~559kb (chr2:50,228,332-50,786,770). In order to confirm this deletion three sets of primers were used; one set of primer was used to amplify genomic region before deletion, second primer set was used to amplify genomic region within deletion and third primer set was used to amplify the genomic region after deletion. The PCR products were then run on 2% agarose gel, the gel images (Fig 3.3 & 3.4) were visualized but these results did not confirm the presence of deletion in the affected members of this family.

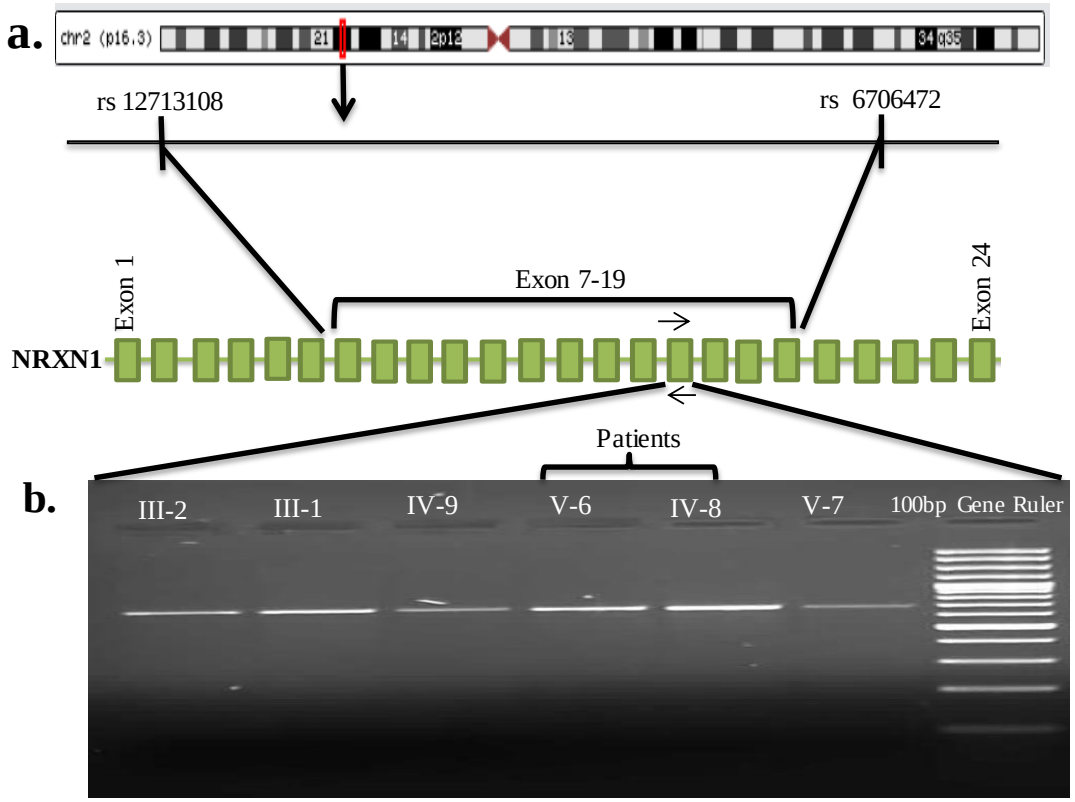

Figure 3.3: (a) Schematic presentation of SNP no call region (rs6706472 & rs12713108) which is present on chr2: p16.3 and found during genome wide genotyping of family A. At this position *NRXN1* gene (exon 7-19) was present which was identified as a probable deleted region. **(b)** The gel image of PCR amplified product obtained from exon 16, to confirm or rule out the deletion of respective genomic region. However, gel image did not show the presence of genomic deletion in the affected individuals of family A.

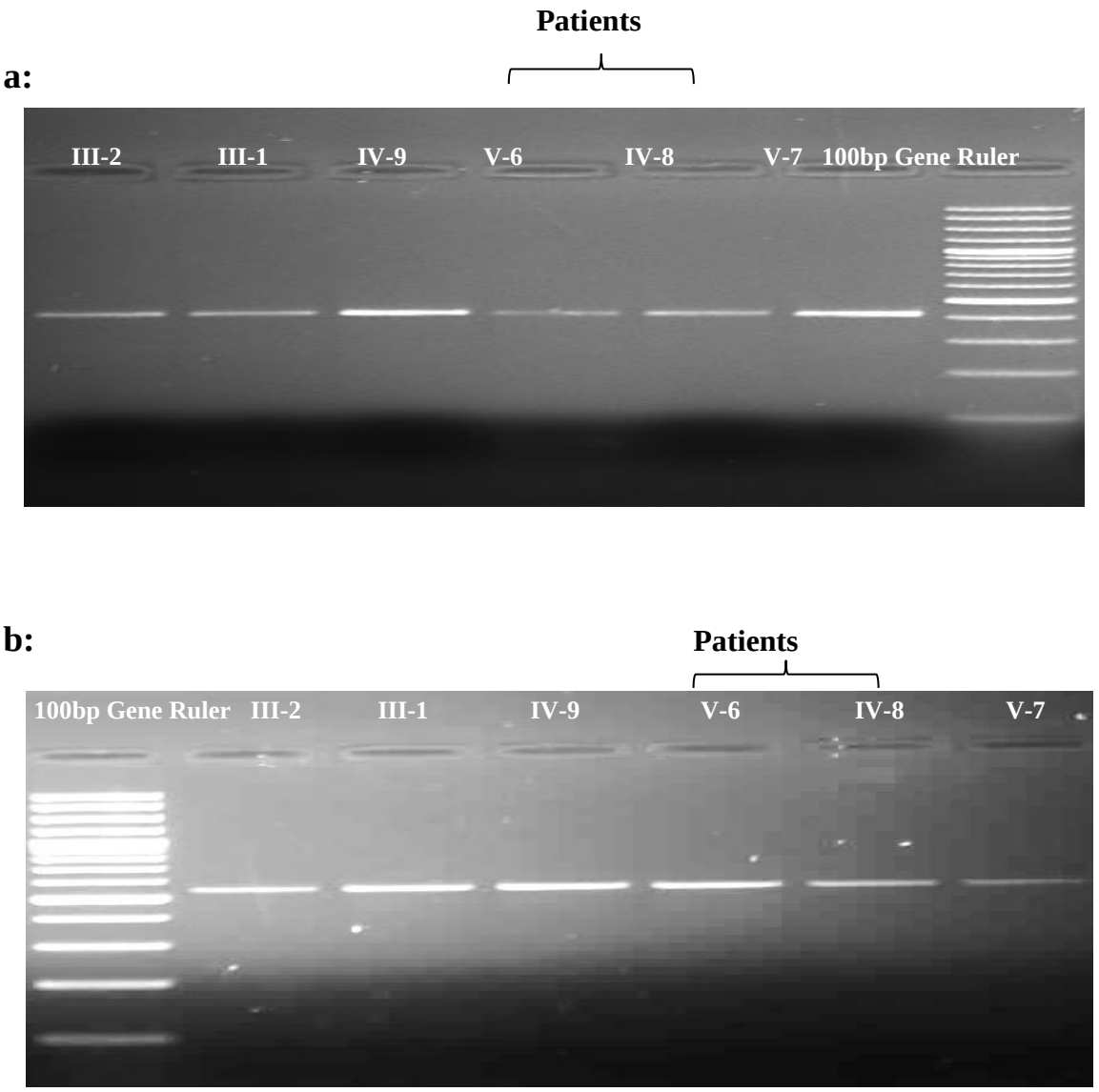

Figure 3.4: (a) Gel image show PCR amplified product of 421bp which was amplified from the genomic region before deletion. **(b)** This shows the PCR amplified product of 544bp that was amplified from the genomic region after deletion.

3.2.1.2 Family B

The mode of inheritance of family B is autosomal recessive therefore all available members of this family were subjected for genome wide genotyping. After genome wide genotyping, the genotyping data was analysed by using Homozygosity mapper tool, four homozygous by descent regions (HBD) were observed (Fig 3.5). Each genomic region was further explored and the genomic coordinates of each genomic region was noted for further analysis of candidate genes. The identified HBD regions are summarized in table 3.3.

3.2.1.2.1 Exome Sequencing in Family B

Exome sequencing was performed on IV-4 member of family B. During BAM file analysis some uncovered regions were observed in two out of four HBD regions. In case of chromosome 1: 38401933-64109264 HBD region two genes *POMGNT1* (exon-23) and *SZT2* (exon-5-15) were uncovered while on chromosome 14:56307478-75947712 HBD region *KIAA0586* (exon-6 & 33) was uncovered (Table 3.4).

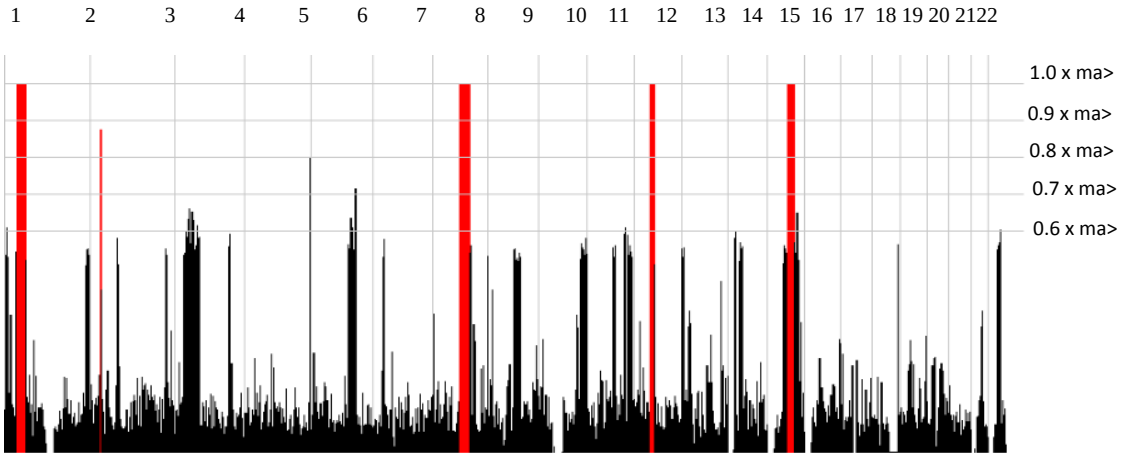

Figure 3.5: Homozygosity Mapper image of HBD regions identified in family B.

Table 3.3: List of homozygous by descent (HBD) regions identified by Homozygosity Mapper.

Score	Chr.	From (bp)	To (bp)	rs	rs
500	1	38401933	64109264	12739026	2269240
500	7	77602832	106391804	848458	7803053
500	11	45540603	57901176	1873059	1868881
500	14	56307478	75947712	34348798	6574237

Table 3.4 Genomics coordinates of uncovered regions identified in individual IV-4 of family B.

Chromosome	Homozygous region mapped	Gene Name	Exon#	Coordinates	Coverage in WES
1	38401933-64109264 (2.5 Mb)	*POMGNT1*	23	46,654,381-46,654,651	Partial covered
		SZT2	5-15	43,880,751-43,889,001	No coverage
14	56307478-75947712	*KIAA0586*	6	58,907,938-58,908,061	No coverage
			33	59,010,601-59,010,685	No coverage

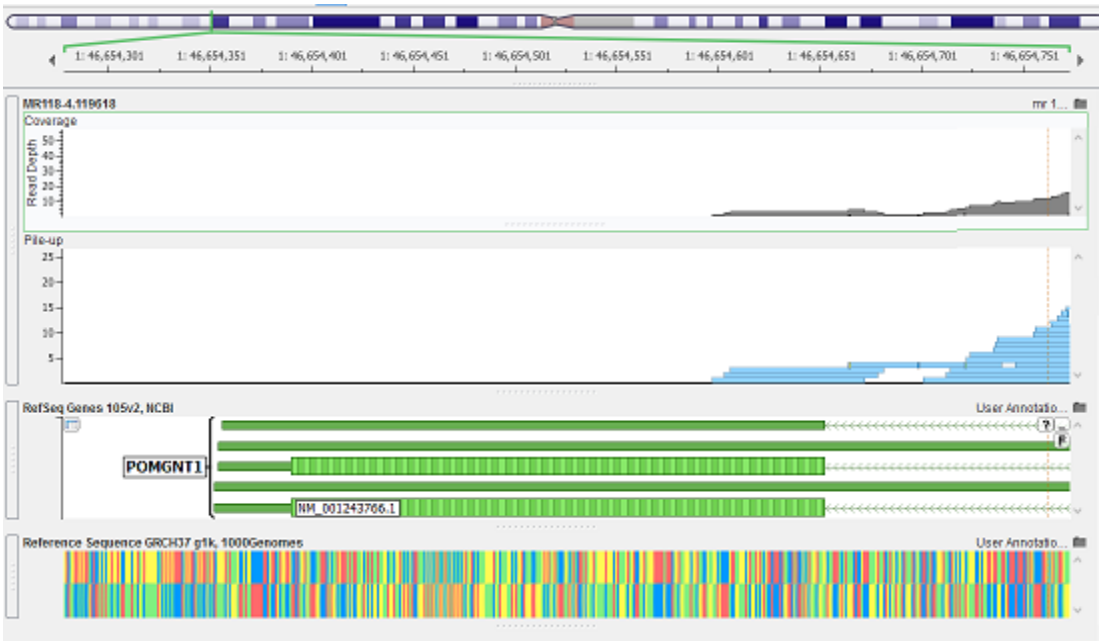

Figure 3.6 The screen shot shows exon-23 of *POMGNT1* gene which is partially covered in the exome sequence obtained from affected individual IV-4.

Figure 3.7 The screen shot shows exon 5-15 of *SZT2* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4.

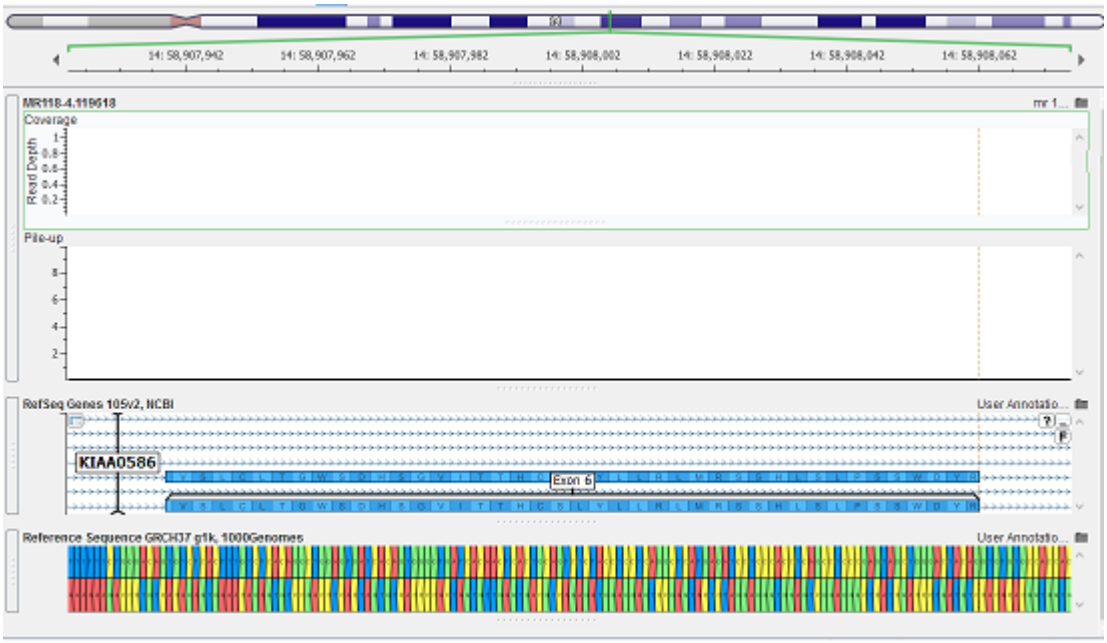

Figure 3.8 The screen shot shows exon-6 of *KIAA0586* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4.

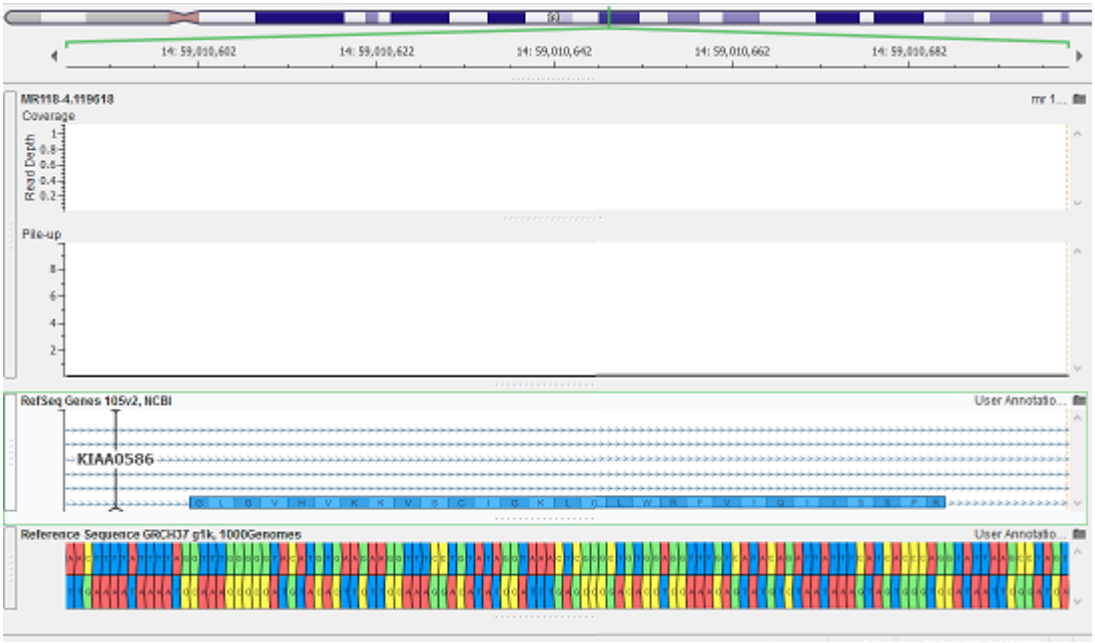

Figure 3.9 The screen shot shows exon 33 of *KIAA0586* gene which is completely uncovered in the exome sequence obtained from affected individual IV-4.

3.2.1.2.1.1 Uncovered Genes during Exome Sequencing of Family B

The uncovered exons of these genes were sequenced through Sanger sequencing by designing the primers from their flanking genomic regions. The sequenced genes were analysed through CLC Genomics Workbench v3.6.5. The sequencing data of *POMGNT1* gene (exon-23) revealed absence of variant in this region (Fig 3.10). *SZT2* gene (exon 5-15) sequencing data also revealed that there is no variant found in this region (Fig 3.11-3.16). *KIAA0586* gene (exon 6 & 33) sequencing data revealed an intronic variation at physical location of chr14:58908158G>A, which is 97bp away from splice donor site of exon 6 (Figure 3.17). This variation was predicted as polymorphism and might affect protein features and changes splice site according to mutation taster. This variation is reported in 1000G (Homozygotes: 2415 & heterozygotes: 89) but is absent in ExAC. Human splice site Finder (HSF) predicts that this polymorphism has no impact on splice site. However, sequence analysis of *KIAA0586* (exon-33) did not identified pathogenic variant (Fig 3.17). The sequencing results of uncovered genes are summarized in table 3.5.

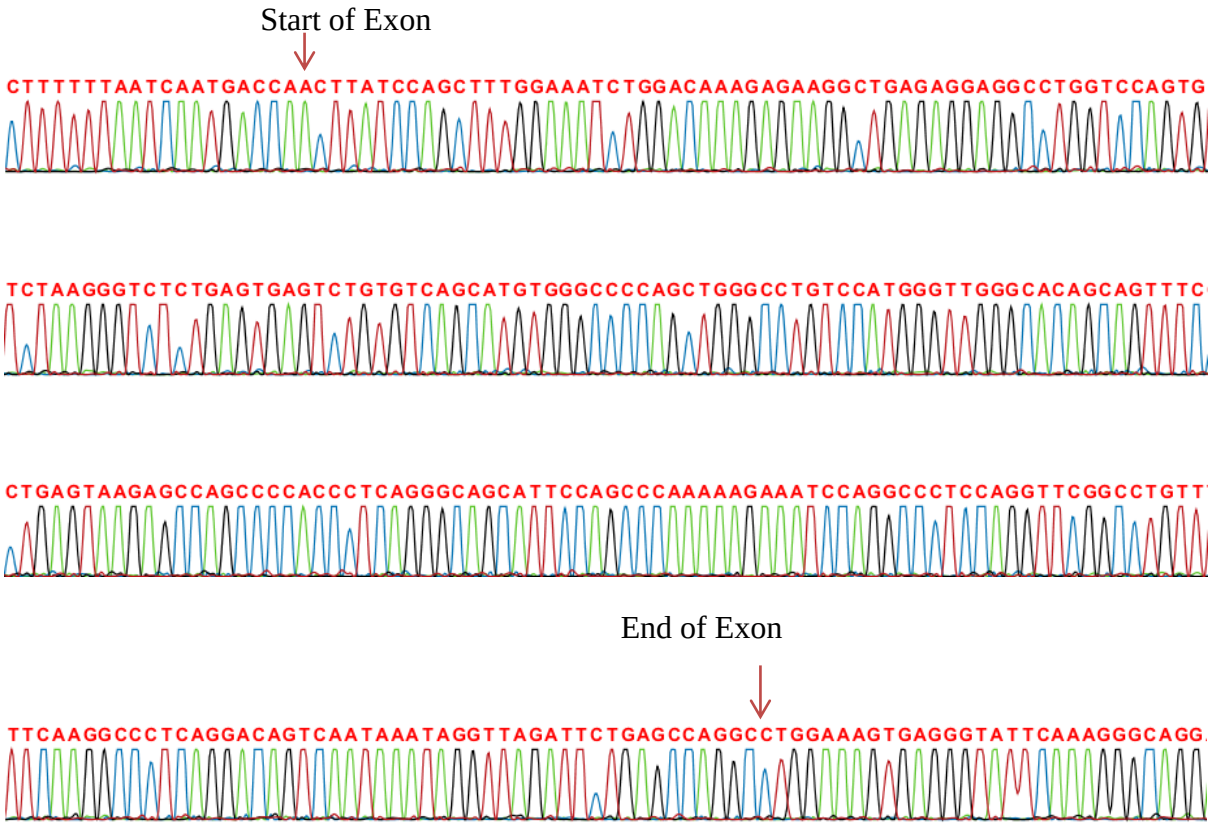

Figure 3.10: Chromatogram of exon 23 of *POMGNT1* gene sequenced in affected individual IV-4 of family B.

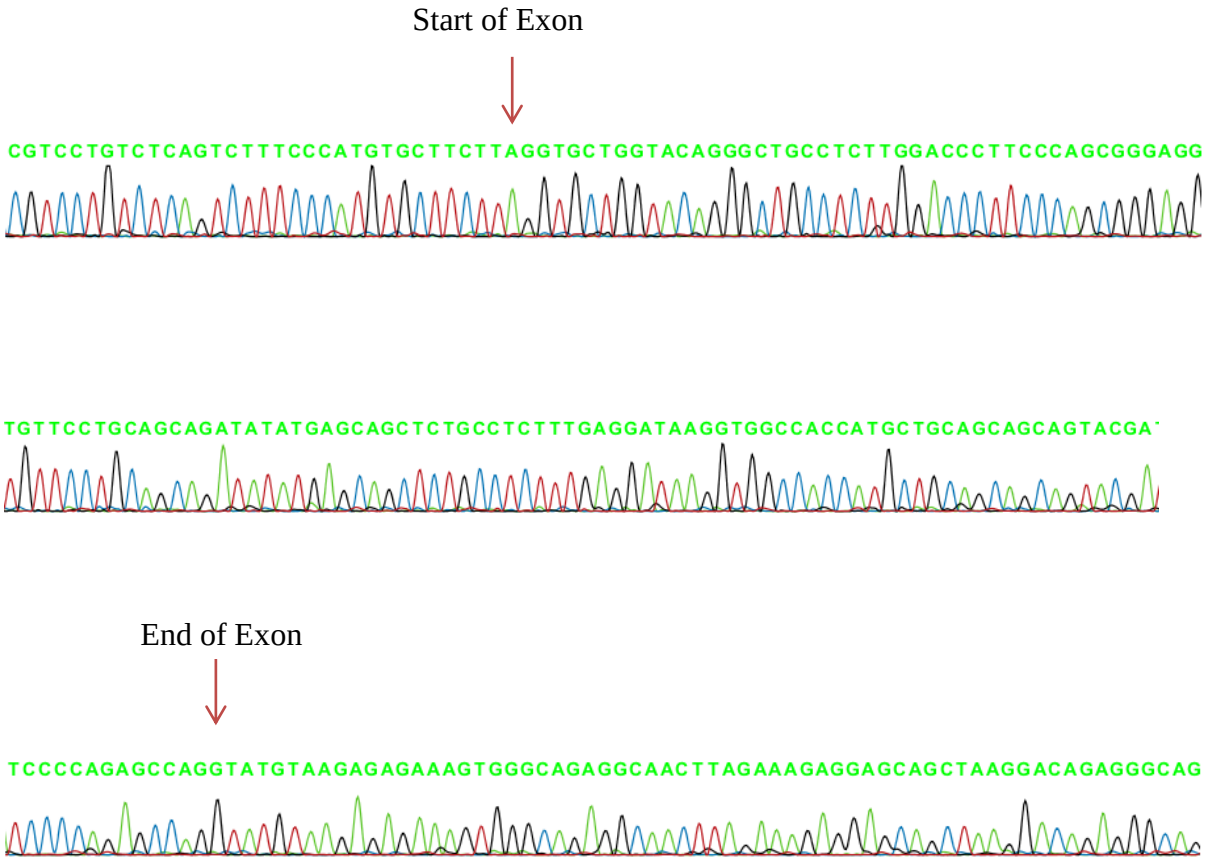

Figure 3.11: Chromatogram of exon 5 of *SZT2* gene sequenced in affected individual IV-4 of family B.

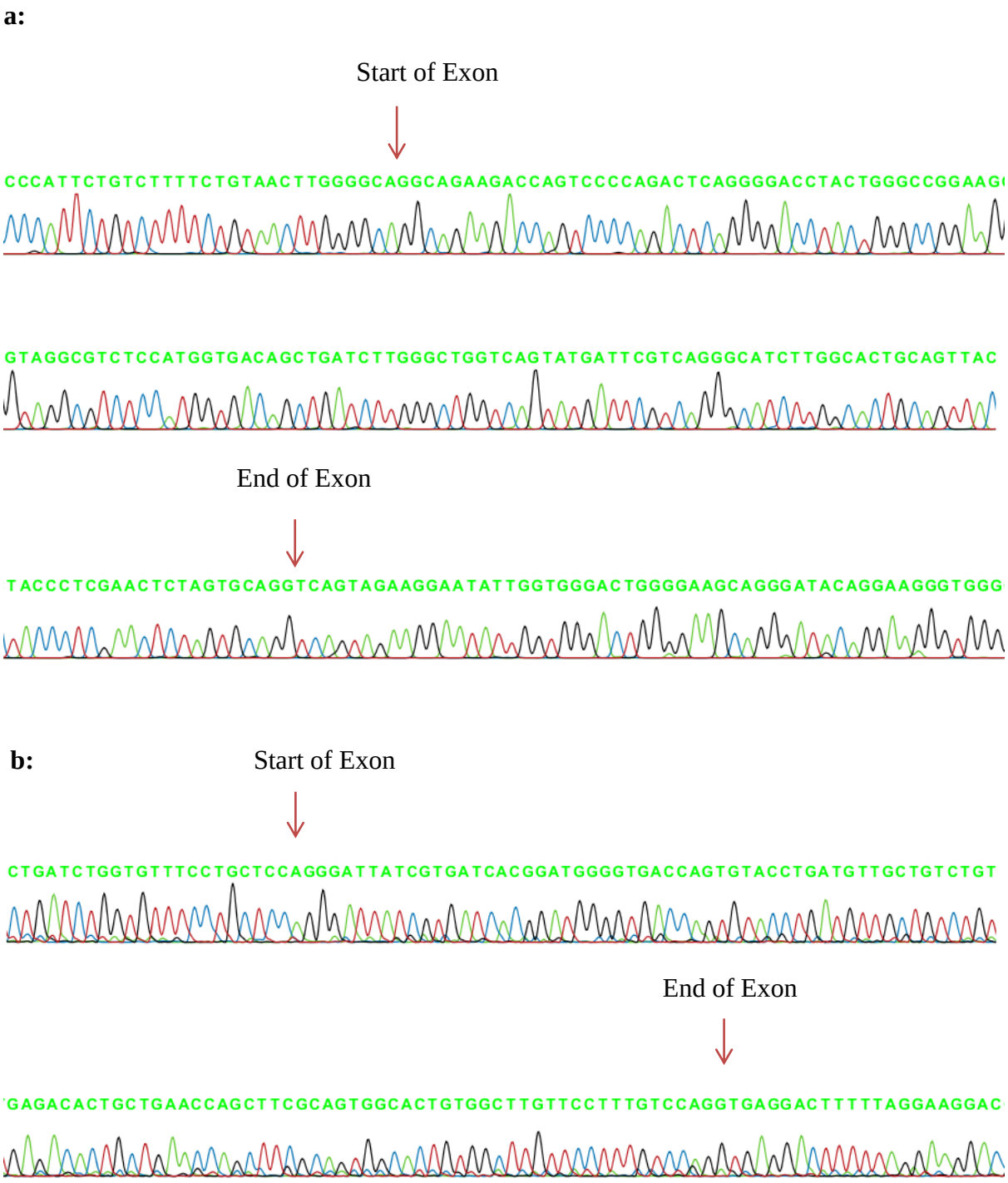

Figure 3.12: (a) Chromatogram of exon 6 of *SZT2* gene sequenced in affected individual IV-4 of family B. **(b)** Chromatogram of exon 7 of *SZT2* gene sequenced in affected individual IV-4 of family B.

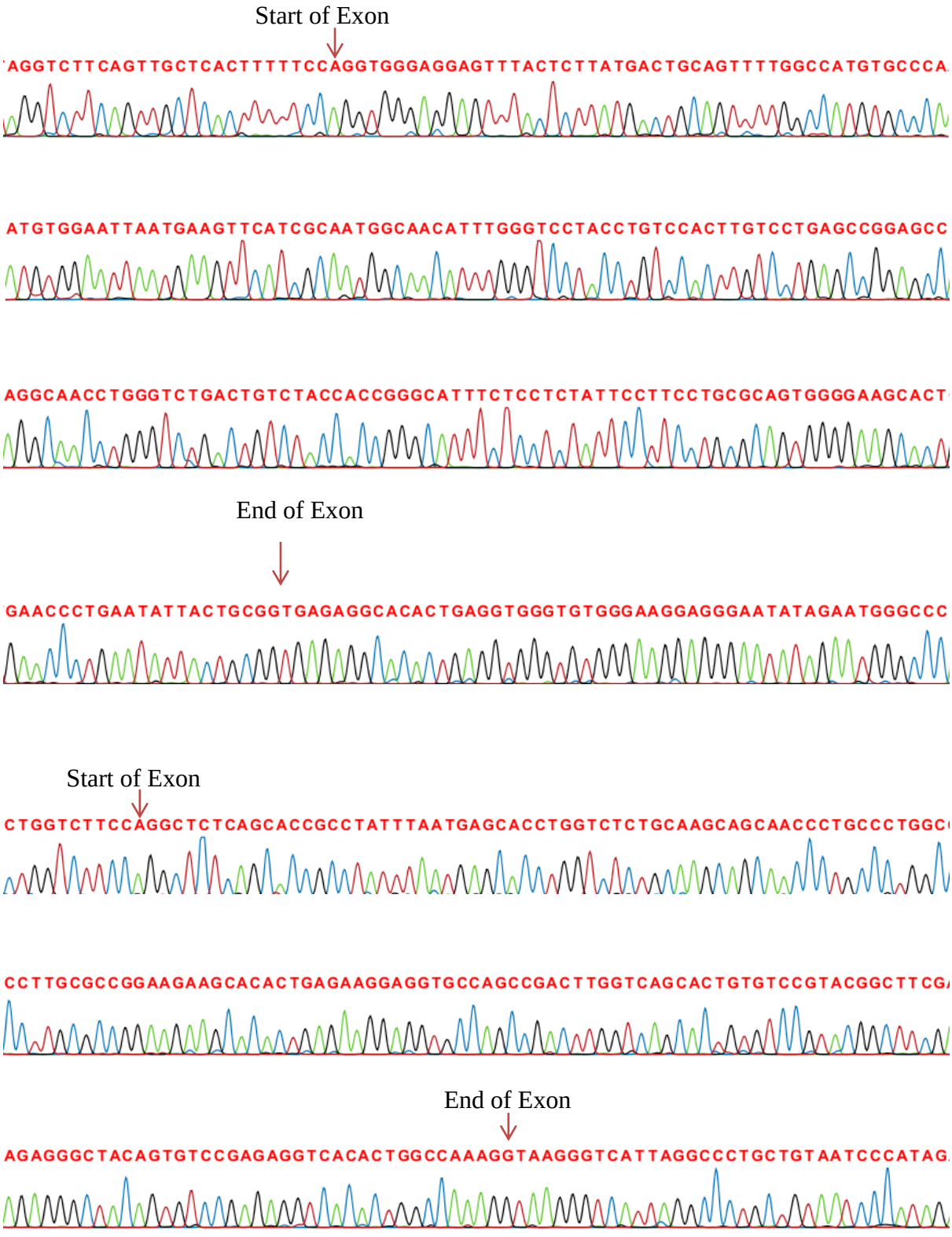

Figure 3.13: Chromatogram of exon 8-9 of *SZT2* gene sequenced in affected individual IV-4 of family B.

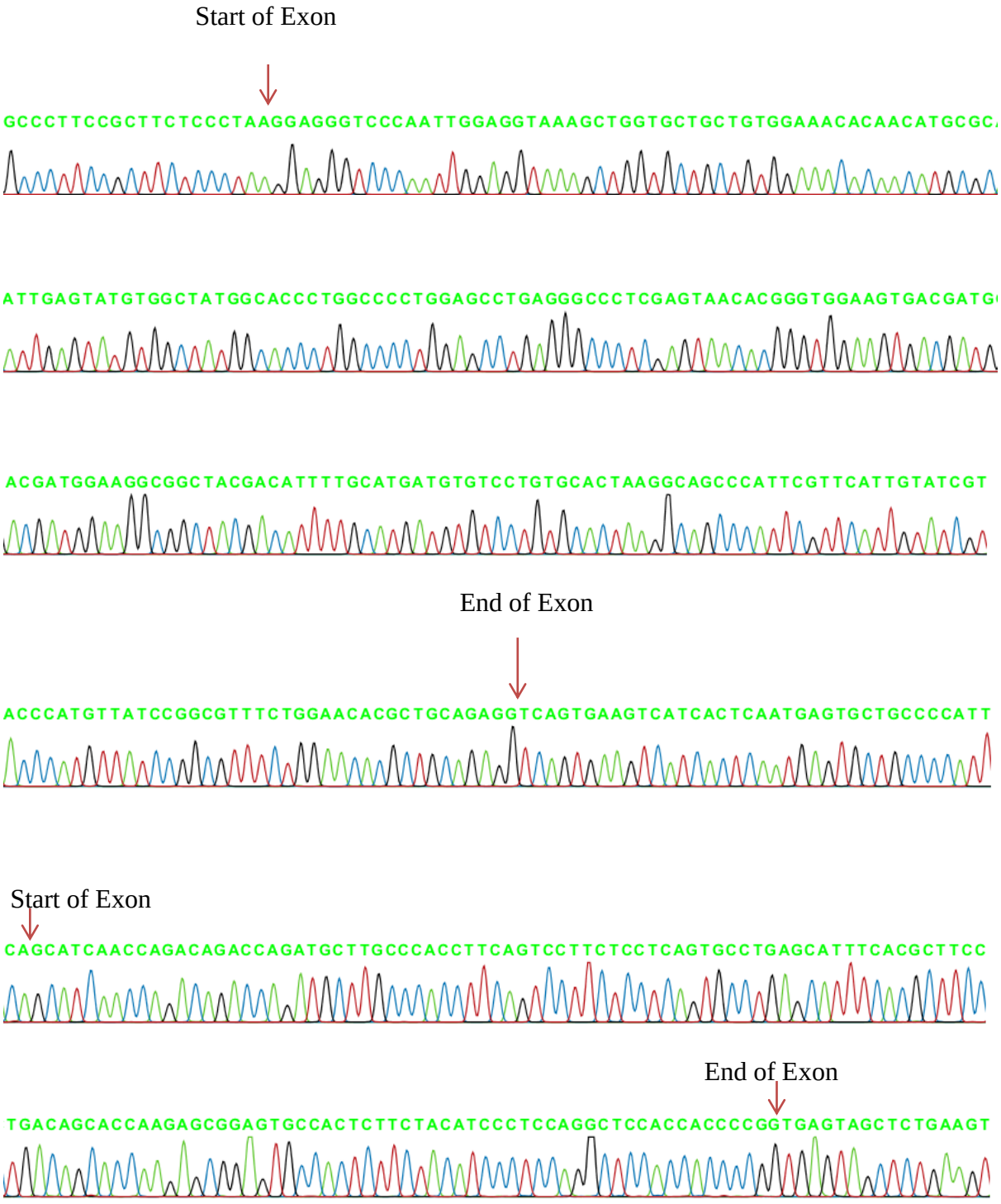

Figure 3.14: Chromatogram of exon 10-11 of *SZT2* gene sequenced in affected individual IV-4 of family B.

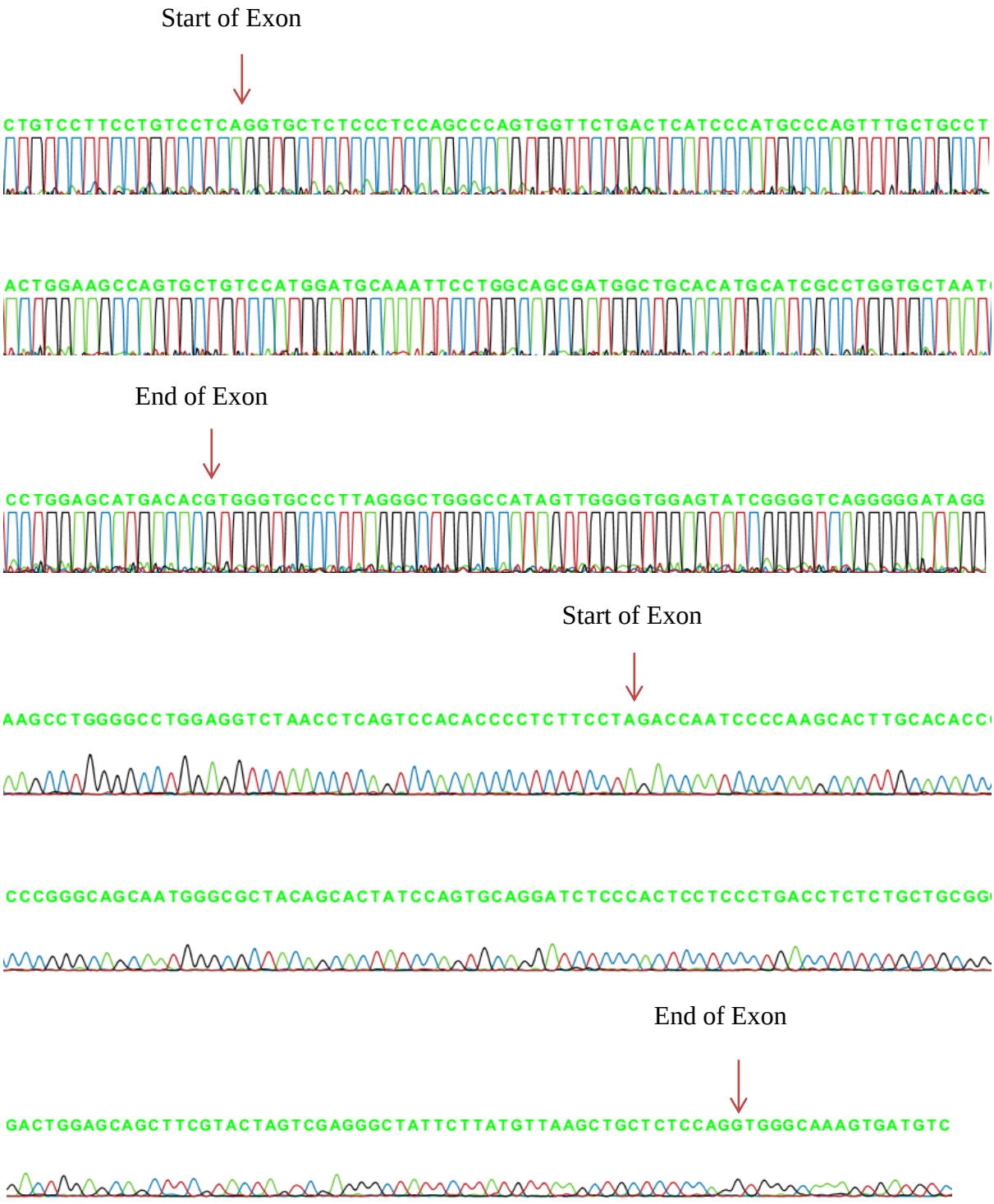

Figure 3.15: Chromatogram of exon 12-13 of *SZT2* gene sequenced in affected individual IV-4 of family B.

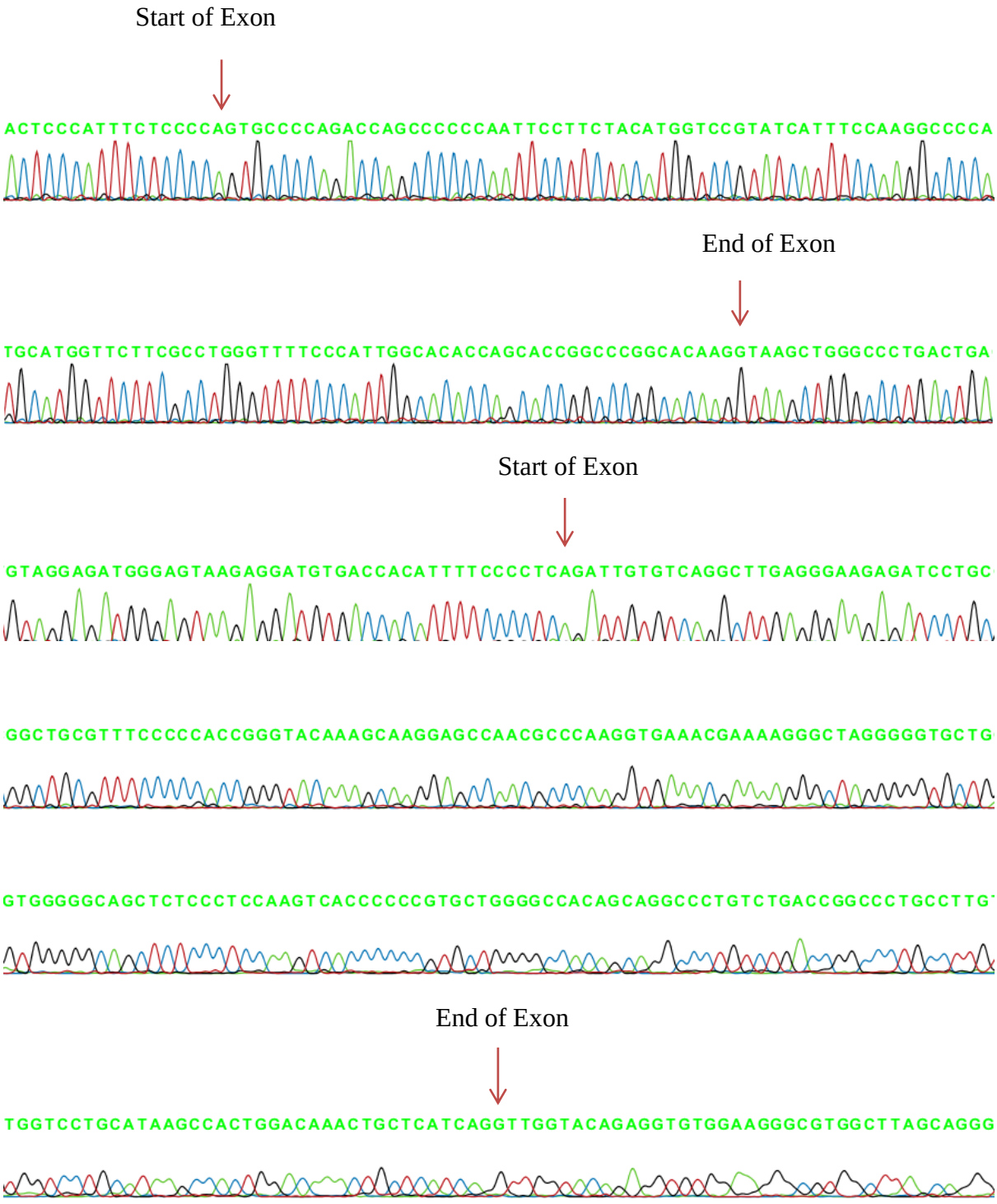

Figure 3.16: Chromatogram of exon 14-15 of *SZT2* gene sequenced in affected individual IV-4 of family B.

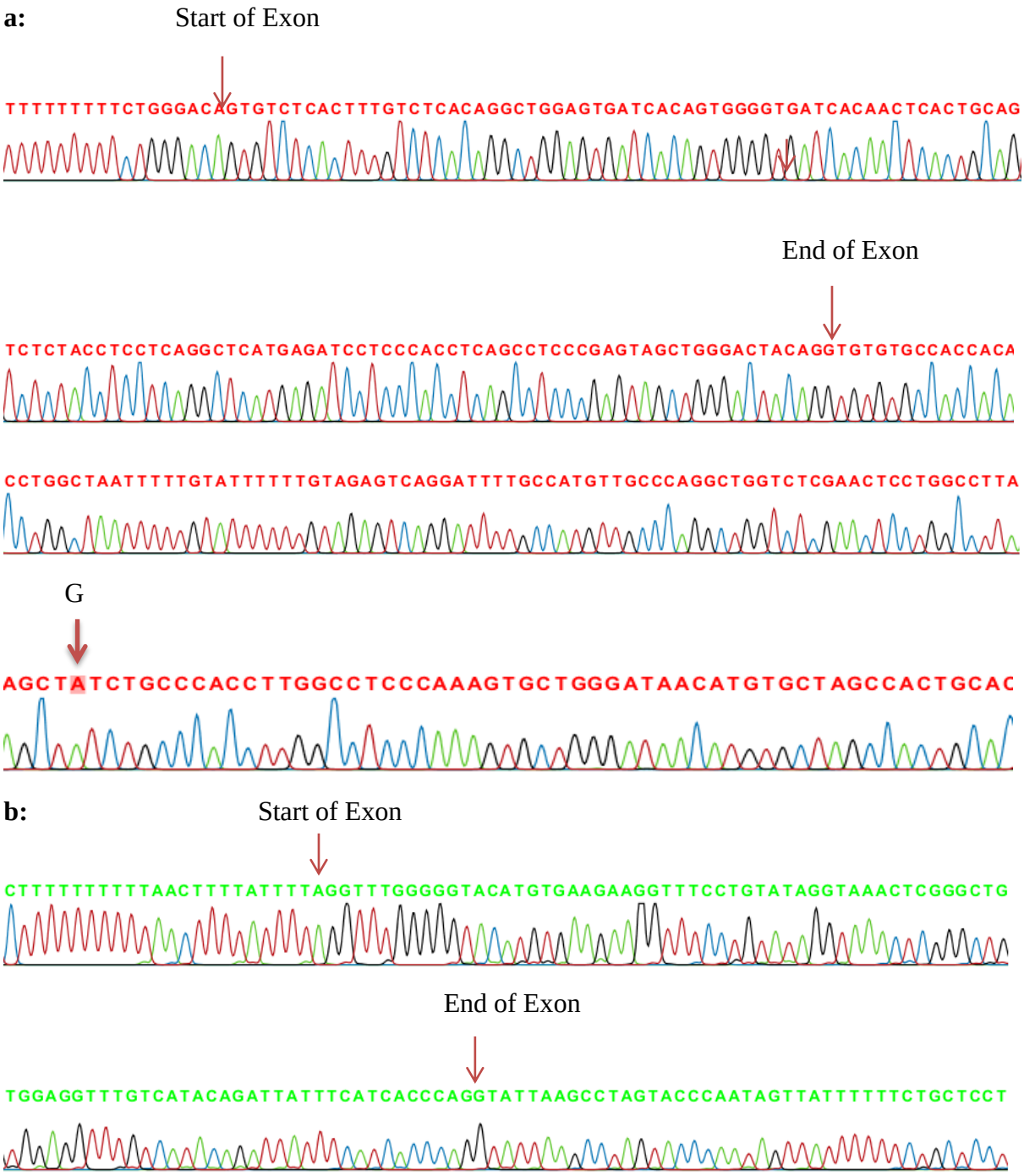

Figure 3.17: (a) Chromatogram of exon 6 of *KIAA0586* gene sequenced in affected individual IV-4 of family B. **(b)** Chromatogram of exon 33 of *KIAA0586* gene sequenced in affected individual of IV-4 of family B.

Table 3.5 Variants identified during the sequencing of *POMGNT1*, *SZT2* and *KIAA0586* genes in family B.

Gene Name	***POMGNT1***	***SZT2***	***KIAA0586***	***KIAA0586***
Exon#	23	5-15	6	33
Variant Physical location	No	No	chr14:58908158G>A	No
Mutation Taster Prediction	NA	NA	Polymorphism	NA
Alteration region	NA	NA	Intron	NA
ExAC frequency	NA	NA	NA	NA
1000G	NA	NA	Homozygote (2415) & Heterozygote (89)	NA
Zygosity	NA	NA	Homozygote	NA
Human Splicing Finder (HSF)	NA	NA	No impact on Splice site	NA
Distance from splice site	NA	NA	97bp away from splice donor site	NA

4. Discussion

Intellectual disability is a neurodevelopmental disorder which is characterized on the basis of deficits in cognition, adaptive behaviour and Intelligence quotient (IQ) below or equal to 70 (American Psychiatric Association, 2000). The worldwide prevalence of ID is 1-3% (Roeleveld *et al*., 1997; Leonard and Wen, 2002). Even though ID occurrence is worldwide but its prevalence is higher in developing countries and in lower socioeconomic areas (Drews *et al*., 1995; Roeleveld *et al*., 1997; Durkin *et al*., 1998; Durkin, 2002; Emerson, 2007). The main factors that cause ID are; environmental insults (such as teratogens, infection and trauma), intoxication during pregnancy, complications of delivery, childhood onset diseases, and premature birth. These factors are responsible for nervous system impairment and cause ID in a significant number of cases. However, major proportion of ID (20-50%) is caused due to genetic abnormalities (Kaufman *et al*., 2010; Liu *et al*., 2010; Ellison *et al*., 2013) which can be categorised as chromosomal abnormalities (duplications, aneuploidies and sub microscopic deletions) and monogenic defects (Ropers *et al*., 2010). According to current estimation, ~15%-20% ID cases are due to sub- microscopic copy number variations (CNVs) (Zahir *et al*., 2007; Hochstenbach *et al*., 2009; Miller *et al*., 2010; Cooper *et al*., 2011). However, ~60% of ID cases have an unknown etiology (Rauch *et al*., 2006).

In the current study, two families (A & B) were recruited from the remote areas of Sindh and Punjab provinces of Pakistan. Affected members of both families show non-syndromic mild ID without any history of fits. They have short term memory and poor speech development. The affected members of family A show non-aggressive behaviour but affected members of family B show aggressive behaviour. Clinical features of affected members of both families are quite similar to ID patients reported in the earlier studies from Iran and Pakistan (Najmabadi *et al*., 2007; Mir *et al*., 2009; Rafiq *et al*., 2010; Khan *et al*., 2012; Riazuddin *et al*., 2017; Harripaul *et al*., 2017). Both families exhibit similar clinical features except the aggressive behaviour that was noted in the members of family B.

Whole genome SNP genotyping was performed on both families (A & B) to find the homozygous by descent region (HBD) containing potential genes responsible for ID phenotype. In family A, analysis of the genotype data did not identify any homozygous region, but in a subsequent analysis the region of continuous SNP no call were identified

to explore the involvement of genomic deletions. Through these analysis, a genomic region with multiple SNP no call region was identified on chromosome 2p16.3, which was flanked by rs6706472 and rs12713108 and spans a genomic region of ~559kb (chr2:50,228,332-50,786,770). This genomic region contain NRXN1 gene which is highly expressed in the brain (Kowalnik *et al*., 2010). Deletions and loss of function mutations of NRXN1 gene are known to cause intellectual disability, autism and schizophrenia (Kim *et al*., 2008; Rujescu *et al*., 2008; Glessner *et al*., 2009; Kirov *et al*., 2009; Gauthier *et al*., 2011). Harrison *et al*. (2011) reported compound heterozygous deletions of *NRXN1* gene (one cause the deletion of exon 20 & 21 and second one affects the promoter and 1-5 exons) in two sisters having severe, early onset epilepsy. Friedman *et al*. (2006) reported a de novo 320 kb deletion on 2p16.3 and removes exons 1-5 of NRXN1 gene in a 7 year old boy with mild facial dysmorphism, vertebral anomalies, cognitive impairment and autistic features. The Autism Genome Project Consortium (2007) also reported *de novo* hemizygous deletions in two siblings with language regression and autism. Zweier *et al*. (2009) reported biallelic NRXN1 deletions in a patient with ID. Gregor *et al*. (2011) reported heterozygous deletion in *NRXN1* in a patient with severe ID. Additionally, CNVs in NRXN1 are implicated in intellectual disability and autism (Simon *et al*., 2005; Feng *et al*., 2006; Szatmari *et al*., 2007; Zahir *et al*., 2007; Marshall *et al*., 2008; Kim *et al*., 2008).

Based on the involvement of *NRXN1* gene in intellectual disability, and its presence within a probably deleted region PCR based assays were used to check its involvement in the patients of family A. However, analysis of PCR results did not confirm the presence of deletion in the affected members of this family. This no call region might be due to base calling and alignment error. Many NGS studies depend on low-coverage sequencing (<5x per site per individual, on average) due to which there is high possibility that only one of two chromosomes has been sampled at a specified site of a diploid individual. Under such conditions, SNP and genotype calling are difficult due to which a considerable uncertainty is often associated with the results (Nielsen *et al*., 2011). Thus, even though linkage studies have found strong evidence of a genetic influence but many rare variants with aetiological significance may be ignored because they do not reach stringent significance parameters and will not be captured by SNP arrays (Chen *et al*., 2017).

In family B, analysis of the genotype data identified four homozygous regions. Exome sequencing of ID patient IV-4 could not identify an ID causing variant. The BAM files of this patient were analysed to check the coverage of the targeted genomic region, especially in the mapped homozygous regions. BAM file analysis detected uncovered regions in two HBD regions. In case of chromosome 1: (38401933-64109264) HBD region two genes *POMGNT1* (exon-23) and *SZT2* (exon-5-15) were not fully uncovered while in chromosome 14:56307478-75947712 HBD region *KIAA0586* (exon-6 & 33) gene was not partially covered.

Mutations in *POMGNT1* gene are responsible for neural migration defects, congenital muscular dystrophy and ocular abnormalities (Yoshida *et al*., 2001). Thirteen different mutations in *POMGNT1* are known to result in muscle eye brain (MEB) disease (Yoshinda *et al*., 2001; Taniguchi *et al*., 2003). Manya *et al*. (2003) reported that mutations near 5′ end of *POMGNT1* coding region causes severe brain symptoms (e.g. hydrocephalus) while at 3′ end it causes milder phenotypes. Based on their involvement in brain and neural migration, the uncovered genomic region (exon 23) of *POMGNT1* was Sanger sequenced in individual IV-4, but could not identify a pathogenic variant.

Mutations in *SZT2* gene are responsible for behavioural deficiencies and mild cognition to profound delays and epileptic encephalopathy (Venkatesan *et al*., 2016). A homozygous non-sense mutation and a compound heterozygous (nonsense mutation along with exonic splice site) mutation results in distinct neuro-radiological anomalies, severe type of autosomal recessive-infantile encephalopathy along with intractable seizures (Vanagaite *et al*., 2013). A 3bp deletion mutation (c.4202_4204delTTC) was reported in two unrelated children having severe intellectual disability and unexplained infantile epileptic encephalopathy. Authors concluded that this deletion causes disruption of *SZT2* with some residual function which might lead to moderate or mild ID without seizures (Falcon *et al*., 2013). On the basis of their involvement in intellectual disability, the uncovered region (exon 5 to 15) of *SZT2* was Sanger sequenced but no pathogenic variant was identified in these exons.

KIAA0586 homozygous nonsense and splicing mutations are responsible for lethal ciliopathies and abnormal SHH (sonic hedgehog) signalling (Alby *et al*., 2015). *KIAA0586* mutations results in mild Joubert syndrome (Gagescu *et al*., 2015a). Sanger sequencing was performed on *KIAA0586* (exon 6 & 33). During sequence analysis of *KIAA0586* uncovered exon 33 pathogenic variant was not identified but an intronic

variation was found at chr14:58908158G>A position. This variation is 97bp away from splice donor site of exon 6 (Figure 3.18) but was predicted as polymorphism. However, variation might affect protein features by altering the splicing process. This variation is reported in 1000G (Homozygotes: 2415 & heterozygotes: 89) but is absent in ExAC. Human splice site Finder (HSF) predicts that this polymorphism has no impact on splice site.

Taken together, this study showed that whole genome genotyping and exome sequencing could not identify the ID causing gene in both families. Additionally, our data showed that detection of genomic deletions from the genotype data may not be reliable, and such cases should be carefully evaluated to conclude their pathogenic nature. Similarly, uncovered regions were found in whole exome sequencing which contain genes already known to cause ID but Sanger sequencing of these uncovered exons of these genes could not identify any pathogenic variant. Therefore, these families can be subjected to whole genome sequencing to identify disease causing mutations.

5. References

Accardo, P. J., & Capute, A. J. (Eds.). (2008). Capute & Accardo's Neurodevelopmental Disabilities in Infancy and Childhood: Neurodevelopmental diagnosis and treatment (Vol. 1). Brookes Pub.

Alby, C., Piquand, K., Huber, C., Megarbané, A., Ichkou, A., Legendre, M., ... & El Chehadeh-Djebbar, S. (2015). Mutations in KIAA0586 cause lethal ciliopathies ranging from a hydrolethalus phenotype to short-rib polydactyly syndrome. The American Journal of Human Genetics, 97(2), 311-318.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders; revised (DSM-III-R). Washington DG.

American Psychiatric Association, & American Psychiatric Association. (2000). DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision. *Washington, DC*: American Psychiatric Association, *75*, 78-85.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders *(DSM-5®)*. American Psychiatric Pub.

Anand, K. S., & Dhikav, V. (2012). Hippocampus in health and disease: An overview. Annals of Indian Academy of Neurology, 15(4), 239.

Bachmann-Gagescu, R., Phelps, I. G., Dempsey, J. C., Sharma, V. A., Ishak, G. E., Boyle, E. A., ... & Doherty, D. (2015a). KIAA0586 is mutated in Joubert syndrome. Human mutation, 36(9), 831-835.

Bachmann-Gagescu, R., Dempsey, J. C., Phelps, I. G., O'roak, B. J., Knutzen, D. M., Rue, T. C., ... & Boyle, E. A. (2015b). Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. Journal of medical genetics, 52(8), 514-522.

Bailey, R. (2013). Anatomy of the Brain-Cerebellum. About Education. http://biology. about. com/od/anatomy/p/frontal-lobes. htm (accessed February 27, 2017).

Ballif, B. C., Sulpizio, S. G., Lloyd, R. M., Minier, S. L., Theisen, A., Bejjani, B. A., & Shaffer, L. G. (2007). The clinical utility of enhanced subtelomeric coverage in array CGH. American Journal of Medical Genetics Part A, 143(16), 1850-1857.

Bangs, F., Antonio, N., Thongnuek, P., Welten, M., Davey, M. G., Briscoe, J., & Tickle, C. (2011). Generation of mice with functional inactivation of talpid3, a gene first identified in chicken. Development, 138(15), 3261-3272.

Basel-Vanagaite, L. (2007). Genetics of autosomal recessive non-syndromic mental retardation: recent advances. Clinical genetics, 72 (3), 167-174.

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

Ben J, Elworthy S, Ng ASM, vanEeden F, Ingham PW. (2011). Targeted mutation of the talpid3 gene in zebrafish reveals its conserved requirement for ciliogenesis and Hedgehog signalling across the vertebrates. Development, 138, 4969–4978.

Benitez, B. A., Alvarado, D., Cai, Y., Mayo, K., Chakraverty, S., Norton, J., ... & Cruchaga, C. (2011). Exome-sequencing confirms DNAJC5 mutations as cause of adult neuronal ceroid-lipofuscinosis. PloS one, 6(11), e26741.

Bennett, R. L. (1999). The Language of the Pedigree. The Practical Guide to the Genetic Family History, 1-12.

Bennett, R. L., French, K. S., Resta, R. G., & Doyle, D. L. (2008). Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. Journal of genetic counseling, 17(5), 424-433.

Berglund, E. C., Kiialainen, A., & Syvänen, A. C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative genetics, 2(1), 23.

Bijlsma, E. K., Collins, A., Papa, F. T., Tejada, M. I., Wheeler, P., Peeters, E. A. J., ... & Broekma, A. J. (2012). Xq28 duplications including MECP2 in five females: Expanding the phenotype to severe mental retardation. European journal of medical genetics, 55(6), 404-413.

Bortolotto, Z. A., Clarke, V. R., Delany, C. M., Parry, M. C., Smolders, I., Vignes, M., ... & Ogden, A. (1999). Kainate receptors are involved in synaptic plasticity. Nature, 402(6759), 297.

Bradley, E. A., Thompson, A., & Bryson, S. E. (2002). Mental retardation in teenagers: prevalence data from the Niagara region, Ontario. The Canadian Journal of Psychiatry, 47(7), 652-659.

Brookes, A. J. (1999). The essence of SNPs. Gene, 234(2), 177-186.

Bucan, M., Abrahams, B. S., Wang, K., Glessner, J. T., Herman, E. I., Sonnenblick, L. I., ... & Kim, C. (2009). Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS genetics, 5(6), e1000536.

Budworth, H., & McMurray, C. T. (2013). A brief history of triplet repeat diseases. In Trinucleotide Repeat Protocols (pp. 3-17). Humana Press, Totowa, NJ.

Butcher, L. M., Meaburn, E., Liu, L., Fernandes, C., Hill, L., Al-Chalabi, A., ... & Craig, I. W. (2004). Genotyping pooled DNA on microarrays: a systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. Behavior Genetics, 34(5), 549-555.

Çalışkan, M., Chong, J. X., Uricchio, L., Anderson, R., Chen, P., Sougnez, C., ... & Matern, D. (2011). Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13. Human molecular genetics, 20(7), 1285-1289.

Campeau, P. M., Kasperaviciute, D., Lu, J. T., Burrage, L. C., Kim, C., Hori, M., ... & Wisniewska, M. (2014). The genetic basis of DOORS syndrome: an exome-sequencing study. The Lancet Neurology, 13(1), 44-58.

Centers for Disease Control and Prevention. (2009). Autism and Developmental Disabilities Monitoring Network, United States, 2006. Surveill Summ MMWR, 58, 1-20.

Chao, L. L., & Knight, R. T. (1998). Contribution of human prefrontal cortex to delay performance. Journal of cognitive neuroscience, 10(2), 167-177.

Chen, X. S., Reader, R. H., Hoischen, A., Veltman, J. A., Simpson, N. H., Francks, C., ... & Fisher, S. E. (2017). Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. Scientific reports, 7, 46105.

Cherniak, C. (1990). The bounded brain: toward quantitative neuroanatomy. Journal of Cognitive Neuroscience, 2(1), 58-68.

Ching, M. S., Shen, Y., Tan, W. H., Jeste, S. S., Morrow, E. M., Chen, X., ... & Austin, C. (2010). Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 153(4), 937-947.

Chiurazzi, P., & Pirozzi, F. (2016). Advances in understanding–genetic basis of intellectual disability. F1000Research, 5.

Collins, F. S., Green, E. D., Guttmacher, A. E., & Guyer, M. S. (2003). A vision for the future of genomics research. Nature, 422(6934), 835.

Collin, R. W., van den Born, L. I., Klevering, B. J., de Castro-Miró, M., Littink, K. W., Arimadyo, K., ... & Siemiatkowska, A. M. (2011). High-resolution homozygosity mapping is a powerful tool to detect novel mutations causative of autosomal recessive RP in the Dutch population. Investigative ophthalmology & visual science, 52(5), 2227-2239.

Contractor, A., Swanson, G., & Heinemann, S. F. (2001). Kainate receptors are involved in short-and long-term plasticity at mossy fiber synapses in the hippocampus. Neuron, 29(1), 209-216.

Cook Jr, E. H., & Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature, 455(7215), 919.

Cooper, D. N., Krawczak, M., & Antonarakis, S. E. (1995). The nature and mechanisms of human gene mutation. Vol. 1. New York: McGraw-Hill, pp. 259-291.

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., ... & Abdel-Hamid, H. (2011). A copy number variation morbidity map of developmental delay. Nature genetics, 43(9), 838.

de Vries, B. B., Pfundt, R., Leisink, M., Koolen, D. A., Vissers, L. E., Janssen, I. M., ... & Smeets, D. (2005). Diagnostic genome profiling in mental retardation. The American Journal of Human Genetics, 77(4), 606-616.

Dobyns, W. B., Reiner, O., Carrozzo, R., & Ledbetter, D. H. (1993). Lissencephaly. Jama, 270, 2838-2842.

Drews, C. D., Yeargin-Allsopp, M., Decoufle, P., & Murphy, C. C. (1995). Variation in the influence of selected sociodemographic risk factors for mental retardation. American journal of public health, 85(3), 329-334.

Dudanova, I., Sedej, S., Ahmad, M., Masius, H., Sargsyan, V., Zhang, W., ... & Missler, M. (2006). Important contribution of α-neurexins to Ca2+-triggered exocytosis of secretory granules. Journal of Neuroscience, 26(41), 10599-10613.

Duong, L., Klitten, L. L., Møller, R. S., Ingason, A., Jakobsen, K. D., Skjødt, C., ... & Tommerup, N. (2012). Mutations in NRXN1 in a family multiply affected with brain disorders: NRXN1 mutations and brain disorders. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 159(3), 354-358.

Durkin, M. S., Hasan, Z. M., & Hasan, K. Z. (1998). Prevalence and correlates of mental retardation among children in Karachi, Pakistan. American journal of epidemiology, 147(3), 281-288.

Durkin, M. (2002). The epidemiology of developmental disabilities in low-income countries. Developmental Disabilities Research Reviews, 8(3), 206-211.

Ellison, J. W., Rosenfeld, J. A., & Shaffer, L. G. (2013). Genetic basis of intellectual disability. Annual review of medicine, 64, 441-450.

Emerson, E. (2007). Poverty and people with intellectual disabilities. Developmental Disabilities Research Reviews, 13(2), 107-113.

Falcone, M., Yariz, K. O., Ross, D. B., Foster II, J., Menendez, I., & Tekin, M. (2013). An amino acid deletion in SZT2 in a family with non-syndromic intellectual disability. PloS one, 8(12), e82810.

Fellermann, K., Stange, D. E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C. L., ... & Radlwimmer, B. (2006). A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. The American Journal of Human Genetics, 79(3), 439-448.

Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., ... & Sommer, S. S. (2006). High frequency of neurexin 1β signal peptide structural variants in patients with autism. Neuroscience letters, 409(1), 10-13.

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. Nature Reviews Genetics, 7(2), 85.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 269(5223), 496-512.

Fox SI (2016). Human Physiology.14th Edition, pp. 207-230, McGraw-Hill Education, New York.

Frankel, W. N., Yang, Y., Mahaffey, C. L., Beyer, B. J., & O'Brien, T. P. (2009). Szt2, a novel gene for seizure threshold in mice. Genes, Brain and Behavior, 8(5), 568-576.

Friedman, J. M., Baross, Á., Delaney, A. D., Ally, A., Arbour, L., Asano, J., ... & Cao, M. (2006). Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. The American Journal of Human Genetics, 79(3), 500-513.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. Nature, 405(6784), 347.

Gauthier, J., Siddiqui, T. J., Huashan, P., Yokomaku, D., Hamdan, F. F., Champagne, N., ... & Fathalli, F. (2011). Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Human genetics, 130(4), 563-573.

Giannandrea, M., Bianchi, V., Mignogna, M. L., Sirri, A., Carrabino, S., D'Elia, E., ... & Ropers, H. H. (2010). Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. The American Journal of Human Genetics, 86(2), 185-195.

Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. European Journal of Human Genetics, 20(5), 490.

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., & Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science, 318(5853), 1136-1140.

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., ... & Imielinski, M. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature, 459(7246), 569.

Gregor, A., Albrecht, B., Bader, I., Bijlsma, E. K., Ekici, A. B., Engels, H., ... & Kohlhase, J. (2011). Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC medical genetics, 12(1), 106.

Gregory, T. R. (2005). Synergy between sequence and size in large-scale genomics. Nature Reviews Genetics, 6(9), 699.

Gschwend, T. P., Krueger, S. R., Kozlov, S. V., Wolfer, D. P., & Sonderegger, P. (1997). Neurotrypsin, a novel multidomain serine protease expressed in the nervous system. Molecular and Cellular Neuroscience, 9(3), 207-219.

Gunay-Aygun, M. (2009). Liver and kidney disease in ciliopathies. In American Journal of Medical Genetics Part C: Seminars in Medical Genetics (Vol. 151, No. 4, pp. 296-306). Wiley Subscription Services, Inc., A Wiley Company.

Hamdan, F. F., Gauthier, J., Spiegelman, D., Noreau, A., Yang, Y., Pellerin, S., ... & D'Anjou, G. (2009). Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. New England Journal of Medicine, 360(6), 599-605.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., ... & Johnson, M. (2017). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Molecular psychiatry.

Harding, K. E., & Robertson, N. P. (2014). Applications of next-generation whole exome sequencing. Journal of neurology, 261(6), 1244-1246.

Harrison, V., Connell, L., Hayesmoore, J., McParland, J., Pike, M. G., & Blair, E. (2011). Compound heterozygous deletion of NRXN1 causing severe developmental delay with early onset epilepsy in two sisters. American Journal of Medical Genetics Part A, 155(11), 2826-2831.

Higgins, J. J., Pucilowska, J., Lombardi, R. Q., & Rooney, J. P. (2004). A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. Neurology, 63(10), 1927-1931.

Hochstenbach, R., van Binsbergen, E., Engelen, J., Nieuwint, A., Polstra, A., Poddighe, P., ... & Poot, M. (2009). Array analysis and karyotyping: workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. European journal of medical genetics, 52(4), 161-169.

Hofman, M. A. (2012). Design principles of the human brain: an evolutionary perspective. In Progress in brain research (Vol. 195, pp. 373-390). Elsevier.

Hollox, E. J., Huffmeier, U., Zeeuwen, P. L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., ... & Reis, A. (2008). Psoriasis is associated with increased β-defensin genomic copy number. Nature genetics, 40(1), 23.

Hu, W. H., Pendergast, J. S., Mo, X. M., Brambilla, R., Bracchi-Ricard, V., Li, F., ... & Bethea, J. R. (2005). NIBP, a novel NIK and IKKβ-binding protein that enhances NF-κB activation. Journal of Biological Chemistry, 280(32), 29233-29241.

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., ... & Lee, C. (2004). Detection of large-scale variation in the human genome. Nature genetics, 3*6*(9), 949.

Ichtchenko, K., Hata, Y., Nguyen, T., Ullrich, B., Missler, M., Moomaw, C., & Südhof, T. C. (1995). Neuroligin 1: a splice site-specific ligand for β-neurexins. Cell, 81(3), 435-443.

Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatter, T., Richard, S., & Scheiffele, P. (2011). SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. Cell, 147(7), 1601-1614.

Iqbal, Z., Neveling, K., Razzaq, A., Shahzad, M., Zahoor, M. Y., Qasim, M., ... & de Brouwer, A. P. (2012). Targeted next generation sequencing reveals a novel intragenic deletion of the TPO gene in a family with intellectual disability. Archives of medical research, 43(4), 312-316.

Iqbal, Z., & Van Bokhoven, H. (2014). Identifying genes responsible for intellectual disability in consanguineous families. Human heredity, 77(1-4), 150-160.

Jo, S., Lee, K. H., Song, S., Jung, Y. K., & Park, C. S. (2005). Identification and functional characterization of cereblon as a binding protein for large –conductance calcium activated potassium channel in rat brain. Journal of neurochemistry, 94(5), 1212-1224.

Joubert, M., Eisenring, J. J., & Andermann, F. (1968). Familial dysgenesis of the vermis: a syndrome of hyperventilation, abnormal eye movements and retardation. Neurology, 18(3), 302.

Katusic, S. K., Colligan, R. C., Beard, C. M., O'Fallon, W. M., Bergstralh, E. J.,Jacobsen, S. J., & Kurland, L. T. (1996). Mental retardation in a birth cohort, 1976-1980, Rochester, Minnesota. American journal of mental retardation: AJMR, 100(4), 335-344.

Katz, G., & Lazcano-Ponce, E. (2008). Intellectual disability: definition, etiological factors, classification, diagnosis, treatment and prognosis. Salud publica de Mexico, 50, s132-41.

Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: a review. Journal of neurodevelopmental disorders, 2(4), 182.

Kennedy, G. C., Matsuzaki, H., Dong, S., Liu, W. M., Huang, J., Liu, G., ... & Liu, W. (2003). Large-scale genotyping of complex DNA. Nature biotechnology, 21(10), 1233.

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., ... & Ishak, G. E. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. The American Journal of Human Genetics, 90(5), 856-863.

Khan, M. A., Khan, S., Windpassinger, C., Badar, M., Nawaz, Z., & Mohammad, R. M. (2016). The molecular genetics of autosomal recessive nonsyndromic intellectual disability: A mutational continuum and future recommendations. Annals of human genetics, 80(6), 342-368.

Kim, H. G., Kishikawa, S., Higgins, A. W., Seong, I. S., Donovan, D. J., Shen, Y., ... & Descartes, M. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. The American Journal of Human Genetics, 82(1), 199-207.

Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46(3), 310.

Kirov, G., Rujescu, D., Ingason, A., Collier, D. A., O'donovan, M. C., & Owen, M. J. (2009). Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophrenia bulletin, 35(5), 851.

Kleefstra, T., Brunner, H. G., Amiel, J., Oudakker, A. R., Nillesen, W. M., Magee, A., ... & Hamel, B. C. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *The* American Journal of Human Genetics, 79(2), 370-377.

Koolen, D. A., Kramer, J. M., Neveling, K., Nillesen, W. M., Moore-Barton, H. L., Elmslie, F. V., ... & Cheung, S. W. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21. 31 microdeletion syndrome. Nature genetics, 44(6), 639.

Kruglyak, L., & Nickerson, D. A. (2001). Variation is the spice of life. Nature genetics, 27(3), 234.

LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic acids research, 37(13), 4181-4193.

Lagae, L. (2008). Learning disabilities: definitions, epidemiology, diagnosis, and intervention strategies. Pediatric Clinics, 55(6), 1259-1268.

Lanciego, J. L., Luquin, N., & Obeso, J. A. (2012). Functional neuroanatomy of the basal ganglia. Cold Spring Harbor perspectives in medicine, 2(12), a009621.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Funke, R. (2001). de la BM, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara .... Initial sequencing and analysis of the human genome. Nature, 409, 860-921.

Ledbetter, D. H., & Martin, C. L. (2007). Cryptic telomere imbalance: A 15-year update. In American Journal of Medical Genetics Part C: Seminars in Medical Genetics (Vol. 145, No. 4, pp. 327-334). Wiley Subscription Services, Inc., A Wiley Company.

Leonard, H., & Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. Developmental Disabilities Research Reviews, 8(3), 117-134.

Lepage, M., & Richer, F. (1996). Inter-response interference contributes to the sequencing deficit in frontal lobe lesions. Brain, 119(4), 1289-1295.

Liu, W., Tanasa, B., Tyurina, O. V., Zhou, T. Y., Gassmann, R., Liu, W. T., ... & Desai, A. (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature, 466(7305), 508.

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. Science, 1162-1164.

Luria, A. R. (1966). Higher cortical functions in man. New York: Consultants Bureau Enterprises. Inc. Google Scholar.

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. Proceedings of the National Academy of Sciences, 107(3), 961-968.

Manya, H., Sakai, K., Kobayashi, K., Taniguchi, K., Kawakita, M., Toda, T., & Endo, T. (2003). Loss-of-function of an N-acetylglucosaminyltransferase, POMGnT1, in muscle–eye–brain disease. Biochemical and biophysical research communications, 306(1), 93-97.

Marieb EN & Hoehn Katja (2010). Human Anatomy & Physiology (8th ed.), pp. 649–50, San Francisco: Benjamin Cummings.

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., ... & Thiruvahindrapduram, B. (2008). Structural variation of chromosomes in autism spectrum disorder. The American Journal of Human Genetics, 82(2), 477-488.

Matson, J. L., Dixon, D. R., Matson, M. L., & Logan, J. R. (2005). Classifying mental retardation and specific strength and deficit areas in severe and profoundly mentally retarded persons with the MESSIER. Research in developmental disabilities, 26(1), 41-45.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. Research in developmental disabilities, 30(6), 1107-1114.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. Research in developmental disabilities, 32(2), 419-436.

Milani, L., Lundmark, A., Nordlund, J., Kiialainen, A., Flaegstad, T., Jonmundsson, G., ... & Syvänen, A. C. (2009). Allele-specific gene expression patterns in primary leukemic cells reveal regulation of gene expression by CpG site methylation. Genome research, 19(1), 1-11.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., ... & Faucett, W. A. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. The American Journal of Human Genetics, 86(5), 749-764.

Mir, A., Kaufman, L., Noor, A., Motazacker, M. M., Jamil, T., Azam, M., ... & Naeem, F. (2009). Identification of mutations in TRAPPC9, which encodes the NIK-and IKK-β-binding protein, in nonsyndromic autosomal-recessive mental retardation. The American Journal of Human Genetics, 85(6), 909-915.

Missler, M., Hammer, R. E., & Südhof, T. C. (1998). Neurexophilin binding to α-Neurexins a single lns domain functions as an independently folding ligand-binding unit. Journal of Biological Chemistry, 273(52), 34716-34723.

Molinari, F., Rio, M., Meskenaite, V., Encha-Razavi, F., Augé, J., Bacq, D., ... & Sonderegger, P. (2002). Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. Science, 298(5599), 1779-1781.

Molinari, F., Foulquier, F., Tarpey, P. S., Morelle, W., Boissel, S., Teague, J., ... & Matthijs, G. (2008). Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. The American Journal of Human Genetics, 82(5), 1150-1157.

Motazacker, M. M., Rost, B. R., Hucho, T., Garshasbi, M., Kahrizi, K., Ullmann, R., ... & Tzschach, A. (2007). A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. The American Journal of Human Genetics, 81(4), 792-798.

Musante, L., & Ropers, H. H. (2014). Genetics of recessive cognitive disorders. Trends in Genetics, 30(1), 32-39.

Najmabadi, H., Motazacker, M. M., Garshasbi, M., Kahrizi, K., Tzschach, A., Chen, W., ... & Vazifehmand, R. (2007). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. Human genetics, 121(1), 43-48.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., ... & Zecha, A. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature, 478(7367), 57.

Naviaux, R. K., & Nguyen, K. V. (2004). *POLG* mutations associated with Alpers' syndrome and mitochondrial DNA depletion. Annals of neurology, 55(5), 706-712.

Nguyen, K. V., Østergaard, E., Ravn, S. H., Balslev, T., Danielsen, E. R., Vardag, A.,& Naviaux, R. K. (2005). *POLG* mutations in Alpers syndrome. Neurology, 65(9), 1493-1495.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics, 12(6), 443.

Parisi MA (2009). Clinical and molecular features of Joubert syndrome and related disorders. American journal of medical genetics Part C, Seminars in medical genetics, 151 C (4), 326-340.

Patel, D. R., Greydanus, D. E., Omar, H. A., & Merrick, J. (Eds.). (2011a). Neurodevelopmental Disabilities: Clinical Care for Children and Young Adults. Springer Science & Business Media.

Patel, D. R., & Merrick, J. (2011b). Neurodevelopmental disabilities: introduction and epidemiology. In Neurodevelopmental Disabilities (pp. 1-13). Springer, Dordrecht.

Piton, A., Redin, C., & Mandel, J. L. (2013). XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. The American Journal of Human Genetics, 93(2), 368-383.

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., ... & Almeida, J. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature, 466(7304), 368.

Rafiq, M. A., Ansar, M., Marshall, C. R., Noor, A., Shaheen, N., Mowjoodi, A., ... & Vincent, J. B. (2010). Mapping of three novel loci for non- syndromic autosomal recessive mental retardation (NSARMR) in consanguineous families from Pakistan. Clinical genetics, 78(5), 478-483.

Rauch, A., Rüschendorf, F., Huang, J., Trautmann, U., Becker, C., Thiel, C., ... & Nürnberg, P. (2004). Molecular karyotyping using an SNP array for genomewide genotyping. Journal of medical genetics, 41(12), 916-922.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., ... & Nürnberg, P. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. American journal of medical genetics Part A, 140(19), 2063-2074.

Ravnan, J. B., Tepperberg, J. H., Papenhausen, P., Lamb, A. N., Hedrick, J., Eash, D., ... & Martin, C. L. (2006). Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. Journal of medical genetics, 43(6), 478-489.

Raymond, F. L. (2006). X linked mental retardation: a clinical guide. Journal of medical genetics, 43(3), 193-200.

Reissner, C., Klose, M., Fairless, R., & Missler, M. (2008). Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. Proceedings of the National Academy of Sciences, 105(39), 15124-15129.

Resta, R. G. (1993). The crane's foot: The rise of the pedigree in human genetics. Journal of genetic counseling, 2(4), 235-260.

Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D. L., ... & Rashid, M. (2017). Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. Molecular psychiatry, 22(11), 1604.

Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., ... & Shendure, J. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science, 328(5978), 636-639.

Roeleveld, N., & Zielhuis, G. A. (1997). The prevalence of mental retardation: a critical review of recent literature. Developmental Medicine & Child Neurology, 39(2), 125-132.

Romine, C. B., & Reynolds, C. R. (2004). Sequential memory: a developmental perspective on its relation to frontal lobe functioning. Neuropsychology review, 14(1), 43-64.

Romani, M., Micalizzi, A., & Valente, E. M. (2013). Joubert syndrome: congenital cerebellar ataxia with the molar tooth. The Lancet Neurology, 12(9), 894-905.

Roosing, S., Hofree, M., Kim, S., Scott, E., Copeland, B., Romani, M., ... & Miccinilli, E. (2015). Functional genome-wide siRNA screen identifies KIAA0586 as mutated in Joubert syndrome. Elife, 4.

Ropers, H. H., & Hamel, B. C. (2005). X-linked mental retardation. Nature reviews genetics, 6(1), 46.

Ropers, H. H. (2006). X-linked mental retardation: many genes for a complex disorder. Current opinion in genetics & development, 16(3), 260-269.

Ropers, H. H. (2008). Genetics of intellectual disability. Current opinion in genetics & development, 18(3), 241-250.

Ropers, H. H. (2010). Genetics of early onset cognitive impairment. Annual review of genomics and human genetics, 11, 161-187.

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

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., ... & Dubas, F. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nature genetics, 38(1), 24.

Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O. P., Barnes, M. R., Toulopoulou, T., ... & Murray, R. (2008). Disruption of the neurexin 1 gene is associated with schizophrenia. Human molecular genetics, 18(5), 988-996.

Saleha, S., Sajid, M., Zafar, S., & Pervaiz, N. (2017). Genetic Defects Underlie the Non-syndromic Autosomal Recessive Intellectual Disability (NS-ARID). Open Life Sciences, 12(1), 167-177.

Sambrook, J., & Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harbor Protocols, 2006(1), pdb-prot4455.

Schuurs-Hoeijmakers, J. H., Hehir-Kwa, J. Y., Pfundt, R., Van Bon, B. W., De Leeuw, N., Kleefstra, T., ... & Van Bokhoven, H. (2011). Homozygosity mapping in outbred families with mental retardation. European Journal of Human Genetics, 19(5), 597.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., ... & Navin, N. (2004). Large-scale copy number polymorphism in the human genome. Science, 305(5683), 525-528.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., ... & Leotta, A. (2007). Strong association of de novo copy number mutations with autism. Science, 316(5823), 445-449. Weiss

Seelow, D., & Schuelke, M. (2012). HomozygosityMapper2012—bridging the gap between homozygosity mapping and deep sequencing. Nucleic acids research, 40(W1), W516-W520.

Siddiqui, T. J., Pancaroglu, R., Kang, Y., Rooyakkers, A., & Craig, A. M. (2010). LRRTMs and neuroligins bind neurexins with a differential code to cooperate in glutamate synapse development. Journal of Neuroscience, 30(22), 7495-7506.

Simon-Sanchez, J., Scholz, S., del Mar Matarin, M., Fung, H. C., Hernandez, D., Gibbs, J. R., ... & Singleton, A. (2008). Genomewide SNP assay reveals mutations underlying Parkinson disease. Human mutation, 29(2), 315-322.

Simon, T. J., Bish, J. P., Bearden, C. E., Ding, L., Ferrante, S., Nguyen, V., ... & Emanuel, B. S. (2005). A multilevel analysis of cognitive dysfunction and psychopathology associated with chromosome 22q11. 2 deletion syndrome in children. Development and psychopathology, 17(3), 753-784.

Stiles, J. (2008). The fundamentals of brain development: Integrating nature and nurture. Harvard University Press.

Strømme, P., & Hagberg, G. (2000). Aetiology in severe and mild mental retardation: a population-based study of Norwegian children. Developmental medicine and child neurology, 42(2), 76-86.

Suckow, A. T., Comoletti, D., Waldrop, M. A., Mosedale, M., Egodage, S., Taylor, P., & Chessler, S. D. (2008). Expression of neurexin, neuroligin, and their cytoplasmic binding partners in the pancreatic β-cells and the involvement of neuroligin in insulin secretion. Endocrinology, 149(12), 6006-6017.

Sudhof, T. C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. Nature, 455(7215), 903.

Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X. Q., ... & Feuk, L. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nature genetics, 39(3), 319.

Taniguchi, K., Kobayashi, K., Saito, K., Yamanouchi, H., Ohnuma, A., Hayashi, Y. K., ... & Falsaperla, R. (2003). Worldwide distribution and broader clinical spectrum of muscle–eye–brain disease. Human molecular genetics, 12(5), 527-534.

Teer, J. K., & Mullikin, J. C. (2010). Exome sequencing: the sweet spot before whole genomes. Human molecular genetics, 19(R2), R145-R151.

Topper, S., Ober, C., & Das, S. (2011). Exome sequencing and the genetics of intellectual disability. Clinical genetics, 80(2), 117-126.

Turner, G. (1996). Finding genes on the X chromosome by which homo may have become sapiens. American journal of human genetics, 58(6), 1109.

Ushkaryov, Y. A., Petrenko, A. G., Geppert, M., & Sudhof, T. C. (1992). Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science, 257(5066), 50-56.

Valenti, D., de Bari, L., De Filippis, B., Henrion-Caude, A., &Vacca, R. A. (2014). Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: an overview of Down syndrome, autism, Fragile X and Rett syndrome. Neuroscience & Biobehavioral Reviews, 46, 202-217.

Van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics, 45, 81-104.

Van Karnebeek, C. D., Jansweijer, M. C., Leenders, A. G., Offringa, M., & Hennekam, R. C. (2005). Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. European journal of human genetics, 13 (1), 6-25.

Van Karnebeek, C. D., & Stockler, S. (2012). Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review. Molecular genetics and metabolism, 105(3), 368-381.

Venkatesan, C., Angle, B., & Millichap, J. J. (2016). Early-life epileptic encephalopathy secondary to SZT2 pathogenic recessive variants. Epileptic Disorders, 18(2), 195-200.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Gocayne, J. D. (2001). The sequence of the human genome. science, 291(5507), 1304-1351.

Vilboux, T., Doherty, D. A., Glass, I. A., Parisi, M. A., Phelps, I. G., Cullinane, A. R., ... & Oden, N. L. (2017). Molecular genetic findings and clinical correlations in 100 patients with Joubert syndrome and related disorders prospectively evaluated at a single center. Genetics in Medicine, 19(8), 875.

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

Vrijenhoek, T., Buizer-Voskamp, J. E., van der Stelt, I., Strengman, E., Sabatti, C., van Kessel, A. G., ... & Veltman, J. A. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. The American Journal of Human Genetics, 83(4), 504-510.

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., ... & Stray, S. M. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. science, 320(5875), 539-543.

Weiss, L. A., Shen, Y., Korn, J. M., Arking, D. E., Miller, D. T., Fossdal, R., ... & Platt, O. S. (2008). Association between microdeletion and microduplication at 16p11. 2 and autism. New England Journal of Medicine, 358(7), 667-675.

Wolfer, D. P., Lang, R., Cinelli, P., Madani, R., & Sonderegger, P. (2001). Multiple roles of neurotrypsin in tissue morphogenesis and nervous system development suggested by the mRNA expression pattern. Molecular and Cellular Neuroscience, 18(4), 407-433.

World Health Organization. (2001). International Classification of Functioning, Disability and Health: ICF. World Health Organization.

Wisniowiecka-Kowalnik B., Nesteruk, M., Peters, S. U., Xia, Z., Cooper, M. L., Savage, S., ... & Duda, A. W. (2010). Intragenic rearrangements in NRXN1 in three families with autism spectrum disorder, developmental delay, and speech delay. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 153(5), 983-993.

Yamamura, Y., Yamashiro, K., Tsuruoka, N., Nakazato, H., Tsujimura, A., & Yamaguchi, N. (1997). Molecular Cloning of a Novel Brain-Specific Serine Protease with a Kringle-like Structure and Three Scavenger Receptor Cysteine-Rich Motifs1. Biochemical and biophysical research communications, 239(2), 386-392.

Yin, Y., Bangs, F., Paton, I. R., Prescott, A., James, J., Davey, M. G., ... & Tickle, C. (2009). The Talpid3 gene (KIAA0586) encodes a centrosomal protein that is essential for primary cilia formation. Development, 136(4), 655-664.

Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., Mizuno, M., ... & Herrmann, R. (2001). Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1. Developmental cell, *1*(5), 717-724.

Zahir, F., & Friedman, J. M. (2007). The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility. Clinical genetics, 72(4), 271-287.

Zahir, F. R., Baross, A., Delaney, A. D., Eydoux, P., Fernandes, N. D., Pugh, T., ... & Friedman, J. M. (2008). A patient with vertebral, cognitive and behavioural

abnormalities and a de novo deletion of NRXN1α. Journal of medical genetics, 45(4), 239-243.

Zahrani, F., Aldahmesh, M. A., Alshammari, M. J., Al-Hazzaa, S. A., & Alkuraya, F. S. (2013). Mutations in c12orf57 cause a syndromic form of colobomatous microphthalmia. The American Journal of Human Genetics, 92(3), 387-391.

Zhang, W., Cui, H., & Wong, L. J. C. (2012). Application of next generation sequencing to molecular diagnosis of inherited diseases. In Chemical Diagnostics (pp. 19-45). Springer Berlin Heidelberg.

Zurflüh, M. R., Zschocke, J., Lindner, M., Feillet, F., Chery, C., Burlina, A., & Blau, N. (2008). Molecular genetics of tetradydrobiopterin-responsive phenylalanine hydroxylase deficiency. Human mutation, 29 (1), 167-175.

Zweier, C., de Jong, E. K., Zweier, M., Orrico, A., Ousager, L. B., Collins, A. L., ... & Schenck, A. (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. The American Journal of Human Genetics, 85(5), 655-666.

Electronic References

Online Mendelian Inheritance in Man (OMIM; www.ncbi.nlm.nih.gov/omim)

Centres for Disease Control and Prevention (http://www.cdc.gov/ncbddd/autism, 2009)

Ensemble Genome Browser (http://asia.ensemble.org/index.html)

Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/)

Blast like alignment tool (https://genome.ucsc.edu/cgi-bin/hgBlat)

UCSC genome browser (https://genome.ucsc.edu/)

In-silico PCR (https://genome.ucsc.edu/cgi-bin/hgPcr)

Golden Helix Genome Browse 2.1.2 (http://goldenhelix.com/products/GenomeBrowse/)

CLC Genomics Workbench (http://www.clcbio.com/blog/clc-genomics-workbench-7-5/)

Sorting Intolerant from Tolerant (SIFT) (http://sift.jcvi.org/)

Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/)

Huaman Splicing Finder (http://www.umd.be/HSF3/HSF.html)

Mutation Taster (http://www.mutationtaster.org/)

1000 Genomes Browser (http://browser.1000genomes.org/index.html)

ExAC (Exome aggregation Consortium) Browser (http://exac.broadinstitute.org/)

dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)