Genetic Basis of Intellectual Disability and Schizophrenia in Selected Omani and UK Families

Ahmed Hamed Hamood Al Amri

BSc (Hons), MSc Medical Microbiology

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

University of Leeds
School of Biomedical Sciences
Faculty of Biological Sciences

August , 2017

The candidate confirms that the work submitted is his/her/their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Jointly authored publications statement

Publications from this thesis

Chapter 3 of this thesis is entirely the work of the author and appears in:

Al-Amri A., Saegh A. A., Al-Mamari W., El-Asrag M. E., Ivorra J. L., Cardno A. G., Inglehearn C. F., Clapcote S. J. and Ali M. 2016. Homozygous single base deletion in TUSC3 causes intellectual disability with developmental delay in an Omani family. American Journal of Medical Genetics Part A.

Al-Amri A., Saegh A. A., Al-Mamari W., El-Asrag M. E., Inglehearn C. F., Clapcote S. J. and Ali M. 2017. Novel LHFPL5 mutation causes non-syndromic hearing loss in an Omani family. BMC Medical Genetics. **In press**

Al-Amri A., Saegh A. A., Al-Mamari W., El-Asrag M. E., Inglehearn C. F., Clapcote S. J. and Ali M. 2017. Evidence supporting the insinuation of *ANKRD2* and *PDZD8* in the pathology of intellectual disability; indication from an Omani family. **In preparation**

J L Ivorra, A Al-Amri, M Ali, J Dachtler, A G Cardno, C Logan, J G L Mullins, T Mahmood, A Tsatsanis, Q Nazar, S Shora, D Du Toit, R Jones, A Abbas, K Khan, S Khan, C A Johnson, J A Duce, M Zhang, E van Wijk, S J Clapcote, C F Inglehearn, Functional characterization of a rare DFNB31 variant and evidence for involvement in schizophrenia. **In preparation**

Mahmood T, Al-Amri A, Poulter J, Ali M, Logan C, Khan S, Johnson C, Cardno AG, Ahmed S, Nazari J, Wilkinson I, Woodruff P, Clapcote S and Inglehearn C. Genetic Imaging of a Consanguineous Family with Schizophrenia Multiplex and a 13q31 Susceptibility Locus. **In preparation**

Publications related to other work during this study

Mcgirr A., Lipina T. V., Mun H.-S., Georgiou J., Al-Amri A. H., Ng E., Zhai D., Elliott C., Cameron R. T. and Mullins J. G. 2016. Specific inhibition of phosphodiesterase-4B results in anxiolysis and facilitates memory acquisition. Neuropsychopharmacology, 41, 1080-1092.

Zollo M., Ahmed M., Ferrucci V., Salpietro V., Asadzadeh F., Carotenuto M., Maroofian R., Al-Amri A., Singh R., Scognamiglio I., Mojarrad M., Musella L., Duilio A., Di Somma A., Karaca E., Rajab A., Al-Khayat A., Mohan Mohapatra T., Eslahi A., Ashrafzadeh F., Rawlins L. E., Prasad R., Gupta R., Kumari P., Srivastava M., Cozzolino F., Kumar Rai S., Monti M., Harlalka G. V., Simpson M. A., Rich P., Al-Salmi F., Patton M. A., Chioza B. A., Efthymiou S., Granata F., Di Rosa G., Wiethoff S., Borgione E., Scuderi C., Mankad K., Hanna M. G., Pucci P., Houlden H., Lupski J. R., Crosby A. H. and Baple E. L. 2017. PRUNE is crucial for normal brain development and mutated in microcephaly with neurodevelopmental impairment. Brain, 140, 940-952.

Acknowledgements

Abstract

Intellectual disability (ID) is devastating condition which is defined using three criteria: reduced intellectual ability, deficit in two or more adaptive behaviours, and diagnosis before the age of 18 years. ID can have various causes, but genetic factors are thought to be responsible for up to 50% of cases. ID is a heterogeneous and complex disorder, and more than 800 genes have been implicated in its pathology. Schizophrenia (SZ) is another complex neurodevelopmental condition that also affects the brain and has a partially overlapping genetic basis with ID. This thesis describes work carried out into the genetic basis of ID and SZ.

The ID project was focused on ID in consanguineous families recruited in Oman. Next generation sequencing allowed the identification of apparently causative mutations in three out of the six families recruited. The mutations are all novel, although some of them occur in previously associated ID genes. The known ID genes in which novel mutations were identified are *TUSC3* (NM_006765: exon2:c.222delA, p.R74 fs) and *NHS* (NM_198270:exon8: c.C4385G, p.S1462C). A novel *LHFPL5* variant (NM_182548:exon2:c.T575C, p.L192P) was also identified in an ID family with hearing loss. In one ID family, mutations in two genes not previously associated with ID were found: *ANKRD2* (NM_001129981:exon8:c.C883T, p.R295W) and *PDZD8* (PDZD8:NM_173791: exon5:c.2197_2200del, p.733_734del).

The SZ project was focused on two SZ families recruited in the UK. Preliminary work on these families had suggested the involvement of homozygosity for genetic variants in the *DFNB31* gene and in a region of Chr13q in the pathogenesis of schizophrenia. Further experiments to validate the initial findings included testing the overexpression of the *DFNB31* variant (R450C) in the SH-SY5Y cell line, pull-down assay and transcript analysis of the genes located at Chr13q.

Identification of the causative mutated gene is an important step in understanding more about the biology of ID and SZ. It also facilitates carrier testing and genetic counselling, and identifies a pathway for potential therapeutic intervention.

Table of Contents

List of Figures

List of Tables

Abbreviations

1000g	1000 genome project/database
Ab	Antibody
ABA	applied behaviour analysis
ABC	transporters ATP-binding cassette
ACC	anterior cingulate cortex
ACOD1	Aconitate Decarboxylase 1
AD	Alzheimer's disease
ADID	autosomal dominant Intellectual disability
ANKRD2	Ankyrin repeat domain 2
APP	Amyloid beta a4 precursor protein gene
AS	Angelman syndrome
ASD	Autism spectrum disorder
ASPM	Abnormal spindle-like, microcephaly-associated gene
BBS	Bardet-biedl syndrome
BD	Binding domain
BF	Bayes factor
BFLS	Borjeson-forssman-lehmann syndrome
bp	Base pair
BPCA	Blocking peptide competition assay
BPD	Bipolar disorder
BSA	Bovine serum albumin
C4	Complement component 4 gene
CADD	Combined annotation dependent depletion
Cas9	CRISPR-associated protein-9 nuclease
CASC5	Cancer susceptibility candidate gene 5 protein gene
CDCV	Common disease, common variant hypothesis
CDK5	Cyclin-dependent kinase 5 gene
cDNA	Complementary DNA
CDRV	Common disease, rare variant hypothesis
CDS	Coding sequence
CENPJ	Centromeric protein J gene
CFB	Complement factor B gene
CFH	Complement factor H gene
CFTR	Cystic fibrosis transmembrane conductance regulator gene
CLN5	Ceroid-lipofuscinosis, neuronal 5
CMV	Cytomegalovirus
CNV	Copy number variants
CR	Cognitive remediation
CRBN	Cereblon gene
CRISPR	Clustered regulatory interspaced short palindromic repeats
CRT	Cyclic reversible termination
cVA	cis-vaccenyl acetate
DA	Dopamine

DAPI	4,6-diamidino-2-phenylindole
dbSNP	Database of Single Nucleotide Polymorphisms
DBT	Dialectical behaviour therapy
ddH2O	Double-distilled water
DFNB31	Deafness, Autosomal Recessive 31 gene
DISC1	Disrupted-in-schizophrenia-1 gene
Dlg1	*Drosophila* disc large tumor suppressor gene
DLPFC	Dorsolateral prefrontal cortex
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's Phosphate Buffered Saline
DPX	Mixture of distyrene, plasticiser and xylene
DS	Down's syndrome
dsDNA	Double stranded deoxyribonucleic acid
DSI	Diffusion spectrum imaging
DSM	Diagnostic and statistical manual of mental disorders
DTI	Diffusion tensor imaging
EDNRB	Endothelin receptor, Type B
EDTA	Ethylene diamine tetra acetic acid
EFNB2	Ephrin B2 gene
EHR	Electronic health records
EIF4G1	Eukaryotic translation initiation factor 4-gamma 1 gene
ENCODE	Encyclopedia of DNA elements project
EP	Extremely preterm
ER	Endoplasmic reticulum
ERT	Enzyme replacement therapy
EVS	Exome variant server
ExAC	Exome aggregation consortium
F	Forward
FAS	Fetal alcohol syndrome
FASD	Fetal alcohol spectrum disorder
FBXL3	F-Box and leucine-rich repeat protein 3
FIN	Finnish
FMR1	Fragile X mental retardation 1 gene
fMRI	Functional magnetic resonance imaging
FXS	Fragile X Syndrome
FXTAS	Fragile X-associated tremor/ataxia syndrome
Gal-4	Galactose-responsive transcription factor 4
GATK	Genome analysis tool kit
GBA	Glucosidase beta acid gene
GD	Gaucher disease
gDNA	Genomic deoxyribonucleic acid
GOLD	Genome online databases
GT	Glutamate

GWAS	Genome-wide association studies
HBD	Homozygous by descent
HER2	Human epidermal growth factor receptor 2
HMGIC	High mobility group isoform I-C
HMW	High molecular weight
HOXB1	Homeobox B1 gene
HTRA1	HTRA Serine peptidase 1 gene
IBD	Identity by descent
ICCG	International collaboration for clinical genomics
ICD	International classification of diseases
ID	Intellectual disability
IF	Immunofluorescence
IGV	Integrative genomics viewer
IHC	Immunohistochemistry
IL1RAPL1	Interleukin 1 receptor accessory protein-like 1 gene
iPS	Induced pluripotent stem cells
IQ	Intelligence quotient
JBTS	Joubert Syndrome
JSRD	Joubert syndrome and related disorders
KCTD12	Potassium channel tetramerization domain containing protein 12
kDa	Kilo Dalton
KIF17	Kinesin family member 17 gene
KLC2	Kinesin light chain 2 gene
KO	Knockout
LHFPL5	Lipoma HMGIC fusion partner-like 5
LOD	Logarithm of the odds
LOF	Loss of function
LPA	Lysophosphatidic acid
LRRK2	Leucine-rich repeat kinase 2 gene
LWH	Leucine, tryptophan and Histidine
MAF	Minor allele frequency
MAMDC1	MAM domain-containing glycosylphosphatidylinositol anchor 1
MAPT	Microtubule-associated protein tau gene
MCH	4-methylcyclohexanol
MCPH1	Microcephalin gene
MD	Middle portion
MECP2	Methyl-CpG-Binding Protein 2 gene
MFS	Marfan syndrome
MgCl2	Magnesium chloride
mGLu5	Metabotropic glutamate receptor subtype 5 inhibitor
MHC	Major histocompatibility complex
MOH	Ministry of health
mRFP	Monomer red fluorescent protein
Mut	Mutant
MYCBP2	Myc-binding protein 2

NA	Not available
NaCl	Sodium chloride
NCBI	National centre for biotechnology information
ND	Not done
NDFIP2	Nedd4 family-interacting protein 2
NGS	Next generation sequencing
NHS	Nance-horan syndrome
NMD	Nonsense mediated decay
NMDAR	N-methyl-D-aspartate receptor
NPAS3	Neuronal Pas Domain Protein 3 gene
NPCs	Neural progenitor cells
NRES	National Research Ethics Service
NRG1	Neuregulin 1 gene
OCT	3-octanol
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline - tween
PCR	Polymerase chain reaction
PDT	Psychodynamic therapy
PDZ	Three proteins; PSD95, Dlg1, Zo-1
PDZD8	PDZ domain containing 8
PFA	Paraformaldehyde
PHC1	Polyhomeotic-like 1 gene
PINK1	Pten-induced putative kinase gene
PM	Personalized medicine
Polyphen-2	Polymorphism phenotyping version-2
POU4F1	Pou domain, class 4, transcription Factor 1
PRKN	Parkin gene
PSD95	Post synaptic density protein 95
PSEN1	Presenilin gene
PSI-BLAST	Position-specific iterative basic local alignment search tool
PVDF	Polyvinylidene difluoride
PWS	Prader-Willi syndrome
QC	Quality control
R	Reverse
RAP2	Regulatory subunit-associated protein 2 gene
RBM26	RNA binding motif protein 26
RFLP	Restriction fragment length polymorphism
RING	Rapid iterative negative geotaxis
RNA	Ribonucleic acid
RNF219	Ring finger protein 219
RT	Room temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
RTT	Rett syndrome

RVIS	Residual variation intolerance score
SBL	Sequencing by ligation
SBS	Sequence by synthesis
SCEL	Sciellin
SCS	Single-cell sequencing
SDM	Site-directed mutagenesis
SETD1A	Set domain-containing protein 1A gene
SHANK3	Sh3 and multiple ankyrin repeat domains 3 gene
SHL	Sensorial hearing loss
SIFT	Sorting intolerant from tolerant tool
SLAIN1	Slain motif family member 1
sMRI	Structural MRI
SNA	Single nucleotide addition
SNCA	Synuclein alpha gene
SNP	Single nucleotide polymorphism
SPTRY2	Sprouty RTK signaling antagonist 2
SQUH	Sultan Qaboos university hospital
STIL	Scl/Tal1-Interrupting Locus gene
STRs	Short tandem repeats
SYA	Sugar yeast agar
SZ	Schizophrenia
TBE	Tris borate EDTA
TBI	Traumatic brain injury
TBS	Tris-buffered saline
TE	Tris and EDTA buffer
TIRF	Total internal reflection fluorescence
TUSC3	Tumor suppressor candidate 3
UAS	Upstream activation sequence
UBR4	Ubiquitin protein ligase E3 component N-recognin 4
UCSC	University of California santa cruz
ug	Microgram
µl	Microlitre
USD	United States dollar
UTRs	Untranslated regions
UV	ultra-violet
VCF	Variant call format
VDSC	Vienna *Drosophila* source centre
VEGA	Vertebrate genome annotation
VEP	Variant effect predictor
WAIS	Wechsler Adult Intelligence Scale
WB	Western Blotting
WDR62	WD Repeat-Containing Protein 62 gene
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World health organisation

WT	Wild type
XL	X-Linked
XLID	X-linked intellectual disability
Y2H	Yeast two hybrid
ZNF335	Zinc Finger Protein 335 gene
Zo-1	Zonula occludens-1 protein

Chapter 1: General Introduction

1.1 Preface

The human brain regulates many of the highly complex life processes and controls the unique behaviours associated with human thought, memory and ingenuity, including feeling, thinking, reasoning and perception. Our brains not only work when we are awake, they are also active while we sleep (Walker and Robertson, 2016). This impressive organ has a cohesive structure involving specialized cells called neurons which are the basic units responsible for processing and transmitting signals and sensory information rapidly throughout the central nervous system (Melillo and Leisman, 2010). Neuronal cells do not look exactly alike as they have different structural themes but they generally share the basic structure of the cell body, dendrites and axons (Figure 1.1). Dendrites are specialised branched projections found at the end of neurons and their function is to receive and propagate the incoming information from other neurons. The information is received as signals at the synapse where the neurotransmitters convey the message. It is then the role of the axons, which are slender-forming cable-like extensions throughout the nervous system, to broadcast the neuronal impulse to other cells (Qu et al., 2017). There are about 100 billion neurons in the adult brain and each of them has the ability to establish up to 10 thousand synaptic connections with other neurons (Herculano-Houzel, 2009).

The dense wiring of the interconnected network makes the brain a highly efficient and complex system, but its soft and delicate tissues make the brain vulnerable to trauma. Because of that, the brain is highly protected by the hard skull bones and also surrounded by the cerebrospinal fluid which provides structural support and extra protection. The number of genes involved in neurodevelopment is more than required to build any other organs like heart or liver. The precise number genes is not actually known but studies suggest that 80% of genes in the human genome are expressed in the brain (Jain, 2013).
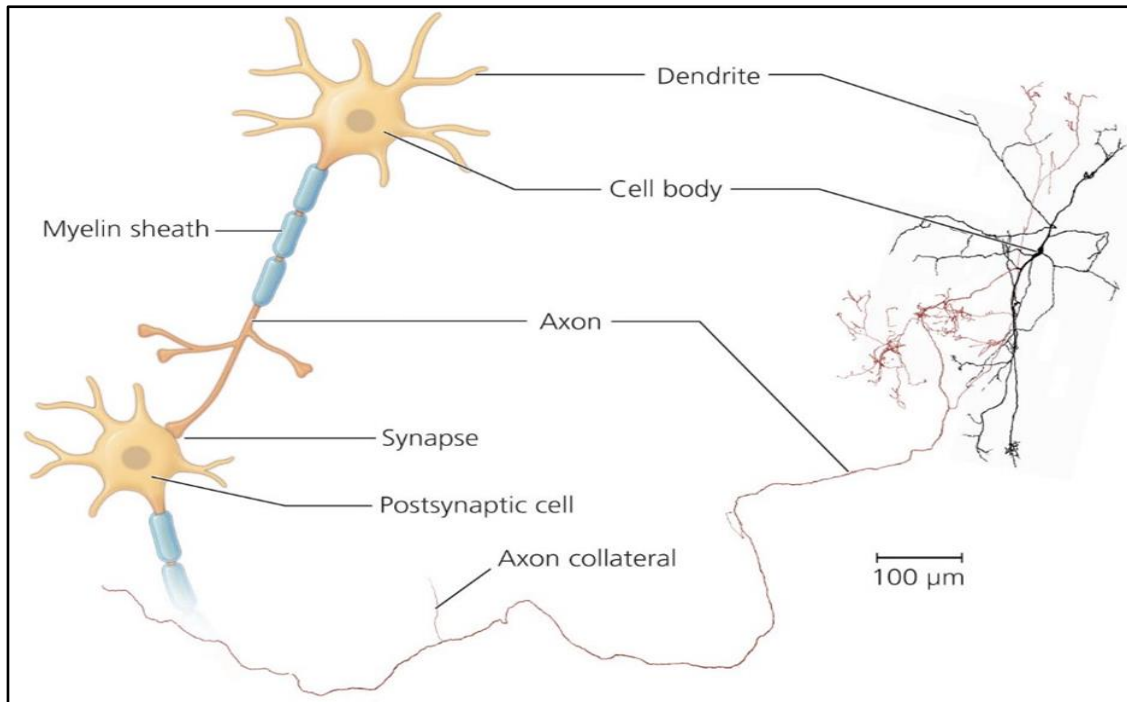
Figure 1.1. Structure of neurons: schematic and real-tracing images. On the left is a picture of a neuron with its multiple branched dendrites. This neuron is firing into another neuron through the axon which forms the synapse. To the right is an outlining of a real neuron when visualised after labelling it with molecules of biocytin, as described by Lee Christian R and Tepper (2007). The body cell and the dendrites are shown in black. The axon is shown in red. (Adapted from Striedter (2015) with written permission).

Not all the genes are expressed at the same time; different genes are expressed at specific developmental stages (Harris, 2015). Pre-natal and post-natal environmental factors can also have important effects on the development of the brain (Homberg et al., 2016). The pre-natal factors include nutrition, infection, and maternal stress during pregnancy (Wingenfeld and Wolf, 2011). After birth, the brain becomes more plastic at different stages of neurodevelopment (Figure 1.2) (Hensch and Bilimoria, 2012). A complex interaction of environmental influences and genetic factors plays a role in the process of neurodevelopment. Given the many genes involved in neurodevelopment and the importance of genetic factors in the maturation and development of the human brain, it is not uncommon for neurodevelopment to be disrupted by genetic errors.
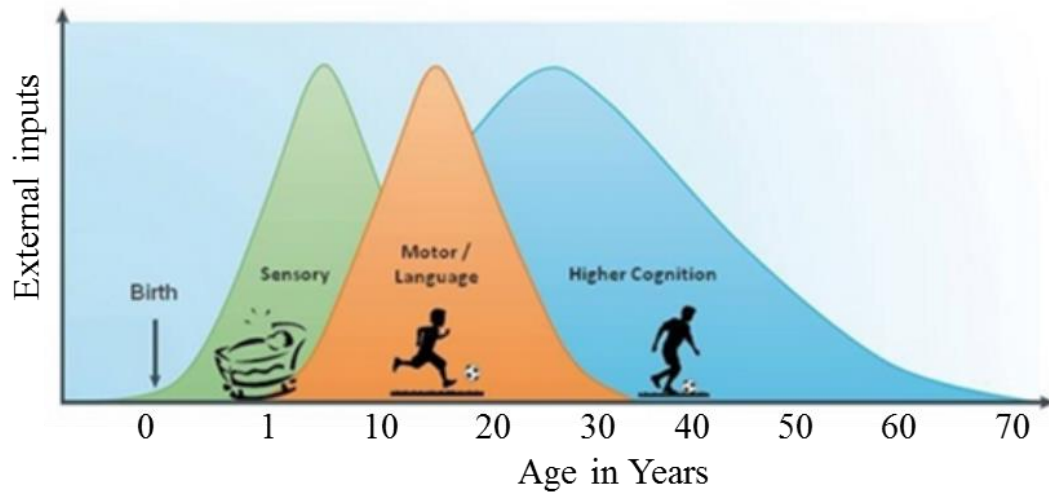
Figure 1.2 Intervals of external inputs essential in enhancing brain plasticity. After birth, there are sensory inputs which include hearing and seeing effects. Subsequently, the motor/language inputs occur and these mainly incorporate perception, movement and speech skills. The third input is crucial to supporting the brain circuit related to developing higher cognitive functions and problem solving. There are no distinct borders between these inputs in which the brain has peaks of plasticity. (Adapted from Hensch and Bilimoria (2012), no permission required).

Most genetic mutations seem to affect the ability of neurons to send signals to other neurons (Grant, 2012; Picker and Walsh, 2013) but there are multiple ways by which brain function can be impaired. Depending on the genes involved and the types of mutation, the impairment of neurodevelopment can range from mild, to moderate, to severe, with or without additional co-morbid clinical features. For example, nonsense mutations in the *CRBN* gene (Cereblon OMIM 609262) cause mild intellectual disability (ID) with no other abnormalities (Kaufman et al., 2010; Xu Guoqiang et al., 2013) whereas mutations in the *PHF6 gene (*Plant HomeoDomain (PHD)-like finger 6, OMIM 300414) *cause* Borjeson-Forssman-Lehmann syndrome (BFLS), which is characterized by a severe type of ID, especially in males, with other manifestations including several dysmorphic features, underdeveloped genitalia and obesity (Berland et al., 2011; Gecz et al., 2006; Jahani-Asl et al., 2016). Sometimes, the gene variants do not directly affect general cognitive ability but disrupt particular thought processes and perception, resulting in psychiatric symptoms such as delusion and hallucination. Examples include mutations in *NRG1* (Neuregulin 1, OMIM 142445), which is linked to schizophrenia, major depressive disorder and bipolar disorder (Barz et al., 2015;

3

Wen Zujia et al., 2016) and *C4*, which was recently associated with schizophrenia (Mayilyan et al., 2008; Sekar et al., 2016).

Identification of the causative genes underlying rare familial forms of these disorders will provide a better understanding of disease pathogenesis and will inform our understanding of the possible causes of more common, sporadic forms of these conditions. For example, the application of linkage analysis to families with Alzheimer's disease (AD) in the 1990s led to the key finding that mutated *APP* (Amyloid Beta a4 Precursor protein OMIM 104760) and *PSEN1* (Presenilin 1, OMIM 104311) are responsible for the development of AD. Without this discovery, it would not have been possible to formulate the prevailing "amyloid cascade hypothesis" which explains the plaque formation in AD pathogenesis (Mohamed et al., 2016). Similarly, the mapping and discovery of *SNCA* (Synuclein, Alpha, OMIM 163890) variants in Parkinson's disease (PD) families (Polymeropoulos et al., 1996; Polymeropoulos et al., 1997) was a milestone in finding an early understanding of the molecular changes associated with PD. With more research and advances in technology, mutations in many other genes, such as *LRRK2* (Leucine-Rich Repeat Kinase 2, OMIM 609007), *PRKN* (Parkin, OMIM 602544), *MAPT* (Microtubule-Associated Protein Tau, OMIM 157140), *EIF4G1* (Eukaryotic Translation Initiation Factor 4-Gamma 1, OMIM 600495) and *PINK1* (Pten-Induced Putative Kinase 1, OMIM 608309) have been implicated in PD pathogenesis (Bekris et al., 2010; Trinh and Farrer, 2013) which resulted in a better understanding of PD.

A better understanding of pathogenic processes will also allow new potential therapeutic targets to be identified. Despite the availability of some educational and management plans to support families with neurodevelopmental disorders, there is much scope for more efficient therapies. A major challenging issue related to the success of the available therapy in neurodevelopmental disorders is the need to identify them very early since they usually begin before birth. Even when applying a particular process to correct an associated phenotype, there are still some complications. For example, it was found that using posterior spinal fusion with unit rod instrumentation could resolve the twisted spinal cord

(scoliosis) in Rett syndrome (RTT, OMIM 312750), one of the syndromic ID caused by mutations in *MECP2* (Methyl-CpG-Binding Protein 2, OMIM 300005)*,* but a high rate of postoperative medical complications is normally expected (Gabos et al., 2012). It is true that some studies on FMR1 knock-out mice have shown worthy amendments in several abnormalities of FXS (Fragile X Syndrome, OMIM 300624) when using mGLuR5 (Metabotropic glutamate receptor subtype 5) inhibitors (Dölen et al., 2007; Michalon et al., 2012) but phase 2 clinical trials with the mGluR5 inhibitors basimglurant and mavoglurant reported no improvement in the clinical phenotype of patients with fragile X syndrome (Scharf et al., 2015).

The search for an effective treatment for neurodevelopmental disorder is accelerating and various techniques are still on trial phases in animal as well as in humans. Progress has already been achieved in some of the conditions with intellectual disabilities. For example, Eisengart and colleagues (Eisengart et al., 2013) found that introducing enzyme replacement therapy (ERT) intravenously could assistance the cognitive ability of patients with Hurler syndrome (Mucopolysaccharidosis type IH, OMIM 607014), which is caused by deficiency in α-L-iduronidase. Furthermore, various trials of different therapeutic approaches to improve cognitive impairment of neurodevelopmental diseases are also ongoing and these include microRNA, stem cell and gene therapy (Picker and Walsh, 2013). More details about the different therapeutic strategies, that have been carried out to find a promising treatment, can be found in chapter 5 (Section 5.4).

Consanguineous families have been good sources to study the genetics of inherited diseases. A consanguineous family is one that results from marriage between two individuals belonging to a recent common ancestor. Because of different reasons including economic benefit, educational level as well as cultural causes, consanguineous marriages have been practiced since the early existence of modern humans (Fuster and Colantonio, 2004; Hamamy, 2012; Na'amnih et al., 2014). Modell and Darr (2002) mention that 20% of the world population lives in communities with a preference for consanguineous marriage.

According to (Bittles and Black, 2010), not less than one billion individuals in the global population live in societies with a preference for consanguineous marriage.

In Arab populations, consanguinity is a long tradition that is a deeply rooted custom because of socio-cultural factors among Muslim communities (Bittles, 2008; Bittles and Black, 2010; Na'amnih et al., 2014). A study by (Sharkia et al., 2016), carried on Arab community in Israel, found the highest rate of consanguineous marriage was among Muslims, followed by Druze, with the lowest rate in Christians. Compared with its very small ratio in the Western population (~1%), Arabs have much higher percentage range consanguineous marriages, which is between 25% and 55% (Sharkia et al., 2016). Many Arab countries display some of the highest rates of consanguineous marriage in the world, specifically first cousin marriages, which are the most common type of consanguineous union. According to Al-Gazali et al. (2006), despite the increased risk of autosomal recessive diseases, congenital malformations and mental retardation, the overall opinion in the Middle East remains in favour of consanguineous marriage.

Oman, which is the second largest territory in the Arabian Peninsula, has a relatively high rate of consanguineous marriage compared with other countries of the world (Rajab Anna et al., 2013). One large scale survey, which included all sections of the Omani population, found that 24.1% of all marriages were between first cousins and 11.8% were between second cousins, while a further 20.4% of marriages were contracted within specific tribal groupings (Rajab A and Patton, 2000). Another study, by Al-Gazali et al. (2006) comparing the rate of first-cousin marriages among Arab countries, found that Oman has a high rate of about 25%. The offspring of consanguineous families are normally at a higher risk of inherited disorders, so this type of family is of great importance in studying the underlying genetics. Consanguineous marriage increases the chance that both members of a couple will carry any recessive variant that is being transmitted in their family, and that this will manifest in the homozygous state in their children (Modell and Darr, 2002).

1.2 Intellectual disability (ID)

1.2.1 ID overview

There are different overlapping terms for intellectual disability including mental handicap, mental retardation, developmental delay and learning disability. The term "mental retardation" has been widely used in the literature to describe individuals with significantly impaired intellectual functioning, but was seen as disparaging, and now "intellectual disability (ID)" is preferred by most researchers (Lai et al., 2012). The brains of individuals with ID do not function within the normal range for their ages. Affected individuals cannot meet the personal and social demands expected of them. They have a significantly reduced ability with regard to intellectual reasoning and adaptive skills (Huang Jichong et al., 2016a). School age children with ID may struggle with memory, problem solving, attention, reading, language, mental arithmetic and visual comprehension. They usually develop more slowly than their peers and soon fail to achieve developmental milestones in some, or all, of the developmental domains (Goharpey et al., 2013; Kok et al., 2016).

ID is a major health issue among young adults, with a worldwide estimated prevalence of about 2-3% (Maulik et al., 2011). It tends to be more common in developing countries due to the effect of environmental factors such as poor medical care, infection and malnutrition (Emerson and Hatton, 2007; Lakhan, 2015; Maris et al., 2013). ID is defined using three criteria: reduced intellectual ability, deficit in two or more adaptive behaviours and diagnosis before the age of 18 years (Carulla et al., 2011).

There has been a gender bias in the prevalence of ID and other neurodevelopmental disorders: more boys are diagnosed than girls by about 30-50% (Abikoff et al., 2002; Polyak et al., 2015; Richardson S. A. et al., 1987). In a study of Taiwanese children aged 3-17 years between 2004 and 2010, it was found that 59.2% of ID patients were boys and 40.8% were girls (Lai et al., 2012). This difference could be partially due to the high contribution of X-linked disorders, as up to 10% of the known genetic causes of ID are located on the X

chromosome (Gandomi et al., 2014; Niranjan et al., 2015). It is also reported that more boys are affected with Fragile X syndrome than girls (Polyak et al., 2015; Saldarriaga et al., 2014) but the alternative intact X-chromosome in females participates in compensating the deficiency (Raymond, 2006).

1.2.2 ID types

The Intelligence quotient (IQ) is the main measure that is used to quantify the severity of intellectual impairment. It is a statistical assessment of the thinking and reasoning ability of a person compared with individuals of the same age. Average IQ within each age group is set at 100, and a person with an IQ below 70 is recognised as having intellectual disability (Webb and Whitaker, 2012). Different systems have been used for IQ tests but the most commonly used ones are the Stanford-Binet intelligence scale and the Wechsler Adult Intelligence Scale (WAIS) (Silverman et al., 2010). Based on IQ score, the Diagnostic and Statistical Manual of Mental Disorders (DSM) divides ID into the following categories: mild, moderate, severe and profound (Carulla et al., 2011; Katz and Lazcano-Ponce, 2008; Schuchardt et al., 2010). Table 1 gives more details of each category.

ID type	IQ score	Description of functional level and support needed	Approximate proportion
Mild	50-69	Intermittent support needed during transitions or periods of uncertainty	85%
Moderate	36-49	Acquire some communication and self-help skills, require limited support in daily situations	10%
Severe	20-35	Acquire only basic self-help and communication skills, require extensive support for daily activities	3.5%
Profound	<20	Require highly structured and supervised living conditions for every aspect of daily routines	1.5%

Table 1.1. Classification of ID based on IQ score as described by Boat and Wu (2015).

Sometimes, the mode of inheritance is also cited beside the ID to indicate its type. Although the genetics of ID is separately highlighted in a subsequent section (1.2.4), some of the genetics are mentioned here to explain the classification point. X-linked ID (XLID, OMIM 309530) is caused by mutations in genes on the X chromosome. Over 150 syndromes have been described (Lubs et al., 2012) but the most common example is fragile X syndrome (FXS), which is caused by an expanded trinucleotide repeat (CGG) in the *FMR1* gene (Fragile X Mental Retardation Protein, OMIM 309550). Unaffected people have between 6 and 54 copies, whereas FXS patients have over 200 copies (Willemsen et al., 2011). Nance-Horan syndrome (NHS, OMIM: 302350) is one of the rare X-linked disorder which is thought to be under diagnosed and not fully evaluated (Toutain, 2003). The condition is characterized by congenital cataracts, teeth anomalies as well as intellectual disability, and it was first described by two independent studies (Horan and Billson, 1974; Nance et al., 1973). The bilateral cataract is normally severe in the affected males and would require an early surgical intervention in order to stop profound vision loss (Burdon et al., 2003). Other reported ocular abnormalities include small eyes (microphthalmia), small cornea (microcornea), recurrent eye movements (nystagmus) and absence of bilateral alignment during object focus (strabismus) (Hong et al., 2014).

The characteristic teeth anomalies which have been described in Nance-Horan syndrome include screwdriver or cone-shaped incisors, supernumerary maxillary incisors (mesiodens) and increased gap between teeth (diastema) (Stevenson et al., 2012; Walpole et al., 1990). About 30-50% of individuals with NHS have intellectual disability (Brooks et al., 2010). In some of the cases, additional specific facial phenotypes have also been reported like inclination in the external part of the ear (anteverted pinnae), shortening in any of the five bones of the hand (metacarpals), and long face (Coccia et al., 2009). The severity of the above-mentioned symptoms would significantly differ from person to person even within members of one affected family. Carrier heterozygous females have been reported with some milder forms of these related manifestations but intellectual disability tends to less likely to occur in those females (Ding et al., 2009; Tug et al., 2013; Zhu D et al., 1990). The reported *NHS* mutations are, so far, not less

than 33 and most of them are either nonsense mutations or InDels (insertions and deletions) but CNVs (copy number variations), missense and splice site mutations are few (Tian et al., 2017).

The autosomal recessive intellectual disability (ARID, OMIM 611093) forms have not been well studied as the search for the underlying causes has been largely hindered by the scarcity of big consanguineous families in the Western communities (Kuss et al., 2011). Joubert Syndrome (JBTS, OMIM 213300) with related disorders (JSRD) (Brancati et al., 2010) and Bardet-Biedl syndrome (BBS1, OMIM 209900) (Forsythe and Beales, 2013) are examples of autosomal recessive ID. JSRD are a group of developmental disabilities and various congenital abnormalities but an essential diagnostic hallmark to distinguish this group is the presence of molar tooth sign (MTS), which is a unique cerebellar and brainstem marker visible on brain imaging (Brancati et al., 2010; Romani et al., 2013). Bardet-Biedl syndrome (BBS) is characterized by defects in multiple organ systems but its main features are rod-cone dystrophy, obesity, hypogonadism, postaxial polydactyly, renal dysfunction and intellectual disability (Zaghloul and Katsanis, 2009).

The autosomal dominant ID (ADID, OMIM 614563) are not commonly seen due to the fact that patients with intellectual disability rarely reproduce (Winnepenninckx et al., 2003). The conditions of autosomal dominant ID are mainly reported with *de novo* mutations (Khan Muzammil Ahmad et al., 2016) and one example is chromosome 2q23.1 deletion syndrome (Van Bon et al., 2010). In fact, duplication of this 2q23.1 region has also been described (Mullegama et al., 2014) to cause a similar effect. Although the autosomal dominant Marfan Syndrome (MFS, OMIM 154700) (Hofman et al., 1988) is a disorder of mainly connective tissue with involvement of cardiovascular, skeletal, ocular and pulmonary systems, some cases of intellectual debility have also been described (Ades et al., 2006; Hoffjan, 2012).

There is also a different ID classification based on whether other clinical features are also present. When low IQ is the sole clinical feature and the only

manifestation of ID, this is classified as non-syndromic ID. Syndromic ID is characterised by learning disabilities in the presence of other, co-morbid clinical features, such as large, protruding ears, long face and adult macro-orchidism of Fragile X syndrome, or the stunted growth, atypical fingerprints and eyelid crease of Down syndrome (DS, OMIM 190685). Other examples of syndromic ID include Rett syndrome (RTT, OMIM 312750), Angelman syndrome (AS, OMIM 105830) and Prader-Willi syndrome (PWS, OMIM 176270). Distinguishing between syndromic and non-syndromic forms of ID is sometimes difficult because the co-morbid features may be subtle (Miclea et al., 2015). Some researchers believe that the same genes are sometimes involved in both syndromic and non-syndromic ID, suggesting that there might be no strict boundaries between these forms (Beaulieu et al., 2013; Birk et al., 2010; Frints et al., 2002; Tejada et al., 2011).

1.2.3 Environmental causes of ID

ID can have various causes, but genetic factors are thought to be responsible for up to 50% of cases (Karam et al., 2015; Kaufman et al., 2010). A wide range of environmental insults occurring during pregnancy, childbirth or infancy may account for the other half of cases. Exposure of the fetus to drugs or alcohol is the most common cause of ID during pregnancy. A significant contributor to the non-genetic causes of ID, with a general prevalence of 1-2 in every 1000 babies, is Fetal Alcohol Syndrome (FAS), in which the fetal brain is affected by alcohol crossing the placenta (Aronson and Hagberg, 1998; Cesconetto et al., 2016; Miller, 2014; Senturias and Asamoah, 2014). A wider range of neuropsychological impairments is associated with Fetal Alcohol Spectrum Disorder (FASD), another consequence of fetal exposure to alcohol. FASD is a major contributor to ID in the Western world and the different levels of FASD affect about 2-5% of the population (May et al., 2013). In the UK, there are 7000 babies born every year at high risk of FAS and FASD (Larcher and Brierley, 2014). The behaviour of these babies is determined by different factors, including the amount of consumed alcohol, maternal metabolism and developmental stage of the fetus during exposure (Basavarajappa and Subbanna, 2016).

12

Environmental causes of ID that might take place during birth are preterm birth, oxygen deprivation and brain trauma. There have been various reports describing different degrees of risk of ID and attention deficit in preterm babies (Anderson Peter J, 2014; Hutchinson et al., 2013; Litt et al., 2005; Soleimani et al., 2014). A UK study by Johnson and coworkers (Johnson et al., 2016), involving 161 children born extremely preterm (EP) and 153 term-born controls, showed that the prevalence of ID was 47% in EP children compared with only 4.6% in the controls. ID is known to occur as a result of hypoxia or oxygen shortage (Sigelman and Rider, 2014), due to inadequate delivery of oxygen to cells. In cases of hypoxia, lysophosphatidic acid (LPA) receptors are activated, leading to overactivation of signalling pathways. As a result, the mitotic neural progenitor cells (NPCs) are driven to migrate from their normal specific regions within the developing brain into more superficial areas and thus fail to grow properly (Herr et al., 2011).

Brain trauma occurring before, during or after birth can also cause long-lasting/permanent harm to the brain (Carone, 2014; Foreman, 2009). A retrospective study was carried out to investigate the intellectual capability of 63 children aged less than 15 years, who had experienced severe traumatic brain injury (TBI) (Chevignard et al., 2016). IQ tests carried out 7-8 years after the injury showed that the children with TBI performed significantly worse than age and ethnicity matched healthy controls (p=0.016).

Other environmental causes of ID that can occur in early childhood include severe malnutrition and some infections. The diet is known to have an important effect on brain function and cognition (Gillette-Guyonnet et al., 2013; Mangialasche et al., 2013). Deficiencies in any of the essential micronutrients have been associated with the development of various functional deficits (Kuper et al., 2015). For example, Waber and colleagues reported impaired IQ in adulthood as a result of malnutrition during infancy (Waber et al., 2014), while Swaminathan and co-workers reported that reduced cognitive and physical performance in Indian children was related to a shortage in vitamin B and folate intake (Swaminathan

et al., 2013). The mechanism by which nutrition affects the development of the brain is complex. The processes of metabolism, enzyme production, energy control, neurogenesis, neurotransmission and synaptic plasticity are all affected by diet (Dauncey, 2014). Certain prenatal infections, like rubella, cytomegalovirus or *Toxoplasma gondii,* can disturb development of the fetus, and are also plausible risk factors for brain damage (Spreen et al., 1995). Various studies suggest that such infections can cause cognitive impairment and autistic spectrum disorders but the evidence is inconsistent (Bilder et al., 2013; Meyer et al., 2007). A study of a big cohort of individuals born in Sweden between 1984 and 2007 to mothers who were diagnosed with infection during pregnancy provides evidence that infections during pregnancy increase the risk of children having Autism spectrum disorder (ASD, OMIM 209850) (Lee Brian K. et al., 2015). Another study which checked children with ASD and ID for CMV infection concluded that congenital CMV infection is a potential contributor to ASD, particularly if the condition is also combined with ID (Engman et al., 2015).

1.2.4 ID genetics

The precise proportion of ID cases caused by gene mutations is unknown, but the literature states proportions ranging from 15 to 50% (Karam et al., 2015; Kaufman et al., 2010; Moeschler and Shevell, 2006). It is not unexpected for genes to have such a significant contribution in ID aetiology, as mutations could affect any of the numerous genes that are crucial for early brain development. The chromosomal aberration in Down's syndrome (DS), trisomy 21, is considered to be the leading genetic cause of ID, as it alone accounts for up to 20% of worldwide cases (Lakhan and Kishore, 2016). Other chromosomal abnormalities frequently associated with ID include trisomy 13 and trisomy 18 (Miclea et al., 2015). The triplet repeat expansion (CGG) in the *FMR1* gene causing fragile X syndrome (FXS) is the major monogenic cause of ID, accounting for 5% of cases (Inaba et al., 2014; Moeschler et al., 2014).

There have been various monogenic genetic causes identified for ID and they are generally related to dysfunction in neurodevelopmental processes or in synaptic

organisation. The processes that regulate neural development during the embryonic stage work in a tightly controlled and organised cascade, resulting in the proliferation of neurons from neuronal precursor. The process of neural proliferation and production of fully differentiated neurons is called neurogenesis, a sophisticated process involving many transcription factors and regulatory genes (Busskamp et al., 2014; Stiles and Jernigan, 2010). Neurogenesis begins very early during embryogenesis, with the neuronal tube being formed about 15 days after fertilization. Neurogenesis continues in most brain regions until early postnatal stages, but the dynamic production of new neurons carries on into adulthood in the ventricular-subventricular zone and the subgranular zone of the dentate gyrus. Adult hippocampal neurogenesis is important in the formation of new memories (Sahay et al., 2011; Urbán and Guillemot, 2014).

Disruption within any steps of neurogenesis could have a significant impact on the cognitive ability of an individual. For example, defects in genes involved in controlling the production of neurons could lead to an imbalance between forming more progenitor cells and forming end differentiated neurons giving rise to microcephaly (OMIM 152950) which is characterized by reduced head circumference at birth and various degrees of ID (Chavali et al., 2014). Microcephaly is a heterogeneous neurodevelopmental disorder, in which mutations have been found in the following genes: *MCPH1* (Microcephalin, OMIM 607117), *WDR62* (WD Repeat-Containing Protein 62, OMIM 613583), *CDK5RAP2* (CDK5 Regulatory Subunit-Associated Protein 2, OMIM 608201), *CASC5* (Cancer Susceptibility Candidate Gene 5 Protein, OMIM 609173), *ASPM* (Abnormal Spindle-Like, Microcephaly-Associated, OMIM 605481), *CENPJ* (Centromeric Protein J, OMIM 609279), *STIL* (Scl/Tal1-Interrupting Locus, OMIM 181590), *CEP135* (Centrosomal Protein, 135-KD, OMIM 611423), *CEP152* (Centrosomal Protein, 152-KD, OMIM 613529), *ZNF335* (Zinc Finger Protein 335, OMIM 610827), *PHC1* (Polyhomeotic-Like 1, OMIM 602978) and *CDK6* (Cyclin-Dependent Kinase 6, OMIM 603368) (Faheem et al., 2015). Problems associated with neuronal migration, restricting the ability of nerve precursor cells to travel from their origin into the cortical plate, have been reported to cause psychomotor disorders, including epilepsy and ID. Disorders that occur

due to mutations in genes involved in neural migration include lissencephaly, polymicrogyria and periventricular heterotopia (Dyment et al., 2013; Filippi, 2015).

The human brain contains a quadrillion (10^{15}) synapses (Srivastava and Schwartz, 2014), which provide communication between neurons. Synapses are very complex junctions that are stabilised by adhesion molecules. Each synapse is made up of three main components: a presynaptic terminal, a post-synaptic terminal, and a cleft, measuring 20-25 nm, between them (Ho Victoria M. et al., 2011b). Communication between neurons through the synapses is activated when an action potential initiates an electrical impulse within the presynaptic terminal, which in turn triggers the release of various neurotransmitters. Neuron to neuron communication is achieved by these neurotransmitters, which diffuse through the synaptic cleft until they reach the post-synaptic terminal, where they attach to specific receptors. If the structure of synapses is disrupted, this could lead to dysfunction of the signalling network, as has been found in many brain disorders (Finnema et al., 2016). An illustration of the extreme complexity of the synapse, with some proteins encoded by genes implicated in ID pathogenicity, is shown in Figure 1.3. Miclea et al. (2015) have categorised ID genes based on their involvement in mechanisms and pathways, such as metabolic pathways, neurogenesis, neuronal migration, synaptic function, intracellular signalling, or epigenetic regulation of transcription. Examples of the different genes found in each group are given in Table 2.
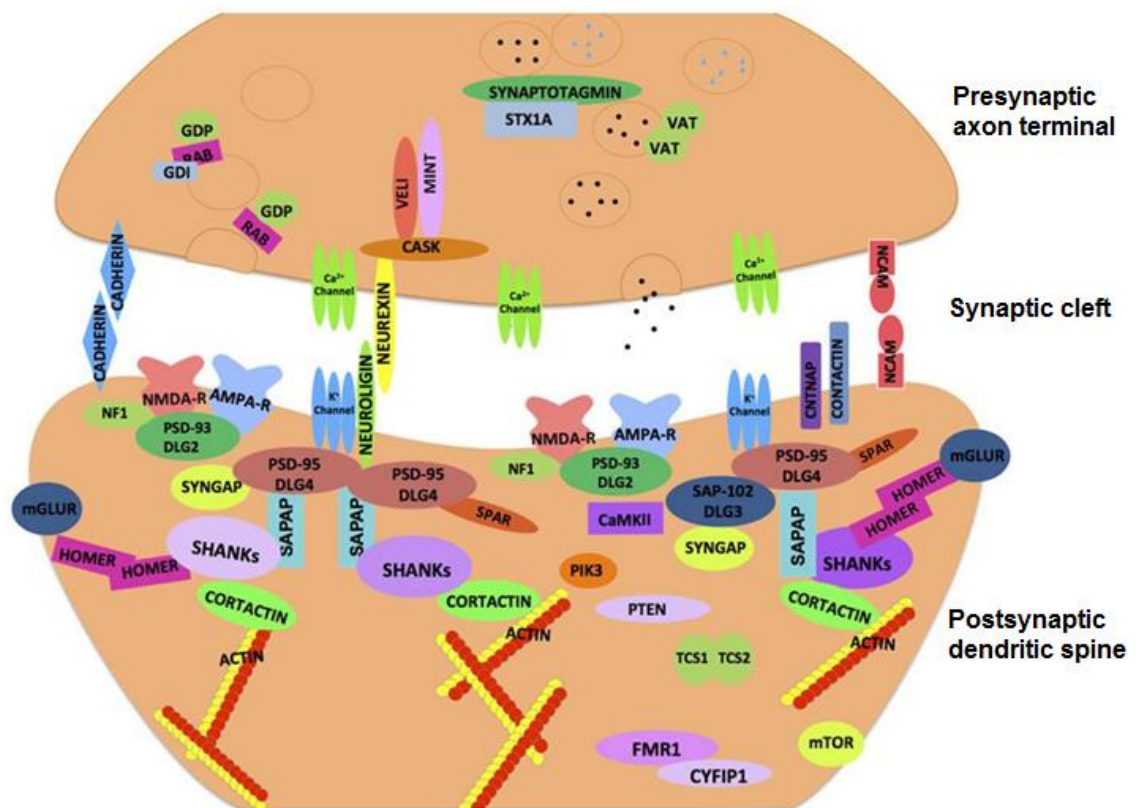
Figure 1.3. Synaptic Complexity and some ID involved proteins. This cartoon represents the presynaptic terminal, the cleft and the postsynaptic components of the glutamatergic synapse in the brain. It shows multiple proteins which are available to propagate the messages through connecting the transmembrane and membrane-associated protein complexes with the underlying actin cytoskeleton. These include some adhesion molecules (like Neurexin, Neuroligin and Cadherins), and scaffolding proteins (like PSD-95, Cask and Shank). Many of the illustrated proteins (such as Neurexin, Cask and Shank) have also been implicated in other neuropsychiatric disorders like ASD. Disruption in any of these proteins would result in a failure to transmit the signals efficiently. (Adapted from Banerjee et al. (2014), no permission required).

Pathway	Genes encoding proteins implicated in ID	
Metabolic Pathways	Organic acid metabolism (*ALDH5A1, L2HGDH*)	Polysaccharide metabolism (*NAGLU, SGSH*)
	Protein glycosylation (*POMGNT1, POMT1, POMT2, FKTN, FKRP, LARGE)*	Purine metabolism (*ADSL*)
	Monocarboxylate transporter (*SLC16A2*)	Creatine transporter (*SLC6A8*)
Neurogenesis	Mitotic spindle regulation (*ASPM, CDK5RAP2, CENPJ*)	DNA repair and mitotic arrest (*MCPH1*)
Intracellular Signalisation	Ras-MAPK-ERK pathway (*SOS1, RAF1, BRAF, SHOC2, HRAS, KRAS, PTPN11, SPRED1, MAP2K1, MAP2K2, NF1, DYRK1A, RPS6KA3*)	PI3K-AKT-mTOR pathway (*TSC1, TSC2, PTEN*)
Neuronal Migration	Microtubule subunits (*TUBA1A, TUBB2B*)	Microtubule regulation (*DCX*)
	Microtubule associated proteins (*PAFAH1B1*)	Transcription factors (*ARX*)
Presynaptic Function	Vesicle traffic (*RAB3GAP1, STXBP1, GDI1, RAB39B*)	Exocytosis inhibition (*IL1RAPL1, CASK*)
	Adhesion between pre- and postsynaptic membranes (*NRXN1, CDH15*)	
Postsynaptic Function	Adhesion between pre- and postsynaptic membranes (*CNTNAP2, NLGN3, NLGN4*)	Neurotransmitter receptor interaction with membrane proteins (*SHANK2, SHANK3*)
	Regulation of postsynaptic proteins (*UBE3A, UBE2A, UBR1; HUWE1, CUL4B*)	Subunits of NMDA receptor (*GRIN2A, GRIN2B*)
	Transport of mRNA from the nucleus to the cytoplasm (*FMR1*)	
Epigenetic Regulation of Transcription	Histone acetyl-transferase (*CREBBP, EP300*)	Histone deacetylase (*HDAC4*)
	Histone methyltransferase (*NSD1, EHMT1, MLL2*)	Histone demethylase (*KDM5C*)
	Transcription factors (*TCF4, RAI1, ZNF711, ZNF41, ZNF674, ZNF81, PHF6, PHF8*)	DNA replication (*SETBP1*)
	DNA methyltransferase (*DNMT3B*)	Chromatin modification (*ATRX, BRWD3*)
	Repression of transcription factors (*BCOR, MECP2*)	

Table 1.2. Types of ID genes.

Based on the fact that ID is more common in males than in females, it was hypothesised that X-linked ID would account for at least 25% but data suggests only 10-20% (Ropers H. and Hamel, 2005; Winnepenninckx et al., 2003). So far, mutations in more than 100 genes located on the X chromosome have been implicated in ID (Lubs et al., 2012), which is consistent with the fact that the human X chromosome has a large amount of brain-specific transcripts that are crucial for learning and memory (Ross et al., 2005). The full figure of X-linked ID genes remains unknown but studying more of the large sample cohorts would help investigating all the other genes reside on the X chromosome. On the other hand, it is believed that the number of genes behind the autosomal forms of ID should not be less than 800-850 genes (Ropers H, 2010). In fact, a more recent study (Maris et al., 2013) suggests that total number of autosomal genes involved in ID could run into thousands, but about 400 of them have already been linked to ID (Lubs et al., 2012; Raymond, 2006). Based on the various performed studies on the non-syndromic ID, 51 loci and 34 genes have been identified with the majority of them residing on either chromosome 6 or 19 (Khan Muzammil Ahmad et al., 2016).

The completion of the human genome sequencing project (Little et al., 2003) has been the milestone which revolutionized the work in biology and medicine. Although it took more than 10 years and 3 billion USD to elucidate the first human genome (Von Bubnoff, 2008), the introduction of next generation sequencing (NGS) technologies has made it possible to do whole genome sequencing with only 1,000 USD (Moens et al., 2014). About a decade ago, around 60% of ID cases had an unknown aetiology, but this percentage has dropped with the introduction and wide usage of next generation sequencing (Vissers Lisenka E. L. M. et al., 2016). The number of identified ID genes is continuously increasing and the figure of 600 genes that was mentioned before (Martínez et al., 2016) has now gone up to 800 genes (Chiurazzi and Pirozzi, 2016).

1.3 Schizophrenia (SZ)

1.3.1 SZ overview

The term "psychosis" refers to an abnormal condition of the mind in which thought and emotions are so impaired that contact is lost with external reality (Gaebel and Zielasek, 2015). In schizophrenia (SZ, OMIM181500), psychosis is a descriptive term for the hallucinations, delusions and impaired insight that may occur. It is a heterogeneous disorder that affects around 1% of the human population worldwide. The onset of SZ is usually in late adolescence or early adulthood (Whiteford et al., 2013) but it can strike at any time of life. Most affected individuals continue to suffer varying degrees of difficulty throughout their lifetime, with a major impact on their quality of life (Hooley, 2010).

About €93.9 billion is spent every year to provide health care for individuals affected with SZ in Europe, while €16.7 billion is spent in the UK alone (Gustavsson et al., 2011). Schizophrenia affects the thoughts, emotions and behaviour of affected individuals, and is often described in terms of positive and negative symptoms. Positive symptoms are those that most individuals do not normally experience, but are present in people with schizophrenia. They can include delusions, hallucinations and paranoid thoughts, typically regarded as manifestations of psychosis. Negative symptoms are deficits of normal emotional responses of other thought processes, and are less responsive to medication. They commonly include apathy (lack of interest), anhedonia (inability to enjoy) and social withdrawal. Deficits in cognitive abilities are also widely recognized as a core feature of schizophrenia (Leigh, 2013; Zhang Jing et al., 2016a). Figure 1.4 summarizes the core symptoms of SZ and also shows their overlaps with other psychiatric disorders.

Figure 1.4. Core schizophrenia symptoms and their overlaps with other disorders. The five core symptoms within the coloured circle are expected to be present in SZ patients for at least one month to warrant a diagnosis of schizophrenia. There is a range of overlaps with other psychiatric disorders, listed outside the segments of the circle. (Adapted from Corvin (2011), no permission required).

Another form of major mental illness is bipolar disorder (BPD), formerly manic depression, which is characterised by episodes of abnormally elevated arousal (mania) and episodes of depression. In severe cases, the individual may develop symptoms of psychosis. When a patient has features of both SZ and a mood disorder—either bipolar disorder or depression—but does not strictly meet diagnostic criteria for either alone, this is called schizoaffective disorder (Craddock et al., 2009).

There has not been a single effective therapy for SZ but antipsychotic drugs have become the cornerstone for the treatment of the associated symptoms (Lieberman et al., 2005). The first-generation antagonists of dopamine D2 receptors have been introduced and used since 1950s to treat SZ symptoms and they are commonly known as conventional antipsychotics or traditional antipsychotics. Despite their acceptable effect, they were found to cause some extrapyramidal side effects such as muscle contractions, tremor and irregular movements and thus many patients have preferred to stop using these drugs after a period of time (Hartling et al., 2012). The introduction of the second-generation antipsychotic drugs or what is called as atypical antipsychotics, like the ones that target serotonin and norepinephrine, was made in 1980s to get a better treatment with less of these nasty side effects. Unfortunately, these drugs have also been reported to cause the unwanted side effects plus their high cost makes it difficult for patients to afford buying them (Davies et al., 2007; Rosenheck et al., 2006). One of the commonly used atypical antipsychotic is clozapine (sold under different brand names including Clozaril, Leponex, Denzapine, Zaponex, Versacloz and others) which is considered to be the gold-standard treatment especially for resistant SZ cases which show no response when using two or more antipsychotics (Siskind et al., 2016; Taylor, 2017). Since their introduction, neither first generation nor the second generation antipsychotics have proved their efficacy and consistency in treating the symptoms of SZ (Hartling et al., 2012; Leucht et al., 2009).

Because glutamatergic dysfunction has been linked to the SZ pathogenesis (Olney and Farber, 1995), increased interest has been focused recently on the use of N-methyl-D-aspartate receptor (NMDAR) to treat this disorder (Merritt et al., 2013). Memantine (sold under different brand names Originally, Axura, Ebixa, Namenda, Namenda XR and others) which is a non-competitive NMDARs antagonist, has been the drug of choice to treat the negative symptoms and prevent further cognitive failure in young patients (Di Iorio et al., 2017). An alternative approach therapy, using cognitive remediation (CR) technique with visualization and narrative structure, has also been a promising method to improve the cognitive functions of the disorder (Hegde, 2017).

1.3.2 SZ types

Depending on the main diagnostic and prominent symptoms, SZ has been classified into different classic subtypes, according to the Diagnostic and Statistical Manual of Mental Disorders version IV (DSM-IV). It is not uncommon for the same affected person to change from one subtype to another with time (Mcglashan and Fenton, 1991). The paranoid SZ subtype is when the affected individual feels extremely suspicious and experiences auditory hallucinations with repetitive delusions about certain persecutions or conspiracies that have been unfairly made against them. As indicated by its name, the main characteristic of the disorganised SZ subtype is disorganization of the patient's thought. Impairment in thinking processes leads to incoherent speech and thus an inability to communicate effectively. In the catatonic SZ subtype, the movement of the patient is affected such that voluntary movements cannot be performed. This subtype is usually associated with a strange and uncomfortable body position/appearance and a tendency to repeat certain pointless movements or facial expressions. If none of the subtypes matches the diagnosis of an affected individual, and different symptoms manifest at different times in a mixed up manner, their condition is described as undifferentiated SZ. The term residual SZ is used when the disorder is at a chronic stage when no specific symptom is prominent but the person has no interest in life. The general features of hallucinations/delusions may be still present but to a lesser extent (Buoli et al., 2013; Dongen et al., 2015; Kito and Suzuki, 2016; Nenadic et al., 2015; Walther et al., 2009).

After the introduction of DSM-IV in 1994, there was some debate on the validity of these SZ subtypes. Kendler and colleagues carried out a study of 416 SZ patients and 1753 of their first-degree relatives in Ireland to see whether there is an outcome difference between the subtypes (Kendler et al., 1994). The study concluded that the subtypes are not aetiologically distinct syndromes. The clinical subtypes of SZ in DSM-IV are not considered to be reliable diagnoses because individuals with SZ may have a mixture of different subtypes. Even if one specific subtype fits better than the others, the stability is short and there is always a fluctuation in the prominent symptoms (Edward and Jhan, 2003). Moreover, the

different subtypes are not a great help in terms of prognosis or treatment response for the patients (Regier, 2007). In many cases, it has not been possible to identify the subtypes as described by DSM-IV, as no essential difference could be made between them (Linscott et al., 2010; Peralta and Cuesta, 2003; Picardi et al., 2012). Because of their inadequate reliability and due to the fact that no distinctive therapeutic approach is used for each subtype, they are absent from the new DSM-V (Tandon Rajiv et al., 2013).

According to the recent version of DSM-V (5th edition, 2013), for a diagnosis of schizophrenia, the individual must have at least two of the following symptoms: delusions; hallucinations; disorganized speech; grossly disorganized or catatonic behaviour; or negative symptoms. These symptoms need to be there for a significant portion of a one-month period, with signs of disturbance for 6 months. Different schizophrenic individuals experience different periods of constant symptoms and symptom-free periods (Erritty and Wydell, 2013). The new 11th edition of the International Classification of Diseases (ICD) has also been updated, with the subtypes being replaced by the following symptom specifiers: positive, negative, depressive, manic, and psychomotor symptoms, and cognitive impairment (Gaebel et al., 2013; Sachdev et al., 2015).

1.3.3 SZ causes

Schizophrenia is a genetically complex disorder which has been shown to run in families. The risk of developing schizophrenia is higher when close relatives also have the disorder (Gottesman I. I. and Erlenmeyer-Kimling, 2001). We do not yet have a complete understanding of the biological changes taking place in schizophrenia, but some theories and supportive findings have been made. For example, various studies nowadays are suggesting that an abnormal abolition of excess neuronal synapses could be behind the pathology of SZ (Boksa, 2012; Hayashi-Takagi et al., 2011) and synaptic pruning has become an interesting target for therapeutic intervention in patients with schizophrenia (Sekar et al., 2016). There is also accumulating and clear evidence for the involvement of both

environmental and genetic factors in the pathology of SZ (Brown Alan S., 2011; Tsuang, 2000).

The fact that concordance of schizophrenia in monozygotic twins (which have identical genomes) is only about 50% raises the potential importance of environmental factors (Owen et al., 2005). Various environmental risk factors have been implicated in schizophrenia, including maternal nutritional deficiency, birth during late winter or early spring, and cannabis use during adolescence (Tandon R. et al., 2008). Infections such as influenza and *Toxoplasma* have also been identified as environmental risk factors for schizophrenia (Fatemi and Folsom, 2009).

Immigration and change of cultures have also been found to increase the risk of developing schizophrenia (Hollander et al., 2016). People who leave their home and settle in other places or countries would normally experience a different atmosphere that affect wide aspects of their life including custom culture, religious practice and social support. Such changes and their associated challenges would collectively participate to express a stressed condition, that can impact their mental well being (Bhugra and Becker, 2005). According to Shekunov and colleagues (Shekunov, 2016), there is an average increase in SZ by about 2.5 times with immigration, but the exact ratio would vary based on ethnicities and settings. There is also a strong evidence suggesting the persistence of schizophrenia into the immigrants' descendants (Cantor-Graae and Selten, 2005; Shekunov, 2016). A meta analysis, based on 4422 schizophrenia and psychosis cases from 21 population-based studies published between 1977 and 2008, found that the increased risk for schizophrenia and related disorders is also extended to the second generation (Bourque et al., 2011). This finding, and others (Cantor-Graae and Pedersen, 2007; Corcoran et al., 2009; Selten et al., 2007), point out to the significant contributions of the post-migration factors and the social context in the occurrence of SZ.

Over a long period of time, there have been several theories on the aetiology of SZ but the most prominent ones propose a disruption in dopamine (DA) and

glutamate (GT) transmission (Howes et al., 2015; Seeman, 2009). The DA and GT neurotransmitter systems (Figure 1.5) have been biologically linked to the development of the different symptoms of SZ (Mcguire et al., 2008). For example, accumulation of DA is hypothesised to cause the positive symptoms of hallucinations and delusions (Brisch et al., 2014; Shen et al., 2012) whereas the accumulation of GT is associated with negative symptoms and cognitive deficits (Merritt et al., 2013). There has also been a recent theory describing synaptic pruning and the role of C4 (complement component 4, OMIM 120810) in SZ pathology (Sekar et al., 2016), and this will be addressed in the next sub-heading about SZ genetics.
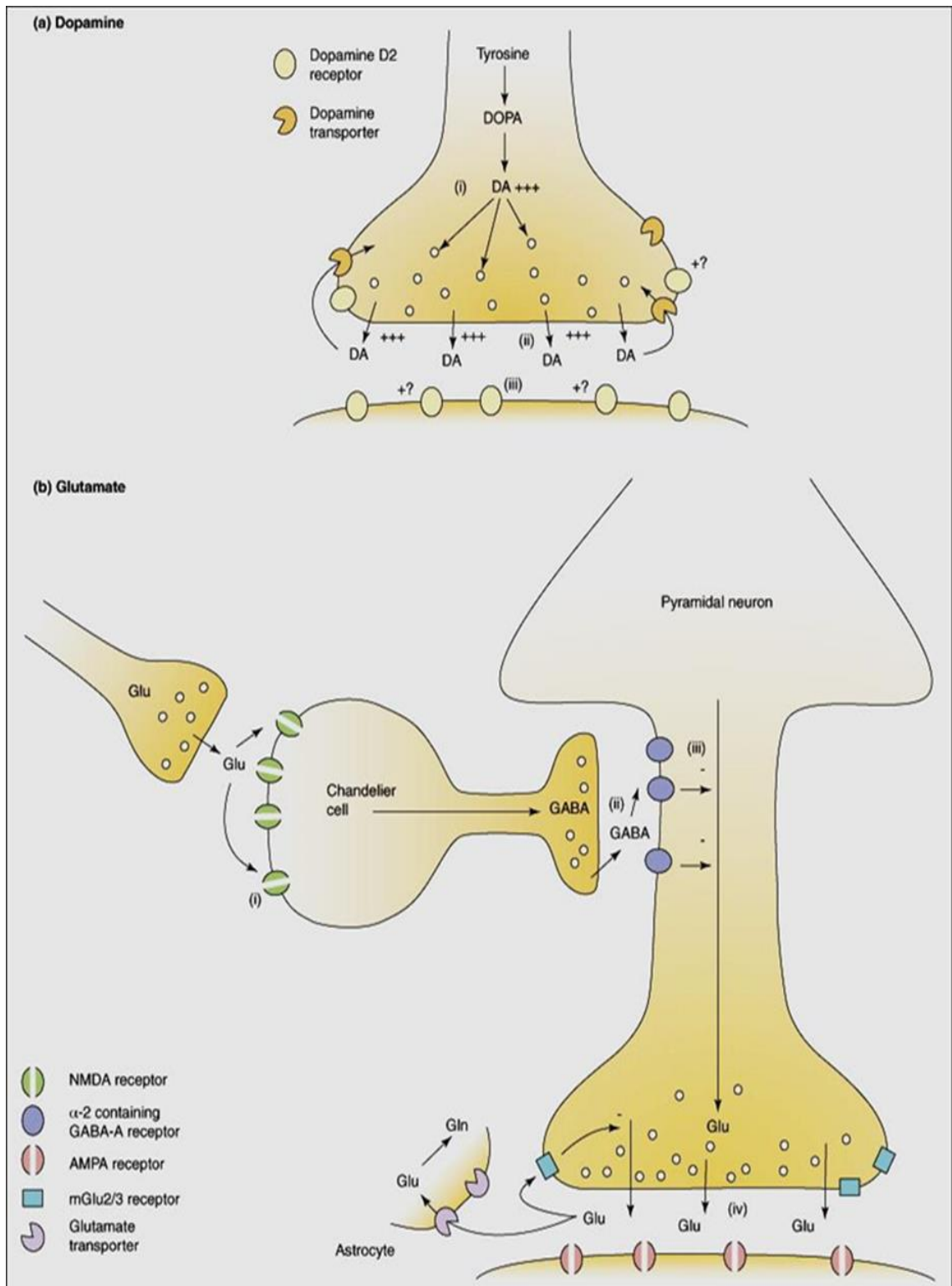
Figure 1.5. Neurotransmitter systems disrupted in schizophrenia. (a) Dopamine neurotransmitter: the increased production of DA (i) from its precursor (tyrosine) leads to an increase in its synaptic release (ii). The dopamine transporters would pump some of dopamine out of the synapse back into cytosol but some amount would still remain. Similarly, the number of DA receptors is unaltered (iii), which means they are not in their active state again and thus not functionally ready to receive the high DA, leading to its accumulation and development of positive symptoms. (b) Glutamate neurotransmitter:

the increased production of glutamate from "messenger-sending" nerve cell causes some damage to interneurons (like chandelier cells) and their NMDA receptors (i) causing a reduction in the quantity of GABA released onto pyramidal neurons (ii). As a compensatory response to GABA shortage, the pyramidal neurons tend to release glutamate onto non-NMDA glutamate receptors (AMPA receptors) (iv). Some glutamate would be taken up by astrocytes and converted to glutamine but most of it would accumulate, causing the negative symptoms and cognitive impairment. (Adapted from Mcguire et al. (2008) with a permission from Elsevier and Copyright Clearance Center, License Number: 4077171303111).

1.3.4 SZ genetics

Relatives of SZ patients have an increased prevalence of SZ compared with the general population (Onstad et al., 1991), such that the chance of developing SZ is increased 10 times for a child who has a first-degree relative already suffering from the disease (Sommer et al., 2016). The SZ concordance rate of monozygotic twins is 40-50%, with a heritability rate of up to 80% (Cardno Alastair G and Gottesman, 2000; Sullivan et al., 2003). SZ is probably the one first condition at which there has been a successful application of polygenic score (Purcell et al., 2009). The polygenic score is a measurement that uses the total trait-associated alleles obtained from genome wide association studies (GWAS) in order to predict a genetic liability to disease risk for an individual (Euesden et al., 2015). It was estimated that the SZ polygenic scores are around 50% in first-degree relatives and 15% in second degree relatives of SZ patients (Power et al., 2015). Relatives of patients with bipolar disorder or depressive disorder also have an increased risk of developing SZ (Rasic et al., 2014).

There are basically two opposite models of genetic architecture that have been proposed in the literature to explain the common disease susceptibility (Figure 1.6). Based on epidemiological studies and the estimated risk to relatives, it is believed that SZ is caused by a group of susceptibility genes at which their proteins interact with each other, under the influence of environmental factors. This is compatible with the 'common disease, common variant (CDCV)' hypothesis, which proposes that SZ is caused by multiple common risk alleles, each of low genetic penetrance, under the influence of the environment. The penetrance of a disease-causing mutation is the proportion of individuals with the mutation who exhibit clinical symptoms. The alternative view is the 'Common

28

Disease, Rare Variant (CDRV)' hypothesis, which proposes that rare risk alleles, each of high penetrance, comprise the major genetic contribution to disease occurrence (Corvin, 2011; Schork et al., 2009). Under this latter model, there is thus an important level of heterogeneity, but the genetic variants are expected to have clearer functional impact by impairing protein function or production (R).

The high level of SZ heritability (about 80%) suggests the implication of genetic factors in the development of this condition. Technological advances in next-generation sequencing approaches have made it possible to explore the whole exome and entire genome of individuals and not only some predefined loci where common variants have been detected. Genome-Wide Association Studies (GWAS) have been successful at discovering numerous novel associations in disease genes that were not suspected to be involved at, demonstrating the role of novel cellular processes. One of the largest meta-analysis (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), at which GWAS was applied to study 37 thousands SZ cases and more than 113 thousands controls, identified 128 associations within 108 loci. Majority (75%) of these loci contain protein-coding genes and additional 8% are less than 20 kb away from a gene. The finding of this above study (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) is consistent with the assumption that the vast majority of GWAS variants are normally involved with changing the expression of the genes rather than altering the organisation of their associated proteins (Maurano et al., 2012; Nicolae et al., 2010). Furthermore, it also correlates with the findings of exome sequencing studies (Purcell et al., 2014) as well as with the reported enrichment of SZ risk loci based on expression quantitative trait loci (eQTL) (Richards et al., 2012). Some of the other featured findings have also come from The International Schizophrenia Consortium (2008) study which showed that the SZ associative candidates are abundant in tissues that are brain-specific, like Hippocampus middle and Cingulate gyrus, rather than other tissues like kidney and bone. Significant associations were also reported in this study for CNVs overlapping with ASD and ID.
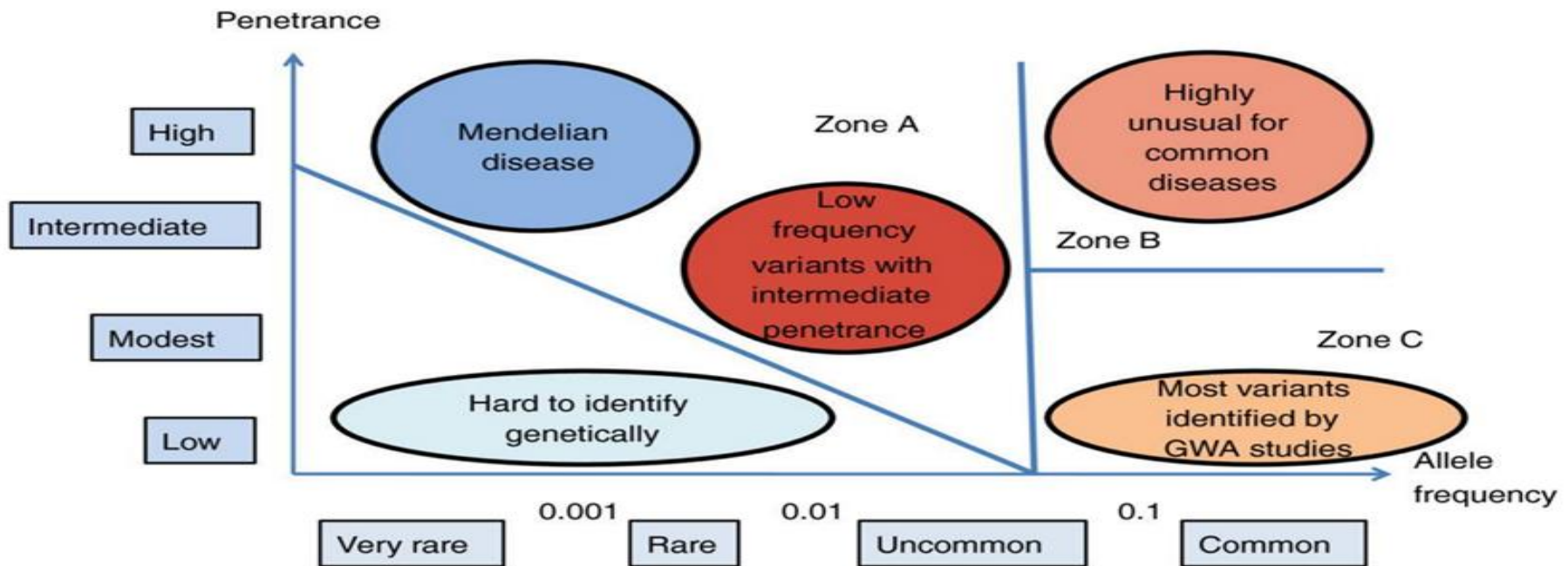
Figure 1.6. Different genetic risk theories of schizophrenia. The x-axis shows the frequency of a given allele in the population while the y-axis indicates their penetrance. Zone A contains relatively rare alleles and zone B contains the variants that are more common than variants in zone A. The variants at Zone C are also common but their penetrance is low comparing with the ones in zone A and B. Under the CDCV model, it is assumed that the major contributors to SZ risk are genetic variants located in zone C and these common variants could be detected with Genome-wide association studies (GWAS). On the other hand, the CDRV model assumes that rare genetic variants in Zone A can be involved in SZ susceptibility. It is known that the alleles causing Mendelian diseases are located in this zone but there is also a group of alleles that cannot be easily identified because of their low frequency and low penetrance. The fast majority cases of schizophrenia alleles are in zone A and C and the utilization of whole exome or whole genome sequencing could significantly help in the identification of the underlying alleles (Adapted from Corvin (2011), no permission required.)

Despite the accumulation of evidence that SZ has a genetic basis, the precise mechanisms are unclear. Some studies have observed potential anticipation, whereby the onset of the disease occurs at an earlier age with each subsequent generation, suggesting the involvement of genes that have trinucleotide repeat expansions (Bassett and Husted, 1997; Stöber et al., 1995). SZ has also been thought to be caused by damaging *de novo* mutations affecting a few genes, including *KLC2* (Kinesin Light Chain 2, OMIM 611729), *KIF17* (Kinesin Family Member 17, OMIM 605037), *SHANK3* (Sh3 and Multiple Ankyrin Repeat Domains 3, OMIM 606230) and *IL1RAPL1* (Interleukin 1 Receptor Accessory Protein-Like 1, OMIM 300206) (Awadalla et al., 2010). Other studies have suggested the involvement of larger scale structural variants. An affected mother and daughter were found to have disruption of the *NPAS3* gene (Neuronal Pas Domain Protein 3, OMIM 609430) as a result of a chromosomal translocation t(9;4) (Kamnasaran et al., 2003). The NPAS3 protein has a significant role in regulating the differentiation process of cells and also in supporting the growth and development of neural cells (Pickard et al., 2005). Subsequent studies have confirmed the association of *NPAS3* with SZ (Bernier et al., 2014; Macintyre et al., 2010). The Disrupted-in-schizophrenia-1 gene (*DISC1,* OMIM 605210) was initially found to be truncated by a chromosomal translocation in an extended Scottish pedigree that includes cases of major depression, schizophrenia and bipolar disorder. The *DISC1* translocation showed more than 70% penetrance for mental disorder. *DISC1* was then regarded as the most promising candidate gene for schizophrenia (Bradshaw and Porteous, 2012).

To achieve a better understanding of the molecular mechanisms of schizophrenia, various genetic studies using different strategies in both humans and animal models have been carried out (Ayalew et al., 2012). Genome-wide association studies (GWAS) have identified various single nucleotide polymorphisms (SNP) as genetic risk factors for SZ. One of the strongest findings is SNP rs1625579 (G/T), located in an intron of a gene encoding a microRNA called MiR-137 (Micro RNA 137, OMIM 614304) (Zhang Ping et al., 2016b). Although the first evidence suggesting correlation between MiR-137 and SZ was in 2010 (Potkin et al., 2010), many successive independent studies have also

reported an association (Lett et al., 2013; Ripke et al., 2014; Whalley et al., 2012). Hence, SNPs in MiR-137 are now a well-documented genetic risk factor for SZ. A gene called *EFNB2* (Ephrin B2, OMIM 600527) has also been considered a persuasive candidate gene for involvement in SZ (Zhang Rui et al., 2010). Interestingly it has been shown that MiR-137 inhibits *EFNB2* expression (Wu et al., 2016), suggesting a possible association mechanism involving both of them (Tooney, 2016).

C4, the complement component 4 gene which has a function in the immune system, has recently been reported to also play a role in brain development, with variants of it showing the biggest hit as risk factor for SZ in GWAS (Sekar et al., 2016). The idea arose from the observation that various markers in the major histocompatibility complex (MHC) are strongly associated with schizophrenia (Dhindsa and Goldstein, 2016). Sekar and colleagues (Sekar et al., 2016) recently showed that the significant association was specifically with certain alleles of C4. After analysing SNP data from 28,799 schizophrenia cases and 35,986 controls, association was found with 7,751 SNPs across the MHC locus. The authors reinforced their finding by demonstrating C4 protein localisation in neuronal synapses, dendrites and axons. They also showed that *C4*-deficient mice have reduced synaptic pruning, which is consistent with the fact that the brains of SZ individuals do not have sufficient synaptic inputs. A plausible mechanistic model for the role of C4 in SZ has been proposed by (Ruzzo and Geschwind, 2016), as illustrated in Figure 1.7.

Based on the findings of researchers at the Wellcome Trust Sanger Institute who performed WES on SZ patients and compared the variants with the ones in healthy individuals, loss of function (LOF) variants in *SETD1A* (Set Domain-Containing Protein 1A, OMIM 611052) increase the risk to SZ by 35 folds (Singh et al., 2016). *SETD1A* encodes a component of a histone methyltransferase that is important for the expression of other genes, particularly those that play roles in synaptic and neurotransmission pathways. In fact, a previously published study (Takata et al., 2014) already reported that *de novo* indels in *SETD1A* could be

responsible for SZ, but the recent study of Singh et al. (2016) strengthens the implication of this gene in the pathogenesis of SZ.
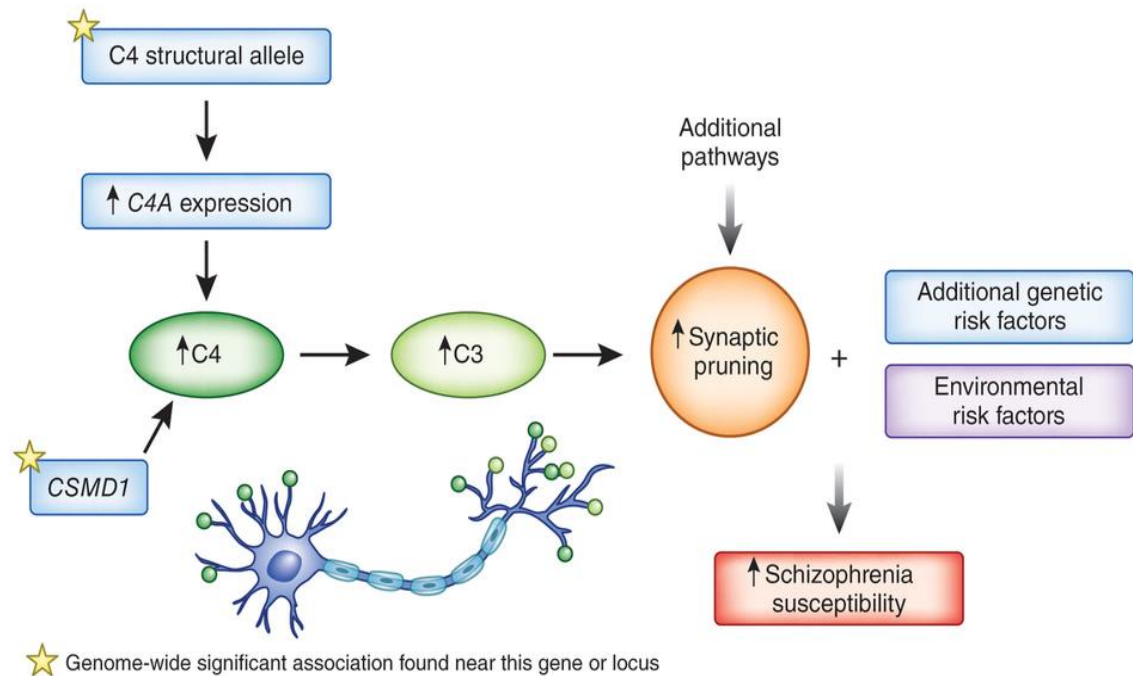
Figure 1.7. Model for the role of C4 in SZ. In GWAS, C4 structural alleles show a significant association with SZ. The above cartoon is an explanatory model which principally links SZ to the increased production of C4A mRNA in the human brain. As a result, this increases C3 production which increases synaptic pruning in the neural cells. *CSMD1* encodes a protein involved in C4 regulation. It also drives the described pathway in the same direction and is also associated with SZ in GWAS studies. The model also demonstrates the effects of other factors, including environmental insults and possible alternative genetic susceptibility factors. (Adapted From Ruzzo and Geschwind (2016) with a permission from Nature Publishing Group, License Number 4077180234698)**.**

Although there is little doubt about the positive impact of the recent findings, such as the associations of *C4* and *SETD1A*, our understanding of this devastating psychiatric illness is still incomplete. Most of the specialists in the field of psychiatric genetics are eager for a better future with more concrete findings (Curtis, 2016b).

1.3.5 Existence of Mendelian SZ alleles

SZ is a complex disorder with susceptibility resulting from the interaction of environmental and genetic factors. However, most of the risk variance associated with SZ is genetic, with heritability estimated to be around 80% (Carroll Liam S. and Owen, 2009; Girard et al., 2011; Lipina and Roder, 2015). This high level of heritability suggests that it should be possible to identify the genetic risk factors that contribute to SZ but results from over 40 years of studies have not, unfortunately, been very successful (Giusti-Rodríguez and Sullivan, 2013). More than a decade ago, it was widely hoped that utilising GWAS would explain the genetic basis of SZ as it successfully has done in other genetically complex conditions like Age-Related Macular Degeneration (AMD). However, the fact that the bulk of the heritability for SZ remains unexplained made it necessary to exploit alternative research strategies. Despite the fact that a part of SZ genetic variance has been unravelled with hypothesis-free screens of GWAS, the implicated common genetic variants have low odds ratios, and even measured collectively they account for no more than 33% of heritability (Lee S. et al., 2012; Sullivan et al., 2003) leaving a large portion unaccounted for. Therefore, there is still an open gap that needs to be addressed between our current knowledge of SZ genetics and the predicted genetic component of the disease. Addressing this should lead to better understanding of the development of brain abnormalities as well as the cognitive deficit associated with this disorder (Murphy and Benítez-Burraco, 2016).

By comparing the high heritability of SZ (80%) with the low proportion of the revealed variants (33%), it could be stated that there is a missing heritability. This missing heritability could in part be due to structural variants missed in existing sequencing studies (Kirov et al., 2009; Levinson et al., 2011; Van Dongen and Boomsma, 2013), in gene-environment interaction (Lipina and Roder, 2015; Van Dongen and Boomsma, 2013) and/or epigenetic factors (Cariaga-Martinez et al., 2016; Pishva et al., 2014). On the large-scale genome-wide study, by The International Schizophrenia Consortium (2008), the underlying rare but highly penetrant CNVs were investigated in 3391 SZ patients in Europe after they were compared with the ones in 3181 ethnically-matched controls. It was found that

CNVs with > 100 kb increase SZ risk by 1.15 folds. Moreover, duplication CNVs showed more consistent pattern in terms of the size as well as how far they occurs from a gene, in comparison with deletions. Significant associations were also reported, for the first time, for 15q13.3 and 1q21.1 regions. However, the genome-wide analysis by Stefansson and colleagues (Stefansson et al., 2008), which was carried out on 2160 trios of SZ and 5558 parent–offspring pairs, have identified 66 *de novo* CNVs, but further testing has also highlighted three predominant deletions all with significant associations; 1q21.1, 15q11.2 and 15q13.3.

Additionally, rare variants with a major effect might account for a portion of the cases, through the so-called common disease-rare variant hypothesis route. These rare variants would be shared by families or small communities, but due to their extremely low frequency in patients, broad association studies with thousands of unrelated individuals would fail to find them (Van Dongen and Boomsma, 2013). The common disease-rare variant hypothesis in psychosis is controversial (Craddock et al., 2007; Mcclellan et al., 2007), but the proven involvement in SZ of mutations in gene like *DISC1* (Millar et al., 2000, Sachs et al., 2005) and rare copy number variants with a high susceptibility risk (Sullivan et al., 2012) suggest that such rare alleles may indeed play a role in SZ.

Recessive alleles with a major effect that predispose to psychosis are likely to be enriched in patients from populations with high levels of consanguineous marriage because such progeny will have inherited two copies of the relevant mutation (maternal and paternal) from a recent common ancestor (Mansour et al., 2010). Evidence for the existence of such alleles comes from the observation that SCZ risk increases with genetic relatedness (Cardno A. G. et al., 1999; Gottesman Irving I, 1991) and consanguinity (Bener et al., 2012; Mansour et al., 2010). The chromosomal region surrounding these alleles will be homozygous identical-by-descent, so regions of homozygosity in affected individuals from consanguineous families are indicators to the mutations involved. Homozygosity mapping has already been applied in families with multiple cases of psychotic

disorders (Iqbal et al., 2013; Knight et al., 2008), showing evidence of increased incidence and provisional genes/chromosome regions of interest.

There are examples in the literature of Mendelian mutations that could account for a proportion of cases in apparently complex diseases. Alzheimer's, Parkinson's and Age-Related Macular Degeneration (AMD) are all diseases which are mostly complex but can also be caused by specific, severe highly penetrant alleles. For example, highly penetrant mutations in amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*), and presenilin 2 (*PSEN2*) have been shown to cause Alzheimer's disease (Mohamed et al., 2016; Van Cauwenberghe et al., 2016). Similarly, p.R1441G mutation in leucine-rich repeat kinase 2 (*LRRK2*) is a potent variant causing Parkinson's disease (Klein Christine and Westenberger, 2012; Mullin and Schapira, 2015). It is true that AMD has been strongly linked to mutations in *CFH* (Complement Factor H, OMIM 134370), *CFB* (Complement Factor B, OMIM 138470) and *HTRA1* (HTRA Serine Peptidase 1, OMIM 602194) (Wang Wenqiu et al., 2016), but Mendelian mode has also been reported in this genetically-complex disorder. For example, mutations in the tissue inhibitor of metalloproteinases-3 (TIMP3) have been found to cause rare Mendelian macular condition known as Sorsby's fundus dystrophy (SFD) (Black and Clark, 2016). The impact of Mendelian mutations on our understanding of the pathobiology of complex diseases is greatly dependent on the nature of the mutations involved. If the mutations involved are rare, highly penetrant, protein damaging variants then they can provide valuable insights to the underlying disease mechanisms. However, variants that are more common, less penetrant and solely affect regulatory processes might not be of significant impact on the pathobiology but they tend to be linked to high risk and susceptibility (Lupski James R. et al., 2011).

The fact that complex diseases do not exhibit large proportion of Mendelian alleles could be largely due to the shared interaction between both genetics as well as environmental factors. However, the involvement of certain variants in different pathological conditions and the extent of the genetic change also plays a role. For example, carriers of a heterozygous mutation in *GBA* (Glucosidase, Beta, Acid, OMIM 606463), a gene in which homozygous mutations cause

Gaucher disease (GD, OMIM 230800), have an increased risk of developing Parkinson disease (Goker-Alpan et al., 2004; Sidransky et al., 2009). Similarly, carrying heterozygous mutations in *CFTR* (Cystic Fibrosis Transmembrane Conductance Regulator, OMIM 602421) is associated with an increased risk of developing idiopathic pancreatitis (Weiss et al., 2005). It has also been shown that different genomic changes in a single gene could lead to the development of conditions of different complexity. A good example is the FMR1 locus, which causes ID when the CGG repeats exceed 200. It has been found that lower CGG repeats (down to 50) are associated with fragile X-associated tremor/ataxia syndrome (FXTAS) in about 33% of males and 10% of females (Jacquemont et al., 2004).

1.4 Genetic overlaps between ID and SZ

ID and SZ are both complex neuropsychiatric disorders, and are two conditions which are part of a group of overlapping disorders. This group also includes major depressive disorder, autism and bipolar disorder, all show varying degrees of overlap within cases and families over time (Homberg et al., 2016; Lehner and Miller, 2016). These disorders are under the influence of multiple genetic and environmental factors that interact with each other in complex ways, producing a broad spectrum of overlapping symptoms. The environmental influence further adds to complexity as this can fluctuate over time. Complex diseases do not normally show correlation between the observed phenotypes and their underlying genotypes as the genetic determinants of disease in different individuals affected by the same disorder will not be the same (Chu et al., 2014; Norbury et al., 2008).

The complexity of these diseases might be explained, in part, by the phenomenon of pleiotropy, where a particular allele can cause varying degrees of a number of different symptoms (Gratten and Visscher, 2016). The different penetrance levels of various risk alleles also add to the complexity. Even in condition when low penetrance alleles are involved, they still could cause a disease when they are together. Low penetrance alleles could also be supported by the environmental

factors in the aetiology of complex diseases (Cooper David N. et al., 2013; Schork et al., 2009).

Figure 1.8 shows the striking overlap between some of the genes with variants that have been implicated in the well-known neuropsychiatric disorders, based on what is called "Hot zone *de novo* mutations", an approach defined by Petrovski and colleagues (Petrovski et al., 2013). To explain more, this zone was established in order to measure the load of *de novo* mutations in cases of neuropsychiatric disorders and compare it with their loads in control individuals. The calculation was made using the results of two measurement designs; $\leq 25^{th}$ residual variation intolerance score (RVIS) and PolyPhen-2 score ≥ 0.95. The RVIS for a particular gene is simply an indicative measure for its functional genetic variation in the normal population based on NHLBI-ESP6500 data set. Therefore, a gene with very low or negative scores has less functional variation and thus will be described as "intolerant" in RVIS. This score has shown great capability to rank the protein coding genes that could cause Mendelian disease (Gussow et al., 2016). PolyPhen-2 score, on the other hand, reflects the ability of the mutation to cause change in the function of the produced protein (Adzhubei et al., 2010).
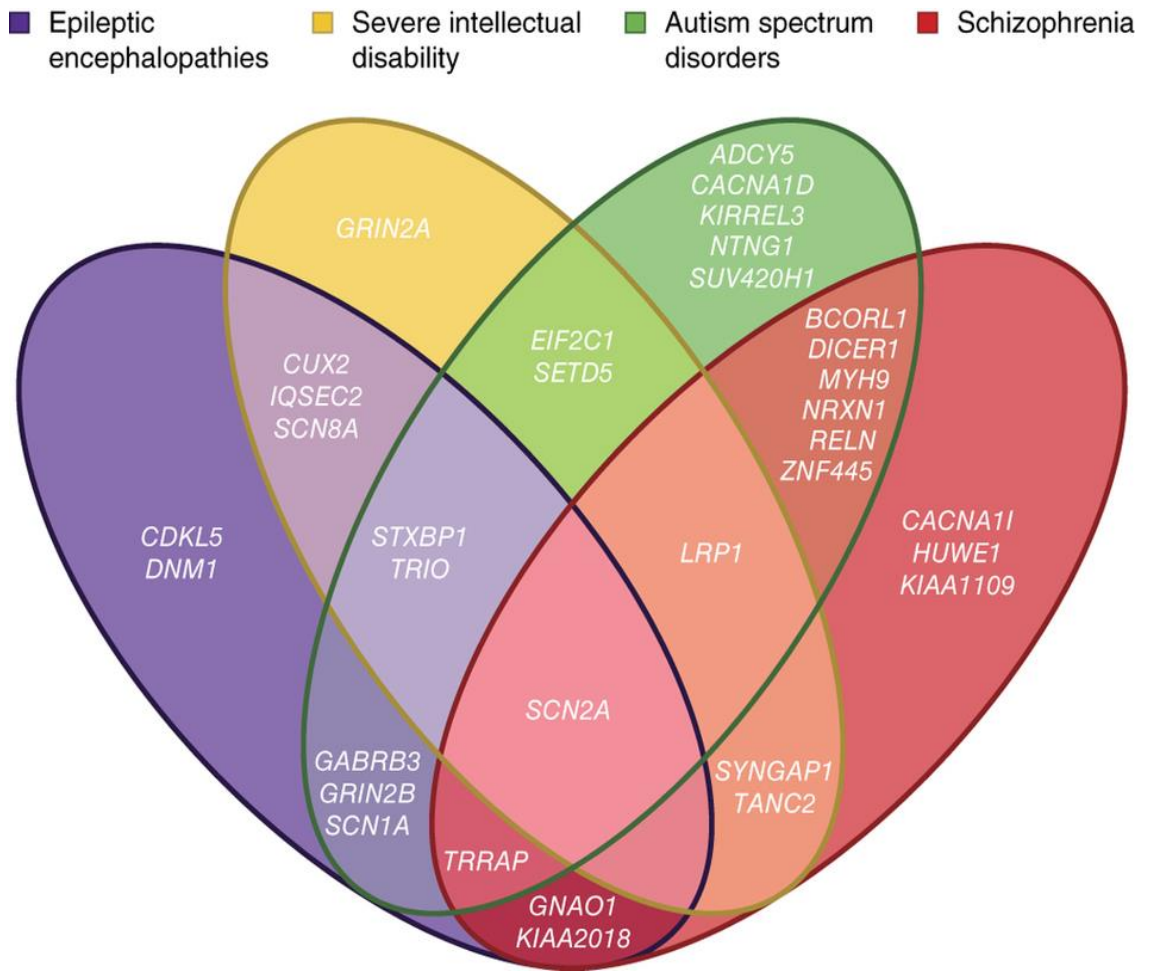
Figure 1.8. Venn diagram showing the outstanding overlap between many of the genes implicated in four neuropsychiatric disorders. This is from a study by Zhu and colleagues and was based on the understanding available in 2014, but is missing some of the more recently implicated genes described above. Overlaps were determined based on "Hot zone" de novo mutations found in four published schizophrenia studies (Girard et al., 2011a; Xu et al., 2012; Gulsuner et al., 2013; Fromer et al., 2014) involving whole exome sequencing of 512 probands. (Adapted From Zhu et al. (2014) with a permission from the Copyright Clearance Center's Rightslink service, License number: 4074171029655).

The idea of shared genetic susceptibility between these complex neurodevelopmental disorders is not new. It was first suggested almost half a century ago (Ornitz, 1969) and it has been supported by various studies over the past 20 years. For example, one study Morrow (2010) describes shared copy number variants (CNV) in intellectual disability, autism spectrum disorder (ASD), and schizophrenia. In addition, Camacho-Garcia and colleagues have reported exonic deletions in the *NRXN1* gene on chromosome 2 in patients with schizophrenia, autism, or mental retardation (Camacho-Garcia et al., 2012). Other examples of overlap between these complex neuropsychiatric disorders include the finding that deletion or duplication at 16p11.2 confers a high risk for ASD or schizophrenia, respectively (Brownstein et al., 2016). Indeed, 16p11.2 microdeletions account for 1% of ASD cases (Fernandez et al., 2009). On the other hand, the reciprocal microduplication of 16p11.2 is associated with a 14.5-fold increased risk of psychosis and 16-fold increased risk of schizophrenia (Giaroli et al., 2014; Mccarthy Shane et al., 2009). Zhiqiang Li and colleagues have reported 17 duplications, with an average length of ~753 kb, at 16p11.2 on genome-wide CNV analysis based on 6,588 Chinese patients with schizophrenia (Li Zhiqiang et al., 2016). Many genome-wide association studies have also reported apparently shared genetic risk factors for bipolar disorder and schizophrenia (Craddock et al., 2009). Recently, 3q29 deletion has been associated with different phenotypes, including ID, SZ and ASD (Biamino et al., 2016).

1.5 Gene identification approaches

Generally, there have been two strategies available for the identification of the genes involved in the causation of complex genetic diseases: genome-wide association and the candidate gene approach (Zhu Mengjin and Zhao, 2007). Genome-wide association is a hypothesis-free approach that involves genotyping a large group of unrelated individuals with a particular disease at thousands of diallelic SNPs across the genome. By identifying loci at which one of the two alleles shows a significantly higher frequency compared with a control group without the disease, this technique identifies loci that might alter the risk for

certain diseases (Mccarthy I. et al., 2008). Using this method, loci associated with type-1 diabetes (Hakonarson et al., 2007; Zeggini et al., 2007), Parkinson's disease (Gandhi and Wood, 2010; Liu Xinmin et al., 2011), breast cancer (Easton et al., 2007; Lindström et al., 2014; Turnbull et al., 2010) and psychosis (Cichon et al., 2011; Ripke et al., 2013) have been identified.

On the other hand, the candidate gene approach is a hypothesis-driven approach that uses prior scientific evidence to select particular 'candidate' genes for study in relation to the disease in question. Apart from the association studies (see above), the evidence for the candidate gene approach could be from linkage studies at which patients of a family share a particular variation that is tested by specific markers (Kwon and Goate, 2000). It could also come from expression studies, which compare the transcript levels of candidate genes in mRNA collected from diseased versus unaffected individuals (Zhu Mengjin and Zhao, 2007). An example of the candidate gene study approach is found in the report by van den Oord and colleagues, which showed a link between a SNP in *MAMDC1* (MAM Domain-Containing Glycosylphosphatidylinositol Anchor 1, OMIM 611128) and neuroticism, a personality trait at which the person has more tendency toward negative mood states (Van Den Oord et al., 2008). Another example is the implication of *HOXB1* (Homeobox B1, OMIM 142968) polymorphisms in the development of autism spectrum disorder (Muscarella et al., 2010).

1.5.1 Genetic markers

Genetic alleles or markers that are located on the same chromosome and are passed on to the next generation as a DNA unit or block are termed "haplotypes". Various SNPs have been identified and then organised according to their allelic haplotypes (Stephens et al., 2001). The use of multiple haplotype markers specific to a chromosomal region in order to evaluate its association with a phenotype of interest has been extensively employed in genetics, particularly in population and gene mapping studies (Crawford et al., 2004; Liu P-Y et al., 2005). Advances in genetic mapping have been facilitated by progress in the

development of microsatellite and SNP markers (Ball et al., 2010; Hearne et al., 1991; Nothnagel and Rohde, 2005).

Microsatellites are repeated simple sequences, sometimes referred as short tandem repeats (STRs). The abundance of these motif sequences throughout the genome, and their high specificity, made microsatellites a valuable tool in the early days of genetic mapping. After their discovery in the late 1980s, microsatellites soon became the markers of choice in genetic mapping (Ellegren, 2004). SNPs are specific known genetic changes that take place on average every 1-2 thousand bases (Sachidanandam et al., 2001). They do not have a high tendency to undergo recurrent mutation and their exact positions in the human genome are known. SNPs allow the genetic makeup of studied samples to be compared with reference samples (Kidd et al., 2006; Tzvetkov and Von Ahsen, 2012). Compared with microsatellites, SNP markers are more numerous in the genome, are stable in mammals and can more easily be incorporated in highly automated systems (Fernández et al., 2013). As a result, SNP arrays have largely replaced microsatellites in mapping studies and are now widely used for mapping various diseases and conditions (Ku C. S. et al., 2010). This has been facilitated by the dramatic improvement in SNP array technology at which the well placed SNPs in haplotype blocks allows getting pretty much association across the whole genome (Netto and Schrijver, 2015).

1.5.2 Homozygosity mapping

Over the past two decades, thousands of genes have been implicated in the causation of human diseases using this approach. Important work on family linkage studies, an approach first proposed in the 1930s (Bernstein and Machol, 1935; Haldane, 1934; Penrose, 1935), revolutionized our understanding of mapping techniques and paved the way for the application of such studies to autosomal recessive disorders. However, the idea of studying consanguineous families and looking in regions which are homozygous by decent for the causative mutations was first proposed by Lander and Botstein in the late 1980s (Lander and Botstein, 1987). Homozygosity mapping, followed by next generation

sequencing is now a well-established method for the identification of mutations that cause Mendelian-inherited disease within consanguineous families (Gilissen et al., 2012). Due to the fact that the technique involves identifying regions that are autozygous, or identical by decent (IBD) as a result of a consanguineous mating, it is also termed "autozygosity" mapping (Mueller and Bishop, 1993).

The main principle of homozygosity mapping (Figure 1.9) is to focus on the genetic regions with homozygous stretches which are shared by the affected individuals, in order to identify the underlying disease allele. The introduction of very high-density SNP genotyping arrays and advances in their analysis allow the entire genome to be screened for homozygous regions relatively easily (Carr et al., 2011). Nowadays, whole genome SNP arrays allow highly informative SNPs to be captured and the most commonly used platforms are from Illumina and Affymetrix (Netto and Schrijver, 2015). The technique has been widely used to identify various loci/genes underlying autosomal recessive diseases and even conditions with extensive genetic heterogeneity in both consanguineous and non-consanguineous families (Edvardson et al., 2016; El-Asrag et al., 2015; Glöckle et al., 2014; Makrythanasis et al., 2016; Ramprasad et al., 2008; Sundaramurthy et al., 2016) .
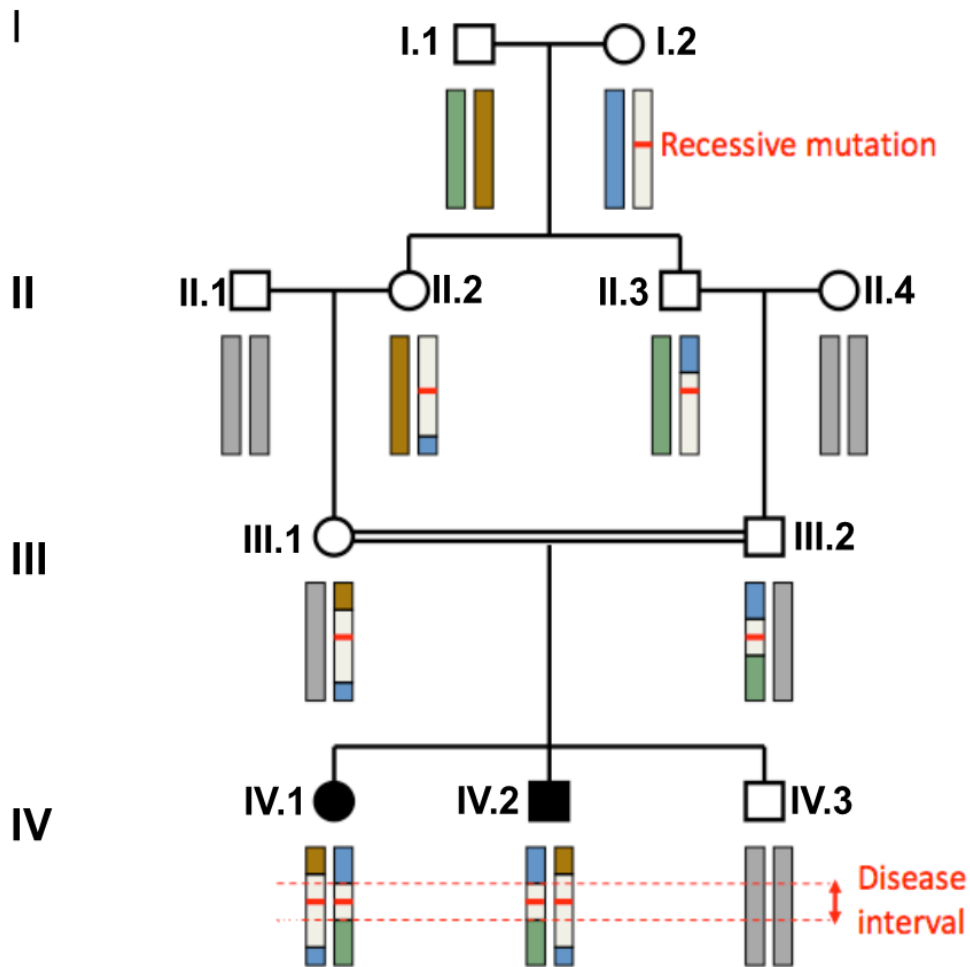
Figure 1.9. Representation of the principle of homozygosity mapping. The great grandmother in the first generation (I.2) of the family carries a mutation, indicated by the red bar, on one copy of one chromosome pair. Colours on the chromosomes denote the grandparental origin of each segment. The mutation is passed on to two of the grandmother's children (II.2 & II.3), who inherit not only the mutation but also the common ancestral region surrounding it, shown in white. The carrier children could then transmit the mutation along with a large portion of the common ancestor regions onto their children (III.1 & III.2). If these grandchildren were then to enter into a consanguineous marriage (depicted by the double line connecting their symbols) then one in four of their children would be expected to inherit two copies the mutation, along with a part of the common ancestral region, from the paternal and maternal sides. In the instance portrayed, two individuals (IV.1 & IV.2) have inherited both mutated copies, resulting in a recessive disease. Examining regions of homozygosity shared by affected individuals for the causative mutations is therefore the main concept of homozygosity mapping. (Adapted from http://autozygosity.org/about/, no permission required)

1.5.3 Next Generation Sequencing, the beginning and current time

The currently-available whole genome sequencing (WGS) has appeared after a large amount of resources over a period of not less than six decades. The discovery of the double helical structure of DNA in 1953 by James Watson and Francis Crick (Watson James D and Crick, 1953) was the first framework on the path towards DNA sequencing. With the availability of restriction enzymes at that time, various trials were made to sequence the genomes of bacteriophages. However, not much progress had been achieved and it was not until 1965, when Robert Holley and his colleagues carried the first RNA sequencing (Holley et al., 1965). At the same time, Frederick Sanger and colleagues were also working separately on a related technique and they showed their exceptional progress in 1977 when they introduced the di-deoxy chain termination method (Sanger, 1977). In fact, intense work was ongoing around this time and various researchers in the field started trying and optimising their techniques. One of the significant moves was made by Allan Maxam and Walter Gilbert who replaced the laborious 2-D fractionation method with the electrophoresis (Maxam and Gilbert, 1977). However, their method involved using some radio-labelled DNA-cleaving chemicals instead of polymerases and di-deoxy chain terminators, which impacted later on it's wide usage. At this stage, it was the real birth of 'first-generation' DNA sequencing and for this work, which facilitated further advancements towards the development of next generation sequencing (NGS), Frederick Sanger and Walter Gilbert were awarded Nobel Prizes in 1980 (Barba et al., 2014). The method of Sanger is the gold-standard way in DNA sequencing and still is used in most laboratories to sequence short DNA of about 200-1000bp.

Various developmental changes then took place gradually for the Sanger sequencing method, which led to the replacement of the radiolabelling materials with fluorescent chemicals (Ansorge et al., 1986; Kambara et al., 1988). Major advances also occurred to the capillary based electrophoresis, which opened the gate for the automation in DNA sequencing (Ansorge et al., 1987). The introduction of the polymerase chain reaction (PCR) (Saiki et al., 1988) and the

subsequent improvement in this technology has participated in the genomics revolution. The golden era of first generation sequencing was achieved with the development of the dideoxy sequencers, such as the ABI PRISM, produced by Applied Biosystems, which at the beginning facilitated the work of the Human Genome Project (Venter et al., 2001) .

The second generation DNA sequencing involved improvements in the strategy used to identify the nucleotides prior to electrophoresis visualisation. The newly introduced method, known as pyrosequencing, is just like Sanger's dideoxy, based on the 'sequence-by-synthesis' (SBS) techniques. This means that for the outcome effect to be detectable, direct action of DNA polymerase is needed (Ju et al., 2006). The pyrosequencing is based on using the enzymatic activities of ATP sulfurylase and luciferase instead of radioactive or florescent materials. The sulfurylase converts pyrophosphate into ATP and the luciferase would then act on the produced ATP, developing light proportional to the amount of pyrophosphate (Nyrén and Lundin, 1985). Because of the major benefits, which include the possibility of using natural nucleotides instead of the heavily-modified dNTPs as well as the fact that it allows a direct result observation without the need for lengthy electrophoreses, this method gained some popularity (Ronaghi et al., 1998). Various improvements were also made to this method especially the incorporation of a special type of dNTPs that allow their enzymatic degradation when they are not consumed, and thus minimizing the need for lengthy washing steps. Pyrosequencing was then licensed to 454 Life Sciences (454; Branford, CT, USA; now Roche, Basel) which developed the first next-generation technology to reach the market , the 454 Sequencer (Rothberg and Leamon, 2008).

The great paradigm move offered by the next-generation technology is its ability to carry out sequencing reactions for a massive number of DNA segments in parallel. With the improvements, the sequencing machines used at that time could produce million or so reads at which each of them is about 400-500bp (Margulies et al., 2005). After the success of the 454 Sequencer, a number of machines have been introduced based on the parallel sequencing techniques

and the field of sequencing has gradually undergone extensive developments during the past 3 decades resulted in the availability of todays' NGS platforms.

The most commonly used NGS technologies are from Roche, Illumina and Life Technologies. Generally speaking, they all use the same concept, based on fragmenting the genome into manageable blocks that can be amplified and sequenced in a massively parallel fashion. Millions of genomic DNA segments are then read simultaneously (Netto and Schrijver, 2015). In fact, recent advances in NGS technology have allowed read length to increase up to 750 bp (Barba et al., 2014). The advances have also reduced the cost of sequencing a human genome. A full genome can now be sequenced for about US$1,000 (Christensen et al., 2015), which has allowed this technology to be applied routinely in the clinical field.

The available benchtop NGS technologies are either based on sequencing by ligation (SBL) or sequencing by synthesis (SBS) technologies (Ho Antoine et al., 2011a). The technology from Life Technologies (such as SOLiD) is based on SBL, in which DNA ligase is utilised to identify which of the dNTPs is available at a particular location within a genomic sequence (Liu Lin et al., 2012). On the other hand, the technologies from both Roche (Ion Torrent) and Illumina (HiSeq) use the SBS approach. According to (Goodwin et al., 2016) Illumina has the top market share and sequencing availability in the clinical laboratories. The Illumina HiSeq, which was the NGS platform used in this study, uses cyclic reversible termination (CRT) and total internal reflection fluorescence (TIRF) microscopy to recognise the identity of the nucleotide (Metzker, 2010), as illustrated in Figure 1.10. The Roche Ion Torrent platform, in contrast, is based on single-nucleotide addition (SNA).
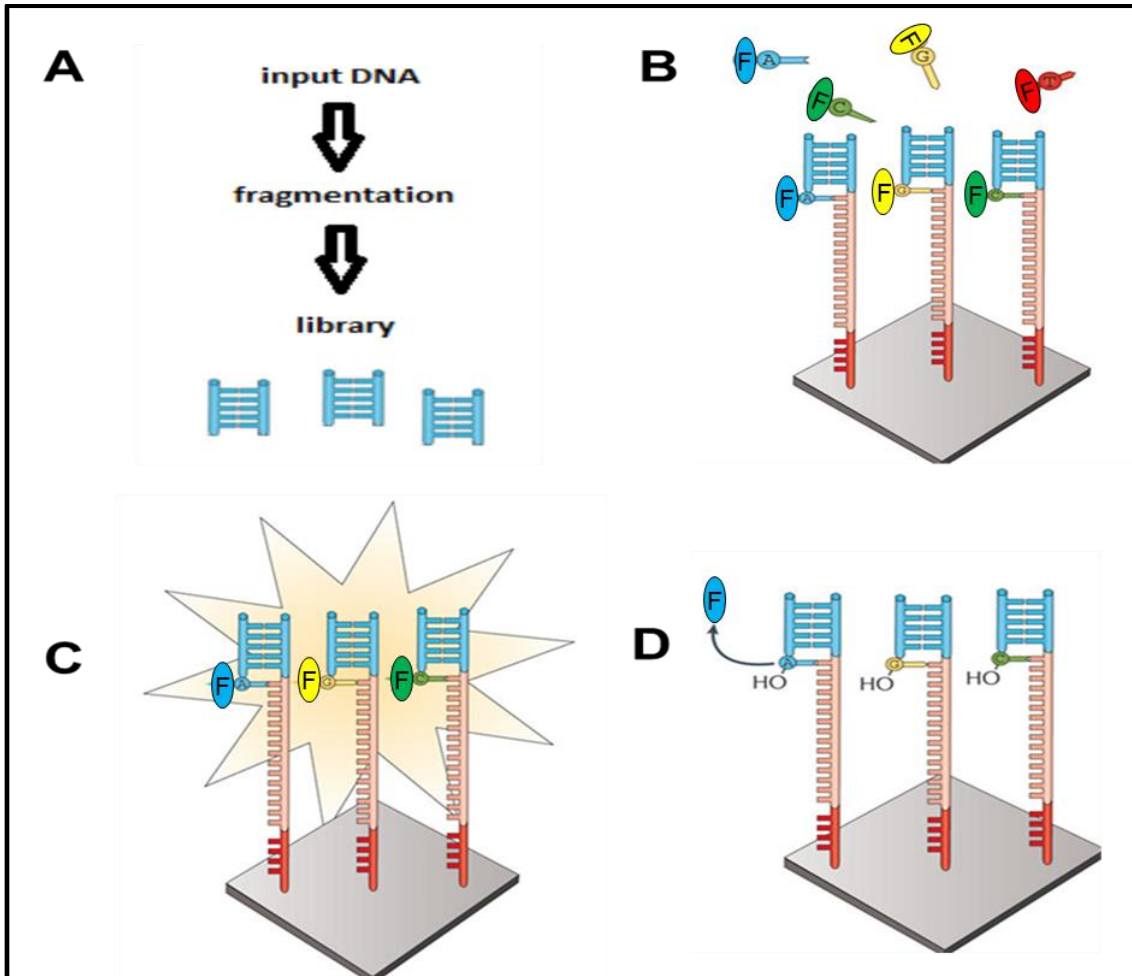
Figure.1.10. Sequencing by synthesis on Illumina platform. (A) Template enrichment step at which the input DNA, which is about 5ug, undergoes fragmentation process followed by end repairing and adaptors ligation steps to form the library. (B) Solid-phase amplification step where the primers, DNA polymerases and modified nucleotides are added to the flow cell. Each nucleotide is specially labelled with cleavable fluorophore (F) and when nucleotides are added in each new cycle, each could only be engaged by one cluster as the 3′ end of each dNTP is already blocked. An adaptor region is available to activate the polymerase binding. (C) Imaging step at which laser channels of the total internal reflection fluorescence (TIRF) microscopy are used. The emitted colour from each cluster is determined by the type of incorporated base at that cycle. (D) Cleavage step where a reducing agent, normally Tris 2-CarboxyEthyl Phosphine (TCEP), is used to break the bonds between the nucleotides and fluorophore (F). This step regenerates a 3′-OH ready for the subsequent cycle. (Adapted from Goodwin et al. (2016) with a permission from the Nature Publishing Group, License number: 4077180975279).

The utilization of NGS has played an extraordinary role in the field of genetics. A tremendous number of studies have been carried out using NGS. The huge expansion in the number of the genes identified for ID in recent years can be attributed to this revolutionary technology (Carvill Gemma L. and Mefford, 2015). NGS technology is also used for applications beyond disease gene identification, including a wide spectrum of analyses of various species of animals, plants and microbes. For example, the acquisition of 27,120 extra projects to the Genome Online Databases (GOLD) by January 2014 was facilitated by NGS technology (El-Metwally et al., 2014).

The identification of certain variants is usually followed by an assessment step in order to evaluate the possible damage caused by these variants. Various *in-silico* tools are freely available to predict the pathogenicity and significance of found variants. They are of great help and becoming more powerful especially in accurate prioritization and selection of the top candidates in Mendelian and complex diseases (Van Der Velde et al., 2015). Examples of these well-known tools are PolyPhen-2, SIFT, mutation taster and CADD. Polymorphism Phenotyping version-2 (Polyphen-2) deducts its score mainly from the footnotes of UCSC genome browser and the vertebrate genome annotation (VEGA) databases. The score of Polyphen-2 ranges from 0 to 1 where higher score means probably damaging and closer to 0 suggests a benign change. The generated value is based on various logarithmic measures including structural features and stability profile in human and vertebrate orthologues (Adzhubei et al., 2010) but a higher impact is also made according to the position of the mutant allele whether it is at a hyper mutable site or not (Schmidt et al., 2008) . SIFT is Sorting Intolerant From Tolerant tool and it uses the annotations of PSI-BLAST (Postion-Specific Iterative Basic Local Alignment Search Tool) to generate the score based on the conservation of the amino acid residues in multiple species. SIFT interpretation is different from Polyphen-2, as higher score indicate tolerance and scores of 0.05 and below mean damaging (Ng and Henikoff, 2003). Mutation taster integrates groups of databases including 1000 Genomes Project, Ensemble, dbSNP and UniPort. This tool provides predication for wide ranges of

variants including the intronic changes, synonymous changes and the ones which span an intron-exon junction (Schwarz et al., 2014).

CADD is Combined Annotation Dependent Depletion tool that considers a huge comparative information from various functional genomic sources like the ENCODE project and UCSC genome browser annotations. Its predication utilizes 63 existing annotations including Ensembl Variant Effect Predictor (VEP), PolyPhen-2 and SIFT to produce one indicative measure called a C-score. This score is a reflection of the deleteriousness, which is more accurate than the pathogenicity score provided by other tools as deleteriousness is not calculated based on particular sets of mutations and thus not affected by ascertainment biases (Kircher et al., 2014). CADD is a great predictor for substitutions and indels and has been used to prioritize and score the variants in this study.

1.5.4 Homozygosity mapping: limitations and the future

Microsatellites have been used effectively in genetics for more than 30 years. However, the use of microsatellites is a labour-intensive process and the analysis is time consuming, even with the use of automated programs (Cantarella and D'agostino, 2015). Moreover, the allele sizes obtained by different research laboratories are sometimes inconsistent (Fernández et al., 2013).

Because SNPs are highly abundant throughout the genome and are amenable to high throughput genotyping methods, they have become the most widely used markers in genetics research (Lijavetzky et al., 2007). Homozygosity mapping in consanguineous families, which uses whole genome SNP arrays, is a very powerful tool for the identification of various disease-causing genes. Affected progeny of consanguineous marriages normally inherit large blocks of homozygous regions, which are likely to harbour the relevant causative gene mutation. However, a significant disadvantage of this approach is the rarity of consanguineous families, especially in Western societies. Although suitable large families can be ascertained from other populations with high endogamy, the health care systems in some countries do not always provide a detailed clinical

assessment. Developing countries tend to have inadequate resources, well-trained specialists, and links between health care institutions, which all affect the quality of the clinical infrastructure (Verma, 2015). An important point that also needs to be considered here is the family size. Big families with multiple affected individuals often lead the researcher straight to the gene because of their ability to provide a solid evidence that co-occurrence of the variant is not likely to be caused by chance but the availability of such families with the DNA from extended members are rare (Weber, 2016). More often now we are screening single consanguineous cases, hence there is a need for a tailored filtering process.

Another disadvantage of homozygosity mapping is the difficulty in identifying the relevant gene, since multiple large regions of homozygosity are often found. Bioinformatics prediction tools prioritise the genes in these regions, but selecting which genes to pursue is a real challenge, especially when the genes have no known function. NGS has facilitated the processes of selecting and identifying the genes when coupled with homozygosity mapping. With the application of tailored filtering criteria, the success rate of NGS for Mendelian disorders during the past few years has been between 60% and 80% (Gilissen et al., 2012). It is now possible to bypass the SNP arrays as it is no more the dominant technology, a powerful approach would involve doing just WES or WGS and then homozygosity mapping on the SNPs that are genotyped in the sequence. Regardless of the applied technology, there will always be thousands of variants that require filtering. Selection of the best candidate genes would also depend on the minor allele frequency (MAF), as rare variants in the population would mean their implication in the studied phenotype is higher. Various literatures have already discussed the significance of gene allele frequency in epidemiological and genetic association studies (Cooper David N. et al., 2013; Goddard et al., 2000; Mattei et al., 2009; Taioli et al., 2004; Xue et al., 2012). Selection of the best candidate is also based on prior knowledge of the function and expression of the genes and of the clinical phenotype in the studied family (Gilissen et al., 2012; Tiffin et al., 2009). Checking whether any of the genes is already known to be associated with the condition under study might quickly identify the most likely causative mutation. Well-established and freely available resources, such as

NCBI, UCSC Genome Browser and Ensembl, help to make this process relatively straightforward.

Compared with whole genome sequencing (WGS), whole exome sequencing (WES) has been more intensively used in clinical diagnostic and research settings (Bamshad et al., 2011; Koboldt et al., 2013). This could be due to the massive optimisation of WES that has been undertaken in the process of detecting thousands of disease-causing variants. The strategy of coupling homozygosity mapping with WES has been widely followed (Alkhiary et al., 2016). However, interest in WGS has increased during the past few years, due to its comprehensive coverage that allows the identification of disease-associated genetic variants outside protein-coding regions (Hrdlickova et al., 2014). WGS is currently more expensive than WES, but costs are quickly dropping to an affordable rate with advances in technology (Gilissen et al., 2014b; Van Nimwegen et al., 2016). WGS is also able to identify the ~3% of coding variants that cannot be detected by WES (Belkadi et al., 2015). However, a major drawback of WGS is the huge number of the variants and the difficulty in interpreting their potential as causative candidates (Cassa et al., 2013).

The rapid revolution among the sequencing technologies made benchmarking the bioinformatics tools of a great necessity. Compared with the widely used WES, at which explicit quality characteristics are available, standardising the bioinformatic tools for WGS can be more challenging. Despite the availability of wide range of bioinformatic tools to facilitate data analysis for next generation sequencing, the interpretation of this new kind of data is accomplished with different issues of exciting and problematic bioinformatic analysis (Pauwels et al., 2015). There have been a need for some innovative computational tools, as the traditional approaches to the classic bioinformatics could not be effective enough to handle this type of data. Moreover, the massively produced raw data by the sequencing platforms could not be underestimated, a single genome comprises hundreds of gigabytes of base calls and quality scores (Oltvai, 2012). Because of that, the huge amount of the generated data implies the need for large data storage and thus all healthcare providers need to consider an extremely generous

data storage before attempting to provide next generation sequencing facilities (Carter Tonia C. and He, 2016; Huang Zhuoyi et al., 2016b). In order to meet the huge storage requirement, cloud-based storage solutions have been suggested and introduced (Nguyen et al., 2011) but the issue of data security has also been raised (Rilak et al., 2014). Security is required for the transfer as well as during the storage of the NGS data. It is very crucial for the applied data management policy to strictly adhere to the privacy laws as well as archival need especially for a long term storage (Schlosberg, 2016). Different systems and platforms have recently been introduced, which have been described to ensure an absolute data protection (Bhuvaneshwar et al., 2015; Chae et al., 2013; Onsongo et al., 2014)

The revolution in the sequencing technologies are resulted on various fascinating outcomes and one of the major advances is the introduction of the Nanopore fourth generation sequencing (Ku Chee Seng and Roukos, 2013; Venkatesan and Bashir, 2011). The idea of using nanopore sensors, a fourth generation technology, to sequence DNA was first proposed during 1990s (Deamer and Akeson, 2000; Kasianowicz et al., 1996). Nanopore sensor technology is based on collecting the physical and chemical properties of a molecule under study by calculating the blockage of current signals initiated by an ion passing through a pore. The flow is influenced by a biased voltage and the properties are mainly considered based on the length of signal interruption as well as the generated amplitude (Venkatesan et al., 2009). Different types of nanopore sensor have been developed and these include biological, solid-state, and hybrid nanopores. The great advantage of this technology is the minimal requirement of sample preparation, which is normally lengthy step in NGS. The technology also implies no amplification and gives results in a much shorter time. It could be the sequencing option of choice in the near future, providing that the limitation of a high error rate is overcome (Feng et al., 2015).

1.6 Project aims

The research described in this thesis was focused on two main areas: intellectual disability (ID) in 6 families recruited in Oman (ID project) and schizophrenia (SZ) in 2 families recruited in the UK (psychosis project). The research hypothesises that "There are genetic causes of ID/SZ in 8 consanguineous families". The following objectives were investigated:

a. To determine regions of shared homozygosity in the affected members of the studied families using SNP array and homozygosity mapping technologies.

b. To apply WES to identify the list of candidate variants and select the ones that are homozygous.

c. To combine the data of homozygosity mapping and WES in order to find the list of causative homozygous variants in these families.

d. To apply various bioinformatics/genotyping methods to filter/select most likely mutations identified.

e. To exclude these mutations if they exist in publically available NGS databases.

f. To Sanger sequence the candidates in the final produced list, so that segregation of the variants could be confirmed.

g. To utilize and carry out certain measures, like position conservation and constraint metrics, that would allow supporting the significance of the identified mutations.

h. To screen extra families with phenotype similar to the studied families, in order to search for a second hit that validate the implication of the identified mutations.

i. To carry out functional analyses of certain mutations to understand more about their implications in the described phenotypes.

j. To use some of the animal models, like flies and mice, to study certain behaviour tasks that are related to learning, memory and cognition.

Chapter 2: Materials and Methods

2.1 Ethical approval, patient recruitment and blood/ saliva sampling

This research involved genetic studies of intellectual disabilities (ID) and schizophrenia (SZ). The study of Omani families was approved by the Sultan Qaboos University Ethical Committee, Muscat, Oman with reference number MRE#751. A copy of the approval is attached in Appendix 1. The families were recruited through consultants at the Genetic and Developmental Medicine Clinic, Sultan Qaboos University Hospital. The main consultants involved in the diagnosis and recruitment of ID patients were Dr Fathiya Al Murshedi, Dr Watfa Al Mamari and Dr Abeer Al Sayegh. Informed consent, using a process that adhered to the tenets of the Declaration of Helsinki, was obtained from the parents of the ID patients. Blood or saliva samples were collected from members of each family, either at the Sultan Qaboos University Hospital or during field trips in different regions of Oman. The collection of blood samples was made by antecubital fossa venepuncture by qualified phlebotomists using EDTA vacutainer tubes whereas saliva was sampled using Oragene saliva collection tubes (DNA Genotek Inc., Ontario, Canada). The collected samples were stored at 4°C not more than 3 days before DNA extraction. The Oragene saliva collection tubes were also used to collect samples from ethnically-matched Omani controls. These controls were recruited from students of the Leeds Omani Society at the University of Leeds, using a process approved by the Leeds Central Research Ethics Committee. A copy of the approval is attached in Appendix 1.

The study of schizophrenia patients within Pakistani families in West Yorkshire was approved by the National Research Ethics Service (NRES) under the title "Genetic investigation of schizophrenia in the Pakistani population of West Yorkshire" with reference number 10/H/1313/37. A copy is attached in Appendix 1. The families were diagnosed by psychiatrists Dr Qadeer Nazar (Lynfield Mount Hospital, Bradford), Dr Alastair Cardno (Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds) and Dr Tariq Mahmood (Leeds and

York Partnership NHS Foundation Trust). Informed consent was obtained from the members of the family to collect blood or saliva. Blood samples from ethnically-matched Pakistani controls (n=154) were collected by Ms Salina Siddiqui (Eye Department, St. James's University Hospital, Leeds). Two-hundred DNA samples from Caucasian subjects with schizophrenia in Glasgow and the West of Scotland were obtained from Dr Catherine Winchester of the University of Strathclyde (Winchester et al., 2012).

2.2 DNA extraction

DNA extraction was done using three different methods, as described below.

2.2.1 Spin column method

DNA was extracted from the blood samples using the Qiagen spin column method according to the manufacturer's instructions (Qiagen Ltd., Manchester, UK). In this method, 20µl of Qiagen proteinase K was put at the bottom of microcentrifuge tubes followed by 200µl of EDTA blood and then 200µl of Buffer AL (lysis solution) before mixing by pulse-vortexing for 15 seconds. The homogeneous mixture was then incubated at 56°C for 15 minutes. Brief centrifugation was given to the tubes after their incubation in order to remove drops from the inside of the lid before adding 200µl ethanol (96–100%) and mixing again by pulse-vortexing. This mixture was then applied to a QIAamp Mini spin column (in a 2ml collection tube), and centrifuged at 8000rpm for 1 minute. After that, the QIAamp Mini spin column was carefully placed into a new 2ml collection tube and the tube containing the filtrate was discarded. Two washes were then applied and the collection tube containing the filtrate was discarded after each wash. The first wash was done using 500µl Buffer AW1 with centrifugation at 8000rpm for 1 minutes and the second wash using 500µl Buffer AW2 with centrifugation at 14,000rpm for 3 minutes. For the sake of eliminating any possible Buffer AW2 carryover from the previous step, 1 minute centrifugation at 14,000rpm was then given to the QIAamp Mini spin column. The QIAamp Mini spin column was placed in a new microcentrifuge tube and 200µl of Buffer AE (elution agent) was added into the column. After 1 minute incubation

at room temperature (RT), 1 minute centrifugation at 8000 rpm was given to elute all the attached DNA. The DNA samples were then stored at -20°C.

2.2.2 Salt precipitation method

This method was used to extract DNA from blood samples of the psychosis families. It was started by mixing 3ml of blood/EDTA with 9 ml of Red Cell Lysis solution (0.32M sucrose, 10mM Tris HCl, 5 mM $MgCl_2$ and 0.75% Triton-X-100) in a polypropylene tube and then centrifugation for 10 minutes at 3,200 rpm. After careful removal of the supernatant, the pellet was re-suspended in 3ml of White Cell Lysis solution (10 mM Tris _HCl, 26 mM EDTA and 0.5% SDS). Multiple and gentle pipetting up and down was applied in order to lyse the cells properly. Then, 1ml of 10M Ammonium Acetate (Protein Precipitation Solution) was added and the samples were vortexed for 20 seconds followed by centrifugation at 3,200rpm for 10 minutes. A new tube was used to collect the supernatant that contained the DNA material. DNA precipitation was done by adding 3ml of isopropanol and centrifugation at 3,200rpm for 10 minutes. The precipitated DNA was washed twice, each with 500 μl of 70% ethanol, and the supernatant was decanted. After air drying, the DNA pellet was redissolved in 200μl of 1 x TE buffer (Tris-EDTA Solution). In order to make sure that all the DNA dissolved in TE buffer, the samples were incubated at 55°C for 4 hours in a shaking water bath and then left for 1 hour at RT before storage at -20°C.

2.2.3 Saliva Oragene kits

Saliva was collected from healthy Omani individuals in Oragene ON-500 tubes (DNA Genotek Inc., Ontario, Canada) and DNA was extracted following the manufacturer's protocol. In this method, the saliva tubes were first incubated at 50°C in a water bath for 1 hour before transferring 750μl of saliva to a 1.5ml tube and adding 30μl of PT-L2P solution (purifying agent) supplied with the kit. After that, the samples were incubated on ice for 10 minutes followed by centrifugation at RT for 10 minutes at 13,000rpm. The clear supernatants of each sample (about 600μl) were then transferred to fresh tubes and the remaining pellets were discarded along with the tubes. Subsequently, 720μl of 100% ethanol were added

to the transferred supernatant and the tubes were inverted 10 times. The samples were allowed to stand at RT for 10 minutes then centrifuged for 10 minutes at 13,000rpm. The supernatant was carefully removed from each tube and discarded, leaving the pellets which were then washed with 500µl of 70% ethanol. After washing, the ethanol was removed without disturbing the DNA pellet and the tubes where kept for about 30 minutes at RT for the pellets to dry. Once dry, 50µl of TE buffer was added to each pellet and the samples were incubated overnight at RT to ensure complete rehydration. Finally, 50µl DNA was transferred to a 1.5ml tube which was later stored at -20°C.

2.3 First strand cDNA synthesis

This method was used to synthesize complementary DNA (cDNA) from the blood lymphocyte RNA, fibroblast RNA as well as the human Total RNA Master Panel II (Clontech Laboratories, Inc). The technique was carried out using a first strand cDNA synthesis kit (Thermo Scientific) following the manufacturer's instructions. The method involved adding 1µl (0.1-5µg) of total RNA and 1µl of random hexamer primer (15-20pmol) to 9µl nuclease-free water in a tube on ice. After gentle mixing and brief centrifugation, the tube was incubated at 65°C for 5 minutes and then chilled on wet ice, followed by a quick spin down before being placed back on ice. While on ice, another tube was placed on ice to prepare a master mix that had 4µl of 5X Reaction Buffer, 1µl of RiboLock RNase Inhibitor (20U/µl), 2µl of 10mM dNTP mix and 2µl of M-MuLV Reverse Transcriptase (200U/µl). After mixing, 9µl of master mix was added to the tube containing the RNA and the primer. This tube was then mixed and briefly centrifuged before incubation for 5 minutes at 25°C, followed by 60 minutes at 45°C and then heating at 70°C for 5 minutes to terminate the reaction. The products of the reverse transcription reaction were stored at -20°C until the PCR was done.

2.4 Primer design

All the sequences of the regions of interest were downloaded from the UCSC Genome Browser, build hg19 (February 2009). Most of the primers were designed using the Primer3 program (http://primer3.ut.ee/) to produce a PCR product of 200-800 base pairs (bp). PCR products were designed to amplify exons of interest as well as the flanking splice site junctions. The primers designed for use in the reverse transcriptase RT-PCR amplification of transcripts were synthesised so that they would span an entire exon, allowing both cDNA and any contaminating genomic DNA to be detected. Some primers to be used in site-directed mutagenesis were designed using the Agilent website (http://www.genomics.agilent.com/primerDesignProgram.jsp). Sequence and details of all the primers are given in Appendix 2.

2.5 Polymerase chain reaction (PCR)

The PCR, including the reverse transcriptase RT-PCR, was normally carried using the standard method, but when this was not successful, the Hot-Shot master mix was used. Both methods are described below.

2.5.1 Standard PCR

PCR was typically performed in 25µl volumes containing 100ng genomic DNA, 1µl of 10x PCR buffer (Invitrogen), 1µl of each forward and reverse primer (10µM stock), 200 µM dATP, dGTP, dCTP and dTTP nucleotides (Invitrogen) and 0.5U Taq DNA polymerase (Invitrogen). Thermal cycling was performed on this mixture with an initial denaturing step of 95°C for 2 minutes then 40 cycles as follows: 94°C for 30 seconds (denaturing), 50-65°C for 30 seconds (annealing) and 72°C for 45 seconds (extension). Following these cycles, a final extension step was performed at 72°C for 5 minutes.

2.5.2 PCR with Hot-Shot master mix

This reaction was prepared by adding 5 µl of Hot-Shot master mix (Clent Life Science), 1 µl of DNA, 3.5 µl of water and 0.25 µl of each forward and reverse

primers. After gentle mixing, the mixture was subjected to the same thermocycling conditions as in the standard PCR method, except that initial denaturation time was 10 minutes instead of 2 minutes.

2.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used to detect the presence and size of PCR products. Gels contained between 1% and 2% (w/v) of ultra-pure agarose (Fisher Scientific, UK). To prepare a gel, the desired amount of agarose was first dissolved in 0.5X TBE buffer by heating in a microwave oven. The solution was cooled to 55°C, mixed with ethidium bromide (Sigma-Aldrich) at a final concentration 0.5 µg/ml, and then poured into a gel tray, to a depth of between 5 and 8mm. Gel combs were inserted. When the gel had solidified, the combs were removed to create wells, and the gel was submerged under 5 mm of 0.5X TBE buffer, in an electrophoresis tank. Wells were loaded with 10ul PCR products that had been diluted with 6X DNA loading dye (30% glycerol 0.25% bromophenol blue and 0.25% xylene cyanol FF). 5µl DNA molecular weight marker (Bioline, UK) was also loaded into a well. The gel was then ran at 100V for 45 minutes. DNA bands within the gel were visualized with a ChemiDoc™ MP System (Bio-Rad) over ultra-violet (UV) light (wave length 320nm). The image was focused through Image Lab™ software (Bio-Rad).

2.7 Genotyping by microsatellite testing

In order to make sure that no mixing of DNA samples had taken place, and for the purpose of verifying sample identity, each DNA sample within a family was subjected to microsatellite testing using three different markers. The markers were D1S478, D4S3042 and D7S483. In brief, the test was done by first carrying out standard PCR (section) using the marker primers (10µM each; Sigma-Aldrich). An aliquot of the PCR products was first visualised on a 1.5% agarose gel by electrophoresis. Then, the PCR products were diluted between 1:10 to 1:70, depending on the intensity of the PCR aliquot on the gel, in distilled water before taking 1 µl and mixing with 8.5µl Hi Di formamide (Applied Biosystems) and 0.5µl of ROX-500 ladder (Applied Biosystems). The diluted PCR products

were then run on an ABI3130xl capillary Genetic Analyser. Finally, the genotyping results were analysed by Gene Mapper software (v3.7). The possible allele sizes for each of the markers used were found on the CEPH website (http://www.cephb.fr/hgdp/main.php). Detailed information about the genotyping markers is given in Appendix 3.

2.8 Clean-up of PCR products before Sanger sequencing

The products were cleaned up using ExoSAP-IT (GE Healthcare Life Science), which contains both exonuclease I and shrimp alkaline phosphatase (SAP). The clean-up reaction was done by adding 2μl of ExoSAP-IT to 5μl of PCR product and incubating on a thermocycler for 30 minutes at 37°C, followed by a deactivation step of 15 minutes at 80°C.

2.9 Sanger DNA sequencing

A plate was set up for sequencing reaction using 0.5 μl of BigDye® Terminator Kit v1.3 (Applied Biosystems) with 1.5μl of 5x sequencing buffer and reverse or forward primer (1μl of 1.6μM) and 1μl of cleaned PCR product in a total volume of 10μl. The mixture was subjected to an initial denaturation step of 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. DNA precipitation was carried out by adding 5μl of 125mM EDTA (pH 8.0) and 60μl of 100% ethanol followed by centrifugation for 30 minutes at 22°C and 3000 g. After that, the fluid contents were replaced by 60 μl of 70% ethanol and centrifuged for 15 minutes at 4°C and 800g. The ethanol was then decanted and the plate was completely dried by centrifugation in an inverted position (1 minute at 10g) followed by air drying of about 20 minutes at RT to allow evaporation of excess ethanol. 10μl of Hi Di Formamide (Applied Biosystems) was used to hydrate each DNA sample before putting the plate on the ABI3130xl Genetic Analyser (Applied Biosystems). Sequencing visualization and analysis was carried out against template sequences from UCSC Genome Browser build hg19 (February 2009) using SeqScape v2.5 and Sequence Analysis v5.2 softwares (Applied Biosystems).

2.10 Whole genome homozygosity mapping

This method of mapping was performed on the DNA of affected family members using Affymetrix SNP 6.0 arrays by a commercial company (AROS Biotechnologies Ltd., Aarhus, Denmark). A flow diagram indicating the main steps involved in the protocol is outlined in Figure 2.1. Basically, this technique genotyped the DNA samples at numerous SNP markers spread across the genome. The Affymetrix® Genome-Wide Human SNP Array 6.0 included 906,600 SNPs and 945,826 copy number probes, making it a powerful and cost-effective tool for the identification of genes involved in rare, monogenic disorders.

2.10.1 Data analysis of the genotyping data

Affymetrix array data, received from Denmark, was analysed using IBDfinder software (http://dna.leeds.ac.uk/ibdfinder/), which produces a graphical plot on which the identical by descent (IBD) regions of all the chromosomes are clearly defined and highlighted. The data was also visualised by extracting SNP data from WES using Agile MultiIdeogram software (http://dna.leeds.ac.uk/auto-zygosity/) which was set to detect runs of homozygosity (ROH) of 1 Mb and above. It generates a circular ideogram in which the genotyping data is displayed as an inner ring of human autosomal chromosomes next to one another. Each ring corresponding to the genotyping data of a single individual. .

2.11 Whole exome next generation sequencing

Library construction was done according to the protocol of 'SureSelect Target Enrichment System for IIlumina Paired-End Sequencing', from Agilent. Figure 2.2 summarises the steps involved.
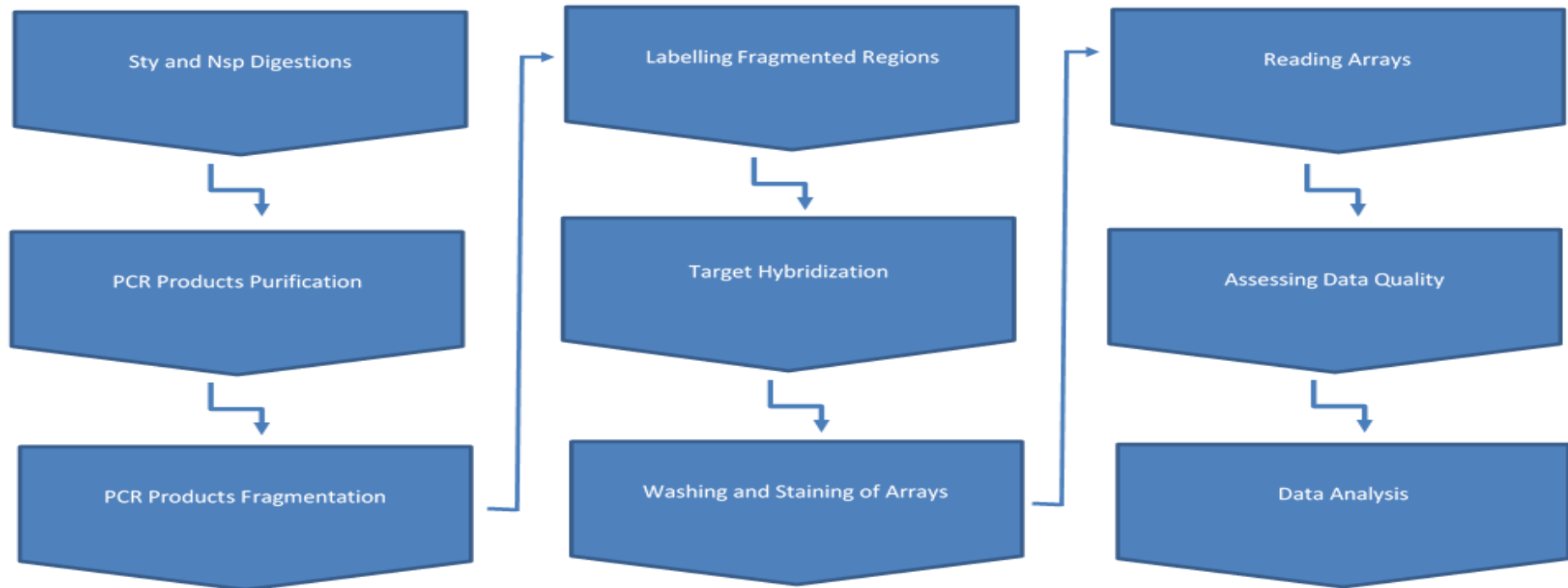
Figure 2.1. A flow diagram indicating the main steps involved in Affymetrix SNP 6.0 arrays. The process was started by setting two plates for Sty and Nsp digestions which were then pooled and purified. After that, the sample was fragmented in a thermocycler and labelled in ice using specific probes. The step of target hybridization was done by injecting each sample into an array chip (Affymetrix® Genome-Wide Human SNP Array 6.0) and placed it into a hybridization oven. Then, hybridization cocktails were exected from the samples followed by washing/staining processes on a GeneChip® Fluidics Station 450. The arrays were read in a GeneChip Scanner 3000 7G and the quality of the genotyping data was assessed by producing a metric Contract QC. Data anlysis was done using IBDfinder and AgileMultiIdeogram softwares.
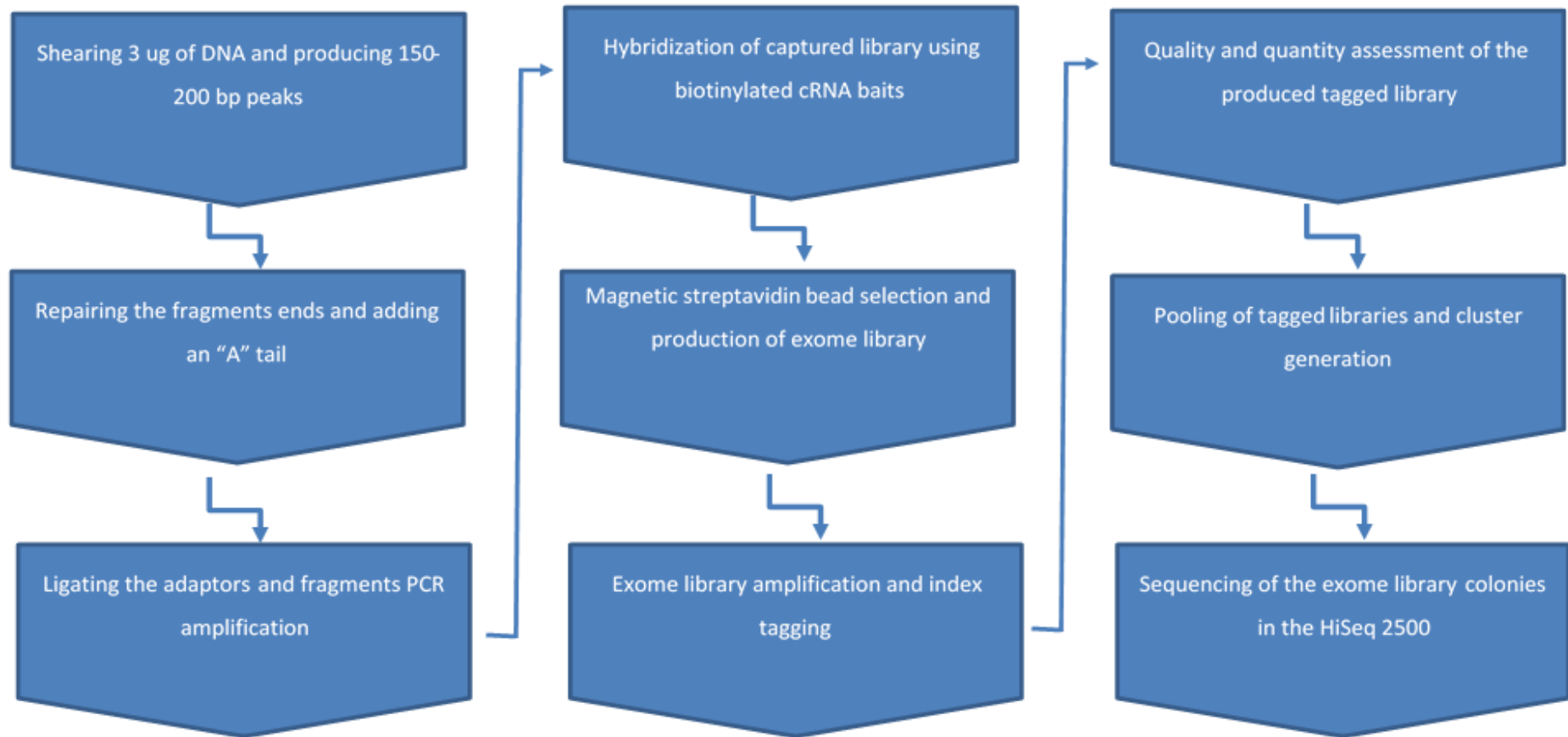
Figure 2.2. A flow diagram showing the main steps involved in constructing library for Whole Exome Next Generation Sequencing. Library preparation is a multitask process and each step normally involves using various reagents. Detailed description about the steps, the used reagents and their quantities could be found in the written paragraphs following this figure.

2.11.1 DNA shearing

In brief, 3μg of genomic DNA was diluted with 250μl 1X TE buffer and put in a secured glass tube (Covaris, USA) which was then inserted into a Covaris s220 machine (Sonicator with Sonolite software, Applied Biosystems) for 5 minutes of sonication to shear the DNA into 150-200bp fragments. The samples were then purified using Agencourt AMPure XP beads (Beckman Coulter Genomics, South Plainfield, USA). After purification, the quality of the sheared DNA was checked by loading 1μl of the sample on a DNA 1000 chip and run on a 2100 Bioanalyzer (Agilent Technology). The results were assessed and visualized using Agilent 2100 Expert Software which produced an electropherogram specific for each sample. Samples which showed a distribution curve with a peak size between 150 to 200 nucleotides in the electropherograms (Appendix 4) were considered to have passed assessment and were thus ready to proceed to the next step.

2.11.2 End-repairing

End-repairing was carried out so that the fragments would have blunt-ends for creating overhangs. This was achieved by taking 48μl of each sample and mixing it with the following reagents from the SureSelect Library Prep Kit (Illumina): 32.2μl of nuclease-free water, 10μl of 10x end repair buffer, 1.6μl of dNTP mix, 1μl of T4 DNA polymerase, 2μl of Klenow DNA polymerase and 2.2μl of T4 polynucleotide kinase. This mixture was incubated for 30 minutes at 20°C. The samples were then purified again using Agencourt AMPure XP beads. After that, adenosine overhangs were added to the 3' ends of the fragments (3'-dA overhangs) by adding 11μl of nuclease-free water, 5μl of 10x Klenow polymerase buffer, 1μl of dATP and 3μl of Exo(-) Klenow to 30μl of the purified samples, and keeping the mixture in a thermal cycler for 30 minutes at 37°C. After that, the samples were purified using Agencourt AMPure XP beads.

2.11.3 Adaptor ligation and library amplification

Ligation of indexing-specific paired-end adaptor was carried out by adding the following reagents to 13μl of the sample: 15.5μl of nuclease-free water, 10μl of 5x T4 DNA ligase buffer, 10μl of SureSelect adaptor oligo Mix and 1.5μl of T4

DNA ligase. The ligation was incubated for 15 minutes at 20°C in a thermal cycler, followed by purification using Agencourt AMPure XP beads. Each adaptor-ligated library was amplified by adding the following reagents to 15ul of the sample: 21µl of nuclease-free water, 1.25µl of SureSelect primer, 1.25µl of SureSelect Illumina Indexing pre-capture PCR reverse primer, 10µl of 5x Herculase II reaction Buffer, 0.5µl of 100mM dNTP mix and 1µl of Herculase II fusion DNA polymerase (all from Agilent kits). The amplification was done using the following thermocycling conditions: initial denaturing step of 98°C for 2 minutes, then 6 cycles of 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute (extension), followed by final extension at 72°C for 10 minutes. There was then a purification step using Agencourt AMPure XP beads.

2.11.4 Library hybridization

Prior to the hybridization step, the samples were assessed using the DNA 1000 chip run on an Agilent 2100 Bioanalyzer. A sample was considered to have passed this check when its electropherogram showed a single peak in the size range 250 to 275bp. After assurance that the samples had the correct quality and quantity, library hybridization was initiated using working reagents of hybridization buffer, SureSelect capture library and SureSelect block mix. The working hybridization buffer was prepared from four different SureSelect buffers: 25µl Hyb#1, 1µl Hyb#2, 10µl Hyb#3 and 13µl Hyb#4. Similarly, the SureSelect Capture Library, which was based on 4 Gb of targeted region, was prepared using 5µl of SureSelectXT Human All Exon V4 mixed with 2µl RNase Block Dilution that was prepared earlier by mixing 75% water with 25% RNase Block solution. The preparation of SureSelect Block Mix was done by mixing 2.5µl SureSelect Indexing Block #1, 2.5µl SureSelect Block #2 and 0.6µl SureSelect Indexing Block #3. After preparation of these reagents, 3.4µl of amplified DNA (750 ng) was mixed with 5.6µl of SureSelect Block Mix and 40µl of the capture libraries. This mixture was incubated for 5 minute at 95°C and then hybridized at 65°C overnight with a heated lid at 105°C.

2.11.5 Selection hybrid capture

Once the hybridization step was complete, hybrid capture selection was done by taking 23µl of hybridised library, while still at 65°C, and adding it to 200µl of prepared magnetic beads (50µl of Dynabeads MyOne Streptavidin T1 washed and resuspended in 200µl SureSelect Binding Buffer). The solution (hybrid-capture and beads) was incubated on a rotator mixer for 30 minutes at RT. During this time, the biotinylated capture-sequence will be attracted to the beads. After that, the tubes were put on a magnetic stand which pulled down all the hybrid captured beads, leaving clear supernatant that was discarded. Then, 500µl of SureSelect Wash1 Buffer was added to the beads and incubated for 15 minutes at RT. Beads were washed three times, each with 500µl of SureSelect Wash2, and the supernatant discarded after putting the tubes on the magnetic stand. After the last wash, the beads were resuspended in 30µl nuclease-free water.

2.11.6 Addition of index tags

For the purpose of sample identification, a specific tag was added to each sample via the PCR. So, the capture library was amplified by taking 14µl of captured on-bead DNA and mixing it with 22.5µl of nuclease-free water, 10µl of 5x Herculase II reaction buffer, 0.5µl 100mM dNTP mix, 1µl of Herculase II fusion DNA polymerase, 1µl of SureSelect Illumina Indexing post-capture forward PCR primer and 1µl from PCR primer Index. After thorough and gentle mixing, the tubes were run on a thermocycler as follows: 2 minutes denaturation at 98°C, 12 cycles of 30 seconds at 98°C, 30 seconds at 57°C and 1 minute at 72°C, followed by an elongation step of 10 minutes at 72°C. There was then a purification step using Agencourt AMPure XP beads before the samples were assessed using a High Sensitivity DNA assay run on the Agilent 2100 Bioanalyzer. The samples which showed electropherograms with a peak in the size range of 300-400 nucleotides were considered suitable for next generation sequencing.

2.11.7 Sample pooling

Samples were pooled in batches of six with a total volume of 50µl. The pooling was performed considering each index-tagged library was present in the mixture

at a concentration of 10 nM. Samples were combined based on the following formula: Volume = (V(f) x C(f)) / (6*C(i)), where V(f) is the final desired volume of the pool, C(f) is the desired final concentration of the pool (10 nM) and C(i) is the initial concentration of each index sample. Once all the samples were pooled, they were taken to the Leeds Sequencing Facility, at which they were prepared for cluster amplification and loaded on an Illumina HiSeq 2000/2500 platform.

2.11.8 Cluster amplification

This step was performed at the NGS facility (University of Leeds. Briefly, sample preparation was done using a TruSeq Cluster Generation kit (IIumina). To start, 3µl (30 fmol) was removed from each pool and diluted with 16µl of Buffer EB (10 mM Tris Cl, ph 8.5) and 1µl of HP3 (2 N NaOH). After brief mixing and pulse centrifugation, the HP3was allowed to denature the DNA by incubating for 5 minutes at RT and then immediately on ice. Final dilution was done by adding 992µl of pre-chilled HT1 (Hybridization Buffer) to 8µl of the denatured DNA. The samples were then treated with TruSeq SBS Kit v5-GA (36-cycle) before loading into an Illumina HiSeq 2500 Genome Analyzer (Illumina Inc. UK).

2.11.9 Analysis of whole exome next generation sequencing output

Analysis of the WES data is a multistep task that involved using certain UNIX console commands (Appendix 5) plus several of the online available facilities. The first step was to score the quality measures of the raw data obtained from the Illumina Higseq 2500

2.11.9.1 Assessing quality

This quality check was performed (see section) by processing the data using FASTQC on the Galaxy platform (Blankenberg et al., 2010). The generated quality scores showed a wide range of measurements which reflected how well was the analysed sample. Examples of these measurement and generated QC charts could be found in Appendix 4. To assure that the raw IIlumina sequencing data were of good quality with no problems or biases, they were loaded in bam

format into FastQC software (http://www.bioinformatics.babraham.ac.uk /projects/fastqc/) for quality checking. This gave a full report with quality scores, sequence length distribution and adaptor contents. In order for a sample to pass the quality control test, the mean of its Phred score had to be above 30, indicating an accuracy of >99.9%. Once the samples passed quality assessment, they were then taken for the main steps of exome analysis.

2.11.9.2 Alignment of reads

Once the sample passed the quality check, it was then taken to the alignment step at which the sequencing reads were aligned against the reference genome (hg19/GRCh37) using Bowtie2 (Langmead and Salzberg, 2012). It should be mentioned here that alignment of the samples for CNVs detection was made using NovoAlign software (http://www.novocraft.com/products/novoalign) as it was preferred for exome depth analysis. Then, the sample was treated in SAM/BAM format using SAMtools (Li Heng et al., 2009) in order to sort and index the sequencing reads. Samtools software (http://samtools.sourceforge.net/) was used to convert the produced SAM (Sequence Alignment/Map) files into BAM (the binary version of a SAM) format. This process involved viewing, sorting and indexing reads before the files were treated with Picard tools (http://broadinstitute.github.io/picard/) to remove all the duplicates. Local realignment was then carried out using the Genome Analysis ToolKit (GATK) software package (https://www.broadinstitute.org/gatk/). After that, all the indels and single substitution changes were called in VCF format using the Unifield Genotyper function of the GATK as described by Depristo et al. (2011). In order to visualise the aligned reads, the sample was then processed with the Integrative Genomics Viewer (IGV) (http://software.broadinstitute.org/software/igv/) (Robinson et al., 2011).

2.11.9.3 Variant annotation

Annovar software (http://annovar.openbioinformatics.org/en/latest/) was used to annotate the generated list of variants (Wang Kai et al., 2010). The list would be in VCF format and Annovar would express the variant's chromosome start

position, end position, reference allele and observed allele. Any amino acid change, conservation and frequency were also annotated.

2.11.9.4 Variant filtering

After all the variants were annotated, they were transferred into Microsoft Excel for manual filtering (Figure 2.3). As consanguineous families were being studied, lots of variants were removed based on their zygosity and whether they mapped outside the homozygous regions. The filtering also excluded any variants not present in the exonic regions or flanking splice sites, and variants that showed synonymous changes. Because of the expected scarcity in mutations causing rare diseases such as intellectual disability, any variants with more than 1% minor allele frequency (MAF) in the 1000 Genomes database were removed. All the variants which showed a small read depth of 10 or less were also excluded as they might not be true candidates. Bioinformatic scores generated by the prediction softwares (Polyphen, SIFT or CADD) were used to prioritize the testing of the variants for segregation. When an X-linked mode of inheritance was suggested, all the X-chromosome variants with CADD scores of >15 were tested.

2.11.9.5 Compound heterozygote variant

For the families in which the causative mutation could not be identified, the results of homozygosity mapping were not used. In fact, the filtering criteria for selecting the compound heterozygous variants involved removing all the homozygous variants from the exome sequencing lists. The genes which showed only one variant were removed from the list. Therefore, the list only included the heterozygous variants at which at least 2 of them are in the same gene. After that, all the synonymous and intronic change were removed. The list only included the variants with less than 1% minor allele frequency in the 1000 Genomes database and read depth of more than 10. Again, bioinformatic scores generated by the prediction softwares (Polyphen, SIFT or CADD) were used to prioritize the testing of the variants for segregation.
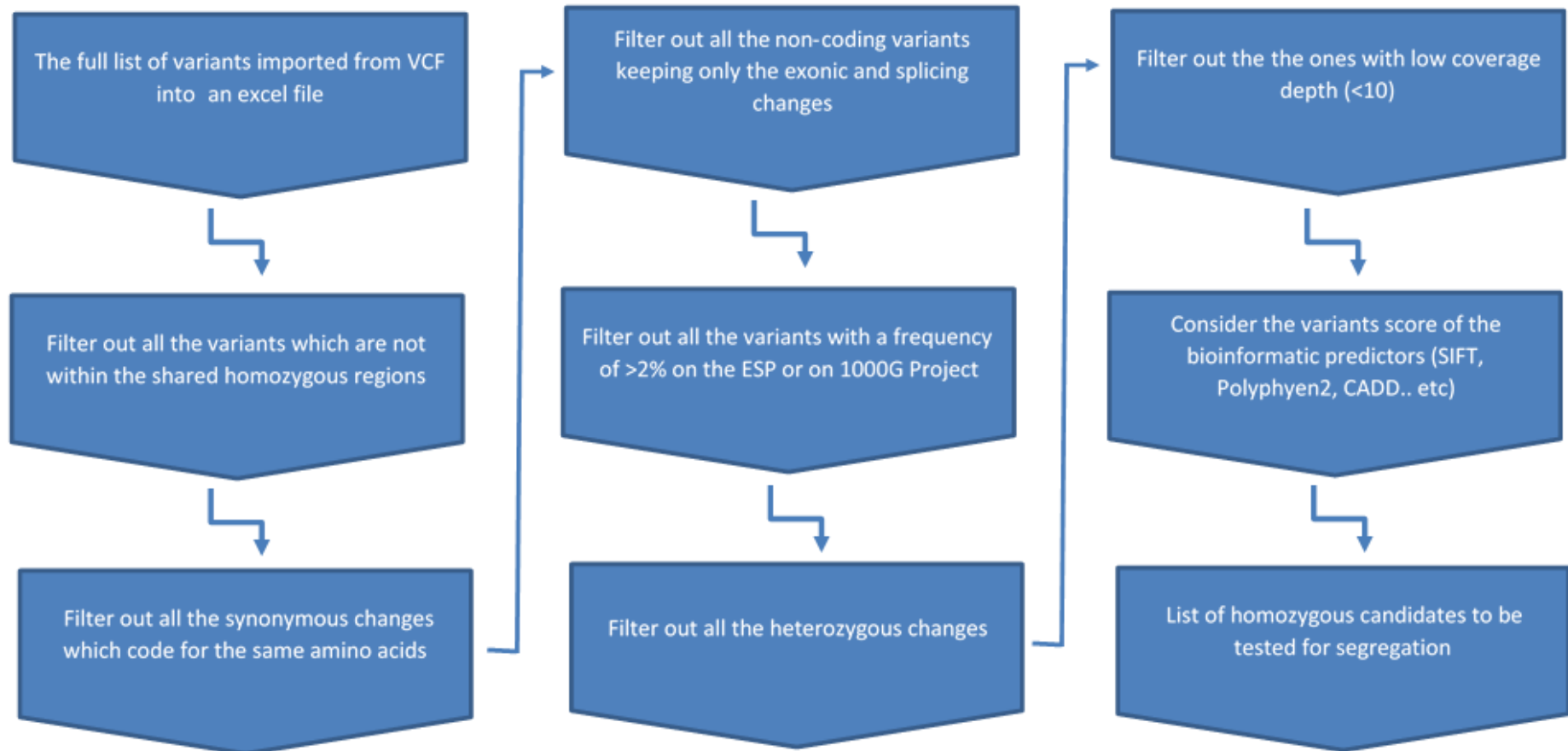
Figure 2.3. A flow diagram showing the steps involved in filtering the variants for recessive disease with homozygous variants. It should be mentioned that the above steps were not necessarily done following a strict order, as the same final list is normally achieved as long as all the steps are carried out.

2.11.9.6 X-linked variants

For the families at which their structure suggested an X-linked inheritance mode, all the variants located on X chromosome were first selected. The filtering criteria was applied to remove all the changes that were synonymous, intronic, with low depth of coverage (<10) or with > 1% MAF. The list included the variants in the exonic regions and flanking splice sites. The variants were then prioritised based on their scores, obtained by the prediction softwares (Polyphen, SIFT or CADD). An updated list of X-linked genes, that had been already known to be associated with ID, was used to highlight some of the prioritised variants. The list of the known X-linked ID genes is managed by the University of Colorado Denver and a copy of the list is attached in Appendix 8. Segregation of the candidate variants was then investigated by Sanger sequencing and checking if any variant was consistent with being the cause of ID in the studied family.

2.11.9.7 Copy number variants

The main principle of the ExomeDepth analysis is to look for structural rearrangements based on a comparison between the read depths of the sample and a small set of control samples. It is important for the used control samples to have been treated with the same version of library reagents and sequenced in the same run. The analysis was carried out as described by Plagnol et al. (2012). Briefly, the fastaq files of the samples were processed as for whole exome analysis (see Section 2.11.9) but the "removal of duplication" step was not done. The indexed bam file of each sample was treated by GATK using the R package ExomeDepth (https://cran.r-project.org/web/packages/ExomeDepth/index.html). The used UNIX console and R commands are shown in Appendix 5. The obtained CNV calls were prioritised based on Bayes factor (BF), which is the $\log_{10}$ of the likelihood ratio of data for the CNV call divided by the call of the normal copy number. There is no ideal threshold for BF but a higher number suggests a potential CNV. Zygosity of the candidate CNV was determined based on the ratio between the expected and observed reads. All the heterozygous variants, with reads ratio of 1 and 1.5, were excluded and the final lists included only the homozygous variants which had read ratio of 0 and 2.

2.12 Download from UK10K database

UK10K (http://www.uk10k.org/) is a database providing whole exome sequences of 4000 healthy individuals and 6000 subjects with neurodevelopmental disorders, obesity and other rare diseases. A request was made to allow the members of Leeds Psychiatric Genetics Group (Appendix 1) to access the UK10K database. After granting the authorisation access, the related cohorts of data were downloaded. Three cases cohorts were selected for the download as they have samples for individuals with various degree of intellectual disabilities/learning disabilities. These cohorts were; UK10K_NEURO_MUIR, UK10K_NEURO_ASD_GALLAGHER and UK10K_RARE_FIND. For the purpose of comparing the changes of significance in the cases, two cohorts of samples for healthy individuals were also included and these were UK10K_COHORTS_TWINSUK and UK10K_COHORT_ALSPAC. The samples of interest within datasets were downloaded using FileZilla download client (https://filezilla-project.org/download.php) in order to get VCF files. The total numbers of downloaded samples were 469 cases and 2432 controls.

Once all the samples were downloaded, multiple analysis was carried out which involved using various UNIX commands (Appendix 5). For security reason, all the samples were only available for downloaded as encrypted data. The first step was to decrypt all the downloaded files and a separately supplied code was used. After that, vcfhacks (https://sourceforge.net/projects/vcfhacks/) was used to search for all the variants that were present in *ANKRD2* or *PDZD8*. This step was very important as it helped to focus only in specific genomic regions of interest and allowing an easy manipulation of the data. The variants were then annotated using variant effect predictor (VEP) (http://www.ensembl.org/info/docs/tools/vep/index.html). After that, all the variants were scored using CADD (http://cadd.gs.washington.edu/) and the ones with <10 were filtered out. Then, the VCF list was converted Microsoft excel and all the non-functional variants (synonymous, intronic, etc.) were filtered out.

2.13 Site-directed mutagenesis (SDM)

A construct, containing wild-type human *DFNB31* transcript 3 (UC004bja.4) cloned into ampicillin-resistant pDEST-733, was received from Dr Erwin van Wijk of Radboud University, Nijmegen, Netherlands. The full sequence of the construct, including an mRFP tag (Monomer Red Fluorescent Protein), is shown in Appendix 6. The main aim of site directed mutagenesis is to introduce a substitution change into a genome (Carter Paul, 1986). The process was done by using QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) to make a change in the DFNB31 insert so that it resembles the reported variants (R450C).

2.13.1 Preparation of ampicillin LB (Luria Broth)

The following components were added to 900 ml of distilled water in a 1 litre Pyrex bottle: 9g of tryptone, 4.5g yeast extract and 9g NaCl. These were thoroughly mixed and autoclaved for 15 minutes at 15 Pa pressure. When the mixture had cooled down, the bottle was labelled and stored at 4°C.

2.13.2 Preparation of ampicillin LB agar

The following components were added to 400ml of distilled water in a 500ml Pyrex bottle: 4g tryptone, 2g yeast extract, 4g NaCl and 6g agar. These were thoroughly mixed and autoclaved for 15 minutes at 15 Pa pressure. When the mixture had cooled down, but before solidification, 400µl of 100mg/ml ampicillin (Sigma-Aldrich) were added and mixed thoroughly. The medium was then poured into petri dishes, labelled and stored at 4°C.

2.13.3 Protocol of SDM

The main steps are shown Figure 2.4 and the construct sequence is shown in Appendix 6. Mutagenic primers (Appendix 2) were designed using the Agilent website (http://www.genomics.agilent.com/primerDesignProgram.jsp). PCR was done using 50ng of the plasmid and high-fidelity DNA polymerase (Pfu Turbo, Agilent Technologies). The pair of primer for the DFNB31 variants (R450C

mutation) was used to introduce the target changes into the wild-type transcript. Thermal cycling was performed with an initial denaturing step of 95°C for 30 seconds, followed by 16 cycles of 95°C for 30 seconds (denaturing), 55°C for 1 minute (annealing) and 68°C for 8 minutes (extension).

The PCR product was then digested with a DpnI restriction enzyme (New England Biolabs) by adding 1µl of DpnI to the PCR product and incubating the reaction at 37°C for one hour. During the incubation time, DpnI recognised the GATC sequence but it only cut the parental DNA plasmid because its "A" was methylated when it was originally prepared from a stain of E.coli. To transform XL1-Blue supercompetent cells, 1µl of the digested PCR product was incubated with 50µl of XL1-Blue supercompetent cells in ice for 30 minutes before exposing to a heat-pulse of 42°C for 45 seconds, followed by immediate storage on ice. The cells were then nourished with 500µl of pre-warmed SOC medium (Invitrogen) and incubation with shaking (230 rpm) at 37°C for one hour. After that, 150µl of the mixture was spread onto ampicillin (100 µg/ml)-LB agar plates (see preparation above). The plates were incubated overnight at 37°C.
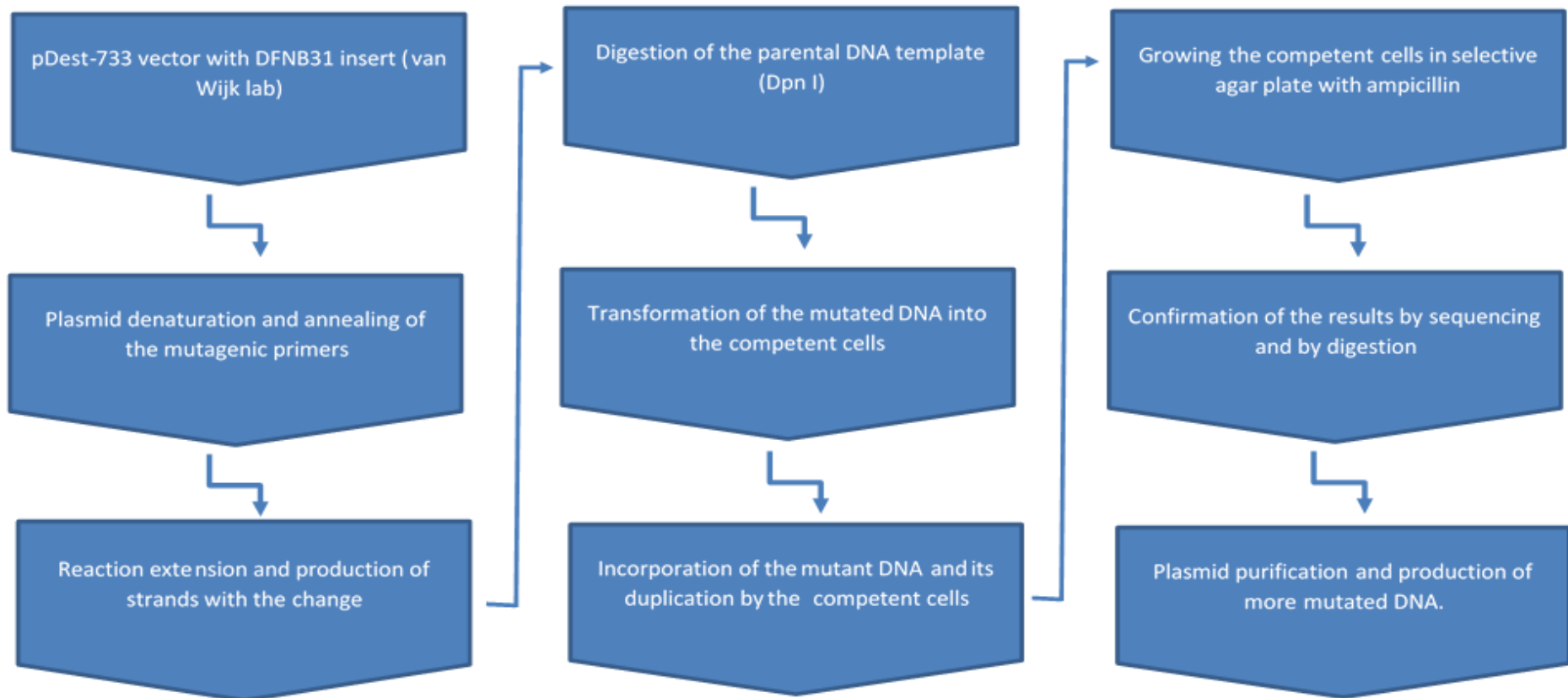
Figure 2.4. A flow diagram showing the steps involved in Site-Directed Mutagenesis assay and the purification of the produced plasmid.
The designed mutagenic primers introduced the required change and the DpnI cleaved the methylated parental DNA. The super competent cells has holes that facilitated the entry of the mutated DNA. The mutated DNA was harvested from the grown cells. Plasmid purification and production of more mutated DNA was made after the confirmation that the change of interest was successfully introduced.

2.13.4 Preparation of bacteria for plasmid preps

Aseptic techniques were followed throughout this process. Starter cultures were first prepared by inoculating a single colony from an ampicillin-LB agar plate into a tube containing 5 ml LB broth (see preparation above) with 5µl of 100 mg/ml ampicillin. The tube was then incubated at 37°C with shaking at 300rpm for about 8 hours. From the starter culture, plasmids were extracted using a Qiagen Miniprep kit, according to the manufacturer's instructions (Qiaprep MiniPrep Handbook). The protocol produced about 45µl of DNA, the concentration of which was quantified using a Nanodrop spectrophotometer.

2.13.5 Testing the prepared plasmids for the introduced changes

A standard PCR was first carried out using a pair of primers (5µM each; Appendix II) that flanked the targeted change (R450C mutation) in the *DFNB31* cDNA. Thermal cycling was performed as follows: 94°C for 3 minutes, then 34 cycles of 94°C for 20 seconds (denaturing), 55°C for 1 minute (annealing) and 72°C for 40 seconds (extension), followed by a final extension step of 72°C for 10 minutes.

2.13.5.1 Digestion

8 µl of the PCR product were mixed with 9.3 µl water, 2 µl Buffer K, 0.2 µl BSA (10 µg/µl) and 0.5 µl ScaI enzyme (Promega, USA). The mixture was incubated at 37°C for 3 hours. The mixture was incubated at 37°C for 1 hour. The digestion was subjected to electrophoresis in a 2% agarose gel at 100V for 45 minutes.

2.13.5.2 Sequencing

Six primers (Appendix 2) were designed to cover the whole sequence of the *DFNB31* cDNA insert. The DNA fragments amplified using these primers were subjected to standard Sanger sequencing, as described earlier. This step was not only to verify the target mutation but also to ensure that no other changes were introduced during the process of SDM.

2.13.6 Plasmid purification using HiSpeed Plasmid Maxi Kit

A starter culture was prepared by aseptically inoculating a single colony from an ampicillin-LB agar plate into a tube that had 5ml LB broth with 5µl of 100 mg/ml ampicillin. The tube was then incubated at 37°C with shaking at 300rpm for about 8 hours. After that, 333µl of the start culture was added to 250ml LB containing 250µl of 100mg/ml ampicillin. The mixture was incubated at 37°C with shaking overnight. Cells from the culture were then harvested by centrifugation at 6000xg for 15 minutes at 4°C. Using a Maxiprep kit (Qiagen), the pellet was first treated by adding 10ml of Buffer P1 (lysis agent), followed by 10ml of Buffer P2 (homogenising agent) and then 10ml of chilled Buffer P3 (precipitation agent). The lysate was then poured into a cartridge and incubated for 10 minutes at room temperature. Using a plunger, the lysate in the cartridge was then applied to a filter column (which had already been equilibrated by applying 10ml of Buffer QBT) and allowed to enter the resin filter by gravity flow. After that, the filter was washed with 60ml of Buffer QC before the DNA was eluted from the filter in 15ml of Buffer QF. The eluted DNA was then precipitated by adding 10.5 ml of isopropanol, followed by 5 minutes of incubation at room temperature. With the aid of a 30ml syringe and constant pressure upon the plunger, the mixture of elute and isopropanol was then filtered through a QIAprecipitator. Two ml of 70% ethanol were then applied to the QIAprecipitator via a syringe. After that, the syringe was twice used to push air quickly and forcefully through the QIAprecipitator membrane, until the membrane was dry. At this step, the DNA was trapped in the QIAprecipitator and ready for elution. One ml of TE buffer was then applied from 5ml syringe to the QIAprecipitator, and the elute was collected in a 1.5ml collection tube. In order to ensure that the maximum amount of DNA was solubilized and recovered, a re-elution step was done by taking the elute and pushing it through the QIAprecipitator again. This Maxiprep process usually produced 1ml of purified DNA with a concentration ranging from 500-1000µg/ul.

2.14 Immunofluorescence on fixed cells

2.14.1 Cell line maintenance

All the cell line culture was performed in a NuAire Labgard 437 ES Class II Biosafety Cabinet using the standard aseptic technique. When required, cells were removed from a -80°C freezer, thawed by immediate swirling in a 37°C water bath, and then rapidly transferred into a tube containing 19ml of Dulbecco's Modified Eagle Medium (DMEM), from Lonza, supplemented with 10% Fetal Bovine Serum (Sigma-Aldrich). The cells were then collected by centrifugation at 3000rpm for 3 minutes, decanting the media, and resuspension in 1 ml DMEM. Cells were grown in T75 flasks (Corning) containing DMEM in a Sanyo MCO 20AIC cell culture incubator set at 37°C and 5% CO_2. Cells were passaged after their confluency by dissociating them from the surface of the flask with 10 ml of 0.25% trypsin (Lonza) and 2-10 minutes incubation at 37°C and 5% CO_2. Sufficient banging into the flask for about 1-2 minutes ensured the maximum removal of the cells attached to the inner flask wall. Using a pipette, the flask wall was flushed thoroughly using the same suspension of cells in the flask and all the cells were transferred into a new Falcon tube containing 15ml of DMEM. The tube containing the 25 ml cell suspension was inverted 3 times to achieve proper mixing of cells. Subsequently, 3 minutes centrifugation at 3000rpm was given to the tube and the media was discarded, leaving the cell pellet, which was resuspended in 1ml of DMEM. Then, the cells were vigorously mixed by pipetting up and down about 20 times before adding 9 ml of fresh DMEM. So, the cells at this stage were suspended in a total of 10ml of DMEM.

2.14.2 Cell counting

For accuracy and consistency, the resuspended cells were counted before their seeding into new flasks. The count was carried out by mixing 50μl of trypan blue stain (Thermo Fisher Scientific) with 50μl of the cell suspension, and then loading into a clean haemocytometer. The haemocytometer was placed under an Olympus CKX41 bright field microscope and cells were counted. The resuspended cells were then diluted based on the desired seeding density.

2.14.3 Transfection

Transient DNA transfection was done to assess the over-expression of plasmids in cultured SHSY5Y cells. Plasmid transfection was carried out using an Effectene Transfection Kit (Qiagen), according to the manufacturer's instruction. The cells were first plated overnight in 24-well plates containing sterile coverslips (Fisher Scientific) in each of the wells. The next day, transfection was done when the growth was about 60-80% confluency. The cells were transfected while growing at 37°C with 5% CO_2 for 16 hours at 0.2µg concentration using 5 µl of Effectene Transfection Reagent and 1.6µl of Enhancer, all diluted in 400µl of DMEM.

After 16 hours of incubation, the media (with the transfection complex) was removed and the cells were immediately resuspended in 500µl of fresh warm (37 oC) DMEM media to prevent cell toxicity.

2.14.4 Cell fixation

The cells were washed three times with 500µl of warm (37 ºC) Dulbecco's phosphate-buffered saline (DPBS, from Sigma) on a rocking platform (Roto-Shake Genie, Qorpak, USA) for 1 minutes at RT. After that, they were fixed with 500µl of 4% paraformaldehyde (Alfa Aesar, UK) for 10 minutes on the Roto-Shake machine. The paraformaldehyde was immediately removed, and 1ml of DPBS containing 50mM NH_4Cl (1.5g in 500ml) was added to the cells, which were then rocked for 1 minute to quench excessive fixative and prevent any subsequent non-specific cross-linking later. The cells were then washed three times with 500µl of DPBS (Lonza) and 5 minutes of rocking.

2.14.5 Cell permeabilisation

Cell permeabilisation was carried out using 0.2% triton X-100 (Sigma-Aldrich) with rocking for 4 minutes, then rinsing the cells twice with 500µl DPBS. The cells were then blocked in 500µl DPBS containing 3% donkey serum (Sigma-Aldrich) for 2 hours with rocking at RT. After that, the cellular coated coverslips were carefully transferred into a new 24-wells plate.

2.14.6 Immunostaining

To each well, 250μl of the appropriate primary antibody, diluted in blocking solution containing 3% donkey serum, was added. Goat anti-whirlin antibody (Santa Cruz, sc-49787) was used at a 1:100 dilution and rabbit anti-UBR4 antibody (Abcam, ab86738) was used at a dilution of 1:250. The plate was transferred into a moist chamber and incubated overnight at 4°C on a rocking platform. The next day, the cells were washed three times with 500μl of 0.1% tween solution (Bio-Rad) for 5 minutes on a rocking platform at RT. The cells were then incubated for 1 hour with 250μl of fluorescently conjugated secondary antibodies in blocking buffer with rocking at RT. Both secondary antibodies (Anti-Goat 488 and Anti-Rabbit 488; Invitrogen) were used at a dilution of 1:500 dilution. The cells were then washed in 500μl DPBS for 10 minutes with rocking before their nuclei were stained with 500μl of 1:1000 diluted DAPI (Sigma-Aldrich). The cells were then given two final washes with 500μl DPBS for 10 minutes with rocking at RT.

2.14.7 Slide mounting

The correct number of microscopic slides was labelled and about 50μl of Fluoromount-G® mounting medium (Southern Biotechnology) were added at the centre of each slide. Then, the correct coverslips were carefully taken from the 24-well plate and mounted upside down on the mounting medium. Any air pebbles present between the slide and the coverslip were driven out by gently pressing on the coverslip. The slides were then kept for at least 12 hours at 4°C before they were taken for confocal microscopy.

2.14.8 Confocal microscopy

The slides were analysed and imaged using both Zeiss LSM510 and Nickon Eclipse TE2000-E microscopes. Each of them has its own associated imaging software. In general, confocal imaging involved using 408nM, 488nM and 543nM lasers for exciting DAPI, Alexa Fluor® 488 and Alexa Fluor® 563 fluorophores, respectively. Multiple photographs were taken for each view of interest using different cellular slices, and co-localization was assessed on 30 cells using the

ImageJ software open source image processing package (http://fiji.sc/). The confocal microscopy and Fiji ImageJ assessment was also independently done by Dr James Dachtler, faculty of biological sciences, university of Leeds.

2.15 Co-immunoprecipitation

2.15.1 Brain tissues homogenising

For minimal disturbance of hypothesised interaction between whirlin and UBR4, ultrasonication was selected as the method for homogenising right brain hemispheres collected from three different wild type adult mice. The tissues were first weighed and then Protease Inhibitor Cocktail (cOmplete^{TM}, EDTA-free; Roche) was added to each tissue based on its weight, 5ml of inhibitor per 1g of tissue. Each sample was then placed under a sonicator tip for 40 seconds using 7 cycles and 40% power. After that, the homogenates were transferred to new Eppendorf tubes on ice before they were centrifuged in a pre-chilled microfuge at 12,000 rpm for 20 minutes at 4°C. After centrifugation, the supernatants were carefully transferred to new tubes and the deposits were discarded along with their tubes. The samples were then stored at -80°C.

2.15.2 Protein quantification

The protein level in each of the samples was measured using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific), following the manufacturer's instructions. At the beginning, five albumin standards were prepared from 2 mg/ml stock. In order to achieve accurate measurement, the assay was done in duplicate for each of the standards and the samples. The prepared standards were 0, 2, 4, 8 and 10mg/ml diluted in ddH$_2$O. The samples were diluted 1:10. After that, the required amount of BCA working reagent was calculated considering that each sample or standard needed 190µl of working reagent containing 50 parts of solution A (sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1M sodium hydroxide) and 1 part of solution B (4% cupric sulphate). The reaction was set up in a microplate (Thermo Fisher Scientific) in which each 10µl of the samples or standards were mixed with

190µl of the prepared BCA working reagent, and the micropate was incubated at 37°C for 30 minutes. The plate was then loaded into an Ultrospec 3000 optic reader (Pharmacia Biotech) and the optical density measured at 750nm. The readings obtained from the different standards were plotted against their protein concentration in order to generate a standard curve that was used to determine the protein concentration of each sample.

2.15.3 Immunoprecipitation

Antibodies were diluted in PBS at a total volume 450 µl. The experiment was designed to include whirlin and blocked whirlin, plus Munc18 as a control antibody. All the antibodies were diluted 1:200 in tubes, except that the blocked whirlin tube also contained a specific peptide (Santa Cruz, sc-49787 P) to block the whirlin antibody. This peptide is the same against which the antibody had originally been raised, and the amount of whirlin peptide was five times the antibody amount in the tube, based on the manufacturer's instructions. The tubes were incubated with rotation overnight at 4°C, with the expectation that the blocking peptide would be effective. In order to attach the prepared antibodies to beads, 50µl of PBS-washed magnetic beads (Novex; Life Technologies) were added to each tube and the mixture (500µl) was incubated with rotation at RT for 2 hours. This preparation was called antibody-bound beads. In the middle of the incubation time, 350µg of protein from each sample was prepared in a total volume of 450µl of PBS and mixed with 50 µl of PBS-washed magnetic beads before being incubated with rotation at RT for 1 hour. The tubes were then placed on a magnetic stand for about 2 minutes until clear supernatant formed. The supernatants were saved and the beads were discarded. This step was done to remove any unwanted materials, leaving clean protein samples known as pre-cleared protein. After that, the antibody-bound beads (500µl) were mixed with the pre-cleared protein (500µl) and kept rotating overnight at 4°C in order for the antibody-bound beads to bind to the protein of interest. The next day, the tubes were placed on the magnetic stand for about 2 minutes before the flow through/supernatant was removed from each sample and saved. The beads were then washed by mixing 500µl of PBST, putting the tubes in the magnetic stand and then removing the supernatant. The beads were then resuspended in 30µl

of 4X loading buffer, boiled at 65°C for 5 minutes in order to dissociate from the proteins, and then placed on ice. Subsequently, the beads were placed in ice and 30µl of eluted supernatant was transferred from each sample to a new Eppendorf tube ready for loading on a gel.

2.16 Western blotting

2.16.1 Samples preparation

The mouse brain lysates were first prepared for Western blotting by making 15µl dilutions containing 60µg of protein from each sample. After that, 5µl of Laemmli buffer with 5% β-mercaptoethanol (see appendix 7 for recipe) was added to each sample in Eppendorf tubes. A cap holder was fixed over the lid of each tube before the samples were boiled in a heating block at 90°C for 5 minutes, then placed on ice. A quick spin was then given to the tubes to bring down any sample that might have adhered to the tube walls during boiling.

2.16.2 Polyacrylamide gel electrophoresis (PAGE)

The samples were run on pre-cast gels of different polyacrylamide concentrations (Bio-Rad) depending on the molecular weight of the protein of interest. Ten percent polyacrylamide gels were used to detect whirlin and 5% gels were used to detect UBR4. The gels were carefully removed from their packaging and the plastic well separator was then removed before inserting the gel into a holder and locked. After that, the gel was placed in a tank and 200ml running buffer (see appendix 7 for recipe) was then added. The wells in the gel were flushed by pipetting running buffer up and down with a Pasteur pipette before loading the samples (20µl each) and 8µl of protein size maker (Hi-Mark pre-stained HMW protein standard, Invitrogen). The gel was then run at 80V for between 60 and 180 minutes using a Power Pac 3000 (Bio-Rad). Meanwhile, 1000ml of transfer buffer (see appendix 7 for recipe) was prepared and stored at 4°C for the next step.

2.16.3 Gel blotting

Before the electrophoresis was stopped, sheets of blotting paper (BioRad) and PVDF membrane (Amersham Biosciences) were cut to the size of the gel. The blotting paper was immersed in transfer buffer and the PVDF membrane was immersed in 100% methanol for 1 minute. Once the electrophoresis had run for the specified time, the gel was removed from its frame and placed over the PVDF membrane. Both the gel and the PVDF membrane were supported by sheets of blotting paper and sponges in the form of a sandwich, which was then fixed and locked in a cassette. The cassette was placed in the tank which was filled transfer buffer and ice blocks. The gel was allowed to transfer at 20V with agitation overnight at 4°C.

2.16.4 Immunostaining

The next day, the membrane was removed and placed in PBST and the correct amount of blocking solution was prepared. Blocking solution consisted of 5% milk (Marvel skimmed milk powder) for the anti-whirlin antibody and 2% BSA (Sigma) for the anti-UBR4 antibody. The membranes were blocked for 1 hour with shaking at RT. After that, the membranes were briefly washed in PBST and then the antibody diluted as 1:200 in the correct amount of blocking solution was added to the membrane and incubated rotating for one hour at RT. Subsequently, the membrane was washed three times in PBST with shaking for 10 minutes. HRP-conjugated secondary antibody specific to the primary antibody, diluted 1 in 5000 of the correct blocking solution, was incubated with the membrane for 1 hour at RT. Again, the membrane was washed three times in PBST with shaking for 10 minutes before it was ready for imaging.

2.16.5 Imaging

Amersham ECL (enhanced chemiluminescent) Prime Western Blotting Detection Reagent (GE Health Care) was used for chemiluminescent detection of proteins. The membrane was covered with equal parts of solutions A (luminol) and B (peroxide) from the kit and left for 1 minute at RT. The membrane was then

removed and excess ECL reagent was drained from the membrane. The membrane was the wrapped with cling film and exposed to UV light.

2.16.6 Membranes stripping

To determine whether samples with no bands in the western blotting did actually contain proteins, membranes were stripped of antibodies using a strong antibody stripping solution (EMD Millipore) before re-probing with a different antibody. This method caused minimal disturbance to the proteins immobilized on the membrane. The membranes were first incubated in 20ml of 1:10 diluted stripping buffer for 30 minutes with rotation at room temperature, then they were washed three times in 20 ml PBST for 5 minutes per wash. The membrane was blocked by incubating it in 20 ml of 5% milk for 30 minutes with rocking at RT. The membranes were incubated with 1:500 diluted mouse anti-β-Actin antibody (Sigma, A5441) in 5% milk for one hour. After that, the membranes were washed three times in 20 ml PBST for 5 minutes before they were incubated in 1:5000 diluted rabbit anti-mouse antibody (Sigma, A9044) in 5% milk for 1 hour at RT. The membranes were then washed three times in 20ml PBST for 5 minutes in preparation for imaging.

2.17 Immunohistochemistry

Immunohistochemistry was carried out on paraffin embedded sagittal sections of adult wild type mouse brain fixed in 4% PFA (obtained from Dr Christine Diggle, University of Leeds). The process used the Novolink Polymer Detection System (Leica Biosystems) and was carried out at RT unless stated otherwise. To melt the paraffin wax, the slides were incubated on a heat block at 70°C for 20 minutes. The sections were then rehydrated using reduced gradient alcohol of 100%, 90% and 70% ethanol, each for 5 minutes. Endogenous peroxidase activity was blocked by keeping the sections in 3-4% (v/v) hydrogen peroxide (H_2O_2) for 20 minutes. The sections were kept under running tap water for 10 minutes in order to wash off all the H_2O_2 before boiling in 1:100 unmasking solution (Vector) for 5 minutes for antigen retrieval. After that, protein blocking was done by adding 100 µl of 0.4% casein to each section and keeping it in a

moisture chamber for 20 minutes. The casein was then removed and the sections were incubated overnight at 4°C in 100µl of primary antibody diluent (Invitrogen) with the optimised concentration of antibody. Both ANKRD2 (rabbit IgG polyclonal, from Santa Cruz) and PDZD8 (rabbit IgG polyclonal, from biorbyt) antibodies were used at 1:100 dilution. The negative slides were incubated with antibody diluent, supplemented with a non-immune immunoglobulin of the same isotype (Rabbit (DA1E) IgG XP® Isotype Control, from Cell Signalling) and concentration as the primary antibodies. The moisture chamber was then kept at RT for 1 hour before sections were washed three times with TBS (Tris-buffered saline) and then incubated in post primary treatment for 30 minutes. This treatment contains the secondary antibody (Rabbit anti Mouse IgG) which aims to detect mouse antibodies and their attracted tissues. The sections were again washed three times with TBS followed by incubation with Polymer HRP for 30 minutes. DAB solution was prepared by mixing one part of DAB chromogen (3,3' – diaminobenzidine) to 19 parts of supplied DAB substrate buffer. Each section was incubated in 100 µl of DAB solution for 5 minutes. The slides were then rinsed for 1 minute under tap water before counterstaining with haematoxylin and then dehydrating the section with 100% ethanol and xylene. Each slide was mounted with the proper coverslip and using DPX mounting medium (contains a mixture of distyrene, plasticiser, and xylene). The sections were then left to dry at RT for at least 3 hours before being examined under Nikon Eclips TE-2000-E light microscope.

2.18 *Drosophila* behaviour experiments

2.18.1 Transgenic fly lines

CG10632 is the *Drosophila melanogaster* orthologue of the human gene *PDZD8*. A CG10632 RNAi *Drosophila* was received from the Vienna *Drosophila* Source Centre (VDSC). They are transgenic white eyed fly lines expressing RNA interferences (RNAi) against CG10632 which is located in chromosome X. VDSC uses the GAL4 system to produce *Drosophila* lines for gene knockdown. The idea of using GAL4 system in the clinical practice goes back to 1987, when O'Kane

and Gehring described a way of regulating the expression of LacZ gene (encodes B-galactosidase) in *Drosophila* genome through its fusion with the P-element promoter (O'kane and Gehring, 1987). This success was later supported by Brand and Perrimon who established the GAL4 system and explained its efficient use as a tool to activate specific target gene in wide ranges of tissues (Brand and Perrimon, 1993). Details on how GAL4 system works are given in Figure 2.5. The system utilizes yeast transcriptional factor GAL4, which was found to maintain its potential to activate the transcription process in other species including human cell lines, plant, zebrafish and *Drosophila* (Caygill and Brand, 2016). GAL4 has a direct interaction SAGA protein complex through Tra1 subunit, allowing it to recruit the Mediator (Jeong et al., 2001) and multiple components of the transcription machinery to generate the transcription process (Lin Ling et al., 2012a). The principle of using GAL4 in the process of knocking down the expression of a gene of interest is explained further in Figure 2.5.

In the experiment described here, a GAL4-actin driver line, donated by Tracey Chapman (University of East Anglia), was used to control CG10632 expression. This GAL4 driver was balancer with curly wing (cyo) to stop homologous recombination in the stock. When crossing to CG10632 RNAi, offspring with curly wings were discarded and only straight-wing males were taken forward into experiments. This is because straight-wing males, and not curly wings male, carry both copies of Actin Gal-4 and CG10632 RNAi in their genome.
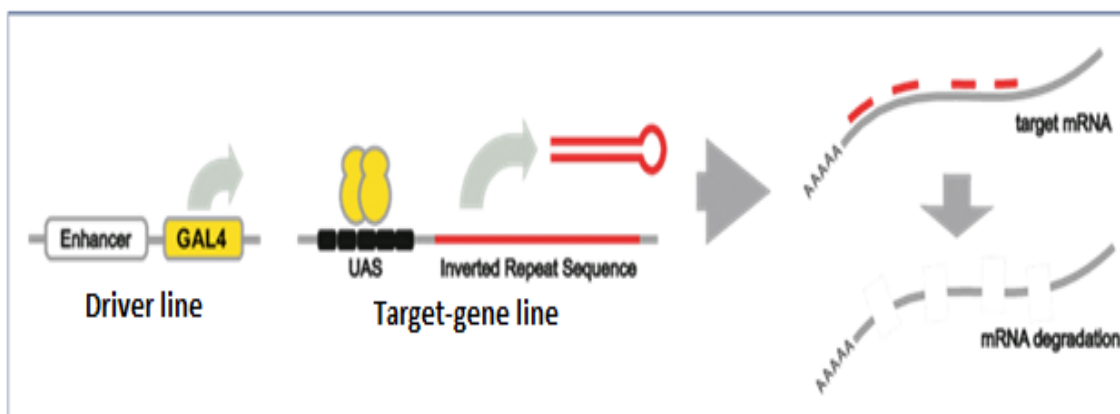
Figure 2.5. Principle of GAL4 to supress gene expression. A yeast transcriptional factor GAL4 is incorporated in a driver line which also has the enhancer of the target gene. The target gene line is generated to contain the upstream activating sequences (UAS) of the GAL4 as well as an inverted repeat sequence of the target gene. The process of knocking the target gene down would not be initiated until both lines comes in contact and the GAL4 get activated by its UAS. The activation of GAL4 system causes the inverted repeats to fold and create a long RNA of the target gene. After that, the produced RNA will be processed by the cellular machinery to make multiple functional siRNAs that would act to repress the target gene. (Adapted from Caygill and Brand (2016) with a permission from Springer and Copyright Clearance Center, License Number: 4105940955052).

The use of chromosome balancers, like Cyo, *FM7h* and *TM3,* is an important toolbox in the field of *Drosophila* genetics due to their capabilities that allow maintaining lethal mutations, avoiding separation of multiple alleles during mitotic recombination and visibility of an effect in multiple generations (Sun et al., 2012). Cyo is a dominant chromosome balancer for the second chromosome of *Drosophila* and once activated, it makes the wings of the flies to be curly or twisted (Lattao et al., 2011; Pina and Pignoni, 2012).

White eye flies, obtained from Elizabeth Duncan (University of Leeds), were used as a control because they have the same genetic background of the RNAi line and would be expected to control the behaviour of physiological differences. In the experiment, white eye and RNAi males were crossed to virgin actin driver females. The crosses were done by knocking the flies down with CO_2, as described by Artiss and Hughes (2007), and putting 5 males with 5 females in the prepared vials after sprinkling some yeast on top of the food (see 2.4.2). The flies were kept to mate and lay eggs at 25°C. After 2-3 days, they were moved into new vials with food and yeast for further egg laying. Larvae hatch from the eggs after about 6 days and then develop into the adult stage. The newly bred *Drosophila* flies were collected under cold anaesthesia (ice), as described by Ratterman (2003). As the CG10632 knockdown occurs at chromosome X, males were selected for the behavioural assays (negative geotaxis, courtship suppression, T-maze). All males were between 4 and 8 days old when used in experiments. Females were only used for the first experiment to compare the relative expression of CG10632 expression.

2.18.2 Food preparation

A standard sugar-yeast agar (SYA) medium was prepared by adding 20g of agar to 1 litre of water, which was then heated in a steamer at 100°C for 90 minutes. Once the agar was dissolved, a smooth mixture of 100g of brewer's yeast and 100 g of sugar was added to the mixed agar. The mixture was stirred until it became homogenous and then allowed to cool to less than 70°C. Thirty ml of nipagin (10% w/v solution from Sigma) and 3ml of propionic acid were then added and mixed thoroughly. Nipagin is a methylparaben which acts as an anti-fungal agent and it also slows the *Drosophila* growth rate in the larval and pupal stages. With the food mixture at about 50°C, a dispenser was used to pour 7ml of the prepared food into 25 x 95mm polystyrene vials. The food was allowed to cool at room temperature for 2 hours before the vials were capped with cotton balls and stored at 4°C for daily usage.

2.18.3 Checking RNA knockdown through qPCR

2.18.3.1 RNA extraction and cDNA Synthesis

To begin with, 5 flies from wild type or RNAi crosses where put into an Eppendorf tube. A hole was made in the lid of each tube before placing it into liquid nitrogen, then transferring it to wet ice. 600µl of QIAzol Lysis Reagent (Qiagen) were added to each tube and flies were homogenised using a micropestle. After that, the tubes were centrifuged for 3 minutes at 12,000rpm and 500µl of the supernatant was transferred into new RNase-free Eppendorf tubes. An equal volume (500µl) of ethanol was then added to each tube and the solution mixed by vortexing. RNA extraction was then carried out following the same steps as described in section 2.2.2 above. All the extracted RNA samples were quantified using a NanoDrop. In order to check the integrity of the RNA samples, they were run on a 1% agarose gel and examined for the appearance of the 18s RNA band. cDNA was synthesized from the RNA using the First Strand cDNA Synthesis method described in section 2.2.3.

2.18.3.2 Relative CG10632 expression

In order to determine the relative expression of CG10632 in the flies, quantitative PCR (qPCR) was carried out using forward and reverse primers (Appendix 2). cDNA was first diluted 1:40 (3μl of cDNA plus 117μl water). A primer mix was prepared with the forward and reverse primers at a concentration of 6 μM each. In addition to the CG10632 primers, an *eEF-1a* (Eukaryotic elongation factor 1-alpha) primer set (Appendix 2) was also used to normalise the expression. Stock cDNA was serially diluted at 1:10, 1:100, 1:1000 and 1:10000. The qPCR reaction was set by making a master mix for CG10632 and for *eEF-1a*. Each sample was tested in triplicate and each reaction would contain 10μl of SYBR Green (Sigma), 4μl of water and 1μl of primer mix. To each well of a 96-well plate, 15μl of master mix and 5μl of the 1:40 diluted cDNA were then added. The plate was then sealed with an optical clear seal and briefly centrifuged followed by 30 second shake at the plate shaker. After that, the plate was loaded into the qPCR thermocycler. The results were analysed using Bio-Rad CFX manger software to determine the qPCR efficiency and the values were plotted using Microsoft Excel.

2.18.4 Rapid iterative negative geotaxis (RING)

Flies were collected in tubes in groups of 10 under short CO_2 anaesthesia (about 1-2 minutes) and then left to recover overnight. The next day, the rapid iterative negative geotaxis (RING) assay was carried as described by Gargano et al. (2005). In brief, the flies were transferred into assay vials (Figure 2.6A) and allowed 1 minute to rest before the tubes were rapidly tapped against the bench three times to stimulate geotaxis. A digital camera (Canon, EOS 1200D) affixed to a tripod 45cm away from the tubes (Figure 2.6B) was used to capture images of the flies at 3 seconds from the last tap when they starting to travel into the top of the tube (known as negative geotaxis). The distances travelled by the flies in a single vial (n=10) were averaged across to generate a single datum. They were 250 RNAi flies and 250 wild type flies tested for RING assay. The data were copied into Microsoft Excel for statistical analysis and graph plotting.
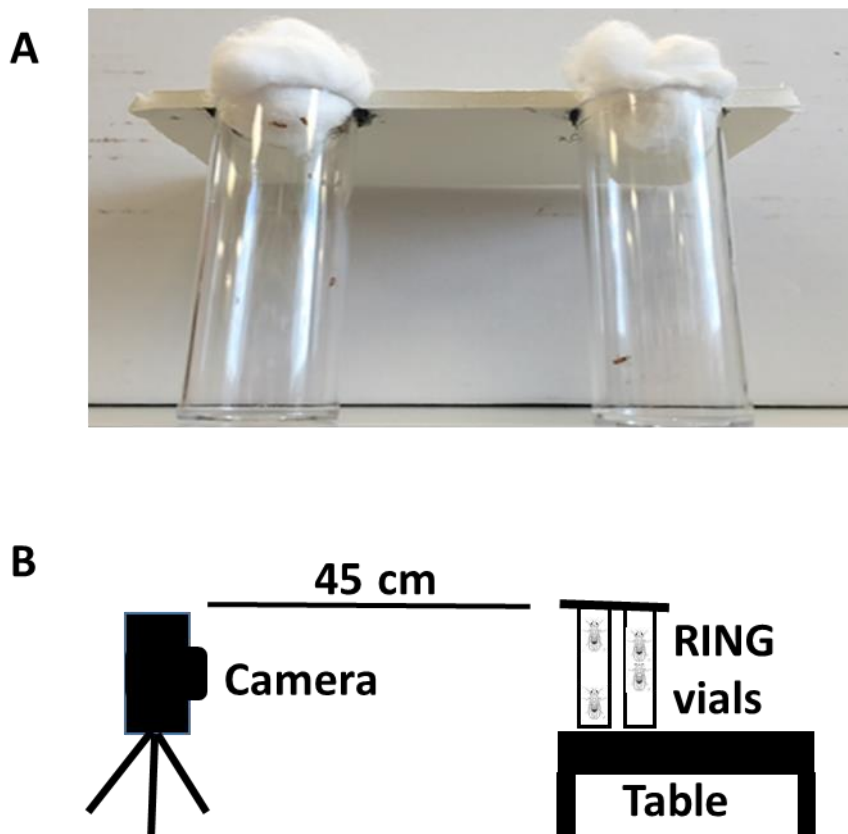
Figure 2.6. The RING assay A. Picture of 2 vials attached to a wooden sheet. Notice that half of the vial's mouth was kept free to put and remove the files. B. A schematic diagram about the set-up and apparatus involved in the RING assay. The taken images were then analysed in order to measure the travelled distance.

2.18.5 Courtship suppression assay

In this assay, female flies were allowed to be mated the day before the experiment. One wild type male and one RNAi male were then put in separate chambers ($0.3cm^3$), each with a mated female separated by a barrier. The males were left for 5 minutes before the chamber barrier was removed and the flies were allowed to interact for 1 hour (Figure 2.7A). The fly interaction was videoed during the first and last 10 minutes (Figure 2.7B). Then, the mated females from both chambers were replaced by virgin ones and the interaction was recorded for 10 minutes. As a control, two other chambers were run in parallel, each containing one wild type or RNAi male, which were left for one hour before a virgin female was put into each chamber and recorded for 10 minutes. The captured videos (n=18) were analysed using JWatcher (http://www.jwatcher.ucla.edu/), a powerful open source tool for the quantitative analysis of behaviour. The analysis

was carried out to measure the courting index base on the the time that males spent out of the 10 minutes.
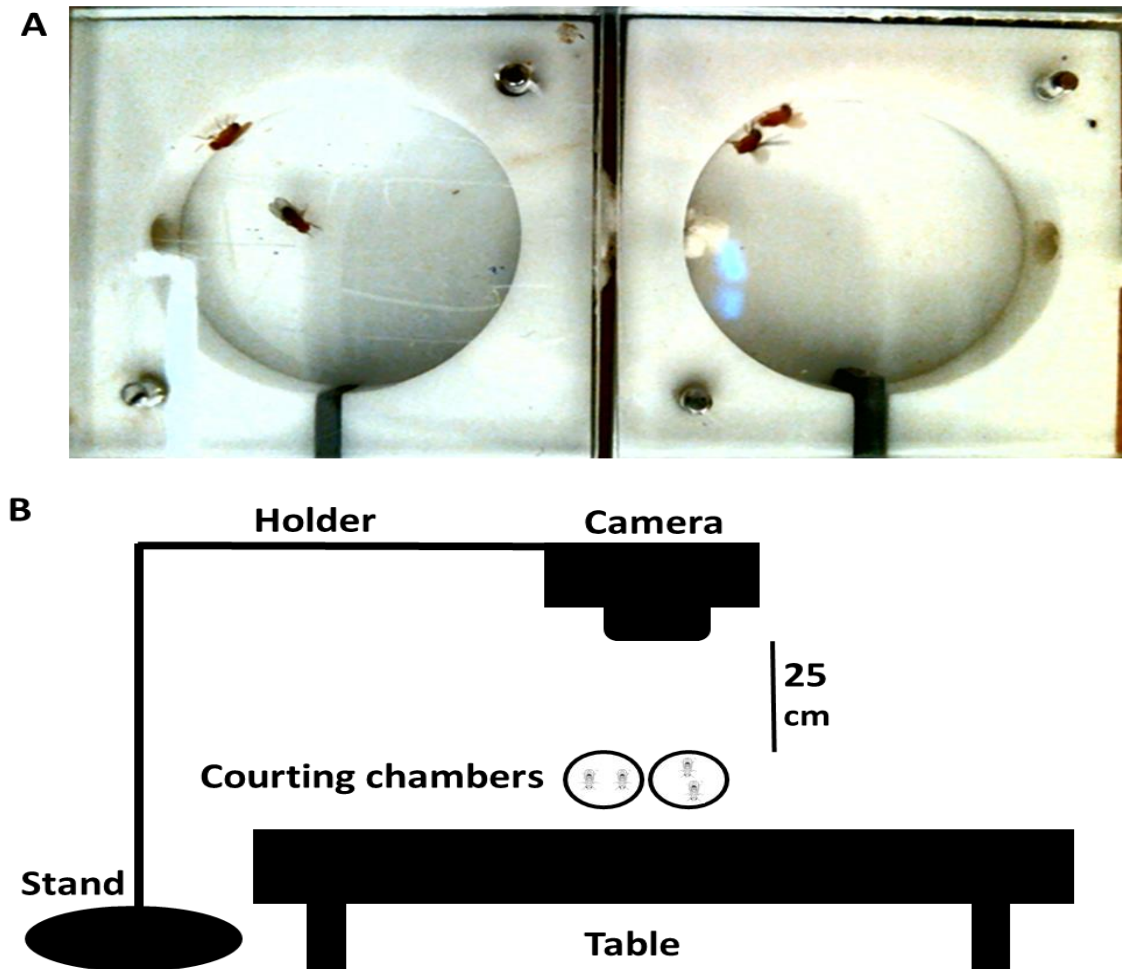
Figure 2.7. The courtship suppression assay. A. Picture of two courting chambers each containing a male and a female. The left chamber showing a non-courting male whereas the right chamber indicates a courting male extending and vibrating its wings towards the female. B. Schematic diagram about the set-up and apparatus involved the courtship suppression assay. The captured videos were then analysed to compare the courting time.

2.18.6 T-maze associative learning

2.18.6.1 Olfactory Memory Test

Olfactory memory is the ability of the brain to link an olfactory cue to an associated event and evoke episodic recollections accordingly. The study of olfactory memory is one of the used assessment in exploring the cognition and

perception functions (Zucco et al., 2012). Our brains process and analyse the different odours in the olfactory bulb which is thoroughly connected with the amygdala and hippocampus, which are the core regions of memory processing in the brain. In *Drosophila*, the olfactory system is equipped with about 1300 sensory neurons located in antenna as well as in the maxillary palp (Larsson et al., 2004).

2.18.6.2 Optimization of odour concentration

The T-maze assay started with an optimization step to select the optimum concentration of the two odours used: 3-octanol (OCT) and 4-methylcyclohexanol (MCH) (Sigma-Aldrich). The odours were diluted in 10 ml of mineral oil and 3 different concentrations (0.1, 0.15 and 0.3ml/l) of each odour were tested. The optimization was carried out using the male progeny of wild type crossed with actin driver. In groups of 50, the flies were given 2 minutes to choose between OCT and MCH that were delivered via air drawn through a vacuum pump connected to the T-maze apparatus. (Figure 2.8). The optimisation step involved more than 2000 flies. The optimum concentration was considered to be achieved, as described by (Murakami et al., 2010), when about half of the flies selected one odour and the other half selected the other odour.
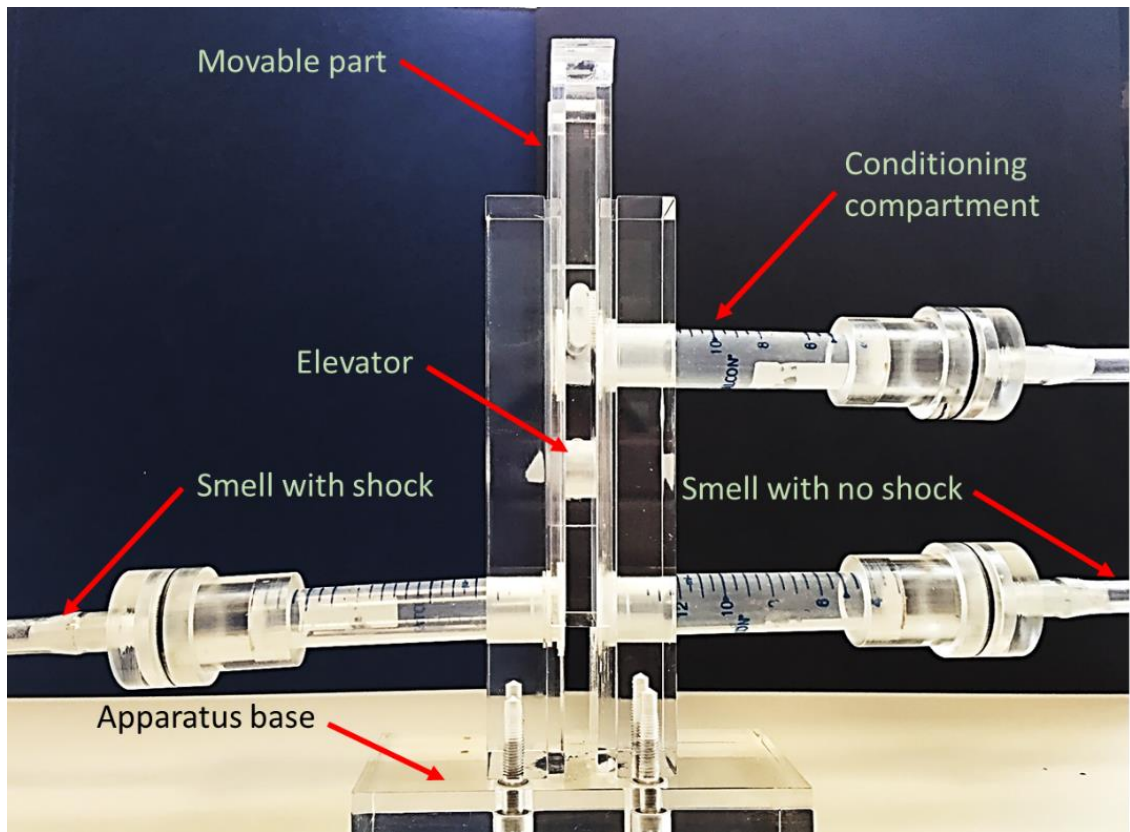
Figure 2.8. A picture of the used T-maze apparatus. The files were first put into the conditioning compartment where they get trained to remember avoiding the smell coupled with the shock . The movable part was then used to bring the elevator up to take the flies to the testing compartment at which the two smells were delivered though a pump.

2.18.6.3 Fly conditioning

All the fly vials were labelled by blind coded marking so that the genetic identity of the flies within could not be known during the testing process. The conditioning was performed as described by Mery and Kawecki (2002) and Frederic mery et al. (2007). Each group of flies was first given 30 seconds to rest and equilibrate in fresh air. After that, they were supplied for 1 minute with one of the two odours along with simultaneous exposure to a mechanical shock (hard vortexing) every 5 seconds that knocked them down and made them struggle. Then, 30 seconds of fresh air was delivered to the flies in order to allow a gap time for the other alternative smell which was also given for 1 minute but without any mechanical shock. The conditioning was exchanged between starting with and without a mechanical shock throughout the experiment. A final 30 seconds period was

allowed before the flies were moved into the testing compartment of the T-maze apparatus (Figure 2.8).

2.18.6.4 T-maze choice test

The testing compartment has a chamber in the middle with right and left connecting arms in which the two odours were supplied (Figure 2.8). As described by Pitman et al. (2009),the flies were given 2 minutes to choose between the two odours (odour paired with the mechanical shock versus the odour not paired with the mechanical shock) before the path that connects the chamber to both arms was closed. After that, the flies from the right and left arms were collected in separate vials. The test was carried without bias so that the location of the odours was exchanged between the left and right arms after testing each set of flies. The number of flies in each vial was recorded in Microsoft Excel in order to calculate the learning index and plot the graphs. The learning index (LI) was calculated by subtracting the number of flies that did not learn from the number of flies which chose the non-risky smell, and then divided by the total number of flies in the test. Statistics were performed in SPSS software version 20.

Chapter 3: Using whole exome sequencing and SNP array to identify pathogenic mutations causing intellectual disability in consanguineous Omani families

3.1 Introduction

Genetic factors are the most common cause of intellectual disability (ID). The condition is extremely heterogeneous and has, so far, been linked to more than 800 genes (Chiurazzi and Pirozzi, 2016). ID has been known to be inherited by different modes including recessive, dominant and X-linked as well as mitochondrial. Consanguineous families in which ID is segregating, like the ones described here, normally have an autosomal recessive mode of inheritance whereas sporadic cases are mostly attributed to dominant, heterozygous *de novo* changes (Musante and Ropers, 2014; Reuter et al., 2017). ID does not always develop because of a single aetiology. Many times, the condition is complex resulting from the interaction of multiple factors. Because of this complexity associated with ID development, genetic screening for it is challenging. Furthermore, a large proportion of ID cases remain unsolved after screening, without a specific and clear diagnosis.

This chapter describes the results of genetic analyses of Omani ID families ascertained by Omani collaborators (see section 2.1). DNA was extracted from peripheral lymphocytes using standard techniques (Section 2.2.1). Whole genome homozygosity mapping was undertaken using Affymetrix SNP 6.0 arrays (Section 2.2.10). The Illumina platform was used for whole exome next generation sequencing (Section 2.2.11). All annotations and physical positions were recorded according to the UCSC Genome Browser GRCh37/hg19 build. Primer pairs for PCR amplification of the regions of interest were designed using Primer3 software. Sequencing of the candidate genes was carried out using the BigDye® Terminator Kit v1.3 and an ABI 3130xl Genetic Analyser (Section 2.2.9). This chapter covers the findings of the genetic analysis of six Omani families

diagnosed with ID. The causative genes were identified in half of the families studied (Families 1, 2 and 3) but could not be defined in the other half (Families 4, 5 and 6). Table 3.1 summarizes the families studied with some details about their diagnosis, inheritance pattern and the number of cases sampled.

Family ID	Diagnosis	Inheritance pattern	Affected members	Sampled members	Sampled unaffected members
Family-1	ID & ASD	Recessive	4	3	4
Family-2	ID & SHL	Recessive / X-linked	2	2	3
Family-3	ID	Recessive	3	3	4
Family-4	ID	Recessive	5	5	2
Family-5	ID	Recessive / X-linked	2	2	2
Family-6	ID	Recessive	2	2	3

Table 3.1. Summary of the studied Omani ID families. All the families are consanguineous and of Omani Arab ethnicity. The expected mode of inheritance as well as the number of the sampled cases are detailed for each family. ID = intellectual disability, ASD = autistic spectrum disorder, SHL= Sensorineural hearing loss.

3.2 Results

3.2.1 Intellectual disability families with identified causes

3.2.1.1 Description and clinical details for Family-1

This consanguineous family originates from a first-cousin marriage (Figure 3.1). Four of the children have been diagnosed with ID as the consensus feature. Table 3.2 summarizes the major features and developmental profile of the affected members in this family. Patient V.5, a male, was reported to have obvious difficulties with reading and writing at school. He was referred to the clinic at 7-years-old for short stature (below the 3rd centile) and developmental delay of language and motor skills. On examination, the patient was described as having a myopathic elongated face with pointed chin and downturned mouth with thin lips. His eye lids (palpebral fissures) were downslanting and the space between his eyes was increased, showing apparent hypertelorism. He also had very thin nails and bilateral fusion of the second and the third toes (syndactyly). Given the similarity of the child's features to Smith–Lemli–Opitz syndrome, caused by a mutation in 7-dehydrocholesterol reductase, this enzyme was tested, but found to be within normal limits. Further investigations including chromatography of plasma amino acids, urine organic acids, lactate, ammonia and creatinine kinase were all within normal reference ranges. Magnetic resonance imaging (MRI) of the brain was normal. His behaviour included a fascination with the sight and sound of water. Family history revealed that his deceased older brother (V.I) also had similar clinical features (Figure 3.1 and Table 3.2) as well as camptodactyly of the right fingers which was confirmed by reading his hospital notes and reviewing family pictures. Patient V.I had also been assessed by the developmental team and was found to fulfil the DSM-5 diagnostic criteria for autism spectrum disorder (ASD). Patient V.I drowned in water at the age of 10 years old.

Patients V.8 and V.9, a female and male, respectively, share a similar developmental and clinical profile as patient V.5 and his deceased brother (see Table 3.2), though patient V.9 does not have any congenital malformations of the hands or toes. Patient V.8 has shown some delay of motor functions including

problems with balance and a poor hand grip and was unable to hold a pencil between the thumb and finger, resulting in her being unable to draw a circle or a straight line. She has had difficulty chewing food and swallowing, and sometimes drools when performing a focused task. Her communication comprises a few single words, vowel sounds and non-verbal gestures, including pointing and facial expressions. It is worth highlighting that the unaffected female sibling V.10 also appears to have some of the facial features described in the affected patients, so the mild dysmorphic features described in this pedigree may not be disease-specific.
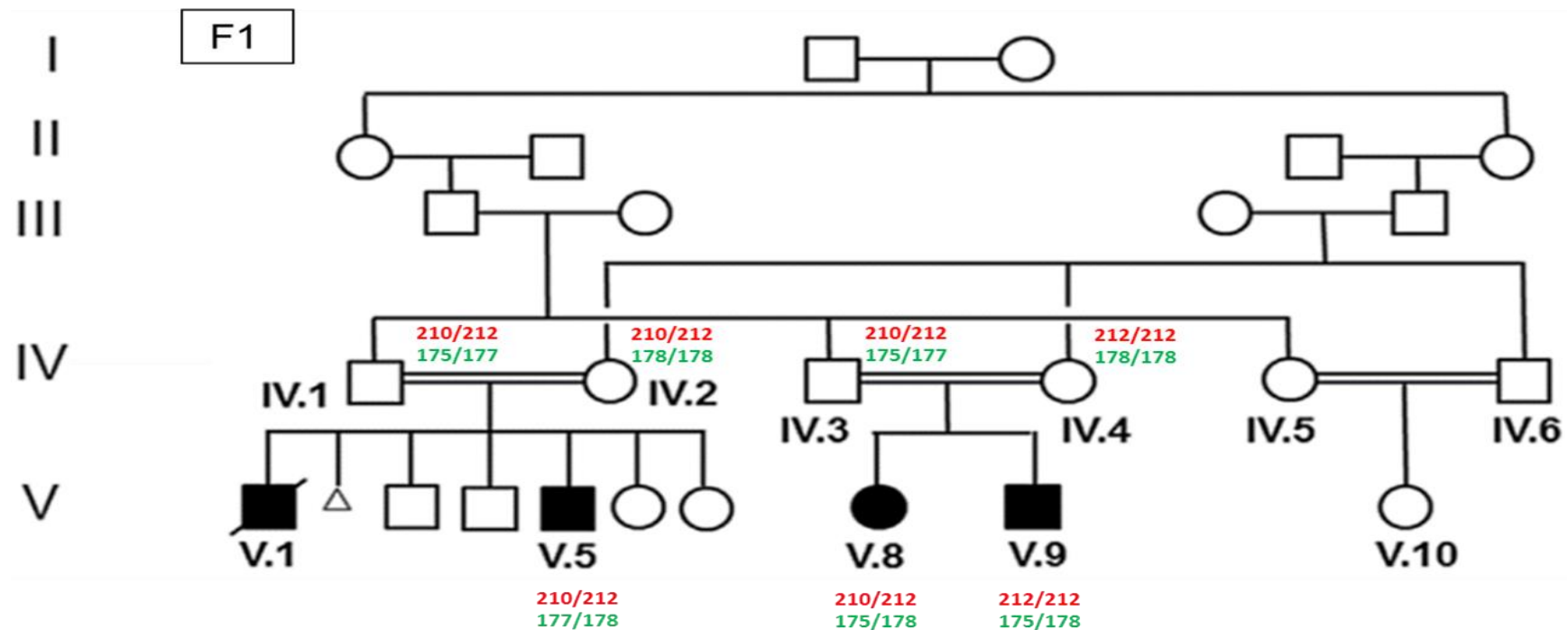
Figure 3.1. Family-1 pedigree with microsatellite genotyping results. This family has multiple consanguineous loops. The three nuclear families in the fourth (IV) generation resulted from marriages between first-cousins. There are four ID affected children: three boys (filled squares), of whom one drowned in water (V.1), and a girl (V.8; filled circle). The small triangle indicates a miscarriage. Detailed clinical information is given in Table 3.2. Genotyping is shown for the cases from whom DNA was available. Red text indicates the alleles detected using marker D4S3042, and the green text indicates the alleles detected with D7S483. SNP array analysis was carried out for case V.8 and WES for case V.9. (Adapted from Al-Amri et al. (2016) with a permission from John Wiley and Sons, and Copyright Clearance Center's RightsLink® service, License number 4055460722605).

ID (Gender):	V.1 (M)	V.5 (M)	V.8 (F)	V.9 (M)	V.10 (F)
TUSC3 genotype (status)	M/M (affected)	M/M (affected)	M/M (affected)	M/M (affected)	M/+ (unaffected)
Growth parameters (at birth)	Wt = 50th centile Ht = 10th centile HC = 25th centile	Wt = 3rd centile Ht = 10th centile HC = 3rd centile	Wt = 3rd centile Ht = $<3^{rd}$ centile HC = $<3^{rd}$ centile	Wt = 3rd centile Ht = $<3^{rd}$ centile HC = $<3^{rd}$ centile	Wt = 90th centile Ht = 50th centile HC = 50th centile
Congenital malformations	Unilateral camptodactyly - hand	Bilateral 2-3 syndactyly - toes	Bilateral 2-3 syndactyly - toes	-ve	-ve
Growth parameters (age)	Wt = 3rd centile Ht = $<3^{rd}$ centile HC = 3rd centile (9 years)	Wt = $<3^{rd}$ centile Ht = $<3^{rd}$ centile HC = 3rd centile (7 years)	Wt = 50th centile Ht = 10th centile HC = 25th centile (3 years)	Wt = 3rd centile Ht = $<3^{rd}$ centile HC = $<3^{rd}$ centile (1 year)	Wt = 3rd centile Ht = 10th centile HC = 50th centile (5 years)
Facial features	Hypertelorism Long face Deep set eyes Thin philtrum Pointed chin	Hypertelorism Long face Thin philtrum Pointed chin	Hypertelorism	Hypertelorism Thin philtrum	Hypertelorism Thin philtrum
Brain imaging	NR	Normal MRI	NR	NR	Normal MRI

ID (Gender):	V.1 (M)	V.5 (M)	V.8 (F)	V.9 (M)	V.10 (F)
IQ	NR	53	40	50	81
Autism spectrum disorder	+ve	-ve	-ve	-ve	+ve
Developmental profile	NR	Motor and language skills delayed Poor reading and writing ability	Motor and language skills delayed	Walked at 20 months Can say ~5 words with meaning at 1 year	Walked at 19 months Single words at 2.5 years Current language normal but has delayed echolalia; also has hyperlaxia

Table 3.2. Summary of clinical features in affected cases from Family-1. Growth parameters were measured using the World Health Organisation child growth standards, and IQ was measured using the Stanford Binet intelligence test (5th edition). M in genotype denotes mutation in *TUSC3*; + in genotype, wild type allele; Wt, weight; Ht, height; HC, head circumference; NR = not recorded; -ve, negative; +ve, positive. (Adapted from Al-Amri et al. (2016) with a permission from John Wiley and Sons, and Copyright Clearance Center's RightsLink® service, License number 4055460722605).

3.2.1.2 Microsatellite genotyping in Family-1

Two polymorphic microsatellite markers (D4S3042 and D7S483) were used to genotype each case to ensure its correct parentage within the pedigree before proceeding with homozygosity mapping (Section 2.10) and exome sequencing (Section 2.11). Genotyping was performed using a ROX-500 ladder and ABI3130xl capillary Genetic Analyser (Section 2.7). An example of the analysed results, generated by the Gene Mapper software is shown in Appendix 3. The results of genotyping in this family are shown in Figure 3.1. The affected boy (V.5) shows a genotype pattern consistent with him being the progeny of IV.1 and IV.2. Similarly, genotypes of the other two affected children (V.8 & V.9) are consistent with them being the progeny of IV.3 and IV.4.

3.2.1.3 Homozygosity mapping in Family-1

Homozygosity mapping in this family was performed using genomic DNA on a SNP array for subject V.8, while WES data of subject V.9 was analysed using Agile MultiIdeogram software as described in section 2.10.1. The results presented as an ideogram are shown in Figure 3.2. Homozygosity mapping indicates the shared regions inherited by descent from a common ancestor, which would be considered most likely harbour a homozygous variant as the cause of disease in a family of this background. When homozygosity mapping and whole exome sequencing are combined (Section 2.11.9.4), this allows a huge number of variants located outside the homozygous regions – and thus unlikely to be the disease-causing mutation – to be discarded.
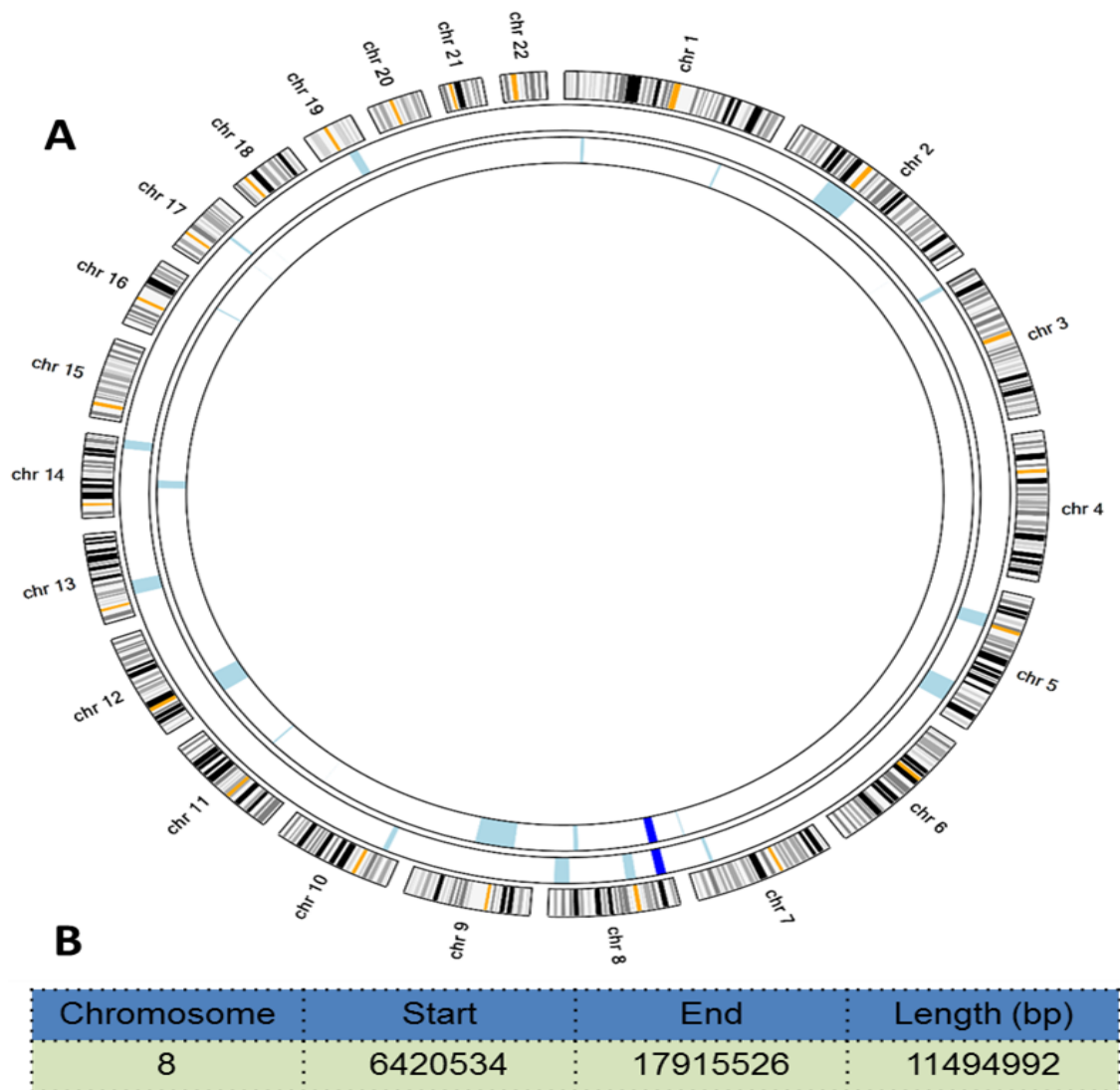
Figure 3.2. Shared homozygous region in Family-1. A. Ideogram where the outermost circle represents the 22 autosomal chromosomes. The middle circle plots homozygosity revealed in the SNP data from subject V.8, while the inner circle shows homozygous segments in WES data from subject V.9. The light blue areas are the homozygous regions within each of the individuals genomic DNA and the dark blue region on chromosome 8 appears to be the only homozygous region shared by both individuals. B. Details of the shared homozygous region. These coordinates were generated for the regions and the corresponding measurements were based on human genome reference hg19. The region measuring about 11.5 Mb in chromosome 8 is expected to contain the mutated gene.

3.2.1.4 Exome sequencing in Family-1

Whole exome next generation sequencing was carried out on subject V.9 using paired end reads on an Illumina HiSeq 2500 (Section 2.11). Before proceeding with the exome analysis, the quality score of the sequenced data was checked

using FastQC software (Section 2.11.9.1) to ensure no obvious problems or biases. An example of the quality score and charts are illustrated in Appendix 4. Table 3.3 shows the basic statistics of exome sequencing for subject V.9's DNA.

Measure	Value
File Name	V.9 of Family-1
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total reads	72326188
Sequence flagged as poor quality	0
Sequence length	101
%GC	50

Table 3.3. Basic statistics of V.9 exome sequencing reads. Simple statistical composition of the analysed sample. V.9 did not show any poor quality sequence. The appearance of only one value for the length of the reads (101) is an indication that all sequences are of the same length. The percentage of the GC content is perfect as it is within the target the GC content is about (49–51%).

After passing the quality check, exome analysis was carried out, as described in section 2.11.9.1. One of the major steps in exome analysis is the removal of PCR duplicates and reads that have not aligned to the reference sequence; this discards a huge number of unwanted reads. By the end of this analysis, V.9 had 61,905 variants that were taken forward for further selection and prioritization (Figure 3.3).

3.2.1.5 Variant filtering and prioritization

Among the 61905 variants, 238 were located in the shared 11.5 Mb homozygous region on chromosome 8 (Figure 3.2). These 238 variants were then filtered based on the likelihood of them being causative. By removing all the synonymous changes that code for the same amino acids and keeping only the exonic and splicing variants, the list of variants was reduced to 44. Because the causative mutation was expected to be rare, another filter was applied to remove all the

variants that were present in more than 1% of the population, based on 1000 Genomes database. The final variant list (Table 3.4) consisted of six variants in *SGK222*, *RP1L1* and *TUSC3*.

- Total number of sequences in the original raw data = 72326188
- Number of variants after analysing the exome data = 61905
- Variants located in the shared homozygous region = 238
- Only non-synonymous exonic variants = 44
- Variants having <1% frequency in the population (1000g) = 6

Figure 3.3. Variants in Family-1 after analysis of exome sequencing and filtering steps. This flow diagram shows the reduction in the number of variants after analysing the exome data (case V.9) , when coupling the analysis with the homozygosity mapping results (case V.8) and the filtering process until getting to the final 6 candidates . All the 238 variants are located in the shared homozygous region of chromosome 8 but the highlighted filtering steps excluded the unlikely causatives.

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	CADD Score	Freq in 1000G
8	8234868	*SGK223*	exonic	NM_001080826:exon2:c.1050insAGCGGC	p.A351delinsSGA	0.985	0
8	10466010	*RP1L1*	exonic	NM_178857:exon4:c.G5598T	p.E1866D	10.03	0
8	10467637	*RP1L1*	exonic	NM_178857:exon4:c.A3971G	p.E1324G	0.79	0
8	10467652	*RP1L1*	exonic	NM_178857:exon4:c.C3956G	p.A1319G	0.013	0
8	10467653	*RP1L1*	exonic	NM_178857:exon4:c.G3955A	p.A1319T	7.54	0
8	15480672	*TUSC3*	exonic	NM_006765:exon2:c.222delA	p.R74fs	31	0

Table 3.4. Final list of candidate variants for Family-1. This list was generated after further filtering that removed all the synonymous changes and the variants that are present in more than 2% of the population. The table illustrates six variants located in three different genes: *SGK222*, *RP1L1* and *TUSC3*. CADD suggests the most likely causative variant and thus is highlighted in light green.

3.2.1.6 Segregation of candidate variants

The six candidates were scored by CADD, a bioinformatic predictor which integrates diverse genome annotations to provide a metric of deleteriousness for the variants of interest. CADD predicts only *TUSC3* variant is significant, others are unlikely to affect the proteins encoded (Table 3.4). Furthermore, examination of the literature showed that *RP1L1* is involved in inherited eye disease, *SGK223* is not known to be involved in any human disease, but *TUSC3* mutations are implicated in ID. Attention therefore focussed on the *TUSC3* variant but other candidates in the final list were all tested for segregation.

A novel homozygous single base pair deletion of an adenine nucleotide at position 222 (Figure 3.4) in exon 2 of *TUSC3* (NM_006765:c.222delA, p.R74fs*3) was the only candidate variant that showed segregation and consistency with the ID diagnosis in Family-1. This variant followed a recessive pattern of inheritance as both parents were found to be heterozygous, all affected children were homozygous and the unaffected children were either heterozygous or wild type (Figure 3.5). This frameshift deletion is predicted to cause the removal of 273 residues from the C-terminus of the predicted protein product as the premature termination of TUSC3 protein synthesis occurs after only 76 amino acids. However, it is not possible without further experiments to determine whether a truncated protein is produced or whether the effect is to cause nonsense mediated decay of the mRNA, meaning there is effectively no protein produced.
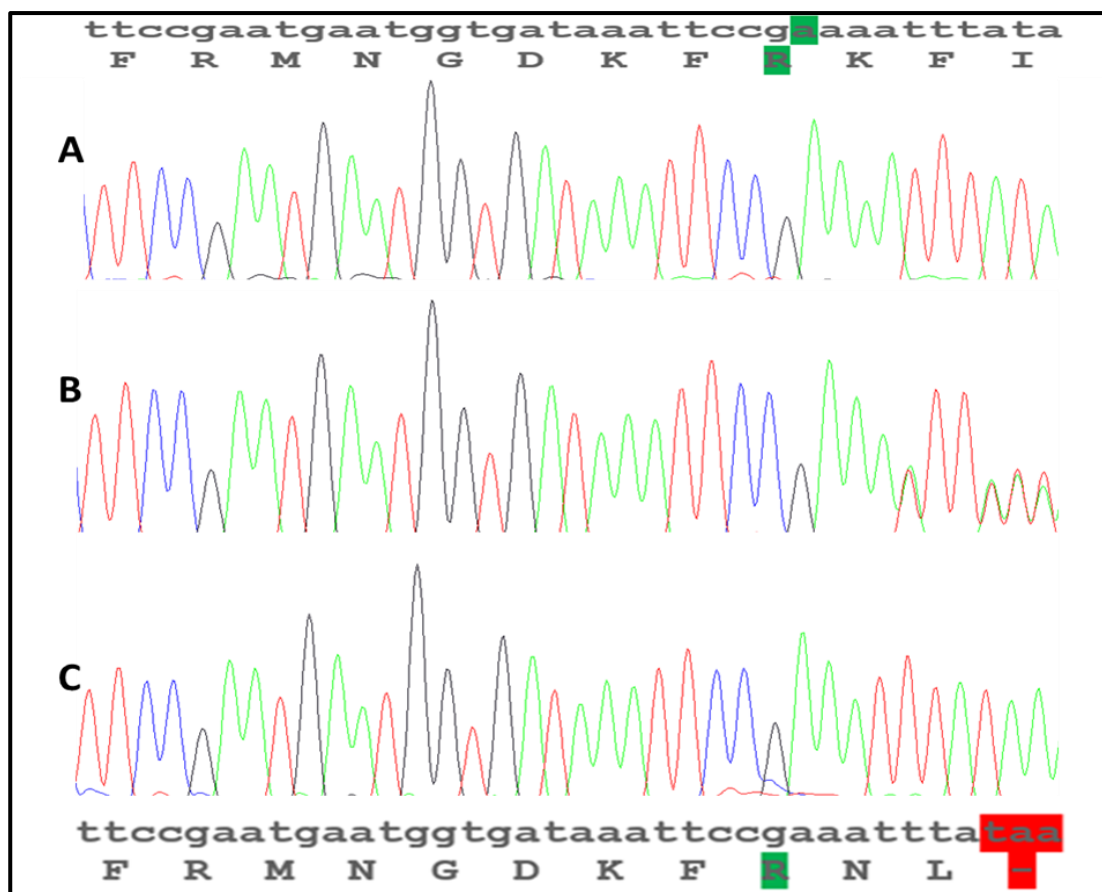
Figure 3.4. Electropherograms of the c.222delA deletion. This mutation occurs in exon 2 of *TUSC3*. The upper alignment highlights, in green, the position of the adenine (A) in the triplet encoding an arginine (R) amino acid. A. Normal wild type sequence found in unaffected controls. B. The pattern observed for an individual who is a heterozygous carrier for the mutation. The deletion in one allele has resulted in the observed overlapping sequencing peaks. C. Homozygous deletion seen in affected cases. The deletion is predicted to result in protein termination after two amino acids from the arginine but alternatively it may cause no protein due to nonsense mediated decay. The bottom alignment shows the arginine (R) highlighted in green, and the stop codon (taa, -) which is indicated by the red shading.
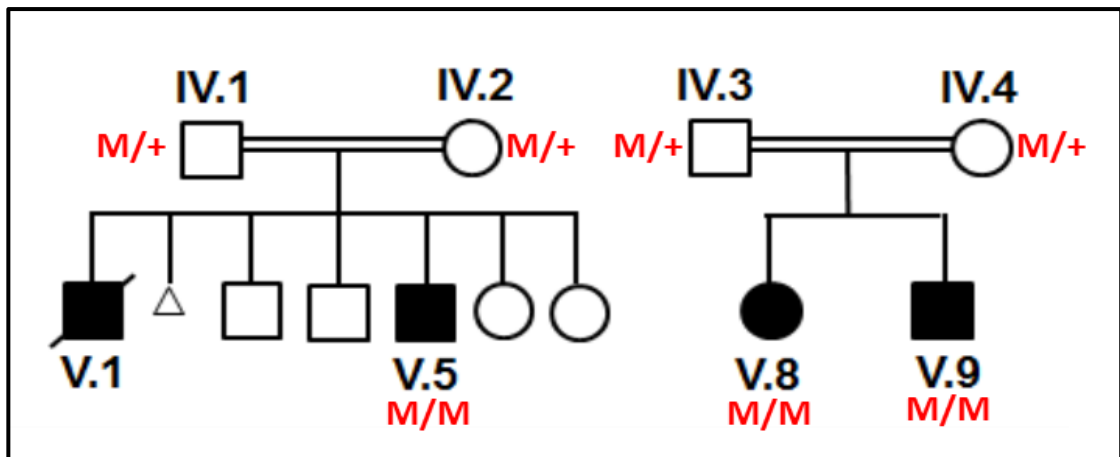
Figure 3.5.Segregation of c.222delA, p.R74fs*3 in Family-1. The nuclear family from the original extended family showing the segregation of the single base deletion mutation in *TUSC3*. + is the wild type allele and M is the deletion (c.222delA, p.R74fs*3).

3.2.1.7 Population frequency of the *TUSC3* mutation

To determine the frequency of the variant, c.222delA, p.R74fs*3, in the Omani population, Sanger sequencing of 100 chromosomes from 50 Omani healthy individuals was performed. The absence of the mutation in these subjects and in 141,353 individuals in the Genome Aggregation Database (gnomAD) suggests that the c.222delA is a very rare change, which is consistent with it being pathogenic.

3.2.1.8 Description and clinical details for Family-2

Family-2 (Figure 3.6) is also a consanguineous Omani family that originates from a first-cousin marriage. The family has two boys (IV.3 and IV.1) diagnosed with ID and sensorineural hearing loss. Prenatal, delivery and neonatal history of IV.3 was unremarkable. He was born by Caesarean section with normal birth parameters and no post-delivery complications, but was noted to have low muscle tone at birth. At 3 years of age, IV.3 was assessed at the Genetics and Developmental Medicine (GDM) clinic for global developmental delay. He did not start sitting unsupported until 12 months of age, started walking at 2 years and was thus diagnosed with fine and gross motor delay. Detailed developmental assessment revealed language delay affecting his expressive speech. He also had poor social interaction and communication and significant stereotypic behaviours and play. Autism diagnostic interview and direct clinical observation

111

revealed that IV.3 fulfilled the DSM-5 diagnostic criteria for autism spectrum disorder.

His 11-year-old brother (IV.1) had severe bilateral sensorineural hearing loss and received a cochlear implant at around 5 years of age. Assessment by the GDM clinic revealed that IV.1 also had autism spectrum disorder, confirmed by autism diagnostic interview and direct clinical observation.

Both brothers were referred to the ophthalmology team and diagnosed with hypermetropia (long-sightedness) and astigmatism (blurred or distorted vision). They also had dental anomalies and similar facial features (summarised in Table 3.5). Both parents were examined and the mother was found to have dental crowding with overlapping incisors and some conical shaped teeth.
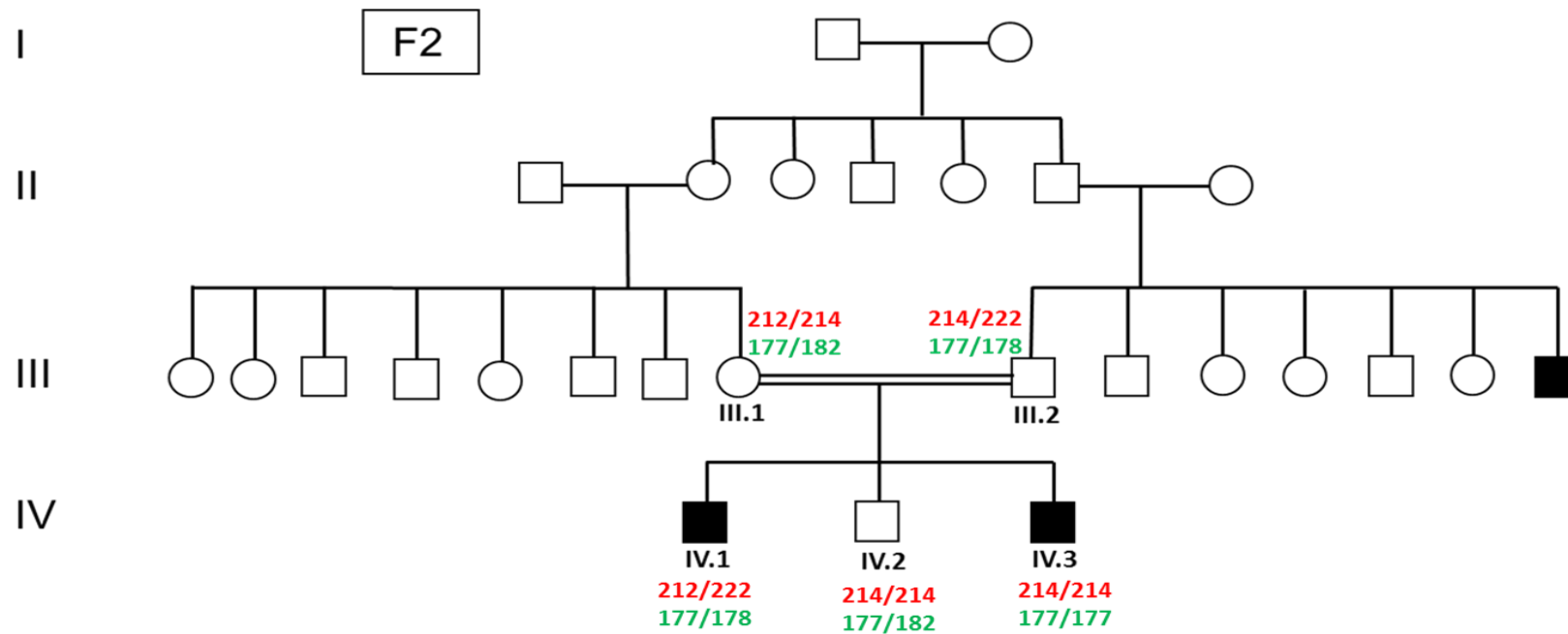
Figure 3.6. Family-2 pedigree with microsatellite genotyping. The fourth generation (IV) progeny from a consanguineous marriage between first-cousins. The incidence of ID suggests an X-linked or recesssive inhertance pattern. IV.1 and IV.3 both have ID and sensorineural hearing loss. There is also a male with ID in the third generation (reported by III.1 & III.2) but no samples or detailed clinical information were available. The red genotypes are the detected alleles when tested using marker D4S3042 and the green genotypes are for marker D7S483. DNA from case IV.1 was analysed by SNP array and case IV.3 DNA was subject to WES.

ID (gender):	IV.1 (male)	IV.3 (male)
Deafness (age at diagnosis)	+ (5 years)	+ (3 years)
Growth parameters (age)	Wt = 25-50th centile Ht = 25-50th centile HC = 50th centile (11 years)	Wt = 50th centile Ht = 50th centile HC = 50th centile (3 years)
Visual problems	Hypermetropia and astigmatism	Hypermetropia and astigmatism
Dental anomalies	Conical shaped teeth, crowding and overlapping incisors	Conical shaped teeth, crowding and overlapping incisors
Skeletal/ muscular development	Normal	Generalised hypotonia
Brain imaging	Normal CT	MRI: non-specific white matter changes in trigone and occipital lobe
IQ	68	45
Speech development	Poor speech and has echolalia	Language skills delayed
LHFPL5 **genotype**	M/M	M/M

Table 3.5. Summary of clinical features in affected members of Family-2. Wt = weight, Ht = height, HC = head circumference, + = presence and - = absence. M represents the c.575T>C variant allele in *LHFPL5*.

3.2.1.9 Microsatellite genotyping in Family-2

Genotyping markers were used on the cases from whom DNA was available to verify the pedigree structure of the family. This was considered to be an important step before proceeding with costly next generation sequencing technology. Figure 3.6 indicates the tested cases and their genotypes. All the children (IV.1,

IV.2 & IV.3) showed genotype patterns that were consistent with III.1 and III.2 being their parents.

3.2.1.10 Homozygosity mapping in Family-2

Homozygosity mapping of Family-2 was carried out using the results of SNP array analysis of subject IV.1 and SNP data extracted from WES of subject IV.3, as described in section 2.10.1. (Figure 3.6). Three shared homozygous regions could be defined (Figure 3.7), and it was assumed, based on the pedigree structure, that one of these ought to harbour the causative mutation.

Chromosome	Start	End	Length
5	146135996	154300940	8164944
6	28474218	44280812	15806594
12	125834395	133777645	7943250

Figure 3.7. Shared homozygous regions in Family-2. A. An ideogram where the outer circle contains template segments of the 22 autosomal chromosomes. The middle circle

is a plot of homozygous regions in SNP data from subject IV.1 while the inner circle shows homozygous regions from WES data from subject IV.3. The light blue areas are the homozygous regions within each individual and the dark blue regions (number=3) are the homozygous regions shared between both affected individuals. B. Details of the shared homozygous regions. The coordinates were generated based on human genome reference hg19. The three shared regions are on different chromosomes and with different lengths. The biggest region, which is on chromosome 6, is highlighted in green.

3.2.1.11 Exome sequencing in Family-2

Whole exome next generation sequencing was carried out on DNA from subject IV.3 using the method described in section 2.11. The data for this sample, generated by Illumina HiSeq 2500, passed the quality check described in section 2.11.9.1. Table 3.6 shows the basic statistics of exome sequencing of IV.3.

Measure	Value
File Name	IV.3 of Family-2
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total reads	36420257
Sequence flagged as poor quality	0
Sequence length	101
%GC	49

Table 3.6. Basic statistics of IV.3 exome sequencing reads. Simple statistical composition of the analysed sample. IV.3 did not show any poor quality sequence. The length of the reads is showing a single value (101) suggesting that all sequences are of the same length. The percentage GC content is within the target (49–51%).

After passing the quality check, exome analysis (Section 2.11.9) was carried out. The sequencing reads were aligned against the reference genome (hg19/GRCh37) using Bowtie2 aligner as described in section 2.11.9.2. PCR duplicates and reads that had not aligned to the reference sequence were removed using Picard tools. By the end of the analysis, IV.3 had 165753 variants that were taken for further selection and prioritization.

3.2.1.12 Variant filtering and prioritization

Among the 165753 variants, 2166 variants were located in the three shared homozygous regions (Figure 3.8). These 2166 variants were then filtered based on the likelihood of them being causative. By removing all the variants with a low depth of coverage depth (<10) as well as synonymous changes that code for the same amino acids, and keeping only the exonic and splicing variants, the list of variants was reduced to 74. Further filtering removed variants that had a frequency higher than 1% in the population and all heterozygous changes. The final variant list (Table 3.7) highlighted three variants in *LHFPL5*, *MRPS10* and *MMP17*.
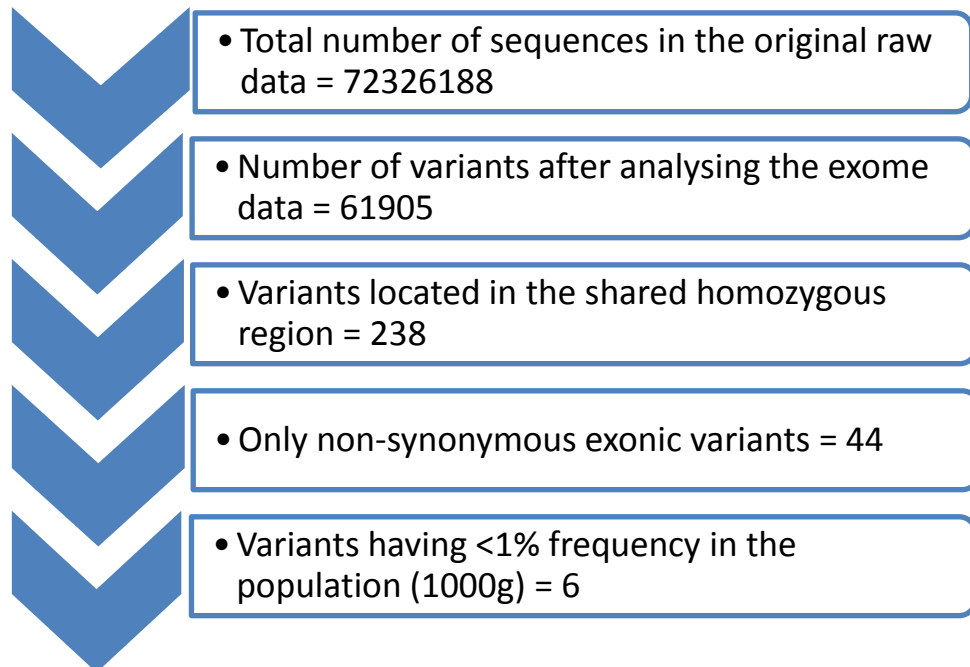

Figure 3.8. Variants of Family-2 after analysis of exome sequencing and filtering steps. This flow diagram shows the reduction in the number of the variants after analysing the exome data (case IV.3), when coupling the analysis with homozygosity mapping results (case IV.1) and after the filtering steps. All 2166 variants are located in the three shared homozygous regions on, chromosomes 5, 6 and 12. filtering steps reduced the variants to and then to only 3.

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	CADD Score	Freq in 1000G
chr6	35782485	*LHFPL5*	exonic	LHFPL5:NM_182548:exon2:c.T575C	p.L192P	28	0
chr6	42185562	*MRPS10*	exonic	MRPS10:NM_018141:exon1:c.C26T	p.A9V	8.928	0.005
chr12	132313098	*MMP17*	exonic	MMP17:NM_016155:exon1:c.59_60insGCTGCCGCT	p.R20delinsRLPL	1.276	0

Table 3.7. The final list of candidate variants for Family-2. The table shows the details of each variant, including its pathogenicity score based on CADD and frequency in the 1000G database. The variant highlighted in green is the strongest candidate.

3.2.1.13 Segregation of candidate variants

Based on CADD score, it is unlikely that the MRPS10 and MMP17 variants have any functional consequences. However the score of the LHFPL5 variant looks pathogenic on CADD (28). Furthermore, it fits with part of the diagnosis because previous researchers found link with deafness (Marková et al., 2016; Shahin et al., 2010). A novel homozygous missense variant (Figure 3.9) in *LHFPL5* (NM_182548:exon2:c.T575C, p.L192P) was the only candidate variant that showed segregation and consistency with the described phenotype in Family-2. This variant was consistent with a recessive pattern of inheritance, as both parents were found to be heterozygous, while both affected boys were homozygous and the unaffected boy was heterozygous (Figure 3.10).
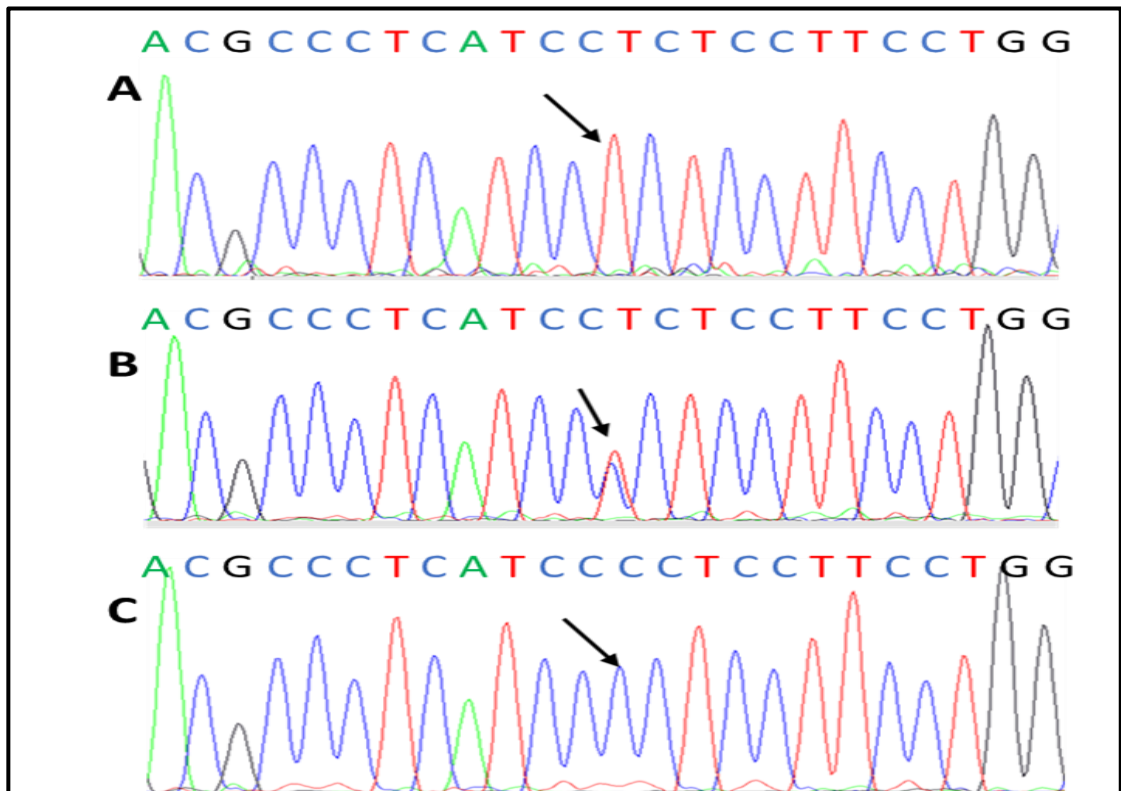
Figure 3.9. Electropherograms of the c.T575C variant. A. The wild type sequence from an ethnically matched control. The arrow indicates the exact location of the thymine (T), at which the substitution takes place. This trinucleotide (CTC) encodes Leucine (L). B. The heterozygous sequence pattern seen in case III.1 (mother), also presents in case III.2 (father) and case IV.2 (unaffected brother). The alternative cytosine (C) could be seen overlapping the reference thymine (T). C. The homozygous variant seen in both affected cases, IV.1 and IV.2. So, this missense mutation resulted in replacing the CTC trinucleotide by CCC, which encodes Proline (P).
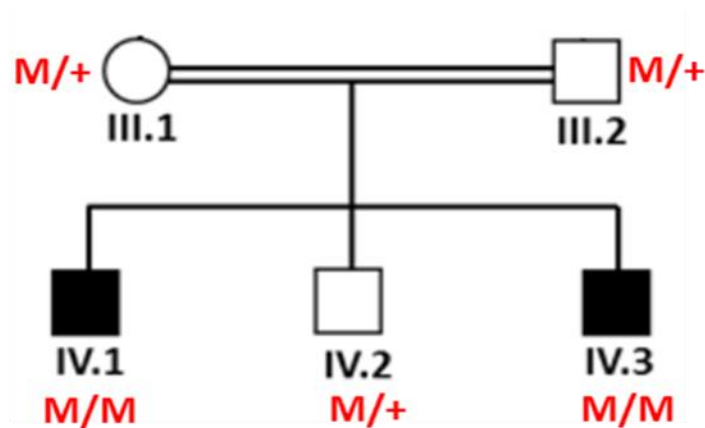
Figure 3.10. Segregation of c.T575C, p.L192P in Family-2. Segregation of the variant in *LHFPL5* within the nuclear family taken from the original extended family. + is the wild type allele; M is c.T575C, p.L192P.

3.2.1.14 Conservation and mutation frequency

The LHFPL5 mutation (NM_182548:exon2:c.T575C, p.L192P) changes the leucine at position 192 to a proline (p.L192P). This position is conserved across multiple species as shown in Figure 3.11, suggesting that the residue may have an important function. In order to rule out the identified variant from being a common SNP in the Omani population, one hundred chromosomes from 50 Omani healthy individuals were screened for the variant, but none was found to carry it. Furthermore, c.T575C is not present in 141,353 unrelated individuals in the Genome Aggregation Database (gnomAD). These findings suggest that c.T575C is a very rare change, which is consistent with it being pathogenic.
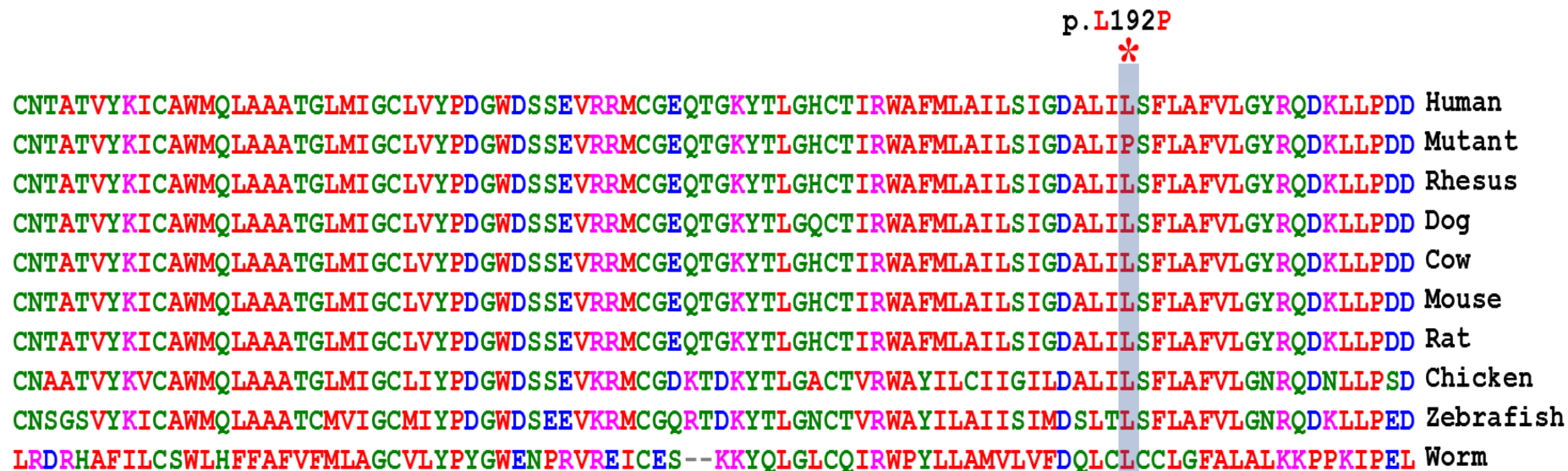
Figure 3.11. ClustalW alignment of p.L192 conservation in LHFPL5. ClustalW alignment of LHFPL5 amino acid sequences around the missense mutation, p.L192P, showing a highly conserved leucine residue at position 192. Accession numbers for LHFPL5 protein are human NP_872354.1, rhesus XP_001112357.2, dog XP_005627321.1, Cow NP_001192875.1, mouse NP_080847.2, rat NP_001013928.1, chicken NP_989729.1, zebrafish XP_686620.5 and worm NP_503066.1.

3.2.1.15 X-linked variants in Family-2

As both affected individuals in Family-2 (Figure 3.6) are male, this raised the possibility of an X-linked mode of inheritance. The existence of the affected male uncle would argue against this as there is no connection between siblings IV1 and IV3 and their paternal uncle. However the clinical and genetic status of this individual has not been assessed, so an X-linked mode of inheritance could not be ruled out. Therefore, variants identified in the WES of subject IV.3 and which were located on the X chromosome were also analysed. Filtering criteria removed synonymous changes, variants with a frequency of more than 1% and those that have coverage depths of less than 10. This filtering shortened the list of variants from 5301 to 664. The remaining changes were then compared to an updated list of X-linked genes that are known to be associated with intellectual disability. This list is managed and updated by the University of Colorado, and a copy is included in Appendix 8. Filtering against the known genes identified three variants as plausible candidates (Table 3.8).

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	1000g	CADD
X chr	17750076	*NHS*	exonic	NHS:NM_198270:exon8:c.C4385G	p.S1462C	0	23
X chr	49846369	*CLCN5*	exonic	CLCN5:NM_000084:exon6:c.C588G	p.I196M	0	15.86
X chr	122336600	*GRIA3*	exonic	GRIA3:NM_001256743:exon3:c.382dupG	p.G127fs	0	1.126

Table 3.8. Final list of X-linked candidate variants for Family-2. This table shows the details of each variant, including its pathogenicity score based on CADD and frequency based on 1000G database. The variant highlighted in green is the strongest candidate.

By combinining the CADD score plus the partial phenotype match, NHS variant was a better candidate. By performing Sanger sequencing, a novel missense variant (Figure 3.12) in part of the *NHS* gene (NM_198270:exon8: c.C4385G, p.S1462C) was the only candidate variant that showed segregation within the family. This variant followed an X-linked pattern of inheritance, as the mother was found to be heterozygous and both affected boys were hemizygous due to the fact that the X chromosome in males has no allelic counterpart (Figure 3.13). The mutation was not present in 50 healthy individuals of the same ethnicity as Family-2 or in 141,353 unrelated individuals in the Genome Aggregation Database (gnomAD).
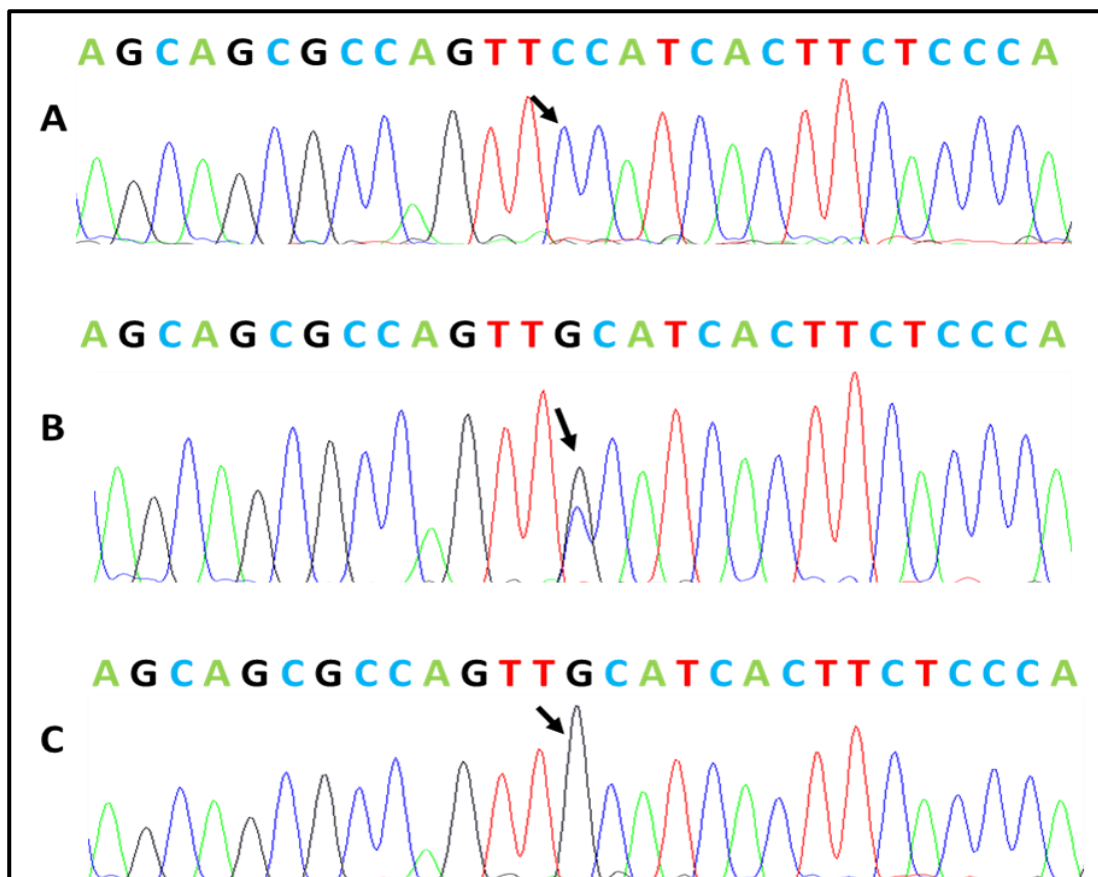
Figure 3.12. Electropherograms of the c.C4385G variant. A. Wild type sequence from the father (III.2), also present in the unaffected brother (IV.2). The arrow indicates the exact location of the cytosine (C), at which the substitution takes place. This trinucleotide (TCC) encodes Serine (S). B. Heterozygous mutation found in the mother (III.1). The alternative guanine (G) could be seen overlapping with the reference cytosine (C) as the substitution occurs in one of the X chromosomes. C. Hemizygous mutation seen in both affected boys (IV.1 & IV.3). Therefore, this missense mutation resulted in replacing the TCC trinucleotide by the TGC that encodes Cysteine (C).
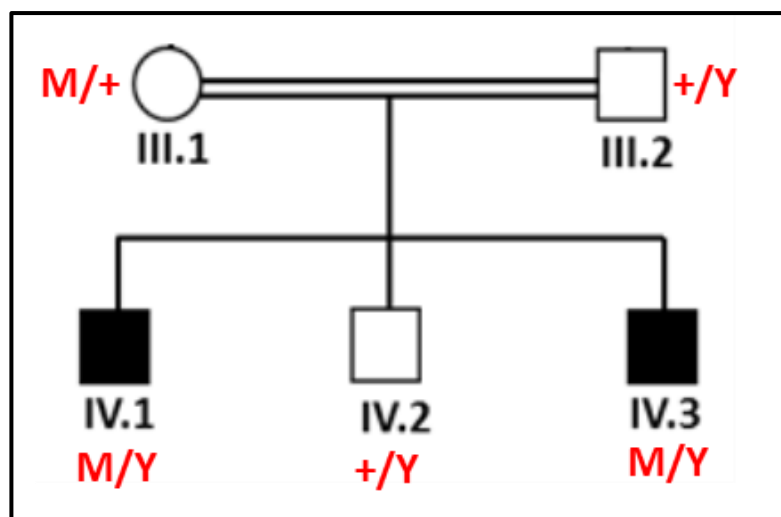
Figure 3.13. Segregation of c.C4385G, p.S1462C in Family-2. Segregation of the *NHS* variant in the nuclear family from the original extended family. + = wild type allele; Y = male chromosome; M = c.C4385G, p.S1462C.

3.2.1.16 Description and clinical details for Family-3

A consanguineous nuclear family is shown in the fourth generation of the extended family (Figure 3.14). The nuclear family resulted from marriage between first-cousin members (III.1 & III.2), and three of their five children are affected with ID and autistic features. They also have scoliosis, myopia and dysmorphic features (long, thin and myopathic face, flat malar region, laxity of joints, bilateral ptosis, hypertelorism, and slender hands and feet). Family structure suggests an autosomal recessive mode of inheritance.
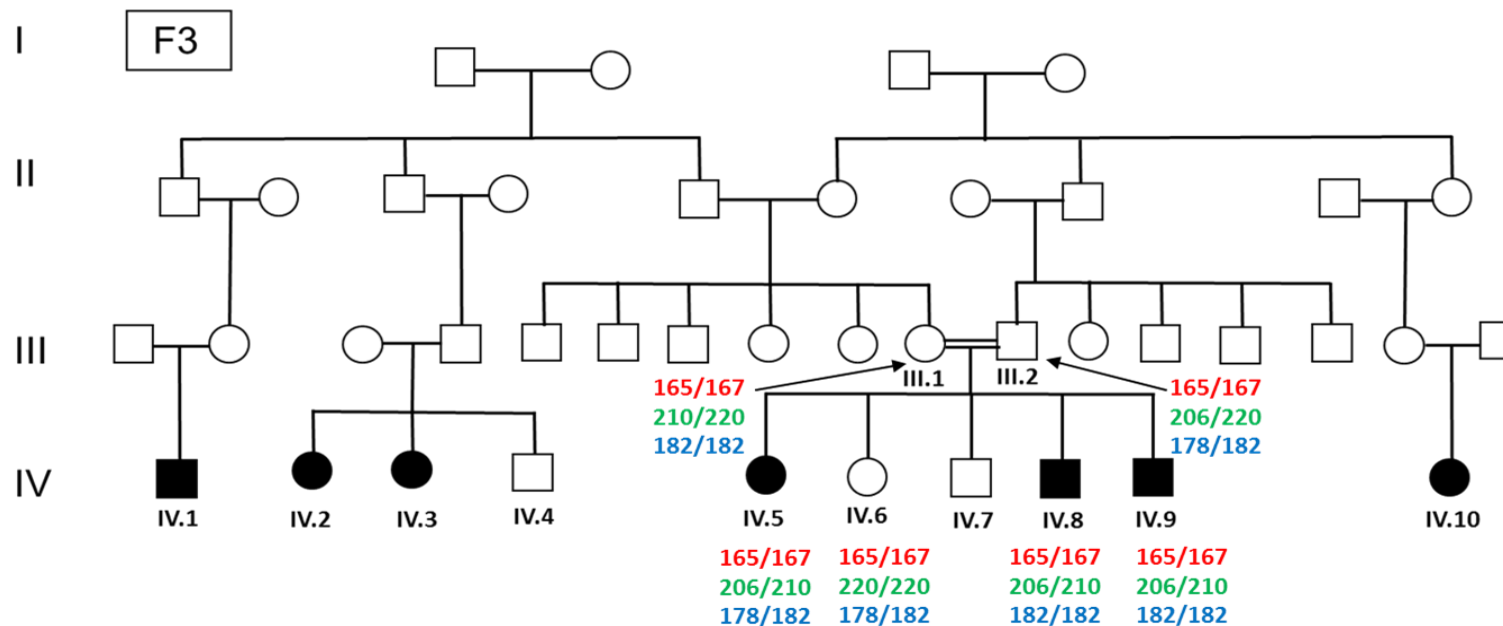
Figure 3.14. Family-3 pedigree with microsatellite genotyping. There are 7 ID affected individuals in the fourth generation of this family but no samples or detailed information was obtainable from cases IV.1, IV.2, IV.3 and IV.10. The nuclear family has three affected children, all with moderate-severe ID. Genotyping is shown of the cases from whom DNA was available. The red results are the detected alleles when genotyping with marker D1S478, the green results are for marker D4S3042 and the blue results are for the D7S483 marker. Whole exome seqencing was done for case IV.5 and SNP array genotyping for cases IV.8 and IV.9.

3.2.1.17 Microsatellite genotyping in Family-3

Three polymorphic microsatellite markers (D1S478, D4S3042 & D7S483) were used to genotype each case where a DNA sample was available to ensure the correct parentage within the pedigree and that no mixing between samples took place. Figure 3.14 shows the tested cases and their genotypes. All the children (IV.5, IV.6, IV.8 & IV.9) have shown genotype patterns that are consistent with being inherited from their parents (III.1 & III.2).

3.2.1.18 Homozygosity mapping in Family-3

Homozygosity mapping of this family was analysed based on the results of the SNP data from subjects IV.8 and IV.9, as well as SNP data extracted from exome sequencing of subject IV.5. Four shared homozygous regions could be defined which are detailed in Figure 3.15. As the family are consanguineous and inheritance is recessive, it was expected that one of these regions would harbour the causative mutation.

A

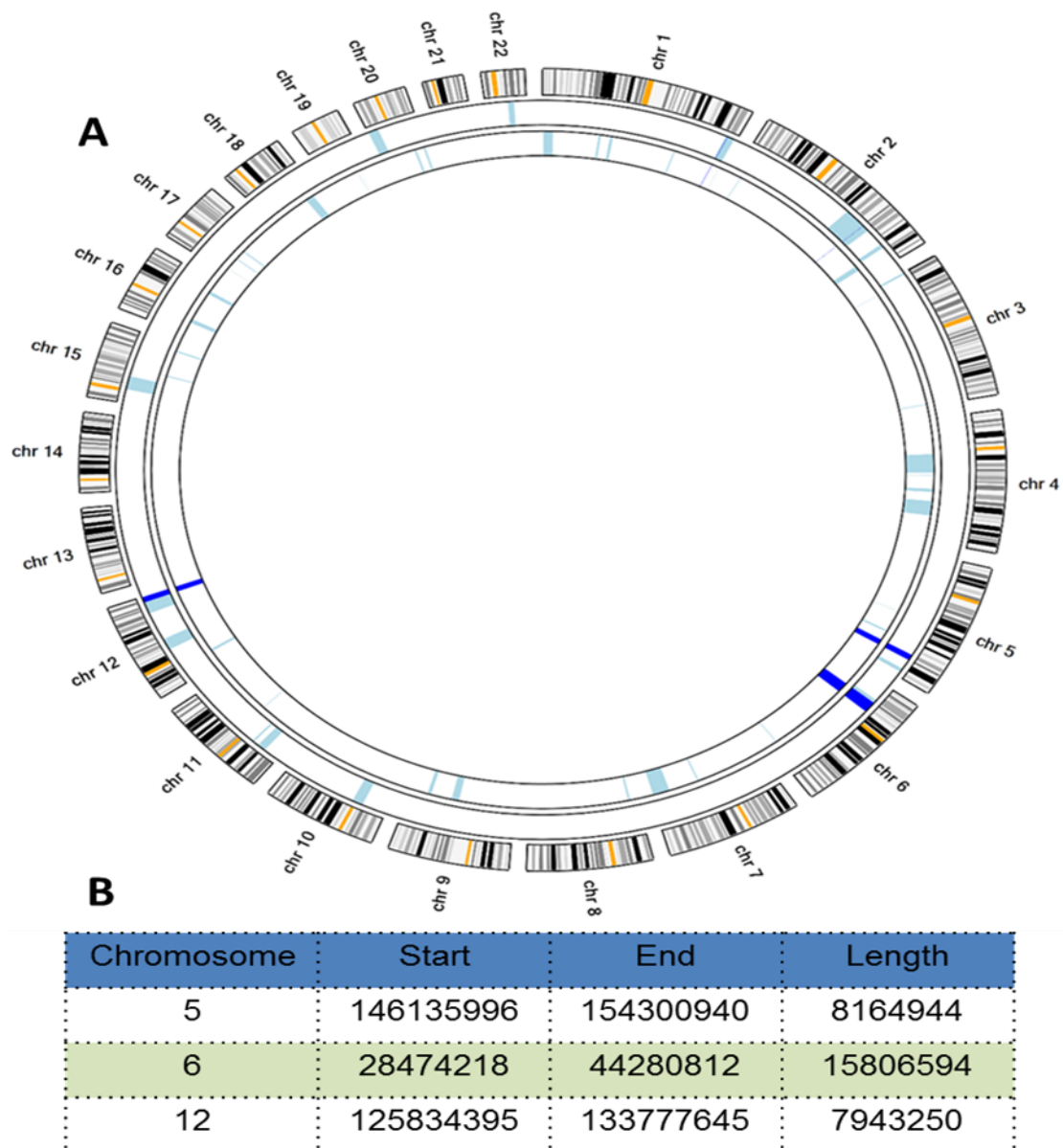| Chromosome	Start	End	Length
6	12344262	14912369	2568107
10	95552653	123845322	28292669
13	32913055	38113532	5200477
17	8802475	17184399	8381924
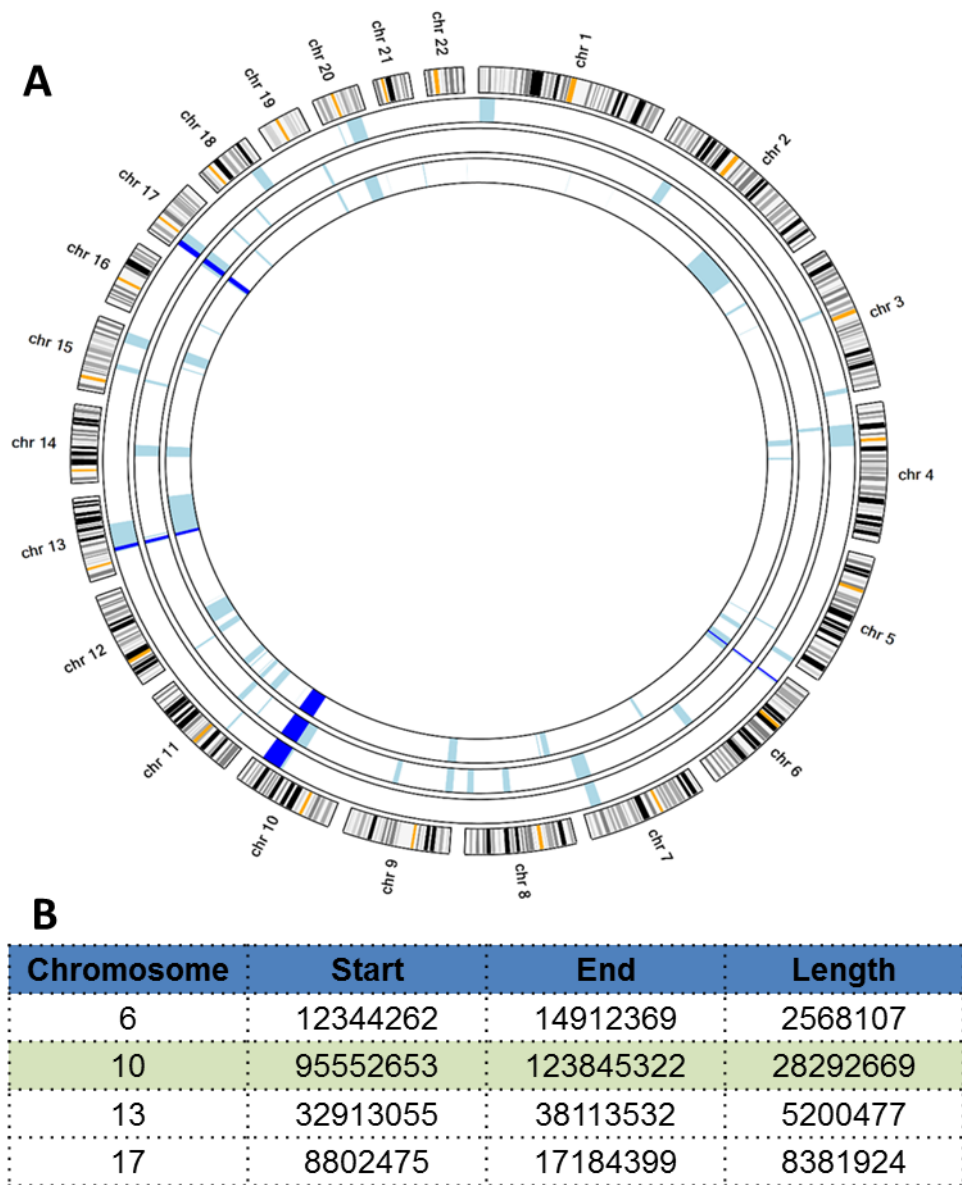
Figure 3.15. Shared homozygous regions in Family-3. A. An ideogram where the outer circle contains template segments of the 22 autosomal chromosomes. The three inner circles plot homozygous regions in SNP data from subject IV.8 and IV.9, and WES data from subject IV.5. The homozygous regions within the genomic DNA of each individual are shown in light blue, and the shared homozygous regions (number=4) common to all affected individuals are shown in dark blue. B. Details of the shared homozygous regions. Corresponding coordinates were based on human genome reference hg19. The four shared regions are on different chromosomes. The biggest region, which is highlighted in green, is on chromosome 10; this region has the highest probability of harbouring the causative mutation.

3.2.1.19 Exome sequencing in Family-3

An Illumina HiSeq 2500 sequencer was used to generate whole exome next generation sequencing (Section 2.11) on subject IV.5. Before proceeding with

exome analysis, the quality score of the next generation sequence data obtained was checked using FastQC software (Section 2.11.9.1) to ensure that the data are of good quality with no problems or biases. Table 3.9 shows the basic statistics of exome sequencing in IV.5.

Measure	Value
File Name	IV.5 of Family-3
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total reads	43218122
Sequence flagged as poor quality	0
Sequence length	101
%GC	49

Table 3.9. Basic statistics of IV.5 exome sequencing reads. NGS analysis of DNA from IV.5 did not show any sequence of poor quality. The single reported value for the length of sequence and the percentage of the GC content suggest that the sample is of good quality.

Exome analysis was carried out using Bowtie2 aligner and hg19/GRCh37 reference genome. IV.5 sample was treated by SAMtools to sort and index the sequencing reads and then to convert into BAM format. Removal of all duplicates was done using Picard tools and Unifield Genotyper was used for calling the variants. Variants annotation was performed by using Annovar software as described in section 2.11.9.3. By the end of the analysis task, had 55491 variants that were taken for further selection and prioritization.

3.2.1.20 Variant selection, filtering and prioritization

The total number of variants (554891) was reduced to 9032 by removing the unlikely causatives which are not located in the four shared homozygous regions (Figure 3.16). The exclusion of those variants was due to the fact that the study was based on a consanguineous family with a rare recessively inherited disorder

(ID), which highly suggest that the disease is caused by a homozygous variant that resides within shared homozygous regions.
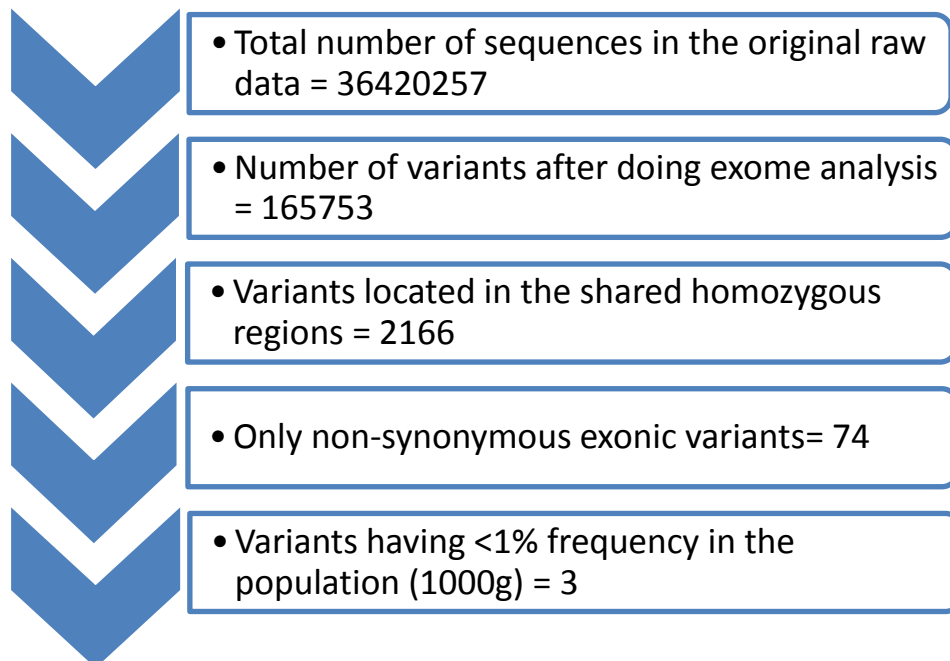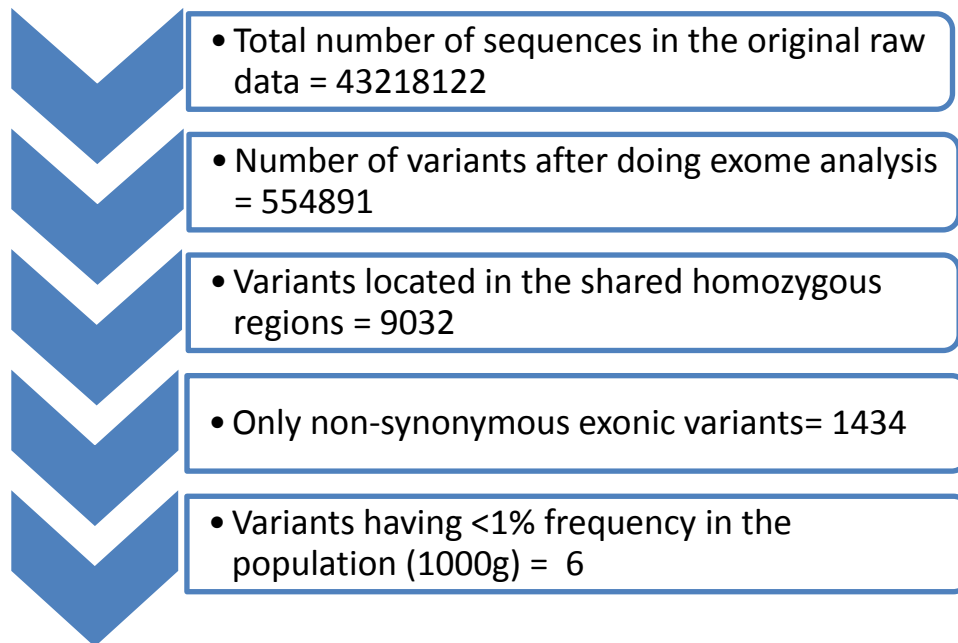

Figure 3.16. Variants of Family-3 after analysis of exome sequencing and homozygosity mapping. This flow diagram shows the reduction in numbers of the variants after analysing the exome data (case IV.1) after coupling the analysis with the homozygosity mapping results (cases IV.8 and IV.9). The power of utilising the results of homozygosity mapping can be seen in the marked reduction in the number of variants. All 9032 variants are located in the four shared homozygous regions on chr6, chr10, chr13 and chr17.

3.2.1.21 Filtering and variant prioritization

The 9032 variants were further filtered by taking out all those with coverage depth <10 as well as synonymous changes that encode the same amino acids, and keeping only exonic and splicing variants. This filtering step brought the list down to 1434 variants. Further filtering involved removing the variants with a frequency more than 1%, and all the heterozygous changes, leaving only 6 candidates (Table 3.10).

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	CADD	1000G
chr10	98155678	TLL2	exonic	TLL2:NM_012465:exon12:c.C1484T	p.T495M	17.83	0.02
chr10	99343381	ANKRD2	exonic	ANKRD2:NM_001129981:exon8:c.C883T	p.R295W	34	0
chr10	102258991	SEC31B	exonic	SEC31B:NM_015490:exon13:c.G1510A	p.V504M	12.80	0.01
chr10	119044044	PDZD8	exonic	PDZD8:NM_173791:exon5:c.2197_2200del	p.733_734del	35	0
chr13	35747681	NBEA	exonic	NBEA:NM_015678:exon27:c.A4504G	p.S1502G	15.85	0.02
chr17	17039562	MPRIP	exonic	MPRIP:NM_015134:exon6:c.534_539del	p.178_180del	18.20	0

Table 3.10. Final list of candidate variants for Family-3. This table shows the details of each variant, including its pathogenicity score based on CADD and frequency based on the 1000G database. The variants highlighted in green are the two strongest candidates.

3.2.1.22 Segregation of candidate variants in Family-3

One of the candidate variants that appeared in the final list is in *NBEA*, which is a well-established autism gene but this candidate did not segregate within the ASD and ID of Family-3. In addition, none of the finally prioritised candidates have a known evidence of involvement in ID. However, the obtained CADD scores as well as literature search about the known functions of these genes indicated that ANKRD2 and PDZD8 are probably the most important among the proposed candidates. Upon Sanger sequencing, both of the variants showed segregation and consistency with the ID diagnosis in Family-3. The first was a missense mutation in *ANKRD2* (NM_001129981:exon8:c.C883T, p.R295W) (Figure 3.17) and the second was a frameshift deletion of four base pairs in *PDZD8* (NM_173791:exon5: c.2197_2200del, p.733_734del) (Figure 3.18). Both mutations followed the suggested recessive pattern of inheritance, as both parents were heterozygous, both affected children were homozygous, and the unaffected children were either heterozygous or wild type (Figure 3.19). These mutations are novel and they are predicted to be pathogenic (CADD score of 34 for c.C883T; 35 for c.2197_2200del).
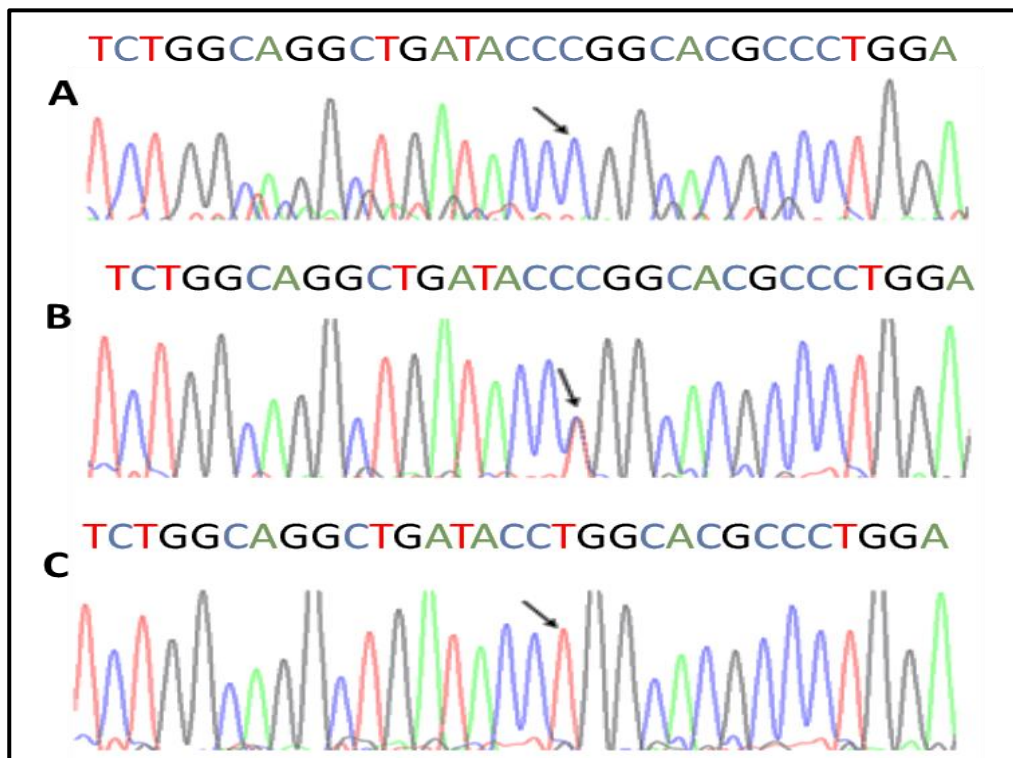
Figure 3.17. Electropherograms of the c.C883T mutation. This mutation occurs in exon 8 of ANKRD2 and the arrow indicates the location of the substituted base;

CGG encodes arginine (R). A. The normal wild type found in an unaffected case. B. Both parents were heterozygous with cytosine (C) to thymine (T) substitution in one copy of ANKRD2. C. All the affected cases have a homozygous substitution replacing a C base with a T. The newly produced trinucleotide (TGG) encodes tryptophan (W).
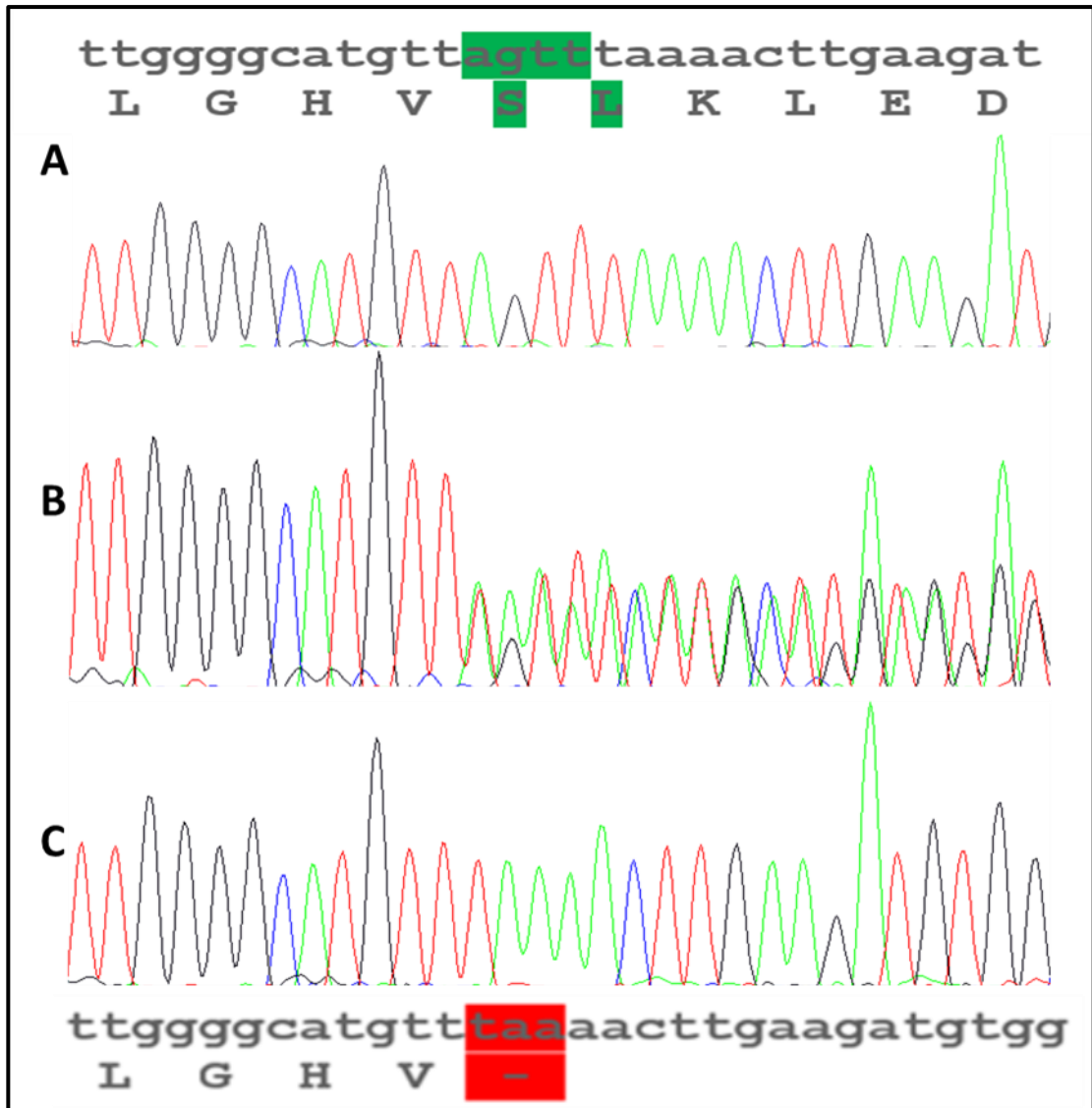
Figure 3.18. Electropherograms of the c.2197_2200del deletion. The upper alignment highlights, in green, the location of the 4-bp deletion (agtt) that forms part of the Serine (S) and Leucine (L) amino acids. A. Wild type *PDZD8* in an unaffected case. B. Both parents were heterozygous with the 4-bp deletion in one copy of *PDZD8*. Deletion in one copy of chromosome 10 causes the overlapping frameshift in the sequencing peaks. C. All the affected cases are homozygous for the frameshift 4-bp deletion, which induces a premature stop mutation that is highlighted in red.
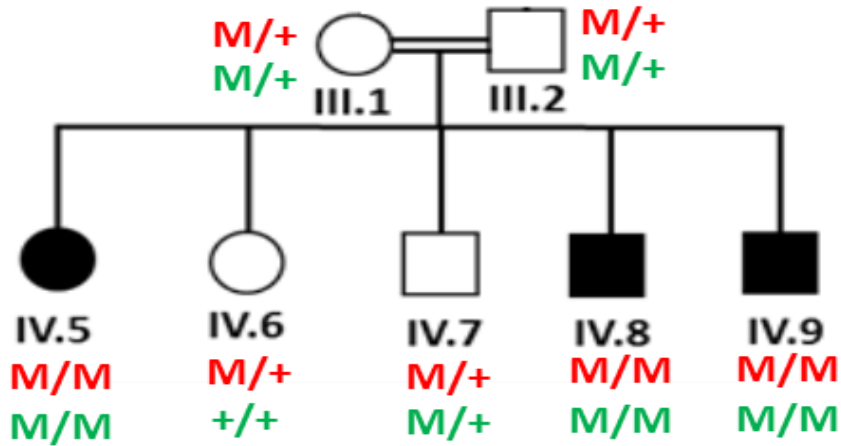
Figure 3.19. Segregation of c.C883T, p.R295W in *ANKRD2* and c.2197_2200del, p.733_734del in *PDZD8* in Family-3. This is only the nuclear family taken from the original extended family. The co-segregation of homozygosity for the mutations in *ANKRD2* and *PDZD8* and ID diagnosis is shown. Red text is *ANKRD2*: c.C883T, p.R295W. Green text is *PDZD8*: c.2197_2200del, p.733_734del. + is wild type allele; M is mutated allele.

3.2.1.23 Conservation and mutation frequency

Measurement of conservation is applied to understand the impact of variation at a given region and it provides an indication as to whether the variation would be damaging or not (Cooper Gregory M and Shendure, 2011). Sequence of a range of species through evolution was observed for the p.R295 position (Figure 3.20) reported in ANKRD2 and the alternative tryptophan amino acid (W) was found to be present in spiny softshell turtle. Although this type of turtle has the same *ANKRD2* variant reported in Family-3, it still function normally. In fact, this finding argues against the *ANKRD2* variant being the causative of the described phenotypes in this family.

Figure 3.20. ClustalW alignment of the normal amino acid sequence of ANKRD2. ClustalW alignment of ANKRD2 amino acid sequences around the missense mutation, p.R295W, showing a highly conserved arginine residue at position 295 and presence of the tryptophan amino acid (W) in the normal sequence at position 295 in the spiny softshell turtle. Accession numbers for ANKRD2 protein are human NP_001123453.1, gorilla ENSGGOP00000014859, chimpanzee XP_507964.3, rhesus XP_001095517.2, dog XP_850948.2, Cow XP_003587918.2, horse ENSECAP00000007038, mouse NP_064417.1, rat NP_001101059.1, chicken XP_004942257.1, cave fish ENSAMXP00000002269, zebrafish XP_005157118.1, spiny softshell turtle APJP00000000.1 and frog NP_001090801.1.

135

Both the ANKRD2 c.C883T variant and the PDZD8 c.2197_2200del variant were screened for in 100 chromosomes from ethnically-matched healthy controls and neither of them was found. They were also not present in 141,353 unrelated individuals in the Genome Aggregation Database (gnomAD).

To search for additional cases with mutations in either of these genes, 11 Belgian ID cases with similar phenotypes were obtained by collaboration from Professor Paul Coucke, Department of Medical Genetics, Ghent University Hospital, Belgium. The exons of both genes were fully screened by Sanger-sequencing. Details of the PCR primers used are shown in Appendix 2. No mutation was found in *PDZD8*, but a homozygous *ANKRD2* variant (NM_001129981:exon1:c.C272T, p.A2V) was found in one Belgian ID case (Figure 3.21), with a CADD score of 20.9.
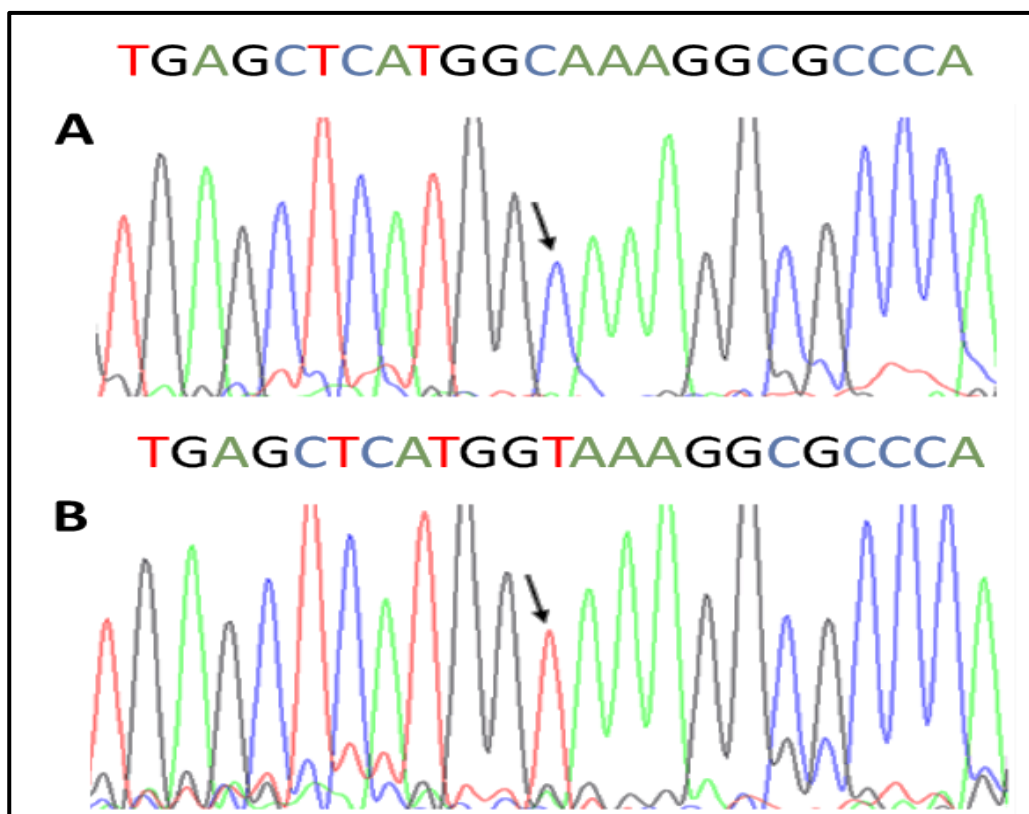
Figure 3.21. Electropherograms of the c.C272T variant in *ANKRD2*. This variant occurs in exon 1 and the arrow indicates the location of the substitution. A. Wild type sequence found in an unaffected case showing the GCA that encodes alanine (A). B. The affected Belgian case showed homozygous substitution of a C to a T. This variant causes the trinucleotide to change resulting in GTA that encodes valine (V). No heterozygous mutation was found for this variant.

Multiple contacts were made with Dr.Coucke to try to get further clinical details about this case and additional family samples so that segregation could be checked, but unfortunately, no reply was received. The identification of this *ANKRD2* variant, but no *PDZD8* variants, in the Belgian ID cases suggests that *ANKRD2* has a role in the phenotypes descried for the affected members of Family-3. However, the fact that segregation testing could not be studied in the Belgian case make this not a strong assumption!.

3.2.1.24 *ANKRD2* and *PDZD8* mutation load in the UK10K database

For the purpose of checking various mutations and comparing their enrichment in the two genes, vcf files from 469 moderate to severe ID cases and 2432 control samples were downloaded from the UK10K, as described in section 2.12. After downloading, vcfhacks programme was used to search for variants of *ANKRD2* and *PDZD8* in the samples. The identified variants were then annotated using variant effect predictor (VEP) and scored with CADD. Variants with >10 CADD located in ANKRD2 and PDZD8 were checked and the mutation load was compared with healthy controls. A total of 93 mutations were found among the ID cases. Details about the identified mutations, including the ones in the control samples, are summarised in Table 3.11.

Gene	Variant	Frequency	CADD	No. of mutations in 938 ID cases alleles	No. of mutations in 4864 control alleles
ANKRD2	ENSP00000306163.5:p.Pro146Leu_rs36020819	0.004264995	26.7	13 heterozygous	11 heterozygous
	ENSP00000306163.5:p.Arg292Cys_rs149812913	0.001613611	33	5 heterozygous	10 heterozygous
	ENSP00000306163.5:p.Val229Leu_n/a	3.96181E-06	24.5	1 heterozygous	0
PDZD8	ENSP00000334642.5:p.Arg897Gln_rs363294	0.047812872	17.34	1 homozygous, 38 heterozygous	4 homozygous, 262 heterozygous
	ENSP00000334642.5:p.Ala11Val_rs78645354	0.051904249	21.1	33 heterozygous	358 heterozygous
	ENSP00000334642.5:p.Ile364Leu_rs138899307	0.000743079	14.46	2 heterozygous	6 heterozygous
	Compound heterozygous rs363294 & rs78645354			3	22

Table 3.11. *ANKRD2* and *PDZD8* mutation load. This list was generated based on 469 ID cases and 2432 control samples in the UK10K database. It contains only plausible mutations based on CADD scores >10. The frequency was based in the Genome Aggregation Database (gnomAD). The number of the heterozygous and homozygous mutations found is indicated for each variant. Student T-test showed no statistically significant difference for ANKRD2 and PDZD8, *p* value was 0.2367 and 0.5393, respectively.

3.2.1.25 Constraint metrics in the ExAC database

Constraint metrics are comparative measurements that could be attained from the ExAC database. For the missense, synonymous changes and CNV changes, a quantification, called a Z score, is created to indicate the deviation in numbers of the found variants from the expected ones (Samocha et al., 2014). An increased Z score (>3.9) means that the gene of interest has recorded a significantly (p-value of 10^{-3}) lower number of variants than the expected, suggesting that disruptive changes in this gene are not tolerated and do have a functional consequence. On the other hand, a gene with a decreased negative Z score is considered as under "low-constraint", suggesting variations are well tolerated. For the loss of function variants (LOF), the number of the observed and expected variants is utilized to calculate a score called pLI (the proportion of genes that are very likely intolerant) (Samocha et al., 2014). When the pLI score for a given gene is high, close to 1, it means that the gene is extremely intolerant to the LOF changes. The constraint metric is thus a useful tool for recognising diseases-causing variants as it provides a valuable indication about the significance of different genetic changes to a given gene (Lek et al., 2016).

The results of constraint metrics for both *ANKRD2* and *PDZD8* genes are summarised in Table 3.12. It was found that *ANKRD2* had observed values that were close to the expected values for all missense, CNVs and LOF variants, and its constraint metrics were not significant. However, *PDZD8* has a significant constraint metric for LOF (pLI = 1.000), which is the nature of the mutation reported in this family.

Gene Name	Changes Type	Expected variants	Observed variants	Constraint Metric
ANKRD2	Synonymous	58.4	51	Z = 0.60
	Missense	119	114	Z = 0.22
	CNV	3.1	1	Z = 0.44
	LOF	11.4	10	pLI = 0.00
PDZD8	Synonymous	189.3	168	Z = 0.96
	Missense	411	322	Z = 2.15
	CNV	2.6	4	Z = 0.25
	LOF	21.5	1	pLI = 1.00

Table 3.12 Constraint metrics of *ANKRD2* and *PDZD8*. These comparative measurements were obtained from the ExAC database. *ANKRD2* has observed values that are close to the expected values but *PDZD8* looks more constrained, especially with loss of function (LOF) mutations, as it showed only one variant compared with the 21.5 expected. The score of 1 in pLI for *PDZD8* means that it is extremely intolerant to the LOF changes. This gives more weight and significance to the *PDZD8* mutation but does not rule out the *ANKRD2* mutation.

3.2.1.26 *ANKRD2* and *PDZD8* expression in human tissues

Reverse transcriptase PCR was done on a human tissue panel (Section 2.5). The primers (Appendix 2) were designed to span at least one exon, allowing both cDNA and any contaminating genomic DNA to be detected. Primers of *ANKRD2* were designed to amplify exons 4 and 5 of the cDNA sequence giving a product size of 355pb. For *PDZD8*, the primers were chosen to amplify exons 2, 3 and 4 across cDNA sequence producing a product size of 406pb. The expression studies. *P53* was used as a control housekeeping gene and its primers were selected to produce 408pb cDNA band or 1057bp genomic band. The expression study showed that ANKRD2 is expressed in all the tested tissues but more abundant in brain, kidney, prostate, salivary gland and skeletal muscle, while PDZD8 is just ubiquitously expressed in all the tested tissues (Figure 3.22).
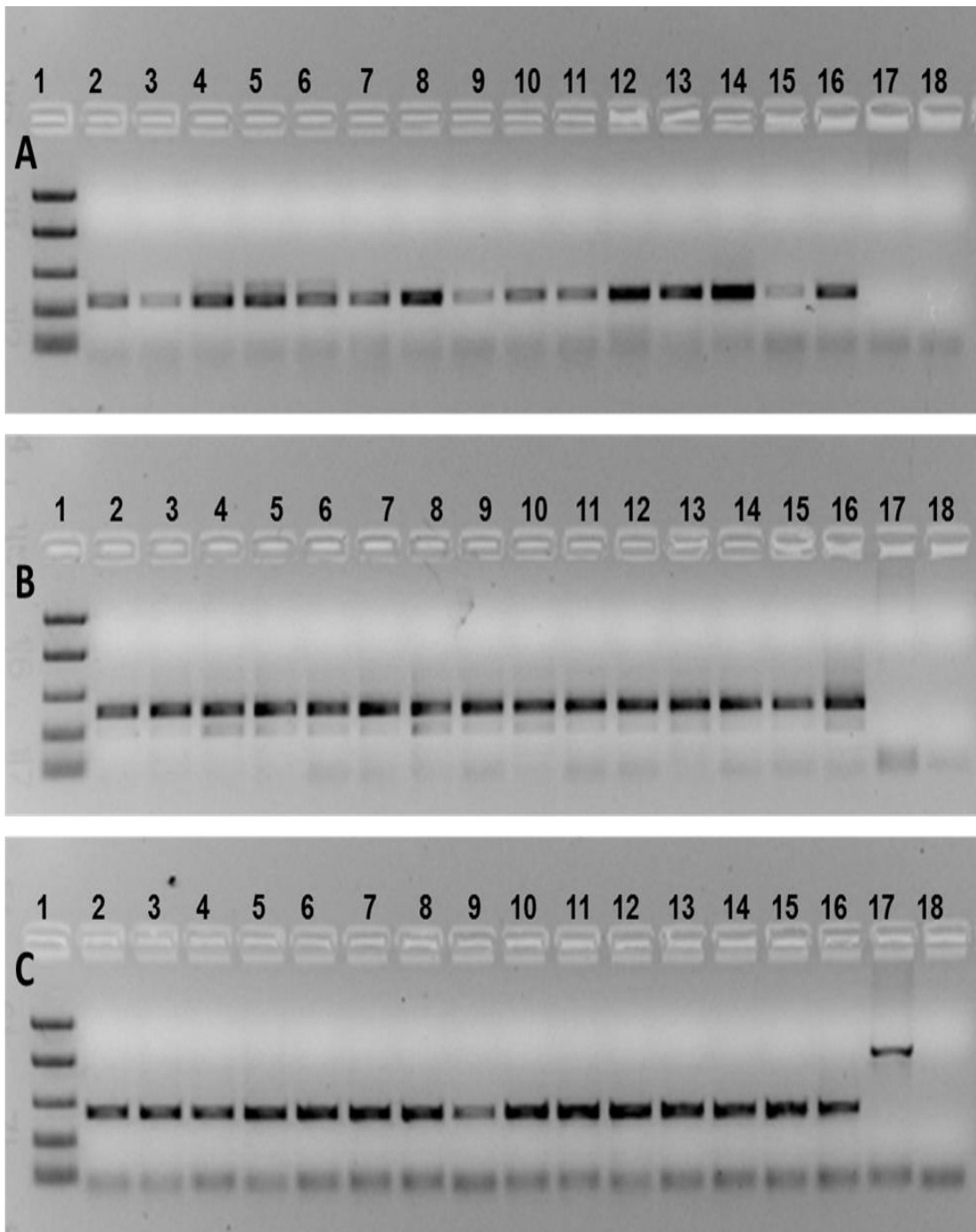
Figure 3.22. Expression of *ANKRD2* and *PDZD8* mRNA in various human tissues.
A. *ANKRD2* expression. B. *PDZD8* expression. C. P53 housekeeping gene expression
(control). Lane 1, 100-bp ladder; 2, Adrenal gland; 3, Bone marrow; 4, Brain
(cerebellum); 5, Brain (whole); 6, Fetal brain; 7, Fetal liver; 8, Kidney; 9, Liver; 10, Lung;
11, Placenta; 12, Prostate; 13, Salivary gland; 14, Skeletal muscle; 15, Spleen; 16,
Testis; 17, Genomic DNA;18, Negative Control.

3.2.1.27 Immunolocalisation in mouse brain regions

The localisation of ANKRD2 and PDZD8 proteins in sagittal sections of mouse brain was studied using immunohistochemistry (IHC). The antibodies were rabbit IgG polyclonal anti-mouse ANKRD2 (from Santa Cruz) and rabbit IgG polyclonal anti-mouse PDZD8 (from biorbyt), both were used at 1:100 dilution. For the negative slides, a non-immune immunoglobulin of the same isotype was as described in Section 2.17. Both ANKRD2 and PDZD8 were found to be moderately expressed in various brain regions including hippocampus and cerebrum (see Figure 3.23 and Table 3.13). In particular, there was strong staining of ANKRD2 in the substantia innominata (Figure 3.23C) and of PDZD8 in the choroid plexus (Figure 3.23F).
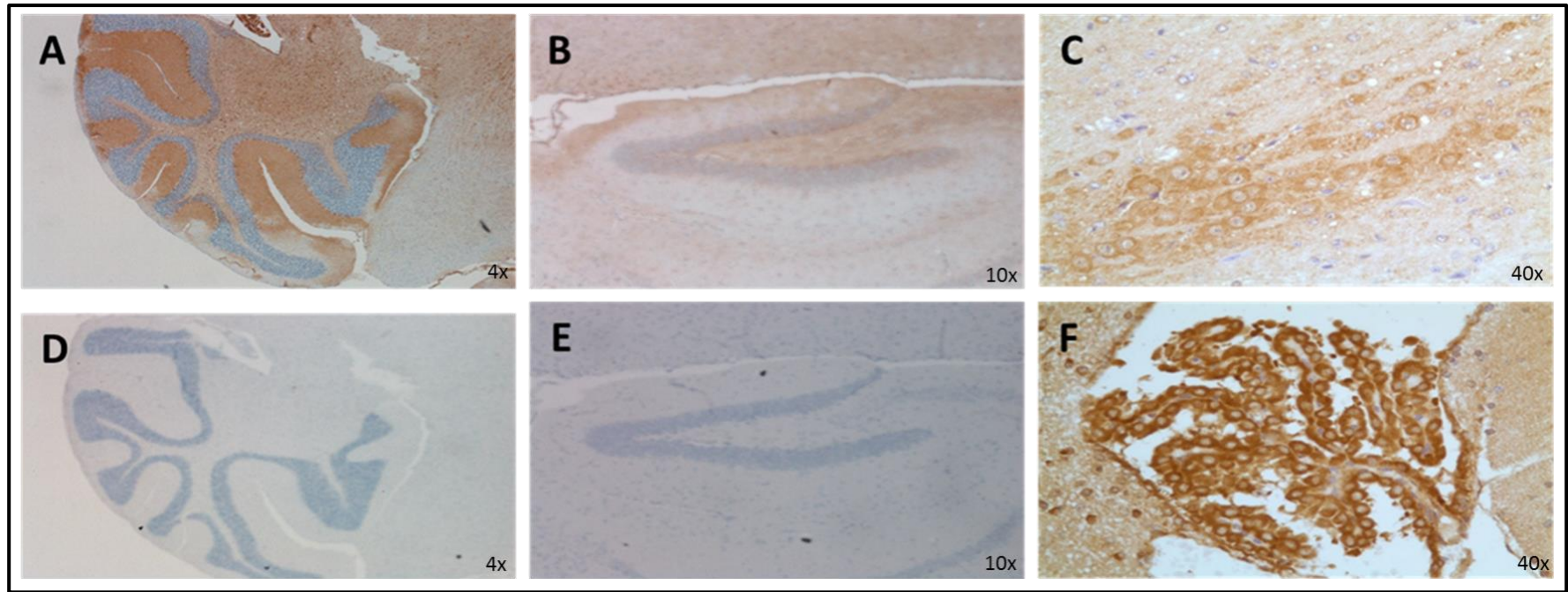
Figure 3.23. Examples of ANKRD2 and PDZD8 localization in regions of mouse brain. A. Strong positive localization of PDZD8 in the cerebellum. B. Moderate staining of ANKRD2 in the hippocampus. C. Strong positive localization of ANKRD2 in the substantia innominata (x40). D. Negative isotype control for cerebellum. E. Negative isotype control for hippocampus. F. Strong PDZD8 staining in the choroid plexus. For more details on the localization, refer to Table 3.13.

Brain Region	ANKRD2 localisation	PDZD8 localisation
Choroid plexus	None	Strong
Substantia innominata	Strong	Very mild
Hippocampus	Moderate in both the granular layer and Pyramidal neurons	Moderate in both the granular layer and Pyramidal neurons
Cerebral cortex	None	Mild in the Neocortical neurons
Cerebellum	Strong in molecular layer; no localisation in the Purkinje cell layer and granular layer	Strong in the molecular layer; moderate localisation in the Purkinje cell layer; no localisation in granular layer
Pontine neurons	Strong	Strong
Basal ganglia Neurons	Mild	Mild
White matter tracts	None	None

Table 3.13. Localization of ANKRD2 and PDZD8 in regions of mouse brain. The degree of staining for ANKRD2 and PDZD8 in the main brain regions is indicated. Choroid plexus and substantial innominata showed a marked difference in localization between these two proteins.

3.2.1.28 Behaviour studies in PDZD8 knock down *Drosophila*

3.2.1.28.1 Background

The equivalent of the *Homo sapiens ANKRD2* in *Drosophila* is *CG44001* but a mutant line with confident and direct orthologue of ANKRD2 was not available.

Given that the mutation in *ANKRD2* in Family-3 is a missense mutation, knock down of *CG44001* in *Drosophila* would be an inadequate model of the human mutation. However, the *PDZD8* orthologue in *Drosophila* is a single gene called CG10632 and a mutant line was available. *Drosophila* in which CG10632 could be knocked down were obtained by collaboration from the Vienna *Drosophila* Source Centre (VDSC). These flies are RNAi responder transgenic flies. The process of knocking down CG10632 was activated by crossing RNAi responder flies with a special type of *Drosophila* which had been created to express GAL4 throughout their bodies*,* as shown in Figure 3.24. Each of these lines had been created in white-eyed flies.
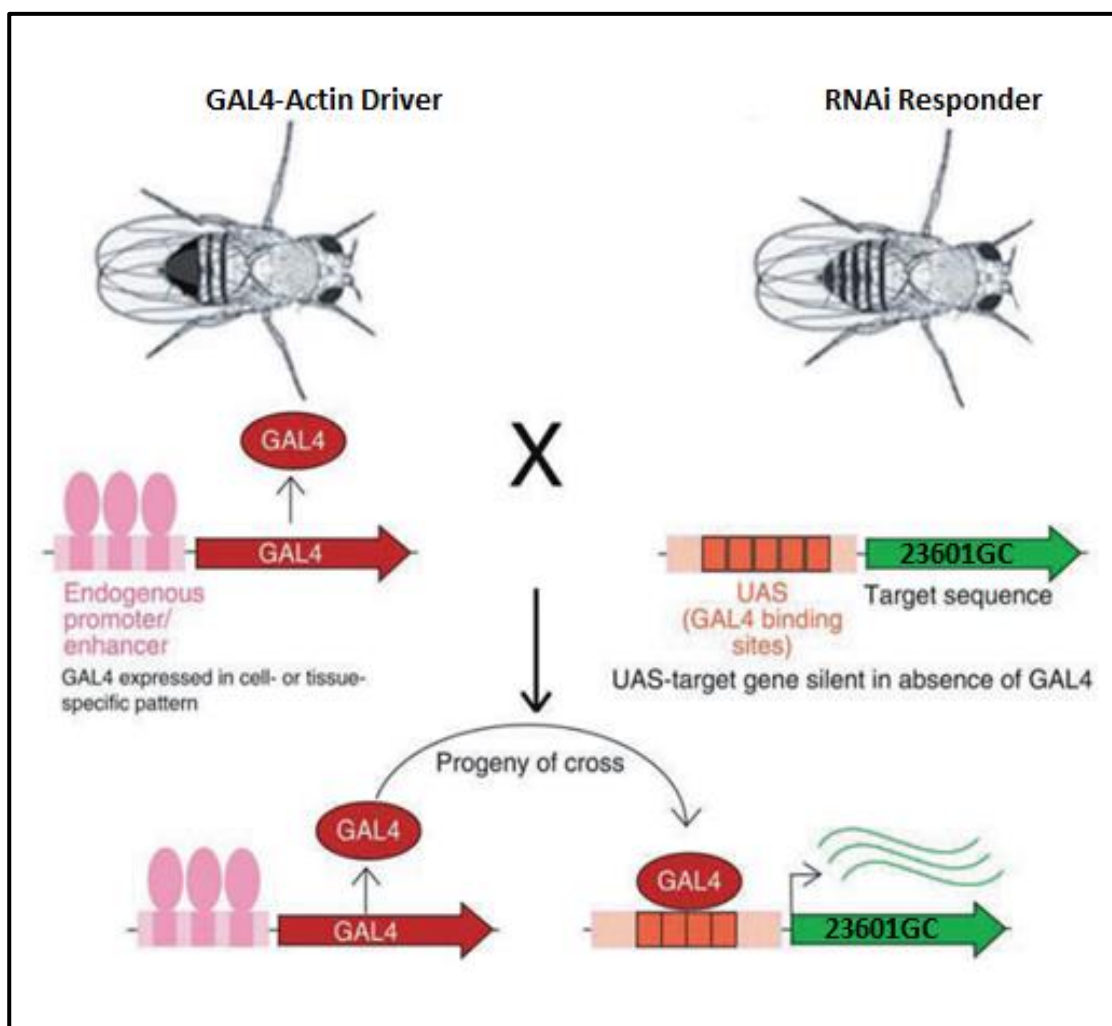
Figure 3.24. Activating the GAL4 system to knock down CG10632 expression. The flies with the GAL4-actin driver were crossed with the RNAi Responder flies. The driver flies also have the promoter/enhancer and the RNAi flies contain the inverted repeat sequence of CG10632 beside GAL4 upstream activating sequences (UAS). The process

that cause CG10632 knock down is silent in the absence of GAL4. However, GAL4 binds to its UAS in the produced progeny, leading to the activation of the GAL4 system, which initiates the transcription of CG10632. The activation of the GAL4 system causes the inverted repeats to create RNA that would be processed by the cellular machinery to make functional siRNAs. The mission of the produced siRNA is to repress CG10632. (Adapted from Caygill and Brand (2016) with a permission from Springer and Copyright Clearance Center, License Number 4106030489352).

3.2.1.28.2 CG10362 expression

White eye flies (W*) were used as a control in the experiment because they have the same genetic background as the RNAi responder line and would be expected to control for behavioural and physiological differences. The RNAi responder and the wild type (W*) were separately crossed to virgin GAL4-Actin driver females, as described in section 2.18.1. RNA extraction and cDNA synthesis was made from the progeny of different types of flies (Section 2.18.3.1). The relative expression of CG10362 was measured by quantitative PCR (qPCR) as described in section 2.8.3.2 and the assay was normalised against *eEF-1a* (Eukaryotic elongation factor 1-alpha) which is a housekeeping gene. Figure 3.25 shows the CG10362 expression in male *Drosophila*. Female progeny were, however, excluded from the behaviour study because they showed increased expression of CG10632 due to compensation. CG10632 expression following RNAi in males and females is shown in Appendix 9. To explain more, the responder line had the transgenic RNAi against CG10362 on chromosome X. The knockdown process was activated when these responder flies were mated with GAL4 driver. The RNAi has, therefore, worked to reduce CG10632 expression but because female flies have two copies of X chromosome, they could maintain a balanced CG10632 expression by amplifying the transcriptional yield of the intact active X chromosome (Heard and Disteche, 2006). Because of the compensation effect of CG10362 expression in females, only male flies showed CG10632 knock down (Figure 3.25) and were selected for behaviour studies.
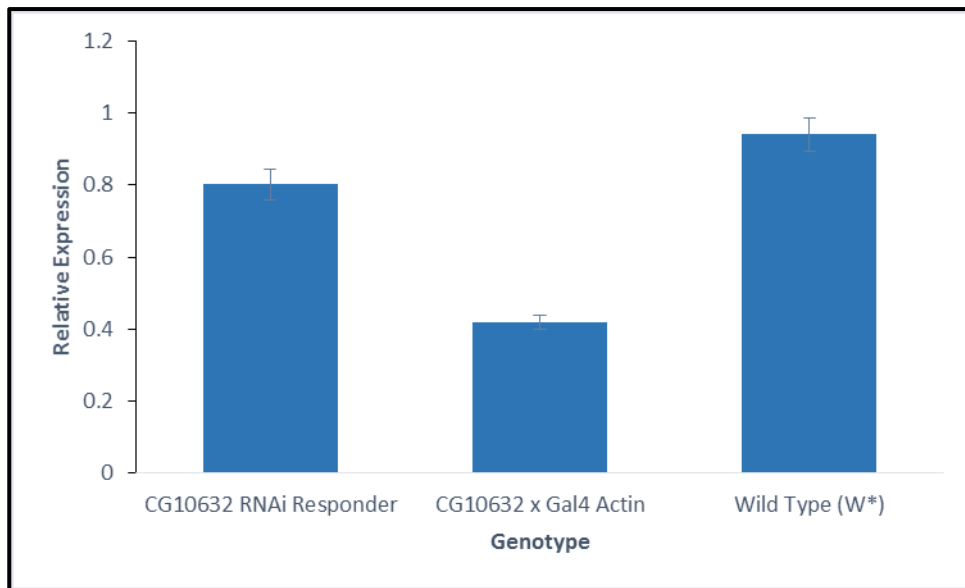
Figure 3.25. Relative expression of CG10632 in _Drosophila_. The expression was normalised against housekeeping _eEF-1a_. The flies with knocked down CG10632, that resulted from crossing CG10632 RNAi responder with the Gal4 Actin driver, showed a reduced expression of CG10632 compared with both CG10632 (RNAi responder) and wild-type flies (W*).

The expression of CG10362 was markedly reduced in the progeny of CG10632 RNAi responder flies crossed with Gal4 Actin flies (the middle block in Figure 3.25). The process appeared to knock down the CG10362 levels to around 40% of the wild type (W*). Therefore, it was confirmed that the knockdown of CG10362 was working and behaviour assays could be undertaken.

3.2.1.28.3 Olfactory learning T-maize assay

An experiment was first carried to select the optimum concentrations of 4-methylcyclohexanol (MCH) and 3-octanol (OCT), at which the naïve flies, which had not previously been exposed to these two smells, had no preference to follow one of the odour more than the other and flies were evenly distributed in the T-maze. For this experiment, male progeny of wild type flies (W*) crossed with actin driver flies were used as described Section 2.18.6.2. The flies were put in the T-maze apparatus, in groups of 50, and allowed 2 minutes to choose between OCT and MCH. The experiment was done using more than 2000 flies against 0.1, 0.15 and 0.3ml/l concentrations for each smell, and the result was achieved with

148

0.1ml/l MCH and 0.15ml/l OCT. Figure 3.26 indicates the percentage of flies that moved to each smell.
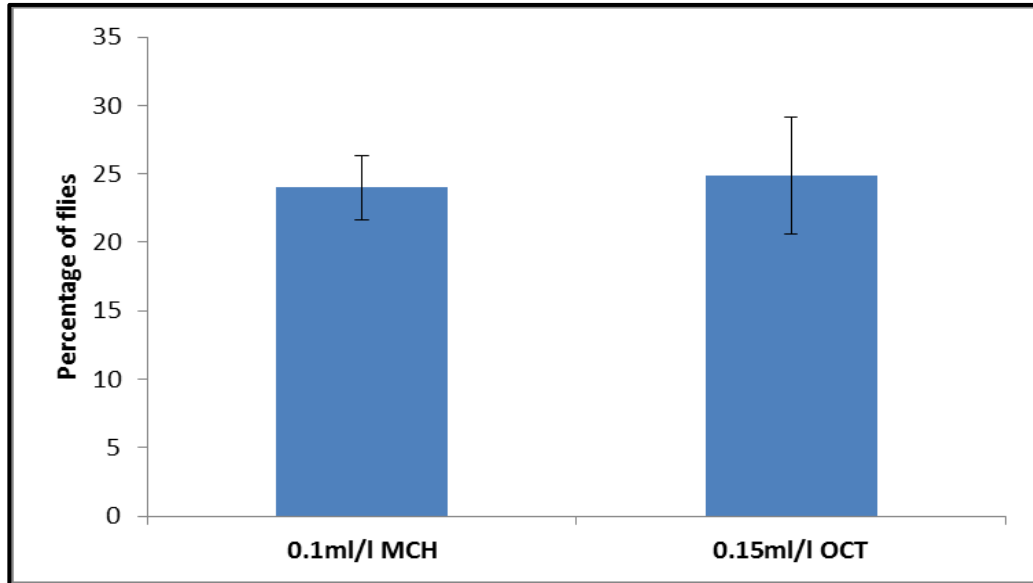
Figure 3.26. Percentage of the flies that moved towards MCH and OCT. The percentage of flies for a smell was calculated by counting the number of flies that selected to follow that particular smell, and dividing this by the total number of flies tested, then multiplying by 100.

The olfactory learning assay was done after training the flies to associate one smell with a shock but not the other. The training (also called conditioning) was carried against the two odours and was exchanged between having OCT or MCH as a safe smell, as described in section 2.18.6.3. After training the flies to remember avoiding one of the smell, because it was coupled with a mechanical shock, and to select the other safe smell, the olfactory learning test was done. This test indicated the associative memory of the flies if they learned from the training and could then avoid the "shock" smell. The assay involved training and testing 800 wild type flies and 800 CG10362 knock down flies. The flies were given 2 minutes in a T-maze apparatus to choose between the two odours (Section 2.18.6.4) and their learning index was calculated. Learning index represents the immediate memory of the flies and its calculation obtained performed using the following equation:

$$Li = \frac{A-B}{C}$$

Li is the Learning index.

A is the number of flies that chose the correct/safe smell.

B is the number of flies that chose the incorrect/shock smell.

C is the total number of flies in the test.

Table 3.14 shows the average Li, the standard deviation as well as the standard error of the mean (SEM) for both wild type and CG10362 knock down flies. The CG10362 knock down flies, that resulted from crossing the CG10632 RNAi responder line with the GAL4 driver line, showed no significant difference in the leaning index (P=0.159) compared with the wild type flies (Figure 3.27).

	CG10632 RNAi Line x Gal4 Actin	Wild Type (W*) x Gal4 Actin
Average	0.18665318	0.063518043
St.dev	0.15040407	0.179160439
S.E.M	0.053175869	0.063342781

Table 3.14. The average of learning index (Li) in the olfactory learning assay.

Although it might suggest that the wild type (W*) flies would achieve a learning index (Li) of 1, this was not the case (Figure 3.27). This is because the used wild type (W*) flies were of white eyed background, which might have interfered with the learning ability of the flies (see Section 3.3.3). It should be also mentioned that there were always some flies which remained in the area between the two smells until the end of the testing period but these flies were not included in the calculation.
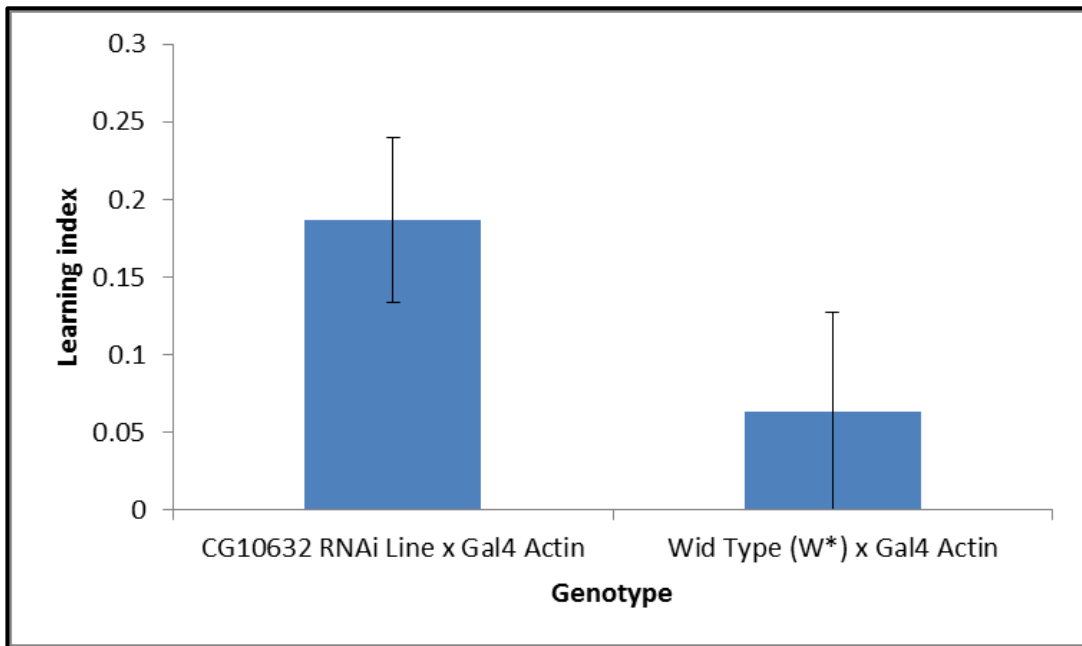
Figure 3.27. Learning index of olfactory learning assay. The CG10362 knock down flies (CG10632 x Gal4 Actin) appeared to have a higher learning index compared with wild type flies (W* x Gal4), but this difference was not significant.

In this assay, the CG10632 knockdown flies show a non-significant trend (T-test: t = -1.489, df = 14, p = 0.159) towards higher ability to remember the mechanical shock and choose the smell that avoids it.

3.2.1.28.4 Rapid iterative negative geotaxis (RING) assay

The rapid iterative negative geotaxis (RING) assay examines the locomotor ability of flies. When flies are startled and knocked down to the bottom of their container by rapid tapping against the bench, normal flies are expected to immediately act and move upward (Section 2.18.4). The measurements in this assay were done based on a total number of 250 CG10362 knock down flies compared with 250 wild type flies. Photographs taken during the test (example in Figure 3.28) clearly indicated the slower movement of the CG10362 knock down flies compared with the wild type flies, which progressed quickly to the top of the tube.

151
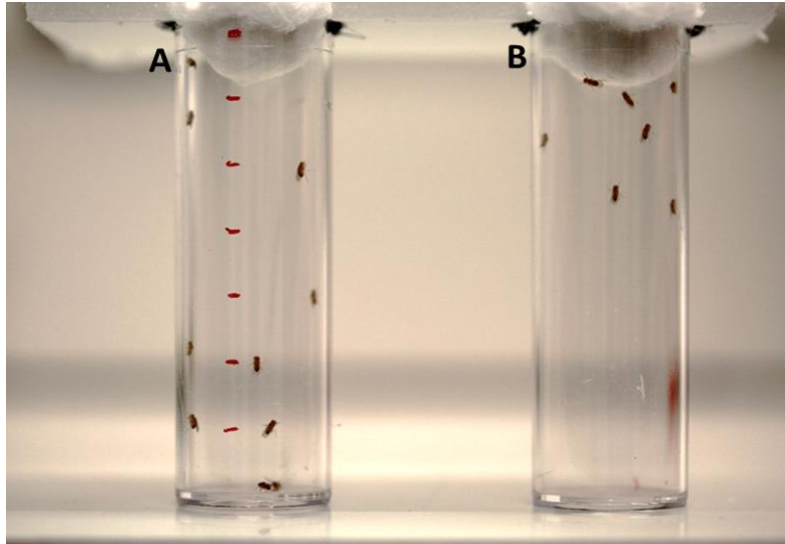
Figure 3.28. Example of RING assay. A. Tube containing the CG10362 knock down flies (CG10632 x Gal4 Actin). B. Tube containing wild type flies (W* x Gal4 Actin). The picture was taken with a digital camera (Canon, EOS 1200D) at 3 seconds after the last tap. The wild-type flies moved significantly faster upward than the knock down flies.
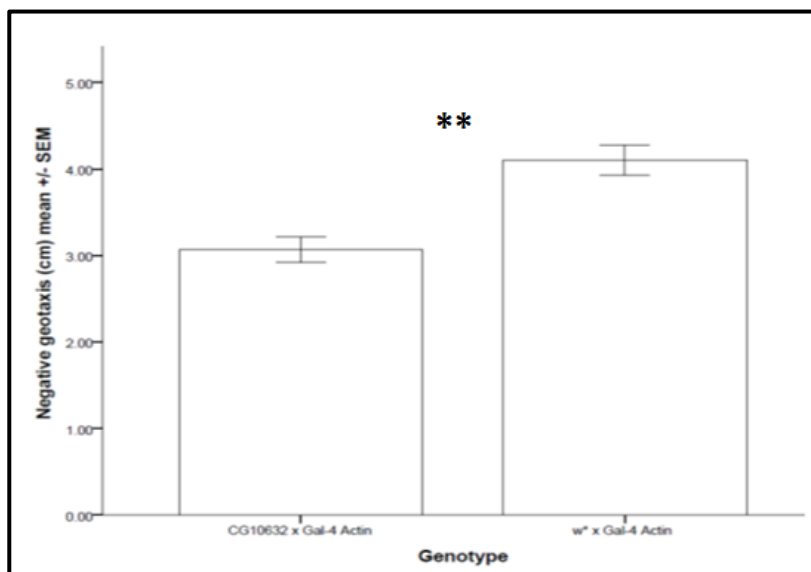
Figure 3.29. Negative geotaxis index. The CG10362 knocked down flies (CG10632 x Gal4 Actin) showed a lower negative geotaxis index compared with wild type flies. **<0.01.

The difference in the negative geotaxis index (Figure 3.29), measured as a mean of distances from the bottom in cm, was analysed using independent sample t-test and found to be significant (p=0.002). This finding is indicative of a loss of locomotor behaviour and could point to CG10362 being important in neurological or physical control of movement. This motor deficit, obtained through RING

assay, also suggests that there is a confounding effect on behavioural measures that require fly movement.

3.2.1.28.5 Courtship suppression assay

Courtship suppression is one of the studies that have been utilised to test the learning paradigm in Drosophila. The courtship suppression assay measures the courtship behaviour of male flies after one-hour exposure to mated females. The development of courting suppression is attributed to the presence of what is so called as conditioned and unconditioned stimuli. To explain more, all female drosophila have the conditioned stimuli, like 9-pentacosene, which would enable the males to recognise the female Drosophila and start courting them (Siwicki et al., 2005). Once a female is get mated, it would have unconditioned stimuli, like 7-trocosene and cis-vaccenyl acetate (cVA), that arise from male cuticular hydrocarbon during copulation process (Lacaille et al., 2007). The unconditioned stimuli last for different times depending on their type but they collectively act to train the male to suppress its courtship for few hours (Ejima A. et al., 2007). Normal male flies are reported to have suppressed courting tendency for four hours after an hour exposure to a mated female (Ejima Aki and Griffith, 2011).

The assay was carried out as described in Section 2.18.5. Male progenies of the CG10632 RNAi responder line crossed with the GAL4-Actin driver as well as of the wild type (W*) x GAL4-Actin driver were kept with mated females for one hour. Their courting activity was videoed during the first and last 10 minutes. The females were then replaced by virgin females and the courting activity of the males was videoed. To compare the courting activity of the males towards the newly placed females, sham-trained males were also exposed to virgin females and videoed. The experiment outlines and different time points, at which videos were recorded, are highlighted in Figure 3.30.
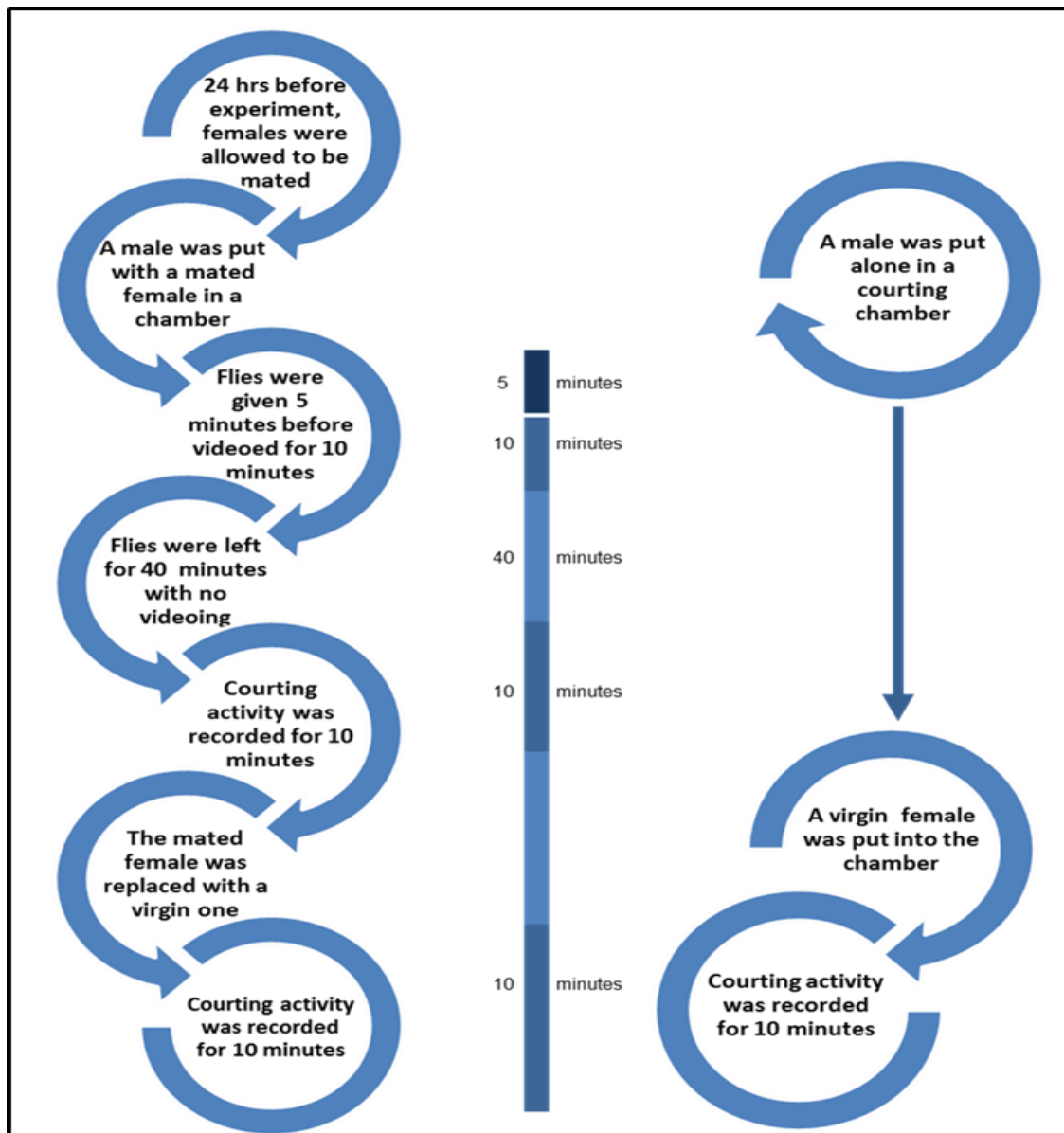
Figure 3.30. A timeline of the experimental procedure and video sample points in the courtship suppression assay. The assay was carried out to measure the courting activity of CG10632 RNAi males on mated flies and compare it with the one of wild type (W*) males. A sham-trained female of either CG10632 RNAi or wild type (W*) flies was used to compare the courting activity of the males towards virgin females that were put instead of the mated ones.

Two indices were calculated based on this assay: the learning index (Figure 3.31) and the memory index (Figure 3.32). The learning index (LI) is the immediate memory of the flies and it was calculated using the equation below. On the other hand, the memory index (Mi) is the one-hour memory that was calculated using the same courting measures but after 60-minutes gap from the training time.

$$Li = \frac{A}{B}$$

A is the percentage of courting over 10 minutes for trained males

B is the percentage of courting over 10 minutes for sham males

The average Li in the courtship suppression assay is shown in Table 3.15 and Figure 3.31 whereas Table 3.16 and Figure 3.32 illustrate the Mi results.

	CG10632 RNAi Line x Gal4 Actin	Wild Type (W*) x Gal4 Actin
Average	0.169472271	0.178209024
St.dev	0.155299082	0.175962431
S.E.M	0.058697536	0.058654144

Table 3.15. The average of learning index (Li) in the courtship suppression assay.
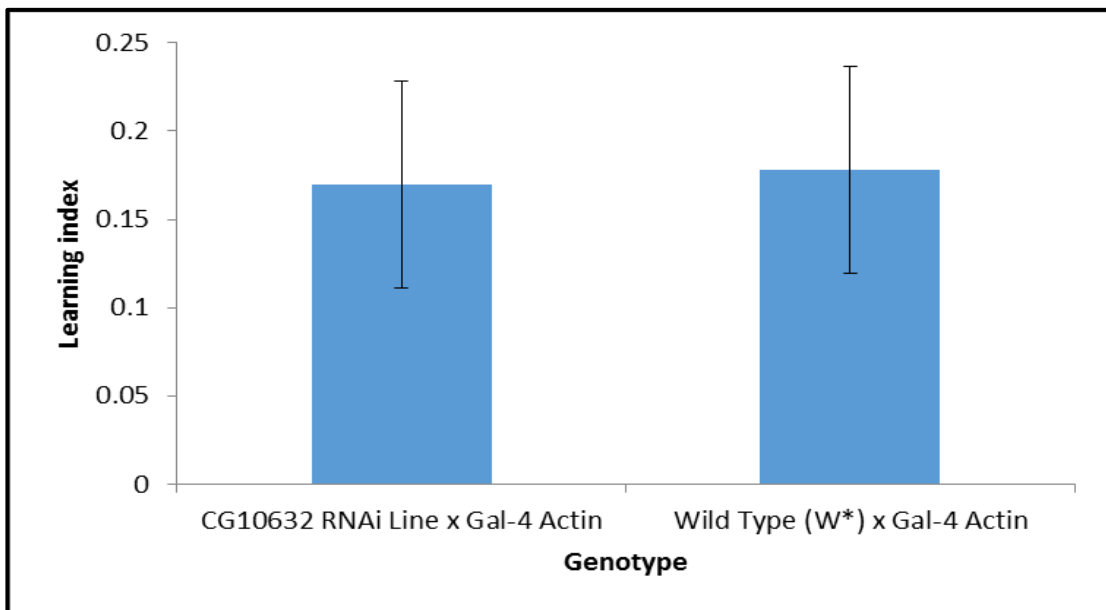
Figure 3.31. Learning index in the courtship suppression assay. This is an indication of the immediate memory of the tested flies. The index is calculated based on the mean time that males spent on courting mated females during the last 10 minutes in comparison with the first 10 minutes of the experiment. Both CG10362 knock down and wild type flies had a similar learning index during the courtship suppression assay.

	CG10632 RNAi Line x Gal4 Actin	Wild Type (W*) x Gal4 Actin
Average	1.004327506	0.90535309
St.dev	0.258936486	0.230636222
S.E.M	0.097868793	0.076878741

Table 3.16. The average of memory index (Li) in the courtship suppression assay.
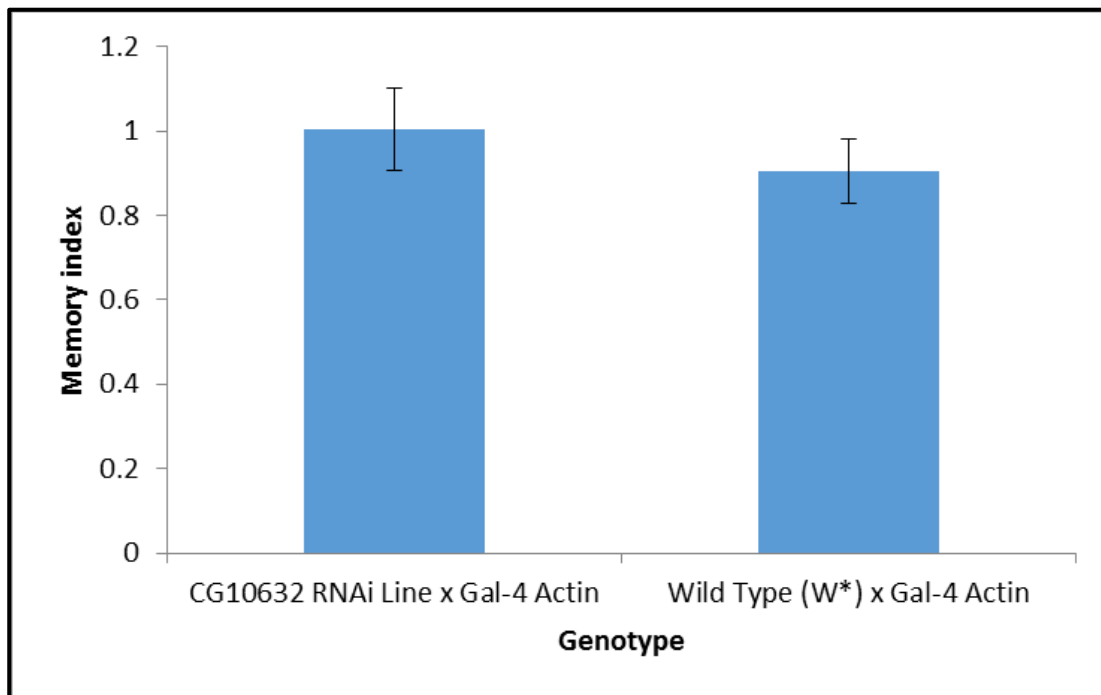
Figure 3.32. Memory index in the courtship suppression assay. This is an indication of the one-hour memory of the tested flies. The index is calculated based on the mean time that experienced males spent in courting virgin females in comparison with the time spent by naive males in courting virgin females. Both CG10362 knock down and wild type flies had a similar memory index during the courtship suppression assay.

In comparison with the courtship index of sham-trained males, who were placed individually in empty chambers, the courtship behaviour of both the CG10362 knock down males and the wild type males was not supressed after one-hour of exposure to mated females (Figure 3.32). The fact that the wild type males have not shown courtship suppression after one hour exposure to mated females rise a question about the validity of the used wild type. This issue and the interference caused by using white-eyed Drosophila in the outcome of courtship suppression assay is discussed in Section 3.3.3. On the other hand, the unsuppressed

courting activity in the CG10362 knock down males, does support the importance of the *PDZD8* in cognitive and memory activities.

3.2.2 Intellectual disabilities families with unidentified causes

3.2.2.1 Clinical features and genetic analysis of Family-4

Family-4 originates from a first-cousin consanguineous marriage (Figure 3.33). The father (case III.1) died of prostate cancer aged 65 years and the mother was reported as healthy. The couple had ten children; two of them (IV.2 and IV.3) died in a road traffic accident. Five of the children were diagnosed with intellectual disability: three (IV.6, IV.8 & IV.10) had moderate level of ID, and the other two (IV.1 & IV.9) had a mild form of ID. Case IV.6 was referred to the Genetic and Developmental Medicine Clinic for speech delay and dysmorphic features. The child was born with normal growth parameters and no history of delivery or neonatal complications. Family history reported that the other four siblings had learning problems. Examination of case IV.6, at the age of 6 years, revealed dysmorphic features including hypertelorism, mild upslanting of palpebral fissures, epicanthal folds, broad nasal tip, brachydactyly of fingers and toes. Brain MRI was reported to be normal.

Microsatellite genotyping was performed to verify the pedigree structure of the family (Figure 3.33) using the methods described in section 2.7. Homozygosity mapping of Family-4 was performed on Affymetrix SNP array data and WES data generated as in section 2.10.1. The data generated were plotted using the Agile MultiIdeogram programme. Results are based on the affected subjects; IV.6, IV.9 and IV.10. These three shared three homozygous regions more than 1Mb in size (Figure 3.34).
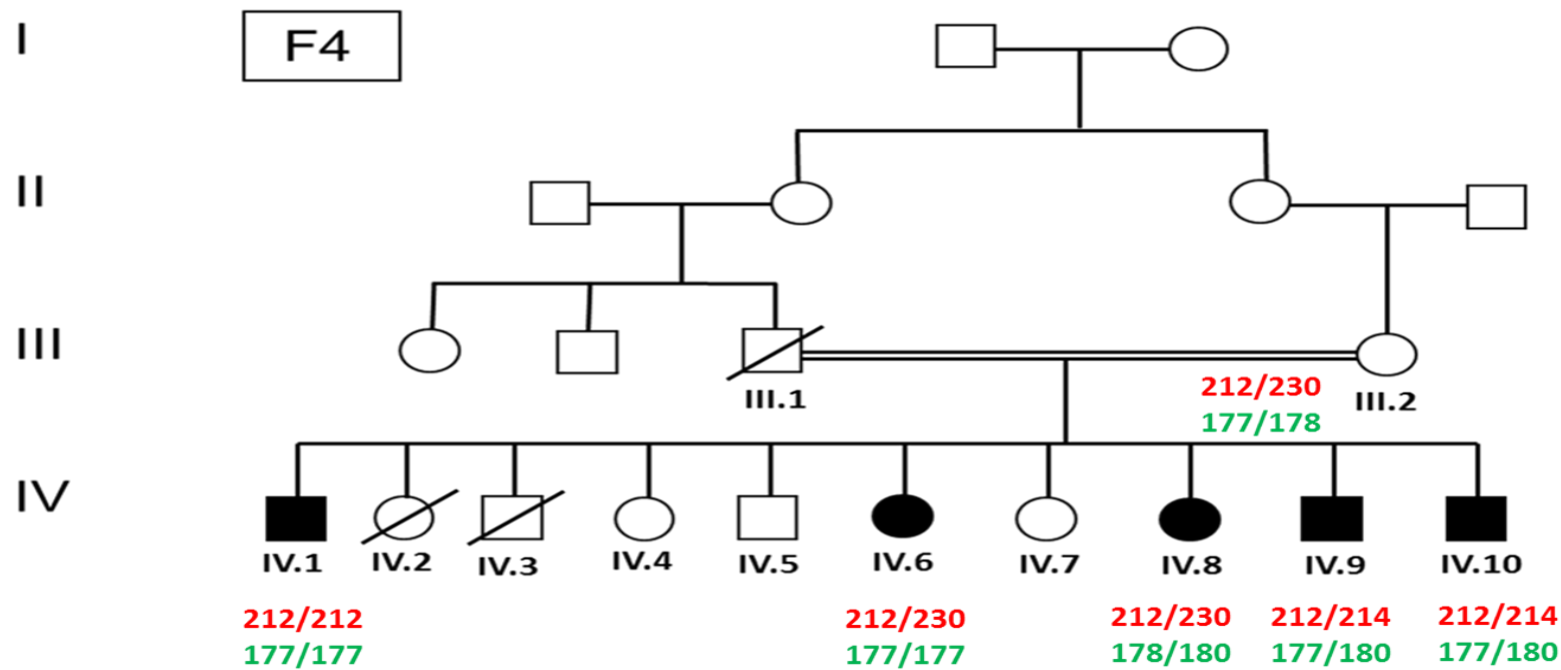
Figure 3.33. Family-4 pedigree with microsatellite genotyping. A consanguineous family from a first-cousin marriage. There are five ID affected individuals: IV.1, 1V.6, IV.8, IV.9 and IV.10. The pattern suggests a recessive inheritance. Genotyping is shown for the cases from whom DNA was available. Red text indicates the alleles for marker D4S3042 and green text indicates the alleles for D7S483. All the affected children (IV.1, IV.6, IV8, IV.9 and IV.10) have shown a genotype pattern that is consistent with being inherited from their parents (III.1 and III.2). WES was done for case IV.6 and SNP array for cases IV.9 and IV.10.

158
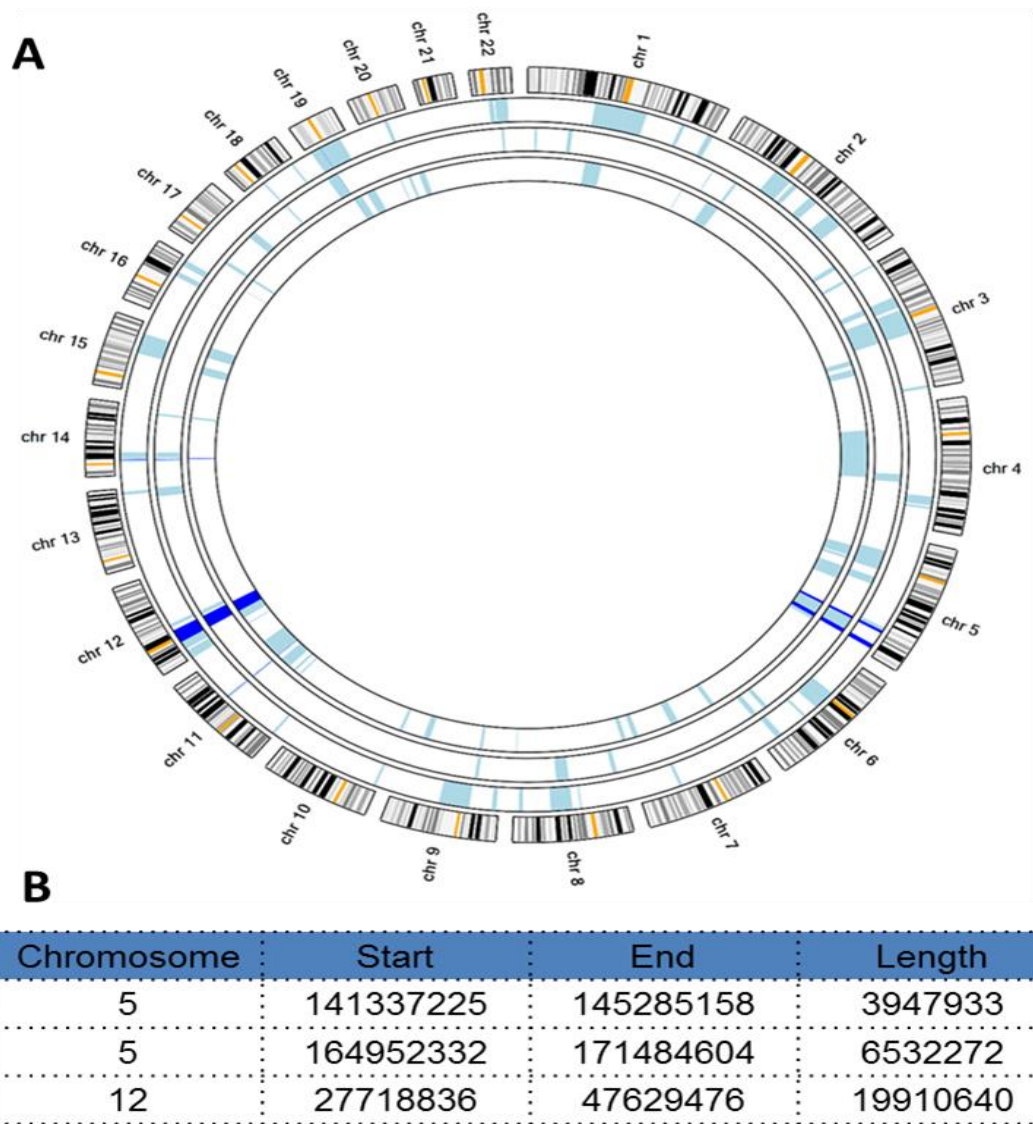
Figure 3.34. Shared homozygous regions in Family-4. A. An ideogram where the outer circle contains template segments of the 22 autosomal chromosomes. The three inner circles are as follows: the SNP data from subjects IV.9 and IV.10 and WES data from subject IV.6. The light blue areas are the homozygous regions within each individual and the dark blue regions (number=3) are the shared homozygous regions between individuals. B. Details of the shared homozygous regions. The coordinates for the regions are based on human genome reference hg19. Two of the shared regions are on chromosome 5 and there is one bigger region on chromosome 12.

Exome sequencing was carried out on case IV.6. After annotation by ANNOVAR and exome analysis, 57996 variants were identified, of which 293 variants were mapped in the homozygous regions shared by the affected members of this family. Only one homozygous variant in the *GPR151* gene on chromosome 5 remained after filtering according to the strategy described in (Section 2.11.9.4,

Table 3.17). This variant did not segregate with the disease in additional family members.

The possibility of compound heterozygous variants in the rest of the genome underlying the ID in Family-4 was also investigated. 393 genes which had more than one variant were selected, but only 9 variants in four genes (Table 3.17) passed the filtering process according to methods described in Section 2.11.9.5. When these were tested in additional family members none of them fully segregated with the disease phenotype.

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	1000g	CADD	Zygosity
chr5	145895285	*GPR151*	exonic	GPR151:NM_194251:exon1:c.390_392del	p.130_131del	0	0.897	Homo
chr11	6411936	*SMPD1*	exonic	SMPD1:NM_000543:exon1:c.108_113del	p.36_38del	0	12.79	Het
chr11	6415259	*SMPD1*	exonic	SMPD1:NM_000543:exon5:c.G1474A	p.G492S	0.0023	9.451	Het
chr12	4735968	*AKAP3*	exonic	AKAP3:NM_006422:exon5:c.T2100G	p.D700E	0	0.001	Het
chr12	4735969	*AKAP3*	exonic	AKAP3:NM_006422:exon5:c.A2099T	p.D700V	0	22	Het
chr12	4736150	*AKAP3*	exonic	AKAP3:NM_006422:exon5:c.T1918G	p.X640E	0	0.001	Het
chr12	58017860	*SLC26A10*	exonic	SLC26A10:NM_133489:exon9:c.G1206A	p.W402X	0.0018	39	Het
chr12	58018668	*SLC26A10*	exonic	SLC26A10:NM_133489:exon10:c.T1247G	p.L416R	0.0032	27.5	Het
chr17	79094162	*AATK*	exonic	AATK:NM_004920:exon10:c.C3265T	p.P1089S	0.0046	0.072	Het
chr17	79094694	*AATK*	exonic	AATK:NM_004920:exon10:c.2731_2733del	p.911_911del	0	8.607	Het

Table 3.17. Final list of homozygous and compound heterozygous candidate variants for Family-4. This table shows one homozygous and 9 heterozygous variants, each with its pathogenicity score based on CADD and frequency based on the 1000G database. Homo= homozygous; Het= heterozygous.

ExomeDepth is a piece of software that searches in genomic ngs data (WES or WGS) for CNVs by using the read depth data from exome sequencing experiments. It works by comparing the exome depth of a given exome in the sample of interest against a set of control samples that have been loaded in the same sequencing run. In this analysis, the control sets comprised 5 samples. The analysis was performed on the exome sequencing results of case IV.6 using the method of Plagnol et al. (2012). The obtained CNV calls were prioritised based on Bayes factor (BF), as described in section 2.11.9.7. This identified a final list of seven putative homozygous CNVs (Table 3.18) that could be involved in the pathogenicity of the ID in Family-4. All the highlighted variants were tested for validation by PCR and IGV (Integrative Genomics Viewer). They were either found to be artefacts or did not segregate within the family. A representative example of the variant validation was a deletion on chromosome 3 that maps within a gene called *EFCAB12*. The deletion appears to remove the full gene. Primers were designed to amplify 356 bp of exon 7 of *EFCAB12* (Appendix 2). All members showed a normal PCR product amplified, like the +ve control sample. The fact that a band was amplified from DNA from all family members, including those affected by ID, suggests this deletion might not be genuine (Figure 3.35).

chr	start	end	Exp Read	Obs Read	Type	BF
chr3	129120437	129127701	163	2	deletion	36.9
chr22	24376424	24384231	228	402	duplication	23.5
chr1	113126595	113126748	68	1	deletion	14.6
chr5	147553798	147553933	66	0	deletion	14.5
chr1	152573209	152586574	43	0	deletion	10.3
chr4	69433480	69434202	25	0	deletion	6.14
chr20	1569207	1569278	18	0	deletion	4.52

Table 3.18. CNVs in case IV.6 of Family-4. This is the final list of seven putative homozygous CNVs that was generated with the programme ExomeDepth after comparing the reads of case IV.6 with a reference set of samples that were run in the same batch on the Illumina 2500 high seq. 'Exp' is the expected reads and 'Obs' is the observed reads. The candidates were prioritised based on Bayes factor (BF).
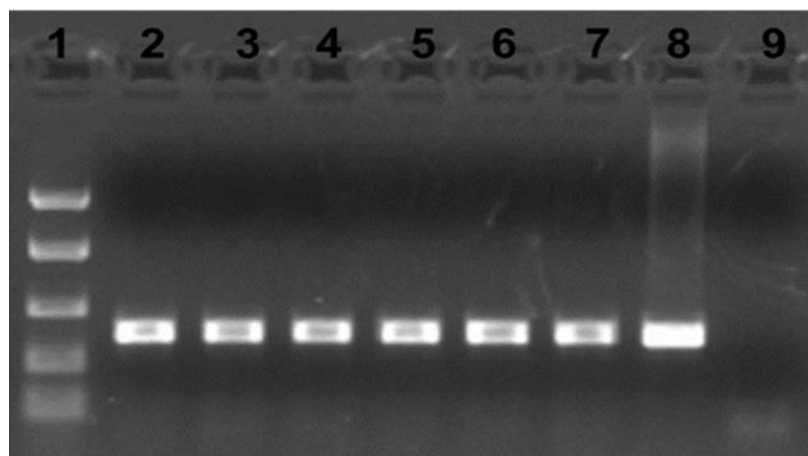
Figure 3.35. Investigation of *EFCAB12* deletion in Family-4. In order to validate the first candidate of the CNVs, DNA samples from family members were tested by PCR for the presence of exon 7 of *EFCAB12*. Lane 1, 100-bp ladder; lane 2, III.2; lane 3, IV.6; lane 4, IV.7; lane 5, IV.8; lane 6, IV.9; lane 7, IV.10; lane 8, +ve control; lane 9, -ve control

3.2.2.2 Clinical features and genetic analysis of family-5.

Family-5 (Figure 3.36) originated from a first-cousin consanguineous marriage. Both parents were reported to be healthy. Two of the children (IV.2 and IV.3) had been diagnosed with moderate ID, ASD and microcephaly. The pedigree structure was verified by microsatellite genotyping (Figure 3.36). Homozygosity mapping was carried out using Affymetrix SNP array data (Section 2.10) generated from genomic DNA in individual IV-3, and from SNP genotypes extracted from WES (section 2.11) from subject IV.2 using the programme Agile MultiIdeogram programme. Five shared homozygous regions were identified (Figure 3.37). Combining the exome sequence in case IV.2 with the homozygosity mapping data identified 538 variants in the shared homozygous regions. Three candidate variants remained after filtering as described in section 2.11.9.4, shown in Table 3.19. However, none showed complete segregation.

Therefore, a further search was undertaken to find possible compound heterozygous variants elsewhere in the exome that could underlie the described phenotypes. 547 variants as possible compound heterozygotes were detected in case IV.2. Only 17 variants in 8 genes passed the filtering process (Table 3.19). However, none of these variants showed complete segregation within the family. The fact that both affected individuals in this family were male (cases IV.2 and

IV.3 in Figure 3.36) suggested possibly an X-linked mode of inheritance for the disease. Therefore, variants located on the X-chromosome were also analysed and 14 variants were highlighted after the filtering and prioritization step (Section 2.11.95). Three of them were in known genes for ID (the full list of known ID genes is shown in Appendix 8). However, all the X-chromosome candidates for this family (Table 3.20) were checked for segregation with the ID in this family and none of them was found to show a complete pattern of co-segregation.

Exome depth, performed on case IV.2, identified 4 putative homozygous CNVs that could be potentially involved in the ID pathogenicity (Table 3.21). PCR, Sanger sequencing and examination on IGV were used as validation methods. None of the tested candidates was consistent with being causative. A representative example using the IGV platform to examine a duplication variant on chromosome 3 in case IV.2 compared with the control is shown in Figure 3.38. The results showed an increased read pattern in the analysed sample.
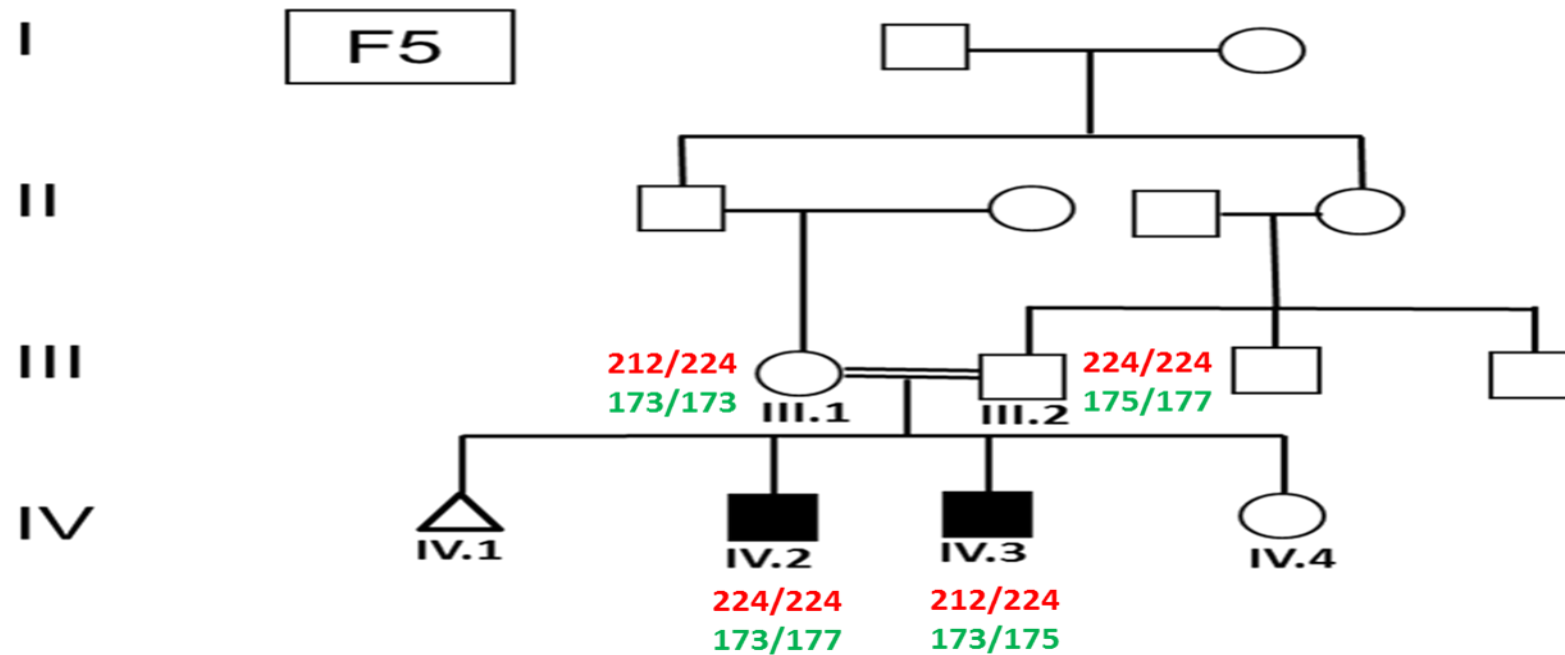
Figure 3.36. Family-5 pedigree with microsatellite genotyping. A consanguineous family originating from a first-cousin marriage. The small triangle indicates a miscarriage. The parents are unaffected but two of the children (IV.2 and IV.3) have ID and this pattern suggests recessive inheritance. Genotyping is shown for the cases from whom DNA was available. Red text indicates the alleles for marker D4S3042 and green text indicates the alleles for D7S483. Both affected cases (IV.2 and IV.3) showed a genotype pattern consistent with being inherited from their parents (III.1 and III.2). Case IV.2 was exome sequenced and case IV.3 SNP genotyped using an Affymetrix SNP chip.

A

B

Chr	Start	End	Length (bp)
5	81572184	95728974	14156790
6	27792221	33768897	5976676
8	139988876	141566311	1577435
11	198510	3242933	3044423
16	59208069	61453609	2245540
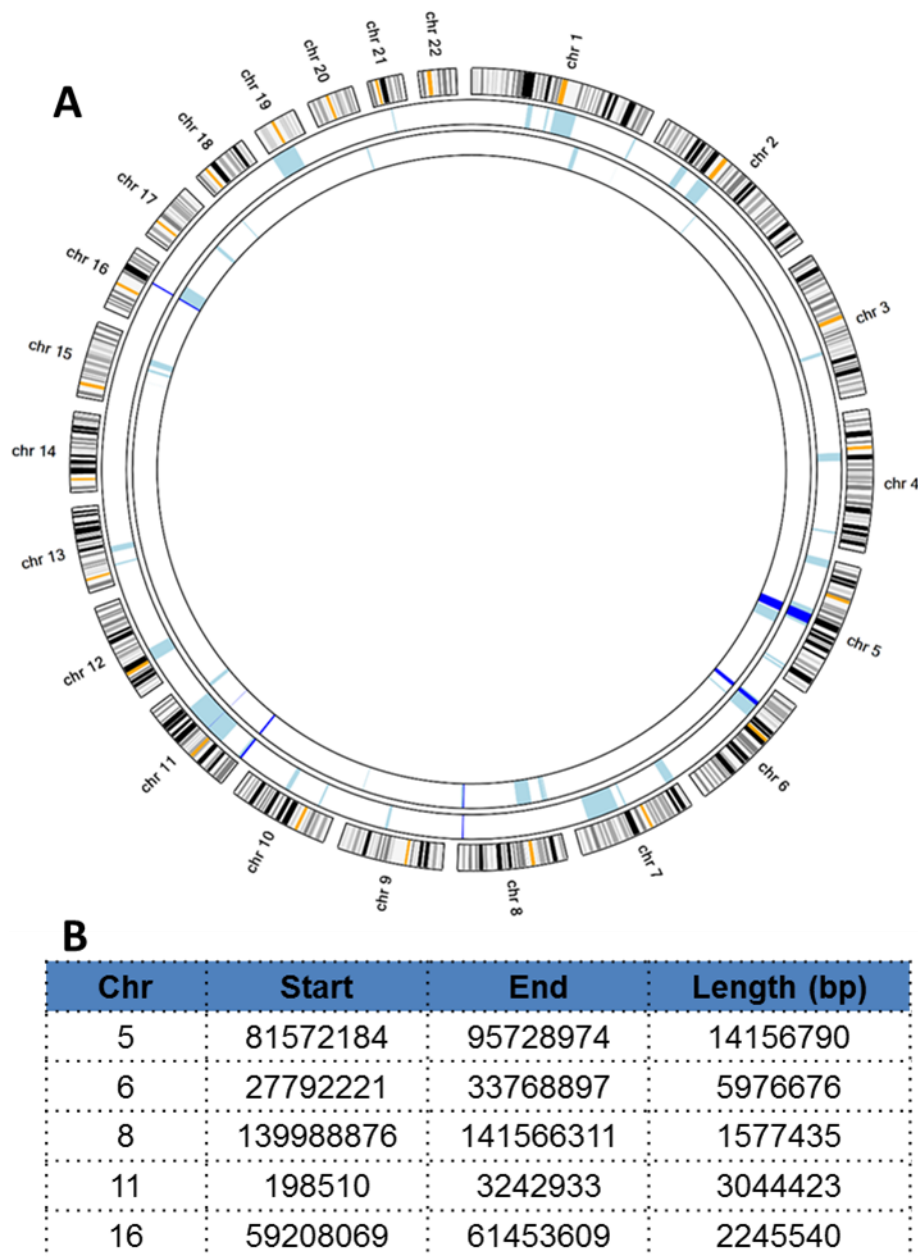
Figure 3.37. Shared homozygous regions in Family-5. A. An ideogram in which the outer circle contains template segments of the 22 autosomal chromosomes. The middle circle is SNP data from subject IV.3 and the inner circle is WES data from subject IV.2. The light blue regions are the homozygous regions within each individual and the dark blue regions (number=5) are shared between the affected individuals. B. Details of the shared homozygous regions. The coordinates were based on human genome reference hg19. The five-shared regions are on different chromosomes and have different lengths. The largest region is on chromosome 5.

Chr	Start	Gene	Coding Effect	cDNA Change	Protein Change	1000g	CADD	Zygosity
chr5	94749787	*FAM81B*	exonic	FAM81B:NM_152548:exon4:c.C430T	p.Q144X	0.02	37	Homo
chr11	405937	*SIGIRR*	exonic	SIGIRR:NM_001135053:exon10:c.C1192G	p.R398G	0	32	Homo
chr11	1298430	*TOLLIP*	exonic	TOLLIP:NM_019009:exon6:c.G664T	p.A222S	0.02	13.98	Homo
chr6	35086272	*TCP11*	Exonic	TCP11:NM_001261818:exon9:c.G1187A	p.R396Q	0	32	Het
chr6	35086275	*TCP11*	Exonic	TCP11:NM_001261818:exon9:c.A1184G	p.Q395R	0	23.8	Het
chr7	100345978	*ZAN*	Exonic	UNKNOWN	?	0	8.718	Het
chr7	100385593	*ZAN*	Exonic	UNKNOWN	?	0	12.85	Het
chr7	100385595	*ZAN*	Exonic	UNKNOWN	?	0	22.5	Het
chr9	140772669	*CACNA1B*	Exonic	CACNA1B:NM_000718:exon1:c.284_285insTCCATTCGAGT	p.P95fs	0	32	Het
chr9	140777306	*CACNA1B*	Exonic	CACNA1B:NM_000718:exon3:c.C501G	p.N167K	0	14.39	Het
chr11	62400108	*GANAB*	Exonic	GANAB:NM_198334:exon9:c.A925T	p.T309S	0	34	Het
chr11	62400570	*GANAB*	Exonic	GANAB:NM_198334:exon8:c.G722T	p.R241M	0	27.3	Het
chr13	24798712	*SPATA13*	Exonic	SPATA13:NM_001166271:exon2:c.A1645G	p.K549E	0	0.005	Het
chr13	24871664	*SPATA13*	Exonic	SPATA13:NM_153023:exon10:c.G1499A	p.G500D	0	27.8	Het
chr15	31521506	*LOC283710*	Exonic	LOC283710:NM_001243538:exon2:c.75_76del	p.25_26del	0	1.944	Het
chr15	31521507	*LOC283710*	Exonic	LOC283710:NM_001243538:exon2:c.75delC	p.P25fs	0	2.124	Het
chr17	39346593	*KRTAP9-1*	Exonic	KRTAP9-1:NM_001190460:exon1:c.455_457del	p.152_153del	0	10.64	Het
chr17	39346622	*KRTAP9-1*	Exonic	KRTAP9-1:NM_001190460:exon1:c.T484A	p.C162S	0	0.623	Het
chr17	67145191	*ABCA10*	Exonic	ABCA10:NM_080282:exon39:c.4515_4516del	p.1505_1506del	0	35	Het
chr17	67149973	*ABCA10*	Exonic	ABCA10:NM_080282:exon33:c.C3964T	p.R1322X	0	28	Het

Table 3.19. Final list of homozygous and heterozygous candidate variants for Family-5. This shows 3 homozygous and 17 heterozygous variants, each with its pathogenicity score based on CADD, and frequency based on the 1000G database.

Chr	Position	Gene	cDNA Change	Protein Change	1000g	CADD
chrX	18917338	*PHKA2*	PHKA2:NM_000292:exon29:c.T3064A	p.S1022T	0	9.778
chrX	26157773	*MAGEB18*	MAGEB18:NM_173699:exon2:c.A671G	p.N224S	0	15.92
chrX	31645821	*DMD*	DMD:NM_004013:exon12:c.A806G	p.K269R	0	21.4
chrX	32404432	*DMD*	DMD:NM_004011:exon5:c.G646A	p.A216T	0	28.9
chrX	35820384	*MAGEB16*	MAGEB16:NM_001099921:exon2:c.G71A	p.S24N	0	0.033
chrX	44088950	*EFHC2*	EFHC2:NM_025184:exon11:c.G1696A	p.V566I	0.0006	0.005
chrX	50121146	*DGKK*	UNKNOWN	?	0	13.01
chrX	53263460	*IQSEC2*	IQSEC2:NM_001111125:exon15:c.G4408A	p.A1470T	0.01	14.64
chrX	69478801	*P2RY4*	P2RY4:NM_002565:exon1:c.G674C	p.R225P	0	25
chrX	70330790	*IL2RG*	IL2RG:NM_000206:exon2:c.A226G	p.S76G	0	13.87
chrX	74494304	*UPRT*	UPRT:NM_145052:exon1:c.C215T	p.S72F	0	14
chrX	108868163	*KCNE1L*	KCNE1L:NM_012282:exon1:c.G87T	p.L29F	0	22.9
chrX	114425217	*RBMXL3*	RBMXL3:NM_001145346:exon1:c.A1213C	p.N405H	0	14.22
chrX	153762634	*G6PD*	G6PD:NM_000402:exon6:c.C653T	p.S218F	0	24.2

Table 3.20. X-Linked variants in Family-5. This table shows 14 variants, each with its pathogenicity score based on CADD, and frequency based on the 1000G database. The three variants highlighted in green are in two genes known to be associated with X-linked ID.

Chr	start	end	Exp Read	Obs Read	Type	BF
chr22	24376424	24384231	173	0	deletion	40.2
chr3	129120437	129127701	124	261	duplication	19.8
chr1	113126595	113126748	51	108	duplication	7.77
chr6	147635401	147635467	46	17	deletion	5.75

Table 3.21. Candidate CNVs in Family-5. This list was generated after comparing the reads of case IV.2 with a reference set of samples that were run in the same batch on the Illumina 2500 high seq. 'Exp' is the expected reads and 'Obs' is the observed reads. The candidates were prioritised based on Bayes factor (BF).
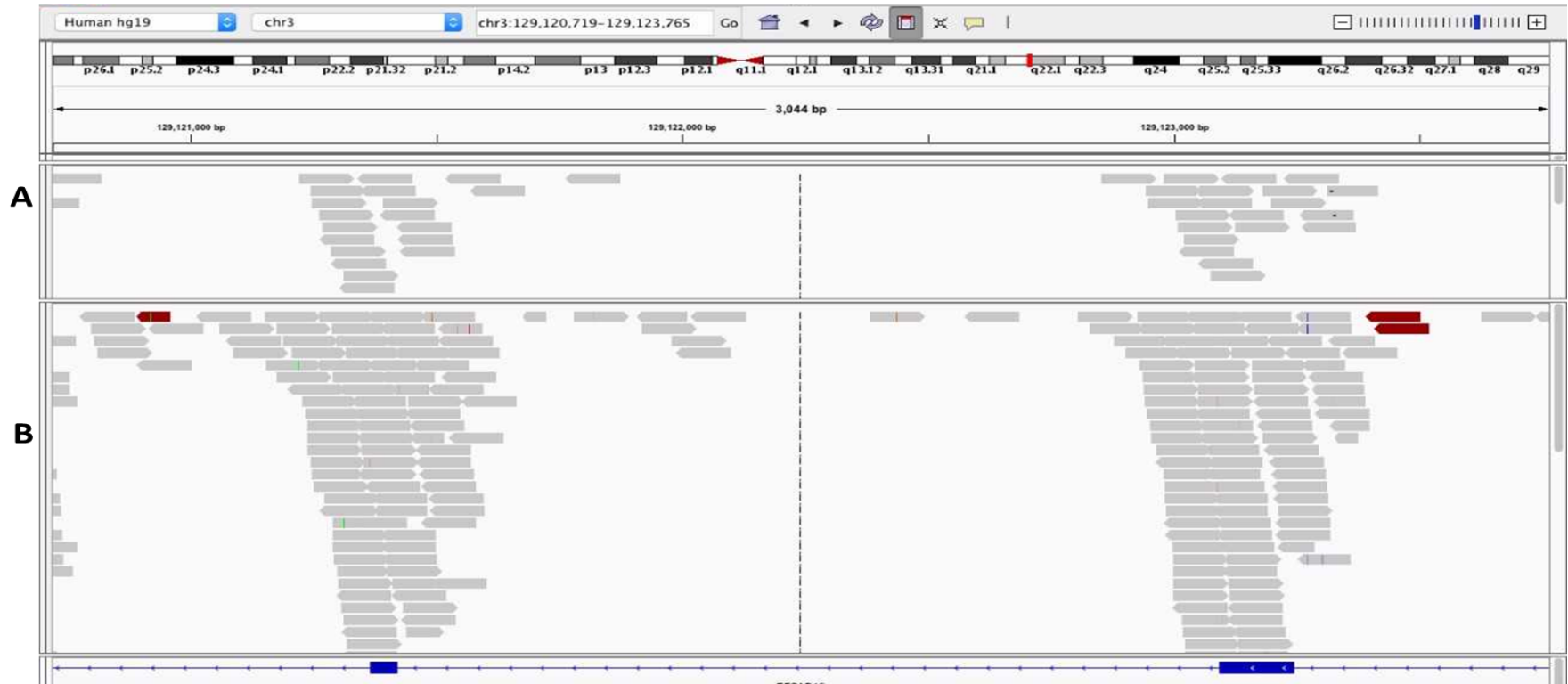
Figure 3.38. Visualisation of chr3 duplication in Family-5. These representations, made by IGV software, cover part of chromosome 3 (129120719-129123765) within which the duplication was expected to occur in ID-affected individuals in Family-5. A. Sequence reads for a control sample that was tested in the same sequencing run as IV.2. B. The sequence of case IV.2 in Family-5. This is an example indicating the duplication CNVs found in the analysed sample but this candidate did not segregate in Family-5 .

3.2.2.3 Clinical features and genetic analysis of Family-6

Family-6 originated from a first-cousin consanguineous marriage (Figure 3.39). There are six children in this family, two (both girls; V.2 & V.5) with severe ID and microcephaly. Genotyping (Section 2.7) was done as a confirmatory step to indicate that no mixing between the samples had occurred. Members of Family-6 were genotyped using two markers (D4S3042 and D7S483) and their results are illustrated in Figure 3.39. Homozygosity mapping of Family-6 was based on subjects V.2 and V.5. Thirteen shared homozygous regions were identified (Figure 3.40 and Table 3.22).

For this family, case V.2 was exome sequenced. Combining SNP data extracted from the exome sequence of case V.2 with the Affymetrix SNP chip data of the case V.5 in homozygosity mapping resulted in thirteen shared regions, within which 10,275 variants were identified in the WES data from case V.2. Only six candidate variants remained (Table 3.23) after carrying out the filtering steps described in Section 2.11.9.4. However, none of the six candidate variants showed complete segregation with ID in this family.

Searching the entire exome for compound heterozygous variants that could possibly underlie the described phenotypes showed 283 variants in case V.2, and 8 variants in four genes were taken forward for segregation testing after they passed the filtering criteria detailed in section 2.11.9.4. Upon testing, none of the proposed compound heterozygous variants showed segregation with ID within the family.

Exome depth was performed as described in section 2.11.9.7 on the exome results of case V.2, based on the method of Plagnol et al. (2012). This procedure identified 5 putative homozygous CNVs that could be involved in the pathogenicity of ID in Family-6 (Table 3.24). All of the proposed CNVs were investigated further using PCR and Sanger sequencing as well as viewing the sequence by eye using IGV software. They were either found to be artefacts or did not segregated within the family. The first ranked homozygous CNV was tested for validation of a representative example and is shown in Figure 3.41. The

putative deletion was found to delete the *PCDHA7* gene. Primers were therefore designed to amplify exon 1 of this gene. The fact that the PCR amplified a normal band from DNA of affected individuals suggests the deletion might not be genuine.
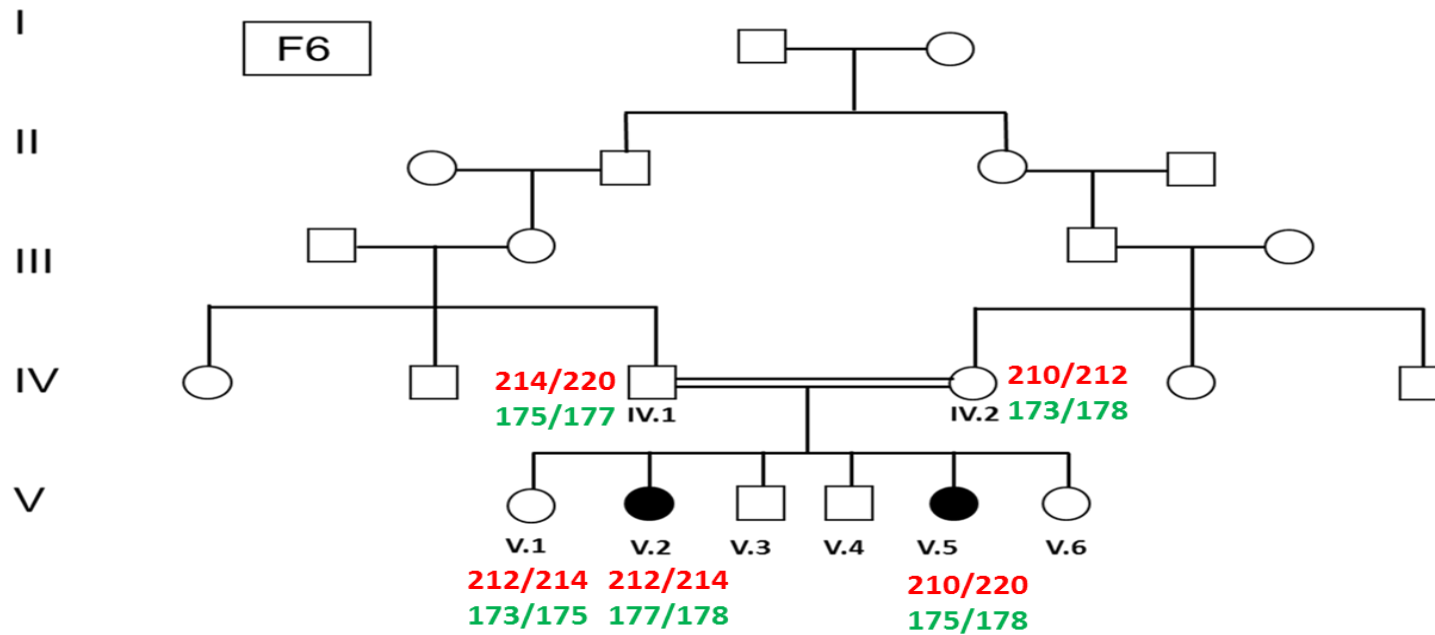
Figure 3.39. Family-6 pedigree: a consanguineous family from a first-cousin marriage. There are six children in this family, of which two (both girls; V.2 & V.5) have severe ID and microcephaly. The pattern suggests recessive inheritance. Genotyping is shown for the cases from whom DNA was available. Red text indicates the detected alleles for marker D4S3042 and green text indicates the alleles for D7S483. All the children including both affected cases (V.2 and V.5) show a genotype pattern consistent with being inherited from their parents (IV.1 and IV.2).

173
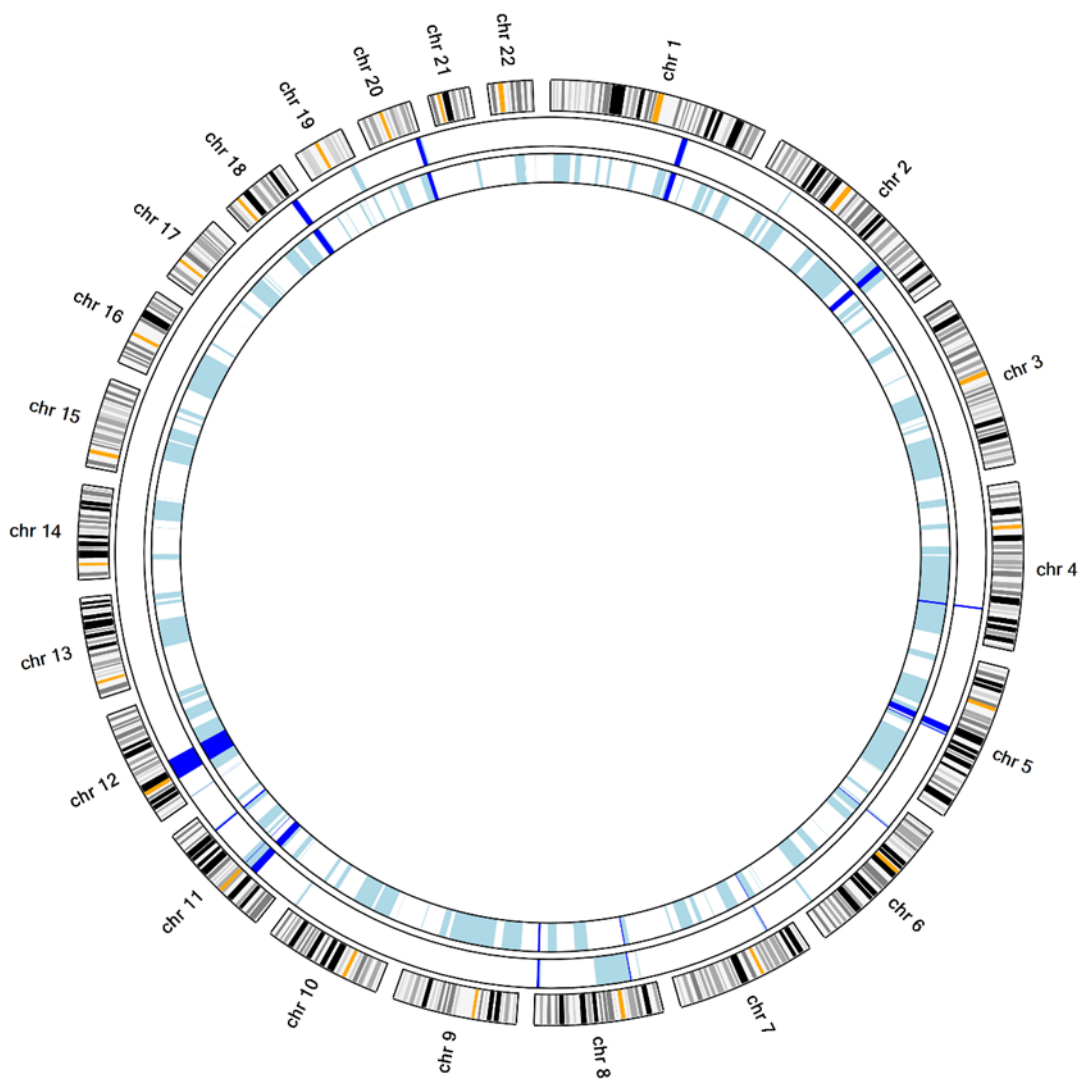
Figure 3.40. Shared homozygous regions in Family-6. An ideogram in which the outer circle contains template segments of the 22 autosomal chromosomes. The middle circle is SNP data from subject V.5 and the inner circle is SNP genotypes extracted from WES data from subject V.2. The light blue areas are homozygous regions within each individual and the dark blue regions (number=13) are the homozygous regions shared between the affected individuals. Chromosome 12 had the largest shared homozygous region.

Chr	Start	End	Length (bp)
1	162504246	169512039	7007793
2	179600563	187487160	7886597
4	146825552	149155928	2330376
5	85682624	93731985	8049361
5	96244638	97442526	1197888
6	28269663	29312762	1043099
8	28452799	29927321	1474522
8	142146708	144779718	2633010
11	37131957	46744925	9612968
11	107895836	110035301	2139465
12	31139966	56928777	25788811
18	62537601	70517206	7979605
20	55436718	60712511	5275793

Table 3.22. Shared homozygous regions in Family-6. Coordinates of the shared homozygous regions between cases V.2 and V.5, based on human genome reference hg19. There are 13-shared regions and the largest region is on chromosome 12.

Chr	Position	Gene	Coding Effect	cDNA Change	Protein Change	1000g	CADD	Zygosity
chr11	46342259	*CREB3L1*	splicing	CREB3L1:NM_052854:exon12:c.1524-1->G	?	0	34	Homo
chr12	48167179	*SLC48A1*	exonic	SLC48A1:NM_017842:exon1:c.T92A	p.V31E	0	23	Homo
chr12	49373410	*WNT1*	exonic	WNT1:NM_005430:exon2:c.T264A	p.S88R	0.01	13.78	Homo
chr12	51678502	*BIN2*	exonic	BIN2:NM_016293:exon12:c.G1666C	p.E556Q	0	24.7	Homo
chr12	51759296	*GALNT6*	exonic	GALNT6:NM_007210:exon5:c.G732T	p.E244D	0	19.07	Homo
chr12	53009992	*KRT73*	exonic	KRT73:NM_175068:exon2:c.A620G	p.E207G	0	23	Homo
chr5	112868579	*YTHDC2*	exonic	YTHDC2:NM_022828:exon5:c.C679T	p.P227S	0	29.5	Het
chr5	112903526	*YTHDC2*	exonic	YTHDC2:NM_022828:exon23:c.A3224G	p.Q1075R	0	17.34	Het
chr11	48286231	*OR4X1*	Stopgain	OR4X1:NM_001004726:exon1:c.T819A	p.Y273X	0.66	35	Het
chr11	48286256	*OR4X1*	exonic	OR4X1:NM_001004726:exon1:c.C844T	p.P282S	0.46	24.1	Het
chr13	25670868	*PABPC3*	exonic	PABPC3:NM_030979:exon1:c.G532A	p.E178K	0	32.2	Het
chr13	25670877	*PABPC3*	exonic	PABPC3:NM_030979:exon1:c.G541A:	p.A181T	0	14.99	Het
chr19	56515447	*NLRP5*	exonic	NLRP5:NM_153447:exon2:c.G428A	p.R143Q	0	23.5	Het
chr19	56544970	*NLRP5*	exonic	NLRP5:NM_153447:exon9:c.G2510A	p.R837H	0.0005	9.27	Het

Table 3.23. Homozygous and heterozygous candidate variants for Family-6. This table shows the six homozygous variants, and the eight heterozygous variants in 4 genes, each with its pathogenicity score based on CADD and frequency based on the 1000G database.

Chr	start	end	Exp Read	Obs Read	type	BF
5	140228082	140238021	209	5	deletion	17
19	1978442	1987252	182	305	duplication	8.36
11	65413928	65422471	191	321	duplication	8.22
19	55603019	55606166	127	223	duplication	7.74
1	43917320	43919464	109	205	duplication	6.4

Table 3.24. CNVs in Family-6. The final list of putative CNVs predicted by the ExomeDepth software after comparing the reads of case IV.2 from family-6 with a reference set of samples that were run in the same batch on the Illumina 2500 high seq. 'Exp' is the expected number of reads and 'Obs' is the observed reads. The candidates were prioritised based on Bayes factor (BF).
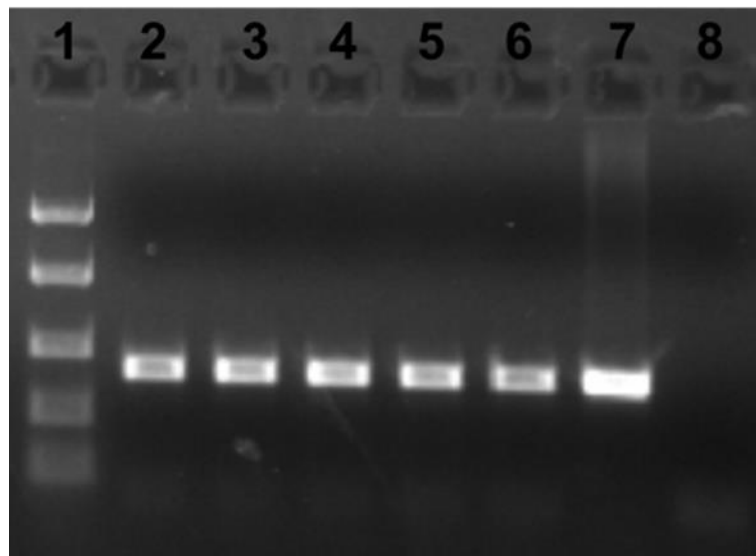
Figure 3.41. Checking for Chr5 deletion in Family-6. This putative deletion predicted by ExomeDepth analysis of WES data from individual IV-2 was found on the genome browser to include a gene called *PCDHA7.* An experiment was carried out to check if it was possible to amplify exon 1 of *PCDHA7* from DNA from the family members. Lane 1, 100 bp ladder; lane 2, IV.1; lane 3, IV.2; lane 4, V.1; lane 5, V.2; lane 6, V.5; lane 7, +ve control; lane 8, -ve control. All members, including the affected cases, showed the normal *PCDHA7* band (394 bp), like the +ve control sample.

3.3 Discussion

3.3.1 Family-1 with *TUSC3* Mutation

Affected individuals from Family-1 presented with developmental, motor and language delay, syndactyly, facial dysmorphology and low IQ. Homozygosity mapping and WES analysis revealed a nonsense mutation in the *TUSC3* gene that cosegregated with the condition in a manner consistent with it being the causative mutation. This mutation would be expected to lead to a truncated protein, but may alternatively lead to nonsense mediated decay, meaning effectively that no protein is produced. Furthermore, previous published reports have shown that other mutations in the same gene have caused similar phenoptypes, detailed below. It was therefore concluded that this mutation is the cause of the condition observed in this family.

TUSC3 is the tumour suppressor candidate-3 gene and is also known as N33 or OST3A (oligosaccharyl transferase 3 homolog A). The gene encodes a 348-amino acid protein comprising a membrane-anchored thioredoxin signal peptide with four transmembrane domains (Figure 3.42).

The TUSC3 protein has been found to participate in the multimeric oligosaccharyl transferase complex (OTC) in the endoplasmic reticulum. The OTC is likely to be involved in post-translational glycosylation of selected proteins essential for normal brain development (Kelleher and Gilmore, 2006; Mohorko et al., 2014). Glycosylation is a sophisticated but very important step for protein modification and without it, a newly synthesised protein would have difficulty with folding and stability (Shental-Bechor and Levy, 2008). Figure 3.43 gives more explanation of the glycosylation process and highlights the step in which TUSC3 is involved.
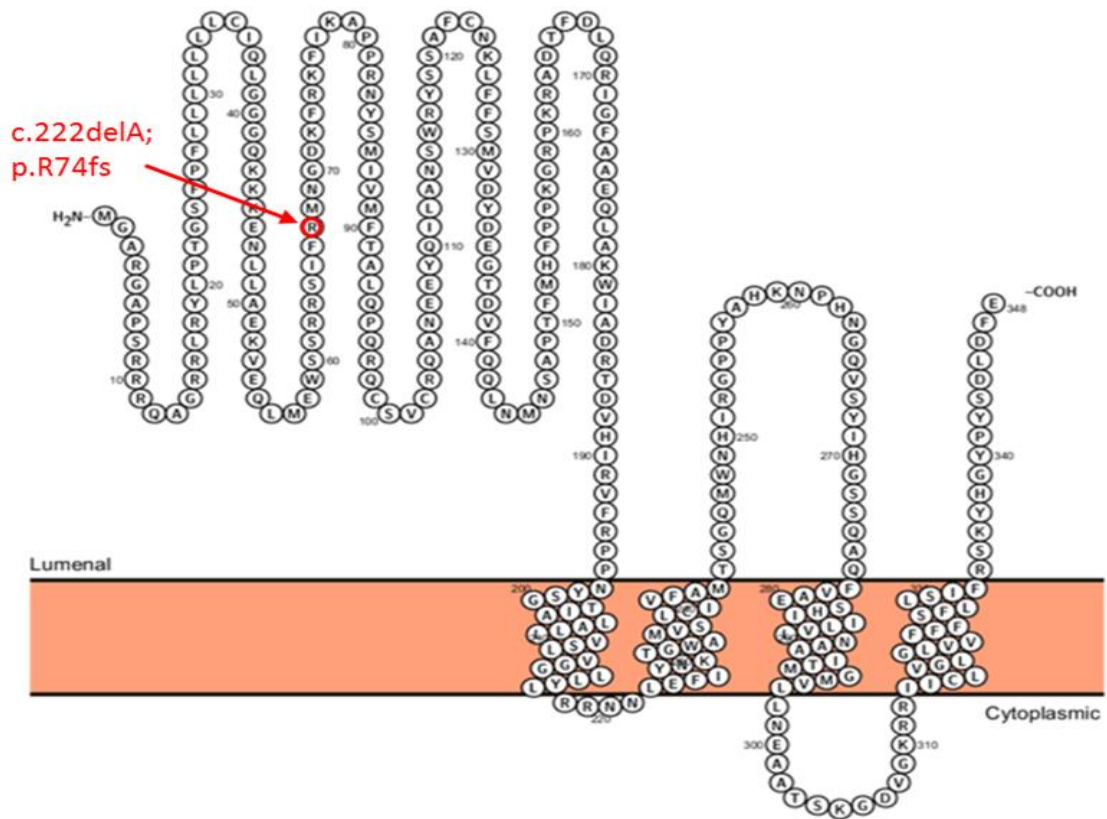
Figure 3.42. Topology prediction of TUSC3 structure. The four transmembrane domains of the protein are shown as clusters of amino acids within the orange-shaded membrane. The amino acid at which the premature truncation occurs due to the c.222delA mutation in Family-1 is indicated with a red arrow. The image was generated using the Protter prediction tool.

It has been suggested that TUSC3 may also play a role as a magnesium transporter (Zhou and Clapham, 2009) but this could be clarified by the fact that TUSC3 could control the efficiency of glycosylation in a membrane transporter. Furthermore, It has previously been found that deletion of the human chromosome 8p22 region, which contains the *TUSC3* gene, is common in cases of prostate (Bergerheim et al., 1991) and ovarian cancers (Pribill et al., 2001). Silencing TUSC3 in ovarian cancer cell lines was also documented to support the proliferation and migration of the cancer cell, which points to the fact that TUSC3 could be a tumour suppressor protein (Vanhara et al., 2013). Because all ID patients in Family-1 are still in their childhood, it is too early to assess the relation of the found variant to tumour development. Although it may be appropriate to monitor these individuals for the presence of tumours in the future, it should also

be noted that the full *TUSC3* deletions described in cancer cells are quite different from the likely premature termination of protein synthesis reported in Family-1. On the other hand, if nonsense mediated decay destroys all truncated mRNAs then this mutation will be an effective knockout just like the situation in tumour cases.

There have been various reports on ID patients, also with speech delay as the prominent features, who have biallelic *TUSC3* mutations (El Chehadeh et al., 2015; Garshasbi Masoud et al., 2008; Garshasbi M. et al., 2011; Khan M. A. et al., 2011; Loddo et al., 2013; Molinari et al., 2008). The affected individuals in Family-1 have symptoms matching the phenotypes described in these reports and the findings are consistent with each other. Moreover, the vast majority of the additional characteristics that have been found in previously reported cases, like short stature, microcephaly and moderate facial dysmorphism, were also present in some of the patients in Family-1 (Table 3.2). In the study of Loddo and colleagues (Loddo et al., 2013), one of the described ID case had congenital malformations in the phalanges and this sign has been observed in three of the four patients in Family-1. Therefore, it seems there is no recognisable genotype-phenotype correlation between the different variants and the additional comorbidity signs (El Chehadeh et al., 2015).
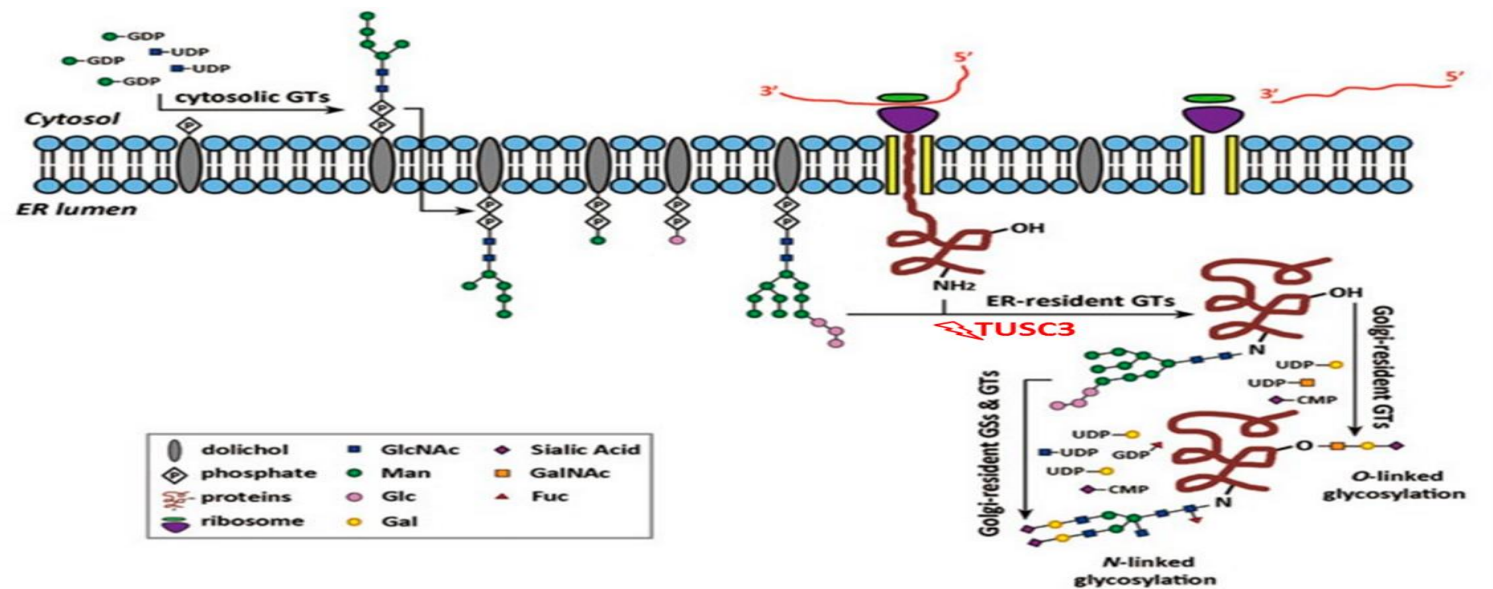
Figure 3.43. The protein glycosylation process and the stage of TUSC3 involvement. The process is started at the cytosolic face of the ER at which mannose-rich oligosaccharide is synthesized. When the newly synthesized protein comes from the ribosomes and starts entering the ER, the developed oligosaccharide chain is transferred from dolichol (a lipid molecule) to the protein and this step is mediated by glycosyltransferases. TUSC3 is involved in the glycosyltransferase complex, facilitating the association of the established oligosaccharide into newly produced protein before it is directed to the Golgi apparatus for further processing and transportation sorting to their final destinations. (Adapted from Wang Yu-Chieh et al. (2014), no permission required).

The previously published studies have reported a premature stop codon mutation (Garshasbi M. et al., 2011), frameshift insertion (Molinari et al., 2008), gross deletion (Garshasbi Masoud et al., 2008; Khan M. A. et al., 2011; Loddo et al., 2013) and an intragenic duplication (El Chehadeh et al., 2015) in *TUSC3*. All of these would be predicated either to yield TUSC3 protein which is not functional, or to lead to complete loss of protein, either through deletion of the gene or nonsense mediated decay. The mutation reported in this current research is a frameshift mutation that could initiate nonsense mediated decay (NMD) of the RNA transcript that leads to its total elimination, or may give rise to a truncated, non-functional protein. Given the published spectrum of *TUSC3* mutations reported in various ID patients, it seems likely that the ID condition in Family-1 is the result of absence of functional TUSC3 protein during neurodevelopment rather than due to the presence of defective mutated protein.

It is worth highlighting that some recent work on primary fibroblast cell cultures derived from patients with a null homozygous mutation in *TUSC3* showed no significant difference from controls in glycosylation of selected proteins (El Chehadeh et al., 2015). This may argue against the key role of TUSC3 in glycosylation of extracerebral proteins.

The finding of a novel *TUSC3* mutation in the consanguineous Family-1 from Oman contributes further data on the spectrum of mutations causing this form of ID and the phenotype associated with it. This is the seventh family with a *TUSC3* mutation causing syndromic or non-syndromic ID that has been identified (Figure 3.44).
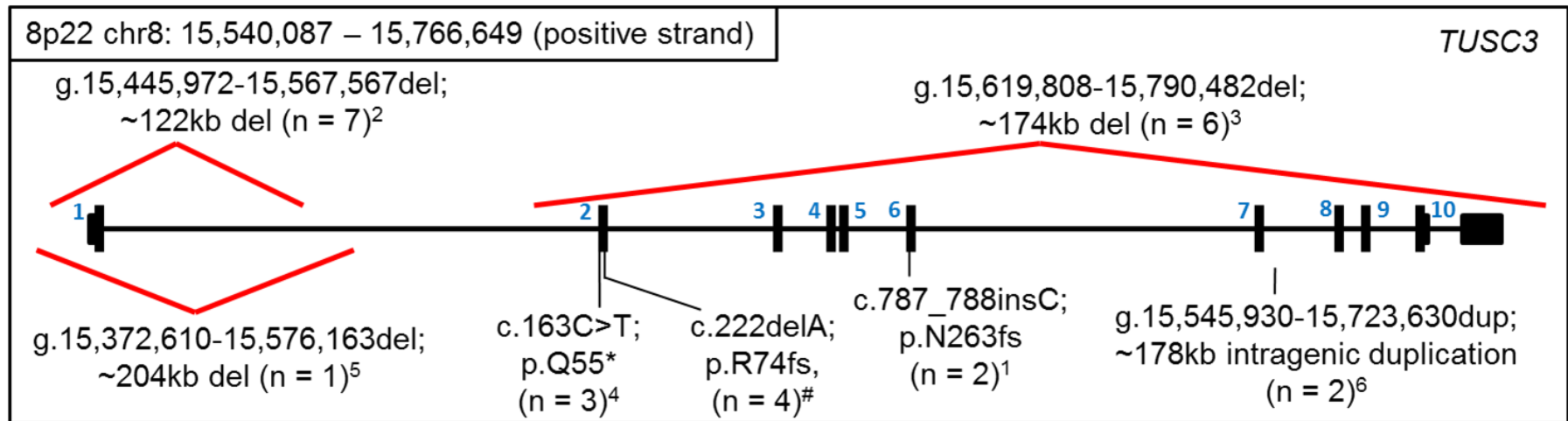
Figure 3.44. Diagram showing the reported *TUSC3* mutations to date. The representation is mapped onto the GRCh38 reference genomic DNA sequence, with exons highlighted according to the transcript with accession code, NM_006765.3. 'n' denotes the number of patients with that mutation. So far, a total of 25 patients have been described with *TUSC3* mutations. The superscript notation assigns the reference that reports the mutation as follows: 1. Molinari et al., 2008; 2. Garshasbi et al., 2008; 3. Khan et al., 2011; 4. Garshasbi et al., 2011; 5. Loddo et al., 2013; 6. El Chehadeh et al., 2015, and # represents the mutation reported in the current study. (Adapted from Al-Amri et al. (2016) with a permission from John Wiley and Sons, and Copyright Clearance Center's RightsLink® service, License number 4055460722605).

3.3.2 Family-2 with *LHFPL5* mutation

The affected members in this family had autistic spectrum disorder, tooth abnormalities and sensorineural deafness as a profound feature. Homozygosity mapping and WES analysis revealed a novel homozygous missense mutation in *LHFPL5* (NM_182548:exon2:c.T575C, p.L192P), co-segregating with the condition in an autosomal recessive fashion, consistent the pattern of inheritance seen in the family and with being involved with the phenotype. In addition, given that both affected individuals are male, an X-linked mode of inheritance was possible. Variants on the X chromosome were therefore identified, filtered and compared to a list of X-linked genes known to be mutated in cases of ID. This identified a missense variant in the NHS gene as a strong candidate for involvement in the phenotype seen in this family.

Mutations in *LHFPL5* are known to cause hearing loss, with eight different mutations already published as causing recessive sensorineural hearing loss (SNHL). The finding in Family-2 is the ninth mutation identified in *LHFPL5* that has been implicated in deafness (Figure 3.45). Different types of *LHFPL5* mutations have been reported and these include a biallelic mutation in the start codon (Shahin et a.l, 2010; Sloan-Heggen et al., 2015) and frameshift mutations (Bensaïd et al., 2011; Kalay et al., 2006; Shabbir et al., 2006), all of which are expected to induce a premature termination in the production of *LHFPL5* mRNA, which might lead to a truncated protein or might cause loss of protein through nonsense mediated decay. Pathogenic mutations that cause missense changes (Ammar-Khodja et al., 2015; Kalay et al., 2006; Marková et al., 2016; Shabbir et al., 2006; Sloan-Heggen et al., 2015); this thesis) and an in-frame deletion (Sloan-Heggen et al., 2015) have also been reported and the protein that is produced as a results of these mutations is not normally functional. By looking at the published reports, there do not seem to be distinct phenotypic differences between the effect of null alleles and missense changes, as all patients had non-syndromic hearing loss, with no reports of the additional symptoms observed in Family-2.
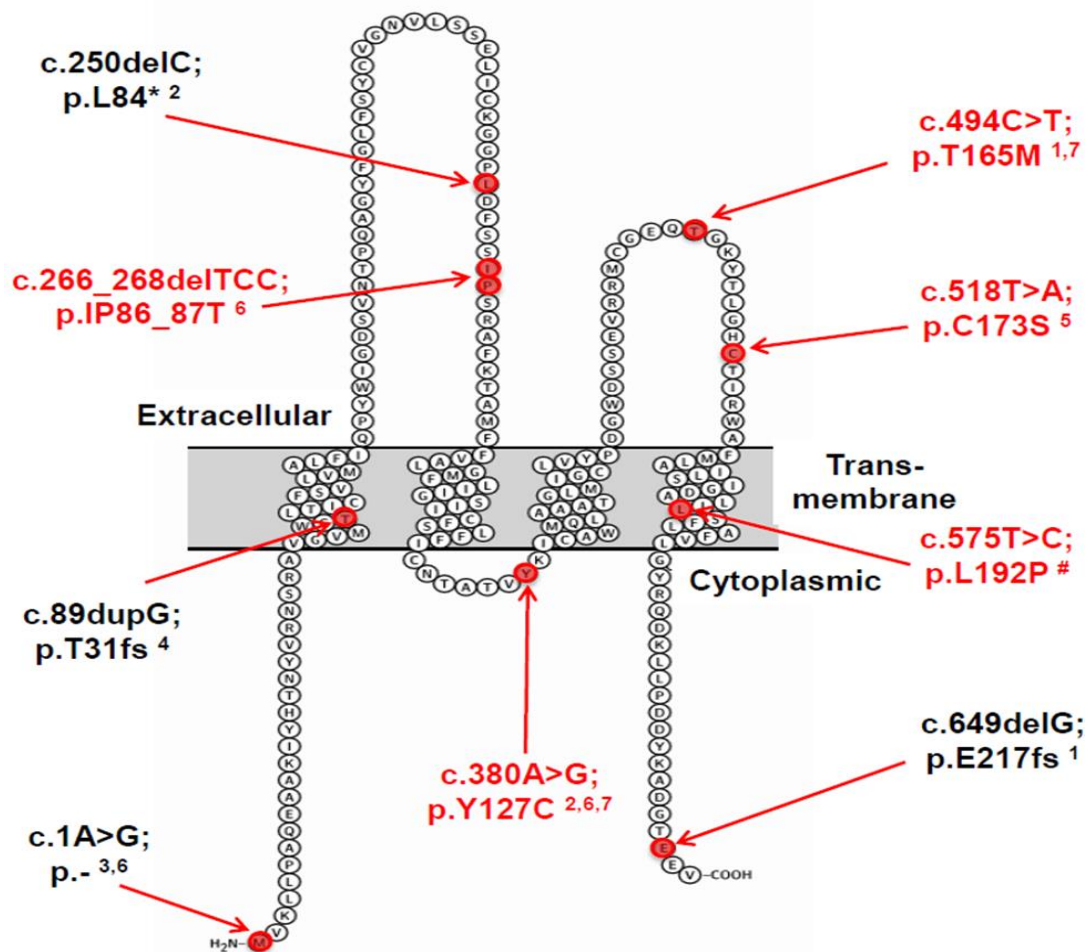
Figure 3.45. Representation of LHFPL5 protein showing the pathogenic mutations identified to date. The protein has four-transmembrane domains. There have been a total of nine *LHFPL5* mutations described up to date, including the one reported here. Missense and in-frame mutations are highlighted in red while frameshift, nonsense and splice mutations are in black. The superscript notation assigns the reference that reports the mutation as follows: 1. Kalay et al, 2006; 2. Shabbir et al, 2006; 3. Shahin et al, 2010; 4. Bensaid et al, 2011; 5. Ammar-Khodja et al, 2015; 6. Sloan-Heggen et al, 2015; 7. Markova et al, 2016 and # represents the mutation reported in the current study. The image was generated from accession code NP_872354.1 using the Protter prediction tool.

LHFPL5 is also known by other aliases including *TMHS* (tetraspan membrane protein of hair cell stereocilia) and *DFNB67* (deafness, autosomal recessive 67). The deduced 219-amino acid protein has four transmembrane domains that form a major part of the mechano-electrical transduction machinery for perception of sound. It is localised to the stereocilia, which are microvilli-like projections of different sizes that are arranged in bundles to make up the hair cells of the inner ear (Longo-Guess et al., 2005; Longo-Guess et al., 2007). LHFPL5 forms a

complex with other transmembrane proteins, TMC1 and TMIE, near the lower end of the stereocilia tip (Fettiplace, 2016). LHFPL5 has also been reported to cooperate with the amino-terminus of PCDH15, which interacts with the top end of the tip where the transduction channel, of unknown identity, is located (Beurg et al., 2015; Zhao B. et al., 2014). LHFPL5 has an important role as a modifier protein since it regulates the localisation of PCDH15 to the stereocilia tip, the formation of the tip-link assembly, and channel conductance (Xiong et al., 2012). Figure 3.46 shows the structure of hair cells and the main molecular components involved in the tip-link assembly. It would be interesting to know the protein interactions that are disturbed by the pathogenic changes in *LHFPL5* that cause human deafness.
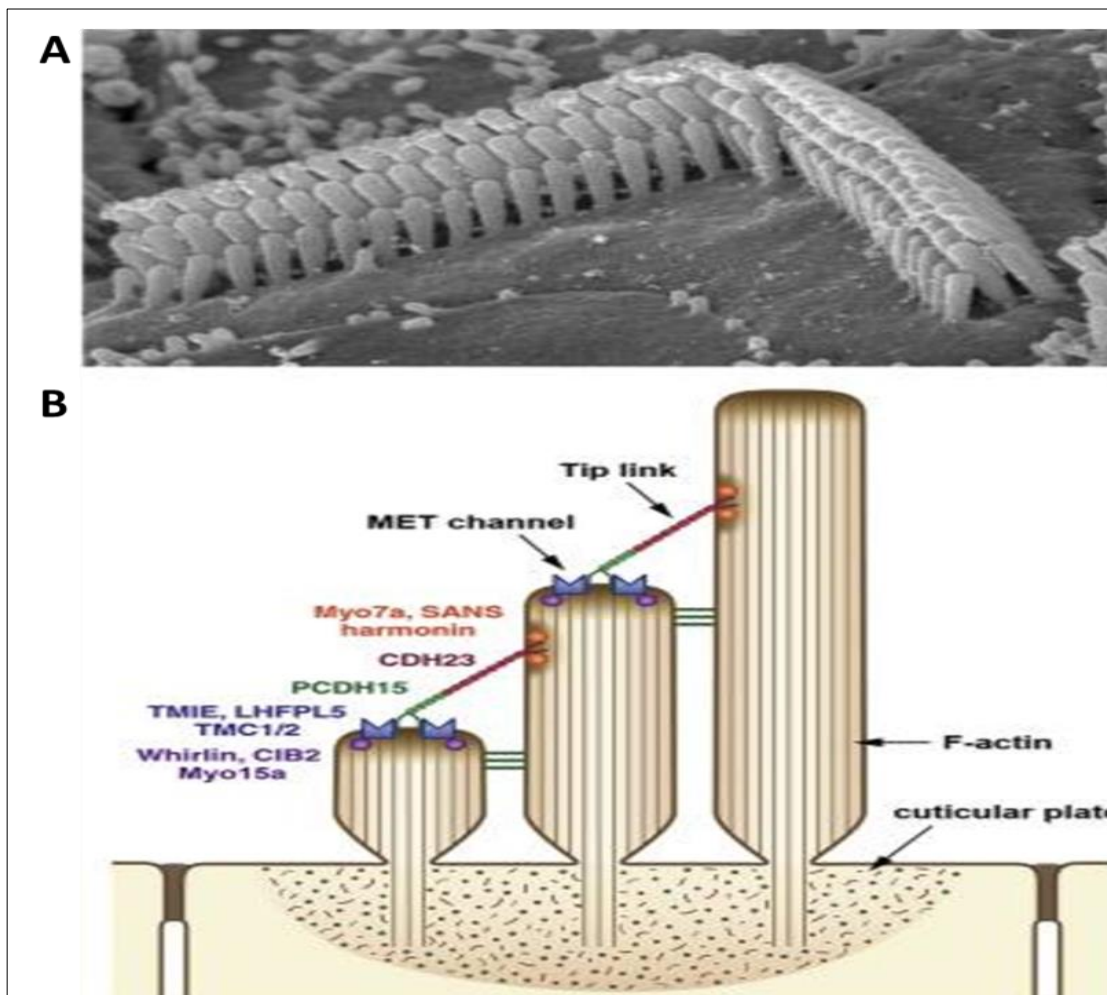
Figure 3.46.The structure of mechano-electrical transduction machinery in hair cells of the inner ear. A. Electron microscope capture of the outer hair cells in V-shaped arrangement. The micrograph is showing compact clusters of stereocilia, each measuring about 2-4 um, projecting in three rows. B. Representation of tip link organisation within the stereocilia and the essentially involved molecular participants. Tip

links, which tie successive stereocilia to each other, are fundamentally made of upper and lower ends. The upper ends consist of homodimers of cadherin 23 (CDH23) whereas the lower ends are formed by homodimers of protocadherin 15 (PCDH15). Components of Usher protein complex (MYO7a, SANS, Harmonin, Whirlin, CIB2) play a role in holding both ends in place. Interconnection of PCDH15, at the lower end of the links, with the mechano-electrical transducer (MET) channel is achieved through its interaction with LHFPL5, along with TMIE and TMC1/2. (Adapted from Fettiplace (2016), no permission required).

LHFPL5 mutations have so far been identified in patients with deafness from Central Europe (Markova et al., 2016), Algeria (Ammar-Khodja et al., 2015), Tunisia , Turkey (Kalay et al., 2006), Palestine (Shahin et al., 2010), Iran (Sloan-Heggen et al., 2015), Pakistan and India (Shabbir et al., 2006), totalling 12 families. Most of the mutations are unique to one family, though it is interesting to note that families were identified with the same mutation in Palestine and Iran (c.1A>G, p.-), Turkey and Central Europe (c.494C>T, p.T165M) and India, Iran and Central Europe (c.380A>G, p.Y127C). Though these findings could reflect mutation hot spots in *LHFPL5*, it is more likely that they represent common founder alleles reflecting population migration over the years. The family described in this report is the 13th with biallelic *LHFPL5* mutations and represents the first time that mutations in this gene have been described in the Arabian Peninsula.

The *NHS* variants reported in Family-2 could not be excluded from being the underlying cause of mental disability in the family especially when considering the fact that about 30-50% of individuals with Nance-Horan syndrome have intellectual disability (Brooks et al., 2010). However, the principle features of Nance-Horan syndrome, including congenital cataract and small mis-shaped hypoplastic teeth, have not been reported in Family-2. The only reported ocular phenotypes in the affected members are hypermetropia and astigmatism, and the tooth anomalies observed were crowding and overlapping incisors (Figure 3.47). Therefore, the affected members do have some tooth abnormalities are but they are not the ones normally seen with NHS mutations. Although a considerable variability of clinical phenotype within Nance-Horan syndrome is known, no case has been so far reported without the characteristic ophthalmic and dental

abnormalities (Liao et al., 2011; Tug et al., 2013). This is also the first time an NHS mutation has been reported in the Arabian Peninsula.
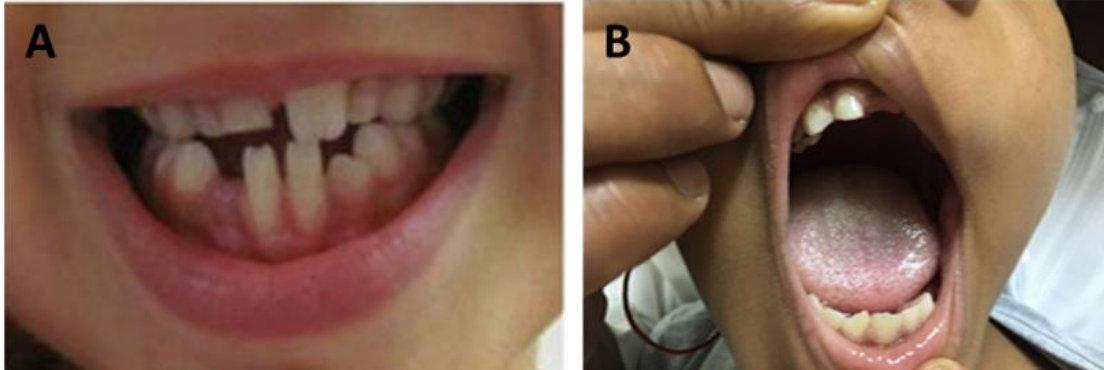
Figure 3.47. Dental features of Nance-Horan syndrome. A. Characteristic screwdriver shaped incisors and diastema in a child with Nance-Horan syndrome (Tug et al. (2013). B. Case IV.3 in Omani Family-2 with mild teeth anomalies including conical shaped teeth, crowding and overlapping incisors. Part of a cochlear implant device is also visible on the left of the picture.

The NHS gene encodes10 exons and is located at Xp22.31–p22.13 (Ding et al., 2009). The biological structure of the NHS protein is unclear but a recent study suggests that there are, at least, four isoforms derived from alternative splicing, the two major isoforms being NHS-A and NHS-1A. (Tian et al., 2017). The pathogenicity of Nance-Horan syndrome is predominantly caused by defects in these two isoforms, which could be due to the functional WAVE homology domain at their N-termini (Brooks et al., 2010). The protein is highly conserved among vertebrates and is expressed in the retina as well as neuronal tissues (Figure 3.48). Although published studies have shown its importance in the development of eye and neural tissues (Burdon et al., 2003; Sharma et al., 2006), the precise function of NHS is still unknown.
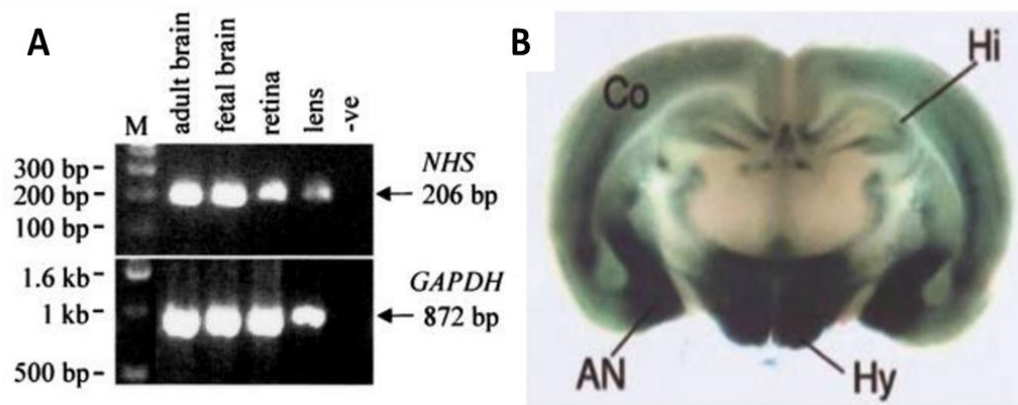
Figure 3.48. Expression of NHS in ocular and neuronal tissue. A. A gel image of reverse-transcriptase PCR on mRNA from human adult brain, fetal brain, retina, and lens, showing the existence of *NHS* transcript in all tested tissues . B. A coronal section of adult mouse brain stained for NHS. The mouse was transgenic for a *lacZ* reporter-gene cassette inserted in the *Nhs* locus. The blue-green colour staining occurred after using β-galactosidase. Co, cortex; Hi , hippocampus; AN, amygdaloid nuclei; Hy, hypothalamus. (Adapted from Burdon et al. (2003), no permission required).

In this family, two novel mutations were discovered: one in *LHFPL5* and the other in *NHS*. It is true that the ultimate significance of the identified *NHS* variant could not be assuredly confirmed in the described phenotype. The ID phenotype in this family cannot therefore be accounted for by genetic findings. It remains possible that this particular LHFPL5 variant also causes ID, as the protein is expressed in the brain, or that the NHS variant or an undetected variant in another gene is causing this condition. Alternatively, the ID in this family may be the result of an undetected environmental factor such as poor maternal nutrition (Abu-Saad and Fraser, 2010; Purandare, 2012) or intrauterine viral infection (Cheeran et al., 2009).

However, the LHFPL5 mutation is likely to account for the deafness phenotype in this family, and family members would be expected to benefit from this finding, through genetic counselling and carrier screening of additional family members and/or potential marriage partners. Affected children within the family can now be identified earlier and appropriate intervention offered, such as cochlear implants, to help with the child's speech and language development.

3.3.3 Family-3 with *ANKRD2* and *PDZD8* mutations

The three affected individuals in Family-3 presented with cognitive deficit, autistic features, scoliosis, myopia and some dysmorphic features. Homozygosity mapping and WES analysis indentified two novel homozygous variants; a missense mutation in *ANKRD2* and a nonsense mutation in *PDZD8*, which co-segregated with the described phenotypes in an autosomal recessive mode of inheritance. Both variants are located within the biggest identified homozygous region (28.3 Mb) on chromosome 10 and they could individually or together be the cause of intellectual disability.

ANKRD2 is a member of the muscle ankyrin repeat protein (MARP) family and is involved in various pathways including muscle stress response, myogenic differentiation and apoptosis (Bean C. et al., 2008; Jasnic-Savovic et al., 2015). Studies have shown that ANKRD2 plays an anti-inflammatory role through a direct interaction with p50 subunit in NF-$_k$B-mediated response (Bean C et al., 2014). Tsukamato et al. (Tsukamoto et al., 2002) described the expression of mouse ANKRD2 in cerebrum and cerebellum. PDZD8, on the other hand, is a cellular cytoskeletal regulator with a poorly understood function. It is known to be a moesin-interacting protein involved in microtubule cross-linking (Henning Matthew S et al., 2011). PDZD8 has also been shown to support human retrovirus infection (HIV-1) via its Gag region which facilitates the interaction with the virus (Henning Matthew S. et al., 2010). Based on this role, PDZD8 has been studied as a potential target for intervention in retrovirus infection, but it was recently shown to be not always required for the infection (Zhang Shijian and Sodroski, 2015). PDZD8 protein was selected as one of the best five candidates (Figure 3.49) for specific interaction with IL13Rα2, a receptor that is overexpressed in various human tumours and associated with poor prognosis (Bartolomé et al., 2015). *PDZD8* was identified as one of the strongest candidates, along with *NRAP* and *HSPA12A,* for a forelimb muscular anomaly in Japanese black cattle (Masoudi et al., 2008) but it was then brought to our knowledge, based on a follow up contact with the author of this paper, that *PDZD8* was ruled out. Recently, significantly reduced PDZD8 expression in the amygdala has been associated with posttraumatic stress disorder in humans (Bharadwaj et al., 2016).
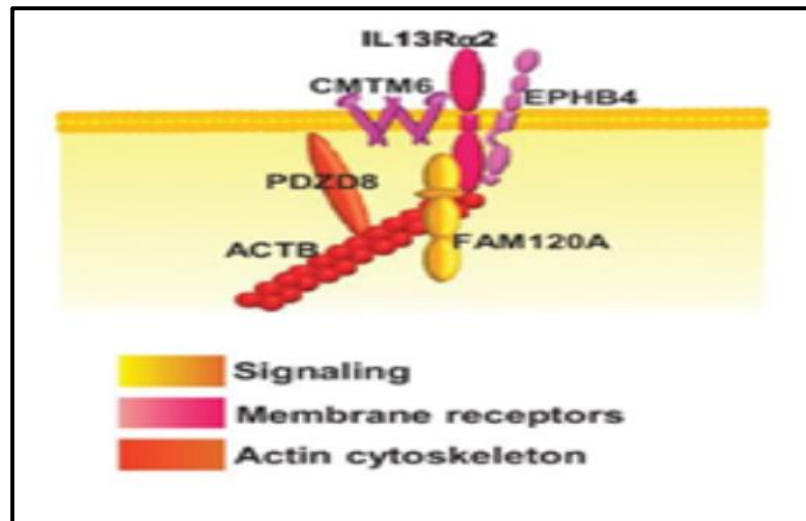
Figure 3.49. PDZD8 in the interaction network for IL13Rα2. Bartolome and colleagues carried out a proteomic approach using colon cancer cells in order to identify the protein partners of IL13Rα2. After precipitating IL13Rα2 from KM12SM cell lysate and fractioning the pulled-down proteins by SDS-PAGE, mass spectrometry was carried out. The analysis was followed by a filtering process at which all the proteins which have no means of contact with the membrane IL13Rα2 (like mitochondrial proteins) or have no relevant protein-protein interactions (like chaperons) were excluded. The final identified IL13Rα2 partners were FAM120A, EPHB4, CMTM6, PDZD8 and ACTB proteins, which are highlighted in this representation. The colour scale indicates the known biological function for each of these proteins. (Adapted from Bartolomé et al. (2015), with a permission from the American Association for Cancer Research and Copyright Clearance Center, License number 4127721023107).

Additional screening of *ANKRD2* and *PDZD8* in 100 chromosomes from healthy individuals of matching ethnicity revealed no homozygous or heterozygous mutations. Both mutations were also absent from 60,706 subjects in the ExAC database, indicating their extreme scarcity. Examination of the data available on ExAC also suggested that, of the two genes, *PDZD8* has by far the most interesting constraint metrics (Table 3.12). This suggests that *PDZD8* is much more constrained (especially for loss-of-function alleles like the one described here) than *ANKRD2*. This does not exclude *ANKRD2* but it gives support to the loss-of-function mutation in *PDZD8* as the possible cause of the ID in Family-3.

Hoischen et al. (2014) mentioned that consideration of suggested candidate genes needs to be investigated for various factors including the location of the mutation, its load in healthy controls, pathogenicity predictions and whether the

191

gene is a known component in one of the networks that are associated with the disease in question. Conservation measures give an indication of the importance and stability of the affected amino acid in different species. A highly conserved position suggests that a mutation could have functional consequences (David and Sternberg, 2015). The location of R295W in ANKRD2 is evolutionarily conserved but the corresponding position in the soft-shelled turtle is occupied by a tryptophan (W), the same residue introduced by the mutation in Family-3 (Figure 3.20). This raises a question as to its significance. This finding suggests that the R295W mutation may not have adverse functional effects since ANKRD2 apparently functions normally in soft-shelled turtles. It is not easy to defend this assumption as, unfortunately, turtle intelligence is not a subject that has been much studied, which could be probably true for most reptiles.

In fact, there are few studies but most of them are in the form of controlling the movements of the animal utilising various cues. For example, in an experiment by Lee and colleagues (Lee Serin et al., 2013), turtles were successful at following a visual stimulus and performing a specific non-linear movement trajectory. Furthermore, red-footed tortoise was also capable of learning and mastering an eight-arm radial maze, at which they needed to remember various different cues (Mueller-Paul et al., 2012). On the other hand, there are also some findings that might suggest turtles are not good at facing navigation challenges. For instance, painted turtles were not able return to their pond when they kept a mile away but they did very well in reaching their pond quickly when they were released 100 meters from home (Emlen, 1969).

It is true that sequence conservation measures have played a significant role in assessing and ranking candidate protein coding and non-coding variants (Adzhubei et al., 2010). However the value of such analyses are impacted by the power of the sequence alignments used to determine conservation. Therefore, it is generally considered best not to depend solely on conservation measures for evidence about the pathogenicity of a particular variant (Macarthur et al., 2014).

It was a shame not to be able to track the segregation of the homozygous *ANKRD2* mutation (NM_001129981:exon1:c.C272T) that was found in one of these Belgian cases (Figure 3.12). The analysis of the *ANKRD2* and *PDZD8* mutation load, based on UK10K cohorts, has not shown a conclusive result, although it could be argued to have favoured the candidacy of *ANKRD2* somewhat (Table 3.11). Furthermore, the strong staining of ANKRD2 in the substantia innominata suggests an important role for this protein in this brain region. The substantia innominata (also known as the nucleus of Meynert) is part of the basal forebrain and is located below the anterior part of the thalamus and contains telencephalic structures (Haines and Ard, 2013). The substantia innominata is believed to have a major role in the neural processing of signals (Noback et al., 2005; Zhu Qiong-Bin et al., 2016). This could be due to the fact that the substantia innominata is enriched in the cholinergic neurons which form extensive connections to the entire neocortex (Zhang Jin-Ping et al., 2006). Moreover, the substantia innominata is also connected to control centres such as the amygdala, hippocampus and hypothalamus, which provide significant network interactions important for electrical activation, recognition of reality, navigation and perception analysis (Kissin, 2012). It has been shown that individuals with Alzheimer's disease (AD) have major cell loss and degradation in the substantia innominata (Zarow et al., 2003). Significant associations have also been reported in AD patients between volumetric loss of the grey matter in substantia innominate and the levels of cognitive impairment (Colloby et al., 2016).

The localization of PDZD8 in the choroid plexus could indicate that PDZD8 has some role in this part of blood-brain barrier. The structure of choroid plexus acts as functional hurdle between the cerebrospinal fluid (CSF) and blood. It is also considered as an access window for immune cells entry into the CNS (Lun et al., 2015). The epithelial cells in the choroid plexus are responsible for certain processes that are crucial for normal embryo development, particularly the ones related to production of CSF, ensuring brain fluid homeostasis (Saunders et al., 2008). Therefore, even a minor disruption to this brain barrier could contribute to development of neurological conditions (Liddelow, 2015).

Drosophila is a simple animal model that has been used in genetic research for more than a century (Morgan, 1910). There are various features of *Drosophila* that make it a good option to be used for experimental studies and these features include the short lifespan (~120 days), the low number of chromosomes (only 4) and sharing some cognition/behaviour aspects with mammals (Prüßing et al., 2013). Moreover, *Drosophila* was one the of the first creatures to have its full genome sequenced and available for research (Adams et al., 2000). *Drosophila* was used in this thesis to study the effect of knocking down CG10362, which is the *PDZD8* orthologue in *Drosophila*, on behaviour. *Drosophila* has a wide range of behaviour that could be studied, like learning, memory and decision making (Bellen et al., 2010; Guo et al., 2010).

Three behaviour assays were done for the CG10362 knocked-down *Drosophila* in order to measure their olfactory learning, negative geotaxis and courtship suppression. Surprisingly, the wild type flies used in this experiment (W[118]) appeared to show reduced learning ability (p = 0.159) to choose the odour that was not paired with the mechanical shock (olfactory learning assay, Section 2.18.6), though this reduction was not statistically significant. Similarly, courtship activity was not suppressed after one hour of courting females in CG10362 knocked-down flies (courtship supression assay, Section 2.18.5). However, the wild type flies used in this experiment (*W*[118]) were not the most appropriate reference, as the white-eyed background on which the knock down flies were bred are not technically wild-type flies. The colour of *Drosophila* eyes is brick-red which is determined by two key pigments, the brown 'ommochromes' and red 'drosopterins' (Kim et al., 2006). *Drosophila* with no eye pigmentation are referred as "white-eyed" which is caused by a mutated *white* (*W*) gene (Mackenzie et al., 1999) that encodes the transmembrane ABC transporter (Ewart and Howells, 1998). Therefore, the obtained findings may in fact be the result of presence of the *white* locus (*W*) mutation, which has been found to have an impact on various aspects of courting behaviour (Krstic et al., 2013). On the other hand, *white* locus (*W*) caused no impact on the rapid iterative negative geotaxis (RING) assay (Section 2.18.4). The significantly lower geotaxis index (p=0.002) in the produced male progeny, compared to wild-type males, is indicative of a loss of locomotor

behaviour and could point to CG10362 being important in controlling physical movement. It is worth mentioning here that the experiment has since been repeated by others with correct controls, and an impact on negative geotaxis and long term memory has been confirmed.

In comparison with invertebrate models such as flies and worms, mice are a better model to study human diseases, especially when studying complex biological systems like immune, cardiovascular and nervous systems. Mice are mammals and have anatomical, physiological and genetic similarities with humans (Rosenthal and Brown, 2007). Moreover, various features, like cheap cost and ease of maintenance, make them the most widely used animal model in research (Eisen, 2005; Justice and Dhillon, 2016). After getting the results from the *Drosophila* experiments, an inquiry was made to the Wellcome Trust Sanger Institute (WTSI) on whether they had a *Pdzd8* knock out (KO) mouse that we could use for behaviour testing. It was found that *Pdzd8*tm1b(EUCOMM)Wtsi line was available. Phenotyping of the *Pdzd8*tm1b(EUCOMM)Wtsi line in the WTSI pipeline (White et al., 2013) indicated that both males and females have reduced body weight, body length and body mass (A. Galli, personal communication). This is biological evidence that KO of PDZD8 does indeed have phenotypic effects in mammals. An arrangement has, therefore, been made to import some of these mice to Leeds in order to do some learning and memory tests (Kirshenbaum et al., 2015). If homozygosity for the *PDZD8* truncation mutation really is responsible for the intellectual disability in Family-3, *Pdzd8*tm1b(EUCOMM)Wtsi mice would be expected to show cognitive deficits.

Based on the genetic analysis of the Omani Family-3, as well the different experimental analyses, mutations in both *ANKRD2* and *PDZD8* remain plausible underlying causes of the ID phenotype in the family. ANKRD2 is a known myogenic factor, and may therefore account for the join laxity described in the affected members of the family. In addition, its contribution to the neurological pathology cannot be excluded. In fact, the *ANKRD2* homozygous mutation found in the Belgian ID case, the substantia innominata localisation, as well as the

enrichment of *ANKRD2* variants in the UK10K ID cohort are all supportive of a role of the *ANKRD2* mutation in this phenotype. Importantly, Tsukamoto and colleagues have already pointed out that ANKRD2 could have an operative role in brain according to their findings on expression of mArpp, a murine gene identical to the human *ANKRD2*, in neurons of the cerebellum and cerebrum (Tsukamoto et al., 2002).

On the other hand, the gene encoding the cytoskeleton regulator PDZD8 showed significant constraint metrics from ExAC, is expressed in the brain, the mutation found is more likely to affect protein function given sequence conservation and the nature of the mutation observed, and the *PDZD8* knock-down *Drosophila* had significantly reduced geotaxis. Thus while it is not possible to discount the possibility that mutations in both genes are causing the condition in Family-3, PDZD8 seems the stronger candidate for causing the ID phenotype. This is the first time that either *ANKRD2* or *PDZD8* have been linked to intellectual disability; it would be interesting to see whether other supportive evidence arises in the future.

3.3.4 Families-4, 5 and 6 with unidentified causes

There are several reasons why genetic testing is a good option for differential diagnosis in certain families. Important factors include having more than one individual within the family affected with the same diagnosis and the presence of particular developmental delay with congenital abnormalities (Trattler et al., 2012). The correct diagnosis of disease conditions is valuable and would aid in finding the casual variant (Rabbani et al., 2014). However, correct identification of the causative gene requires good physical examination, detailed medical family history and clinical laboratory findings. It might not be always possible to make a confident diagnosis for particular cases, especially with subtle phenotypes which might not have been previously associated with the disease of interest (Ropers H, 2006).

Technological advances in DNA sequencing have significantly helped in identifying the causes of various genetic diseases in humans. In intellectual disability, the positive impact of high-throughput sequencing can be seen in the growing number of the newly discovered ID genes and publications; ID papers have increased from <20 in 2006 to around 190 in 2015 (Chiurazzi and Pirozzi, 2016). Despite these successes, there is still a large proportion of about 50% of ID cases that have unidentified causes (Alexander-Bloch et al., 2016).

Families-4, 5 and 6 have all been generally diagnosed with ID but each of them has a different number of affected individuals with different manifestations. Cognitive impairment is not a direct diagnosis, especially with the presence of certain comorbidities which make a correct diagnosis more challenging (Kaufman et al., 2010). The families were all subjected to homozygosity mapping and whole exome sequencing. Testing of all the filtered homozygous and compound heterozygous variants did not reveal any candidate that could be considered as the underlying cause. ExomeDepth also did not reveal any underlying CNVs in these families.

One potential reason for the inability to find the underlying variants in these three families is the possibility that some of the family members were not diagnosed correctly. It is known that incorrect attribution of disease status among family members can misguide researchers, resulting in failure to identify the underlying causative factors (Duncan et al., 2014). This could occur when a family member who is actually unaffected, but within a borderline status showing features of very mild ID, is erroneously classified as affected. A causative mutation could then be excluded because it is not carried by that particular individual, even if it shows complete co-segregation with ID the rest of the family.

Incorrect attribution of disease status could also involve the opposite scenario whereby an affected individual is classified as unaffected, which could negatively affect the ability to find the causative variants. For example, this could occur in cases of reduced or incomplete penetrance, when an affected person fails to express most features of the disease in question (Cooper David N. et al., 2013).

In fact, various studies have shown that ID could be caused by disease risk variants that display incomplete penetrance (Ehret et al., 2015; Kuechler et al., 2015; Ropers H and Wienker, 2015; Vorstman and Ophoff, 2013); finding this type of variant is a challenge and would require larger data sets for statistical significance (Wang Jian and Shen, 2014).

Another possible reason for the failure to identify causative variants in these families is that they may not be detected by the technologies utilised. If the mutations are present in the middle of an intron, or in a promoter or enhancer, it is not possible for exome sequencing to detect them, as this test focuses on exons (Warr et al., 2015). Such mutations would require whole genome sequencing in order to identify them (Belkadi et al., 2015) because they are not conventionally captured by whole exome sequencing (Warr et al., 2015).

During the recent few years, somatic mutations have been found to play a pathogenicity role in human diseases other than cancer (Frank, 2014; Kennedy et al., 2012). Somatic mutations, that arise in the germ cell lineage during the early eight to fourteen divisions after fertilization, may not be detected using offspring blood or saliva (Acuna-Hidalgo et al., 2016). These genetic variations are not originally present neither in the parent's germs (ovum or sperm) nor in the fertilized egg but they take place later. So, a somatic mutation that occurs, for example, to a neuronal cell during the time of prenatal brain development could lead to mosaicism since not all divided cells will have this mutation (Biesecker and Spinner, 2013). The ultimate result about whether this somatic mutation would alter the function of the associated gene causing a human disease is based on the number of the affected cells as mutation can still remain restricted to small number of cells which are not phenotypically significant (Acuna-Hidalgo et al., 2016). Assessment of somatic mutation causing neurodevelopmental disorders would always require highly sensitive technologies (Poduri et al., 2013). A study based on high-coverage whole-genome sequencing (WGS) and single-cell sequencing (SCS) from the cerebral cortex of three normal individuals found thousands of somatic single nucleotide variations, that tends to occur during transcription process rather than due to errors in DNA replication (Lodato et al.,

2015). This high enrichment of somatic mutations and their predominance in the brain has pointed out that they could be at the source of developing psychiatric disorders (Insel, 2014; Poduri et al., 2013).

Chapter 4: Genetic and functional studies to assess the implication of Mendelian alleles in causing Schizophrenia in consanguineous UK families

4.1 Introduction

Psychiatric disorders are among the top causes of disability in all countries, regardless of their living standard. More disability cases are caused by depression than by HIV infection or dietary problems (Lopez and Murray, 1996). As a result, psychiatric disorders have been increasingly the focus of new medical approaches and research throughout the world. In the UK, there has been a radical change in approach to these conditions, with mental health services shifting from an asylum-based care towards care in a community setting (Herring et al., 2006).

Psychotic disorders include a wide range of conditions, but the most common are schizophrenia (SZ) and bipolar disorder (BPD), both with a largely unknown aetio-pathogenesis. Patients with a psychotic condition frequently meet criteria of more than one particular disorder, as many of the associated signs and symptoms overlap between disorder categories (Paris, 2007). Schizophrenia patients normally have positive as well as negative symptoms. Positive symptoms include delusions, hallucinations and paranoid thoughts, whereas negative symptoms include lack of interest, inability to enjoy fun activities and social withdrawal (Patel et al., 2014). Accumulating evidence suggests a shared genetic architecture among psychotic disorders (Doherty and Owen, 2014); and some of the implicated overlapping genes have already been mentioned in Chapter 1 (Section 1.4).

The influence of consanguinity, when a marriage occurs between two individuals belonging to a recent common ancestor, on human diseases has been well-documented (Bittles and Black, 2010; Hamamy, 2012; Modell and Darr, 2002;

Mokhtari and Bagga, 2003; Shawky et al., 2013) and also highlighted in Chapter 1 (Section 1.1). The Pakistani population of West Yorkshire, comprises 189,000 individuals, the vast majority of whom derive from Mirpur in Azad Kashmir, Pakistan (Shaw, 2014). Marriage is normally within the community and approximately 70% of marriages are between relatives (Darr and Modell, 1988). As a result, this community is both endogamous and consanguineous, leading to enrichment for homozygosity of recessive alleles, high background autozygosity (Woods et al., 2006) and recessive disease. SZ and other psychoses are common in this population, with a higher prevalence of mental disorders than the general population (Mcgrath et al., 2004).

One hypothesis proposed for this increase is the change of culture associated with immigration (Bhugra, 2004; Bhugra and Becker, 2005; Hollander et al., 2016; Shekunov, 2016), as discussed in Section 1.3.3 of Chapter 1. Based on the 2001 Census, Pakistanis (747,000) comprised the second biggest group, after Indians (over one million), of immigrants living in the UK (Charsley, 2013; Schierup et al., 2006). We hypothesise that a proportion of the higher prevalence of mental disorders is caused by recessive alleles of major effect predisposing to psychosis in the Pakistani population of West Yorkshire. The existence of Mendelian SZ alleles has been discussed in Section 1.3.5 of Chapter 1.

This chapter describes research in two families in which a psychotic disorder appears to be segregating in a Mendelian fashion. Both families were collected along with others in 2012 for a project entitled "Genetic investigation of schizophrenia in the Pakistani population of West Yorkshire". Both had been the subject of prior research by others, which the author sought to extend as part of this project. The first family (Family-7) has a putative mutation in *WHRN* that appears to co-segregate with psychosis. In previous work detailed in Section 1.2.1, others had carried out both genetic and functional assessment of that variant, revealing that it weakens binding between whirlin (encoded by *WHRN*) and the ubiquitin-ligase UBR4. The second family (Family-8) is consanguineous and has multiple cases of schizophrenia, a pattern of inheritance highly suggestive of a recessively inherited Mendelian allele. Homozygosity mapping

suggests that one or more genes in a locus on chromosome 13q31 is a plausible underlying cause of the increased incidence of schizophrenia observed in this family. Previous work on Family-8 is detailed in Section 1.3.1.

4.2 Family-7: the possible involvement of a *WHRN* variant in psychotic disorder

4.2.1 Work done previously

The initial genetic work was undertaken by Dr Jose Ivorra Martinez (Formerly of Leeds University but now moved back to his country Spain), who showed that the two brothers affected with psychosis (IV.7 & IV.8) in Family-7 (Figure 4.1) were homozygous for a mutation in *WHRN*. The clinical information and the work performed before this doctoral research was initiated are explained below. *In-silico* modelling was carried out by Dr Jonathan Mullins of Swansea University, while protein-protein interaction studies were carried out by Dr Erwin van Wijk, Radboud University, Nijmegen.

4.2.1.1 Family collection and diagnosis

Family-7 is a consanguineous family derived from a first cousin marriage (Figure 4.1). The family lives in Bradford but originates from Mirpur in Pakistan-administered Kashmir. All the members were initially assessed by Dr Qadeer Nazar at Lynfield Mount Hospital, Bradford and the affected individuals were diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-4). The diagnosis was confirmed by Dr Alastair G Cardno of the Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds.
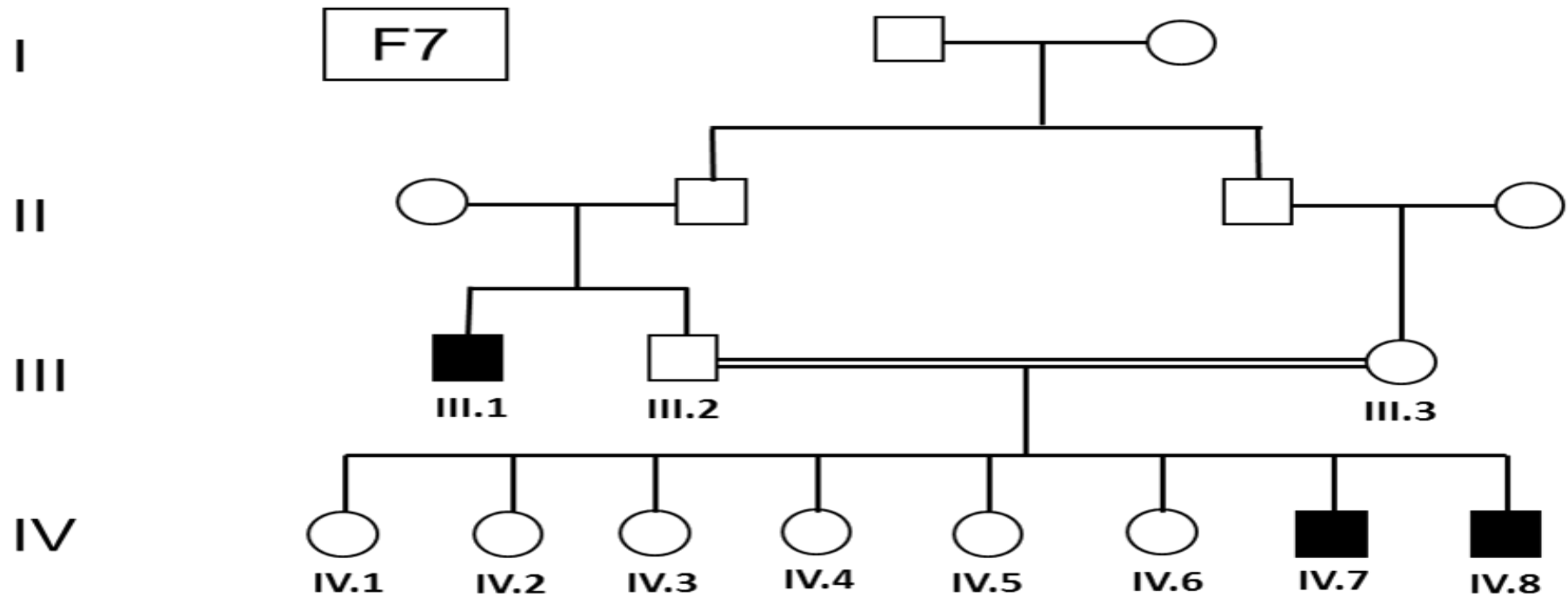
Figure 4.1. Pedigree of Family-7. A consanguineous marriage between first-cousins in the third generation (III.2 and III.3) resulted in eight children. There are six unaffected sisters in this family (IV.1-IV.6) and two sons affected with psychosis: one with schizophrenia (IV.7) and the other one with schizoaffective disorder (IV.8). The mode of inheritance in this family is suggestive of X-linked inheritance, given that only males are affected and all females are unaffected. However, this does not exclude autosomal recessive inheritance.

4.2.1.2 *WHRN* as a causative gene

To originally identify putative causative mutations in the family, Dr Jose Ivorra Martinez used homozygosity mapping combined with whole-exome sequencing. Microarray analysis of individuals IV.7 and IV.8 was performed using an Affymetrix Genome-Wide SNP 6.0 Array, and the results were analysed for identity-by-descent (IBD). Two shared homozygous regions were found on chromosomes 5 and 9 (Figure 4.2).
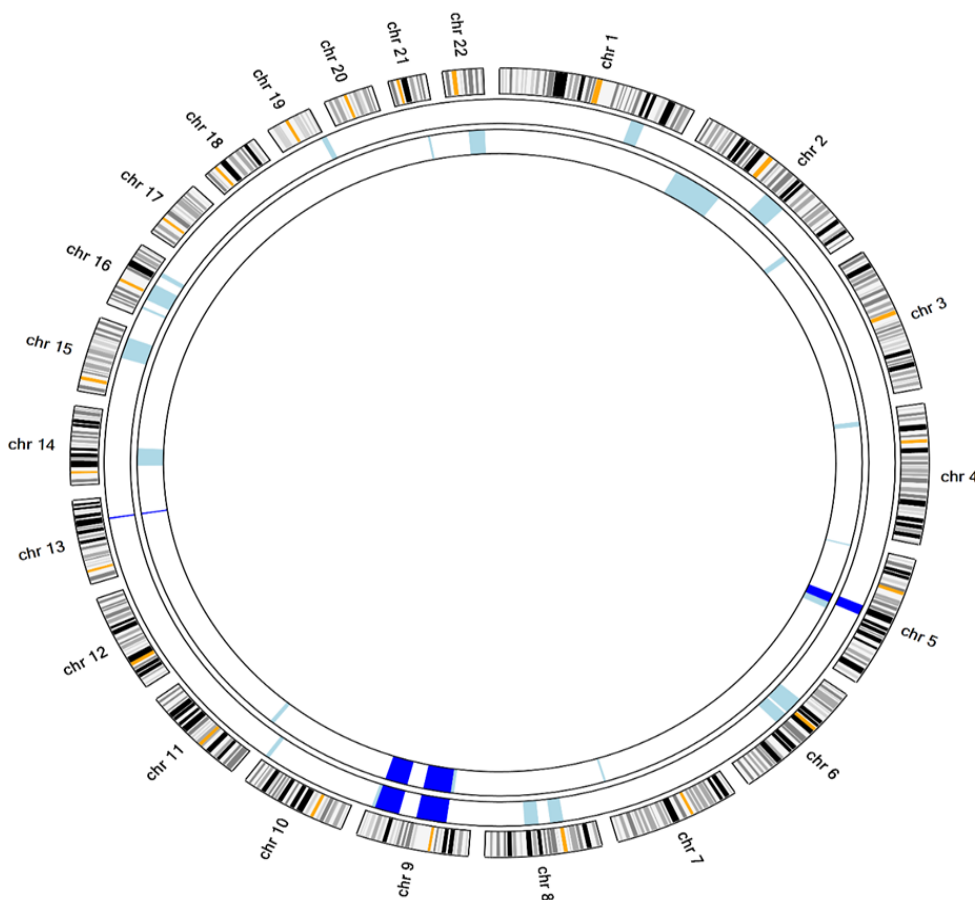
Figure 4.2. Shared homozygous regions in Family-7. An ideogram where the outer circle contains template segments of the 22 autosomal chromosomes. The middle circle is the SNP data from subject IV.7 and the inner circle is SNP data from subject IV.8. The light blue areas are the homozygous regions within each of the individuals and the dark blue regions on chromosomes 5 and 9 are the homozygous regions shared between both affected individuals. The region in chromosome 13 is small, less than 1mb.

Whole-exome sequencing was carried out in one of the affected individuals (IV.8) using an Illumina HiSeq 2500 platform according to the standard Agilent

SureSelect Library Preparation protocol. Data were filtered to select only homozygous splicing, missense, nonsense or frameshift variants with a minimum depth coverage of 20, less than 1% frequency in the 1000 Genomes or 6500 Exomes projects, and predicted to be pathogenic according SIFT and Polyphen2 scores. After checking the segregation of the candidate changes in the parents and six unaffected siblings under autosomal recessive or X-linked models, only the *WHRN* variant appeared to co-segregate with psychosis in the family. The unaffected parents were heterozygous, the affected sons were homozygous, and the unaffected daughters were a mixture of heterozygous and wild-type. This mutation, in exon 6 of NM_015404 *WHRN* (cDNA, c.C1348T), is a substitution of an arginine by a cysteine (p.R450C). PhyloP score (Pollard et al., 2010) indicated that the arginine at position 450 is conserved (0.999), which suggests that a change at this position could have functional consequences (David and Sternberg, 2015)

WHRN is a gene with 12 exons located on chromosome 9q32, and encodes a protein, whirlin, the structure of which is shown in Figure 4.3. Like other PDZ domain containing proteins, whirlin is involved in protein-protein interactions (Green et al., 2013). Whirlin has two main isoforms: short (NM_001083885) and long (NM_015404; NM_001173425) and their protein measures 49kDa and 97kDa, respectively. The main difference between them is that the first 5 exons are missing in the short isoform, so that it has only one PDZ domain (PDZ3) and a proline-rich domain (Aller et al., 2010). Both short and long isoforms of whirlin are expressed in brain, retina and cochlea of fetal and adult tissues (Van Wijk E. et al., 2006b).
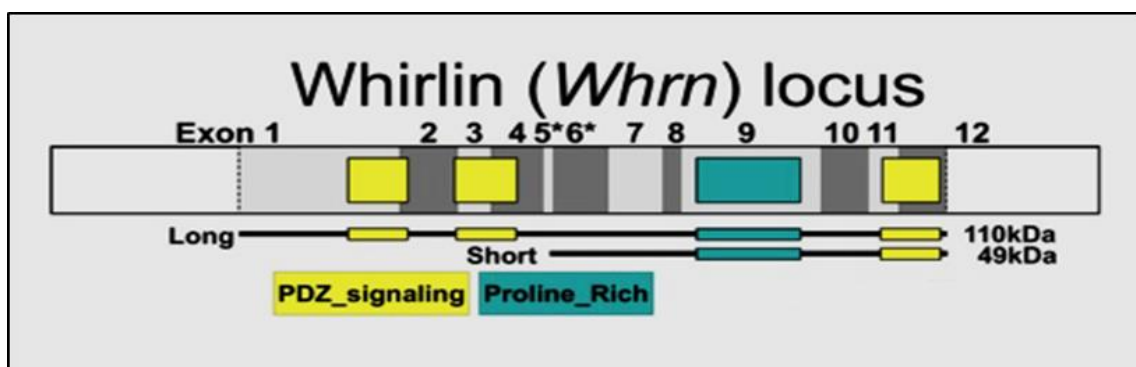
Figure 4.3. Structure of long and short forms of Whirlin. Whirlin is composed of three PDZ domains (yellow) and a proline-rich region (blue/green). The short form of 49 kDa

has alternative transcriptional start sites (asterisks). The exons are indicated by alternately shaded boxes scaled according to their lengths. Full-length whirlin is 110 kDa. (Adapted from Green et al. (2013), no permission required).

Homozygous mutations that cause premature termination of the protein near the N-terminus of whirlin have been found to cause the deaf-blindness condition, type 2D Usher syndrome (Ebermann et al., 2006). On the other hand, truncating mutations near the C-terminus have been reported to cause autosomal recessive deafness (Mustapha et al., 2002). As far as we know, no mutation has, so far, been identified in the middle domain of whirlin, where R450 is located. None of the patients in Family-7 have any obvious problems with hearing, vision or vestibular function.

GWAS studies have already suggested that *WHRN* is a candidate for involvement in bipolar disorder (Baum et al., 2008; Sklar et al., 2008). The gene was also independently replicated in a Finnish bipolar disorder cohort (Ollila et al., 2009). In addition, the chromosome region where *WHRN* is located (9q31-32) has been identified as a linked region for bipolar disorder (Detera-Wadleigh Sevilla D. et al., 1997; Liu J. et al., 2003; Venken et al., 2005) and schizophrenia (Freedman et al., 2001; Kaufmann et al., 1998) in family studies. Therefore, evidence is already available to support the hypothesis that variants in *WHRN* increase the risk of developing psychotic disorders.

4.2.1.3 Screening for the R450C mutation

In light of this preliminary evidence suggesting that the *WHRN* R450C variant might be a Mendelian allele predisposing to schizophrenia, it was then screened for in DNA from 28 Pakistani and 58 Caucasian local schizophrenia patients but no further variants were identified. Further scanning of online sources also excluded this variant from 1,442 schizophrenia/psychosis cases available via the UK10K resource. A number of control cohorts containing a total of 10,047 healthy individuals were scanned (Table 4.1) and an unaffected Afro-American individual from the NHLBI GO Exome Sequencing Project was found to be a *WHRN* R450C

heterozygote. These findings, detailed in Table 4.1, suggest the scarcity of the R450C mutation.

Used cohort	Type of sample	Number of samples	Ethnicity
Leeds/Bradford (DNA)	Patients	80	Caucasian/British Pakistanis
UK10K (Online)	Patients	1442	Caucasian
UK10K (Online)	Controls	2453	Caucasian
NHLBI GO ESP (Online)	Controls	6502	American-European
1000 Genomes Project (Online)	Controls	1092	Various origins

Table 4.1. Samples screened for the *WHRN* R450C variant. A total of 1,522 patients with various psychoses were checked for the R450C variant. The control samples used were available via online sources from different databases, and the total number of screened samples was 10,047.

In order to investigate the middle region of whirlin and to check its relation to schizophrenia risk, an association study was done with rs4978584 (c.G1318A; p.A440T), a common missense polymorphism located only 10 amino acid residues away from R450 (Figure 4.4). Genotyping of rs4978584 in the patient and control cohorts revealed that it is in Hardy-Weinberg equilibrium, and is significantly associated with schizophrenia ($p = 0.017$) under both additive (odds ratio for minor allele: 1.22; $p = 0.017$) and codominant models ($p=0.027$).
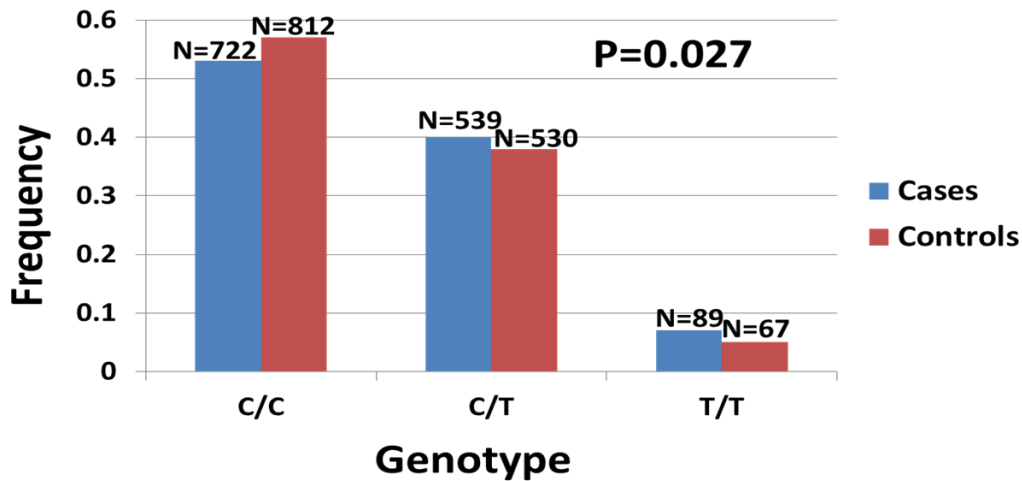
Figure 4.4. Association of rs4978584 in schizophrenia. The association was done using cases and controls from UK10K database. The reference genotype (C/C) is more in the controls whereas both heterozygous (C/T) and homozygous (T/T) are more in cases with significant association.

4.2.1.4 Structural modelling of the R450C whirlin variant

Given the apparent linkage of the R450C whirlin variant with psychotic disorder in Family-7, the association of rs4978584 with psychosis in the cohorts in Table 4.1, and the published evidence of a link between *WHRN* variants and bipolar disorder (Baum et al., 2008; Ollila et al., 2009; Sklar et al., 2008; Venken et al., 2005), *in-silico* analysis was used to predict the effect of R450C mutation on whirlin protein structure to assess its likely pathogenicity. This analysis was carried out by Dr Jonathan Mullins of Swansea University, who used the program, I-Tasser (Roy et al., 2010) to model the R450C variant (Figure 4.5).
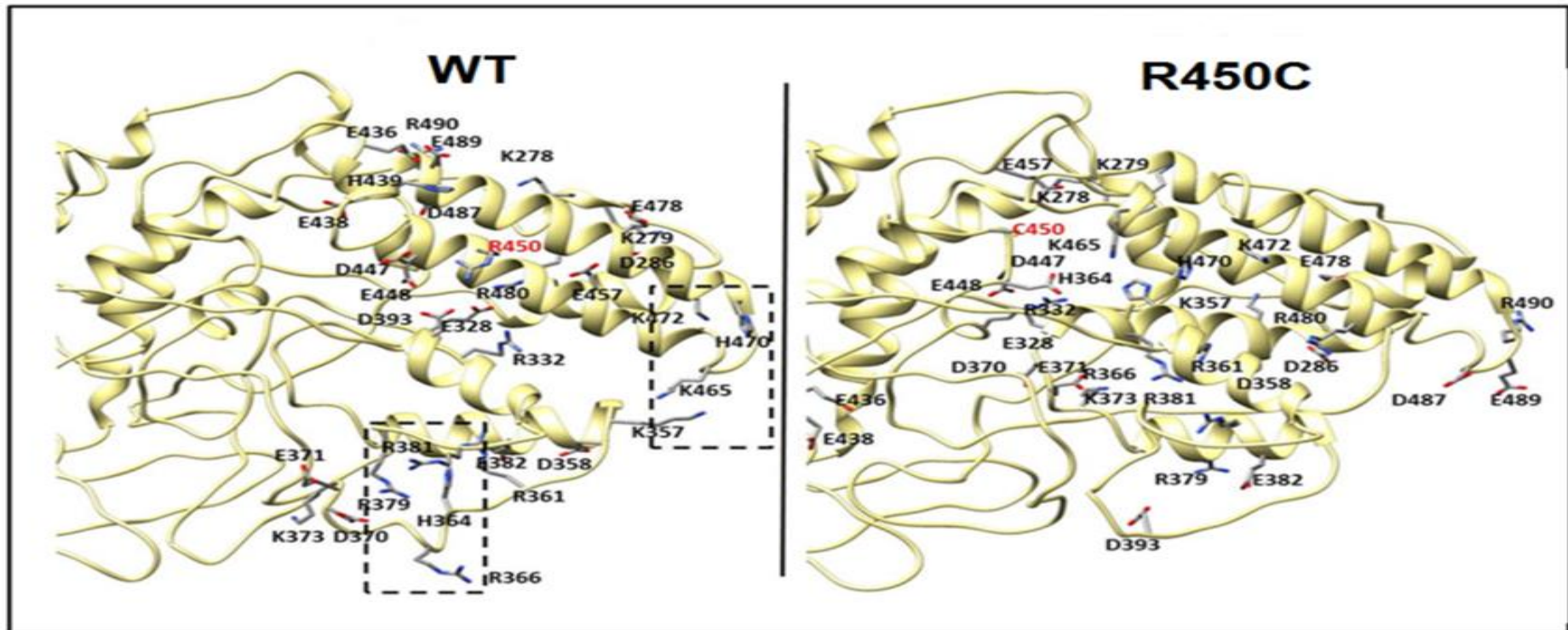
Figure 4.5. Structural modelling of the R450C variant in whirlin. Left: wild type (WT) whirlin protein showing an approximately equal distribution of positively and negatively charged residues around the R450 residue, all in highly accessible positions. Right: whirlin with the R450C mutation, which results in severe disruption of the organisation of adjacent residues and withdrawal from their highly accessible peripheral location. (This image was produced by Dr Jonathan Mullins and is used with his permission).

When the wild type and R450C models of whirlin were compared, the regions surrounding the R450 position were found to undergo substantial conformational rearrangement in the R450C variant, particularly with regard to reorganisation of positively charged outward facing regions that may form interaction surfaces for other proteins to bind with whirlin. In contrast, other parts of the protein remain largely unchanged. The apparent withdrawal of loops from their highly accessible peripheral locations suggests that interactions between this region of whirlin and other proteins are likely to be greatly impaired.

4.2.1.5 Whirlin interaction with UBR4 is disrupted by the R450C mutation

After the structural modelling suggested that the R450C mutation disrupt the binding ability of whirlin, members of the Leeds Psychosis Group contacted a research team in Nijmegen who had published on whirlin protein structure and function. Dr Erwin van Wijk of Radboud University, Nijmegen informed the Leeds group that a yeast two-hybrid assay (Coates and Hall, 2003) had identified UBR4 as interacting with the middle portion of whirlin containing the R450 residue. The Netherlands group investigated this protein-protein interaction assay using constructs containing the middle domain (MD) of both wild-type and R450C mutant whirlin, screened against a human brain cDNA library. An interaction of wild-type whirlin with UBR4 was identified, and found to be weakened by the R450C mutation.

4.2.2 Follow-up work

A number of lines of evidence, both genetic and protein modelling, suggested the involvement of *WHRN* variants in psychoses, and more specifically that the R450C mutation is likely to have an impact on whirlin function. The author therefore carried out further experiments, as detailed below, to investigate this link.

4.2.2.1 Screening more schizophrenia patients for the *WHRN* R450C variant

To increase the sample size for R450C screening and to look for other possible mutations, Sanger sequencing of the key exon 6 of *WHRN* was undertaken (Section 2.9 in Chapter 2) on a further 200 Caucasian schizophrenia patients ascertained and sampled in Glasgow. R450C was not found in these patients. Because Family-7 is of Pakistani ethnicity, it was also important to check the frequency of the R450C variant in the unaffected Pakistani population. Therefore, 154 unaffected Pakistani controls collected in the Leeds/Bradford area were also sequenced. All the tested healthy controls showed normal sequences, and the R450C mutation was not found in any of these samples either. Results are shown in Table 4.2.

Used cohort	Type of sample	Number of samples	Ethnicity
Glasgow (DNA)	Patients	200	Caucasian
Leeds/Bradford (DNA)	Controls	154	Asian

Table 4.2. Screening of extra patient and control samples for *WHRN* R450C. 200 Caucasian schizophrenia patients ascertained in Glasgow and 154 normal Pakistani control samples collected from the Leeds/Bradford area were screened for variants in exon 6 of *WHRN*, including the R450C variant. No variants were found in any of the tested samples.

4.2.2.2 WHRN expression in human tissues

In order to determine which tissues express WHRN, reverse transcriptase PCR was carried out as described in Sections 2.3-2.5 on cDNA prepared from samples of 20 different human tissues. The results showed that WHRN is expressed in a large number of tissues, and is particularly highly expressed in fetal and adult brain (Figure 4.6).
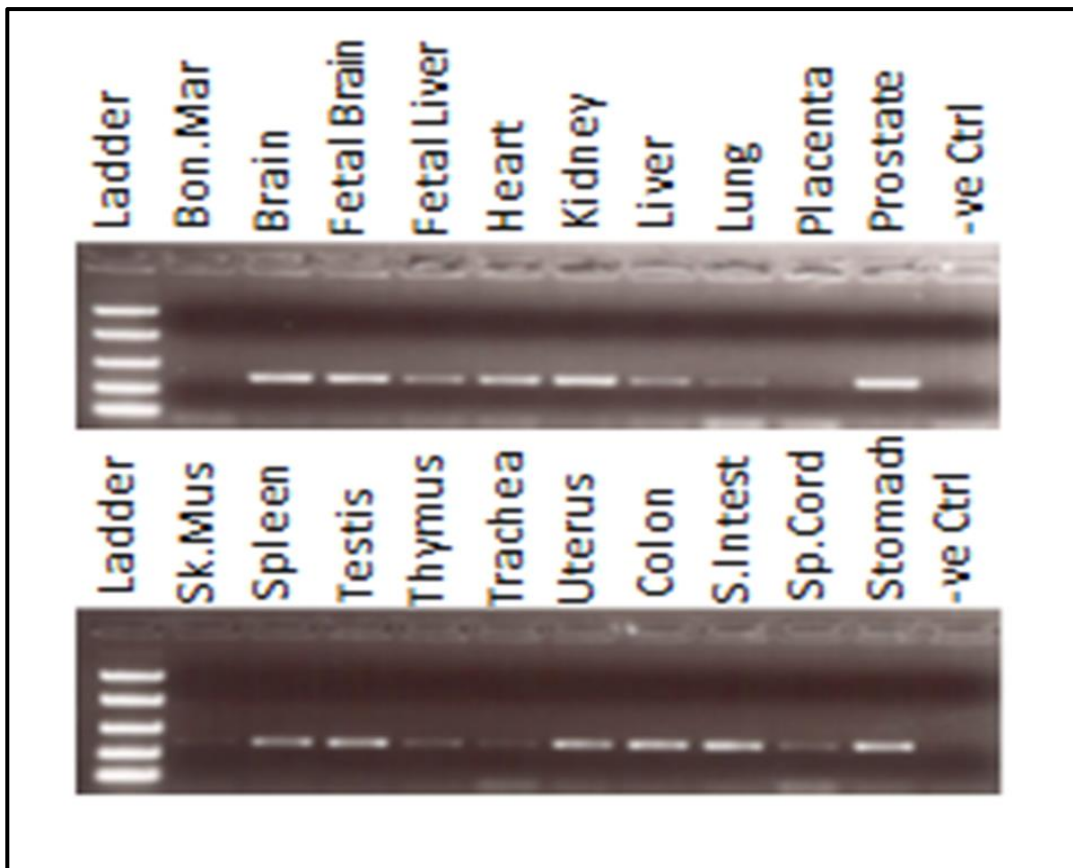
Figure 4.6. Expression of WHRN in human tissues. WHRN expression was assessed in a Human Total RNA Master Panel II (Clontech Laboratories, Inc) containing cDNA from 20 different human tissues. The amplification products were size fractionated on a 1.5% ultra-pure agarose gel (Fisher Scientific, UK), stained with ethidium bromide and visualized with a ChemiDoc™ MP System (Bio-Rad) over ultra-violet (UV) light (wave length 320 nm). The image was focused through Image Lab™ software (Bio-Rad). This analysis revealed that WHRN is expressed in all tissues but at widely varying levels, with particularly clear expression in fetal and adult brain.

4.2.2.3 Site directed mutagenesis

For the purpose of carrying out further functional analyses of the *in vitro* effect of the R450C mutation on subcellular localisation and protein interactions of whirlin, a WHRN construct was obtained as a gift from Dr Erwin van Wijk. This construct contains wild-type human WHRN transcript 3 (UC004bja.4) that has been cloned by Gateway destination cloning (Katzen, 2007) into an ampicillin-resistant pDEST-733 mammalian expression vector (Invitrogen) (Van Wijk Erwin et al., 2006a). The map of the pDEST-733 is shown in Figure 4.7. A monomer Red Fluorescent Protein (mRFP) tag (Campbell et al., 2002) was already fused upstream of the attB1 recombination site of the vector.
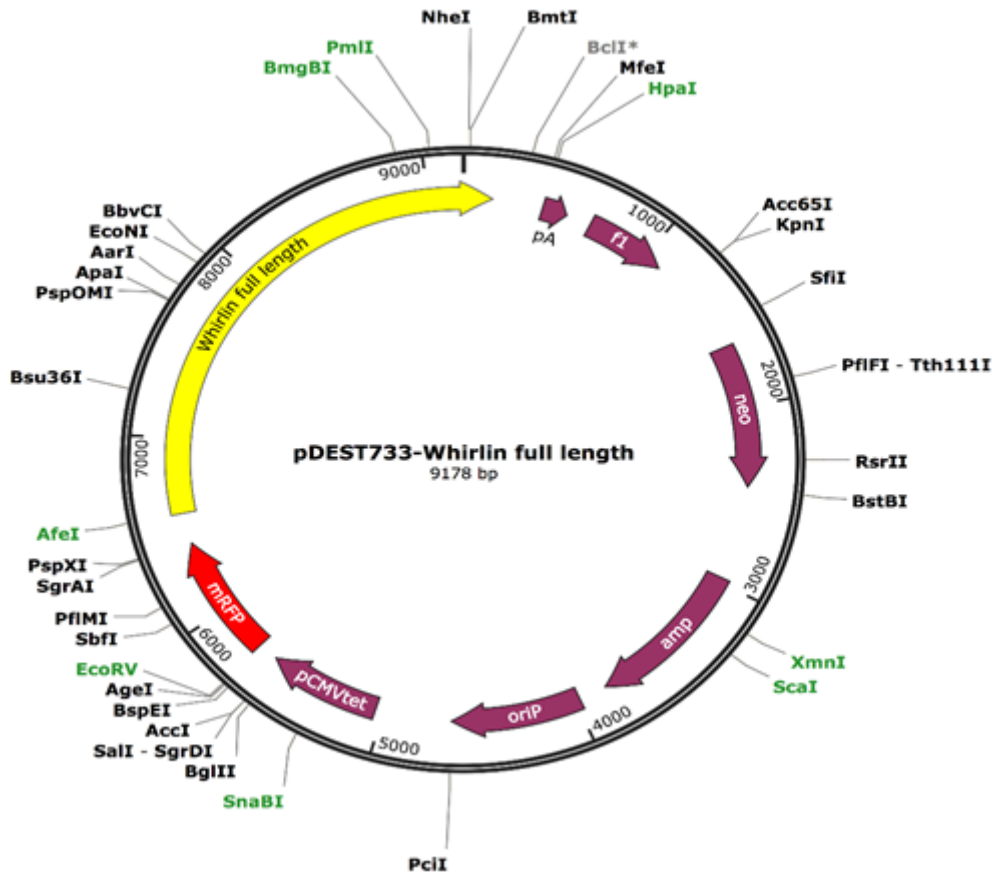
Figure 4.7. Map of pDEST-733 vector. WHRN transcript 3 (2748 bp) was cloned into this vector. The full sequence of the construct, including the mRFP tag, is shown in Appendix 6.

In order to compare the wild-type 450R and mutant 450C proteins, site directed mutagenesis (SDM) was carried out to generate the 450C clone. The mutagenesis experiment was specially designed to target base number 1348 in *WHRN* and change the wild-type "C" nucleotide into a "T" base. The method used is detailed in Figure 4.8, and two 450C clones were generated.
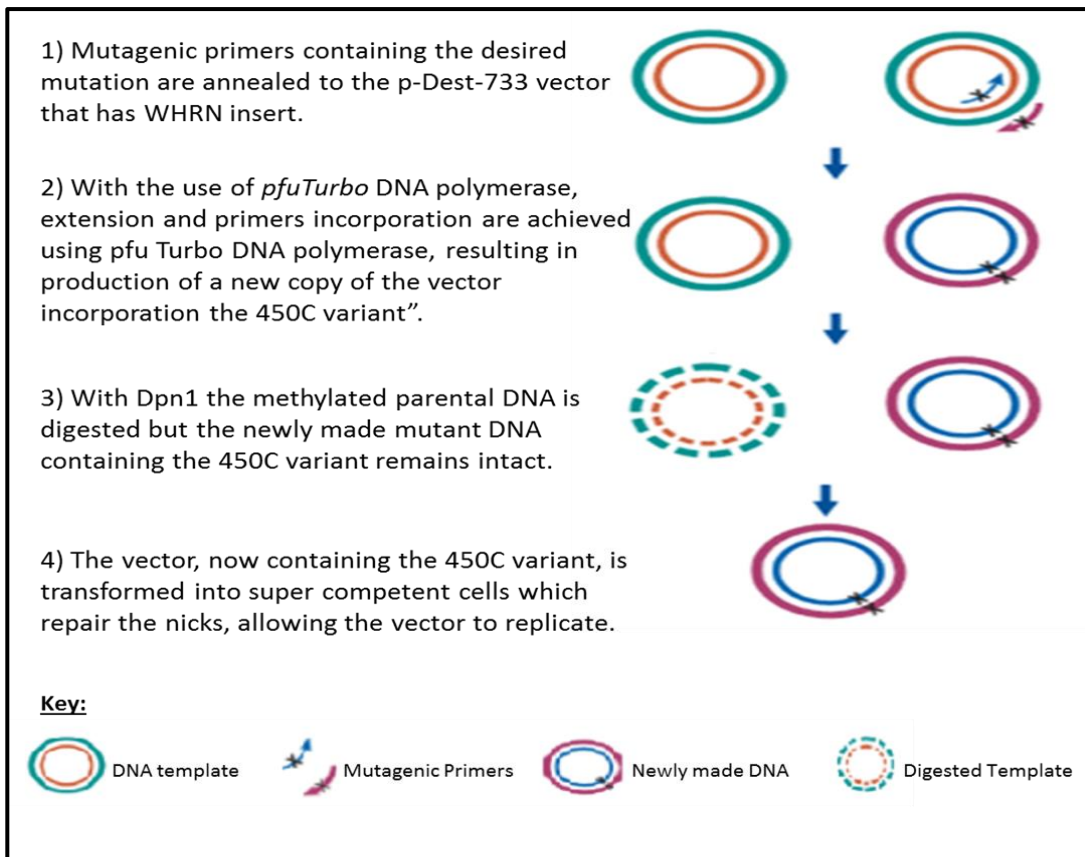
Figure 4.8. The major steps to introduce the 450C mutation into the *WHRN* gene construct. The use of mutagenic primers and *pfuTurbo* DNA polymerase results in the introduction of the change of interest into both stands of the newly-synthesised DNA plasmid. *DpnI*, which recognises GATC, will cleave the sequence only if the "A" is methylated. The *in vitro*-synthesised DNA plasmid (the new plasmid with the mutation) will not be methylated and remains intact, while the parental DNA plasmid originally prepared from a strain of *E. coli* is methylated and will be digested by *DpnI*. For the transformation, XL1-Blue supercompetent cells were used. Because of the increased membrane permeability, these cells facilitate the entry of the plasmid. (Adapted from (Smith, 2007) with permission from Nature Publishing Group, License Number: 4078900153135).

4.2.2.3.1 Confirming SDM by enzymatic digestion

The clones (two wild-type with 450R and two mutants with 450C) were selected for growth on agar plates using ampicillin, as explained in Section 2.14. Plasmids were then extracted and standard PCR was carried out using a pair of primers (5 μM) that flanked the targeted change in *WHRN,* as described in Section 2.14.5. To ensure that the site-directed mutagenesis was successful, *WHRN* cDNA clones were tested for the presence of a restriction fragment length polymorphism

(RFLP). The *ScaI* restriction enzyme cut the mutant 450C clone but not wild-type *WHRN* (Figure 4.9).
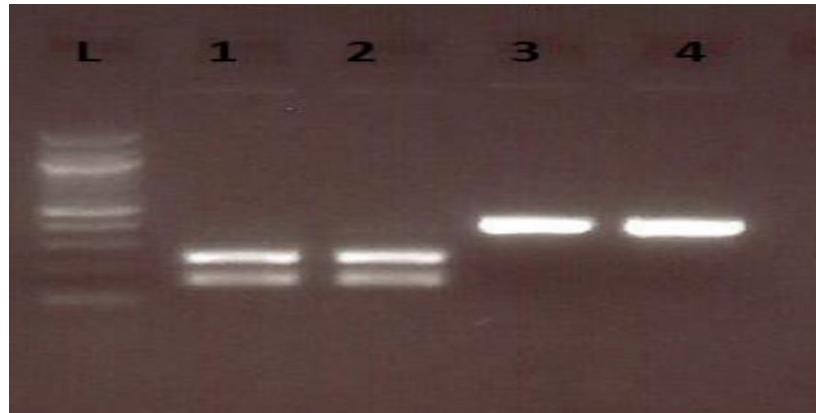
Figure 4.9. *ScaI* digestion of R450C mutant *WHRN*. 'L' indicates is a 100-bp ladder. Lanes 1 and 2 are the mutated R450C *WHRN* and lanes 3 and 4 are wild-type (450R) *WHRN. Sca*I recognises the AGTACT sequence and cuts the mutated product, giving two bands of 238 bp and 157 bp (lanes 1 and 2). The wild-type was not digested, leaving the 395 bp band intact (lanes 3 and 4).

4.2.2.3.2 Confirmation of SDM by Sanger sequencing

Sanger sequencing was carried out as described in Section 2.9, not only to confirm that the SDM was successful but also to find out if there had been any other mutations introduced during the SDM process. It was confirmed that the required change had been inserted at the correct location (Figure 4.10) and that no other change was present.
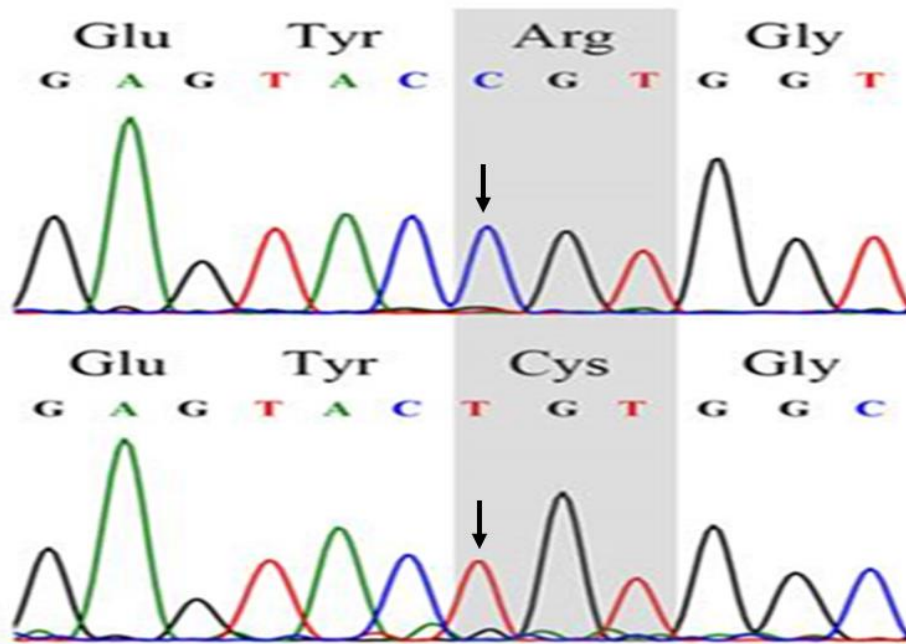
Figure 4.10. Sequencing of the site of the targeted R450C mutation in *WHRN*. The upper sequence is the wild-type and the lower sequence is the mutated *WHRN*. The wild-type blue "C" base has been changed to red "T", indicated by the arrows. Apart from the deliberately introduced change (R450C), the full insert was found to match the normal wild-type sequence.

4.2.2.4 Whirlin over-expression and its effect on UBR4

Transfection studies were undertaken to see whether the presence of the 450C mutation in whirlin affected the cellular localisation of the tagged protein when it was overexpressed in a human neuroblastoma cell line (SH-SY5Y), as described in Section 2.14.3. This experiment also tested whether the localisation of endogenous UBR4 was altered by overexpression of the mutant versus the wild-type construct. Based on independent analyses by two different individuals, the transfection studies showed no localisation difference between the 450C mutant and wild-type whirlin. It was also found that strong co-localisation existed between whirlin and UBR4 (Figure 4.11).
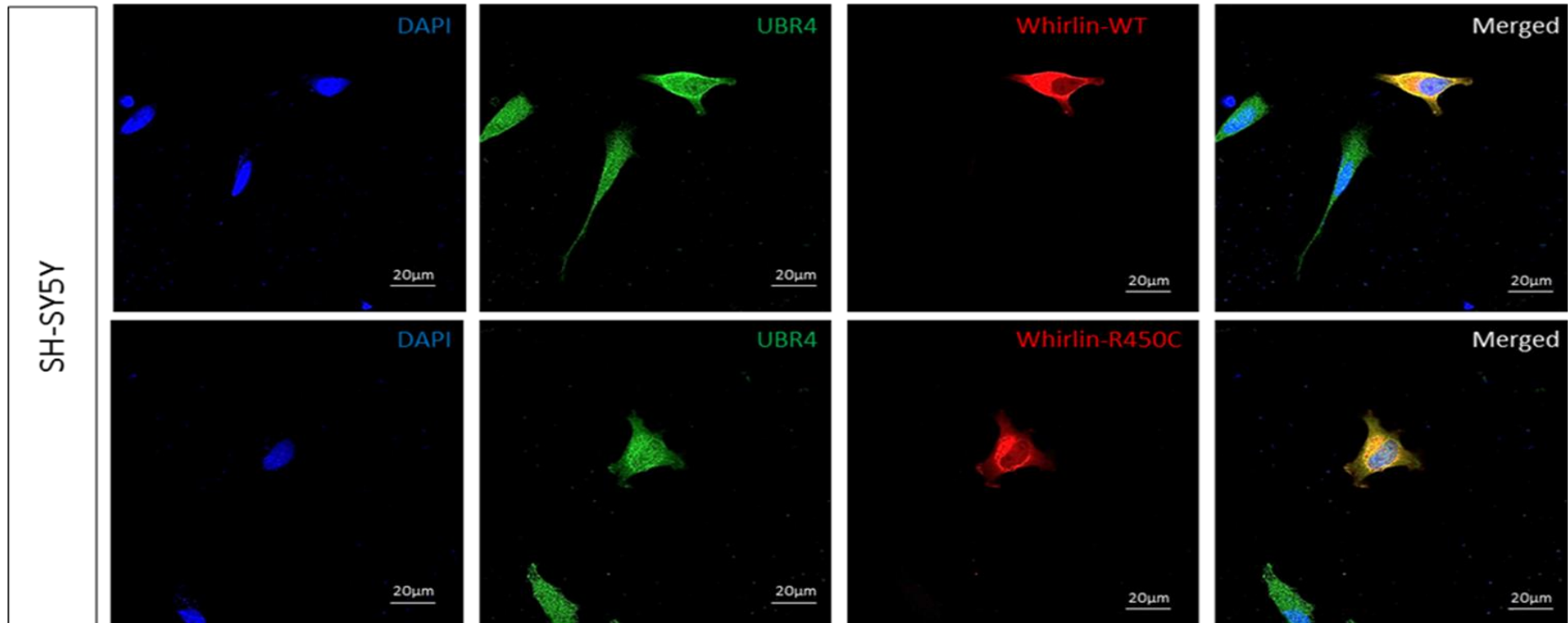
Figure 4.11. Co-localisation of whirlin and UBR4. Thirty SH-SY5Y cells transfected with wild-type (WT) or mutant 450C whirlin were captured and independently analysed by two individuals. SH-SY5Y cells transfected with wild-type whirlin (upper panel) or R450C mutant whirlin (lower panel) construct with a red fluorescent tag are shown. The nuclei are stained blue with DAPI stain and endogenous UBR4 is stained in green. Both WT and mutant whirlin show consistent co-localisation with the green UBR4.

The hypothesis proposed as a result of the findings of the Nijmegen group were that the 450C mutation weakens but does not abolish binding between whirlin and UBR4. It might not be possible to detect a subtle difference in UBR4 co-localisation by simple visual examination of cells transfected with wild-type or mutant whirlin. Therefore, the ratio of UBR4-whirlin co-localisation to non-co-localisation in SH-SY5Y cells transfected with either the WT or 450C mutant whirlin constructs was assessed using Fiji ImageJ (Schindelin et al., 2012) software. This software allows the user to measure each pattern of cellular fluorescence separately and then calculate the extent to which two different signals overlap, in order to obtain a measure of the extent of co-localisation in the form of a Pearson R square value (Mukaka, 2012). The analysis revealed that whirlin and UBR4 co-localisation is strong in the perinuclear area but weaker at the membrane. Co-localisation values for UBR4 with wild type and 450C mutant whirlin are given in Table 4.3. The mean co-localisation values and significance value are shown in Figure 4.12.

Image No	Pearson R square WT	Pearson R square R450C
1	0.78	0.49
2	0.78	0.61
3	0.56	0.52
4	0.48	0.65
5	0.83	0.46
6	0.63	0.56
7	0.59	0.55
8	0.6	0.62
9	0.8	0.63
10	0.71	0.6
11	0.62	0.46
12	0.5	0.59
13	0.53	0.6
14	0.55	0.61
15	0.54	0.58
Mean	0.633333333	0.568666667
STD	0.112999508	0.058636356
t-test	0.067694391	

Table 4.3. Co-localisation of tagged whirlin and endogenous UBR4 in SH-SY5Y cells overexpressing WT or mutant 450C tagged whirlin. Thirty cells, half transfected with wild-type (WT) whirlin and the other half transfected with 450C mutant (R450C) whirlin, were assessed. Co-localisation with UBR4 was measured using the Pearson R square values in Fiji ImageJ software. The results suggest a difference, as shown in Figure 5.10.
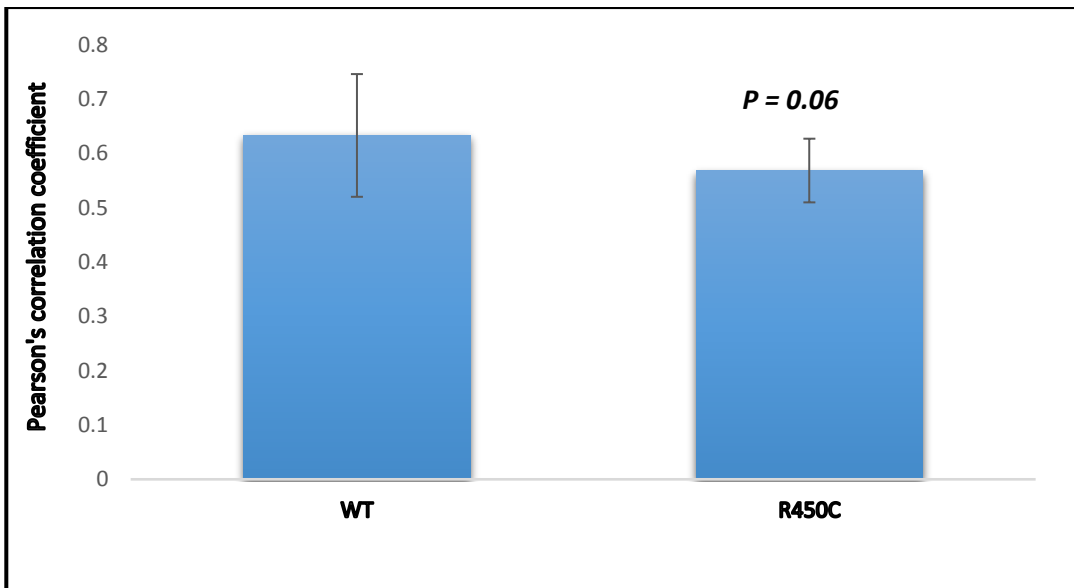
Figure 4.12. Pearson's correlation coefficient measurement of whirlin and UBR4 co-localisation in whirlin construct-transfected SH-SY5Y cells. The extent of co-localisation between endogenous UBR4 and overexpressed tagged wild-type (WT) or 450C mutant (R450C) whirlin. The difference in co-localisation is not significant (P = 0.067).

4.2.2.5 Further testing to investigate that whirlin interacts with UBR4

Yeast two-hybrid (Y2H) analyses carried out by colleagues in Nijmegen (described in Section 4.2.1.5) suggested that whirlin interacts with UBR4. Furthermore, comparison of co-localisation of WT and mutant 450C whirlin with endogenous UBR4 in SH-SY5Y cells suggested a trend of some reduction in UBR4 co-localisation with mutant 450C whirlin, though this difference did not reach statistical significance. It should be, however, mentioned that Y2H can sometimes give false positives that could result due to various reasons including nonspecific interactions, inefficient filtering criteria and considering the test outcome with out carrying out replicates under alternative conditions (Serebriiskii and Golemis, 2001). The overexpression of baits/preys in Y2H is potential for nonspecific interaction and should be considered. Excluding the proteins that normally show multiple interactions, such as chaperones, and carrying out replicates using different settings, such as using different reporters, are also crucial to minimise false positive results in Y2H (Koegl and Uetz, 2007). In order to provide further independent evidence of WHRN/UBR4 interaction, a mouse

whirlin antibody was used to pull down whirlin from mouse brain lysate, then the resulting protein complex was tested to determine whether it contained Ubr4.

4.2.2.5.1 Checking the whirlin antibody

Protein extracts were prepared from brain hemispheres from three different wild-type mice. These were subjected to western blotting to test the ability of the whirlin antibody to detect whirlin protein. Western blot analysis was carried out using whirlin antibody (1 in 200 dilution) and it showed the expected band size of 97-kDa (Figure 4.13).
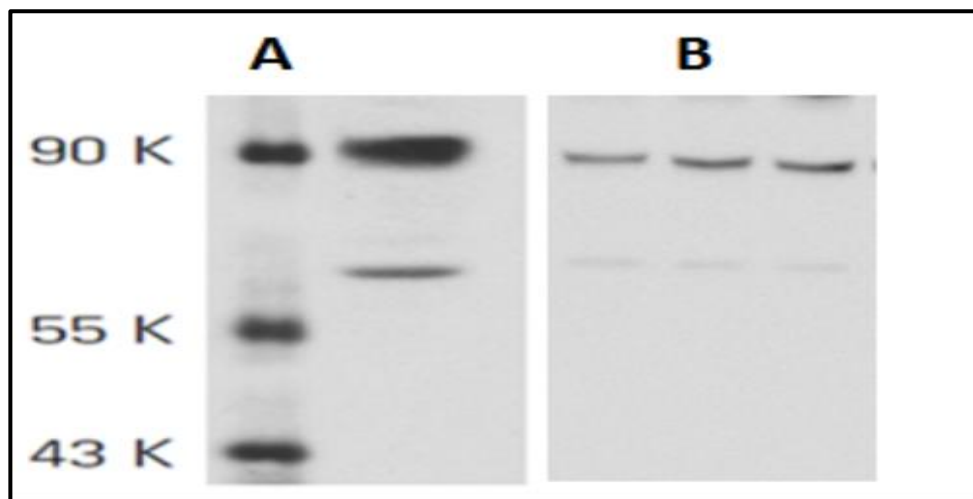
Figure 4.13. Western blot of whirlin in mouse brain. A. Image taken from whirlin antibody datasheet (Santa Cruz, sc-49787) showing a marker as well as the expected 97-kDa band of whirlin. **B.** Image of western blot of three samples showing the corresponding whirlin bands. A lower molecular weight band appears in each of the samples and is also present in the image from the whirlin antibody datasheet. This could indicate a shorter isoform of the protein.

4.2.2.5.2 Peptide blocking check

As a control for the specificity of the whirlin antibody (Santa Cruz, sc-49787), western blotting was carried out on two of the previously tested mouse brain hemispheres. The experiment was started by performing whirlin antibody blocking using the specific peptide against which it was raised, referred to as Santa Cruz, sc-49787P, as described in Section 2.16.3. Western blotting was then done on the two samples using blocked and unblocked whirlin antibody. The

absence of bands when the antibody was peptide-blocked and the appearance of bands with the normal unblocked antibody (Figure 4.14) indicated the good quality and specificity of the antibody used.
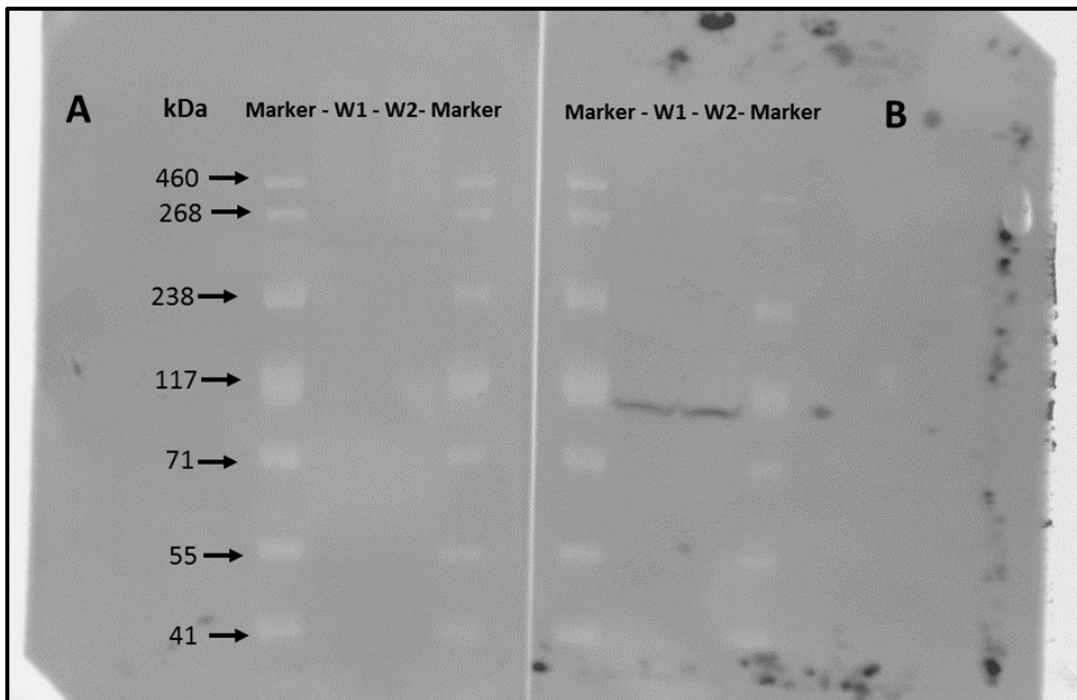
Figure 4.14. Peptide blocking to confirm the specificity of the whirlin antibody. A. Western blot with size marker Hi-Mark pre-stained HMW protein standard (Invitrogen) in the first and last lanes, in between two samples of mouse brain extract (WT1 and WT2) were loaded and size fractionated. This was then probed with whirlin antibody after peptide blocking. **B.** An identically loaded western blot membrane probed with the normal unblocked whirlin antibody showing bands of the expected size for whirlin (~97-kDa). The failure of the peptide-blocked antibody to detect whirlin and its successful detection when the antibody was not blocked suggests that the antibody used has good specificity.

4.2.2.5.3 Detection of β-Actin

In order to exclude the possibility that the disappearance of bands on the membrane probed with peptide-blocked whirlin antibody was not due to a loading artefact or a difference in the samples or methods used to prepare the western blots, both membranes were stripped and stained for mouse β-Actin (Sigma, A5441) (Figure 4.15). A band of the expected size and of equal intensity was detected in both membranes.
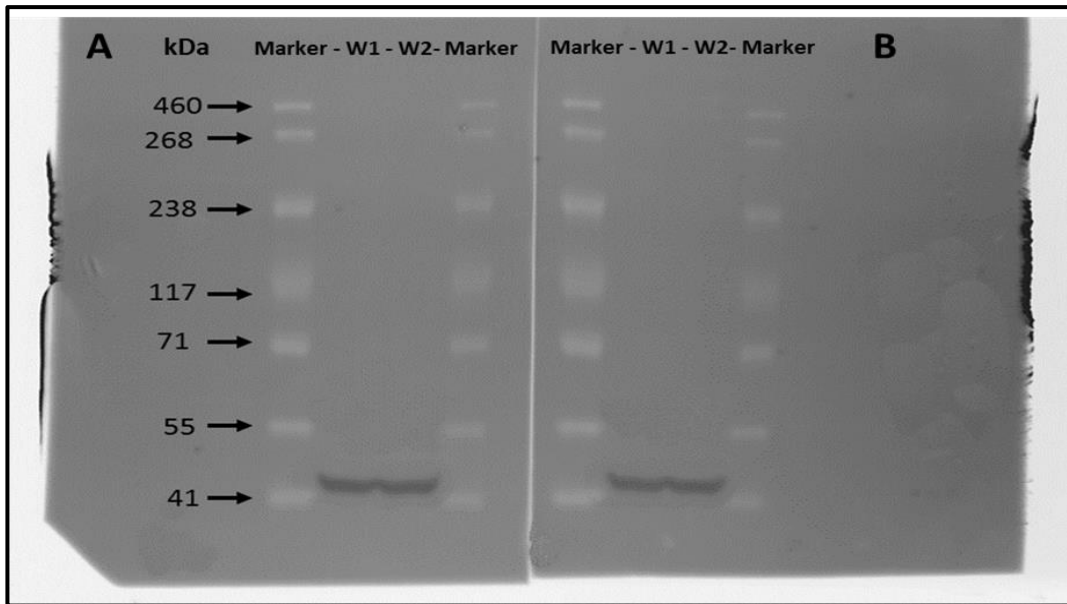
Figure 4.15. Detection of β-Actin as a loading control. The same membranes shown in Figure 4.14 after stripping with antibody stripping solution (Section 2.16.6) and then staining for β-Actin (1:500, Sigma, A5441). Both membranes showed bands of the expected size for β-Actin (42-kDa).

4.2.2.6 Co-immunoprecipitation to investigate Whrn and Ubr4 interaction

Co-immunoprecipitation analysis was carried out in an attempt to further investigate Whrn/Ubr4 interaction in mouse brain lysates (Section 2.15.1). As a control, another antibody (Munc18) was also used in the same assay. Munc18 is a member of the Sec1/Munc18 protein family and is a neuron-specific protein of 67-kDa that is abundant in the brain (Kalidas et al., 2000). It has been found that Munc18 interacts with syntaxin-1 and enhances its stability (Toonen et al., 2005). The antibody for Munc18 protein has been tested and verified previously by Dachtler et al. (2015). There were no expectations that this protein will interact with Ubr4 and it was therefore included as a control in this experiment.

Whirlin antibody and Munc18 antibody (both 1:200 dilution) were attached separately to magnetic beads. During incubation with purified brain samples, each antibody was expected to bind specifically to its corresponding protein. Before eluting the materials from the beads, the complex of antibodies binding to

their proteins was washed to remove the unbound protein portion. This material was saved and called "unbound". The proteins, which were bound to their antibodies, were then eluted from the beads as described in Section 2.16.3. The material eluted from the beads of whirlin antibody was called "whirlin bound" and the material eluted from the beads of Munc18 antibody was called "Munc18 bound". Beside the bound and unbound materials, the experiment also included the starting material of the samples, which was an untreated mouse brain lysate expected to have all the protein contents. The buffer (PBST), used in the washing steps, was also included in the test to give an indication if amount of the whirlin bound got lost during the washing steps. The elutes were then processed for western blot (2.16) followed by immunostaining for Ubr4 using 1:200 rabbit anti-Ubr4 antibody (Abcam, ab86738)

Although a light band of 574-kDa, which is equivalent to Ubr4 size, was observed when the WB membrane was viewed under the UV, the printed image hardly show anything in that position. Furthermore, the appearance nonspecific bands suggests poor quality of Ubr4 antibody. Therefore, it has not been possible to draw a firm conclusion based on the obtained image (Figure 4.16). The full experiment, therefore, needs to be repeated after getting a better Ubr4 antibody and testing its specificity though peptide blocking assay. Another possible experiment would be to use a cell line that does not have endogenous WHRN and UBR4, such as HEK293, to transfect human WT WHRN and UBR4 into the cells and do immunoprecipitation while, in parallel transfecting human mutant WHRN and UBR4 into other cells and doing pull-down immunoprecipitation. If our hypothesis, that WHRN/UBR4 interaction is present but disturbed with mutated WHRN, is correct, then a normal UBR4 band will be detected in the first case but the intensity of the band will be smaller in the second case. Such finding will, therefore, support the raised hypothesis and would also point out that R450C is probably a functional/pathological mutation.
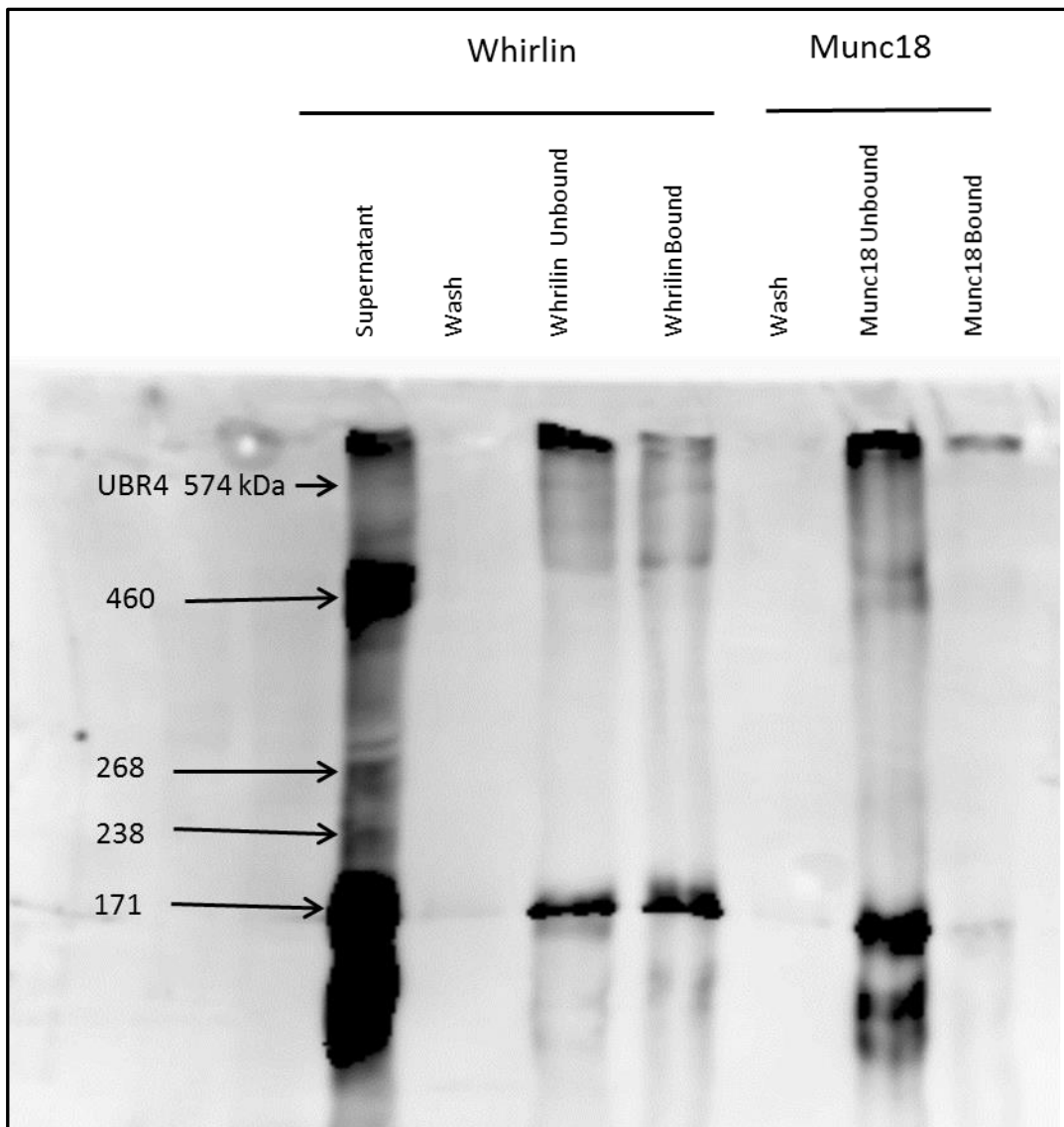
Figure 4.16. Testing Ub4 in protein pulled-down by whirlin antibody from mouse brain lysate. The lane labelled "Supernatant" is the untreated mouse brain lysate which forms the starting material. The lane labelled "wash" is the PBST used to wash the material that was not captured by the antibody in the beads. The "unbound" lane contains the material that was not held by the antibody bound to beads, and "bound" is the lysate generated from the beads holding the antibody. A Ubr4 band of the expected size was not clearly detected and the appearance of non-specific bands argues about the quality of the used Ubr4 antibody. No confirmed result has been considered and this experiment would, therefore, need further validation and repeating.

4.3 Family-8: a Mendelian allele for schizophrenia on chromosome 13q?

4.3.1 Preliminary work

Family-8 was ascertained by Dr Tariq Mahmood after affected family members attended clinics and respite care at the Becklin Centre, St James's University Hospital, Leeds. The initial genetic analyses described below were undertaken by Prof. Chris Inglehearn (homozygosity mapping) and Dr Jose Ivorra-Martinez (whole exome sequencing).

4.3.1.1 Family description

A large extended consanguineous family (Figure 4.17) with multiple individuals diagnosed with schizophrenia by DSM-IV criteria was identified and ascertained when family members attended clinics in the Becklin Centre, St James's Hospital, Leeds. Six affected members of the fourth generation had been monitored for up to 15 years by Dr Mahmood. The family consisted of four nuclear families, each formed by intermarriages of first cousins from the same two related families in the previous generation. Three of these nuclear families live in Yorkshire and the fourth in northern Pakistan. The six affected individuals in the fourth generation were diagnosed with schizophrenia and are indicated by the filled symbols in Figure 5.15. This pedigree structure is highly suggestive of recessive inheritance for an allele causing or predisposing to schizophrenia.
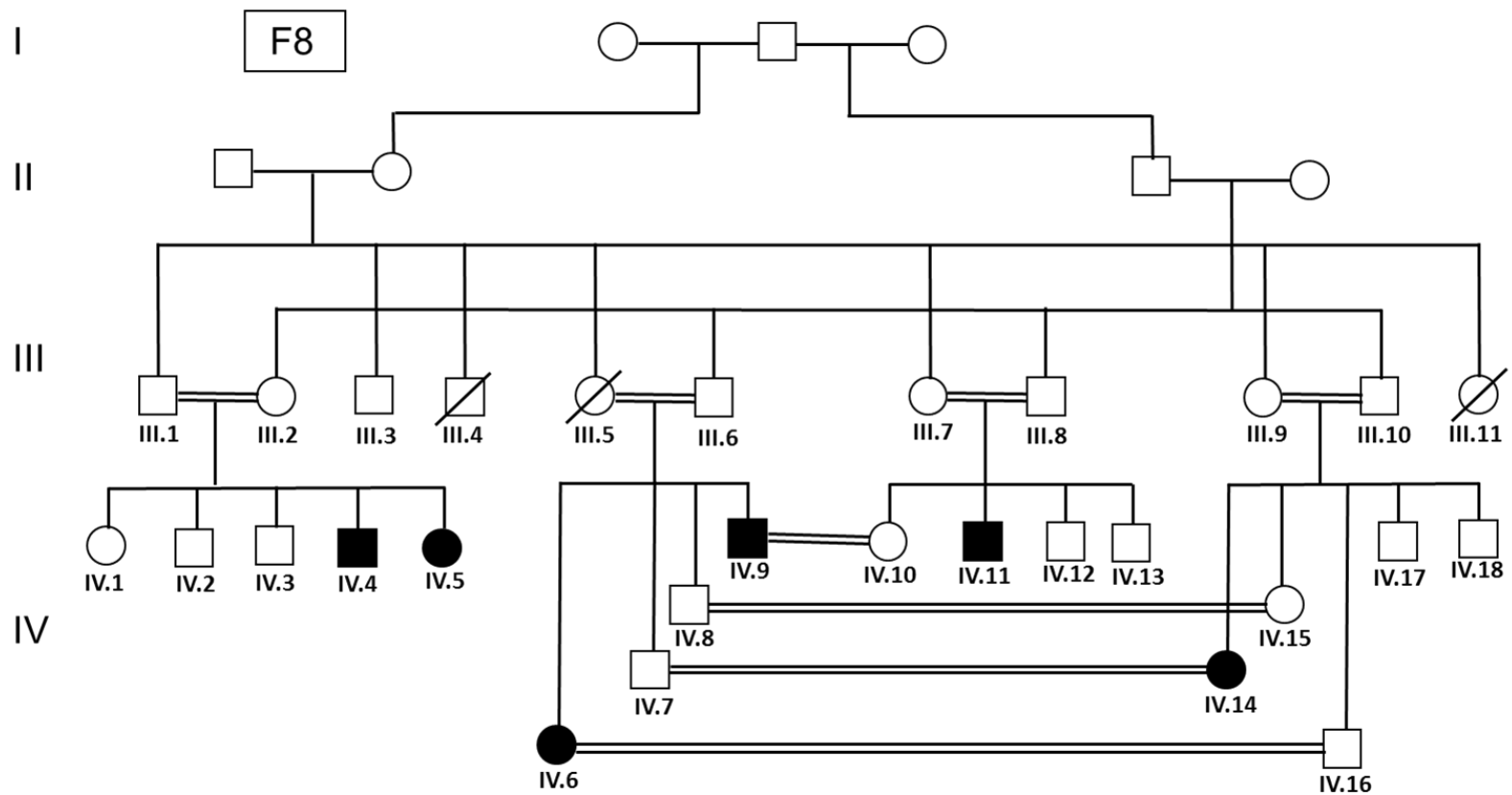
Figure 4.17. Pedigree structure of Family-8. A large family with four consanguineous marriages between first-cousins in the third generation and six cases of schizophrenia. One of the consanguineous families (marriage between III.1 and III.2) lives in northern Pakistan but the other three live in Yorkshire. The pattern of inheritance in this family suggests an autosomal recessive mode.

The onset of the disorder in the six affected individuals was in the late teens and early twenties. At the beginning, these individuals had positive symptoms such as auditory hallucinations, paranoid delusions and disordered thoughts. However, in later stages they developed negative symptoms, and cognitive decline was evident in cases IV.4, IV.5, IV.6 and IV.9. The three males have shown marked sensitivity to cannabis, one reported to become extremely thought disordered with minute quantities of cannabis. The females have not been exposed to cannabis. One affected female (IV.14) has shown some affective features such as grandiose ideas in post-natal relapses. Four of the affected cases are being treated with and are responsive to clozapine. Table 4.4 summarises the clinical findings in this family and the drug treatments used.

Case ID	Sex	Age of onset	Delusions	Auditory hallucinations	Thought disorder	Grandiose ideas	Negative symptoms	Drug treatment
IV.4	M	19 y	+	+	n/a	n/a	+	Clozapine
IV.5	F	23 y	+	+	n/a	n/a	+	Clozapine
IV.6	F	17 y	+	n/a	n/a	n/a	n/a	Olanzapine
IV.9	M	22 y	+	+	+	n/a	++	Clozapine
IV.11	M	21 y	+	+	++	n/a	+	Clozapine
IV.14	F	23 y	+	n/a	n/a	+	n/a	Quetiapine

Table 4.4. Symptoms of the affected cases in Family-8. Symptoms and current treatments for the six affected individuals diagnosed with schizophrenia from Family-8. n/a = not available.

4.3.1.2 Homozygosity mapping results

Homozygosity mapping of this family was carried out using SNP genotypes generated on an Affymetrix 6.0 Whole genome SNP array. DNA from four affected family members, IV4, IV.5, IV.9 and IV.11, was sent to AROS Applied Biotechnology (Aarhus, Denmark) for genotyping. Data were analysed using the Agile ideogram software available via http://dna.leeds.ac.uk/autozygosity. Results revealed only one homozygous region shared between these affected individuals, on chromosome 13q, shown in Figure 4.18. It was therefore hypothesised that a gene in this region would be found to carry a mutation of recessive effect causing or predisposing to schizophrenia.
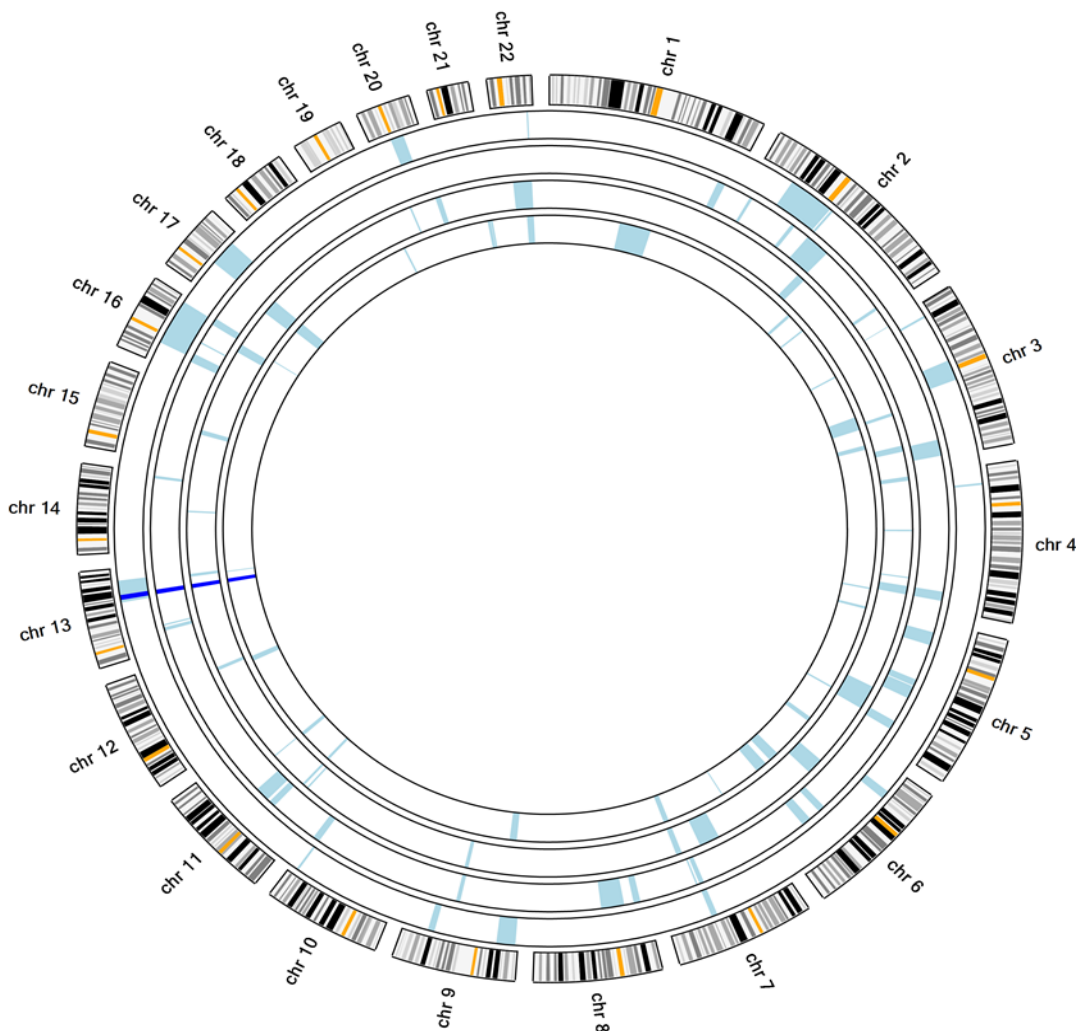
Figure 4.18. Shared homozygous region in Family-8. Ideogram plotting SNP data from the four affected genotyped members of Family-8. The outer circle contains template segments of the 22 autosomal chromosomes. The four inner circles plot the SNP data from affected subjects IV.4, IV.5, IV.9 and IV.11. The light blue areas are the homozygous regions in each individual and the dark blue region on chromosome 13 is

the homozygous region shared between all four individuals. This region, chr13:75367473-81504615, measures about 5 Mb.

The relationship between the chromosome13q region and neuropsychiatric disorders is not new. Early evidence from a linkage study of 540 individuals with bipolar disorder from 97 families, assessed by National Institute of Mental Health (NIMH) Diagnostic Centres, suggested 13q32, besides two more regions on chromosome Xp, as interesting regions that need further investigation (Stine et al., 1997). Subsequently, further support has come from linkage studies by Detera-Wadleigh Sevilla D et al. (1999), Badenhop et al. (2001), Badenhop et al. (2002) and Ferraren et al. (2005). Association studies have also provided evidence implicating 13q in BPD. For example, (Hattori et al., 2003) found an association between the *G72*/*G30* gene locus, located at 13q33, and bipolar disorder, after performing transmission/disequilibrium testing (TDT) and haplotype analysis on two independent pedigree sets, from Clinical Neurogenetics (CNG) (22 families) and the National Institute of Mental Health (NIMH) Genetics Initiative (152 families). A 65-kb genomic region at 13q32-33 had already been associated with schizophrenia (Chumakov et al., 2002). The 13q32 locus has, therefore, been a potential locus of susceptibility for both bipolar disorder and schizophrenia (Berrettini, 2003).

4.3.1.3 Exome Sequencing in Family-8

Whole exome next generation sequencing was carried out on subject IV.6 using paired end reads on an Illumina HiSeq 2500 (Section 2.12). The shared homozgous region contains some likely neuronal candidates, including *CLN5,* which is associated with neuronal ceroid lipofuscinoses (NCL), also known as Batten disease (Williams et al., 2006); *SLAIN*, which is implicated in susceptibility to schizophrenia (Curtis, 2016a) and intellectual disability (Harripaul et al., 2017); and *SPRY2*, which has a known regulatory role in neurogenesis (Dow et al., 2015). However, after doing all the exome analysis, combining it with the result of homozygosity mapping and applying the filtering criteria, no obvious candidate variant was identified.

4.3.2 Follow-up work

After the finding of a homozygous region on chromosome 13q, followed by exome sequencing that identified a list of candidate variants in the region but was unable to highlight any obvious causative variant, the author undertook follow-up work in the search for an underlying causative recessive allele in Family-8, as detailed below.

4.3.2.1 13q region under investigation

Because no obvious candidate variant was identified by whole exome analysis, it was hypothesised that a variant could be in a non-coding region such as a promoter, UTR, intron or intergenic region. Therefore, all the genes located in the region (chr13:76,482,752-82,585,964) were interrogated by genome sequencing. Table 4.5 lists the RefSeq annotated coding genes located in the 13q region. A full list, including the pseudogenes and microRNA genes, is shown in Appendix 10

No	Start	Stop	Symbol	Description
1	77454304	77460540	*KCTD12*	Potassium channel tetramerization domain containing 12
2	77522694	77532777	*ACOD1*	Aconitate decarboxylase 1
3	77566059	77576652	*CLN5*	Ceroid-lipofuscinosis, neuronal 5
4	77579389	77601331	*FBXL3*	F-box and leucine rich repeat protein 3
5	77618792	77901177	*MYCBP2*	MYC binding protein 2, E3 ubiquitin protein ligase
6	78109809	78219398	*SCEL*	Sciellin
7	78271989	78338377	*SLAIN1*	SLAIN motif family member 1
8	78469616	78549664	*EDNRB*	Endothelin receptor type B
9	79173227	79177695	*POU4F1*	POU class 4 homeobox 1
10	79188421	79233314	*RNF219*	Ring finger protein 219
11	79885962	79980393	*RBM26*	RNA binding motif protein 26
12	80055259	80130212	*NDFIP2*	Nedd4 family interacting protein 2
13	80910111	80915086	*SPRY2*	sprouty RTK signaling antagonist 2

Table 4.5. Fully annotated RefSeq genes located in the 13q region implicated by homozygosity mapping in Family-8.

Non-coding variants missed in the whole exome sequencing could have many possible consequences. One possible effect might be to decrease or increase the transcription of one or more of the 13 genes in the region, leading to a change in the level of one or more mRNAs from these genes in the brain. Such variants might be more likely to be in promotor regions but could be in 5' or 3' UTRs, introns or intergenic DNA. Another possible consequence, particularly of intronic variants, is that splicing might be altered to include a cryptic exon or cause exon skipping. This would manifest as variation in the sequence and possibly affect the length of the mRNA from one or more of these 13 genes.

To screen for both of these possibilities, RT-PCR was carried out on each of the genes located in the 13q31 locus. PCR of a near full-length mRNA could reveal length variants implying mis-splicing, while semi-quantitative RT-PCR (Section 2.3) could reveal large differences in the level of one or more RNAs. Ideally, this work would be carried out on patient brain tissue, but this was not available. Fresh lymphocytes or fibroblasts can be obtained from the patients, so a preliminary test on control samples was done to assess whether transcripts could be amplified from these tissues. Sets of primers (Appendix 2) were designed as described in Section 2.4. RT-PCR products were ran through 1.5% agarose gels and visualised over ultra-violet (UV) light (Figure 4.19). It was found that cDNA from lymphocytes and fibroblasts were equally likely to express the genes in the 13q31 locus. Based on this experiment, blood lymphocytes were selected as a source of cDNA for transcript analysis.
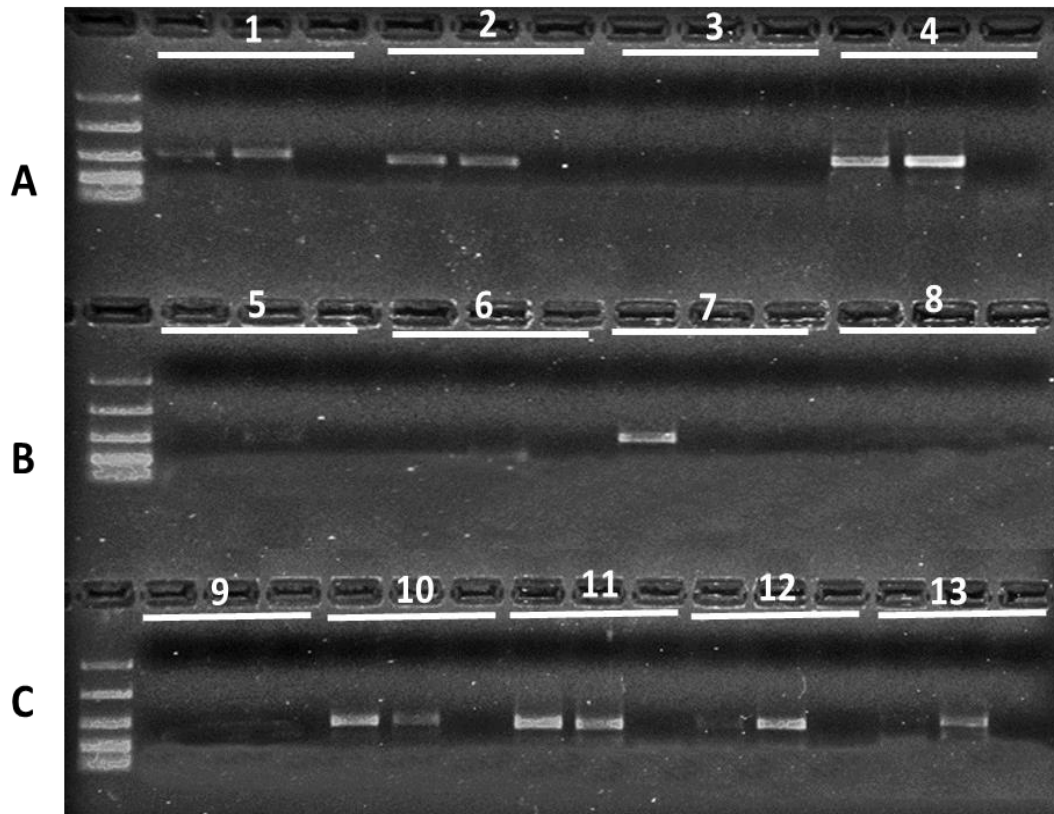
Figure 4.19. Reverse transcriptase RT-PCR to assess whether transcripts in the 13q region could be amplified from lymphocytes or fibroblasts. A. 1, *KCTD12*; 2, *ACOD1*; 3, *CLN5*; 4, *FBXL3*. **B.** 5, *MYCBP2*; 6, *SCEL*; 7, *SLAIN1*; 8, *EDNRB*. **C.** 9, *POU4F1*; 10, *RNF219*; 11, *RBM26*; 12, *NDFIP2*; 13, *SPRY2*. Each row has a 100-bp marker in the first lane. For each gene, 1st lane, lymphocyte cDNA; 2nd lane, fibroblast cDNA; 3rd lane, ddH2O as a negative control.

Since RT-PCR did not amplify any band for some of the tested genes (such as *CLN5*, *MYCBP2*, and *SCEL*), while those bands present were rather weak, RT-PCR with Hot-Shot master mix was tried on cDNA from blood lymphocytes, as described in Section 2.3. Figure 4.20 illustrates the amplified bands after visualisation over ultra-violet (UV) light. The HotShot master mix resulted in clearer, more intense bands. Furthermore, *CLN5*, which did not work in the previous experiment, now showed a band.
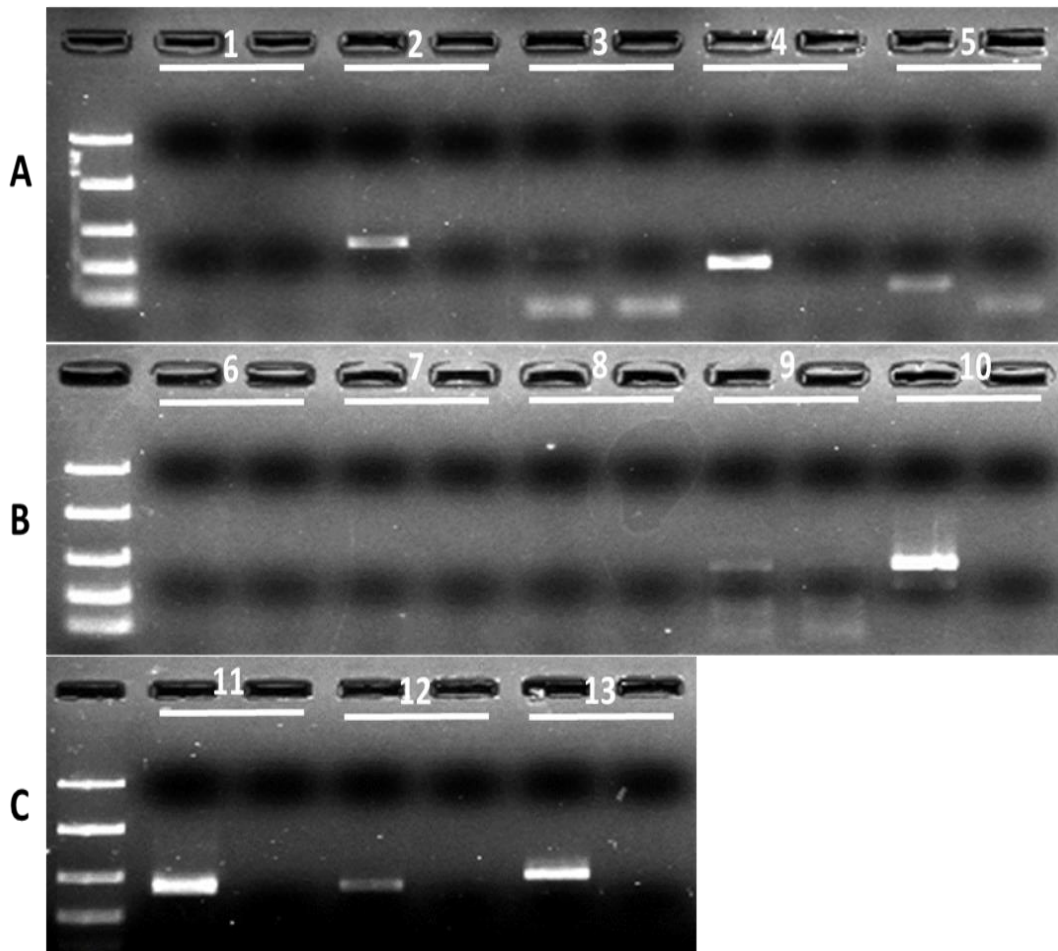
Figure 4.20. Reverse transcriptase-PCR with HotShot master mix. For each gene, 1st lane, lymphoblast cDNA; 2nd lane, negative control. The genes were loaded in the same order as in Figure 4.17. Most of the transcripts could be detected in blood but there were still some that could not be amplified.

4.4 Discussion

Investigating the causes of genetically complex neuropsychiatric disorders is a challenging task. Individuals with schizophrenia have abnormalities in several brain regions and the condition is extremely heterogeneous, with both genetic and environmental factors influencing clinical outcome. Some neuropsychiatric disorders, however, follow a Mendelian pattern of inheritance (Huang Yue et al., 2014), such that their underlying genetic causes could be determined relatively simply using family studies. The pedigree structures of Family-7 and Family-8 described here suggested the possibility that a Mendelian allele might account for inheritance of schizophrenia in them; if true, this might mean that a similar approach in these families could reveal new insights into schizophrenia.

4.4.1 Putative involvement of a *WHRN* variant in schizophrenia in Family-7

This family has two affected males with six unaffected sisters and unaffected parents. At first glance the pedigree might suggest either autosomal recessive or X-linked inheritance. Genetic testing did identify a candidate variant in the *TENM1* gene on the X chromosome present in both affected brothers, but segregation analyses excluded the *TENM1* variant (analysis by Dr Jose Ivorra Martinez). The alternative hypothesis is that the disorder is the result of an autosomal recessive allele that is homozygous in the affected brothers.

The initial work of a colleague (Dr Jose Ivorra Martinez) implicated *WHRN* variant in schizophrenia in Family-7 (see Section 4.2.1.2). Homozygosity mapping in the affected individuals identified two homozygous regions, and within these a missense mutation, c.C1348T:p.R450C in the *WHRN* gene encoding whirlin, was found to segregate in a manner consistent with it having a causative role in psychosis. Based on the preliminary work on this family, additional genetic studies were carried out to further test the involvement of *WHRN* variants in schizophrenia. Furthermore, functional studies of the identified *WHRN* coding variant was initiated.

In a genetic study, it is important to screen as large a number of patients as possible to achieve significance, and include age and ethnicity matched controls. The control genotypes used in the preliminary study were largely of Caucasian origin and it was therefore important to check the frequency of the variant in samples from the same ethnic background. Under-representation of an ethnic group in particular databases could lead researchers to overstate the significance of their findings, and this needed to be ruled out by using controls from the same ethnic group. If the variant occurs in unaffected controls from the same ethnic group, it is likely to be ethnic group specific and not associated with the disease phenotype (Casanova et al., 2014). Screening of additional cohorts of schizophrenia patients and control samples, including ethnically matched controls, did not identify the *WHRN* R450C variant or any similar changes within

the region. This suggests that this variant is rare and is a "private mutation" in this family.

WHRN encodes the protein whirlin, which is involved in the formation of scaffolding protein complexes in the central nervous system (Yap et al., 2003) and photoreceptor synapses (Van Wijk E. et al., 2006b). The protein has also been found to play an essential role in actin filament packing and stereocilia elongation in the hair cells of the inner ear (Mogensen et al., 2007). Figure 4.21 demonstrates the stereocilia interactome, in which whirlin exerts a central role among other proteins involved in the growth and cohesion of stereocillia. Hair cells have highly organised synaptic sites, like the ones in photoreceptor cells, called ribbon synapses (Khimich et al., 2005). These are essential for neurons that transmit graded signals. Ribbon synapses are characterized by an electron-dense structure of tight calcium channels and vesicles to enhance immediate release of the associated neurotransmitters (Sterling and Matthews, 2005).
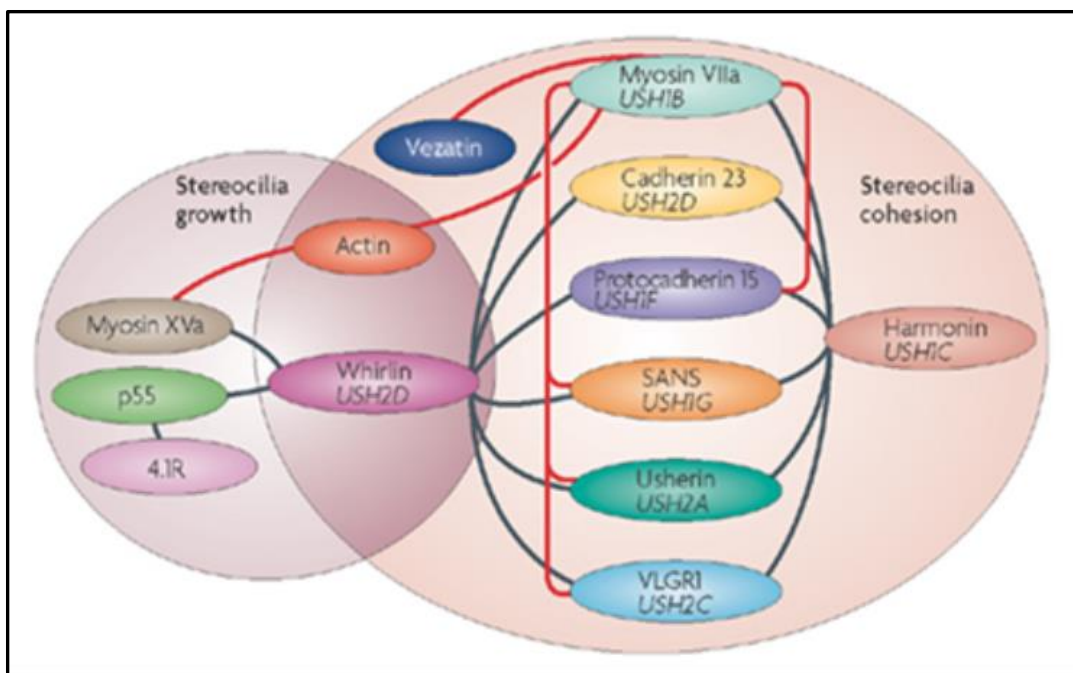
Figure 4.21. The role of whirlin within the 'stereocilia interactome'. The left circle contains the proteins that are involved in the growth of the stereocillia and the right circle contains the proteins that are crucial for the stereocillia cohesion. Whirlin is a scaffolding protein, located in the overlapping region of the two circles, required to establish a connection between these two groups, and forming a complex that is crucial for actin polymerization and stereocilia elongation. Evidence for this complex interactome comes from mouse genetic and protein

interaction studies. (Adapted from Brown Steve D. M. et al. (2008) and used with permission from Nature Publishing Group, License Number: 4078241032964).

The C1348T:p.R450C change is a missense variant located in the central portion of *WHRN*, encoding a Harmonin-N-like domain which is a putative protein binding domain (Pan et al., 2009). In contrast, the variants causing Usher syndrome or non-syndromic sensorineural deafness are nonsense mutations, clustered at the N-terminus or C-terminus, respectively. The N- and C-terminal PDZ domains are thought to be sites through which the protein interacts with other known components of the Usher protein network (Van Wijk E. et al., 2006b). For these reasons, research on the function of this protein has tended to focus on the N- and C-terminal domains, while the central region has not been extensively studied. The genetic findings in Family-7 suggest that missense variants in the central part of whirlin could predispose to psychiatric disorders such as bipolar disorder and schizophrenia. Nonsense mutations in the N- and C-termini are known to cause Usher syndrome and deafness, respectively.

It was previously found that whirlin is expressed in a large number of mouse tissues, with predominant expression in the brain, the eye and the inner ear (Van Wijk Erwin et al., 2006a). Whirlin expression in human tissues, carried out in this thesis (Figure 4.5), showed a similar finding. Whirlin was expressed in a wide range of tissues, particularly neuronal tissues of fetal and adult brain. A construct carrying the 450C variant was made as described in Section 2.14 and it was used to study the localisation of the 450C mutant in comparison with wild type whirlin. The R450C mutation did not appear to affect the cellular localization of the protein in overexpression studies (Section 4.2.2.4). However, structural modelling (Section 4.2.1.4) predicted that the change disrupts protein conformation, specifically disrupting two clusters of peripheral charged residues with potential protein binding properties. It was found, through yeast two-hybrid assay (Section 2.4.1.5), that the central region of whirlin, where the R450 residue is located, interacts with UBR4 in an interaction not previously described in the literature, and that the R450C mutation weakens this interaction.

239

UBR4 is an E3 ubiquitin-protein ligase of 5,183 amino acids that has been shown to interact with the smooth endoplasmic reticulum, a major intracellular calcium store, in neurons (Shim et al., 2008). It has also been shown to interact with calmodulin, a calcium sensor protein, in the cytoplasm (Nakatani et al., 2005), highlighting a potential role for UBR4 in calcium signalling. Consistent with the known role of UBR4 in neurogenesis (Parsons et al., 2015), neuronal migration (Shim et al., 2008) and neuronal Ca^{2+} signalling (Nakatani et al., 2005), an R5091H mutation in *UBR4* is reported to co-segregate with a novel form of autosomal dominant episodic ataxia in a large three-generation Irish family (Conroy et al., 2014). Interestingly, UBR4 is also a time-of-day-dependent and light-inducible protein in the suprachiasmatic nucleus of the mouse brain during circadian rhythm (Ling et al., 2014). Circadian rhythm disruption is a common feature in many neuropsychiatric diseases including schizophrenia, bipolar disorder and depression (Wulff et al., 2010). Although the precise mechanisms remain unclear, recent evidence suggests that circadian rhythm disruption is not simply a product of medication or an absence of social routine, but instead reflects commonly affected underlying pathways and mechanisms (Jagannath et al., 2013; Monti et al., 2013).

Based on the experiments and assays carried out, it is concluded that whirlin is highly expressed in neuronal tissue and that the R450C mutation observed in Family-7 is rare. Preliminary data suggest that the interaction of whirlin with UBR4 is affected by the R450C mutation, but further experiments are needed to confirm this. The R450C mutation might cause some alteration to the surrounding residues in whirlin, thus reducing but not abolishing the binding to UBR4. Therefore, a small disruption of UBR4 function, caused by reduced binding to whirlin, could conceivably confer risk for mental disorders.

4.4.2 Family-8 with a putative Mendelian allele in a homozygous region on chromosome 13q

The structure of this large consanguineous Pakistani family, with six schizophrenia cases across multiple first cousin marriages, is suggestive of a

recessive mode of inheritance. Although homozygosity mapping suggests a single shared region at chromosome 13 (13q31), the analysis of whole exome sequencing (WES) in this family has not been able to identify a single clear causative mutation. Despite its known power in diagnostic genetics, WES still has some limitations and is not always the best application (Klein Hanns-Georg et al., 2014; Warr et al., 2015). Indeed, there have been some cases where the underlying genetic causes were missed by WES, and other techniques were more useful. For example, a study of Inherited Retinal Dystrophy (IRD) did not detect a splicing mutation in the *CDHR1* gene (c.439-17G>A) by the standard WES analysis. This is because the mutation was filtered out due to not being in a coding region or obvious splice site. However, transcript analysis of cDNA from a patient's fibroblasts indicated the splicing of *CDHR1* by a mutation activating a cryptic splice-acceptor site (Tiwari et al., 2016).

In order to further investigate the genes located in the 13q locus, additional biological samples are required from the family members. Availability of samples is dependent on the consent of family members, and the feasibility of collection. For example, neuropsychistric disorders would ideally be investigated by determining the expression of the genes in the 13q homozygous region in brain samples. However, as brain samples are not available, other tissues were considered. Blood is one of the most easily accessible tissues (Kukurba and Montgomery, 2015). Rollins et al. (2010) found that majority of the brain transcripts (about 90%) are also expressed in blood. Peripheral blood mononuclear cells (PBMCs) have therefore been suggested as a good tissue in which to study gene expression in individuals with psychiatric disorders. The present study found that fibroblasts and blood lymphocytes gave similar results in the amplification of the 13 genes in the 13q31 locus (Figure 4.18). RT-PCR showed that most of the transcripts could be amplified from blood lymphocyte cDNA (Section 4.3.2.1), but it was not possible to amplify the transcripts of *SCEL*, *SLAIN1* and *EDNRB*.

A mutation in any one of the 13 genes located on 13q31-32 could potentially be a plausible cause of schizophrenia in this family. Some of the genes have already

been linked to some types of neuropsychiatric disorder. For example, mutations in *CLN5*, encoding a soluble lysosomal protein, have been associated with neuronal ceroid lipofuscinosis, an autosomal recessive, progressive encephalopathy in children (Jalanko and Braulke, 2009; Lyly, 2008). *SPRY2* encodes a protein belonging to the sprouty family that is involved in signal transduction in multiple processes, including axon growth and brain development. Decreased expression of *SPRY2* has been observed in post-mortem samples of dorsolateral prefrontal cortex (DLPFC) from schizophrenia patients (Pillai, 2008).

Genetic analysis indicated 13q31 as the region most likely linked to the schizophrenia in Family-8, and all the genes within it require thorough investigation. Further sampling of blood for transcript analysis is thus necessary. Whole genome sequencing could also be done, as it might reveal a hidden mutation in this family. Experimental studies have shown the capability of genome sequencing to detect causative mutations missed by techniques such as exome sequencing and genomic microarrays (Gilissen et al., 2014a; Lupski James R et al., 2010).

One further planned investigation in Family-8 is neuroimaging to obtain images of brain structure and activation patterns (Wen Wei et al., 2011). Neuroimaging is a non-invasive method to obtain information on brain structure and function in living subjects (Rumsey and Ernst, 2009). Magnetic resonance imaging (MRI) technique, which is based on using various magnetic fields, radio waves and fields gradients, has been used to attain such valuable information.

There are two main neuroimaging techniques to study brain structure. Structural MRI (sMRI) utilises three-dimensional weighted scans to measure the thickness of the cortical regions (Barton and Harvey, 2000) and to illustrate the distribution of anatomical networks in the human brain (He et al., 2007). Diffusion tensor imaging (DTI) of grey matter is used to reconstruct the anatomical connections between entire brain regions, permitting the study of the highly complex network properties of the brain and the production of detailed 3D maps (Hagmann et al., 2007). Diffusion spectrum imaging (DSI) followed by computational tractography

has been used to study the structural core of cortical networks (Hagmann et al., 2008).

Functional magnetic resonance imaging (fMRI) is also a method for studying brain networks. fMRI has been used to show the changes in the functional organization of the brain in Alzheimer's disease (Kivistö et al., 2014; Sperling, 2011) and schizophrenia (Liu Y et al., 2008), indicating specific alterations in the frontal, parietal and temporal lobes. With technical advances to provide a higher anatomical resolution and improved quantitation, fMRI has become the dominant technique to investigate brain networks (Dauvermann et al., 2014; Gur and Gur, 2010; Karlsgodt et al., 2010). By fMRI, schizophrenia subjects have been shown to have a characteristic physiological response of the synaptic activity within the dorsolateral prefrontal cortex (DLPFC) and anterior cingulate cortex (ACC) (Thermenos et al., 2013). In a study by (Wagner et al., 2013), schizophrenia patients (n=36) showed a significant reduction ($p < 0.001$) in the functional activation of the DLPFC and ACC compared with healthy control subjects (n=36) (Figure 4.22).
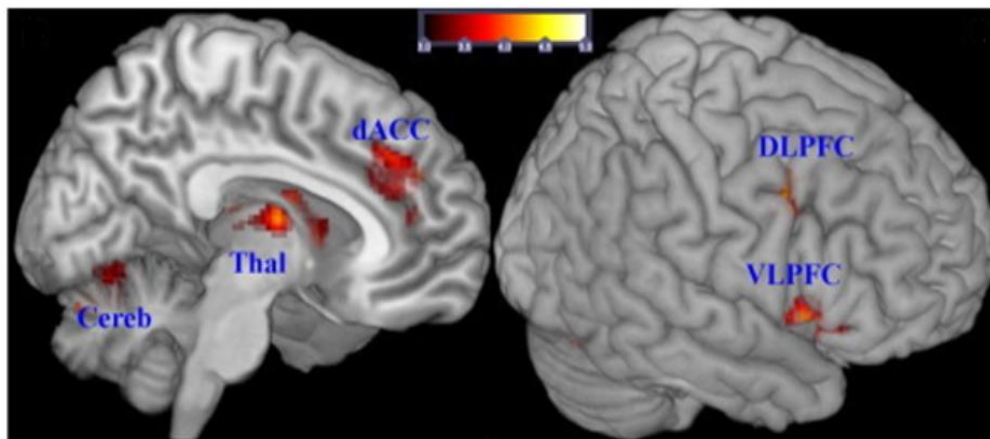
Figure 4.22. Brain regions that have reduced functional activity in schizophrenia patients. The orange-yellow scale represents the degree of hypoactivity in schizophrenia patients compared with healthy control subjects. Significantly lower activation was found in cerebellum, mediodorsal thalamus, dorsal anterior cingulate cortex (dACC), dorsolateral prefrontal cortex (DLPFC) and ventrolateral prefrontal cortex (VLPFC). (Adapted from Wagner et al. (2013); no permission required).

Certain tasks might require more concentration and longer time from the brain networks, depending on the task's difficulty (Gómez et al., 2014). A normal person requires a certain period of elapsed time to do daily routine activities, but schizophrenia patients have timing disturbance that negatively impacts on their cognition and daily behaviours (Alústiza et al., 2016). Various neuroimaging studies have shown that dysfunctional neural timing is the core deficiency in schizophrenia patients (Macar and Vidal, 2009; Volz et al., 2001). (Alústiza et al., 2016) carried out a multimodal analysis using data from 43 fMRI studies of schizophrenia patients. They concluded that schizophrenia patients have overloaded networks, resulting in a broad range of task-related timing and response deficits (Figure 4.23).

Neuroimaging of Family-8 would help to determine whether the affected cases have a pattern of brain region activation similar to the patterns seen in unaffected family members who are also homozygous for the 13q31 region, and in other, unrelated schizophrenia patients.

Four of the affected in Family-8 are responding to clozapine without any reported resistance or major side effects. Clozapine is normally used to treat schizophrenia when other antipsychotics are ineffective, but an estimated 40-70% of schizophrenia patients show resistance to clozapine (Papetti et al., 2007). Clozapine is reported to be helpful in reducing self-harming and suicidal behaviour (Meltzer et al., 2003; Zarzar and Mcevoy, 2013), but is also associated with the development of agranulocytosis, a life-threating side effect (Abanmy et al., 2014). The usage of clozapine has, therefore, become restricted and the haematological parameters of patients treated with clozapine are monitored (Miyamoto et al., 2004).
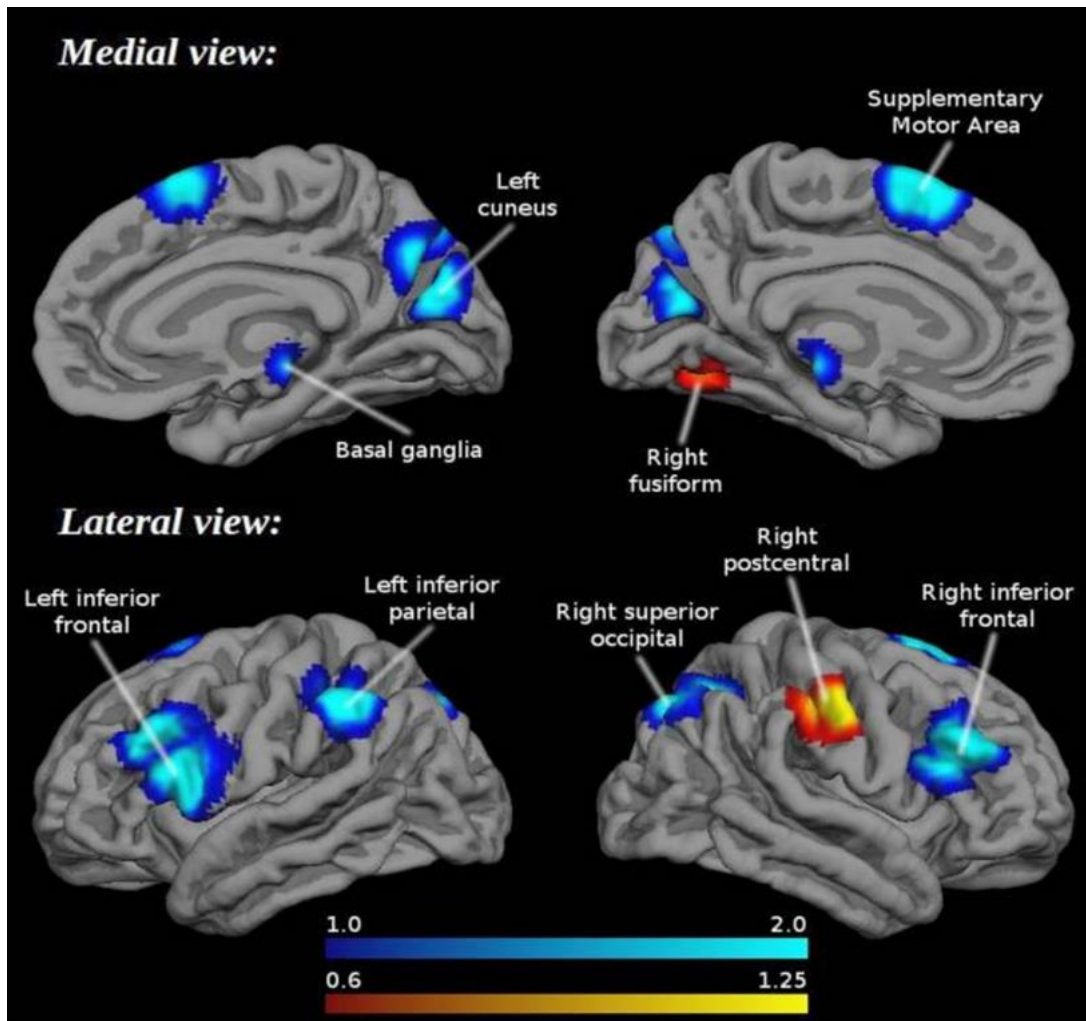
Figure 4.23. Neuroimaging of schizophrenia patients showing brain regions involved in cognitive tasks. Alústiza et al. (2016) carried out a signed differential mapping (SDM) meta-analysis on 954 schizophrenia patients and 999 healthy volunteers from 43 fMRI studies. The blue scale indicates the level of hypoactivity whereas the orange scale represents the degree of hyperactivity in schizophrenia patients. Bilateral inferior frontal and superior occipital gyri, right supplementary motor area, left inferior parietal gyri, left cuneus, and red nucleus all showed reduced activity. Hyperactivity was recorded in right postcentral and fusiform gyri. (Adapted from Alústiza et al. (2016); no permission required).

Affected individuals in Family-7 and Family-8 are reported to use cannabis, which can cause a wide range of psychiatric symptoms (Van Gastel et al., 2013). The active compound in cannabis is called Δ9-tetrahydrocannabinol (Δ9-THC), which produces behavioural effects within minutes (Anderson Beth M. et al., 2010; Khan Masood A. and Akella, 2009). When smoking cannabis, the psychoactive agent Δ9-THC gets immediately absorbed into the blood stream and then rapidly delivered to the brain. Being extremely lipid soluble, the Δ9-THC accumulates

within the fatty tissues and then is slowly released back into the body areas. After smoking cannabis, a high concentration is found in the brain and sensory and motor parts of the body (Ashton, 2001). Cannabis has not been approved by the US Food and Drug Administration (FDA), though there is still some conflict between federal and state laws about its legal status in the USA. In the UK, cannabis was made illegal in 1925 (Chadwick, 2008). Despite it having potentially harmful effects, cannabis has gradually become the most commonly used illicit drug, accounting for 80% of worldwide drug abuse (Jha et al., 2006; Macleod et al., 2004).

Hearing impairment (Carvill S., 2001; Stefanis et al., 2006) and Usher syndrome (Dammeyer, 2012; Domanico et al., 2015) are believed to be risk factors for the development of mental disorders, with a reported frequency of 23% of Usher syndrome patients presenting psychiatric conditions (Dammeyer, 2012). This could be the result of sensory deprivation, but alternatively it may imply a common pathology linking defects of hearing and vision with psychotic disorders. The involvement of the central nervous system (CNS) in Usher syndrome (Koizumi et al., 1988; Schaefer et al., 1998) supports this hypothesis.

To conclude, the follow up work on Family-7 has contributed further findings supporting our hypothesis that a proportion of the increased psychosis in the Pakistani population is caused by recessive alleles. In Family-8, it was not possible to progress the project further due to lack of the tissue samples from the family, but the preliminary results obtained should help when those samples become available in the future.

5. General Discussion

5.1 Summary of key findings

The research described herein used NGS technology to identify the genetic causes of intellectual disability (ID) in six consanguineous Omani families. It also includes some follow-up functional work on loci identified in two consanguineous families in the UK with schizophrenia.

The combination of whole genome homozygosity mapping and whole exome sequencing (WES) revealed mutations underlying ID in Omani families. Homozygosity mapping identified homozygous regions shared between the ID affected members in each family. The shared homozygous-by-descent (HBD) regions potentially could contain the mutation that causes the recessive ID phenotype in each family (Woods et al., 2006). To investigate an X-linked mode of inheritance, the affected cases in each family were checked for X chromosome variants. Families in which the underlying mutations could not be identified by the above strategy were also tested for possible copy number variants (CNVs) using the method of Plagnol et al. (2012). CNVs are a major cause of neurological disease and their estimated prevalence in ID is about 10-30% (Vissers Lisenka Elm et al., 2010). NGS technology was able to identify the causative mutation in three of the six ID families studied, a detection rate of 50%. Detection rates vary widely in the reported literature. For example, in 50 patients with ataxia, Nemeth et al. (Németh et al., 2013) reported a detection rate of 8.3% for patients with an adult onset progressive disorder, and 40% for patients with a childhood or adolescent onset progressive disorder. This suggests that applying NGS on samples of early onset disorders might increase the detection rate, which is not surprising as early-onset phenotypes are more likely to have a genetic basis (Acuna-Hidalgo et al., 2016; Ropers H, 2010; Zou et al., 2014). A WES study of an ID cohort of 100 trios (patients and their healthy parents) had a detection rate of 16% (De Ligt et al., 2012). A rate of 55% was achieved by the study of Rauch et al. (2012) on ID children and their parents from Germany and Switzerland. By

comparison, a WES study of 17 families with inherited retinal disorders achieved a mutation detection rate of 57% (Audo et al., 2012).

The mutations reported in this thesis are all novel, although some of them occur in genes previously associated with ID. The known ID genes in which novel mutations were identified are *TUSC3* and NHS. *TUSC3* is a known ID gene with various reported mutations but the variant identified in Family-1 (NM_006765:exon2:c.222delA, p.R74 fs) has not been reported before. This is the first time that a mutation in TUSC3 has been identified in an Arab population. The family are expected to benefit from this knowledge through genetic counselling and carrier screening of additional family members and/or potential marriage partners. The reported NHS change (NM_198270:exon8:c.C4385G, p.S1462C) might not be the cause of the cognitive impairment in Family-2 because the affected members do not have the ophthalmological and dental anomalies characteristic of Nance-Horan syndrome (NHS). One of the patients has some very mild teeth anomalies, including conical shaped teeth, crowding and overlapping incisors, but the NHS characteristic abnormalities are not seen in ID affected siblings. Moreover, as only about 30% of patients with pathogenic *NHS* mutations have intellectual disability, Nance-Horan syndrome is not necessarily the cause of ID in Family-2. The possibility remains that the *NHS* mutation reported in this study could be a SNP in the Omani population, even though none of the 50 tested ethnically-matched controls had this change. The significance of the NHS variant to the cognitive deficit in Family-2 thus remains uncertain. LHFPL5 is a gene that has been previously reported to be mutated in hearing loss. However, the LHFPL5 variant identified in this thesis (NM_182548:exon2:c.T575C, p.L192P) is novel, and could not be excluded from being the cause of intellectual disability in Family-2. In Family-3, two new mutated genes – *ANKRD2* and *PDZD8* – were identified as being potentially associated with ID. The subsequent analysis and screening suggested that a mutation in either, or both, of them could be the underlying cause of the phenotype in Family-3.

The genetic cause of ID in the remaining three families could not be solved even after searching for possible compound heterozygous mutations. These families were also tested for CNVs, which are a common genetic cause of neurological diseases. CNVs are genetic regions of more than 50bp that are duplicated or deleted in comparison with the reference genome sequence (Tattini et al., 2015). During the last decade, the importance of CNVs in risk of diseases has become increasingly apparent and is expected to be important in the development of personalized medicine in the future (Love et al., 2011; Sharp et al., 2005).

Numerous logarithmic and statistical tools exist for discovering CNVs from array intensities and next generation sequencing data. The review of Pirooznia et al. (2015) suggests appropriate methods, such as read depth and combined approach, to detect CNVs in whole genome sequencing data. Examples of commonly used tools are HMM-Hidden Markov models (Fromer et al., 2012; Love et al., 2011), CoNIFER-copy number inference from exome reads (Krumm et al., 2012) and EXCAVATOR (D'aurizio et al., 2016; Magi et al., 2013). An evaluation of 37 available tools concluded that a combination of different tools would help in detecting CNVs with higher confidence (Zhao Min et al. (2013). An analysis tool called ExomeDepth (Section 2.11.9.7) was used in the present study to look for possible CNVs in the families for which the causative mutation was not identified by homozygosity mapping and whole exome sequencing. ExomeDepth has shown a higher sensitivity than five other tools (ExomeCNV, CONTRA, ExomeCopy, CoNIFER and HMM) for detecting rare CNVs (Samarakoon et al., 2014), but did not identify any CNVs in the present study. In consequence, three families remain unsolved, and there are various potential reasons for these negative findings (explained in Chapter 3, Section 3.3.4).

Follow-up work was undertaken based on previous findings in two SZ families (Family-7 & Family-8). In Family-7, a variant in DFNB31 (R450C) was previously identified by a University of Leeds colleague, post-doctoral fellow Dr Jose Ivorra Martinez. Whirlin, the protein which encoded by *DFNB31*, was over-expressed in SH-SY5Y cells. A pull-down assay has not shown a conclusive evidence that whirlin interacts with UBR4 and it is, therefore, important to repeat the

experiment, using good quality of UBR4 antibody. UBR4/WHRN interaction need to be further tested and once that is confirmed, then it would be interesting to find out if the R450C mutant whirlin have reduced binding to UBR4, as this partial loss of function could potentially underlie the schizophrenia in Family-7. In Family-8, a single homozygous locus at chromosome 13q31 was previously found to be shared by six affected members. Whole exome sequencing did not find any potentially causative mutations in the coding regions, so the expression of the 13 genes in the shared region was analysed as the next step. Transcript analysis using control samples showed no difference in transcript amplification between fibroblasts and lymphocytes. Sample collection from Family-8 is the next step required for further transcript analysis. Other types of analysis of Family-8 could include neuroimaging and whole genome sequencing (WGS).

5.2 Implications for the diagnosis of ID and neuro-psychiatric disorders

Mental illnesses and intellectual disabilities are overlapping cognitive disorders that cause a high proportion of disability in the world. It has been estimated that mental disorders occur in 10–20% of teenagers in the world, and half of these conditions start before 14 years of age (Charara et al., 2017). People with mental illness/intellectual disability have limitations in their daily life and less opportunity for a good quality of life (Lecroy and Holschuh, 2012). As such, they need appropriate attention and health care. However, the study of mental illnesses has various challenges in terms of diagnosis and treatment (Wittchen et al., 2003).

Depending on the severity of the illness, patients with ID or mental diseases may appear normal if their condition is mild. For example, it could be hard for a specialist clinician to confidently diagnose a person with mild ID, when their IQ score is 69. Conversely, a normal person who does not perform well in IQ tests could be diagnosed with intellectual disability. The score of IQ is not a reflection of every aspect of a person's intellectual capacity, so caution is required before making a diagnosis based solely on the IQ test (Richardson Ken and Norgate, 2015; Strauss et al., 2006). Similarly, it can be difficult to recognize small changes

that occur in an individual with schizophrenia before the condition manifests in its complete form. Some of the first signs in young adults, like occasional spontaneous movements, sleep problems, and much time doing nothing, are not specific to SZ and could be misguiding. Furthermore, evaluating the negative symptoms, like lack of interest and social withdrawal, is hard because they are not as grossly abnormal as positives ones, like hallucinations and delusions, and thus could sometimes be missed by the patients and their relatives (Sarkar et al., 2015; Stahl and Buckley, 2007). Therefore, a definite diagnosis can sometimes not be easily achieved through observing the symptoms. In fact, this type of uncertainty about diagnosis is considered one of the major problems that psychiatrists face (Aboraya et al., 2006).

Despite the fact that mental illnesses can affect many areas of life, most affected individuals hesitate to seek medical help for various reasons. One of the more obvious reasons is poor awareness of the condition. A person with poor insight might think that nothing is wrong with them and the situation is not so bad as to require treatment (Reddy, 2016). The negative stigma and discrimination associated with having a mental illness is another reason for people not getting treatment (Corrigan and Watson, 2002). There is a fear of being labelled as "mentally retarded" or "mad", which could bring shame and unwanted attention. Another reason is the avoidance of sharing personal information, known as "distrust" (Henderson et al., 2015; Suite et al., 2007). Distrust could involve worry about maintaining confidentiality of the provided information or it could be attributed to hearing about others' bad experiences of treatment (Armstrong et al., 2013). Other factors documented to form barriers to achieving medical help include limited access to the professional centres that provide mental health care (Caldas and Killaspy, 2011; Ofori-Atta et al., 2010), as well as the cost of treatment (Mccrone et al., 2008; Trachtenberg et al., 2013).

Numerous studies have been done in the field of neurogenetics, including those on intellectual disability, schizophrenia and bipolar disorder. The findings of these studies have generally resulted in a better understanding of the biology and genes underlying these conditions (Chiurazzi and Pirozzi, 2016). For example,

Najmabadi et al. (2011) studied 136 consanguineous families with autosomal recessive ID and found mutations in 23 genes that were already linked to ID, plus 50 novel candidate genes. Similarly, De Ligt et al. (2012) used exome sequencing to identify *de novo* mutations in 12 previously known ID genes and 24 novel candidate genes amongst 100 trios in which the affected individuals had unexplained severe ID. Furthermore, WGS enabled the identification of 33 novel candidate genes in 143 multiplex families mainly with ID (Alazami et al., 2015). Likewise, a study of 53 schizophrenia sporadic cases, 22 unaffected controls and their parents resulted in the identification of *de novo* germ-line mutations in 40 genes (Xu Bin et al., 2011). A recent study of 192 consanguineous families with autosomal recessive ID from Pakistan and Iran identified 46 known ID genes and 26 new ID genes (Harripaul et al., 2017).

One of the genetic advances that has been introduced in the clinical setting to diagnose ID and neurodevelopmental diseases is targeted gene capture technology. Unlike standard NGS technology, at which all genomic regions have an equal chance to be sequenced, gene capture technology works by creating an enriched DNA capture to significantly increase the chance for a particular region or gene to be sequenced (Lin Xi et al., 2012b). There has been a range of diagnostic testing panels offered by different commercial companies. One example is a panel designed for detecting >90 X-linked genes of ID, presented by Eurofins Clinical Diagnostics (http://www.egl-eurofins.com/tests/MXLI1). Another example is the devACT® Clinical Management Panel by Courtagen (http://www.courtagen.com/test-menu-genetic-testing-devseek-clinical-management-panel.htm) which is a panel of 226 genes for patients with developmental delay, intellectual disability and Autism Spectrum Disorders (ASD). The results are normally received from Courtagen in less than three weeks at a cost of about $300. Given their reasonably acceptable cost as well as short handling time, the diagnostic panels offered by different vendors have been of increased use (Schnekenberg and Németh, 2013). In a study by Lim et al. (2015), a panel called Haloplex ICCG (International Collaboration for Clinical Genomics) was used to screen 180 genes associated with congenital neuromuscular and neurodevelopmental disorders. This panel achieved >97%

breadth of coverage at which the targeted bases are sequenced >20x average depth, and a 40% detection rate of an abnormality in one of the 180 genes. A limitation of diagnostic panels is that they screen only a particular set of genes, and thus are not suitable for identifying novel mutated genes (Winkler and Wiemann, 2016). For a panel to be efficient, it needs to be updated with more and more genes as they are identified and validated, but these are expensive and time consuming processes (Watson Christopher M et al., 2014). These panels are also not the best option to use in cases where the patients do not show a previously described disease phenotype.

Because of the limitations associated with the pre-designed panels, there has been a lot of interest in the use of whole genome sequencing (WGS) technology for mutation detection. WGS has several advantages that make it the method of choice in DNA sequencing (Pollack, 2014; Van El et al., 2013). These advantages include its consistent coverage as well as its greater capability to identify structural variants, in comparison with WES (Belkadi et al., 2015; Ekblom and Wolf, 2014). Some biases commonly take place in WES due to artifacts formed during library preparation (Aird et al., 2011), but these are eliminated with WGS, as no PCR step is involved in library preparation (Meynert et al., 2014). Furthermore, WGS has the capability to sequence regions that are difficult to sequence, such as GC-rich (Marx, 2013) and repetitive regions (Sipos et al., 2012). The cost of WGS is also gradually reducing and it now costs less than $1000 (Nagasaki et al., 2015; Warr et al., 2015). However, WGS still has some limitations, including difficulty in bioinformatic analysis, the interpretation of findings, the need of massive servers for data storage, and the security of stored data (Christensen et al., 2015).

Limitations and dilemmas are also present in NGS technologies. For instance, WES in the present study resulted in the identification of the *NHS* variant in Family-2 but it was not possible to confirm its causation of the phenotype based on the existing information and samples. Interpretation of variant lists obtained from NGS analysis is one of the biggest challenges (Dienstmann et al., 2014) and the availability of broad knowledge, about the affected patients, and resources,

about the candidate variants, would significantly help in making the appropriate interpretation of variants. Another important issue in the use of NGS is whether to report secondary or incidental findings that are unexpectedly observed by researchers (Wolf et al., 2008). During a particular study, researchers might surprisingly come across a finding that is not related to the purposed study but is important and needs medical action. For example, mutations in *BRCA1* or *BRCA2* predispose to cancer risk (Petrucelli et al., 2013) and it would be significantly valuable to report such findings to the clinicians who might consequently consider life-saving mastectomy for the patient (Metcalfe et al., 2014). Despite the obvious benefits of reporting such incidental findings, various issues related to confidentiality, discrimination and health insurance need to be considered (Clift et al., 2015; Meacham et al., 2010; Simon et al., 2011). Certain measures related to reporting incidental findings are being discussed by patient groups, researchers and clinicians. The American College of Medical Genetics (ACMG) has made regulations and guidelines to ensure best practice in reporting incidental findings (Jorde et al., 2015; Kalia et al., 2017).

Single cell sequencing (SCS) studies DNA or RNA taken from one cell rather than from a whole tissue sample (Gawad et al., 2016; Wang Yong and Navin, 2015). Neurobiology has shown increased utilization of SCS, after the fields of cancer, embryonic development and microbiology (Wang Jian and Song, 2017). Examples of its uses in neurobiology include neuron classification and neuronal diversity (Wang Yong and Navin, 2015). RNA sequencing in a single neuron, for exploring neuronal morphology, is one of the earliest applications of SCS (Qiu et al., 2012). LINE-1 retrotransposition is one of the processes believed to participate prominently in causing genomic variation among neuronal cells (Singer et al., 2010). Evrony et al. (2012) studied LINE-1 retrotransposition in neurons of the cerebral cortex, and found an occurrence of 0.6 insertion events in each neuron, which could be relevant to the neuronal diversity within the genome. Another study using SCS in normal and diseased brains reported that CNVs occur during the normal developmental stages in about 5% of the neuronal cell population (Cai et al., 2014).

The diagnostic services for mental health are underdeveloped in developing countries (Ngui et al., 2010). Since Sultan Qaboos acceded to the throne in Oman in 1970, enormous improvement the number of clinical centres/laboratories and experienced personnel dedicated to mental illness has occurred (Al-Azri et al., 2013). Mental health services were first founded in Oman through Al-Rahama Hospital which was opened in Muscat in the mid-1970s. There were also some out-patient services in different parts of the country during the late 1970s. The Ibn Sina hospital, opened in 1983, was the second specialised hospital to offer a high-standard of service for mental illnesses. The Royal Hospital, opened in 1987 with 630 beds, is now the largest hospital in the country that provides tertiary health care with genetic services (Al-Gazali et al., 2006). This was later followed by psychiatric services at the Sultan Qaboos University Hospital (SQUH) in 1991. By 2005, the World Health Organisation (WHO) ranked the service offered by the health institutions in Oman as "the most well-organized in the world" (Ghodse, 2011). The National Genetic Centre, inaugurated in 2013 as a supportive department for the genetic service of the Royal Hospital, is a recent development by the Ministry of Health. The centre has cytogenetic and molecular laboratories facilitated with advanced genetic analysis including NGS (Rajab Anna et al., 2015).

5.3 Future direction in ID and neuropsychiatric disorder treatment

Most individuals with ID have major complications not only in terms of the cognitive impairment but also with comorbid psychiatric disorders (Brown Julie F. et al., 2013) as well as stereotypic and aggressive behaviours (Medeiros, 2015). For example, a study based in the counties of Leicestershire and Rutland found that ID patients had 14% aggression rate, >3 times episodes of severe physical episodes towards other people, and about 3% of the population reported major injuries from ID patients that required medical attention (Tyrer F et al., 2006). Studies have also shown that people with ID have an increased incidence of psychological stress (Bramston and Fogarty, 2000; Janssen et al., 2002), mental

health illness and antisocial tendencies (Crocker et al., 2007; Wheeler et al., 2013).

Treatment of the manifestations associated with ID patients have been considered an important and effective target and there have been many studies (Luyben, 2009; Prout H Thompson and Nowak-Drabik, 2003; Willner, 2005) reporting the positive effect of implementing concurrent psychosocial interventions to decrease the rate of these complications. However, the study of ID populations is associated with certain restrictions. Beside their nonrandomized designs, most of the studies in this field have limitations in measuring the outcomes, as no indication is normally given about whether there has been a reduction in the rate or severity of the manifestations (Brown Julie F. et al., 2013). Therefore, it might be too early to objectively judge how useful these sequential interventions are in reducing the abnormal behaviours in ID individuals, especially in the absence of solid clinical evidence.

Another approach that has been in wide use as a treatment intervention to reduce the challenging behaviours in ID patients is called Applied Behaviour Analysis (ABA). This technique was described about 50 years ago (Baer et al., 1968) and its attributed mainly on the utilisation of social background and surroundings to sort out a social issue. "ABA has produced remarkably powerful interventions in fields such as education, developmental disabilities and autism, clinical psychology, behavioural medicine, organizational behaviour management, and a host of other fields and populations" (Slocum et al., 2014). There is a huge amount of evidence on the success of ABA in improving the challenges of ID individuals (Borrero and Vollmer, 2006; Hassiotis et al., 2009; Medeiros, 2015; Neef, 2001; Robertson et al., 2005). Despite the large number of studies that collectively demonstrate the effectiveness of the ABA procedure, some researchers have proposed certain arguments related to this discipline. Examples of their valid arguments include that the therapy is based on a cruel premise, the stimuli in the real life are not like in application, the difficulty to notice low-frequency behaviour and no clear measurements for relaxation that occurs after the treatment (Brown Julie F. et al., 2013; Sigafoos and Schlosser, 2008).

Dialectical Behaviour Therapy (DBT) is also a treatment approach that has been used in cases of ID and mental illnesses including conditions of borderline personality disorder (BPD). DBT was initially developed to help women with biopolar disorder and a previous history of a suicide attempt (Linehan, 1993). DBT is a type of cognitive-behavioural approach at which comprehensive life skills are incorporated in a programme and used to help the person to control their challenging emotions and behaviours (Chapman, 2006). The method showed promising results when applied in a pilot study of six individuals with ID, as general progress was recorded in all the studied functional measures (Sakdalan et al., 2010). Promising results were also reported for six adolescents, with features of biploar disorder (BPD), after receiving DBT treatment for a period of 26 weeks (Geddes et al., 2013). DBT intervention has also been used to reduce the manifestations associated with sucidal thoughts (Luoma and Villatte, 2012) as well as depression (Berking et al., 2009).

Apart from the cognitive and behavioural therapies, psychodynamic therapy (PDT) is one of the more recent approaches that has been researched to improve the condition of individuals with mental illness (Prout Tracy A and Wadkins, 2014; Wittenberg, 2006). PDT, which is also sometimes called insight-oriented therapy, is based on considering the negative impact of past occurrences to help the client understand their influence on present behaviours and thus control them (Schnyder and Cloitre, 2015). An investigative analysis on 528 individuals with schizophrenia and severe mental illness from four randomized trials (Malmberg et al., 2001) showed less hospitalization and medication after using PDT. Although there have been also some other studies suggesting the helpful value of PDT for psychosis, there is little evidence of longer term benefits and cost-effectiveness (Fonagy, 2015).

Because ID patients have 4-5 times the rate psychiatric illness compared with the general population, the use of antipsychotic drugs to treat ID patients is not uncommon (Rush et al., 2004). Various studies have reported that inappropriate use of antipsychotic drugs by ID patients is a concern (Branford, 1994; Nyunt Tin et al., 2008; Shah, 2017). Public Health England reports that 58% of ID and/or

autism patients are receiving antipsychotics but have no definite diagnosis of mental illness (Glover et al., 2015). The inappropriate use of antipsychotics also includes their overuse and extreme dosage without checking their adverse effects (Deb et al., 2009; Mcgillivray and Mccabe, 2004; Tyrer Peter et al., 2014). About 22.5% of ID and/or autism individuals have been issued with more than one antipsychotic, and 5.5% of them have been found with doses higher than the maximum (Glover et al., 2015). Over the last three decades, there has not been remarkable improvement in the development of new antipsychotics. Most of the currently used antipsychotics are more than 40 years old, and majority of the newly introduced drugs are either in the same category as the old ones or have no better effects (Gelenberg et al., 2013). The issue of not having an efficient antipsychotic and the fact that a gap is present between the currently available knowledge and the offered therapy is discussed further in section 5.4.

From the genetic point of view, there have been a number of therapeutic strategies that could potentially be applied to ID individuals. For example, the use of siRNAs to reduce the expression of the candidate gene (Rondal Jean-Adolphe and Rasore-Quartino, 2007), the use of immunoglobulins to target the product of the candidate gene (Pritchard and Kola, 2007), and the use of certain biochemical compounds that could alter either the activity of the associated protein or a step in its associated pathway (Korenberg et al., 1996) have all been tried in experimental animals as possible therapeutic approaches (Picker and Walsh, 2013; Rondal Jean-A and Lang, 2010). Hayashi et al. (2017) have described a novel approach to restore many of the cellular and behavioural failures seen in Fragile X syndrome (FXS). In their method, they focused on obstructing the catalytic activity of p21-activated kinase (PAK) in FMR1 knockout (KO) mice by crossing them with transgenic mice with a dominant negative form of the kinase (dnPAK transgene) specifically expressed in the postnatal forebrain (Hayashi et al., 2004). A wide range of abnormalities and FXS phenotypes were rescued in the offspring, including phenotypes associated with synaptic defects, cortical spine abnormalities, locomotor activity, stereotypy and anxiety. A successful therapeutic strategy has been described via introducing an extra copy of MECP2, a gene mutated in a neurodevelopmental disorder called Rett syndrome (RTT),

in a mouse model with a knock-in nonsense mutation (p.R255X) (Pitcher et al., 2015).

The technologies of genome editing systems like zinc finger nucleases (ZFNs), TALEN (Transcription Activator-Like Effector Nucleases) and CRISPR/Cas9 (Clustered Regulatory Interspaced Short Palindromic Repeats) offer great therapeutic potential. Their direct interventions to the genome are expected to yield a considerable influence on the treatment of various inherited conditions in humans (Carroll Dana, 2016; Nemudryi et al., 2014). ZFNs are being used in mice to develop a cure for Duchenne muscular dystrophy (DMD), a disease that occurs due to lose of dystrophin expression. Ousterout et al. (2015) demonstrated the strategy of ZFNs to remove particular sequences of dystrophin in order to skip exon 51 and thus repair the reading frame for the expression in mice. Similarly, the use of CRISPR-Cas9 to cut the full CGG-repeats of FMR1 in several cell lines, including human FXS iPS cells, is believed to restore the activity of this gene and maybe an effective treatment for individuals with FXS in the future (Xie et al., 2016).

The employment of stem cells, which have incredible potential to differentiate into other cell types and remarkable self-renew capacity, has also been explored as a possible therapeutic means for various brain conditions, including neuropsychiatric diseases (Benninghoff, 2009). The utilisation of stem cell technologies has been applied to many of human degenerative diseases including stroke, Parkinson's and Huntington's diseases (Lindvall and Kokaia, 2010; Lindvall et al., 2004). One of the great implications of this approach is the modelling that has been successfully done for various neuropsychiatric diseases of monogenic as well as non-monogenic causes (Soliman et al., 2017). For instance, published work describes the generation of neurons from iPS cells from patients with bipolar disorder (Madison et al., 2015; Mertens et al., 2015) and schizophrenia (Brennand et al., 2011; Chiang et al., 2011). Wen et al. (2014) generated iPS cells from four patients of a psychiatric disorder family with mutated DISC1 (Sachs et al., 2005); these cells showed various synaptic abnormalities and provide a cellular model on which to test potential therapies.

The biology of ID/neuropsychiatric diseases is being increasingly revealed by such studies, paving the way towards better therapies for these conditions.

5.4 The impact of up-to-date findings on the success of therapy

Great progress has been achieved in revealing the underlying genetic causes of intellectual disability, with more than 800 genes being linked to the condition (Chiurazzi and Pirozzi, 2016). Twin studies, adoption studies, linkage studies as well as association studies have collectively highlighted that neuropsychiatric diseases have at least a partial genetic basis (Cardno Alastair G, 2014; Kremen et al., 2016; O'donovan et al., 2009; Pepper and G Cardno, 2014; Ronald, 2015). The substantial progress in gene identification, coupled with advances in brain imaging techniques, has implicated multiple biological mechanisms/pathways. Furthermore, knowledge of the cellular functions of proteins in the brain has improved. The expectation of an effective therapy for ID has increased with encouraging research outcomes in recent years. For example, Cachón-González and colleagues (Cachón-González et al., 2012) have used gene transfer and reported promising results on treating mouse model of Tay- Tay-Sachs disease, a neurodegenerative disorder caused mutations in HEXA gene that resulted to no functional β-hexosaminidase A. In their experiments, Cachón-González and colleagues infused recombinant adeno-associated viral vectors encoding human β-hexosaminidase α and β subunits (rAAVα and rAAVβ) into the brain cranial of the mouse model at 1 month of age and they could rescue GM2 gangliosidosis in the CNS for up to 2 years. Similarly, it has become possible to prevent the occurrence of conditions like phenylketonuria and hyperthyroidism (Menolascino and Stark, 2012).

Despite the massive improvement in the identified genes and the comprehensive knowledge of many associated pathways, there is still a gap between identification of the affected gene and the availability of technology to eradicate the causes or to establish an efficient therapy, particularly for neuropsychiatric disorders (Krystal and State, 2014). Barely any antipsychotics were discovered

between 1975 and the introduction of second generation drugs in 1990s (Krystal and State, 2014; Tamminga, 2000). Later, some newer psychiatric medications have been discovered but none of the currently available drugs has demonstrated an improvement in the symptoms of cognitive deficits in schizophrenia patients (Stolzer, 2016). Furthermore, the newly introduced drugs have their own limitations, including a high price, major side effects and increased body weight (Hoffman et al., 2011; Stolzer, 2016). A randomised controlled trial (Kahn et al., 2008) comparing the typical antipsychotic haloperidol versus four newer second generation drugs (amisulpride, olanzapine, quetiapine and ziprasidone) in 498 patients with first episode SZ did not indicate any better response for the new drugs, as both groups showed a similar cessation rate. Similarly, Leucht et al. (2009) found no major differences in a wide range of aspects, including efficacy, side effects and sedation, between first-generation and second-generation drugs in their meta-analysis of 21,533 patients with SZ. Leucht et al. (2009) also pointed out that there is a huge variation among the antipsychotic drugs, and their classification might be confusing. Gøtzsche (2015) state that: "The drug industry has invented catchy terms such as 'second generation antipsychotics' and 'atypical antipsychotics' but there is nothing special about the new drugs".

It is believed that the deficiency in getting and providing adequate drug therapy for ID and neuropsychiatric diseases could be partially attributed to various limitations when researching this area of medicine. Moreover, the incomplete understanding of the underlying biology at the current time and the existing management measures are not of great help to select the best treatment option. In other words, they do not really direct whether treating a patient with some cognitive, behavioural or psychodynamic therapies would be advantageous, or the patient is not likely to benefit. According to Professor Helen Mayberg of Emory University, "Syndromes are so nonspecific by our current criteria that the best we can do now is flip a coin" (Weir, 2012). Many studies are being carried out using revolutionary technologies, such that the current gap in knowledge is closing and the underlying biology is being better elucidated. As knowledge increases, personalized medicine (PM) could be one of the best options.

Personalized medicine (PM) has been has been put forward as an alternative to the "one-size-fits-all" approach that has been ineffective in many conditions (Pavelić et al., 2015). There are certain factors, specific for each patient, which play a significant role in determining his/her likely response to the prescribed drugs. These factors include the age, health status and diet of the person, but genetic make-up is also a main contributor. PM works by considering the knowledge of molecular basis underlies drug response in order search for the best option of drugs that would give maximum benefit to the patients (Vogenberg et al., 2010). PM also allows pharmaceutical companies to take a forward step in producing some medicines that are tailored for particular groups of patients who no not seem to mprove with the available agents. The positive impact of knowing the patient's genetic variation on the treatment strategy could be in the form of choosing alternative drugs with less side effects or considering a preventive monitoring plan to avoid the occurrence of susceptible diseases. One good example of PM is *HER2* (human epidermal growth factor) testing for women diagnosed with breast cancer. The best treatment is based on finding out whether the tumour is *HER2* positive, as this is known to propagate more rapidly and would thus needs more attention (Lidgren et al., 2008; Weigel and Dowsett, 2010). In the field of neurology, PM has been utilised for some conditions and promising outcomes are expected with Alzheimer's disease and Parkinson's disease (Huang Xuemei et al., 2004), as well as in psychiatric diseases (Cutter and Liu, 2012). Progress in PM has already been made in various areas, as an approach to selecting the best treatment for each individual patient (Pavelić et al., 2015; Schee Genannt Halfmann et al., 2017).

In order to benefit from the PM strategy, it is important for the patients to have their genetic profile uploaded into a system, like the Personal Health Record (PHR), that allows the integration and retrieval of critically complex data in a practically easy manner (Van Gorp and Comuzzi, 2014). The newly introduced smart cards, like electronic health records (EHR) (Latha et al., 2012) and Genome (G) cards (Jauhari and Rizvi, 2015), are examples of the applications that have been developed to facilitate PM work. Besides the DNA sequencing profile, the electronic system needs to contain some lifestyle information, like details related

to exercise, sleep, diet and medication, which is entered by the patients themselves. It has been realised that patient engagement is very advantages and a routine review of their electronic medical records would allow them to check and update the information simultaneously (Harman et al., 2012; Mishuris et al., 2016). Documentation of any other information would be good and might be vital for certain analysis in the future. Physicians can access the records and utilize the stored information in order to select the best treatment strategy that could be used for the patients. These records are also important for clinical research as they facilitate the ultimate aim of delivering innovative medicines to society (Beresniak et al., 2015).

5.5 Concluding remarks and future work

Homozygosity mapping has considerably contributed to identifying recessive mutations in consanguineous families. Combining homozygosity mapping with next generation sequencing is a powerful strategy, applied in this current research that enables researchers to map the locus and identify the underlying mutation. Although next generation sequencing is still a relatively new technology, it has already made a potential impact on many diverse fields. In this thesis different novel variants, including the ones in *TUSC3*, *LHFPL5* and *PDZD8*, have been identified in families with intellectual disabilities and the implication of *WHRN* and 13q region in psychiatric diseases has also been further strengthened. The clinicians, who are treating the studied families, have already been updated with these findings. It is expected that the members of the families would benefit from the feedback as it imparts knowledge and makes them more aware of their genetic make-up to help make informed choices such as facilitating carrier testing in cases were interrelated marriages are intended. Furthermore, it helps them to consider prenatal or pre-implantation testing for future pregnancy.

In the field of neuroscience, NGS has revolutionized approaches of genetic analysis and allowed the detection of many rare and *de novo* mutations. Furthermore, a huge number of novel neurological disease genes have been discovered, resulting in the elucidation of their principle aetiologies. The great

wealth of the obtained data has provided greater insights into understanding the biology of intellectual disability and psychiatric diseases. Much greater knowledge is now available about the mechanisms involved in these complex disorders and the susceptibility risks to their development. Despite the great success of next generation sequencing technologies, some difficulties still exist with regard to bioinformatic analysis, result interpretation and ethical dilemmas.

There is no doubt that the mutations identified in this thesis have generally extended the information related to the causes of the diseases but it is too early to consider any therapy based on these findings. For example, TUSC3 is involved in the glycosylation process which belongs to the metabolic pathways of the genes implicated in ID. It might be worth doing a future study to find out more about the exact role of TUSC3 in glycosylation, particularly in relation to the proteins that are known to have a role in the brain. *TUSC3* gene is also known as a tumour suppressor and it would be good to monitor the affected members of Family-1 for the presence of tumours in the future. The variant identified in *NHS* gene was reported in patients who did not have the characteristic ophthalmological features seen in Nance-Horan syndrome. Although it would be useful to follow and re-evaluate the vision of affected members in the family, more work needs to be also done to further assess the pathogenic significance of the reported *NHS* variant.

The implication of new ID genes, like *PDZD8*, in this small set of families suggests that the field of neurological diseases is probably still enriched with numerous undiscovered mutations It would be interesting to see if cognitive deficits are present in PDZD8 KO mice (Pdzd8[tm1b(EUCOMM)Wtsi]) that have already been arranged to be imported to Leeds from the Wellcome Trust Sanger Institute (Hinxton, UK). PDZD8 is a cytoskeletal regulator but not much information is available for this protein. The fact that PDZD8 is of poorly understood function makes it difficult to predict the exact pathways that links it to the ID. Because of that, functional characterisation is actually needed to follow the exact role of PDZD8. One of the follow up work for the three unsolved ID families is to do whole genome sequencing (WGS), which might identify some underlying mutations that

have not been possible to be picked by the performed strategies. For the studied psychosis families, mutated *WHRN* appears to have a role in the development of schizophrenia and evidence of mutations from other future reports would be important for its implication to this disorder. Similarly, the identified 13q region have strengthened the importance of this region in the occurrence of psychiatric diseases and reverse transcriptase RT-PCR on patients' lymphocytes for the 13 genes in this region could suggest the presence of an underlying mutations. Sanger sequencing could then identify large deletion/insertion, promotor variant or intronic mutation in one of the genes.

In terms of therapy, a considerable impact has generally been achieved and the commonly prevailing view that no treatment is available for cognitively impaired individuals has been changed in recent years. Different types of therapies have been utilised to improve the intellectual ability of patients, including cognitive, behavioural and psychodynamic therapies. Although not much improvement has occurred in the medication side, the application of siRNAs and CRISPR/Cas9 has proved the possibility to restore the action of the mutated gene using cells or animal models. Furthermore, stem-cell technology is another exciting application and its advances have already allowed modelling conditions like schizophrenia and bipolar disorder, providing important information for their therapy. The move to personalised medicine, which is facilitated by smart cards and electronic records, is a pronounced evolution that would allow selecting the best treatment option that matches the patient's need. It is therefore obvious that research has the potential to yield massive outcomes in both technology and treatment but the demand is still there for lives to be free from disease.

References

Abanmy N. O., Al-Jaloud A., Al-Jabr A., Al-Ruwaisan R., Al-Saeed W. and Fatani S. 2014. Clozapine-induced blood dyscrasias in saudi arab patients. *International Journal of Clinical Pharmacy,* 36**,** 815-820.

Abikoff H. B., Jensen P. S., Arnold L. L. E., Hoza B., Hechtman L., Pollack S., Martin D., Alvir J., March J. S., Hinshaw S., Vitiello B., Newcorn J., Greiner A., Cantwell D. P., Conners C. K., Elliott G., Greenhill L. L., Kraemer H., Pelham W. E., Severe J. B., Swanson J. M., Wells K. and Wigal T. 2002. Observed classroom behavior of children with adhd: Relationship to gender and comorbidity. *Journal of Abnormal Child Psychology,* 30**,** 349-359.

Aboraya A., Rankin E., France C., El-Missiry A. and John C. 2006. The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis. *Psychiatry (Edgmont),* 3**,** 41-50.

Abu-Saad K. and Fraser D. 2010. Maternal nutrition and birth outcomes. *Epidemiologic Reviews***,** 32, 5–25.

Acuna-Hidalgo R., Veltman J. A. and Hoischen A. 2016. New insights into the generation and role of de novo mutations in health and disease. *Genome Biology,* 17**,** 241.

Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G., Scherer S. E., Li P. W., Hoskins R. A. and Galle R. F. 2000. The genome sequence of drosophila melanogaster. *Science,* 287**,** 2185-2195.

Ades L., Sullivan K., Biggin A., Haan E., Brett M., Holman K., Dixon J., Robertson S., Holmes A. and Rogers J. 2006. Fbn1, tgfbr1, and the marfan-craniosynostosis/mental retardation disorders revisited. *American Journal of Medical Genetics Part A,* 140**,** 1047-1058.

Adzhubei I. A., Schmidt S., Peshkin L., Ramensky V. E., Gerasimova A., Bork P., Kondrashov A. S. and Sunyaev S. R. 2010. A method and server for predicting damaging missense mutations. *Nature Methods,* 7**,** 248-249.

Aird D., Ross M. G., Chen W.-S., Danielsson M., Fennell T., Russ C., Jaffe D. B., Nusbaum C. and Gnirke A. 2011. Analyzing and minimizing pcr amplification bias in illumina sequencing libraries. *Genome Biology,* 12**,** R18-R18.

Al-Azri M. H., Al-Awisi H., Al-Rasbi S. and Al-Moundhri M. 2013. Coping with a diagnosis of breast cancer among omani women. *Journal of Health Psychology,* 19**,** 836-846.

Al-Gazali L., Hamamy H. and Al-Arrayad S. 2006. Genetic disorders in the arab world. *British Medical Journal,* 333**,** 831-834.

Al-Amri A., Saegh A. A., Al-Mamari W., El-Asrag M. E., Ivorra J. L., Cardno A. G., Inglehearn C. F., Clapcote S. J. and Ali M. 2016. Homozygous single base deletion in tusc3 causes intellectual disability with developmental delay in an omani family. *American Journal of Medical Genetics Part A*. doi: 10.1002.

Alazami A. M., Patel N., Shamseldin H. E., Anazi S., Al-Dosari M. S., Alzahrani F., Hijazi H., Alshammari M., Aldahmesh M. A. and Salih M. A. 2015. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Reports,* 10**,** 148-161.

Alexander-Bloch A. F., Mcdougle C. J., Ullman Z. and Sweetser D. A. 2016. Iqsec2 and x-linked syndromal intellectual disability. *Psychiatric Genetics,* 26**,** 101-108.

Alkhiary Y. M., Jelani M., Almramhi M. M., Mohamoud H. S. A., Al-Rehaili R., Al-Zahrani H. S., Serafi R., Yang H. and Al-Aama J. Y. 2016. Whole-exome sequencing reveals a recurrent mutation in the cathepsin c gene that causes papillon–lefevre syndrome in a saudi family. *Saudi Journal of Biological Sciences,* 23**,** 571-576.

Aller E., Jaijo T., Van Wijk E., Ebermann I., Kersten F., García-García G., Voesenek K., Aparisi M. J., Hoefsloot L. and Cremers C. 2010. Sequence variants of the dfnb31 gene among usher syndrome patients of diverse origin. *Molecular Vision*, 16, 495–500.

Alústiza I., Radua J., Albajes-Eizagirre A., Domínguez M., Aubá E. and Ortuño F. 2016. Meta-analysis of functional neuroimaging and cognitive control studies in schizophrenia: Preliminary elucidation of a core dysfunctional timing network. *Frontiers in Psychology,* 7**,** 192.

Ammar-Khodja F., Bonnet C., Dahmani M., Ouhab S., Lefèvre G. M., Ibrahim H., Hardelin J.-P., Weil D., Louha M. and Petit C. 2015. Diversity of the causal genes in hearing impaired algerian individuals identified by whole exome sequencing. *Molecular Genetics & Genomic Medicine,* 3**,** 189-196.

Anderson B. M., Rizzo M., Block R. I., Pearlson G. D. and O'leary D. S. 2010. Sex, drugs, and cognition: Effects of marijuana. *Journal of Psychoactive Drugs,* 42**,** 413-424.

Anderson P. J. Neuropsychological outcomes of children born very preterm. Seminars in Fetal and Neonatal Medicine, 2014. Elsevier, 90-96.

Ansorge W., Sproat B., Stegemann J., Schwager C. and Zenke M. 1987. Automated DNA sequencing: Ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Research,* 15**,** 4593-4602.

Ansorge W., Sproat B. S., Stegemann J. and Schwager C. 1986. A non-radioactive automated method for DNA sequence determination. *Journal of Biochemical and Biophysical Methods,* 13**,** 315-323.

Armstrong K., Putt M., Halbert C. H., Grande D., Schwartz J. S., Liao K., Marcus N., Demeter M. B. and Shea J. A. 2013. Prior experiences of racial discrimination and racial differences in health care system distrust. *Medical Care,* 51**,** 144.

Aronson M. and Hagberg B. 1998. Neuropsychological disorders in children exposed to alcohol during pregnancy: A follow-up study of 24 children to alcoholic mothers in göteborg, sweden. *Alcoholism: Clinical and Experimental Research,* 22**,** 321-324.

Artiss T. and Hughes B. 2007. Taking the headaches out of anesthetizing drosophila: A cheap & easy method of constructing carbon dioxide staging. *The American Biology Teacher,* 69**,** e77-e80.

Ashton C. H. 2001. Pharmacology and effects of cannabis: A brief review. *The British Journal of Psychiatry,* 178**,** 101-106.

Audo I., Bujakowska K. M., Léveillard T., Mohand-Saïd S., Lancelot M.-E., Germain A., Antonio A., Michiels C., Saraiva J.-P., Letexier M., Sahel J.-A., Bhattacharya S. S. and Zeitz C. 2012. Development and application of a next-generation-sequencing (ngs) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet Journal of Rare Diseases,* 7**,** 8.

Awadalla P., Gauthier J., Myers R. A., Casals F., Hamdan F. F., Griffing A. R., Côté M., Henrion E., Spiegelman D., Tarabeux J., Piton A., Yang Y., Boyko A., Bustamante C., Xiong L., Rapoport J. L., Addington A. M., Delisi J. L. E., Krebs M.-O., Joober R., Millet B., Fombonne É., Mottron L., Zilversmit M., Keebler J., Daoud H., Marineau C., Roy-Gagnon M.-H., Dubé M.-P., Eyre-Walker A., Drapeau P., Stone

E. A., Lafrenière R. G. and Rouleau G. A. 2010. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *American Journal of Human Genetics,* 87**,** 316-324.

Ayalew M., Le-Niculescu H., Levey D. F., Jain N., Changala B., Patel S. D., Winiger E., Breier A., Shekhar A., Amdur R., Koller D., Nurnberger J. I., Corvin A., Geyer M., Tsuang M. T., Salomon D., Schork N. J., Fanous A. H., O'donovan M. C. and Niculescu A. B. 2012. Convergent functional genomics of schizophrenia: From comprehensive understanding to genetic risk prediction. *Molecular Psychiatry,* 17**,** 887-905.

Badenhop R., Moses M., Scimone A., Mitchell P., Ewen-White K., Rosso A., Donald J., Adams L. and Schofield P. 2002. A genome screen of 13 bipolar affective disorder pedigrees provides evidence for susceptibility loci on chromosome 3 as well as chromosomes 9, 13 and 19. *Molecular psychiatry,* 7**,** 594.

Badenhop R., Moses M., Scimone A., Mitchell P., Ewen K., Rosso A., Donald J., Adams L. and Schofield P. 2001. A genome screen of a large bipolar affective disorder pedigree supports evidence for a susceptibility locus on chromosome 13q. *Molecular psychiatry,* 6**,** 396-403.

Baer D. M., Wolf M. M. and Risley T. R. 1968. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* 1**,** 91-97.

Ball A. D., Stapley J., Dawson D. A., Birkhead T. R., Burke T. and Slate J. 2010. A comparison of snps and microsatellites as linkage mapping markers: Lessons from the zebra finch (taeniopygia guttata). *BMC Genomics,* 11**,** 1-15.

Bamshad M. J., Ng S. B., Bigham A. W., Tabor H. K., Emond M. J., Nickerson D. A. and Shendure J. 2011. Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews Genetics,* 12**,** 745-755.

Banerjee S., Riordan M. and Bhat M. A. 2014. Genetic aspects of autism spectrum disorders: Insights from animal models. *Frontiers in Cellular Neuroscience,* 8**,** 58.

Barba M., Czosnek H. and Hadidi A. 2014. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses,* 6**,** 106-136.

Bartolomé R. A., García-Palmero I., Torres S., López-Lucendo M., Balyasnikova I. V. and Casal J. I. 2015. Il13 receptor α2 signaling requires a scaffold protein, fam120a, to activate the fak and pi3k pathways in colon cancer metastasis. *Cancer Research,* 75**,** 2434-2444.

Barton R. A. and Harvey P. H. 2000. Mosaic evolution of brain structure in mammals. *Nature,* 405**,** 1055-1058.

Barz C. S., Bessaih T., Abel T., Feldmeyer D. and Contreras D. 2015. Altered resonance properties of somatosensory responses in mice deficient for the schizophrenia risk gene neuregulin 1. *Brain Structure and Function***,** 1-16.

Basavarajappa B. S. and Subbanna S. 2016. Epigenetic mechanisms in developmental alcohol-induced neurobehavioral deficits. *Brain Sciences,* 6**,** 12.

Bassett A. S. and Husted J. 1997. Anticipation or ascertainment bias in schizophrenia? Penrose's familial mental illness sample. *American Journal of Human Genetics,* 60**,** 630.

Baum A. E., Akula N., Cabanero M., Cardona I., Corona W., Klemens B., Schulze T. G., Cichon S., Rietschel M., Nothen M. M., Georgi A., Schumacher J., Schwarz M., Abou Jamra R., Hofels S., Propping P., Satagopan J., Detera-Wadleigh S. D., Hardy J. and Mcmahon F. J. 2008. A genome-wide association study implicates

diacylglycerol kinase eta (dgkh) and several other genes in the etiology of bipolar disorder. *Molecular Psychiatry,* 13**,** 197-207.

Bean C., Facchinello N., Faulkner G. and Lanfranchi G. 2008. The effects of ankrd2 alteration indicate its involvement in cell cycle regulation during muscle differentiation. *Biochim Biophys Acta,* 1783**,** 1023-1035.

Bean C., Verma N., Yamamoto D., Chemello F., Cenni V., Filomena M., Chen J., Bang M. and Lanfranchi G. 2014. Ankrd2 is a modulator of nf-κb-mediated inflammatory responses during muscle differentiation. *Cell Death & Disease,* 5**,** e1002.

Beaulieu C. L., Huang L. J., Innes A. M., Akimenko M. A., Puffenberger E. G., Schwartz C., Jerry P., Ober C., Hegele R. A., Mcleod D. R., Schwartzentruber J., Majewski J., Bulman D. E., Parboosingh J. S., Boycott K. M. and Consortium F. C. 2013. Intellectual disability associated with a homozygous missense mutation in thoc6. *Orphanet Journal of Rare Diseases,* 8.

Bekris L. M., Mata I. F. and Zabetian C. P. 2010. The genetics of parkinson disease. *Journal of Geriatric Psychiatry and Neurology,* 23**,** 228-242.

Belkadi A., Bolze A., Itan Y., Cobat A., Vincent Q. B., Antipenko A., Shang L., Boisson B., Casanova J.-L. and Abel L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences,* 112**,** 5473-5478.

Bellen H. J., Tong C. and Tsuda H. 2010. 100 years of drosophila research and its impact on vertebrate neuroscience: A history lesson for the future. *Nature reviews. Neuroscience,* 11**,** 514.

Bener A., Dafeeah E. E. and Samson N. 2012. The impact of consanguinity on risk of schizophrenia. *Psychopathology,* 45**,** 399-400.

Benninghoff J. 2009. Stem cell approaches in psychiatry - challenges and opportunities. *Dialogues in Clinical Neuroscience,* 11**,** 397-404.

Bensaïd M., Hmani-Aifa M., Hammami B., Tlili A., Hakim B., Charfeddine I., Ayadi H., Ghorbel A., Castillo I. D. and Masmoudi S. 2011. Dfnb66 and dfnb67 loci are non allelic and rarely contribute to autosomal recessive nonsyndromic hearing loss. *European Journal of Medical Genetics,* 54**,** e565-e569.

Beresniak A., Schmidt A., Dupont D., Sundgren M., Kalra D. and De Moor G. J. 2015. Improving performance of clinical research: Development and interest of electronic health records. *BioMed Research International,* doi./10.1155/864549.

Bergerheim U. S. R., Kunimi K., Collins V. P. and Ekman P. 1991. Deletion mapping of chromosomes 8 10 and 16 in human prostatic carcinoma. *Journal of Urology,* 145**,** 350A.

Berking M., Neacsiu A., Comtois K. A. and Linehan M. M. 2009. The impact of experiential avoidance on the reduction of depression in treatment for borderline personality disorder. *Behaviour Research and Therapy,* 47**,** 663-670.

Berland S., Alme K., Brendehaug A., Houge G. and Hovland R. 2011. Phf6 deletions may cause borjeson-forssman-lehmann syndrome in females. *Molecular Syndromology,* 1**,** 294-300.

Bernier D., Macintyre G., Bartha R., Hanstock C. C., Mcallindon D., Cox D., Purdon S., Aitchison K. J., Rusak B. and Tibbo P. G. 2014. Npas3 variants in schizophrenia: A neuroimaging study. *BMC Medical Genetics,* 15**,** 37-37.

Bernstein F. and Machol R. 1935. The detection of linkage in human families. *Proceedings of the Royal Society of London B: Biological Sciences,* 117**,** 63-68.

Berrettini W. Evidence for shared susceptibility in bipolar disorder and schizophrenia. American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2003. Wiley Online Library, 59-64.

Beurg M., Xiong W., Zhao B., Müller U. and Fettiplace R. 2015. Subunit determination of the conductance of hair-cell mechanotransducer channels. *Proceedings of the National Academy of Sciences of the United States of America*, 112**,** 1589-1594.

Bharadwaj R. A., Jaffe A. E., Chen Q., Deep-Soboslay A., Goldman A. L., Mighdoll M. I., Cotoia J. A., Brandtjen A. C., Shin J. and Hyde T. M. 2016. Genetic risk mechanisms of posttraumatic stress disorder in the human brain. *Journal of Neuroscience Research*, doi: 10.1002/23957

Bhugra D. 2004. Migration, distress and cultural identity. *British Medical Bulletin,* 69**,** 129-141.

Bhugra D. and Becker M. A. 2005. Migration, cultural bereavement and cultural identity. *World Psychiatry,* 4**,** 18-24.

Bhuvaneshwar K., Sulakhe D., Gauba R., Rodriguez A., Madduri R., Dave U., Lacinski L., Foster I., Gusev Y. and Madhavan S. 2015. A case study for cloud based high throughput analysis of ngs data using the globus genomics system. *Computational and Structural Biotechnology Journal,* 13**,** 64-74.

Biamino E., Di Gregorio E., Belligni E. F., Keller R., Riberi E., Gandione M., Calcia A., Mancini C., Giorgio E. and Cavalieri S. 2016. A novel 3q29 deletion associated with autism, intellectual disability, psychiatric disorders, and obesity. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171**,** 290-299.

Biesecker L. G. and Spinner N. B. 2013. A genomic view of mosaicism and human disease. *Nature Reviews Genetics,* 14**,** 307-320.

Bilder D. A., Pinborough-Zimmerman J., Bakian A. V., Miller J. S., Dorius J. T., Nangle B. and Mcmahon W. M. 2013. Prenatal and perinatal factors associated with intellectual disability. *American Journal on Intellectual and Developmental Disabilities,* 118**,** 156-176.

Birk E., Har-Zahav A., Manzini C. M., Pasmanik-Chor M., Kornreich L., Walsh C. A., Noben-Trauth K., Albin A., Simon A. J., Colleaux L., Morad Y., Rainshtein L., Tischfield D. J., Wang P., Magal N., Maya I., Shoshani N., Rechavi G., Gothelf D., Maydan G., Shohat M. and Basel-Vanagaite L. 2010. Sobp is mutated in syndromic and nonsyndromic intellectual disability and is highly expressed in the brain limbic system. *American Journal of Human Genetics,* 87**,** 694-700.

Bittles A. H. 2008. A community genetics perspective on consanguineous marriage. *Public Health Genomics,* 11**,** 324-330.

Bittles A. H. and Black M. L. 2010. Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 1779-1786.

Black J. R. M. and Clark S. J. 2016. Age-related macular degeneration: Genome-wide association studies to translation. *Genetics in Medicine,* 18**,** 283-289.

Blankenberg D., Gordon A., Von Kuster G., Coraor N., Taylor J., Nekrutenko A. and The Galaxy T. 2010. Manipulation of fastq data with galaxy. *Bioinformatics,* 26**,** 1783-1785.

Boat T. F. and Wu J. T. 2015. *Mental disorders and disabilities among low-income children*, National Academies Press.

270

Boksa P. 2012. Abnormal synaptic pruning in schizophrenia: Urban myth or reality? *Journal of Psychiatry & Neuroscience,* 37**,** 75-77.

Borrero C. S. and Vollmer T. R. 2006. Experimental analysis and treatment of multiply controlled problem behavior: A systematic replication and extension. *Journal of Applied Behavior Analysis,* 39**,** 375-379.

Bourque F., Van Der Ven E. and Malla A. 2011. A meta-analysis of the risk for psychotic disorders among first-and second-generation immigrants. *Psychological medicine,* 41**,** 897-910.

Bradshaw N. J. and Porteous D. J. 2012. Disc1-binding proteins in neural development, signalling and schizophrenia. *Neuropharmacology,* 62**,** 1230-1241.

Bramston P. and Fogarty G. 2000. The assessment of emotional distress experienced by people with an intellectual disability: A study of different methodologies. *Research in Developmental Disabilities,* 21**,** 487-500.

Brancati F., Dallapiccola B. and Valente E. M. 2010. Joubert syndrome and related disorders. *Orphanet Journal of Rare Diseases,* 5**,** 20.

Brand A. H. and Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development,* 118**,** 401-415.

Branford D. 1994. A study of the prescribing for people with learning disabilities living in the community and in national health service care. *Journal of Intellectual Disability Research,* 38**,** 577-586.

Brennand K. J., Simone A., Jou J., Gelboin-Burkhart C., Tran N., Sangar S., Li Y., Mu Y., Chen G., Yu D., Mccarthy S., Sebat J. and Gage F. H. 2011. Modelling schizophrenia using human induced pluripotent stem cells. *Nature,* 473**,** 221-225.

Brisch R., Saniotis A., Wolf R., Bielau H., Bernstein H.-G., Steiner J., Bogerts B., Braun K., Jankowski Z., Kumaratilake J., Henneberg M. and Gos T. 2014. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: Old fashioned, but still in vogue. *Frontiers in Psychiatry,* 5**,** 47.

Brooks S. P., Coccia M., Tang H. R., Kanuga N., Machesky L. M., Bailly M., Cheetham M. E. and Hardcastle A. J. 2010. The nance–horan syndrome protein encodes a functional wave homology domain (whd) and is important for co-ordinating actin remodelling and maintaining cell morphology. *Human Molecular Genetics,* 19**,** 2421-2432.

Brown A. S. 2011. The environment and susceptibility to schizophrenia. *Progress in Neurobiology,* 93**,** 23-58.

Brown J. F., Brown M. Z. and Dibiasio P. 2013. Treating individuals with intellectual disabilities and challenging behaviors with adapted dialectical behavior therapy. *Journal of Mental Health Research in Intellectual Disabilities,* 6**,** 280-303.

Brown S. D. M., Hardisty-Hughes R. E. and Mburu P. 2008. Quiet as a mouse: Dissecting the molecular and genetic basis of hearing. *Nature Reviews Genetics,* 9**,** 277-290.

Brownstein C. A., Kleiman R. J., Engle E. C., Towne M. C., D'angelo E. J., Yu T. W., Beggs A. H., Picker J., Fogler J. M. and Carroll D. 2016. Overlapping 16p13. 11 deletion and gain of copies variations associated with childhood onset psychosis include genes with mechanistic implications for autism associated pathways: Two case reports. *American Journal of Medical Genetics Part A*,170A(5):1165-73.

Buoli M., Dell'osso B., Zaytseva Y., Gurovich I. Y., Movina L., Dorodnova A., Shmuckler A. and Altamura A. C. 2013. Duration of untreated illness (dui) and schizophrenia

sub-types: A collaborative study between the universities of milan and moscow. *International Journal of Social Psychiatry,* 59**,** 765-770.

Burdon K. P., Mckay J. D., Sale M. M., Russell-Eggitt I. M., Mackey D. A., Wirth M. G., Elder J. E., Nicoll A., Clarke M. P., Fitzgerald L. M., Stankovich J. M., Shaw M. A., Sharma S., Gajovic S., Gruss P., Ross S., Thomas P., Voss A. K., Thomas T., Gécz J. and Craig J. E. 2003. Mutations in a novel gene, nhs, cause the pleiotropic effects of nance-horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *The American Journal of Human Genetics,* 73**,** 1120-1130.

Busskamp V., Lewis N. E., Guye P., Ng A. H. M., Shipman S. L., Byrne S. M., Sanjana N. E., Murn J., Li Y., Li S., Stadler M., Weiss R. and Church G. M. 2014. Rapid neurogenesis through transcriptional activation in human stem cells. *Molecular Systems Biology,* 10**,** 760.

Cachón-González M. B., Wang S. Z., Mcnair R., Bradley J., Lunn D., Ziegler R., Cheng S. H. and Cox T. M. 2012. Gene transfer corrects acute gm2 gangliosidosis—potential therapeutic contribution of perivascular enzyme flow. *Molecular Therapy,* 20, 1489-1500.

Cai X., Evrony G. D., Lehmann H. S., Elhosary P. C., Mehta B. K., Poduri A. and Walsh C. A. 2014. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Reports,* 8**,** 1280-1289.

Caldas J. M. and Killaspy H. 2011. Long-term mental health care for people with severe mental disorders. *European Union Publication***,** 5.

Camacho-Garcia R. J., Planelles M. I., Margalef M., Pecero M. L., Martinez-Leal R., Aguilera F., Vilella E., Martinez-Mir A. and Scholl F. G. 2012. Mutations affecting synaptic levels of neurexin-1 beta in autism and mental retardation. *Neurobiology of Disease,* 47**,** 135-143.

Campbell R. E., Tour O., Palmer A. E., Steinbach P. A., Baird G. S., Zacharias D. A. and Tsien R. Y. 2002. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences,* 99**,** 7877-7882.

Cantarella C. and D'agostino N. 2015. Psr: Polymorphic ssr retrieval. *BMC Research Notes,* 8**,** 525.

Cantor-Graae E. and Pedersen C. B. 2007. Risk of schizophrenia in second-generation immigrants: A danish population-based cohort study. *Psychological medicine,* 37**,** 485-494.

Cantor-Graae E. and Selten J.-P. 2005. Schizophrenia and migration: A meta-analysis and review. *American Journal of Psychiatry,* 162**,** 12-24.

Cardno A. G. 2014. Genetics and psychosis. *Advances in psychiatric treatment,* 20**,** 69-70.

Cardno A. G. and Gottesman I. I. 2000. Twin studies of schizophrenia: From bow-and-arrow concordances to star wars mx and functional genomics. *American Journal of Medical Genetics,* 97**,** 12-17.

Cardno A. G., Marshall E., Coid B. and Et Al. 1999. Heritability estimates for psychotic disorders: The maudsley twin psychosis series. *Archives of General Psychiatry,* 56**,** 162-168.

Cariaga-Martinez A., Saiz-Ruiz J. and Alelú-Paz R. 2016. From linkage studies to epigenetics: What we know and what we need to know in the neurobiology of schizophrenia. *Frontiers in Neuroscience,* 10**,** 202.

Carone D. A. 2014. Young child with severe brain volume loss easily passes the word memory test and medical symptom validity test: Implications for mild tbi. *Clinical Neuropsychologist,* 28**,** 146-162.

Carr I. M., Markham S. a. F. and Pena S. D. 2011. Estimating the degree of identity by descent in consanguineous couples. *Human Mutation,* 32**,** 1350-1358.

Carroll D. 2016. Genome editing: Progress and challenges for medical applications. *Genome Medicine,* 8**,** 120.

Carroll L. S. and Owen M. J. 2009. Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Medicine,* 1**,** 1-7.

Carter P. 1986. Site-directed mutagenesis. *Biochemical Journal,* 237**,** 1.

Carter T. C. and He M. M. 2016. Challenges of identifying clinically actionable genetic variants for precision medicine. *Journal of Healthcare Engineering,* 2016**,** 14.

Carulla L. S., Reed G. M., Vaez-Azizi L. M., Cooper S.-A., Leal R. M., Bertelli M., Adnams C., Cooray S., Deb S., Dirani L. A., Girimaji S. C., Katz G., Kwok H., Luckasson R., Simeonsson R., Walsh C., Munir K. and Saxena S. 2011. Intellectual developmental disorders: Towards a new name, definition and framework for "mental retardation/intellectual disability" in icd-11. *World Psychiatry,* 10**,** 175-180.

Carvill G. L. and Mefford H. C. 2015. Next-generation sequencing in intellectual disability. *Journal of Pediatric Genetics,* 4**,** 128-135.

Carvill S. 2001. Sensory impairments, intellectual disability and psychiatry. *Journal of Intellectual Disability Research,* 45**,** 467-483.

Casanova J.-L., Conley M. E., Seligman S. J., Abel L. and Notarangelo L. D. 2014. Guidelines for genetic studies in single patients: Lessons from primary immunodeficiencies. *Journal of Experimental Medicine*, jem. 20140520.

Cassa C. A., Tong M. Y. and Jordan D. M. 2013. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. *Human Mutation,* 34**,** 1216-1220.

Caygill E. E. and Brand A. H. 2016. The gal4 system: A versatile system for the manipulation and analysis of gene expression. *Drosophila: Methods and Protocols*, 33-52.

Cesconetto P. A., Andrade C. M., Cattani D., Domingues J. T., Parisotto E. B., Filho D. W. and Zamoner A. 2016. Maternal exposure to ethanol during pregnancy and lactation affects glutamatergic system and induces oxidative stress in offspring hippocampus. *Alcoholism-Clinical and Experimental Research,* 40**,** 52-61.

Chadwick P. 2008. *Schizophrenia: The positive perspective: Explorations at the outer reaches of human experience*, Routledge.

Chae H., Jung I., Lee H., Marru S., Lee S.-W. and Kim S. 2013. Bio and health informatics meets cloud : Biovlab as an example. *Health Information Science and Systems,* 1**,** 6.

Chapman A. L. 2006. Dialectical behavior therapy: Current indications and unique elements. *Psychiatry (Edgmont),* 3**,** 62-68.

Charara R., Forouzanfar M., Naghavi M., Moradi-Lakeh M., Afshin A., Vos T., Daoud F., Wang H., El Bcheraoui C. and Khalil I. 2017. The burden of mental disorders in the eastern mediterranean region, 1990-2013. *PLoS One,* 12**,** e0169575.

Charsley K. 2013. *Transnational pakistani connections: Marrying 'back home'*, Routledge.

Chavali P. L., Pütz M. and Gergely F. 2014. Small organelle, big responsibility: The role of centrosomes in development and disease. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 369**,** 20130468.

Cheeran M. C. J., Lokensgard J. R. and Schleiss M. R. 2009. Neuropathogenesis of congenital cytomegalovirus infection: Disease mechanisms and prospects for intervention. *Clinical Microbiology Reviews,* 22**,** 99-126.

Chevignard M., Francillette L., Toure H., Brugel D., Meyer P., Vannier A. L., Opatowski M. and Watier L. 2016. Intellectual outcome following childhood severe traumatic brain injury: Results of a prospective longitudinal study: The seven-year follow-up of the tge cohort. *Annals of Physical and Rehabilitation Medicine,* 59, Supplement**,** e132-e133.

Chiang C. H., Su Y., Wen Z., Yoritomo N., Ross C. A., Margolis R. L., Song H. and Ming G. L. 2011. Integration-free induced pluripotent stem cells derived from schizophrenia patients with a disc1 mutation. *Molecular Psychiatry,* 16**,** 358-360.

Chiurazzi P. and Pirozzi F. 2016. Advances in understanding – genetic basis of intellectual disability. *F1000Research,* 5**,** F1000 Faculty Rev-599.

Christensen K. D., Dukhovny D., Siebert U. and Green R. C. 2015. Assessing the costs and cost-effectiveness of genomic sequencing. *Journal of Personalized Medicine,* 5**,** 470-486.

Chu J.-H., Hersh C. P., Castaldi P. J., Cho M. H., Raby B. A., Laird N., Bowler R., Rennard S., Loscalzo J., Quackenbush J. and Silverman E. K. 2014. Analyzing networks of phenotypes in complex diseases: Methodology and applications in copd. *BMC Systems Biology,* 8**,** 1-12.

Chumakov I., Blumenfeld M., Guerassimenko O., Cavarec L., Palicio M., Abderrahim H., Bougueleret L., Barry C., Tanaka H. and La Rosa P. 2002. Genetic and physiological data implicating the new human gene g72 and the gene for d-amino acid oxidase in schizophrenia. *Proceedings of the National Academy of Sciences,* 99**,** 13675-13680.

Cichon S., Mühleisen T. W., Degenhardt F. A., Mattheisen M., Miró X., Strohmaier J., Steffens M., Meesters C., Herms S., Weingarten M., Priebe L., Haenisch B., Alexander M., Vollmer J., Breuer R., Schmäl C., Tessmann P., Moebus S., Wichmann H. E., Schreiber S., Müller-Myhsok B., Lucae S., Jamain S., Leboyer M., Bellivier F., Etain B., Henry C., Kahn J.-P., Heath S., Bipolar Disorder Genome Study C., Hamshere M., O'donovan M. C., Owen M. J., Craddock N., Schwarz M., Vedder H., Kammerer-Ciernioch J., Reif A., Sasse J., Bauer M., Hautzinger M., Wright A., Mitchell P. B., Schofield P. R., Montgomery G. W., Medland S. E., Gordon S. D., Martin N. G., Gustafsson O., Andreassen O., Djurovic S., Sigurdsson E., Steinberg S., Stefansson H., Stefansson K., Kapur-Pojskic L., Oruc L., Rivas F., Mayoral F., Chuchalin A., Babadjanova G., Tiganov A. S., Pantelejeva G., Abramova L. I., Grigoroiu-Serbanescu M., Diaconu C. C., Czerski P. M., Hauser J., Zimmer A., Lathrop M., Schulze T. G., Wienker T. F., Schumacher J., Maier W., Propping P., Rietschel M. and Nöthen M. M. 2011. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American Journal of Human Genetics,* 88**,** 372-381.

Clift K. E., Halverson C. M., Fiksdal A. S., Kumbamu A., Sharp R. R. and Mccormick J. B. 2015. Patients' views on incidental findings from clinical exome sequencing. *Applied & Translational Genomics,* 4**,** 38-43.

Coates P. and Hall P. 2003. The yeast two-hybrid system for identifying protein–protein interactions. *The Journal of Pathology,* 199**,** 4-7.

Coccia M., Brooks S. P., Webb T. R., Christodoulou K., Wozniak I. O., Murday V., Balicki M., Yee H. A., Wangensteen T., Riise R., Saggar A. K., Park S.-M., Kanuga N., Francis P. J., Maher E. R., Moore A. T., Russell-Eggitt I. M. and Hardcastle A. J. 2009. X-linked cataract and nance-horan syndrome are allelic disorders. *Human Molecular Genetics,* 18**,** 2643-2655.

Colloby S. J., Elder G. J., Rabee R., O'brien J. T. and Taylor J. P. 2016. Structural grey matter changes in the substantia innominata in alzheimer's disease and dementia with lewy bodies: A dartel-vbm study. *International Journal of Geriatric Psychiatry*.

Conroy J., Mcgettigan P., Murphy R., Webb D., Murphy S. M., Mccoy B., Albertyn C., Mccreary D., Mcdonagh C. and Walsh O. 2014. A novel locus for episodic ataxia: Ubr4 the likely candidate. *European Journal of Human Genetics,* 22**,** 505-510.

Cooper D. N., Krawczak M., Polychronakos C., Tyler-Smith C. and Kehrer-Sawatzki H. 2013. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics,* 132**,** 1077-1130.

Cooper G. M. and Shendure J. 2011. Needles in stacks of needles: Finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics,* 12**,** 628-640.

Corcoran C., Perrin M., Harlap S., Deutsch L., Fennig S., Manor O., Nahon D., Kimhy D., Malaspina D. and Susser E. 2009. Incidence of schizophrenia among second-generation immigrants in the jerusalem perinatal cohort. *Schizophrenia Bulletin,* 35**,** 596-602.

Corrigan P. W. and Watson A. C. 2002. Understanding the impact of stigma on people with mental illness. *World Psychiatry,* 1**,** 16-20.

Corvin A. P. 2011. Two patients walk into a clinic...A genomics perspective on the future of schizophrenia. *BMC Biology,* 9**,** 77-77.

Craddock N., O'donovan M. C. and Owen M. J. 2007. Phenotypic and genetic complexity of psychosis. Invited commentary on ... Schizophrenia: A common disease caused by multiple rare alleles. *The British journal of psychiatry,* 190**,** 200-3.

Craddock N., O'donovan M. C. and Owen M. J. 2009. Psychosis genetics: Modeling the relationship between schizophrenia, bipolar disorder, and mixed (or "schizoaffective") psychoses. *Schizophrenia Bulletin,* 35**,** 482-490.

Crawford D. C., Carlson C. S., Rieder M. J., Carrington D. P., Yi Q., Smith J. D., Eberle M. A., Kruglyak L. and Nickerson D. A. 2004. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. *The American Journal of Human Genetics,* 74**,** 610-622.

Crocker A. G., Mercier C., Allaire J. F. and Roy M. E. 2007. Profiles and correlates of aggressive behaviour among adults with intellectual disabilities. *Journal of Intellectual Disability Research,* 51**,** 786-801.

Curtis D. 2016a. Practical experience of the application of a weighted burden test to whole exome sequence data for obesity and schizophrenia. *Annals of Human Genetics,* 80**,** 38-49.

Curtis D. 2016b. Schizophrenia genetics moves into the light. *The British Journal of Psychiatry,* 209**,** 93-94.

Cutter G. R. and Liu Y. 2012. Personalized medicine: The return of the house call? *Neurology. Clinical Practice,* 2**,** 343-351.

D'aurizio R., Pippucci T., Tattini L., Giusti B., Pellegrini M. and Magi A. 2016. Enhanced copy number variants detection from whole-exome sequencing data using excavator2. *Nucleic Acids Research,* 44**,** e154-e154.

Dachtler J., Ivorra J. L., Rowland T. E., Lever C., Rodgers R. J. and Clapcote S. J. 2015. Heterozygous deletion of α-neurexin i or α-neurexin ii results in behaviors relevant to autism and schizophrenia. *Behavioral Neuroscience,* 129**,** 765.

Dammeyer J. 2012. Children with usher syndrome: Mental and behavioral disorders. *Behavioral and Brain Functions,* 8**,** 16.

Darr A. and Modell B. 1988. The frequency of consanguineous marriage among british pakistanis. *Journal of Medical Genetics,* 25**,** 186-190.

Dauncey M. J. 2014. Nutrition, the brain and cognitive decline: Insights from epigenetics. *European Journal of Clinical Nutrition,* 68**,** 1179-1185.

Dauvermann M. R., Whalley H. C., Schmidt A., Lee G. L., Romaniuk L., Roberts N., Johnstone E. C., Lawrie S. M. and Moorhead T. W. 2014. Computational neuropsychiatry–schizophrenia as a cognitive brain network disorder. *Frontiers in Psychiatry*, doi:10.3389/00030.

David A. and Sternberg M. J. E. 2015. The contribution of missense mutations in core and rim residues of protein–protein interfaces to human disease. *Journal of Molecular Biology,* 427**,** 2886-2898.

Davies L., Lewis S., Jones P., Barnes T., Gaughran F., Hayhurst K., Markwick A. and Lloyd H. 2007. Cost-effectiveness of first-v. Second-generation antipsychotic drugs: Results from a randomised controlled trial in schizophrenia responding poorly to previous therapy. *The British Journal of Psychiatry,* 191**,** 14-22.

De Ligt J., Willemsen M. H., Van Bon B. W., Kleefstra T., Yntema H. G., Kroes T., Vulto-Van Silfhout A. T., Koolen D. A., De Vries P. and Gilissen C. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine,* 367**,** 1921-1929.

Deamer D. W. and Akeson M. 2000. Nanopores and nucleic acids: Prospects for ultrarapid sequencing. *Trends in Biotechnology,* 18**,** 147-151.

Deb S., Kwok H., Bertelli M., Salvador-Carulla L., Bradley E., Torr J. and Barnhill J. 2009. International guide to prescribing psychotropic medication for the management of problem behaviours in adults with intellectual disabilities. *World Psychiatry,* 8**,** 181-186.

Depristo M. A., Banks E., Poplin R., Garimella K. V., Maguire J. R., Hartl C., Philippakis A. A., Del Angel G., Rivas M. A. and Hanna M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics,* 43**,** 491-498.

Detera-Wadleigh S. D., Badner J. A., Berrettini W. H., Yoshikawa T., Goldin L. R., Turner G., Rollins D. Y., Moses T., Sanders A. R. and Karkera J. D. 1999. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11. 2. *Proceedings of the National Academy of Sciences,* 96**,** 5604-5609.

Detera-Wadleigh S. D., Badner J. A., Yoshikawa T., Sanders A. R., Goldin L. R., Turner G., Rollins D. Y., Moses T., Guroff J. J., Kazuba D., Maxwell M. E., Edenberg H. J., Foroud T., Lahiri D., Nurnberger J. I., Stine O. C., Mcmahon F., Meyers D. A.,

Mackinnon D., Simpson S., Mcinnis M., Depaulo J. R., Rice J., Goate A., Reich T., Blehar M. C. and Gershon E. S. 1997. Initial genome scan of the nimh genetics initiative bipolar pedigrees: Chromosomes 4, 7, 9, 18, 19, 20, and 21q. *American Journal of Medical Genetics,* 74**,** 254-262.

Dhindsa R. S. and Goldstein D. B. 2016. Schizophrenia: From genetics to physiology at last. *Nature,* 530**,** 162-163.

Di Iorio G., Baroni G., Lorusso M., Montemitro C., Spano M. C. and Di Giannantonio M. 2017. Efficacy of memantine in schizophrenic patients: A systematic review. *Journal of Amino Acids,* 2017**,** 7021071.

Dienstmann R., Dong F., Borger D., Dias-Santagata D., Ellisen L. W., Le L. P. and Iafrate A. J. 2014. Standardized decision support in next generation sequencing reports of somatic cancer variants. *Molecular Oncology,* 8**,** 859-873.

Ding X., Patel M., Herzlich A. A., Sieving P. C. and Chan C.-C. 2009. Ophthalmic pathology of nance-horan syndrome: Case report and review of the literature. *Ophthalmic Genetics,* 30**,** 127-135.

Doherty J. L. and Owen M. J. 2014. Genomic insights into the overlap between psychiatric disorders: Implications for research and clinical practice. *Genome Medicine,* 6**,** 29-29.

Dölen G., Osterweil E., Rao B. S., Smith G. B., Auerbach B. D., Chattarji S. and Bear M. F. 2007. Correction of fragile x syndrome in mice. *Neuron,* 56**,** 955-962.

Domanico D., Fragiotta S., Cutini A., Grenga P. L. and Vingolo E. M. 2015. Psychosis, mood and behavioral disorders in usher syndrome: Review of the literature. *Medical Hypothesis, Discovery and Innovation in Ophthalmology,* 4**,** 50.

Dongen J., Buck N. and Van Marle H. 2015. Unravelling offending in schizophrenia: Factors characterising subgroups of offenders. *Criminal Behaviour and Mental Health,* 25**,** 88-98.

Dow A. L., Lin T. V., Chartoff E. H., Potter D., Mcphie D. L., Van't Veer A. V., Knoll A. T., Lee K. N., Neve R. L. and Patel T. B. 2015. Sprouty2 in the dorsal hippocampus regulates neurogenesis and stress responsiveness in rats. *PloS One,* 10**,** e0120693.

Duncan E., Brown M. and Shore E. M. 2014. The revolution in human monogenic disease mapping. *Genes,* 5**,** 792-803.

Dyment D. A., Sawyer S. L., Chardon J. W. and Boycott K. M. 2013. Recent advances in the genetic etiology of brain malformations. *Current Neurology and Neuroscience Reports,* 13**,** 364.

Easton D. F., Pooley K. A., Dunning A. M., Pharoah P. D. P., Thompson D., Ballinger D. G., Struewing J. P., Morrison J., Field H., Luben R., Wareham N., Ahmed S., Healey C. S., Bowman R., Meyer K. B., Haiman C. A., Kolonel L. K., Henderson B. E., Le Marchand L., Brennan P., Sangrajrang S., Gaborieau V., Odefrey F., Shen C.-Y., Wu P.-E., Wang H.-C., Eccles D., Evans D. G., Peto J., Fletcher O., Johnson N., Seal S., Stratton M. R., Rahman N., Chenevix-Trench G., Bojesen S. E., Nordestgaard B. G., Axelsson C. K., Garcia-Closas M., Brinton L., Chanock S., Lissowska J., Peplonska B., Nevanlinna H., Fagerholm R., Eerola H., Kang D., Yoo K.-Y., Noh D.-Y., Ahn S.-H., Hunter D. J., Hankinson S. E., Cox D. G., Hall P., Wedren S., Liu J., Low Y.-L., Bogdanova N., Schurmann P., Dork T., Tollenaar R. a. E. M., Jacobi C. E., Devilee P., Klijn J. G. M., Sigurdson A. J., Doody M. M., Alexander B. H., Zhang J., Cox A., Brock I. W., Macpherson G., Reed M. W. R.,

Couch F. J., Goode E. L., Olson J. E., Meijers-Heijboer H., Van Den Ouweland A., Uitterlinden A., Rivadeneira F., Milne R. L., Ribas G., Gonzalez-Neira A., Benitez J., Hopper J. L., Mccredie M., Southey M., Giles G. G., Schroen C., Justenhoven C., Brauch H., Hamann U., Ko Y.-D., Spurdle A. B., Beesley J., Chen X., Mannermaa A., Kosma V.-M., Kataja V., Hartikainen J., Day N. E., et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447**,** 1087-1093.

Ebermann I., Scholl H. P. N., Charbel Issa P., Becirovic E., Lamprecht J., Jurklies B., Millán J. M., Aller E., Mitter D. and Bolz H. 2006. A novel gene for usher syndrome type 2: Mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Human Genetics*, 121**,** 203-211.

Edvardson S., Murakami Y., Nguyen T. T. M., Shahrour M., St-Denis A., Shaag A., Damseh N., Le Deist F., Bryceson Y. and Abu-Libdeh B. 2016. Mutations in the phosphatidylinositol glycan c (pigc) gene are associated with epilepsy and intellectual disability. *Journal of Medical Genetics*, jmedgenet-2016-104202.

Edward H. and Jhan L. 2003. Subtypes of schizophrenia: A cluster analytic approach. *The Canadian Journal of Psychiatry*, 48.

Ehret J. K., Engels H., Cremer K., Becker J., Zimmermann J. P., Wohlleber E., Grasshoff U., Rossier E., Bonin M. and Mangold E. 2015. Microdeletions in 9q33. 3-q34. 11 in five patients with intellectual disability, microcephaly, and seizures of incomplete penetrance: Is stxbp1 not the only causative gene? *Molecular cytogenetics*, 8**,** 72.

Eisen E. J. 2005. *The mouse in animal genetics and breeding research*, Imperial College Press.

Eisengart J. B., Rudser K. D., Tolar J., Orchard P. J., Kivisto T., Ziegler R. S., Whitley C. B. and Shapiro E. G. 2013. Enzyme replacement is associated with better cognitive outcomes after transplant in hurler syndrome. *The Journal of Pediatrics*, 162**,** 375-380.e1.

Ejima A. and Griffith L. C. 2011. Assay for courtship suppression in drosophila. *Cold Spring Harb Protoc*, 2011**,** t5575.

Ejima A., Smith B. P., Lucas C., Van Der Goes Van Naters W., Miller C. J., Carlson J. R., Levine J. D. and Griffith L. C. 2007. Generalization of courtship learning in drosophila is mediated by cis-vaccenyl acetate. *Current Biology*, 17**,** 599-605.

Ekblom R. and Wolf J. B. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7**,** 1026-1042.

El-Asrag Mohammed e., Sergouniotis Panagiotis i., Mckibbin M., Plagnol V., Sheridan E., Waseem N., Abdelhamed Z., Mckeefry D., Van schil K., Poulter James a., Black G., Hall G., Ingram S., Gillespie R., Ramsden S., Manson F., Hardcastle A., Michaelides M., Cheetham M., Arno G., Thomas N., Bhattacharya S., Moore T., Nemeth A., Downes S., Lise S., Lord E., Johnson C. A., Carr I. M., Leroy B. P., De Baere E., Inglehearn C. F., Webster A. R., Toomes C. and Ali M. 2015. Biallelic mutations in the autophagy regulator dram2 cause retinal dystrophy with early macular involvement. *The American Journal of Human Genetics*, 96**,** 948-954.

El-Metwally S., Ouda O. M. and Helmy M. 2014. *Next generation sequencing technologies and challenges in sequence assembly*, Springer Science & Business.

El Chehadeh S., Bonnet C., Callier P., Béri M., Dupré T., Payet M., Ragon C., Mosca-Boidron A. L., Marle N., Mugneret F., Masurel-Paulet A., Thevenon J., Seta N., Duplomb L., Jonveaux P., Faivre L. and Thauvin-Robinet C. 2015. Homozygous

truncating intragenic duplication in tusc3 responsible for rare autosomal recessive nonsyndromic intellectual disability with no clinical or biochemical metabolic markers. *Journal of Inherited Metabolic Disease Reports,* 20**,** 45-55.

Ellegren H. 2004. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics,* 5**,** 435-445.

Emerson E. and Hatton C. 2007. Contribution of socioeconomic position to health inequalities of british children and adolescents with intellectual disabilities. *American Journal on Mental Retardation,* 112**,** 140-150.

Emlen S. T. 1969. Homing ability and orientation in the painted turtle chrysemys picta marginata. *Behaviour,* 33**,** 58-76.

Engman M.-L., Sundin M., Miniscalco C., Westerlund J., Lewensohn-Fuchs I., Gillberg C. and Fernell E. 2015. Prenatal acquired cytomegalovirus infection should be considered in children with autism. *Acta Paediatrica,* 104**,** 792-795.

Erritty P. and Wydell T. N. 2013. Are lay people good at recognising the symptoms of schizophrenia? *PLoS One,* 8.

Euesden J., Lewis C. M. and O'reilly P. F. 2015. Prsice: Polygenic risk score software. *Bioinformatics,* 31**,** 1466-1468.

Evrony G. D., Cai X., Lee E., Hills L. B., Elhosary P. C., Lehmann H. S., Parker J. J., Atabay K. D., Gilmore E. C., Poduri A., Park P. J. and Walsh C. A. 2012. Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. *Cell,* 151**,** 483-496.

Ewart G. D. and Howells A. J. 1998. [15] abc transporters involved in transport of eye pigment precursors in drosophila melanogaster. *Methods in Enzymology,* 292**,** 213-224.

Faheem M., Naseer M. I., Rasool M., Chaudhary A. G., Kumosani T. A., Ilyas A. M., Pushparaj P. N., Ahmed F., Algahtani H. A., Al-Qahtani M. H. and Saleh Jamal H. 2015. Molecular genetics of human primary microcephaly: An overview. *BMC Medical Genomics,* 8**,** S4-S4.

Fatemi S. H. and Folsom T. D. 2009. The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophrenia Bulletin,* 35**,** 528-548.

Feng Y., Zhang Y., Ying C., Wang D. and Du C. 2015. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics & Bioinformatics,* 13, 4-16.

Fernandez B. A., Roberts W., Chung B., Weksberg R., Meyn S., Szatmari P., Joseph-George A. M., Mackay S., Whitten K. and Noble B. 2009. Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11. 2 in individuals ascertained for diagnosis of autism spectrum disorder. *Journal of Medical Genetics,* 47(3):195-203.

Fernández M. E., Goszczynski D. E., Lirón J. P., Villegas-Castagnasso E. E., Carino M. H., Ripoli M. V., Rogberg-Muñoz A., Posik D. M., Peral-García P. and Giovambattista G. 2013. Comparison of the effectiveness of microsatellites and snp panels for genetic identification, traceability and assessment of parentage in an inbred angus herd. *Genetics and Molecular Biology,* 36**,** 185-191.

Ferraren D. O., Liu C., Badner J. A., Corona W., Rezvani A., Monje V. D., Gershon E. S., Bonner T. I. and Detera-Wadleigh S. D. 2005. Linkage disequilibrium analysis in the loc93081-kdelc1-bivm region on 13q in bipolar disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 133**,** 12-17.

Fettiplace R. 2016. Is tmc1 the hair cell mechanotransducer channel? *Biophysical Journal,* 111**,** 3-9.

Filippi M. 2015. *Oxford textbook of neuroimaging*, Oxford University Press.

Finnema S. J., Nabulsi N. B., Eid T., Detyniecki K., Lin S.-F., Chen M.-K., Dhaher R., Matuskey D., Baum E. and Holden D. 2016. Imaging synaptic density in the living human brain. *Science Translational Medicine,* 8**,** 348ra96-348ra96.

Fonagy P. 2015. The effectiveness of psychodynamic psychotherapies: An update. *World Psychiatry,* 14**,** 137-150.

Foreman P. 2009. *Education of students with an intellectual disability: Research and practice*, IAP.

Forsythe E. and Beales P. L. 2013. Bardet–biedl syndrome. *European Journal of Human Genetics,* 21**,** 8-13.

Frank S. A. 2014. Somatic mosaicism and disease. *Current Biology,* 24**,** R577-R581.

Frederic mery, Juliette pont, Thomas preat and Tadeusz j. kawecki 2007. Experimental evolution of olfactory memory in drosophila melanogaster. *Physiological and Biochemical Zoology,* 80**,** 399-405.

Freedman R., Leonard S., Olincy A., Kaufmann C. A., Malaspina D., Cloninger C. R., Svrakic D., Faraone S. V. and Tsuang M. T. 2001. Evidence for the multigenic inheritance of schizophrenia. *American Journal of Medical Genetics,* 105**,** 794-800.

Frints S. G. M., Froyen G., Marynen P. and Fryns J. P. 2002. X-linked mental retardation: Vanishing boundaries between non-specific (mrx) and syndromic (mrxs) forms. *Clinical Genetics,* 62**,** 423-432.

Fromer M., Moran J. L., Chambert K., Banks E., Bergen S. E., Ruderfer D. M., Handsaker R. E., Mccarroll S. A., O'donovan M. C. and Owen M. J. 2012. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *The American Journal of Human Genetics,* 91**,** 597-607.

Fuster V. and Colantonio S. E. 2004. Socioeconomic, demographic, and geographic variables affecting the diverse degrees of consanguineous marriages in spain. *Human Biology,* 76**,** 1-14.

Gabos P. G., Inan M., Thacker M. and Borkhu B. 2012. Spinal fusion for scoliosis in rett syndrome with an emphasis on early postoperative complications. *Spine,* 37**,** E90-E94.

Gaebel W. and Zielasek J. 2015. Focus on psychosis. *Dialogues in Clinical Neuroscience,* 17**,** 9-18.

Gaebel W., Zielasek J. and Cleveland H.-R. 2013. Psychotic disorders in icd-11. *Asian Journal of Psychiatry,* 6**,** 263-265.

Gandhi S. and Wood N. W. 2010. Genome-wide association studies: The key to unlocking neurodegeneration? *Nature Neuroscience,* 13**,** 789-794.

Gandomi S. K., Farwell Gonzalez K. D., Parra M., Shahmirzadi L., Mancuso J., Pichurin P., Temme R., Dugan S., Zeng W. and Tang S. 2014. Diagnostic exome sequencing identifies two novel iqsec2 mutations associated with x-linked intellectual disability with seizures: Implications for genetic counseling and clinical diagnosis. *Journal of Genetic Counseling,* 23**,** 289-298.

Gargano J. W., Martin I., Bhandari P. and Grotewiel M. S. 2005. Rapid iterative negative geotaxis (ring): A new method for assessing age-related locomotor decline in drosophila. *Experimental Gerontology,* 40**,** 386-395.

Garshasbi M., Hadavi V., Habibi H., Kahrizi K., Kariminejad R., Behjati F., Tzschach A., Najmabadi H., Ropers H. H. and Kuss A. W. 2008. A defect in the tusc3 gene is associated with autosomal recessive mental retardation. *The American Journal of Human Genetics,* 82**,** 1158-1164.

Garshasbi M., Kahrizi K., Hosseini M., Nouri Vahid L., Falah M., Hemmati S., Hu H., Tzschach A., Ropers H. H., Najmabadi H. and Kuss A. W. 2011. A novel nonsense mutation in tusc3 is responsible for non-syndromic autosomal recessive mental retardation in a consanguineous iranian family. *American Journal of Medical Genetics Part A,* 155A**,** 1976-80.

Gawad C., Koh W. and Quake S. R. 2016. Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics,* 17**,** 175-188.

Gecz J., Turner G., Nelson J. and Partington M. 2006. The borjeson-forssman-lehman syndrome (bfls, mim [num]301900). *European Journal of Human Genetics,* 14**,** 1233-1237.

Geddes K., Dziurawiec S. and Lee C. W. 2013. Dialectical behaviour therapy for the treatment of emotion dysregulation and trauma symptoms in self-injurious and suicidal adolescent females: A pilot programme within a community-based child and adolescent mental health service. *Psychiatry journal,* doi.10.1155/145219.

Gelenberg A. J., Bassuk E. L. and Schoonover S. C. 2013. *The practitioner's guide to psychoactive drugs*, Springer Science & Business Media.

Ghodse H. 2011. *International perspectives on mental health*, RCPsych Publications.

Giaroli G., Bass N., Strydom A., Rantell K. and Mcquillin A. 2014. Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: A systematic review of literature and meta-analysis. *Schizophrenia Research,* 159**,** 340-346.

Gilissen C., Hehir-Kwa J. Y., Thung D. T., Van De Vorst M., Van Bon B. W. M., Willemsen M. H., Kwint M., Janssen I. M., Hoischen A., Schenck A., Leach R., Klein R., Tearle R., Bo T., Pfundt R., Yntema H. G., De Vries B. B. A., Kleefstra T., Brunner H. G., Vissers L. E. L. M. and Veltman J. A. 2014b. Genome sequencing identifies major causes of severe intellectual disability. *Nature,* 511**,** 344-347.

Gilissen C., Hoischen A., Brunner H. G. and Veltman J. A. 2012. Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics,* 20**,** 490-497.

Gillette-Guyonnet S., Secher M. and Vellas B. 2013. Nutrition and neurodegeneration: Epidemiological evidence and challenges for future research. *British Journal of Clinical Pharmacology,* 75**,** 738-755.

Girard S. L., Gauthier J., Noreau A., Xiong L., Zhou S., Jouan L., Dionne-Laporte A., Spiegelman D., Henrion E., Diallo O., Thibodeau P., Bachand I., Bao J. Y. J., Tong A. H. Y., Lin C.-H., Millet B., Jaafari N., Joober R., Dion P. A., Lok S., Krebs M.-O. and Rouleau G. A. 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature Genetics,* 43**,** 860-863.

Giusti-Rodríguez P. and Sullivan P. F. 2013. The genomics of schizophrenia: Update and implications. *The Journal of Clinical Investigation,* 123**,** 4557-4563.

Glöckle N., Kohl S., Mohr J., Scheurenbrand T., Sprecher A., Weisschuh N., Bernd A., Rudolph G., Schubach M. and Poloschek C. 2014. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *European Journal of Human Genetics,* 22**,** 99-104.

Glover G., Williams R., Branford D., Avery R., Chauhan U., Hoghton M. and Bernard S. 2015. Prescribing of psychotropic drugs to people with learning disabilities and/or autism by general practitioners in england. *Public Health England*, FR/ID/09.

Goddard K. A., Hopkins P. J., Hall J. M. and Witte J. S. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *The American Journal of Human Genetics*, 66, 216-234.

Goharpey N., Crewther D. P. and Crewther S. G. 2013. Problem solving ability in children with intellectual disability as measured by the raven's colored progressive matrices. *Research in Developmental Disabilities*, 34, 4366-4374.

Goker-Alpan O., Schiffmann R., Lamarca M., Nussbaum R., Mcinerney-Leo A. and Sidransky E. 2004. Parkinsonism among gaucher disease carriers. *Journal of Medical Genetics*, 41, 937-940.

Gómez J., Marín-Méndez J. J., Molero P., Atakan Z. and Ortuño F. 2014. Time perception networks and cognition in schizophrenia: A review and a proposal. *Psychiatry research*, 220, 737-744.

Goodwin S., Mcpherson J. D. and Mccombie W. R. 2016. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333-351.

Gottesman I. I. 1991. *Schizophrenia genesis: The origins of madness*, WH Freeman/Times Books/Henry Holt & Co.

Gottesman I. I. and Erlenmeyer-Kimling L. 2001. Family and twin strategies as a head start in defining prodromes and endophenotypes for hypothetical early-interventions in schizophrenia. *Schizophrenia Research*, 51, 93-102.

Gøtzsche P. C. 2015. *Deadly psychiatry and organised denial*, Art People.

Grant S. G. N. 2012. Synaptopathies: Diseases of the synaptome. *Current Opinion in Neurobiology*, 22, 522-529.

Gratten J. and Visscher P. M. 2016. Genetic pleiotropy in complex traits and diseases: Implications for genomic medicine. *Genome Medicine*, 8, 1-3.

Green J. A., Yang J., Grati M. H., Kachar B. and Bhat M. A. 2013. Whirlin, a cytoskeletal scaffolding protein, stabilizes the paranodal region and axonal cytoskeleton in myelinated axons. *BMC neuroscience*, 14, 96.

Guo A., Zhang K., Peng Y. and Xi W. 2010. Research progress on drosophila visual cognition in china. *Science China Life Sciences*, 53, 374-384.

Gur R. E. and Gur R. C. 2010. Functional magnetic resonance imaging in schizophrenia. *Dialogues in Clinical Neuroscience*, 12, 333-343.

Gussow A. B., Petrovski S., Wang Q., Allen A. S. and Goldstein D. B. 2016. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. *Genome Biology*, 17, 9.

Gustavsson A., Svensson M., Jacobi F., Allgulander C., Alonso J., Beghi E., Dodel R., Ekman M., Faravelli C., Fratiglioni L., Gannon B., Jones D. H., Jennum P., Jordanova A., Jönsson L., Karampampa K., Knapp M., Kobelt G., Kurth T., Lieb R., Linde M., Ljungcrantz C., Maercker A., Melin B., Moscarelli M., Musayev A., Norwood F., Preisig M., Pugliatti M., Rehm J., Salvador-Carulla L., Schlehofer B., Simon R., Steinhausen H.-C., Stovner L. J., Vallat J.-M., Den Bergh P. V., Van Os J., Vos P., Xu W., Wittchen H.-U., Jönsson B. and Olesen J. 2011. Cost of disorders of the brain in europe 2010. *European Neuropsychopharmacology*, 21, 718-779.

Hagmann P., Cammoun L., Gigandet X., Meuli R., Honey C. J., Wedeen V. J. and Sporns O. 2008. Mapping the structural core of human cerebral cortex. *PLoS Biol,* 6**,** e159.

Hagmann P., Kurant M., Gigandet X., Thiran P., Wedeen V. J., Meuli R. and Thiran J.-P. 2007. Mapping human whole-brain structural networks with diffusion mri. *PloS One,* 2**,** e597.

Haines D. E. and Ard M. D. 2013. *Fundamental neuroscience for basic and clinical applications, with student consult online access, 4: Fundamental neuroscience for basic and clinical applications*, Elsevier Health Sciences.

Hakonarson H., Grant S. F. A., Bradfield J. P., Marchand L., Kim C. E., Glessner J. T., Grabs R., Casalunovo T., Taback S. P., Frackelton E. C., Lawson M. L., Robinson L. J., Skraban R., Lu Y., Chiavacci R. M., Stanley C. A., Kirsch S. E., Rappaport E. F., Orange J. S., Monos D. S., Devoto M., Qu H.-Q. and Polychronakos C. 2007. A genome-wide association study identifies kiaa0350 as a type 1 diabetes gene. *Nature,* 448**,** 591-594.

Haldane J. B. S. 1934. Methods for the detection of autosomal linkage in man. *Annals of Eugenics,* 6**,** 26-65.

Hamamy H. 2012. Consanguineous marriages: Preconception consultation in primary health care settings. *Journal of Community Genetics,* 3**,** 185-192.

Harman L. B., Flite C. A. and Bond K. 2012. Electronic health records: Privacy, confidentiality, and security. *Virtual Mentor,* 14**,** 712.

Harripaul R., Vasli N., Mikhailov A., Rafiq M. A., Mittal K., Windpassinger C., Sheikh T., Noor A., Mahmood H. and Downey S. 2017. Mapping autosomal recessive intellectual disability: Combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry*, doi.10.1038/.60.

Harris E. E. 2015. *Ancestors in our genome: The new science of human evolution*, Oxford University Press, USA.

Hartling L., Abou-Setta A. M., Dursun S., Mousavi S. S., Pasichnyk D. and Newton A. S. 2012. Antipsychotics in adults with schizophrenia: Comparative effectiveness of first-generation versus second-generation medications: A systematic review and meta-analysis. *Annals of Internal Medicine,* 157**,** 498-511.

Hassiotis A., Robotham D., Canagasabey A., Romeo R., Langridge D., Blizard R., Murad S. and King M. 2009. Randomized, single-blind, controlled trial of a specialist behavior therapy team for challenging behavior in adults with intellectual disabilities. *American Journal of Psychiatry,* 166**,** 1278-1285.

Hattori E., Liu C., Badner J. A., Bonner T. I., Christian S. L., Maheshwari M., Detera-Wadleigh S. D., Gibbs R. A. and Gershon E. S. 2003. Polymorphisms at the g72/g30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. *The American Journal of Human Genetics,* 72**,** 1131-1140.

Hayashi-Takagi A., Barker P. B. and Sawa A. 2011. Readdressing synaptic pruning theory for schizophrenia: Combination of brain imaging and cell biology. *Communicative & Integrative Biology,* 4**,** 211-212.

Hayashi M. L., Choi S.-Y., Rao B. S. S., Jung H.-Y., Lee H.-K., Zhang D., Chattarji S., Kirkwood A. and Tonegawa S. 2004. Altered cortical synaptic morphology and impaired memory consolidation in forebrain- specific dominant-negative pak transgenic mice. *Neuron,* 42**,** 773-787.

He Y., Chen Z. J. and Evans A. C. 2007. Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cerebral Cortex,* 17**,** 2407-2419.

Heard E. and Disteche C. M. 2006. Dosage compensation in mammals: Fine-tuning the expression of the x chromosome. *Genes & Development,* 20**,** 1848-1867.

Hearne C. M., Mcaleer M. A., Love J. M., Aitman T. J., Cornall R. J., Ghosh S., Knight A. M., Prins J.-B. and Todd J. A. 1991. Additional microsatellite markers for mouse genome mapping. *Mammalian Genome,* 1**,** 273-282.

Hegde S. 2017. A review of indian research on cognitive remediation for schizophrenia. *Asian Journal of Psychiatry,* 25**,** 54-59.

Henderson R., Williams P., Gabbidon J., Farrelly S., Schauman O., Hatch S., Thornicroft G., Bhugra D. and Clement S. 2015. Mistrust of mental health services: Ethnicity, hospital admission and unfair treatment. *Epidemiology and Psychiatric Sciences,* 24**,** 258.

Henning M. S., Morham S. G., Goff S. P. and Naghavi M. H. 2010. Pdzd8 is a novel gag-interacting factor that promotes retroviral infection. *Journal of Virology,* 84**,** 8990-8995.

Henning M. S., Stiedl P., Barry D. S., Mcmahon R., Morham S. G., Walsh D. and Naghavi M. H. 2011. Pdzd8 is a novel moesin-interacting cytoskeletal regulatory protein that suppresses infection by herpes simplex virus type 1. *Virology,* 415**,** 114-121.

Hensch T. K. and Bilimoria P. M. 2012. Re-opening windows: Manipulating critical periods for brain development. *Cerebrum: the Dana Forum on Brain Science,* 2012**,** 11.

Herculano-Houzel S. 2009. The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience,* 3.

Herr K. J., Herr D. R., Lee C.-W., Noguchi K. and Chun J. 2011. Stereotyped fetal brain disorganization is induced by hypoxia and requires lysophosphatidic acid receptor 1 (lpa(1)) signaling. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 15444-15449.

Herring S., Gray K., Taffe J., Tonge B., Sweeney D. and Einfeld S. 2006. Behaviour and emotional problems in toddlers with pervasive developmental disorders and developmental delay: Associations with parental mental health and family functioning. *Journal of Intellectual Disability Research,* 50**,** 874-882.

Ho A., Murphy M., Wilson S., Atlas S. R. and Edwards J. S. 2011a. Sequencing by ligation variation with endonuclease v digestion and deoxyinosine-containing query oligonucleotides. *BMC Genomics,* 12**,** 1-8.

Ho V. M., Lee J.-A. and Martin K. C. 2011b. The cell biology of synaptic plasticity. *Science (New York, N.y.),* 334**,** 623-628.

Hoffjan S. 2012. Genetic dissection of marfan syndrome and related connective tissue disorders: An update 2012. *Molecular Syndromology,* 3**,** 47-58.

Hoffman D. A., Schiller M., Greenblatt J. M. and Iosifescu D. V. 2011. Polypharmacy or medication washout: An old tool revisited. *Neuropsychiatric Disease and Treatment,* 7**,** 639-648.

Hofman K. J., Bernhardt B. A., Pyeritz R. E., Opitz J. M. and Reynolds J. F. 1988. Marfan syndrome: Neuropsychological aspects. *American Journal of Medical Genetics,* 31, 331-338.

Hoischen A., Krumm N. and Eichler E. E. 2014. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nature Neuroscience,* 17**,** 764-772.

Hollander A.-C., Dal H., Lewis G., Magnusson C., Kirkbride J. B. and Dalman C. 2016. Refugee migration and risk of schizophrenia and other non-affective psychoses: Cohort study of 1.3 million people in sweden. *British Medical Journal,* 352**,** i1030.

Holley R. W., Apgar J., Everett G. A., Madison J. T., Marquisee M., Merrill S. H., Penswick J. R. and Zamir A. 1965. Structure of a ribonucleic acid. *Science,* 147**,** 1462-1465.

Homberg J. R., Kyzar E. J., Scattoni M. L., Norton W. H., Pittman J., Gaikwad S., Nguyen M., Poudel M. K., Ullmann J. F. P., Diamond D. M., Kaluyeva A. A., Parker M. O., Brown R. E., Song C., Gainetdinov R. R., Gottesman I. I. and Kalueff A. V. 2016. Genetic and environmental modulation of neurodevelopmental disorders: Translational insights from labs to beds. *Brain Research Bulletin,* 125**,** 79-91.

Hong N., Chen Y.-H., Xie C., Xu B.-S., Huang H., Li X., Yang Y.-Q., Huang Y.-P., Deng J.-L., Qi M. and Gu Y.-S. 2014. Identification of a novel mutation in a chinese family with nance-horan syndrome by whole exome sequencing. *Journal of Zhejiang University. Science. B,* 15**,** 727-734.

Hooley J. M. 2010. Social factors in schizophrenia. *Current Directions in Psychological Science,* 19**,** 238-242.

Horan M. B. and Billson F. 1974. X-linked cataract and hutchinsonian teeth. *Journal of Paediatrics and Child Health,* 10**,** 98-102.

Howes O., Mccutcheon R. and Stone J. 2015. Glutamate and dopamine in schizophrenia: An update for the 21(st) century. *Journal of Psychopharmacology (Oxford, England),* 29**,** 97-115.

Hrdlickova B., De Almeida R. C., Borek Z. and Withoff S. 2014. Genetic variation in the non-coding genome: Involvement of micro-rnas and long non-coding rnas in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1842**,** 1910-1922.

Huang J., Zhu T., Qu Y. and Mu D. 2016a. Prenatal, perinatal and neonatal risk factors for intellectual disability: A systemic review and meta-analysis. *PLoS ONE,* 11**,** e0153655.

Huang X., Chen P. C. and Poole C. 2004. Apoe-ε2 allele associated with higher prevalence of sporadic parkinson disease. *Neurology,* 62**,** 2198-2202.

Huang Y., Yu S., Wu Z. and Tang B. 2014. Genetics of hereditary neurological disorders in children. *Translational Pediatrics,* 3**,** 108-119.

Huang Z., Rustagi N., Veeraraghavan N., Carroll A., Gibbs R., Boerwinkle E., Venkata M. G. and Yu F. 2016b. A hybrid computational strategy to address wgs variant analysis in> 5000 samples. *BMC bioinformatics,* 17**,** 361.

Hutchinson E. A., De Luca C. R., Doyle L. W., Roberts G., Anderson P. J. and Group V. I. C. S. 2013. School-age outcomes of extremely preterm or extremely low birth weight children. *Pediatrics***,** peds. 2012-2311.

Inaba Y., Schwartz C. E., Bui Q. M., Li X., Skinner C., Field M., Wotton T., Hagerman R. J., Francis D., Amor D. J., Hopper J. L., Loesch D. Z., Bretherton L., Slater H. R. and Godler D. E. 2014. Early detection of fragile x syndrome: Applications of a novel approach for improved quantitative methylation analysis in venous blood and newborn blood spots. *Clinical Chemistry,* 60**,** 963-973.

Insel T. 2014. Brain somatic mutations: The dark matter of psychiatric genetics? *Molecular Psychiatry,* 19**,** 156-158.

Iqbal Z., Vandeweyer G., Van Der Voet M., Waryah A. M., Zahoor M. Y., Besseling J. A., Roca L. T., Vulto-Van Silfhout A. T., Nijhof B., Kramer J. M., Van Der Aa N., Ansar M., Peeters H., Helsmoortel C., Gilissen C., Vissers L. E. L. M., Veltman J. A., De Brouwer A. P. M., Frank Kooy R., Riazuddin S., Schenck A., Van Bokhoven H. and Rooms L. 2013. Homozygous and heterozygous disruptions of ank3: At the crossroads of neurodevelopmental and psychiatric disorders. *Human Molecular Genetics,* 22**,** 1960-1970.

Jacquemont S., Hagerman R. J., Leehey M. A. and Et Al. 2004. Penetrance of the fragile x–associated tremor/ataxia syndrome in a premutation carrier population. *The Journal of the American Medical Association, 2*91**,** 460-469.

Jagannath A., Peirson S. N. and Foster R. G. 2013. Sleep and circadian rhythm disruption in neuropsychiatric illness. *Current Opinion in Neurobiology,* 23**,** 888-894.

Jahani-Asl A., Cheng C., Zhang C. and Bonni A. 2016. Pathogenesis of börjeson-forssman-lehmann syndrome: Insights from phf6 function. *Neurobiology of Disease,* 96**,** 227-235.

Jain K. K. 2013. *Applications of biotechnology in neurology*, Springer Science & Business Media.

Jalanko A. and Braulke T. 2009. Neuronal ceroid lipofuscinoses. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1793**,** 697-709.

Janssen C., Schuengel C. and Stolk J. 2002. Understanding challenging behaviour in people with severe and profound intellectual disability: A stress-attachment model. *Journal of Intellectual Disability Research,* 46**,** 445-453.

Jasnic-Savovic J., Nestorovic A., Savic S., Karasek S., Vitulo N., Valle G., Faulkner G., Radojkovic D. and Kojic S. 2015. Profiling of skeletal muscle ankrd2 protein in human cardiac tissue and neonatal rat cardiomyocytes. *Histochemistry and Cell Biology,* 143**,** 583-597.

Jauhari S. and Rizvi S. 2015. An indian eye to personalized medicine. *Computers in Biology and Medicine,* 59**,** 211-220.

Jeong C.-J., Yang S.-H., Xie Y., Zhang L., Johnston S. A. and Kodadek T. 2001. Evidence that gal11 protein is a target of the gal4 activation domain in the mediator. *Biochemistry,* 40**,** 9421-9427.

Jha P., Chaloupka F. J., Moore J., Gajalakshmi V., Gupta P. C., Peck R., Jamison D., Breman J., Measham A. and Alleyne G. 2006. Disease control priorities in developing countries. *Disease Control Priorities in Developing Countries.*

Johnson S., Strauss V., Gilmore C., Jaekel J., Marlow N. and Wolke D. 2016. Learning disabilities among extremely preterm children without neurosensory impairment: Comorbidity, neuropsychological profiles and scholastic outcomes. *Early Human Development,* 103**,** 69-75.

Jorde L. B., Carey J. C. and Bamshad M. J. 2015. *Medical genetics*, Elsevier Health Sciences.

Ju J., Kim D. H., Bi L., Meng Q., Bai X., Li Z., Li X., Marma M. S., Shi S., Wu J., Edwards J. R., Romu A. and Turro N. J. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences,* 103**,** 19635-19640.

Justice M. J. and Dhillon P. 2016. Using the mouse to model human disease: Increasing validity and reproducibility. The Company of Biologists Ltd.

Kahn R. S., Fleischhacker W. W., Boter H., Davidson M., Vergouwe Y., Keet I. P., Gheorghe M. D., Rybakowski J. K., Galderisi S. and Libiger J. 2008. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: An open randomised clinical trial. *The Lancet,* 371**,** 1085-1097.

Kalay E., Li Y., Uzumcu A., Uyguner O., Collin R. W., Caylan R., Ulubil-Emiroglu M., Kersten F. F., Hafiz G., Van Wijk E., Kayserili H., Rohmann E., Wagenstaller J., Hoefsloot L. H., Strom T. M., Nurnberg G., Baserer N., Den Hollander A. I., Cremers F. P., Cremers C. W., Becker C., Brunner H. G., Nurnberg P., Karaguzel A., Basaran S., Kubisch C., Kremer H. and Wollnik B. 2006. Mutations in the lipoma hmgic fusion partner-like 5 (lhfpl5) gene cause autosomal recessive nonsyndromic hearing loss. *Humun Mutation,* 27**,** 633-9.

Kalia S. S., Adelman K., Bale S. J., Chung W. K., Eng C., Evans J. P., Herman G. E., Hufnagel S. B., Klein T. E., Korf B. R., Mckelvey K. D., Ormond K. E., Richards C. S., Vlangos C. N., Watson M., Martin C. L. and Miller D. T. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (acmg sf v2.0): A policy statement of the american college of medical genetics and genomics. *Genetics in Medicine,* 19**,** 249-255.

Kalidas S., Santosh V., Shareef M. M., Shankar S. K., Christopher R. and Shetty K. T. 2000. Expression of p67 (munc-18) in adult human brain and neuroectodermal tumors of human central nervous system. *Acta Neuropathologica,* 99**,** 191-198.

Kambara H., Nishikawa T., Katayama Y. and Yamaguchi T. 1988. Optimization of parameters in a DNA sequenator using fluorescence detection. *Nature Biotechnology,* 6**,** 816-821.

Kamnasaran D., Muir W., Ferguson-Smith M. and Cox D. 2003. Disruption of the neuronal pas3 gene in a family affected with schizophrenia. *Journal of Medical Genetics,* 40**,** 325-332.

Karam S. M., Riegel M., Segal S. L., Félix T. M., Barros A. J. D., Santos I. S., Matijasevich A., Giugliani R. and Black M. 2015. Genetic causes of intellectual disability in a birth cohort: A population-based study. *American Journal of Medical Genetics Part A,* 167**,** 1204-1214.

Karlsgodt K. H., Sun D. and Cannon T. D. 2010. Structural and functional brain abnormalities in schizophrenia. *Current Directions in Psychological Science,* 19**,** 226-231.

Kasianowicz J. J., Brandin E., Branton D. and Deamer D. W. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences,* 93**,** 13770-13773.

Katz G. and Lazcano-Ponce E. 2008. Intellectual disability: Definition, etiological factors, classification, diagnosis, treatment and prognosis. *Salud Pública de México,* 50**,** s132-s141.

Katzen F. 2007. Gateway (r) recombinational cloning: A biological operating system. *Expert Opinion on Drug Discovery,* 2**,** 571-589.

Kaufman L., Ayub M. and Vincent J. B. 2010. The genetic basis of non-syndromic intellectual disability: A review. *Journal of Neurodevelopmental Disorders,* 2**,** 182-209.

Kaufmann C. A., Suarez B., Malaspina D., Pepple J., Svrakic D., Markel P. D., Meyer J., Zambuto C. T., Schmitt K., Matise T. C., Friedman J. M. H., Hampe C., Lee H.,

Shore D., Wynne D., Faraone S. V., Tsuang M. T. and Cloninger C. R. 1998. Nimh genetics initiative millennium schizophrenia consortium: Linkage analysis of african-american pedigrees. *American Journal of Medical Genetics,* 81**,** 282-289.

Kelleher D. J. and Gilmore R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology,* 16**,** 47R-62R.

Kendler K. S., Mcguire M., Gruenberg A. M. and Walsh D. 1994. Outcome and family study of the subtypes of schizophrenia in the west of ireland. *American Journal of Psychiatry,* 151**,** 849-856.

Kennedy S. R., Loeb L. A. and Herr A. J. 2012. Somatic mutations in aging, cancer and neurodegeneration. *Mechanisms of Ageing and Development,* 133**,** 118-126.

Khan M. A. and Akella S. 2009. Cannabis-induced bipolar disorder with psychotic features: A case report. *Psychiatry (Edgmont),* 6**,** 44-48.

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

Khan M. A., Rafiq M. A., Noor A., Ali N., Ali G., Vincent J. B. and Ansar M. 2011. A novel deletion mutation in the tusc3 gene in a consanguineous pakistani family with autosomal recessive nonsyndromic intellectual disability. *BMC Medical Genetics,* 12**,** 56.

Khimich D., Nouvian R., Pujol R., Tom Dieck S., Egner A., Gundelfinger E. D. and Moser T. 2005. Hair cell synaptic ribbons are essential for synchronous auditory signalling. *Nature,* 434**,** 889-894.

Kidd K. K., Pakstis A. J., Speed W. C., Grigorenko E. L., Kajuna S. L. B., Karoma N. J., Kungulilo S., Kim J.-J., Lu R.-B., Odunsi A., Okonofua F., Parnas J., Schulz L. O., Zhukova O. V. and Kidd J. R. 2006. Developing a snp panel for forensic identification of individuals. *Forensic Science International,* 164**,** 20-32.

Kim J., Suh H., Kim S., Kim K., Ahn C. and Yim J. 2006. Identification and characteristics of the structural gene for the drosophila eye colour mutant sepia, encoding pda synthase, a member of the omega class glutathione s-transferases. *Biochemical Journal,* 398**,** 451-460.

Kircher M., Witten D. M., Jain P., O'roak B. J., Cooper G. M. and Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics,* 46**,** 310-315.

Kirov G., Grozeva D., Norton N., Ivanov D., Mantripragada K. K., Holmans P., Consortium I. S., Consortium T. W. T. C. C., Craddock N., Owen M. J. and O'donovan M. C. 2009. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human Molecular Genetics,* 18**,** 1497-1503.

Kissin B. 2012. *Conscious and unconscious programs in the brain*, Springer Science & Business Media.

Kito K. and Suzuki K. 2016. Research on the effect of the foot bath and foot massage on residual schizophrenia patients. *Archives of Psychiatric Nursing,* 30**,** 375-381.

Kivistö J., Soininen H. and Pihlajamaki M. 2014. Functional mri in alzheimer's disease. *Advanced brain neuroimaging topics in health and disease-methods and applications.* InTech.

Klein C. and Westenberger A. 2012. Genetics of parkinson's disease. *Cold Spring Harbor Perspectives in Medicine,* 2**,** a008888.

Klein H.-G., Bauer P. and Hambuch T. 2014. Whole genome sequencing (wgs), whole exome sequencing (wes) and clinical exome sequencing (ces) in patient care. *LaboratoriumsMedizin*, 38**,** 221-230.

Knight H. M., Maclean A., Irfan M., Naeem F., Cass S., Pickard B. S., Muir W. J., Blackwood D. H. R. and Ayub M. 2008. Homozygosity mapping in a family presenting with schizophrenia, epilepsy and hearing impairment. *European Journal of Human Genetics,* 16**,** 750-758.

Koboldt D. C., Steinberg K. M., Larson D. E., Wilson R. K. and Mardis E. R. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell,* 155**,** 27-38.

Koegl M. and Uetz P. 2007. Improving yeast two-hybrid screening systems. *Briefings in Functional Genomics,* 6**,** 302-312.

Koizumi J., Ofuku K., Sakuma K., Shiraishi H., Iio M. and Nawano S. 1988. Cns changes in usher's syndrome with mental disorder: Ct, mri and pet findings. *Journal of Neurology, Neurosurgery, and Psychiatry,* 51**,** 987-990.

Kok L., Van Der Waa A., Klip H. and Staal W. 2016. The effectiveness of psychosocial interventions for children with a psychiatric disorder and mild intellectual disability to borderline intellectual functioning: A systematic literature review and meta-analysis. *Clinical Child Psychology and Psychiatry,* 21**,** 156-171.

Korenberg J., Aaltonen J., Brahe C., Cabin D., Creau N., Delabar J., Doering J., Gardiner K., Hubert R. and Ives J. 1996. Report and abstracts of the sixth international workshop on human chromosome 21 mapping 1996. Cold Spring Harbor, New York, USA.

Kremen W. S., Panizzon M. S. and Cannon T. D. 2016. Genetics and neuropsychology: A merger whose time has come. *Neuropsychology,* 30**,** 1-5.

Krstic D., Boll W. and Noll M. 2013. Influence of the white locus on the courtship behavior of drosophila males. *PloS One,* 8**,** e77904.

Krumm N., Sudmant P. H., Ko A., O'roak B. J., Malig M., Coe B. P., Quinlan A. R., Nickerson D. A., Eichler E. E. and Project N. E. S. 2012. Copy number variation detection and genotyping from exome sequence data. *Genome Research,* 22**,** 1525-1532.

Krystal J. H. and State M. W. 2014. Psychiatric disorders: Diagnosis to therapy. *Cell,* 157**,** 201-214.

Ku C. S., Loy E. Y., Salim A., Pawitan Y. and Chia K. S. 2010. The discovery of human genetic variations and their use as disease markers: Past, present and future. *Journal of Human Genetics,* 55**,** 403-15.

Ku C. S. and Roukos D. H. 2013. From next-generation sequencing to nanopore sequencing technology: Paving the way to personalized genomic medicine. *Expert Review of Medical Devices,* 10**,** 1-6.

Kuechler A., Zink A. M., Wieland T., Lüdecke H.-J., Cremer K., Salviati L., Magini P., Najafi K., Zweier C., Czeschik J. C., Aretz S., Endele S., Tamburrino F., Pinato C., Clementi M., Gundlach J., Maylahn C., Mazzanti L., Wohlleber E., Schwarzmayr T., Kariminejad R., Schlessinger A., Wieczorek D., Strom T. M., Novarino G. and Engels H. 2015. Loss-of-function variants of setd5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *European Journal of Human Genetics,* 23**,** 753-760.

Kukurba K. R. and Montgomery S. B. 2015. Rna sequencing and analysis. *Cold Spring Harbor protocols,* 2015**,** 951-969.

Kuper H., Nyapera V., Evans J., Munyendo D., Zuurmond M., Frison S., Mwenda V., Otieno D. and Kisia J. 2015. Malnutrition and childhood disability in turkana, kenya: Results from a case-control study. *Plos One,* 10.

Kuss A. W., Garshasbi M., Kahrizi K., Tzschach A., Behjati F., Darvish H., Abbasi-Moheb L., Puettmann L., Zecha A., Weißmann R., Hu H., Mohseni M., Abedini S. S., Rajab A., Hertzberg C., Wieczorek D., Ullmann R., Ghasemi-Firouzabadi S., Banihashemi S., Arzhangi S., Hadavi V., Bahrami-Monajemi G., Kasiri M., Falah M., Nikuei P., Dehghan A., Sobhani M., Jamali P., Ropers H. H. and Najmabadi H. 2011. Autosomal recessive mental retardation: Homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Human Genetics,* 129**,** 141-148.

Kwon J. M. and Goate A. M. 2000. The candidate gene approach. *Alcohol Research and Health,* 24**,** 164-168.

Lacaille F., Hiroi M., Twele R., Inoshita T., Umemoto D., Maniere G., Marion-Poll F., Ozaki M., Francke W., Cobb M., Everaerts C., Tanimura T. and Ferveur J. F. 2007. An inhibitory sex pheromone tastes bitter for drosophila males. *PLoS One,* 2**,** e661.

Lai D.-C., Tseng Y.-C., Hou Y.-M. and Guo H.-R. 2012. Gender and geographic differences in the prevalence of intellectual disability in children: Analysis of data from the national disability registry of taiwan. *Research in Developmental Disabilities,* 33**,** 2301-2307.

Lakhan R. 2015. Profile of social, environmental and biological correlates in intellectual disability in a resource-poor setting in india. *Indian Journal of Psychological Medicine,* 37**,** 311-316.

Lakhan R. and Kishore M. T. 2016. Down syndrome in tribal population in india: A field observation. *Journal of Neurosciences in Rural Practice,* 7**,** 40-43.

Lander E. S. and Botstein D. 1987. Homozygosity mapping - a way to map human recessive traits with the DNA of inbred children. *Science,* 236**,** 1567-1570.

Langmead B. and Salzberg S. L. 2012. Fast gapped-read alignment with bowtie 2. *Nature methods,* 9**,** 357-359.

Larcher V. and Brierley J. 2014. Fetal alcohol syndrome (fas) and fetal alcohol spectrum disorder (fasd)—diagnosis and moral policing; an ethical dilemma for paediatricians. *Archives of Disease in Childhood,* 99**,** 969-970.

Larsson M. C., Domingos A. I., Jones W. D., Chiappe M. E., Amrein H. and Vosshall L. B. 2004. Or83b encodes a broadly expressed odorant receptor essential for drosophila olfaction. *Neuron,* 43**,** 703-714.

Latha N. A., Murthy B. R. and Sunitha U. 2012. Smart card based integrated electronic health record system for clinical practice. *Editorial Preface,* 3.

Lattao R., Bonaccorsi S., Guan X., Wasserman S. A. and Gatti M. 2011. Tubby-tagged balancers for the drosophila x and second chromosomes. *Fly,* 5**,** 369-370.

Lecroy C. W. and Holschuh J. 2012. *First person accounts of mental illness and recovery*, John Wiley & Sons.

Lee B. K., Magnusson C., Gardner R. M., Blomström Å., Newschaffer C. J., Burstyn I., Karlsson H. and Dalman C. 2015. Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain, Behavior, and Immunity,* 44**,** 100-105.

Lee C. R. and Tepper J. M. 2007. Morphological and physiological properties of parvalbumin-and calretinin-containing γ-aminobutyric acidergic neurons in the substantia nigra. *Journal of Comparative Neurology,* 500**,** 958-972.

Lee S., Decandia T. R., Ripke S., Yang J., Sullivan P. F., Goddard M. E., Keller M. C., Visscher P. M. and Wray N. R. 2012. Estimating the proportion of variation in susceptibility to schizophrenia captured by common snps. *Nature Genetics,* 44**,** 247-250.

Lee S., Kim C.-H., Kim D.-G., Kim H.-G., Lee P.-S. and Myung H. 2013. Remote guidance of untrained turtles by controlling voluntary instinct behavior. *PloS One,* 8**,** e61798.

Lehner T. and Miller B. L. 2016. *Genomics, circuits, and pathways in clinical neuropsychiatry*, Academic Press.

Leigh H. 2013. *The patient: Biological, psychological, and social dimensions of medical practice*, Springer Science & Business Media.

Lek M., Karczewski K. J., Minikel E. V., Samocha K. E., Banks E., Fennell T., O'donnell-Luria A. H., Ware J. S., Hill A. J., Cummings B. B., Tukiainen T., Birnbaum D. P., Kosmicki J. A., Duncan L. E., Estrada K., Zhao F., Zou J., Pierce-Hoffman E., Berghout J., Cooper D. N., Deflaux N., Depristo M., Do R., Flannick J., Fromer M., Gauthier L., Goldstein J., Gupta N., Howrigan D., Kiezun A., Kurki M. I., Moonshine A. L., Natarajan P., Orozco L., Peloso G. M., Poplin R., Rivas M. A., Ruano-Rubio V., Rose S. A., Ruderfer D. M., Shakir K., Stenson P. D., Stevens C., Thomas B. P., Tiao G., Tusie-Luna M. T., Weisburd B., Won H.-H., Yu D., Altshuler D. M., Ardissino D., Boehnke M., Danesh J., Donnelly S., Elosua R., Florez J. C., Gabriel S. B., Getz G., Glatt S. J., Hultman C. M., Kathiresan S., Laakso M., Mccarroll S., Mccarthy M. I., Mcgovern D., Mcpherson R., Neale B. M., Palotie A., Purcell S. M., Saleheen D., Scharf J. M., Sklar P., Sullivan P. F., Tuomilehto J., Tsuang M. T., Watkins H. C., Wilson J. G., Daly M. J., Macarthur D. G. and Exome Aggregation C. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536**,** 285-291.

Lett T. A., Chakavarty M. M., Felsky D., Brandl E. J., Tiwari A. K., Gonçalves V. F., Rajji T. K., Daskalakis Z. J., Meltzer H. Y. and Lieberman J. A. 2013. The genome-wide supported microrna-137 variant predicts phenotypic heterogeneity within schizophrenia. *Molecular psychiatry,* 18**,** 443-450.

Leucht S., Corves C., Arbter D., Engel R. R., Li C. and Davis J. M. 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: A meta-analysis. *The Lancet,* 373**,** 31-41.

Levinson D. F., Duan J., Oh S., Wang K., Sanders A. R., Shi J., Zhang N., Mowry B. J., Olincy A., Amin F., Cloninger C. R., Silverman J. M., Buccola N. G., Byerley W. F., Black D. W., Kendler K. S., Freedman R., Dudbridge F., Pe'er I., Hakonarson H., Bergen S. E., Fanous A. H., Holmans P. A. and Gejman P. V. 2011. Copy number variants in schizophrenia: Confirmation of five previous findings and new evidence for 3q29 microdeletions and vipr2 duplications. *The American journal of psychiatry,* 168**,** 302-316.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and Genome Project Data Processing S. 2009. The sequence alignment/map format and samtools. *Bioinformatics,* 25, 2078-2079.

Li Z., Chen J., Xu Y., Yi Q., Ji W., Wang P., Shen J., Song Z., Wang M., Yang P., Wang Q., Feng G., Liu B., Sun W., Xu Q., Li B., He L., He G., Li W., Wen Z., Liu K., Huang F., Zhou J., Ji J., Li X. and Shi Y. 2016. Genome-wide analysis of the role of copy number variation in schizophrenia risk in chinese. *Biological Psychiatry,* 80**,** 331-337.

Liao H.-M., Niu D.-M., Chen Y.-J., Fang J.-S., Chen S.-J. and Chen C.-H. 2011. Identification of a microdeletion at xp22. 13 in a taiwanese family presenting with nance-horan syndrome. *Journal of Human Genetics,* 56**,** 8-11.

Liddelow S. A. 2015. Development of the choroid plexus and blood-csf barrier. *Frontiers in Neuroscience,* 9**,** 32.

Lidgren M., Jönsson B., Rehnberg C., Willking N. and Bergh J. 2008. Cost-effectiveness of her2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. *Annals of oncology,* 19**,** 487-495.

Lieberman J. A., Stroup T. S., Mcevoy J. P., Swartz M. S., Rosenheck R. A., Perkins D. O., Keefe R. S. E., Davis S. M., Davis C. E., Lebowitz B. D., Severe J. and Hsiao J. K. 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine,* 353**,** 1209-1223.

Lijavetzky D., Cabezas J. A., Ibáñez A., Rodríguez V. and Martínez-Zapater J. M. 2007. High throughput snp discovery and genotyping in grapevine (vitis vinifera l.) by combining a re-sequencing approach and snplex technology. *BMC Genomics,* 8**,** 424-424.

Lin L., Chamberlain L., Zhu L. J. and Green M. R. 2012a. Analysis of gal4-directed transcription activation using tra1 mutants selectively defective for interaction with gal4. *Proceedings of the National Academy of Sciences,* 109**,** 1997-2002.

Lin X., Tang W., Ahmad S., Lu J., Colby C. C., Zhu J. and Yu Q. 2012b. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hearing Research,* 288**,** 10.1016/j.heares.2012.01.004.

Lindström S., Thompson D. J., Paterson A. D., Li J., Gierach G. L., Scott C., Stone J., Douglas J. A., Dos-Santos-Silva I., Fernandez-Navarro P., Verghase J., Smith P., Brown J., Luben R., Wareham N. J., Loos R. J. F., Heit J. A., Shane Pankratz V., Norman A., Goode E. L., Cunningham J. M., Deandrade M., Vierkant R. A., Czene K., Fasching P. A., Baglietto L., Southey M. C., Giles G. G., Shah K. P., Chan H.-P., Helvie M. A., Beck A. H., Knoblauch N. W., Hazra A., Hunter D. J., Kraft P., Pollan M., Figueroa J. D., Couch F. J., Hopper J. L., Hall P., Easton D. F., Boyd N. F., Vachon C. M. and Tamimi R. M. 2014. Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. *Nature Communications,* 5.

Lindvall O. and Kokaia Z. 2010. Stem cells in human neurodegenerative disorders—time for clinical translation? *The Journal of Clinical Investigation,* 120**,** 29-40.

Lindvall O., Kokaia Z. and Martinez-Serrano A. 2004. Stem cell therapy for human neurodegenerative disorders–how to make it work.

Linehan M. 1993. *Cognitive-behavioral treatment of borderline personality disorder*, Guilford press.

Ling H. H., Beaulé C., Chiang C.-K., Tian R., Figeys D. and Cheng H.-Y. M. 2014. Time-of-day-and light-dependent expression of ubiquitin protein ligase e3 component n-

recognin 4 (ubr4) in the suprachiasmatic nucleus circadian clock. *PloS One,* 9**,** e103103.

Linscott R. J., Allardyce J. and Van Os J. 2010. Seeking verisimilitude in a class: A systematic review of evidence that the criterial clinical symptoms of schizophrenia are taxonic. *Schizophrenia Bulletin,* 36**,** 811-829.

Lipina T. V. and Roder J. C. 2015. *Drug discovery for schizophrenia*, Royal Society of Chemistry.

Litt J., Taylor H. G., Klein N. and Hack M. 2005. Learning disabilities in children with very low birthweight prevalence, neuropsychological correlates, and educational interventions. *Journal of Learning Disabilities,* 38**,** 130-141.

Little J., Khoury M. J., Bradley L., Clyne M., Gwinn M., Lin B., Lindegren M.-L. and Yoon P. 2003. The human genome project is complete. How do we develop a handle for the pump? : Oxford Univ Press.

Liu J., Juo S. H., Dewan A., Grunn A., Tong X., Brito M., Park N., Loth J. E., Kanyas K., Lerer B., Endicott J., Penchaszadeh G., Knowles J. A., Ott J., Gilliam T. C. and Baron M. 2003. Evidence for a putative bipolar disorder locus on 2p13-16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21-24, 13q32, 14q21 and 17q11-12. *Molecular Psychiatry,* 8**,** 333-342.

Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L. and Law M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International,* 2012.

Liu P.-Y., Zhang Y.-Y., Lu Y., Long J.-R., Shen H., Zhao L.-J., Xu F.-H., Xiao P., Xiong D.-H., Liu Y.-J., Recker R. R. and Deng H.-W. 2005. A survey of haplotype variants at several disease candidate genes: The importance of rare variants for complex diseases. *Journal of Medical Genetics,* 42**,** 221-227.

Liu X., Cheng R., Verbitsky M., Kisselev S., Browne A., Mejia-Sanatana H., Louis E. D., Cote L. J., Andrews H., Waters C., Ford B., Frucht S., Fahn S., Marder K., Clark L. N. and Lee J. H. 2011. Genome-wide association study identifies candidate genes for parkinson's disease in an ashkenazi jewish population. *BMC Medical Genetics,* 12**,** 1-16.

Liu Y., Liang M., Zhou Y., He Y., Hao Y., Song M., Yu C., Liu H., Liu Z. and Jiang T. 2008. Disrupted small-world networks in schizophrenia. *Brain,* 131**,** 945-961.

Lodato M. A., Woodworth M. B., Lee S., Evrony G. D., Mehta B. K., Karger A., Lee S., Chittenden T. W., D'gama A. M., Cai X., Luquette L. J., Lee E., Park P. J. and Walsh C. A. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science (New York, N.Y.),* 350**,** 94-98.

Loddo S., Parisi V., Doccini V., Filippi T., Bernardini L., Brovedani P., Ricci F., Novelli A. and Battaglia A. 2013. Homozygous deletion in tusc3 causing syndromic intellectual disability: A new patient. *American Journal of Medical Genetics Part A,* 161A**,** 2084-7.

Longo-Guess C. M., Gagnon L. H., Cook S. A., Wu J., Zheng Q. Y. and Johnson K. R. 2005. A missense mutation in the previously undescribed gene tmhs underlies deafness in hurry-scurry (hscy) mice. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 7894-9.

Longo-Guess C. M., Gagnon L. H., Fritzsch B. and Johnson K. R. 2007. Targeted knockout and lacz reporter expression of the mouse tmhs deafness gene and characterization of the hscy-2j mutation. *Mammalian Genome,* 18**,** 646-56.

Lopez A. D. and Murray C. J. 1996. *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020; summary*, Harvard School of Public Health.

Love M. I., Mysickova A., Sun R., Kalscheuer V., Vingron M. and Haas S. A. 2011. Modeling read counts for cnv detection in exome sequencing data. *Statistical Applications in Genetics and Molecular Biology*, 10**,** 1.

Lubs Herbert a., Stevenson Roger e. and Schwartz Charles e. 2012. Fragile x and x-linked intellectual disability: Four decades of discovery. *The American Journal of Human Genetics*, 90**,** 579-590.

Lun M. P., Monuki E. S. and Lehtinen M. K. 2015. Development and functions of the choroid plexus-cerebrospinal fluid system. *Nature Reviews Neuroscience*, 16**,** 445-457.

Luoma J. B. and Villatte J. L. 2012. Mindfulness in the treatment of suicidal individuals. *Cognitive and Behavioral Practice*, 19**,** 265-276.

Lupski J. R., Belmont J. W., Boerwinkle E. and Gibbs R. A. 2011. Clan genomics and the complex architecture of human disease. *Cell*, 147**,** 32-43.

Lupski J. R., Reid J. G., Gonzaga-Jauregui C., Rio Deiros D., Chen D. C., Nazareth L., Bainbridge M., Dinh H., Jing C. and Wheeler D. A. 2010. Whole-genome sequencing in a patient with charcot–marie–tooth neuropathy. *New England Journal of Medicine*, 362**,** 1181-1191.

Luyben P. D. 2009. Applied behavior analysis: Understanding and changing behavior in the community—a representative review. *Journal of Prevention & Intervention in the Community*, 37**,** 230-253.

Lyly A. 2008. Molecular interactions underlying neuronal ceroid lipofuscinoses cln1 and cln5.

Macar F. and Vidal F. 2009. Timing processes: An outline of behavioural and neural indices not systematically considered in timing models. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 63**,** 227.

Macarthur D., Manolio T., Dimmock D., Rehm H., Shendure J., Abecasis G., Adams D., Altman R., Antonarakis S. and Ashley E. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508**,** 469-476.

Macintyre G., Alford T., Xiong L., Rouleau G. A., Tibbo P. G. and Cox D. W. 2010. Association of npas3 exonic variation with schizophrenia. *Schizophrenia Research*, 120**,** 143-149.

Mackenzie S. M., Brooker M. R., Gill T. R., Cox G. B., Howells A. J. and Ewart G. D. 1999. Mutations in the white gene of drosophila melanogaster affecting abc transporters that determine eye colouration. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1419**,** 173-185.

Macleod J., Oakes R., Copello A., Crome I., Egger M., Hickman M., Oppenkowski T., Stokes-Lampard H. and Smith G. D. 2004. Psychological and social sequelae of cannabis and other illicit drug use by young people: A systematic review of longitudinal, general population studies. *The Lancet*, 363**,** 1579-1588.

Madison J. M., Zhou F., Nigam A., Hussain A., Barker D. D., Nehme R., Van Der Ven K., Hsu J., Wolf P., Fleishman M., O'dushlaine C., Rose S., Chambert K., Lau F. H., Ahfeldt T., Rueckert E. H., Sheridan S. D., Fass D. M., Nemesh J., Mullen T. E., Daheron L., Mccarroll S., Sklar P., Perlis R. H. and Haggarty S. J. 2015.

Characterization of bipolar disorder patient-specific induced pluripotent stem cells from a family reveals neurodevelopmental and mrna expression abnormalities. *Molecular Psychiatry*, 20**,** 703-717.

Magi A., Tattini L., Cifola I., D'aurizio R., Benelli M., Mangano E., Battaglia C., Bonora E., Kurg A. and Seri M. 2013. Excavator: Detecting copy number variants from whole-exome sequencing data. *Genome Biology*, 14**,** R120.

Makrythanasis P., Guipponi M., Santoni F. A., Zaki M., Issa M. Y., Ansar M., Hamamy H. and Antonarakis S. E. 2016. Exome sequencing discloses kalrn homozygous variant as likely cause of intellectual disability and short stature in a consanguineous pedigree. *Human Genomics,* 10**,** 26.

Malmberg L., Fenton M. and Rathbone J. 2001. Individual psychodynamic psychotherapy and psychoanalysis for schizophrenia and severe mental illness. *The Cochrane Database of Systematic Reviews***,** CD001360-CD001360.

Mangialasche F., Solomon A., Kåreholt I., Hooshmand B., Cecchetti R., Fratiglioni L., Soininen H., Laatikainen T., Mecocci P. and Kivipelto M. 2013. Serum levels of vitamin e forms and risk of cognitive impairment in a finnish cohort of older adults. *Experimental Gerontology,* 48**,** 1428-1435.

Mansour H., Fathi W., Klei L., Wood J., Chowdari K., Watson A., Eissa A., Elassy M., Ali I., Salah H., Yassin A., Tobar S., El-Boraie H., Gaafar H., Ibrahim N. E., Kandil K., El-Bahaei W., El-Boraie O., Alatrouny M., El-Chennawi F., Devlin B. and Nimgaonkar V. L. 2010. Consanguinity and increased risk for schizophrenia in egypt. *Schizophrenia Research,* 120**,** 108-112.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y.-J. and Chen Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

Maris A. F., Barbato I. T., Trott A. and Montano M. a. E. 2013. Familial mental retardation: A review and practical classification. *Ciência & Saúde Coletiva,* 18**,** 1717-1729.

Marková S., Šafka Brožková D., Mészárosová A., Neupauerová J., Groh D., Křečková G., Laššuthová P. and Seeman P. 2016. Mutations in eight small dfnb genes are not a frequent cause of non-syndromic hereditary hearing loss in czech patients. *International Journal of Pediatric Otorhinolaryngology,* 86**,** 27-33.

Martínez F., Caro-Llopis A., Roselló M., Oltra S., Mayo S., Monfort S. and Orellana C. 2016. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *Journal of Medical Genetics*.

Marx V. 2013. Next-generation sequencing: The genome jigsaw. *Nature,* 501**,** 263-268.

Masoudi A., Uchida K., Yokouchi K., Ohwada K., Abbasi A., Tsuji T., Watanabe T., Hirano T., Sugimoto Y. and Kunieda T. 2008. Linkage mapping of the locus responsible for forelimb-girdle muscular anomaly of japanese black cattle on bovine chromosome 26. *Animal Genetics,* 39**,** 46-50.

Mattei J., Parnell L. D., Lai C.-Q., Garcia-Bailo B., Adiconis X., Shen J., Arnett D., Demissie S., Tucker K. L. and Ordovas J. M. 2009. Disparities in allele frequencies and population differentiation for 101 disease-associated single nucleotide polymorphisms between puerto ricans and non-hispanic whites. *BMC Genetics,* 10**,** 45.

Maulik P. K., Mascarenhas M. N., Mathers C. D., Dua T. and Saxena S. 2011. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities,* 32**,** 419-436.

Maurano M. T., Humbert R., Rynes E., Thurman R. E., Haugen E., Wang H., Reynolds A. P., Sandstrom R., Qu H. and Brody J. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science,* 337**,** 1190-1195.

Maxam A. M. and Gilbert W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences,* 74**,** 560-564.

May P. A., Blankenship J., Marais A. S., Gossage J. P., Kalberg W. O., Joubert B., Cloete M., Barnard R., De Vries M., Hasken J., Robinson L. K., Adnams C. M., Buckley D., Manning M., Parry C. D., Hoyme H. E., Tabachnick B. and Seedat S. 2013. Maternal alcohol consumption producing fetal alcohol spectrum disorders (fasd): Quantity, frequency, and timing of drinking. *Drug Alcohol Depend,* 133**,** 502-12.

Mayilyan K. R., Weinberger D. R. and Sim R. B. 2008. The complement system in schizophrenia. *Drug News & Perspectives,* 21**,** 200-210.

Mccarthy I., Abecasis G. R., Cardon L. R., Goldstein D. B., Little J., Ioannidis J. P. A. and Hirschhorn J. N. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics,* 9**,** 356-369.

Mccarthy S., Makarov V., Kirov G., Addington A., Mcclellan J., Yoon S., Perkins D., Dickel D. E., Kusenda M., Krastoshevsky O., Krause V., Kumar R. A., Grozeva D., Malhotra D., Walsh T., Zackai E. H., Kaplan P., Ganesh J., Krantz I. D., Spinner N. B., Roccanova P., Bhandari A., Pavon K., Lakshmi B., Leotta A., Kendall J., Lee Y.-H., Vacic V., Gary S., Iakoucheva L., Crow T. J., Christian S. L., Lieberman J., Stroup S., Lehtimäki T., Puura K., Haldeman-Englert C., Pearl J., Goodell M., Willour V. L., Derosse P., Steele J., Kassem L., Wolff J., Chitkara N., Mcmahon F. J., Malhotra A. K., Potash J. B., Schulze T. G., Nöthen M. M., Cichon S., Rietschel M., Leibenluft E., Kustanovich V., Lajonchere C. M., Sutcliffe J. S., Skuse D., Gill M., Gallagher L., Mendell N. R., Wellcome Trust Case Control C., Craddock N., Owen M. J., O'donovan M. C., Shaikh T. H., Susser E., Delisi L. E., Sullivan P. F., Deutsch C. K., Rapoport J., Levy D. L., King M.-C. and Sebat J. 2009. Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genetics,* 41**,** 1223-1227.

Mcclellan J. M., Susser E. and King M.-C. 2007. Schizophrenia: A common disease caused by multiple rare alleles. *The British Journal of Psychiatry,* 190**,** 194-199.

Mccrone P. R., Dhanasiri S., Patel A., Knapp M. and Lawton-Smith S. 2008. *Paying the price: The cost of mental health care in england to 2026,* King's Fund.

Mcgillivray J. A. and Mccabe M. P. 2004. Pharmacological management of challenging behavior of individuals with intellectual disability. *Research in Developmental Disabilities,* 25**,** 523-537.

Mcglashan T. H. and Fenton W. S. 1991. Classical subtypes for schizophrenia: Literature review for dsm-iv. *Schizophrenia Bulletin,* 17**,** 609-632.

Mcgrath J., Saha S., Welham J., El Saadi O., Maccauley C. and Chant D. 2004. A systematic review of the incidence of schizophrenia: The distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC medicine,* 2**,** 13.

Mcguire P., Howes O. D., Stone J. and Fusar-Poli P. 2008. Functional neuroimaging in schizophrenia: Diagnosis and drug discovery. *Trends in Pharmacological Sciences,* 29**,** 91-98.

Meacham M. C., Starks H., Burke W. and Edwards K. 2010. Researcher perspectives on disclosure of incidental findings in genetic research. *Journal of Empirical Research on Human Research Ethics,* 5**,** 31-41.

Medeiros K. 2015. Behavioral interventions for individuals with intellectual disabilities exhibiting automatically-reinforced challenging behavior: Stereotypy and self-injury. *Journal of Psychological Abnormalities*, doi:10.4172/2329-9525.1000141.

Melillo R. and Leisman G. 2010. *Neurobehavioral disorders of childhood: An evolutionary perspective*, Springer Science & Business Media.

Meltzer H. Y., Alphs L., Green A. I., Altamura A. C., Anand R., Bertoldi A., Bourgeois M., Chouinard G., Islam M. Z. and Kane J. 2003. Clozapine treatment for suicidality in schizophrenia: International suicide prevention trial (intersept). *Archives of General Psychiatry,* 60**,** 82-91.

Menolascino F. J. and Stark J. 2012. *Handbook of mental illness in the mentally retarded*, Springer Science & Business Media.

Merritt K., Mcguire P. and Egerton A. 2013. Relationship between glutamate dysfunction and symptoms and cognitive function in psychosis. *Frontiers in Psychiatry,* 4**,** 151.

Mertens J., Wang Q.-W., Kim Y., Yu D. X., Pham S., Yang B., Zheng Y., Diffenderfer K. E., Zhang J., Soltani S., Eames T., Schafer S. T., Boyer L., Marchetto M. C., Nurnberger J. I., Calabrese J. R., Ødegaard K. J., Mccarthy M. J., Zandi P. P., Alba M., Nievergelt C. M., The Pharmacogenomics of Bipolar Disorder S., Mi S., Brennand K. J., Kelsoe J. R., Gage F. H. and Yao J. 2015. Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature,* 527**,** 95-99.

Mery F. and Kawecki T. J. 2002. Experimental evolution of learning ability in fruit flies. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 14274-14279.

Metcalfe K., Gershman S., Ghadirian P., Lynch H. T., Snyder C., Tung N., Kim-Sing C., Eisen A., Foulkes W. D. and Rosen B. 2014. Contralateral mastectomy and survival after breast cancer in carriers of brca1 and brca2 mutations: Retrospective analysis. *British Medical Journal,* 348**,** g226.

Metzker M. L. 2010. Sequencing technologies—the next generation. *Nature Reviews genetics,* 11**,** 31-46.

Meyer U., Yee B. K. and Feldon J. 2007. The neurodevelopmental impact of prenatal infections at different times of pregnancy: The earlier the worse? *The Neuroscientist,* 13**,** 241-256.

Meynert A. M., Ansari M., Fitzpatrick D. R. and Taylor M. S. 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics,* 15**,** 247.

Michalon A., Sidorov M., Ballard T. M., Ozmen L., Spooren W., Wettstein J. G., Jaeschke G., Bear M. F. and Lindemann L. 2012. Chronic pharmacological mglu5 inhibition corrects fragile x in adult mice. *Neuron,* 74**,** 49-56.

Miclea D., Peca L., Cuzmici Z. and Pop I. V. 2015. Genetic testing in patients with global developmental delay / intellectual disabilities. A review. *Clujul Medical,* 88**,** 288-292.

Miller M. 2014. Commentary. *Current Problems in Pediatric and Adolescent Health Care,* 44**,** 105-106.

Mishuris R. G., Yoder J., Wilson D. and Mann D. 2016. Integrating data from an online diabetes prevention program into an electronic health record and clinical workflow, a design phase usability study. *BMC Medical Informatics and Decision Making,* 16**,** 88.

Miyamoto S., Duncan G. E., Marx C. E. and Lieberman J. A. 2004. Treatments for schizophrenia: A critical review of pharmacology and mechanisms of action of antipsychotic drugs. *Molecular Psychiatry,* 10**,** 79-104.

Modell B. and Darr A. 2002. Genetic counselling and customary consanguineous marriage. *Nature Review Genetics,* 3**,** 225-229.

Moens L. N., Falk-Sörqvist E., Asplund A. C., Bernatowska E., Smith C. E. and Nilsson M. 2014. Diagnostics of primary immunodeficiency diseases: A sequencing capture approach. *PLoS One,* 9**,** e114901.

Moeschler J. B. and Shevell M. 2006. Clinical genetic evaluation of the child with mental retardation or developmental delays. *Pediatrics,* 117**,** 2304-2316.

Moeschler J. B., Shevell M., Moeschler J. B., Shevell M., Saul R. A., Chen E., Freedenberg D. L., Hamid R., Jones M. C., Stoler J. M. and Tarini B. A. 2014. Comprehensive evaluation of the child with intellectual disability or global developmental delays. *Pediatrics,* 134**,** e903-e918.

Mogensen M. M., Rzadzinska A. and Steel K. P. 2007. The deaf mouse mutant whirler suggests a role for whirlin in actin filament dynamics and stereocilia development. *Cell Motility and the Cytoskeleton,* 64**,** 496-508.

Mohamed T., Shakeri A. and Rao P. P. N. 2016. Amyloid cascade in alzheimer's disease: Recent advances in medicinal chemistry. *European Journal of Medicinal Chemistry,* 113**,** 258-272.

Mohorko E., Owen R. L., Malojcic G., Brozzo M. S., Aebi M. and Glockshuber R. 2014. Structural basis of substrate specificity of human oligosaccharyl transferase subunit n33/tusc3 and its role in regulating protein n-glycosylation. *Structure,* 22**,** 590-601.

Mokhtari R. and Bagga A. 2003. Consanguinity, genetic disorders and malformations in the iranian population. *Acta Biologica Szegediensis,* 47**,** 47-50.

Molinari F., Foulquier F., Tarpey P. S., Morelle W., Boissel S., Teague J., Edkins S., Futreal P. A., Stratton M. R., Turner G., Matthijs G., Gecz J., Munnich A. and Colleaux L. 2008. Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. *The American Journal of Human Genetics,* 82**,** 1150-7.

Monti J. M., Bahammam A. S., Pandi-Perumal S. R., Bromundt V., Spence D. W., Cardinali D. P. and Brown G. M. 2013. Sleep and circadian rhythm dysregulation in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 43**,** 209-216.

Morgan T. H. 1910. Sex limited inheritance in drosophila. *Science,* 32**,** 120-122.

Morrow E. M. 2010. Genomic copy number variation in disorders of cognitive development. *Journal of the American Academy of Child and Adolescent Psychiatry,* 49**,** 1091-1104.

Mueller-Paul J., Wilkinson A., Hall G. and Huber L. 2012. Radial-arm-maze behavior of the red-footed tortoise (geochelone carbonaria). *Journal of Comparative Psychology,* 126**,** 305.

Mueller R. F. and Bishop D. T. 1993. Autozygosity mapping, complex consanguinity, and autosomal recessive disorders. *Journal of Medical Genetics,* 30**,** 798-799.

Mukaka M. M. 2012. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi,* 24**,** 69-71.

Mullegama S. V., Rosenfeld J. A., Orellana C., Van Bon B. W. M., Halbach S., Repnikova E. A., Brick L., Li C., Dupuis L., Rosello M., Aradhya S., Stavropoulos D. J., Manickam K., Mitchell E., Hodge J. C., Talkowski M. E., Gusella J. F., Keller K., Zonana J., Schwartz S., Pyatt R. E., Waggoner D. J., Shaffer L. G., Lin A. E., De Vries B. B. A., Mendoza-Londono R. and Elsea S. H. 2014. Reciprocal deletion and duplication at 2q23.1 indicates a role for mbd5 in autism spectrum disorder. *European Journal of Human Genetics,* 22**,** 57-63.

Mullin S. and Schapira A. 2015. The genetics of parkinson's disease. *British Medical Bulletin,* 114**,** 39-52.

Murakami S., Dan C., Zagaeski B., Maeyama Y., Kunes S. and Tabata T. 2010. Optimizing drosophila olfactory learning with a semi-automated training device. *Journal of Neuroscience Methods,* 188**,** 195-204.

Murphy E. and Benítez-Burraco A. 2016. Bridging the gap between genes and language deficits in schizophrenia: An oscillopathic approach. *Frontiers in Human Neuroscience,* 10**,** 422.

Musante L. and Ropers H. H. 2014. Genetics of recessive cognitive disorders. *Trends in Genetics,* 30**,** 32-39.

Muscarella L. A., Guarnieri V., Sacco R., Curatolo P., Manzi B., Alessandrelli R., Giana G., Militerni R., Bravaccio C., Lenti C., Saccani M., Schneider C., Melmed R., D'agruma L. and Persico A. M. 2010. Candidate gene study of hoxb1 in autism spectrum disorder. *Molecular Autism,* 1**,** 9-9.

Mustapha M., Chouery E., Chardenoux S., Naboulsi M., Paronnaud J., Lemainque A., Mégarbané A., Loiselet J., Weil D. and Lathrop M. 2002. Dfnb31, a recessive form of sensorineural hearing loss, maps to chromosome 9q32-34. *European Journal of human genetics,* 10**,** 210.

Na'amnih W., Romano-Zelekha O., Kabaha A., Rubin L. P., Bilenko N., Jaber L., Honovich M. and Shohat T. 2014. Prevalence of consanguineous marriages and associated factors among israeli bedouins. *Journal of Community Genetics,* 5**,** 395-398.

Nagasaki M., Yasuda J., Katsuoka F., Nariai N., Kojima K., Kawai Y., Yamaguchi-Kabata Y., Yokozawa J., Danjoh I. and Saito S. 2015. Rare variant discovery by deep whole-genome sequencing of 1,070 japanese individuals. *Nature Communications,* 6.

Nakatani Y., Konishi H., Vassilev A., Kurooka H., Ishiguro K., Sawada J.-I., Ikura T., Korsmeyer S. J., Qin J. and Herlitz A. M. 2005. P600, a unique protein required for membrane morphogenesis and cell survival. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 15093-15098.

Nance W., Warburg M., Bixler D. and Helveston E. 1973. Congenital x-linked cataract, dental anomalies and brachymetacarpalia. *Birth Defects Original Article Series,* 10**,** 285-291.

Neef N. A. 2001. The past and future of behavior analysis in developmental disabilities: When good news is bad and bad news is good. *The Behavior Analyst Today,* 2**,** 336.

Németh A. H., Kwasniewska A. C., Lise S., Parolin Schnekenberg R., Becker E. B. E., Bera K. D., Shanks M. E., Gregory L., Buck D., Zameel Cader M., Talbot K., De Silva R., Fletcher N., Hastings R., Jayawant S., Morrison P. J., Worth P., Taylor M., Tolmie J., O'regan M., Consortium U. K. A., Valentine R., Packham E., Evans J., Seller A. and Ragoussis J. 2013. Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain,* 136**,** 3106-3118.

Nemudryi A. A., Valetdinova K. R., Medvedev S. P. and Zakian S. M. 2014. Talen and crispr/cas genome editing systems: Tools of discovery. *Acta Naturae,* 6**,** 19-40.

Nenadic I., Yotter R. A., Sauer H. and Gaser C. 2015. Patterns of cortical thinning in different subgroups of schizophrenia. *The British Journal of Psychiatry,* 206**,** 479-483.

Netto G. J. and Schrijver I. 2015. *Genomic applications in pathology*, Springer.

Ng P. C. and Henikoff S. 2003. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Research,* 31**,** 3812-3814.

Ngui E. M., Khasakhala L., Ndetei D. and Roberts L. W. 2010. Mental disorders, health inequalities and ethics: A global perspective. *International Review of Psychiatry (Abingdon, England),* 22**,** 235-244.

Nguyen T., Shi W. and Ruden D. 2011. Cloudaligner: A fast and full-featured mapreduce based tool for sequence mapping. *BMC Research Notes,* 4**,** 171-171.

Nicolae D. L., Gamazon E., Zhang W., Duan S., Dolan M. E. and Cox N. J. 2010. Trait-associated snps are more likely to be eqtls: Annotation to enhance discovery from gwas. *PLoS Genetics,* 6**,** e1000888.

Niranjan T. S., Skinner C., May M., Turner T., Rose R., Stevenson R., Schwartz C. E. and Wang T. 2015. Affected kindred analysis of human x chromosome exomes to identify novel x-linked intellectual disability genes. *PLoS ONE,* 10**,** e0116454.

Noback C. R., Strominger N. L., Demarest R. J. and Ruggiero D. A. 2005. *The human nervous system: Structure and function*, Springer Science & Business Media.

Norbury C. F., Tomblin J. B. and Bishop D. V. 2008. *Understanding developmental language disorders: From theory to practice*, Psychology press.

Nothnagel M. and Rohde K. 2005. The effect of single-nucleotide polymorphism marker selection on patterns of haplotype blocks and haplotype frequency estimates. *The American Journal of Human Genetics,* 77**,** 988-998.

Nyrén P. and Lundin A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical biochemistry,* 151**,** 504-509.

Nyunt Tin N., Devapriam J., Bala Raju L. and Bhaumik S. 2008. Use of atypical antipsychotics in people with intellectual disability–is it atypical? *The British Journal of Development Disabilities,* 54**,** 101-111.

O'kane C. J. and Gehring W. J. 1987. Detection in situ of genomic regulatory elements in drosophila. *Proceedings of the National Academy of Sciences,* 84**,** 9123-9127.

O'donovan M. C., Craddock N. J. and Owen M. J. 2009. Genetics of psychosis; insights from views across the genome. *Human Genetics,* 126**,** 3-12.

Ofori-Atta A., Read U. and Lund C. 2010. A situation analysis of mental health services and legislation in ghana: Challenges for transformation. *African Journal of Psychiatry,* 13.

Ollila H. M., Soronen P., Silander K., Palo O. M., Kieseppa T., Kaunisto M. A., Lonnqvist J., Peltonen L., Partonen T. and Paunio T. 2009. Findings from bipolar disorder

genome-wide association studies replicate in a finnish bipolar family-cohort. *Molecular Psychiatry,* 14**,** 351-3.

Olney J. W. and Farber N. B. 1995. Glutamate receptor dysfunction and schizophrenia. *Archives of General Psychiatry,* 52**,** 998-1007.

Oltvai Z. 2012. *Conceptual advances in pathology, an issue of clinics in laboratory medicine,* Elsevier Health Sciences.

Onsongo G., Erdmann J., Spears M. D., Chilton J., Beckman K. B., Hauge A., Yohe S., Schomaker M., Bower M., Silverstein K. a. T. and Thyagarajan B. 2014. Implementation of cloud based next generation sequencing data analysis in a clinical laboratory. *BMC Research Notes,* 7**,** 314-314.

Onstad S., Skre I., Edvardsen J., Torgersen S. and Kringlen E. 1991. Mental disorders in first-degree relatives of schizophrenics. *Acta Psychiatrica Scandinavica,* 83**,** 463-467.

Ornitz E. M. 1969. Disorders of perception common to early infantile autism and schizophrenia. *Comprehensive Psychiatry,* 10**,** 259-274.

Owen M., Craddock N. and O'donovan M. 2005. Schizophrenia: Genes at last? *TRENDS in Genetics,* 21**,** 518-525.

Pan L., Yan J., Wu L. and Zhang M. 2009. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 5575-5580.

Papetti F., Morel-Pingault V., Buisse V., Maziere L., Banayan M., Thauby S., Besnard T., Darcourt G. and Pringuey D. 2007. Clozapine-resistant schizophrenia related to an increased metabolism and benefit of fluvoxamine: Four case reports. *L'Encephale,* 33**,** 811-818.

Paris J. 2007. Why psychiatrists are reluctant to diagnose: Borderline personality disorder. *Psychiatry (Edgmont),* 4**,** 35-39.

Parsons K., Nakatani Y. and Nguyen M. D. 2015. P600/ubr4 in the central nervous system. *Cell Mol Life Sci,* 72**,** 1149-60.

Patel K. R., Cherian J., Gohil K. and Atkinson D. 2014. Schizophrenia: Overview and treatment options. *Pharmacy and Therapeutics,* 39**,** 638.

Pauwels K., De Keersmaecker S. C. J., De Schrijver A., Du Jardin P., Roosens N. H. C. and Herman P. 2015. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not? *Trends in Food Science & Technology,* 45**,** 319-326.

Pavelić K., Martinović T. and Pavelić S. K. 2015. Do we understand the personalized medicine paradigm? *EMBO Reports,* 16**,** 133-136.

Penrose L. S. 1935. The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. *Annals of Eugenics,* 6**,** 133-138.

Pepper E. and G Cardno A. 2014. Genetics of schizophrenia and other psychotic disorders. *Current Psychiatry Reviews,* 10**,** 133-142.

Peralta V. and Cuesta M. J. 2003. The nosology of psychotic disorders: A comparison among competing classification systems. *Schizophrenia Bulletin,* 29**,** 413.

Petrovski S., Wang Q., Heinzen E. L., Allen A. S. and Goldstein D. B. 2013. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet,* 9**,** e1003709.

Petrucelli N., Daly M. B. and Feldman G. L. 2013. Brca1 and brca2 hereditary breast and ovarian cancer.

Picardi A., Viroli C., Tarsitani L., Miglio R., De Girolamo G., Dell'acqua G. and Biondi M. 2012. Heterogeneity and symptom structure of schizophrenia. *Psychiatry Research,* 198**,** 386-394.

Pickard B. S., Malloy M. P., Porteous D. J., Blackwood D. H. R. and Muir W. J. 2005. Disruption of a brain transcription factor, npas3, is associated with schizophrenia and learning disability. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics,* 136B**,** 26-32.

Picker J. D. and Walsh C. A. 2013. New innovations: Therapeutic opportunities for intellectual disabilities. *Annals of neurology,* 74**,** 10.1002/ana.24002.

Pillai A. 2008. Decreased expression of sprouty2 in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder: A correlation with bdnf expression. *PLoS One,* 3**,** e1784.

Pina C. and Pignoni F. 2012. Tubby-rfp balancers for developmental analysis: Fm7c 2xtb-rfp, cyo 2xtb-rfp and tm3 2xtb-rfp. *Genesis (New York, N.Y. : 2000),* 50**,** 119-123.

Pirooznia M., Goes F. S. and Zandi P. P. 2015. Whole-genome cnv analysis: Advances in computational approaches. *Frontiers in genetics,* 6.

Pishva E., Kenis G., Van Den Hove D., Lesch K.-P., Boks M. P. M., Van Os J. and Rutten B. P. F. 2014. The epigenome and postnatal environmental influences in psychotic disorders. *Social Psychiatry and Psychiatric Epidemiology,* 49**,** 337-348.

Pitcher M. R., Herrera J. A., Buffington S. A., Kochukov M. Y., Merrit J. K., Fisher A. R., Schanen N. C., Costa-Mattioli M. and Neul J. L. 2015. Rett syndrome like phenotypes in the r255x mecp2 mutant mouse are rescued by mecp2 transgene. *Human molecular genetics***,** ddv030.

Pitman J. L., Dasgupta S., Krashes M. J., Leung B., Perrat P. N. and Waddell S. 2009. There are many ways to train a fly. *Fly,* 3**,** 3.

Plagnol V., Curtis J., Epstein M., Mok K. Y., Stebbings E., Grigoriadou S., Wood N. W., Hambleton S., Burns S. O., Thrasher A. J., Kumararatne D., Doffinger R. and Nejentsev S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics,* 28**,** 2747-2754.

Poduri A., Evrony G. D., Cai X. and Walsh C. A. 2013. Somatic mutation, genomic variation, and neurological disease. *Science (New York, N.Y.),* 341**,** 1237758-1237758.

Pollack H. A. 2014. Coverage of whole genome sequencing in the affordable care act. *Journal of health politics, policy and law,* 39**,** 237-238.

Pollard K. S., Hubisz M. J., Rosenbloom K. R. and Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research,* 20**,** 110-121.

Polyak A., Rosenfeld J. A. and Girirajan S. 2015. An assessment of sex bias in neurodevelopmental disorders. *Genome Medicine,* 7**,** 1-11.

Polymeropoulos M. H., Higgins J. J., Golbe L. I., Johnson W. G., Ide S. E., Iorio G. D., Sanges G., Stenroos E. S., Pho L. T., Schaffer A. A., Lazzarini A. M., Nussbaum R. L. and Duvoisin R. C. 1996. Mapping of a gene for parkinson's disease to chromosome 4q21-q23. *Science,* 274**,** 1197-1199.

Polymeropoulos M. H., Lavedan C., Leroy E., Ide S. E., Dehejia A., Dutra A., Pike B., Root H., Rubenstein J., Boyer R., Stenroos E. S., Chandrasekharappa S., Athanassiadou A., Papapetropoulos T., Johnson W. G., Lazzarini A. M., Duvoisin R. C., Di Iorio G., Golbe L. I. and Nussbaum R. L. 1997. Mutation in the α-synuclein gene identified in families with parkinson's disease. *Science,* 276**,** 2045-2047.

Potkin S. G., Macciardi F., Guffanti G., Fallon J. H., Wang Q., Turner J. A., Lakatos A., Miles M. F., Lander A. and Vawter M. P. 2010. Identifying gene regulatory networks in schizophrenia. *Neuroimage,* 53**,** 839-847.

Power R. A., Steinberg S., Bjornsdottir G., Rietveld C. A., Abdellaoui A., Nivard M. M., Johannesson M., Galesloot T. E., Hottenga J. J., Willemsen G., Cesarini D., Benjamin D. J., Magnusson P. K. E., Ullen F., Tiemeier H., Hofman A., Van Rooij F. J. A., Walters G. B., Sigurdsson E., Thorgeirsson T. E., Ingason A., Helgason A., Kong A., Kiemeney L. A., Koellinger P., Boomsma D. I., Gudbjartsson D., Stefansson H. and Stefansson K. 2015. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nature Neuroscience,* 18**,** 953-955.

Pribill I., Speiser P., Leary J., Leodolter S., Hacker N. F., Friedlander M. L., Birnbaum D., Zeillinger R. and Krainer M. 2001. High frequency of allelic imbalance at regions of chromosome arm 8p in ovarian carcinoma. *Cancer Genetics and Cytogenetics,* 129**,** 23-29.

Pritchard M. and Kola I. 2007. The biological bases of pharmacological therapies in down syndrome. *Therapies and Rehabilitation in Down syndrome***,** 18-27.

Prout H. T. and Nowak-Drabik K. M. 2003. Psychotherapy with persons who have mental retardation: An evaluation of effectiveness. *American Journal on Mental Retardation,* 108**,** 82-93.

Prout T. A. and Wadkins M. J. 2014. *Essential interviewing and counseling skills: An integrated approach to practice*, Springer Publishing Company.

Prüßing K., Voigt A. and Schulz J. B. 2013. Drosophila melanogaster as a model organism for alzheimer's disease. *Molecular Neurodegeneration,* 8**,** 35.

Purandare C. N. 2012. Maternal nutritional deficiencies and interventions. *Journal of Obstetrics and Gynaecology of India,* 62**,** 621-623.

Purcell S. M., Moran J. L., Fromer M., Ruderfer D., Solovieff N., Roussos P., O'dushlaine C., Chambert K., Bergen S. E. and Kähler A. 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature,* 506**,** 185.

Purcell S. M., Wray N. R., Stone J. L., Visscher P. M., O'donovan M. C., Sullivan P. F., Sklar P., Ruderfer D. M., Mcquillin A. and Morris D. W. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature,* 460**,** 748-752.

Qiu S., Luo S., Evgrafov O., Li R., Schroth G. P., Levitt P., Knowles J. A. and Wang K. 2012. Single-neuron rna-seq: Technical feasibility and reproducibility. *Frontiers in Genetics,* 3**,** 124.

Qu Y., Hahn I., Webb S. E., Pearce S. P. and Prokop A. 2017. Periodic actin structures in neuronal axons are required to maintain microtubules. *Molecular Biology of the Cell,* 28**,** 296-308.

Rabbani B., Tekin M. and Mahdieh N. 2014. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics,* 59**,** 5-15.

Rajab A., Al Rashdi I. and Al Salmi Q. 2013. Genetic services and testing in the sultanate of oman. Sultanate of oman steps into modern genetics. *Journal of Community Genetics,* 4**,** 391-397.

Rajab A., Hamza N., Al Harasi S., Al Lawati F., Gibbons U., Al Alawi I., Kobus K., Hassan S., Mahir G. and Al Salmi Q. 2015. Repository of mutations from oman: The entry point to a national mutation database. *F1000Research,* 4.

Rajab A. and Patton M. 2000. Short report: A study of consanguinity in the sultanate of oman. *Annals of Human Biology,* 27**,** 321-326.

Ramprasad V. L., Soumittra N., Nancarrow D., Sen P., Mckibbin M., Williams G. A., Arokiasamy T., Lakshmipathy P., Inglehearn C. F. and Kumaramanickavel G. 2008. Identification of a novel splice-site mutation in the lebercilin (lca5) gene causing leber congenital amaurosis.

Rasic D., Hajek T., Alda M. and Uher R. 2014. Risk of mental illness in offspring of parents with schizophrenia, bipolar disorder, and major depressive disorder: A meta-analysis of family high-risk studies. *Schizophrenia Bulletin,* 40**,** 28-38.

Ratterman D. M. 2003. Eliminating ether by using ice for drosophila labs. *Tested Studies For Laboratory Teaching,* 24**,** 259-265.

Raymond F. L. 2006. X linked mental retardation: A clinical guide. *Journal of Medical Genetics,* 43**,** 193-200.

Reddy M. S. 2016. Lack of insight in psychiatric illness: A critical appraisal. *Indian Journal of Psychological Medicine,* 38**,** 169-171.

Regier D. A. 2007. Time for a fresh start? Rethinking psychosis in dsm-v. *Schizophrenia Bulletin,* 33**,** 843-845.

Reuter M. S., Tawamie H., Buchert R. and Et Al. 2017. Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *The Journal of the American Medical Association Psychiatry*, 74(3):293-299.

Richards A. L., Jones L., Moskvina V., Kirov G., Gejman P. V., Levinson D. F., Sanders A. R., Purcell S., Visscher P. M. and Craddock N. 2012. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Molecular psychiatry,* 17**,** 193.

Richardson K. and Norgate S. H. 2015. Does iq really predict job performance? *Applied Developmental Science,* 19**,** 153-169.

Richardson S. A., Koller H. and Katz M. 1987. Sex differences in the classification of children as mildly mentally retarded. *Upsala Journal of Medical Sciences. Supplement*, 44**,** 83-8.

Rilak Z., Wernicke S. and Bogicevic I. 2014. Keeping genomic data safe on the cloud. *Journal of Biomolecular Techniques : JBT,* 25**,** S5-S5.

Ripke S., Neale B. M., Corvin A., Walters J. T., Farh K.-H., Holmans P. A., Lee P., Bulik-Sullivan B., Collier D. A. and Huang H. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421.

Ripke S., O'dushlaine C., Chambert K., Moran J. L., Kähler A. K., Akterin S., Bergen S., Collins A. L., Crowley J. J., Fromer M., Kim Y., Lee S. H., Magnusson P. K. E., Sanchez N., Stahl E. A., Williams S., Wray N. R., Xia K., Bettella F., Børglum A. D., Bulik-Sullivan B. K., Cormican P., Craddock N., De Leeuw C., Durmishi N., Gill M., Golimbet V., Hamshere M. L., Holmans P., Hougaard D. M., Kendler K. S., Lin K., Morris D. W., Mors O., Mortensen P. B., Neale B. M., O'neill F. A., Owen M. J.,

Milovancevic M., Posthuma D., Powell J., Richards A. L., Riley B. P., Ruderfer D., Rujescu D., Sigurdsson E., Silagadze T., Smit A. B., Stefansson H., Steinberg S., Suvisaari J., Tosato S., Verhage M., Walters J. T., Multicenter Genetic Studies of Schizophrenia C., Psychosis Endophenotypes Consortium W. T. C.-C. C., Bramon E., Corvin A. P., O'donovan M. C., Stefansson K., Scolnick E., Purcell S., Mccarroll S., Sklar P., Hultman C. M. and Sullivan P. F. 2013. Genome-wide association analysis identifies 14 new risk loci for schizophrenia. *Nature Genetics,* 45**,** 1150-1159.

Robertson J., Emerson E., Pinkney L., Caesar E., Felce D., Meek A., Carr D., Lowe K., Knapp M. and Hallam A. 2005. Treatment and management of challenging behaviours in congregate and noncongregate community-based supported accommodation. *Journal of Intellectual Disability Research,* 49**,** 63-72.

Robinson J. T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E. S., Getz G. and Mesirov J. P. 2011. Integrative genomics viewer. *Nature Biotechnology,* 29**,** 24-26.

Rollins B., Martin M. V., Morgan L. and Vawter M. P. 2010. Analysis of whole genome biomarker expression in blood and brain. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : the Official Publication of the International Society of Psychiatric Genetics,* 153B**,** 919-936.

Romani M., Micalizzi A. and Valente E. M. 2013. Joubert syndrome: Congenital cerebellar ataxia with the "molar tooth". *Lancet Neurology,* 12**,** 10.1016/S1474-4422(13)70136-4.

Ronaghi M., Uhlén M. and Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science,* 281**,** 363.

Ronald A. 2015. Recent quantitative genetic research on psychotic experiences: New approaches to old questions. *Current Opinion in Behavioral Sciences,* 2**,** 81-88.

Rondal J.-A. and Lang S. 2010. Towards hybrid therapeutic strategies in intellectual disabilities. *International Journal of Behavioral Consultation and Therapy,* 5**,** 270.

Rondal J.-A. and Rasore-Quartino A. 2007. *Therapies and rehabilitation in down syndrome*, John Wiley & Sons.

Ropers H. 2006. X-linked mental retardation: Many genes for a complex disorder. *Current Opinion in Genetics & Development,* 16**,** 260-269.

Ropers H. 2010. Genetics of early onset cognitive impairment. *Annual Review of Genomics and Human Genetics,* 11**,** 161-187.

Ropers H. and Hamel B. C. J. 2005. X-linked mental retardation. *Nature Review Genetics,* 6**,** 46-57.

Ropers H. and Wienker T. 2015. Penetrance of pathogenic mutations in haploinsufficient genes for intellectual disability and related disorders. *European Journal of Medical Genetics,* 58**,** 715-718.

Rosenheck R. A., Leslie D. L., Sindelar J., Miller E. A., Lin H., Stroup T. S., Mcevoy J., Davis S. M., Keefe R. S. and Swartz M. 2006. Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia. *American Journal of Psychiatry,* 163**,** 2080-2089.

Rosenthal N. and Brown S. 2007. The mouse ascending: Perspectives for human-disease models. *Nature Cell Biology,* 9**,** 993-999.

Ross M. T., Grafham D. V., Coffey A. J., Scherer S., Mclay K., Muzny D., Platzer M., Howell G. R., Burrows C. and Bird C. P. 2005. The DNA sequence of the human x chromosome. *Nature,* 434**,** 325-337.

Rothberg J. M. and Leamon J. H. 2008. The development and impact of 454 sequencing. *Nature Biotechnology,* 26**,** 1117-1124.

Roy A., Kucukural A. and Zhang Y. 2010. I-tasser: A unified platform for automated protein structure and function prediction. *Nature Protocols,* 5**,** 725-738.

Rumsey J. M. and Ernst M. 2009. *Neuroimaging in developmental clinical neuroscience,* Cambridge University Press.

Rush K. S., Bowman L. G., Eidman S. L., Toole L. M. and Mortenson B. P. 2004. Assessing psychopathology in individuals with developmental disabilities. *Behavior Modification,* 28**,** 621-637.

Ruzzo E. K. and Geschwind D. H. 2016. Schizophrenia genetics complements its mechanistic understanding. *Nature Neuroscience,* 19**,** 523-525.

Sachdev P. S., Mohan A., Taylor L. and Jeste D. V. 2015. Dsm-5 and mental disorders in older individuals: An overview. *Harvard Review of Psychiatry,* 23**,** 320-328.

Sachidanandam R., Weissman D., Schmidt S. C., Kakol J. M., Stein L. D., Marth G., Sherry S., Mullikin J. C., Mortimore B. J. and Willey D. L. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-933.

Sachs N., Sawa A., Holmes S., Ross C., Delisi L. and Margolis R. 2005. A frameshift mutation in disrupted in schizophrenia 1 in an american family with schizophrenia and schizoaffective disorder. *Molecular psychiatry,* 10**,** 758-764.

Sahay A., Scobie K. N., Hill A. S., O'carroll C. M., Kheirbek M. A., Burghardt N. S., Fenton A. A., Dranovsky A. and Hen R. 2011. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature,* 472**,** 466-470.

Saiki R. K., Gelfand D. H., Stoffel S., Scharf S. J. and Higuchi R. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487.

Sakdalan J. A., Shaw J. and Collier V. 2010. Staying in the here-and-now: A pilot study on the use of dialectical behaviour therapy group skills training for forensic clients with intellectual disability. *Journal of Intellectual Disability Research,* 54**,** 568-572.

Saldarriaga W., Tassone F., González-Teshima L. Y., Forero-Forero J. V., Ayala-Zapata S. and Hagerman R. 2014. Fragile x syndrome. *Colombia Médica : CM,* 45**,** 190-198.

Samarakoon P. S., Sorte H. S., Kristiansen B. E., Skodje T., Sheng Y., Tjønnfjord G. E., Stadheim B., Stray-Pedersen A., Rødningen O. K. and Lyle R. 2014. Identification of copy number variants from exome sequence data. *BMC genomics,* 15**,** 661.

Samocha K. E., Robinson E. B., Sanders S. J., Stevens C., Sabo A., Mcgrath L. M., Kosmicki J. A., Rehnström K., Mallick S., Kirby A., Wall D. P., Macarthur D. G., Gabriel S. B., Depristo M., Purcell S. M., Palotie A., Boerwinkle E., Buxbaum J. D., Cook E. H., Gibbs R. A., Schellenberg G. D., Sutcliffe J. S., Devlin B., Roeder K., Neale B. M. and Daly M. J. 2014. A framework for the interpretation of de novo mutation in human disease. *Nature Genetics,* 46**,** 944-950.

Sanger F. 1977. Nucleotide sequence of bacteriophage (d x174 DNA.

Sarkar S., Hillner K. and Velligan D. I. 2015. Conceptualization and treatment of negative symptoms in schizophrenia. *World Journal of Psychiatry,* 5**,** 352.

Saunders N. R., Ek C. J., Habgood M. D. and Dziegielewska K. M. 2008. Barriers in the brain: A renaissance? *Trends in Neurosciences,* 31**,** 279-286.

Schaefer G. B., Bodensteiner J. B., Thompson J. N., Kimberling W. J. and Craft J. M. 1998. Volumetric neuroimaging in usher syndrome: Evidence of global involvement. *American Journal of Medical Genetics,* 79**,** 1-4.

Scharf S. H., Jaeschke G., Wettstein J. G. and Lindemann L. 2015. Metabotropic glutamate receptor 5 as drug target for fragile x syndrome. *Current Opinion in Pharmacology,* 20**,** 124-134.

Schee Genannt Halfmann S., Evangelatos N., Schröder-Bäck P. and Brand A. 2017. European healthcare systems readiness to shift from 'one-size fits all'to personalized medicine. *Personalized Medicine,* 14**,** 63-74.

Schierup C.-U., Hansen P. and Castles S. 2006. *Migration, citizenship, and the european welfare state: A european dilemma*, Oxford University Press on Demand.

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez J.-Y., White D. J., Hartenstein V., Eliceiri K., Tomancak P. and Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nature Methods,* 9**,** 676-682.

Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421-427.

Schlosberg A. 2016. Data security in genomics: A review of australian privacy requirements and their relation to cryptography in data storage. *Journal of Pathology Informatics,* 7**,** 6.

Schmidt S., Gerasimova A., Kondrashov F. A., Adzuhbei I. A., Kondrashov A. S. and Sunyaev S. 2008. Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet,* 4**,** e1000281.

Schnekenberg R. P. and Németh A. H. 2013. Next-generation sequencing in childhood disorders. *Archives of Disease in Childhood,* archdischild-2012-302881.

Schnyder U. and Cloitre M. 2015. *Evidence based treatments for trauma-related psychological disorders: A practical guide for clinicians*, Springer.

Schork N. J., Murray S. S., Frazer K. A. and Topol E. J. 2009. Common vs. Rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development,* 19**,** 212-219.

Schuchardt K., Gebhardt M. and Mäehler C. 2010. Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research,* 54**,** 346-353.

Schwarz J. M., Cooper D. N., Schuelke M. and Seelow D. 2014. Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods,* 11**,** 361-362.

Seeman P. 2009. Glutamate and dopamine components in schizophrenia. *Journal of Psychiatry & Neuroscience : JPN,* 34**,** 143-149.

Sekar A., Bialas A. R., De Rivera H., Davis A., Hammond T. R., Kamitaki N., Tooley K., Presumey J., Baum M., Van Doren V., Genovese G., Rose S. A., Handsaker R. E., Schizophrenia Working Group of the Psychiatric Genomics C., Daly M. J., Carroll M. C., Stevens B. and Mccarroll S. A. 2016. Schizophrenia risk from complex variation of complement component 4. *Nature,* 530**,** 177-183.

Selten J.-P., Cantor-Graae E. and Kahn R. S. 2007. Migration and schizophrenia. *Current opinion in psychiatry,* 20**,** 111-115.

Senturias Y. and Asamoah A. 2014. Fetal alcohol spectrum disorders: Guidance for recognition, diagnosis, differential diagnosis and referral. *Current Problems in Pediatric and Adolescent Health Care,* 44**,** 88-95.

Serebriiskii I. G. and Golemis E. A. 2001. Two-hybrid system and false positives. *In:* Macdonald P. N. (ed.) *Two-hybrid systems: Methods and protocols.* Totowa, NJ: Humana Press.

Shabbir M. I., Ahmed Z. M., Khan S. Y., Riazuddin S., Waryah A. M., Khan S. N., Camps R. D., Ghosh M., Kabra M., Belyantseva I. A., Friedman T. B. and Riazuddin S. 2006. Mutations of human tmhs cause recessively inherited non-syndromic hearing loss. *Journal of Medical Genetics,* 43**,** 634-40.

Shah C. 2017. Reducing antipsychotic prescribing in people with learning disabilities. *Depression,* 10, 58.

Shahin H., Walsh T., Rayyan A. A., Lee M. K., Higgins J., Dickel D., Lewis K., Thompson J., Baker C., Nord A. S., Stray S., Gurwitz D., Avraham K. B., King M. C. and Kanaan M. 2010. Five novel loci for inherited hearing loss mapped by snp-based homozygosity profiles in palestinian families. *European Journal of Human Genetics,* 18**,** 407-13.

Sharkia R., Mahajnah M., Athamny E., Khatib M., Sheikh-Muhammad A. and Zalan A. 2016. Changes in marriage patterns among the arab community in israel over a 60-year period. *Journal of Biosocial Science,* 48**,** 283-287.

Sharma S., Ang S. L., Shaw M., Mackey D. A., Gécz J., Mcavoy J. W. and Craig J. E. 2006. Nance–horan syndrome protein, nhs, associates with epithelial cell junctions. *Human Molecular Genetics,* 15**,** 1972-1983.

Sharp Andrew j., Locke Devin p., Mcgrath Sean d., Cheng Z., Bailey Jeffrey a., Vallente Rhea u., Pertz Lisa m., Clark Royden a., Schwartz S., Segraves R., Oseroff Vanessa v., Albertson Donna g., Pinkel D. and Eichler Evan e. 2005. Segmental duplications and copy-number variation in the human genome. *American Journal of Human Genetics,* 77**,** 78-88.

Shaw A. 2014. *Kinship and continuity: Pakistani families in britain*, Routledge.

Shawky R. M., Elsayed S. M., Zaki M. E., El-Din S. M. N. and Kamal F. M. 2013. Consanguinity and its relevance to clinical genetics. *Egyptian Journal of Medical Human Genetics,* 14**,** 157-164.

Shekunov J. 2016. Immigration and risk of psychiatric disorders: A review of existing literature. *American Journal of Psychiatry Residents' Journal,* 11**,** 3-5.

Shen L.-H., Liao M.-H. and Tseng Y.-C. 2012. Recent advances in imaging of dopaminergic neurons for evaluation of neuropsychiatric disorders. *Journal of Biomedicine and Biotechnology,* 2012**,** 14.

Shental-Bechor D. and Levy Y. 2008. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 8256-8261.

Shim S. Y., Wang J., Asada N., Neumayer G., Tran H. C., Ishiguro K., Sanada K., Nakatani Y. and Nguyen M. D. 2008. Protein 600 is a microtubule/endoplasmic reticulum-associated protein in cns neurons. *Journal of Neuroscience,* 28**,** 3604-14.

Sidransky E., Nalls M. A., Aasly J. O., Aharon-Peretz J., Annesi G., Barbosa E. R., Bar-Shira A., Berg D., Bras J. and Brice A. 2009. Multicenter analysis of glucocerebrosidase mutations in parkinson's disease. *New England Journal of Medicine,* 361**,** 1651-1661.

Sigafoos J. and Schlosser R. W. 2008. Applied behavior analysis is not an autism therapy. Taylor & Francis.

Sigelman C. K. and Rider E. A. 2014. *Life-span human development*, Cengage Learning.

Silverman W., Miezejeski C., Ryan R., Zigman W., Krinsky-Mchale S. and Urv T. 2010. Stanford-binet & wais iq differences and their implications for adults with intellectual disability (aka mental retardation). *Intelligence,* 38**,** 242-248.

Simon C. M., Williams J. K., Shinkunas L., Brandt D., Daack-Hirsch S. and Driessnack M. 2011. Informed consent and genomic incidental findings: Irb chair perspectives. *Journal of Empirical Research on Human Research Ethics,* 6**,** 53-67.

Singer T., Mcconnell M. J., Marchetto M. C., Coufal N. G. and Gage F. H. 2010. Line-1 retrotransposons: Mediators of somatic variation in neuronal genomes? *Trends in Neurosciences,* 33**,** 345-354.

Singh T., Kurki M. I., Curtis D., Purcell S. M., Crooks L., Mcrae J., Suvisaari J., Chheda H., Blackwood D., Breen G., Pietilainen O., Gerety S. S., Ayub M., Blyth M., Cole T., Collier D., Coomber E. L., Craddock N., Daly M. J., Danesh J., Diforti M., Foster A., Freimer N. B., Geschwind D., Johnstone M., Joss S., Kirov G., Korkko J., Kuismin O., Holmans P., Hultman C. M., Iyegbe C., Lonnqvist J., Mannikko M., Mccarroll S. A., Mcguffin P., Mcintosh A. M., Mcquillin A., Moilanen J. S., Moore C., Murray R. M., Newbury-Ecob R., Ouwehand W., Paunio T., Prigmore E., Rees E., Roberts D., Sambrook J., Sklar P., Clair D. S., Veijola J., Walters J. T. R., Williams H., Swedish Schizophrenia S., Study I., Study D. D. D., Consortium U. K., Sullivan P. F., Hurles M. E., O'donovan M. C., Palotie A., Owen M. J. and Barrett J. C. 2016. Rare loss-of-function variants in setd1a are associated with schizophrenia and developmental disorders. *Nature Neuroscience,* 19**,** 571-577.

Sipos B., Massingham T., Stütz A. M. and Goldman N. 2012. An improved protocol for sequencing of repetitive genomic regions and structural variations using mutagenesis and next generation sequencing. *PloS One,* 7, e43359.

Siskind D., Mccartney L., Goldschlager R. and Kisely S. 2016. Clozapine v. First- and second-generation antipsychotics in treatment-refractory schizophrenia: Systematic review and meta-analysis. *The British Journal of Psychiatry,* 209**,** 385-392.

Siwicki K. K., Riccio P., Ladewski L., Marcillac F., Dartevelle L., Cross S. A. and Ferveur J. F. 2005. The role of cuticular pheromones in courtship conditioning of drosophila males. *Learning and Memory,* 12**,** 636-45.

Sklar P., Smoller J. W., Fan J., Ferreira M. A., Perlis R. H., Chambert K., Nimgaonkar V. L., Mcqueen M. B., Faraone S. V., Kirby A., De Bakker P. I., Ogdie M. N., Thase M. E., Sachs G. S., Todd-Brown K., Gabriel S. B., Sougnez C., Gates C., Blumenstiel B., Defelice M., Ardlie K. G., Franklin J., Muir W. J., Mcghee K. A., Macintyre D. J., Mclean A., Vanbeck M., Mcquillin A., Bass N. J., Robinson M., Lawrence J., Anjorin A., Curtis D., Scolnick E. M., Daly M. J., Blackwood D. H., Gurling H. M. and Purcell S. M. 2008. Whole-genome association study of bipolar disorder. *Molecular Psychiatry,* 13**,** 558-69.

Sloan-Heggen C. M., Babanejad M., Beheshtian M., Simpson A. C., Booth K. T., Ardalani F., Frees K. L., Mohseni M., Mozafari R., Mehrjoo Z., Jamali L., Vaziri S., Akhtarkhavari T., Bazazzadegan N., Nikzat N., Arzhangi S., Sabbagh F., Otukesh H., Seifati S. M., Khodaei H., Taghdiri M., Meyer N. C., Daneshi A., Farhadi M., Kahrizi K., Smith R. J., Azaiez H. and Najmabadi H. 2015. Characterising the spectrum of autosomal recessive hereditary hearing loss in iran. *Journal of Medical Genetics,* 52**,** 823-9.

Slocum T. A., Detrich R., Wilczynski S. M., Spencer T. D., Lewis T. and Wolfe K. 2014. The evidence-based practice of applied behavior analysis. *The Behavior Analyst,* 37**,** 41-56.

Smith C. 2007. Cloning and mutagenesis: Tinkering with the order of things. *Nature Methods,* 4**,** 455.

Soleimani F., Zaheri F. and Abdi F. 2014. Long-term neurodevelopmental outcomes after preterm birth. *Iranian Red Crescent Medical Journal,* 16**,** e17965.

Soliman M. A., Aboharb F., Zeltner N. and Studer L. 2017. Pluripotent stem cells in neuropsychiatric disorders. *Molecular Psychiatry*, doi: 10.1038/40.

Sommer I. E., Bearden C. E., Van Dellen E., Breetvelt E. J., Duijff S. N., Maijer K., Van Amelsvoort T., De Haan L., Gur R. E., Arango C., Díaz-Caneja C. M., Vinkers C. H. and Vorstman J. a. S. 2016. Early interventions in risk groups for schizophrenia: What are we waiting for? *NPJ Schizophrenia,* 2**,** 16003.

Sperling R. 2011. The potential of functional mri as a biomarker in early alzheimer's disease. *Neurobiology of Aging,* 32**,** S37-S43.

Spreen O., Risser A. H. and Edgell D. 1995. *Developmental neuropsychology*, Oxford University Press, USA.

Srivastava A. K. and Schwartz C. E. 2014. Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms. *Neuroscience and Biobehavioral Reviews,* 46 Pt 2**,** 161-174.

Stahl S. and Buckley P. F. 2007. Negative symptoms of schizophrenia: A problem that will not go away. *Acta Psychiatrica Scandinavica,* 115**,** 4-11.

Stefanis N., Thewissen V., Bakoula C., Van Os J. and Myin-Germeys I. 2006. Hearing impairment and psychosis: A replication in a cohort of young adults. *Schizophrenia Research,* 85**,** 266-272.

Stefansson H., Rujescu D., Cichon S., Pietiläinen O. P., Ingason A., Steinberg S., Fossdal R., Sigurdsson E., Sigmundsson T. and Buizer-Voskamp J. E. 2008. Large recurrent microdeletions associated with schizophrenia. *nature,* 455**,** 232.

Stephens J. C., Schneider J. A., Tanguay D. A., Choi J., Acharya T., Stanley S. E., Jiang R., Messer C. J., Chew A., Han J.-H., Duan J., Carr J. L., Lee M. S., Koshy B., Kumar A. M., Zhang G., Newell W. R., Windemuth A., Xu C., Kalbfleisch T. S., Shaner S. L., Arnold K., Schulz V., Drysdale C. M., Nandabalan K., Judson R. S., Ruaño G. and Vovis G. F. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science,* 293**,** 489-493.

Sterling P. and Matthews G. 2005. Structure and function of ribbon synapses. *Trends in Neurosciences,* 28**,** 20-29.

Stevenson R. E., Schwartz C. E., Rogers R. C. and Rogers R. C. 2012. *Atlas of x-linked intellectual disability syndromes*, Oxford University Press.

Stiles J. and Jernigan T. L. 2010. The basics of brain development. *Neuropsychology Review,* 20**,** 327-348.

Stine O. C., Mcmahon F. J., Chen L. S., Xu J., Meyers D. A., Mackinnon D. F., Simpson S., Mcinnis M. G., Rice J. P. and Goate A. 1997. Initial genome screen for bipolar disorder in the nimh genetics initiative pedigrees: Chromosomes 2, 11, 13, 14, and x. *American Journal of Medical Genetics Part A,* 74**,** 263-269.

Stöber G., Franzek E., Lesch K. P. and Beckmann H. 1995. Periodic catatonia: A schizophrenic subtype with major gene effect and anticipation. *European Archives of Psychiatry and Clinical Neuroscience,* 245**,** 135-141.

Stolzer J. M. 2016. The meteoric rise of mental illness in america and implications for other countries. *The European Journal of Counselling Psychology,* 4.

Strauss E., Sherman E. M. and Spreen O. 2006. *A compendium of neuropsychological tests: Administration, norms, and commentary,* American Chemical Society.

Striedter G. F. 2015. *Neurobiology, a functional approach, first edition,* Oxford, Oxford University Press.

Suite D. H., La Bril R., Primm A. and Harrison-Ross P. 2007. Beyond misdiagnosis, misunderstanding and mistrust: Relevance of the historical perspective in the medical and mental health treatment of people of color. *Journal of the National Medical Association,* 99**,** 879-885.

Sullivan P. F., Kendler K. S. and Neale M. C. 2003. Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of general psychiatry,* 60**,** 1187-1192.

Sun F. F., Johnson J. E., Zeidler M. P. and Bateman J. R. 2012. Simplified insertion of transgenes onto balancer chromosomes via recombinase-mediated cassette exchange. *G3: Genes| Genomes| Genetics,* 2**,** 551-553.

Sundaramurthy S., Swaminathan M., Sen P., Arokiasamy T., Deshpande S., John N., Gadkari R. A., Mannan A. U. and Soumittra N. 2016. Homozygosity mapping guided next generation sequencing to identify the causative genetic variation in inherited retinal degenerative diseases. *Journal of Human Genetics.*

Swaminathan S., Edward B. and Kurpad A. 2013. Micronutrient deficiency and cognitive and physical performance in indian children. *European journal of clinical nutrition,* 67**,** 467-474.

Taioli E., Pedotti P. and Garte S. 2004. Importance of allele frequency estimates in epidemiological studies. *Mutation Research/Reviews in Mutation Research,* 567**,** 63-70.

Takata A., Xu B., Ionita-Laza I., Roos J. L., Gogos J. A. and Karayiorgou M. 2014. Loss-of-function variants in schizophrenia risk and setd1a as a candidate susceptibility gene. *Neuron,* 82**,** 773-780.

Tamminga C. A. 2000. Treatment mechanisms: Traditional and new antipsychotic drugs. *Dialogues in Clinical Neuroscience,* 2**,** 281-286.

Tandon R., Gaebel W., Barch D. M., Bustillo J., Gur R. E., Heckers S., Malaspina D., Owen M. J., Schultz S. and Tsuang M. 2013. Definition and description of schizophrenia in the dsm-5. *Schizophrenia Research,* 150**,** 3-10.

Tandon R., Keshavan M. S. and Nasrallah H. A. 2008. Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. *Schizophrenia Research,* 102**,** 1-18.

Tattini L., D'aurizio R. and Magi A. 2015. Detection of genomic structural variants from next-generation sequencing data. *Frontiers in Bioengineering and Biotechnology,* 3**,** 92.

Taylor D. M. 2017. Clozapine for treatment-resistant schizophrenia: Still the gold standard? *CNS Drugs,* 31**,** 177-180.

Tejada M.-I., Martínez-Bouzas C., García-Ribes A., Larrucea S., Acquadro F., Cigudosa J.-C., Belet S., Froyen G. and López-Aríztegui M.-A. 2011. A child with mild x-linked intellectual disability and a microduplication at xp22.12 including rps6ka3. *Pediatrics,* 128**,** e1029-e1033.

The International Schizophrenia Consortium 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature,* 455**,** 237-241.

Thermenos H., Keshavan M., Juelich R., Molokotos E., Whitfield-Gabrieli S., Brent B., Makris N. and Seidman L. 2013. A review of neuroimaging studies of young relatives of individuals with schizophrenia: A developmental perspective from schizotaxia to schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 162**,** 604-635.

Tian Q., Li Y., Kousar R., Guo H., Peng F., Zheng Y., Yang X., Long Z., Tian R., Xia K., Lin H. and Pan Q. 2017. A novel nhs mutation causes nance-horan syndrome in a chinese family. *BMC Medical Genetics,* 18**,** 2.

Tiffin N., Andrade-Navarro M. A. and Perez-Iratxeta C. 2009. Linking genes to diseases: It's all in the data. *Genome Medicine,* 1**,** 77-77.

Tiwari A., Lemke J., Altmueller J., Thiele H., Glaus E., Fleischhauer J., Nürnberg P., Neidhardt J. and Berger W. 2016. Identification of novel and recurrent disease-causing mutations in retinal dystrophies using whole exome sequencing (wes): Benefits and limitations. *PloS One,* 11**,** e0158692.

Toonen R. F., De Vries K. J., Zalm R., Südhof T. C. and Verhage M. 2005. Munc18–1 stabilizes syntaxin 1, but is not essential for syntaxin 1 targeting and snare complex formation. *Journal of neurochemistry,* 93**,** 1393-1400.

Tooney P. A. 2016. Attention: Schizophrenia risk gene product mir-137 now targeting efnb2. *EBioMedicine,* 12**,** 10-11.

Trachtenberg M., Parsonage M., Shepherd G. and Boardman J. 2013. Peer support in mental health care: Is it good value for money?. Centre for Mental Health, London, UK.

Trattler W. B., Kaiser P. K. and Friedman N. J. 2012. *Review of ophthalmology: Expert consult-online and print,* Elsevier Health Sciences.

Trinh J. and Farrer M. 2013. Advances in the genetics of parkinson disease. *Nature Reviews Neurology,* 9**,** 445-454.

Tsuang M. 2000. Schizophrenia: Genes and environment. *Biological Psychiatry,* 47**,** 210-220.

Tsukamoto Y., Senda T., Nakano T., Nakada C., Hida T., Ishiguro N., Kondo G., Baba T., Sato K. and Osaki M. 2002. Arpp, a new homolog of carp, is preferentially expressed in type 1 skeletal muscle fibers and is markedly induced by denervation. *Laboratory Investigation,* 82**,** 645-655.

Tug E., Dilek N. F., Javadiyan S., Burdon K. P. and Percin F. E. 2013. A turkish family with nance-horan syndrome due to a novel mutation. *Gene,* 525**,** 141-145.

Turnbull C., Ahmed S., Morrison J., Pernet D., Renwick A., Maranian M., Seal S., Ghoussaini M., Hines S., Healey C. S., Hughes D., Warren-Perry M., Tapper W., Eccles D., Evans D. G., Hooning M., Schutte M., Van Den Ouweland A., Houlston R., Ross G., Langford C., Pharoah P. D. P., Stratton M. R., Dunning A. M., Rahman N. and Easton D. F. 2010. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genetics,* 42**,** 504-507.

Tyrer F., Mcgrother C., Thorp C., Donaldson M., Bhaumik S., Watson J. and Hollin C. 2006. Physical aggression towards others in adults with learning disabilities: Prevalence and associated factors. *Journal of Intellectual Disability Research,* 50**,** 295-304.

Tyrer P., Cooper S.-A. and Hassiotis A. 2014. Drug treatments in people with intellectual disability and challenging behaviour. *British Medical Journal,* 349**,** g4323.

Tzvetkov M. and Von Ahsen N. 2012. Pharmacogenetic screening for drug therapy: From single gene markers to decision making in the next generation sequencing era. *Pathology-Journal of the RCPA,* 44**,** 166-180.

Urbán N. and Guillemot F. 2014. Neurogenesis in the embryonic and adult brain: Same regulators, different roles. *Frontiers in Cellular Neuroscience,* 8**,** 396.

Van Bon B. W. M., Koolen D. A., Brueton L., Mcmullan D., Lichtenbelt K. D., Adès L. C., Peters G., Gibson K., Novara F., Pramparo T., Bernardina B. D., Zoccante L., Balottin U., Piazza F., Pecile V., Gasparini P., Guerci V., Kets M., Pfundt R., De Brouwer A. P., Veltman J. A., De Leeuw N., Wilson M., Antony J., Reitano S., Luciano D., Fichera M., Romano C., Brunner H. G., Zuffardi O. and De Vries B. B. A. 2010. The 2q23.1 microdeletion syndrome: Clinical and behavioural phenotype. *European Journal of Human Genetics,* 18**,** 163-170.

Van Cauwenberghe C., Van Broeckhoven C. and Sleegers K. 2016. The genetic landscape of alzheimer disease: Clinical implications and perspectives. *Genetics in Medicine,* 18**,** 421-430.

Van Den Oord E. J., Kuo P.-H., Hartmann A. M., Webb B. T., Möller H.-J., Hettema J. M., Giegling I., Bukszár J. and Rujescu D. 2008. Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. *Archives of General Psychiatry,* 65**,** 1062-1071.

Van Der Velde K. J., Kuiper J., Thompson B. A., Plazzer J. P., Van Valkenhoef G., De Haan M., Jongbloed J. D. H., Wijmenga C., De Koning T. J., Abbott K. M., Sinke R., Spurdle A. B., Macrae F., Genuardi M., Sijmons R. H., Swertz M. A. and Insi G. H. T. G. 2015. Evaluation of cadd scores in curated mismatch repair gene variants yields a model for clinical validation and prioritization. *Human Mutation,* 36**,** 712-719.

Van Dongen J. and Boomsma D. I. 2013. The evolutionary paradox and the missing heritability of schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 162**,** 122-136.

Van El C. G., Cornel M. C., Borry P., Hastings R. J., Fellmann F., Hodgson S. V., Howard H. C., Cambon-Thomsen A., Knoppers B. M. and Meijers-Heijboer H. 2013. Whole-genome sequencing in health care. *European Journal of Human Genetics,* 21**,** S1-S5.

Van Gastel W. A., Tempelaar W., Bun C., Schubart C. D., Kahn R. S., Plevier C. and Boks M. P. M. 2013. Cannabis use as an indicator of risk for mental health problems in adolescents: A population-based study at secondary schools. *Psychological Medicine,* 43**,** 1849-1856.

Van Gorp P. and Comuzzi M. 2014. Lifelong personal health data and application software via virtual machines in the cloud. *IEEE Journal of Biomedical and Health Informatics,* 18**,** 36-45.

Van Nimwegen K. J. M., Van Soest R. A., Veltman J. A., Nelen M. R., Van Der Wilt G. J., Vissers L. E. L. M. and Grutters J. P. C. 2016. Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing. *Clinical Chemistry*.

Van Wijk E., Van Der Zwaag B., Peters T., Zimmermann U., Te Brinke H., Kersten F. F., Märker T., Aller E., Hoefsloot L. H. and Cremers C. W. 2006a. The dfnb31 gene

product whirlin connects to the usher protein network in the cochlea and retina by direct association with ush2a and vlgr1. *Human Molecular Genetics,* 15**,** 751-765.

Van Wijk E., Van Der Zwaag B., Peters T., Zimmermann U., Te Brinke H., Kersten F. F. J., Marker T., Aller E., Hoefsloot L. H., Cremers C. W. R. J., Cremers F. P. M., Wolfrum U., Knipper M., Roepman R. and Kremer H. 2006b. The dfnb31 gene product whirlin connects to the usher protein network in the cochlea and retina by direct association with ush2a and vlgr1. *Human Molecular Genetics,* 15**,** 751-765.

Vanhara P., Horak P., Pils D., Anees M., Petz M., Gregor W., Zeillinger R. and Krainer M. 2013. Loss of the oligosaccharyl transferase subunit tusc3 promotes proliferation and migration of ovarian cancer cells. *International Journal of Oncology,* 42**,** 1383-1389.

Venkatesan B. M. and Bashir R. 2011. Nanopore sensors for nucleic acid analysis. *Nature Nanotechnology,* 6**,** 615-624.

Venkatesan B. M., Dorvel B., Yemenicioglu S., Watkins N., Petrov I. and Bashir R. 2009. Highly sensitive, mechanically stable nanopore sensors for DNA analysis. *Advanced Materials,* 21**,** 2771-2776.

Venken T., Claes S., Sluijs S., Paterson A. D., Van duijn C., Adolfsson R., Del-Favero J. and Van broeckhoven C. 2005. Genomewide scan for affective disorder susceptibility loci in families of a northern swedish isolated population. *American Journal of Human Genetics,* 76**,** 237-248.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A. and Holt R. A. 2001. The sequence of the human genome. *Science,* 291**,** 1304-1351.

Verma A. 2015. Empowering the neurogenetic testing services in developing countries: Use the basic skills with speed and scale. *Annals of Neurosciences,* 22**,** 1-3.

Vissers L. E., De Vries B. B. and Veltman J. A. 2010. Genomic microarrays in mental retardation: From copy number variation to gene, from research to diagnosis. *Journal of Medical Genetics,* 47**,** 289-297.

Vissers L. E. L. M., Gilissen C. and Veltman J. A. 2016. Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics,* 17**,** 9-18.

Vogenberg F. R., Isaacson Barash C. and Pursel M. 2010. Personalized medicine: Part 1: Evolution and development into theranostics. *Pharmacy and Therapeutics,* 35**,** 560-576.

Volz H.-P., Nenadic I., Gaser C., Rammsayer T., Häger F. and Sauer H. 2001. Time estimation in schizophrenia: An fmri study at adjusted levels of difficulty. *Neuroreport,* 12**,** 313-316.

Von Bubnoff A. 2008. Next-generation sequencing: The race is on. *Cell,* 132**,** 721-723.

Vorstman J. A. and Ophoff R. A. 2013. Genetic causes of developmental disorders. *Current Opinion in Neurology,* 26**,** 128-136.

Waber D. P., Bryce C. P., Girard J. M., Zichlin M., Fitzmaurice G. M. and Galler J. R. 2014. Impaired iq and academic skills in adults who experienced moderate to severe infantile malnutrition: A forty-year study. *Nutritional Neuroscience,* 17**,** 58-64.

Wagner G., Koch K., Schachtzabel C., Schultz C. C., Gaser C., Reichenbach J. R., Sauer H., Bär K.-J. and Schlösser R. G. 2013. Structural basis of the fronto-thalamic dysconnectivity in schizophrenia: A combined dcm-vbm study(). *NeuroImage : Clinical,* 3**,** 95-105.

Walker Matthew p. and Robertson Edwin m. 2016. Memory processing: Ripples in the resting brain. *Current Biology,* 26**,** R239-R241.

Walpole I. R., Hockey A. and Nicoll A. 1990. The nance-horan syndrome. *Journal of Medical Genetics,* 27**,** 632-634.

Walther S., Horn H., Razavi N., Koschorke P., Müller T. J. and Strik W. 2009. Quantitative motor activity differentiates schizophrenia subtypes. *Neuropsychobiology,* 60**,** 80-86.

Wang J. and Shen Y. 2014. When a "disease-causing mutation" is not a pathogenic variant. *Clinical Chemistry,* 60**,** 711-713.

Wang J. and Song Y. 2017. Single cell sequencing: A distinct new field. *Clinical and Translational Medicine,* 6**,** 10.

Wang K., Li M. and Hakonarson H. 2010. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research,* 38**,** e164-e164.

Wang W., Gawlik K., Lopez J., Wen C., Zhu J., Wu F., Shi W., Scheibler S., Cai H. and Vairavan R. 2016. Genetic and environmental factors strongly influence risk, severity and progression of age-related macular degeneration. *Signal Transduction and Targeted Therapy,* 1**,** 16016.

Wang Y.-C., Peterson S. E. and Loring J. F. 2014. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Research,* 24**,** 143-160.

Wang Y. and Navin N. E. 2015. Advances and applications of single-cell sequencing technologies. *Molecular Cell,* 58**,** 598-609.

Warr A., Robert C., Hume D., Archibald A., Deeb N. and Watson M. 2015. Exome sequencing: Current and future perspectives. *G3: Genes|Genomes|Genetics,* 5**,** 1543-1550.

Watson C. M., Crinnion L. A., Morgan J. E., Harrison S. M., Diggle C. P., Adlard J., Lindsay H. A., Camm N., Charlton R. and Sheridan E. 2014. Robust diagnostic genetic testing using solution capture enrichment and a novel variant-filtering interface. *Human Mutation,* 35**,** 434-441.

Watson J. D. and Crick F. H. 1953. Molecular structure of nucleic acids. *Nature,* 171**,** 737-738.

Webb J. and Whitaker S. 2012. Defining learning disability. *Psychologist,* 25**,** 440-443.

Weber T. K. 2016. *The role of genetic testing in surgical oncology, an issue of surgical oncology clinics of north america*, Elsevier Health Sciences.

Weigel M. T. and Dowsett M. 2010. Current and emerging biomarkers in breast cancer: Prognosis and prediction. *Endocrine-related Cancer,* 17**,** R245-R262.

Weir K. 2012. The roots of mental illness. *Monitor on Psychology,* 43**,** 30-33.

Weiss F. U., Simon P., Bogdanova N., Mayerle J., Dworniczak B., Horst J. and Lerch M. M. 2005. Complete cystic fibrosis transmembrane conductance regulator gene sequencing in patients with idiopathic chronic pancreatitis and controls. *Journal of the British Society of Gastroenterology,* 54**,** 1456-1460.

Wen W., He Y. and Sachdev P. 2011. Structural brain networks and neuropsychiatric disorders. *Current Opinion in Psychiatry,* 24**,** 219-225.

Wen Z., Chen J., Khan R. a. W., Song Z., Wang M., Li Z., Shen J., Li W. and Shi Y. 2016. Genetic association between nrg1 and schizophrenia, major depressive disorder,

bipolar disorder in han chinese population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 171**,** 468-478.

Whalley H. C., Papmeyer M., Romaniuk L., Sprooten E., Johnstone E. C., Hall J., Lawrie S. M., Evans K. L., Blumberg H. P. and Sussmann J. E. 2012. Impact of a microrna mir137 susceptibility variant on brain function in people at high genetic risk of schizophrenia or bipolar disorder. *Neuropsychopharmacology,* 37**,** 2720-2729.

Wheeler J. R., Clare I. C. and Holland A. J. 2013. Offending by people with intellectual disabilities in community settings: A preliminary examination of contextual factors. *Journal of Applied Research in Intellectual Disabilities,* 26**,** 370-383.

White J. K., Gerdin A.-K., Karp N. A., Ryder E., Buljan M., Bussell J. N., Salisbury J., Clare S., Ingham N. J. and Podrini C. 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell,* 154**,** 452-464.

Whiteford H. A., Degenhardt L., Rehm J., Baxter A. J., Ferrari A. J., Erskine H. E., Charlson F. J., Norman R. E., Flaxman A. D. and Johns N. 2013. Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *The Lancet,* 382**,** 1575-1586.

Willemsen R., Levenga J. and Oostra B. A. 2011. Cgg repeat in the fmr1 gene: Size matters. *Clinical Genetics,* 80**,** 214-225.

Williams R. E., Aberg L., Autti T., Goebel H. H., Kohlschütter A. and Lönnqvist T. 2006. Diagnosis of the neuronal ceroid lipofuscinoses: An update. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1762**,** 865-872.

Willner P. 2005. The effectiveness of psychotherapeutic interventions for people with learning disabilities: A critical overview. *Journal of Intellectual Disability Research,* 49**,** 73-85.

Winchester C. L., Ohzeki H., Vouyiouklis D. A., Thompson R., Penninger J. M., Yamagami K., Norrie J. D., Hunter R., Pratt J. A. and Morris B. J. 2012. Converging evidence that sequence variations in the novel candidate gene map2k7 (mkk7) are functionally associated with schizophrenia. *Human Molecular Genetics,* 21**,** 4910-4921.

Wingenfeld K. and Wolf O. T. 2011. Hpa axis alterations in mental disorders: Impact on memory and its relevance for therapeutic interventions. *CNS Neuroscience & therapeutics,* 17**,** 714-722.

Winkler E. C. and Wiemann S. 2016. Findings made in gene panel to whole genome sequencing: Data, knowledge, ethics–and consequences? *Expert Review of Molecular Diagnostics,* 16**,** 1259-1270.

Winnepenninckx B., Rooms L. and Kooy R. F. 2003. Mental retardation: A review of the genetic causes. *The British Journal of Development Disabilities,* 49**,** 29-44.

Wittchen H.-U., Mühlig S. and Beesdo K. 2003. Mental disorders in primary care. *Dialogues in Clinical Neuroscience,* 5**,** 115-128.

Wittenberg J. 2006. Evidence-based psychotherapies for children and adolescents. *Journal of the Canadian Academy of Child and Adolescent Psychiatry,* 15**,** 186-188.

Wolf S. M., Lawrenz F. P., Nelson C. A., Kahn J. P., Cho M. K., Clayton E. W., Fletcher J. G., Georgieff M. K., Hammerschmidt D., Hudson K., Illes J., Kapur V., Keane M. A., Koenig B. A., Leroy B. S., Mcfarland E. G., Paradise J., Parker L. S., Terry S. F., Van Ness B. and Wilfond B. S. 2008. Managing incidental findings in human

subjects research: Analysis and recommendations. *The Journal of Law, Medicine & Ethics : a Journal of the American Society of Law, Medicine & Ethics,* 36**,** 219-211.

Woods C. G., Cox J., Springell K., Hampshire D. J., Mohamed M. D., Mckibbin M., Stern R., Raymond F. L., Sandford R., Malik Sharif S., Karbani G., Ahmed M., Bond J., Clayton D. and Inglehearn C. F. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *The American Journal of Human Genetics,* 78**,** 889-896.

Wu S., Zhang R., Nie F., Wang X., Jiang C., Liu M., Valenzuela R. K., Liu W., Shi Y. and Ma J. 2016. Microrna-137 inhibits efnb2 expression affected by a genetic variant and is expressed aberrantly in peripheral blood of schizophrenia patients. *EBioMedicine,* 12**,** 133-142.

Wulff K., Gatti S., Wettstein J. G. and Foster R. G. 2010. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience,* 11**,** 589-599.

Xie N., Gong H., Suhl J. A., Chopra P., Wang T. and Warren S. T. 2016. Reactivation of fmr1 by crispr/cas9-mediated deletion of the expanded cgg-repeat of the fragile x chromosome. *PloS One,* 11**,** e0165499.

Xiong W., Grillet N., Elledge H. M., Wagner T. F., Zhao B., Johnson K. R., Kazmierczak P. and Muller U. 2012. Tmhs is an integral component of the mechanotransduction machinery of cochlear hair cells. *Cell,* 151**,** 1283-95.

Xu B., Roos J. L., Dexheimer P., Boone B., Plummer B., Levy S., Gogos J. A. and Karayiorgou M. 2011. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature Genetics,* 43**,** 864-868.

Xu G., Jiang X. and Jaffrey S. R. 2013. A mental retardation-linked nonsense mutation in cereblon is rescued by proteasome inhibition. *The Journal of Biological Chemistry,* 288**,** 29573-29585.

Xue Y., Chen Y., Ayub Q., Huang N., Ball Edward v., Mort M., Phillips Andrew d., Shaw K., Stenson Peter d., Cooper David n., Tyler-Smith C. and The Genomes Project C. 2012. Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing. *American Journal of Human Genetics,* 91**,** 1022-1032.

Yap C. C., Liang F., Yamazaki Y., Muto Y., Kishida H., Hayashida T., Hashikawa T. and Yano R. 2003. Cip98, a novel pdz domain protein, is expressed in the central nervous system and interacts with calmodulin-dependent serine kinase. *Journal of Neurochemistry,* 85**,** 123-134.

Zaghloul N. A. and Katsanis N. 2009. Mechanistic insights into bardet-biedl syndrome, a model ciliopathy. *The Journal of Clinical Investigation,* 119**,** 428-437.

Zarow C., Lyness S. A., Mortimer J. A. and Chui H. C. 2003. Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in alzheimer and parkinson diseases. *Archives of Neurology,* 60**,** 337-341.

Zarzar T. and Mcevoy J. 2013. Clozapine for self-injurious behavior in individuals with borderline personality disorder. *Therapeutic Advances in Psychopharmacology,* 3**,** 272-274.

Zeggini E., Weedon M. N., Lindgren C. M., Frayling T. M., Elliott K. S., Lango H., Timpson N. J., Perry J. R. B., Rayner N. W., Freathy R. M., Barrett J. C., Shields B., Morris A. P., Ellard S., Groves C. J., Harries L. W., Marchini J. L., Owen K. R., Knight B.,

Cardon L. R., Walker M., Hitman G. A., Morris A. D., Doney A. S. F., Mccarthy M. I. and Hattersley A. T. 2007. Replication of genome-wide association signals in uk samples reveals risk loci for type 2 diabetes. *Science,* 316**,** 1336-1341.

Zhang J.-P., Wei L.-C., Cao R. and Chen L.-W. 2006. Differential co-expression of ampa receptor subunits in substance p receptor-containing neurons of basal forebrain regions of c57/bl mice. *Neurochemistry International,* 49**,** 319-326.

Zhang J., Pan Z., Gui C., Zhu J. and Cui D. 2016a. Clinical investigation of speech signal features among patients with schizophrenia. *Shanghai Archives of Psychiatry,* 28**,** 95-102.

Zhang P., Bian Y., Liu N., Tang Y., Pan C., Hu Y. and Tang Z. 2016b. The snp rs1625579 in mir-137 gene and risk of schizophrenia in chinese population: A meta-analysis. *Comprehensive Psychiatry,* 67**,** 26-32.

Zhang R., Zhong N.-N., Liu X.-G., Yan H., Qiu C., Han Y., Wang W., Hou W.-K., Liu Y. and Gao C.-G. 2010. Is the efnb2 locus associated with schizophrenia? Single nucleotide polymorphisms and haplotypes analysis. *Psychiatry Research,* 180**,** 5-9.

Zhang S. and Sodroski J. 2015. Efficient human immunodeficiency virus (hiv-1) infection of cells lacking pdzd8. *Virology,* 481**,** 73-78.

Zhao B., Wu Z., Grillet N., Yan L., Xiong W., Harkins-Perry S. and Muller U. 2014. Tmie is an essential component of the mechanotransduction machinery of cochlear hair cells. *Neuron,* 84**,** 954-67.

Zhao M., Wang Q., Wang Q., Jia P. and Zhao Z. 2013. Computational tools for copy number variation (cnv) detection using next-generation sequencing data: Features and perspectives. *BMC Bioinformatics,* 14**,** S1.

Zhou H. and Clapham D. E. 2009. Mammalian magt1 and tusc3 are required for cellular magnesium uptake and vertebrate embryonic development. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 15750-5.

Zhu D., Alcorn D., Antonarakis S., Levin L., Huang P., Mitchell T., Warren A. C. and Maumenee I. H. 1990. Assignment of the nance-horan syndrome to the distal short arm of the x chromosome. *Human Genetics,* 86**,** 54-58.

Zhu M. and Zhao S. 2007. Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences,* 3**,** 420-427.

Zhu Q.-B., Unmehopa U., Bossers K., Hu Y.-T., Verwer R., Balesar R., Zhao J., Bao A.-M. and Swaab D. 2016. Microrna-132 and early growth response-1 in nucleus basalis of meynert during the course of alzheimer's disease. *Brain,* awv383.

Zou Z., Liu C., Che C. and Huang H. 2014. Clinical genetics of alzheimer's disease. *BioMed Research International,* 2014.

Zucco G. M., Herz R. S. and Schaal B. 2012. *Olfactory cognition: From perception and memory to environmental odours and neuroscience,* John Benjamins Publishing.

Appendices

Appendix 1- Ethical approval documents for the study and sample collection.

Sultan Qaboos University
COLLEGE OF MEDICINE & HEALTH SCIENCES

جامعة السلطان قابوس
كلية الطب
والعلوم الصحية

26th June 2013

To   :   Dr. Watfa Al Mamari
                    Department of Child Health

Subject: Research Proposal "Genetic Studies of Intellectual Developmental Disabilites
                (IDD) in affected Omani Families "(MREC#751)

Thank you for submitting the above mentioned proposal. It is my pleasure to inform you on behalf

of the Ethics Committee that your proposal has been approved, and you can now start your

research.

Wish you a productive study.

Best Regards,

P.P. **Prof. Badreldin Hamid Ali**
Assistant Dean for postgraduate Studies & Research

P.O. Box: 35
Postal Code 123
Al-Khod, Sultanate of Oman
Telephone: (+968) 24141103; Fax: (+968) 24413419

صندوق البريد:٣٥
الرمز البريدي:١٢٣
الخوض – سلطنة عمان
هاتف: ٢٤١٤١١٠٣ (٩٦٨+)، فاكس: ٢٤٤١٣٤١٩ (٩٦٨+)

July 14, 2013

To Whom It May Concern

It is to declare that we are conducting a research study approved by Sultan Qaboos University ethical committee by the title **Genetics Studies of Intellectual Developmental Disabilities in Affected Families in Oman**.

Mr. Ahmed Al Amri, PhD students Leeds University, is our external collaborator on this research.

This will necessitate shipping DNA samples of patients and families on regular bases to the laboratory where Mr. Ahmed is conducting his research in Leeds University.

Sincerely,

Dr. Fathiya Al Murshedi
Consultant, Clinical &Biochemical Genetics
Genetic & Developmental Medicine Clinic

Dr. Watfa Al Mamri
Consultant, Developmental Child Health
Genetic & Developmental Medicine Clinic

Sultan Qaboos University

COLLEGE OF MEDICINE & HEALTH SCIENCES

جامعة السلطان قابوس
كلية الطب
والعلوم الصحية

18th July 2013
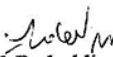
To   :   Dr. Hamad Al Sinawi
           Department of Behavior Medicine

Subject: **Research Proposal** "Genetic Basis of Schizophrenia in Oman "(MREC#760)

Thank you for submitting the above mentioned proposal. It is my pleasure to inform you on behalf of the Ethics Committee that your proposal has been approved, and you can now start your research.

Wish you a productive study.

Best Regards,

Prof. Badreldin Hamid Ali
Assistant Dean for postgraduate Studies & Research
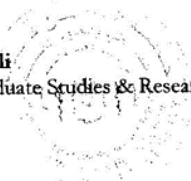
NRES Committee Yorkshire & The Humber - Leeds Central

10 October 2012

Dr Alastair G Cardno
Senior Lecturer in Psychiatry
University of Leeds
Charles Thackrah Building
101 Clarendon Road
Leeds
LS2 9LJ

Dear Dr Cardno

Study title:	Genetic Investigation of Schizophrenia in the Pakistani Population of West Yorkshire
REC reference:	10/H1313/37
Amendment number:	AM2
Amendment date:	19 September 2012

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Protocol	1.2	19 September 2012
Notice of Substantial Amendment (non-CTIMPs)	AM2	19 September 2012
Covering Letter		26 September 2012

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

322

UNIVERSITY OF LEEDS

Professor Chris Inglehearn
Section of Ophthalmology and Neuroscience
Leeds Institute of Molecular Medicine
Wellcome Trust Brenner Building
St James's University Hospital
Leeds LS9 7TF
UK
T +44 (0) 113 343 8646
F +44 (0) 113 343 8603
c.inglehearn@leeds.ac.uk
14th January 2015

To whom it may concern

Re: Request for extended UK10K access for the Leeds Psychiatric Genetics Group. CFI 14.1.15

The Leeds Psychiatric Genetics Group uses homozygosity mapping and whole exome sequencing in nuclear families to identify genes involved in a variety of human inherited brain disorders. Initially we ascertained and sampled nuclear families and consanguineous cases with schizophrenia, schizoaffective or bipolar disorder, from the local Pakistani community which has a high level of first cousin marriage, to test the hypothesis that recessive alleles of large effect should be more prevalent in such families. This was the basis of our application to the UK10K data, and for that project we were granted access to the following datasets:

As controls
UK10K_COHORTS_TWINSUK
UK10K_COHORT_ALSPAC
As cases
UK10K_NEURO_MUIR
UK10K_NEURO_FSZ
UK10K_NEURO_FSZNK
UK10K_NEURO_IOP_COLLIER
UK10K_NEURO_UKSCZ
UK10K_NEURO_ABERDEEN
UK10K_NEURO_EDINBURGH
UK10K_NEURO_GURLING

More recently we have initiated a new collaborative project with Mr Ahmed Al-Amri, Dr Watfa AL-Mamari and Dr Abeer Al-Sayegh in the Child Health Department of Sultan Qaboos University Hospital, Oman, to look at consanguineous families and cases with autistic spectrum disorder, intellectual disability and/or psychosis. Sampling for this project is carried out with local ethical approval in Oman and samples are analysed in Leeds by Mr Al-Amri, who is studying for his PhD with Dr Manir Ali, Dr Steve Clapcote and myself (all named on the original UK10K application) as supervisors. Ahmed's current working thesis title is "The genetic basis of ID and psychosis in Omani families".

The Leeds Teaching Hospitals **NHS**
NHS Trust

Using the same approach – homozygosity mapping and WES – in these families, we have already identified a mutation in the *TUSC3* gene, mutations in which are known to cause syndromic intellectual disability. In addition we have identified several further potential novel candidate ID/ASD genes. It would very much help us in seeking to prove the involvement of the new candidates if we could look for similar changes in the UK10K cohorts of patients with intellectual disability and ASD, both for rare alleles of large effect and for evidence of association with more common alleles – a similar approach to that we have used in the initial schizophrenia project for which two papers are now being drafted.

We therefore request additional access to the following UK10K datasets to support this new project:

ASD_SKUSE
ASD_TAMPERE
ASD_BIONED
ASD_MGAS
ASD_FI
IMGSAC
ASD_GALLAGHER

In addition to myself, our original application listed the following as co-applicants, and we would very much appreciate the additional access requested for the same team of researchers:

Name with title:	Dr Jose Ivorra Martinez
Position:	Post-doctoral fellow
Institutional email address:	J.L.IvorraMartinez@leeds.ac.uk

Name with title:	Dr Manir Ali
Position:	Group Leader
Institutional email address:	m.ali@leeds.ac.uk

Name with title:	Dr Steve Clapcote
Position:	Lecturer in Pharmacology
Institutional email address:	S.J.Clapcote@leeds.ac.uk

Name with title:	Dr Alastair Cardno
Position:	Senior lecturer
Institutional email address:	A.G.Cardno@leeds.ac.uk

Name with title:	Mr Ahmed Al-Amri
Position:	PhD student
Institutional email address:	bsahha@leeds.ac.uk

Yours faithfully

Professor Chris Inglehearn
Professor of Molecular Ophthalmology
Section Head, Section of Ophthalmology and Neuroscience

Appendix 2- Details of the primers used in this study

F= Family, temp= temperature, S= standard, HS= hot-shot master mix, PCR= polymerase chain reaction, RT= reverse transcriptase, Q= quantitative, CNV= copy number variation, CH= compound heterozygous, BLG= Belgian sample and SDM= site directed mutagenesis.

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (ºC)	Product size (BP)	Remarks
OR4L1	TTTTTGGACGATGGGAACTT	CAGCTTGTGGCTCATGATTG	57	371	F4, S-PCR
HSH2D	CAAAGCCAGTCCTGTGTCAC	GGCTTTTGAGAGATGTGGCC	57	349	F4, S-PCR
TUSC3	GAAATTCAAACAAAAGGACTTTGA	ACATGCCCCCAATAAAAGAA	56	424	F1, S-PCR
ALOX5AP	CACACCTCCTCTCATCTCCC	TAGATCCTGGGGCCACTCTA	57	391	F1, S-PCR
LTBP4	ATGGTCCCGAGATTTGTGGA	CCACAACCCGAAAATTCCCA	60	419	F1, S-PCR
CREB3L1	CAGATGGCTGGGAAATCAAC	GGAGAGTTTGATGGTGGTGTT	56	395	F6, S-PCR
DIXDC1	ACCTGAACCTGTGAACGCTA	ACAAGGGCAAACTACAAGGG	56	269	F6, S-PCR
LHFPL5	AGCCACAGGCCTAATGATTG	TGATTTCAGGGAGGACAAGG	57	392	F2, S-PCR
DEFB126	AAAGAATGGTTGGGCAATGT	AAGCCTCTTTGCTTTAATGAGTCA	57	260	F2, S-PCR
SIGIRR	AACTTCCTGGGCCTCACAG	ATTTTCTGGCACTGGGAGGC	59	442	F5, S-PCR
SLC35C1	TTCATGTGGAGTGCTTTGC	CAGCTGGGCAAAGTCACG	60	433	F5, S-PCR
CLCA4	GCCAACTCCAAGGAAAGCTT	TGAGTTTATGATCACACATGGGA	58	422	F5, S-PCR
DLX6	AAACCCGGGAGAAGGCTTT	TAGGACTGGAGGTAAGGGCT	57	432	F5, S-PCR
MUC17	ACTCAGGTGGCCACTTCTAC	GCCTCAGGACTGGTCACC	57	320	F5, S-PCR
MMP12	GCACAGAACAGCCCAGACTA	GCCACGTATGTCATCAGCAG	57	330	F5, S-PCR
CELA1	TGTGATTTTCCCTGCCTTGC	TCCAGGAGAGTAAAGGACGC	59	325	F5, S-PCR
RBPJ	ACCTACACACCAGAACCAGG	CCAACCAACCAACAAAACAGC	57	447	F3, S-PCR

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (ºC)	Product size (BP)	Remarks
PODXL	AGCGGCAGGGAGGAAGAG	GTGGATGGTGCAAGGTCAG	57	431	F3, S-PCR
AKR1B10	GCGATTGTACCTGAAGCCAC	TCCCTGGGCTGGAAAATGAT	57	335	F3, S-PCR
TRPV6	CTTCCTCTGGTTCTGCAGGA	CACTGGCCTAAAGTCCCTGA	60	366	F3, S-PCR
TRPV5	GGACAGGGTAAGGAGTGGAG	GGAATAGGAAGCAAGGGACC	57	438	F3, S-PCR
OR2A14	CCAGGAAAGCACCATCTCCT	CAAGTAGGAGACCAGCACCA	57	450	F3, S-PCR
CYP2C9	ACCCCTGAATTGCTACAACA	CCCGGTGATGGTAGAGGTTT	57	345	F3, S-PCR
ANKRD2	CCCATCCCCTAGTGTGACAA	TGCTCCTGGATCATGCAGAA	57	430	F3, S-PCR
PDZD8	TAAGGACAGTTCGGACGACC	ACATGGTGGTCTGATTCTCCT	57	408	F3, S-PCR
ALOX5AP	CACACCTCCTCTCATCTCCC	CTGGGAGCCTCGGTTCTTAA	62	354	F3, S-PCR
KBTBD6	ATCTCCACCATGCAGTCCC	TCACTGAGAGACACACGACC	60	422	F3, S-PCR
MYH8	GTGTGTGTGTAAAGCCTGCA	AGCTTCAGTAACATGGTGCA	58	448	F3, S-PCR
CENPV	CCCTGGGCCTGAAGACTG	GAACGTCTCCCAGCGCTC	58	447	F3, HS-PCR
MPRIP	CTGACATTGTGTAGCTGGCC	CATTCCCGTGCATACAAGCC	57	384	F3, S-PCR
KCNJ18	GAGGAGGGCGAGTACATCC	TGTGGAAGTGCGAGTAGTCA	55	325	F3, S-PCR
MUC16	GCAAGGCAGCCCAACATATT	ACAAACCTGCACATGTTCCC	58	424	F5, S-PCR
ANKLE1	CTCCTTGTCTTCCTGGCTGA	GCTTCCCTTCTTGTCCTCCT	57	344	F5, S-PCR
FAM47C	CCCACGGAGCCTGGTAAATA	CCAGGCAGAGATCAGGCAC	59	403	F5, S-PCR
DGKK	ACATCCTTTCCAGTCCTGCT	GGCACACAGACCTCCTTTTG	57	360	F5, S-PCR
IL2RG	GTCTCAAACTCCTGGCCTCT	TGGTCTCTGATCCAACCCAC	57	355	F5, S-PCR
FAM171B	CTTCCTCCTGCCGGTGAG	GAGGCCCTAAGAGATCGCC	59	439	F6, S-PCR
OR4C45	GCACAGACACTCACACCTTG	AGGTGTAAATAGGGTGCAAAACA	57	273	F6, S-PCR
NR1H2	GATGAGGGAGCAGTGTGAGT	CCGTCACACTCAAGGTTGTG	55	404	F6, S-PCR
ZNF654	TGTTCAGACAGACTCAAATCCT	TTTGGTTTGGAAGGTGCCTG	57	432	F4, S-PCR
MYH6	CCCTCTTTGTCTCTCGCTCT	ATTCCCCACCTCAGTTCTGG	58	381	F4, S-PCR
NGDN	GAACTTGGGAAGGGATGCTG	GGAGCTGTGGGACCTTAGAG	57	349	F4, S-PCR
FLJ22148	CTCCGAGTTCTGGGAGCTG	GACTCTGTGACGCTAACCCT	57	417	F4, S-PCR

326

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (°C)	Product size (BP)	Remarks
PS3	ACTTGCCTCTGGAACACCTT	TTATCAAATGTCCGTCGGCA	57	352	F2, S-PCR
TTC30B	GGATGCCAGGCCTACAGAA	ATCCCTGCTAGCCCATCAAG	59	420	F2, S-PCR
MSX2	TTACCACATCCCAGCTCCTC	TGGGAAGCACAGGTCTATGG	58	393	F2, S-PCR
KRT6A	TCTTAGGCTCTCTGTGGCAG	TGCCAATTCTCCTCTCCCAG	59	294	F2, S-PCR
LRCOL1	ACATTCCCTTTCTCAGCCCC	GCACTGCAGGAAGATGGTCT	56	341	F2, S-PCR
RFPL2	GGATATGGGCATGCAGAACG	GAGCGTTCCTAAGTCTACCCA	57	276	F2, S-PCR
NHS	TGAGTGAGATGTTTGCCCCA	CAGGATCTCAGTGGCAGACA	58	310	F2, S-PCR
CLCN5	CTTCTTAGCCTTGTTGACTTCCT	CGCTTGGCTTCATTCTTCCT	57	259	F2, S-PCR
PSMD10	CTCCTTGCTAACCCTTCCTAAA	CCGCCTTCCAGTAACATGAC	60	356	F2, S-PCR
MAMLD1	GAATTTTCACCTCCAGCCCC	CTTGGCTGAGGGGTGTTTTC	60	288	F2, S-PCR
MAGEC1	GATTTCTCAGAGCCCTCCTGA	GTGGAGGAGAAGGAGCGG	61	288	F2, S-PCR
GPR151	TGTGTGATTGGCATCCTCCT	TCCACACCTTCATGATGCCT	57	389	F2, S-PCR
ACIN1	ATTCATCCTCTAGCCGGTCC	GGGGCTGAAATGAAAACCACT	58	306	F4, S-PCR
HOMEZ	CACCAACAGCTACGGGAAAC	ACCCACCCTGAAGAACACAA	58	370	F4, S-PCR
FSIP2	GCCAGTGCCATTTTGAAGCT	GGCCTGTGACTTCCCTTACT	57	445	F6, S-PCR
ASB8	ATTATGGCGCAGAGGTCAGA	AGGAGTACTGCAGGGACAAC	57	417	F6, S-PCR
WNT1	CTCCTCCACGAACCTGCTTA	AACGCCTGGTAGAAGTCTGG	59	410	F6, S-PCR
BIN2	TGAGAGGCTGGAAAGGTTGA	TAGGAAAATGGGCACGGAGA	60	318	F6, S-PCR
CELA1	AGTGATGGATGACAAGGGGT	ATAAAGGACCAGGGTTGCGA	62	379	F5, HS-PCR
EFHC2	GTGTGAAGTCCCTTTGCTGG	CCTTGAGGGGAAAACAGGTG	59	276	F5, S-PCR
IQSEC2	CCAGTCACCATTGTCACCAC	ACATGGGGAGGAGGACATTG	61	333	F5, S-PCR
UPRT	GTTACAGTGTCCGGACTCCA	ATGATGGTCTGTAGCTCCCG	62	366	F5, S-PCR
GPR98	AGCAGCATGAGAATGGACCA	CCGATGAGATTCTGGGTGGA-	57	301	F5, S-PCR
FAM81B	GATTCTAGTTCTGGTTCCTGGC	TTTGACGATGCTGGTGATGG	57	300	F5, S-PCR
HLA-DRB5	TCGGGGTTATGGTTGGGATC	TTCCAGTACTCAGCGTCAGG	62	303	F5, S-PCR
TOLLIP	GACTGGTGTTTGCTGAGACC	GATGGCGGCATCCTTGTTC	57	322	F5, S-PCR

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (°C)	Product size (BP)	Remarks
SUPT20HL1	AGGGCATTGCTGGACTACTT	TACAGAGGGCCACGAATACC	57	414	F5, S-PCR
GRIA3	CTTCCCCTCACCTTGTCACC	AACTCTGGGCTAAGGGGTTC	57	364	F2, S-PCR
KCNK17	CTCTCCCGCTATGTACCGAC	TTCCCGACGTTTCCCTGAA	60	376	F2, S-PCR
MRPS10	GCACAGGTTCATGTCTCTATGG	ACTAGCCCCTTTCTCGTCAC	62	331	F2, S-PCR
TMEM132D	TTTTCCTTGCTGCCCACCTA	CCGGCTCAGGTAGACTTTGT	62	309	F2, S-PCR
MMP17	GCCCATGTGAGCGCCATG	TGCACTGCTGGGGTCTTC	62	270	F2, S-PCR
ANKRD2_1	CAGCCAGCAGTTCCCTTTTG	GATCTCCCTCCCCTTGACAG	60	494	BLG, S-PCR
ANKRD2_2	GATGTCTCTGTGGGGTGTCC	AGATTCTGCAGGGCTAGCAA	60	327	BLG, S-PCR
ANKRD2_3	GCTTAGCTCAGAGGTCTCCC	GGTCTCCTCATCCACAGGG	60	356	BLG, S-PCR
ANKRD2_4	TCCAGAACCTCATCGAGCTG	GGTCAATTCCTGTACGGGGA	60	476	BLG, S-PCR
ANKRD2_5	TAGAGCATAAGGGAGCTGGC	GGAGGAGAAGGGGCTTACTC	60	417	BLG, S-PCR
ANKRD2_6	ACTTTCTTCCCTGCAGCTCT	TGCAGCCAGAAACCAGACTA	57	269	BLG, S-PCR
ANKRD2_7	TGTACCATTCTCTGCCCCTC	CACTCACTATGGCCCTTGTG	60	421	BLG, S-PCR
ANKRD2_8	CCTGGGCCTGGAAATCAATG	ATAGAGAATGGGGCGGACAG	57	500	BLG, S-PCR
ANKRD2_9	CTGGTGGTGTCAATGAGGGT	AGTTTCCCTCCTCCCCAAAC	60	492	BLG, S-PCR
PDZD8_1	AAGGGCCGCAGAGAAGTAAA	GGAAGAACTGGGCGAGGAG	57	353	BLG, S-PCR
PDZD8_2	GCCCGATGTCCTCAAAATGG	ATGGTGGCGTTGAGGAAGTA	57	402	BLG, S-PCR
PDZD8_3	GACGCGGGAGACTTGCTA	CCCGTTGTACTCCACCTCC	57	298	BLG, S-PCR
PDZD8_4	GCCCTTCATCAAGACCATCC	TCCCCGACCTCCAGCTCA	57	450	BLG, S-PCR
PDZD8_5	TGTAGAGCACGTGTAGTATTAGG	TTTTCACTGCACTGTTCTTCCC	57	472	BLG, S-PCR
PDZD8_6	ACAGGTATGATACTCTCCCTTGT	TGAAAGGAAAGCTAGTCGGACT	57	327	BLG, S-PCR
PDZD8_7	ATCAGCCATGGTGCTCTCTG	GAAGGTGGGAAGGGAAGGAT	59	400	BLG, S-PCR
PDZD8_8	TCAGCACATTCAGGGTTTGA	ATCTACTGTCAACCCGGCAG	57	411	BLG, S-PCR
PDZD8_9	GGCCAGAGTAATCAAGGTGC	TGGCACAGGTGGTTTGAAAG	57	388	BLG, S-PCR
PDZD8_10	AGCTGTAGGAAGTCACCCAC	GGTCGTCCGAACTGTCCTTA	57	290	BLG, S-PCR
PDZD8_11	GTGCCACCTCCTCTTGTAGA	CCGTGTTTGATGTAGCTAGGC	57	502	BLG, S-PCR

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (°C)	Product size (BP)	Remarks
PDZD8_12	TGGTGCAGGGATCCTTTCAA	CTGTTAAACCCATCTGTCCAACA	60	320	BLG, S-PCR
PDZD8_13	AGGAGAATCAGACCACCATGT	TCCTGTCCTTCCAGCCTAAG	55	346	BLG, S-PCR
PDZD8_14	TGGAGCAACTGATAGGCGAA	CCCTGTACAGATCACGACCA	57	473	BLG, S-PCR
PDZD8_15	CAGACAACGAAGGCAGTGAC	TGCCTGCTCTGTAGTGAATCA	57	426	BLG, S-PCR
PDZD8_16	TCACTTCTTTCTGCTGCCTT	AGTAACTGGAGAAGCAGGCC	57	279	BLG, S-PCR
PDZD8_17	ATCCTCGTATGCACTCTGGC	TTGTCAGTTTGGGCATGCAG	59	354	BLG, S-PCR
A2ML	GGAGACATTATGGGCACAGC	TACCCTGGGAAGCCAAATGT	58	348	F4, S-PCR, CH
BZRAP	GAATGGGGAGCTGCGACA	GCATGGGCCAAGTTGCTATT	60	255	F4, S-PCR, CH
CACNA1B	GAACTGCTTCACCGTCAACC	CGCACTCACGGAAATGCTC	58	270	F4, S-PCR, CH
HSPB7	AACCTCTTCCACCTTCCGAG	ACATCTGTCTTGGTGGCCA	58	334	F4, S-PCR, CH
SLC26A9	ACACACACACACACACACAC	GGAGAGAGGAACCCAAGCTC	57	319	F4, S-PCR, CH
ZAN	AGGCTGGTCTCAAACTCCTG	GGATCCATCTTGTTGCGTCC	60	281	F4, S-PCR, CH
NLRP5	CAACGTGGAATGTCTGGCAC	CCCATTTCTCCCTCCTCCAT	62	336	F6, S-PCR, CH
OR4X1	AATTGGAGTGAGCAGAGGGT	ATTCGCTGGTTCATGATGGC	58	362	F6, S-PCR, CH
PABPC3	CGTGGTTTGTGATGAAAATGGT	CTTTCACACTTAAGGCGGGC	60	278	F6, S-PCR, CH
YTHDC2	CAGCATTGATTATCCCCATCCA	TTGACCAATCCTTTCCCGTC	61	474	F6, S-PCR, CH
FUCA1	TCCCTTCTTGAGCACCAGAG	GATGAACACCCCGAACTTGG	59	360	F5, S-PCR, CH
KIAA0040	TCAGTGCCTTCTTCAGCTCT	TGGTCAGAAGGAGGCTTAGC	57	343	F5, S-PCR, CH
KRT4	AGAAGCCTCTACAACCTCAGG	ACAGCTCACCTTGTCGATGA	57	324	F5, S-PCR, CH
MROH8	ACTCCCACCAGCTCGCAG	TTAGAACTCATCCTGGCCCG	55	320	F5, S-PCR, CH
MYO18B	CCACTGCTTTGCCACCTAAG	GGGAGGGACAGCAATTCTCT	57	274	F5, S-PCR, CH
MYO5C	ACACGCAACGCTGTAATTGT	GTTTCCTTAGCAGCGACACA	57	259	F5, S-PCR, CH
PABPC3	GACGGAACTTAAGCGCACAT	GTGGCCACAATTCTACCGTT	57	261	F5, S-PCR, CH
SGSM2	TCTCATCCCACTGTGAGCTG	GTAGTGGCCAAGCAGAAAGG	60	274	F5, S-PCR, CH
SPATA13	GGGGAGCTGACCAAAATCAC	CCTCGTCTGTGGTCCTACTG	60	235	F5, S-PCR, CH
SYNE1	CCCGGCCCTAAACCAATTTT	GAAAAGTGCAGTCCTGAGCC	60	394	F5, S-PCR, CH

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (°C)	Product size (BP)	Remarks
SLC26A9	CCTCTCCGAGTCCTTCATCC	TGAAGGAGGAGGGATGGAGA	59	425	F5, S-PCR, CH
CACNA1B-a	GGGTCCTCTACAAGCAATCG	CGCACTCACGGAAATGCTC	57	343	F5, S-PCR, CH
CACNA1B-b	CAGTTGTCTCAGCCCCTCC	TGTTCTCAGGCACAGACCAT	57	305	F5, S-PCR, CH
CACNA1B-c	TGGAGAATCTGCAGGGTGTT	AGCTTGCCTACCCTGTGAG	60	309	F5, S-PCR, CH
ZAN-a	GATTACAGGCACGCACCAC	GGATCCATCTTGTTGCGTCC	58	355	F5, S-PCR, CH
ZAN-b	CCTCACATCCCTTTCTCCCA	GGGAAGGAGAACAGGCGAA	59	342	F5, S-PCR, CH
MROH8	GAGCGGCGACTAGTAACCTC	AACTGCAAGGCCCTGTGA	57	299	F5, S-PCR, CH
NBPF4	TCTCACTTGTGGATGGTGCT	GCATTGCTCATACTCTGCCC	58	568	F4, S-PCR, CNV
DMD	GCCTAAGCCTTGGTTTAGCG	AAAGTAGCCTCCCCTCAACC	58	598	F4, S-PCR, CNV
C3ORF33	CGTAGAAAAGGGTGAAACGGT	AGAATCGCTTGAACCCAGGA	58	461	F6, S-PCR, CNV
ARNT2	TCTGTAGGTCTCCAGCTCCT	GGTTCCCTTGTCTGCCTTTG	55	600	F5, S-PCR, CNV
TCF4	GCCATTGCCATTTCCAAAAGG	GAGATGGGGAGAAGGCCATT	62	440	F5, S-PCR, CNV
LAMA2	CTTGCCAGCTCCACTTTCTC	CAGTAACAACCCGATTAACTGGA	58	410	F5, S-PCR, CNV
EFCAB12	CAGAGGCTTAGGGAGGGATG	CCCACACTGAAGGGATAGCA	58	287	F4, S-PCR, CNV
SPINK14	AATGGAACGGTCAACCCCT	AGGGCTCAGATACCAGAACG	60	333	F4, S-PCR, CNV
LCE3C	CTGCAAAGATGTCCTGCCAG	TCATCAGGATCCAGGTCAGC	60	308	F4, S-PCR, CNV
UGT2B17	TCCAATAGCTCCTGACTTTCTCA	TCATTTAGTAGCCATCCACAGTT	55	277	F4, S-PCR, CNV
SIRPB1	AGGTAGCCTCAGATGGGGTA	GGTTCCCTGAGACTCCTGAG	55	536	F4, S-PCR, CNV
GSTT1	AGAATCACTTGAACCCGGGA	AAGAGACTGGTGCCTGAACA	57	389	F5, S-PCR, CNV
PCDHA7	CTGACTGGCACCGTTCAATT	TGAGAGTCAACTGTGGAGCA	57	394	F6, S-PCR, CNV
DFNB31-1	GCCACTGTACCCGTGCTC	CTGCCTGTAGGGGGTGGT	57	517	F7, S-PCR, SDM
DFNB31-2	TGCTCTTCGACCAATACACG	CCTTTTTCTCATCCCCTCCT	57	528	F7, S-PCR, SDM
DFNB31-3	AGGAGGGGATGAGAAAAAGG	GGGTGTTGAGCAGCTTGAA	57	590	F7, S-PCR, SDM
DFNB31-4	CCCTGTTCAAGCTGCTCAA	AGGCATAGATGGGGGAAGAG	57	573	F7, S-PCR, SDM
DFNB31-5	TCTTCTCGGCTCCACAGAAC-	GCAGCTCCTTGCTACTCCTG	60	522	F7, S-PCR, SDM
DFNB31-6	CCACCAAGAGCAGGAGTAGC	CCCAGTACAGCACCAGTTCC	57	592	F7, S-PCR, SDM

Oligo name	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temp (°C)	Product size (BP)	Remarks
ANKRD2	TCCAGAACCTCATCGAGCTG	CCACAGTGGCCCCATTATCT	61	288	F3, RT-PCR
PDZD8	CAAGCACACCCTACCGAATT	CGAGTTTGGAGCCACAGTTT	61	367	F3, RT-PCR
TP53	GTACTCCCCTGCCCTCAACA	CTGGAGTCTTCCAGTGTGA	61	600	F3, RT-PCR
CG10632	GACCTTCTGCTGGACATCAAC	CAGTGCGTGTAGGGCTTTCT	60	152	F3, Q-PCR
eEF-1a	GTCTGGAGGCAATGTGCTTT	AATATGATGTCGCCCTGGTT	60	97	F3, Q-PCR
KCTD12	GAGCGCACTGGAACTCAAG	TGTAGCGGAAGAGGAAGCC	57	500	F8, RT-PCR
ACOD1	CCATGGCTTGAAAGTGGGAC	GTAATCGGCCTTGCACTTCA	60	600	F8, RT-PCR
CLN5	GAGAGGCTCCGGAAGTACTG	CCATTGTGTAGTTCTTGCCAGT	60	431	F8, RT-PCR
FBXL3	AAAGCTTCTTTCTGGGCAGC	TCTCCACAAGTCAGGCATGT	60	450	F8, RT-PCR
MYCBP2	GTCGTTGGAAAAGGGCTGTG	TCAGACTTGCTTCTTGTCTTCAC	60	472	F8, RT-PCR
SCEL	TACTGGAAGGCAGCATGTCC	GTGTTGGAGGTGTTGGCATT	60	451	F8, RT-PCR
SLAIN1	GCCCTTCGTCTACTTCAAGC	CTTGCAGACGAGCCATGATC	60	476	F8, RT-PCR
EDNRB	GCTGAACACTGGGAAGGAAC	AACACAAGGCAGGACACAAC	60	437	F8, RT-PCR
POU4F1	TCCAGCTCCGAGGCCATC	ATATGCGGGTGAGGGTGC	60	500	F8, RT-PCR
RNF219	GAAGAGTAGCGGTAGGTCGG	CCCTGCACCAAGGTTAAAGG	60	463	F8, RT-PCR
RBM26	ACCACCCATGCCAAAGAAAC	GGAGGAGGCTGGTGATGAAT	60	393	F8, RT-PCR
NDFIP2	CAGGAACGGAGGCGGAAG	AGCCTTCTCAGCTTCATCGT	60	409	F8, RT-PCR
SPRY2	GTGTTCATCAGCGGGGAATC	GTGTTTGTGCTGAGTGGAGG	60	471	F8, RT-PCR

Appendix 3- Genotyping microsatellites.

The genotyping markers used in this thesis were D1S478, D4S3042 and D7S483. The following is a representative example of the analysis report obtained from Gene-Mapper software version 3.7. This reading was generated for sample IV.6 of family-3 when tested with D4S3042 marker. The image shows the double alleles 220/220.
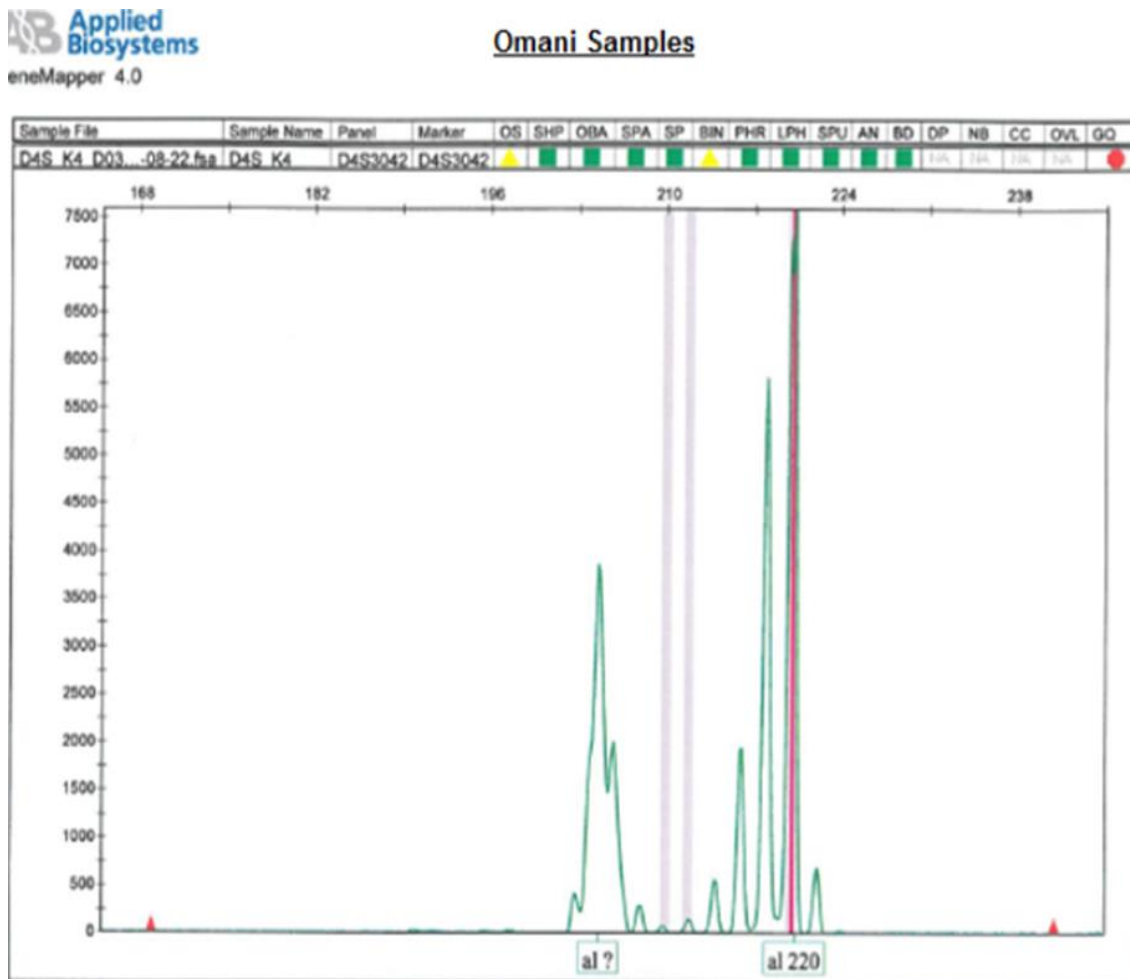
Appendix 4- QC measures of NGS

Various QC check was performed during the library preparation for the NGS samples. This is one example of the quality check and it was done after DNA shearing step, at which 1µl of the sample was loaded on a DNA 1000 chip and run on a 2100 Bioanalyzer. This report was generated by Agilent 2100 Expert Software, and it shows the expected distribution curves for two samples, with a peak size between 150 to 200 nucleotides in their electropherograms.

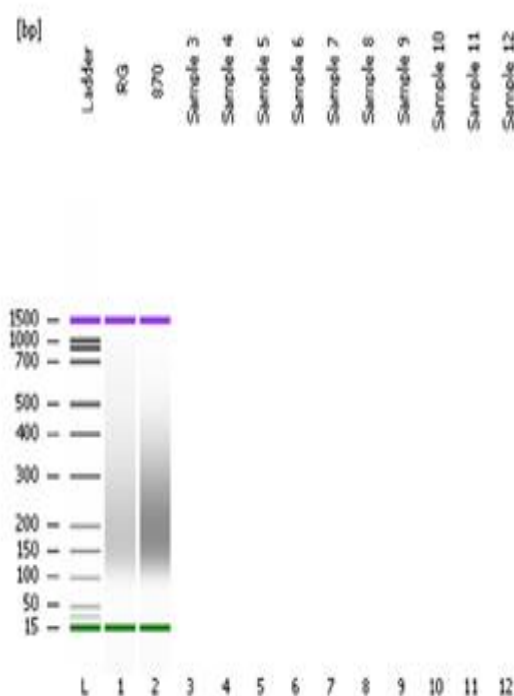2100 expert_DNA 1000_DE72904747_2013-09-12_17-06-40.xad
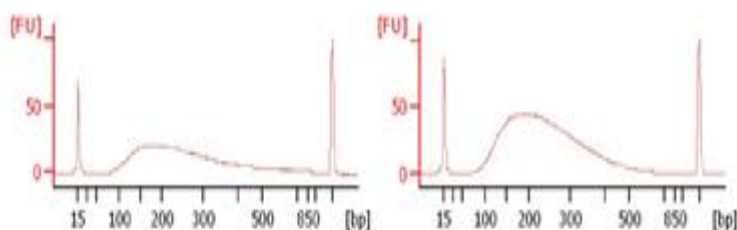
The quality check charts are examples of the quality measures that were investigated after getting the results of exome sequencing and before doing the analysis process. The scores which are shown here belongs to the case V.9 in Family-1, at which the data was processed using FASTQC on the Galaxy platform. A. whisker-type plot showing a general overview of bases positions. Green area is for the very good quality calls, orange area is for the acceptable calls and red area is for the unacceptable poor quality calls. B. plot of the average quality per read. C. Plot of the average sequence content for the four DNA bases (A, T, G and C). D. Plot of the GC content distribution compared to a reference. This sample showed a very good quality calls as its reads were located in the green area. The sample has 72,326,188 reads and none of them was flagged as poor quality. The average quality per read is 39 and the distribution of GC content is slightly different than the reference distribution but still within the expected range.
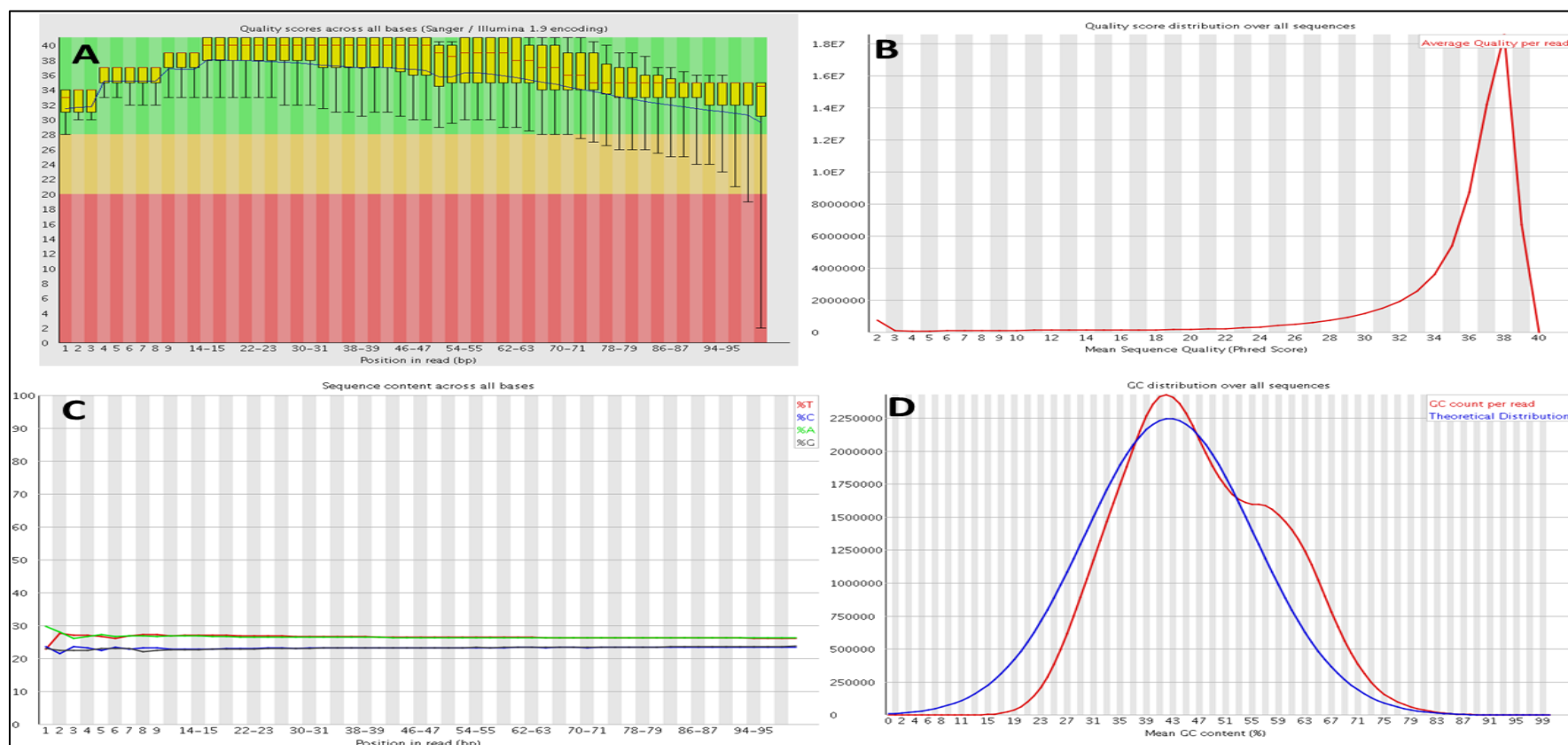
Appendix 5: UNIX console commands

A) Whole exome analysis

1.1 Alignment (Novoalign/Bowtie2)

```
$ novoalign -c 12 -d <path> b37/human_g1k_v37.nix -f sample_R1_001.fastq.gz
sample_R2_001.fastq .gz -o SAM $'@RG\tID:sample_novoID\tSM:sample\tPL: ILLUMINA\tLB:
sample_ exome' -k –K mismatches_sample_novoID.txt 2> novostats_sample_novoID.txt >
sample_novoID.sam
```

```
$ <path> /bowtie2- version no /bowtie2 -x <path> /ucsc.hg19.idx -p 6 -q -1 <path>
Sample_R1.fastq -2 <path> Sample_R2.fastq -S <path> Sample.sam --sam-rg ID: IN--sam-rg
SM:IN --sam-rg PL:Illumina --sam-rg PU:HiSeq
```

1.2 Remove reads which don't map uniquely
```
$ sed '/XS:/d' <path> Sample.sam >  <path> Sample_uniqueAlignment.sam
```

1.3 Sorting and idexing alignment sam file (Samtools or Picard)

```
$ <path> /samtools-version no/samtools view -bt <path> /DataFiles/ucsc.hg19.fasta.fai <path>
Sample_ uniqueAlignment.sam > Sample_uniqueAlignment.bam
```

```
$ <path> /samtools-version no/samtools sort <path> Sample_uniqueAlignment.bam <path>
Sample_ uniqueAlignment_sort
```

```
$ <path> /samtools-version no/samtools index <path> Sample_uniqueAlignment_sort.bam
```

```
$ java –Xmx8g -jar <path> /picard/picard-tools- version no/SortSam.jar I=
Sample_uniqueAlignment .sam O= Sample_uniqueAlignment_sort.bam SO=coordinate
CREATE_INDEX=TRUE
```

1.4 Remove duplicates (Picard)
```
$ java –Xmx8g -jar <path> /picard/picard-tools- version no/MarkDuplicates.jar I= <path>
Sample_unique Alignment_sort.bam O= <path>  Sample_uniqueAlignment_sort.rmdups.bam
M=Sample. rmdups. metrics CREATE_INDEX=TRUE
```

Alternative

```
$ java -Xmx8g -jar <path> /picard/picard-tools- version no/MarkDuplicates.jar INPUT= <path>
Sample_              uniqueAlignment_sort.bam              REMOVE_DUPLICATES=true
VALIDATION_STRINGENCY            =LENIENT           AS=true            METRICS_FILE=
```

Sample_uniqueAlignment_sort_metrics_file.dups   OUTPUT=   <path>   Sample
_uniqueAlignment_sort.rmdups.bam

1.5 Create indel realigner targets (GATK)

$ java -Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T Realigner
TargetCreator -R <path> /human_g1k_ v37.fasta –known <path> /b37/1000G_phase1.
indels.b37.vcf -known <path> /Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I <path>
Sample_uniqueAlignment _sort.rmdups.bam -o <path> Sample_uniqueAlignment_sort.rmdups.
indelrealign.intervals

Alternative

$ java -Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T
RealignerTargetCreator -R <path> /ucsc.hg19.fasta -I <path> Sample_uniqueAlignment_sort.
rmdups.bam   -known   <path>   /1000G_biallelic.indels.hg19.vcf   -log   <path>
Sample_uniqueAlignment   _sort.rmdups.   indelrealign.intervals.log   -o   <path>
Sample_uniqueAlignment_sort.rmdups.indelrealign.intervals

1.6 Perform indel realignment

$ java –Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T
IndelRealigner –R <path> / human_g1k_v37.fasta –known <path> /1000G_phase1.indels.b37.vcf
–known   <path>   /Mills_and_1000G_   gold_standard.indels.b37.sites.vcf   –I   <path>
Sample_uniqueAlignment_sort.   Rmdups.bam   -targetIntervals   <path>
Sample_uniqueAlignment_sort.rmdups.indelrealign.intervals   –o   <path>
Sample_uniqueAlignment_sort.rmdups.indelrealign.bam

Alternative

$ java -Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T
IndelRealigner -R <path> /ucsc.hg19.fasta -I <path> Sample_uniqueAlignment_sort.
rmdups.bam -targetIntervals path> Sample_uniqueAlignment_sort.rmdups.indelrealign.intervals
-log <path> Sample_uniqueAlignment _sort .rmdups. indelrealign.intervals.log –o <path>
Sample_uniqueAlignment_sort.rmdups.indelrealign .bam

1.7. Recalibrate base quality scores (GATK)
1.7.1 Get the recalibration model

$ java   -Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T
BaseRecalibrator –I <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.bam -R <path>
/human_g1k_v37.fasta –o <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.grp

–knownSites <path> /dbSnp version no.b37.vcf.gz –knownSites <path> /1000G_phase1.indels.b37.vcf –knownSites <path> /Mills_and_1000G _gold_standard.indels.b37.sites.vcf -nct [no.threads (8)]

1.7.2 Check the recalibration model

$ java –Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T BaseRecalibrator –I <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.bam -R <path> /human_g1k_v37.fasta –BQSR <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.grp -o <path> Sample_unique Alignment_sort.rmdups.indelrealign.postrecal.grp -knownSites <path> /dbSnp version no.b37.vcf.gz –knownSites <path> /1000G_phase1.indels.b37.vcf –knownSites <path> /Mills_and_1000G_gold_ standard.indels .b37.sites.vcf –nct 8

$ java –Xmx8g -jar <path>/GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T Analyze Covariates–R <path> /human_g1k_v37. fasta -before <path> Sample_uniqueAlignment_sort. rmdups. indelrealign.recal.grp –after <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.postrecal.grp–plots <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.postrecal.plots.pdf

1.7.3 Apply the recalibration
$ java –Xmx8g -jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T PrintReads –R <path> /human_g1k_v37.fasta -o <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.bam –I <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.bam –BQSR <path> Sample_unique Alignment_sort.rmdups.indelrealign.recal.grp –nct 8

1.8 Variant Calling (SNVs and indels) (UnifiedGenotyper/HaplotypeCaller)

$ java -Xmx4g -jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T HaplotypeCaller –R <path> /human_g1k_v37.fasta -D <path> /dbSnp version no.b37.vcf.gz -stand_call_conf 30 -stand_ emit_conf 10 –I <path> Sample_uniqueAlignment_ sort.rmdups.indelrealign.recal –o <path> Sample_ uniqueAlignment_sort.rmdups.indelrealign.recal.raw.vcf

$ java -Xmx8g -jar <path> /GenomeAnalysisTK- version no/GenomeAnalysisTK.jar –T UnifiedGenotyper -R <path> /ucsc.hg19.fasta –I <path> Sample_uniqueAlignment_ sort.rmdups. indelrealign.recal –D <path> /dbsnp_version no.hg19.vcf -L <path> /hg19_exome.interval_list -nt 8 -stand_call_conf 50.0 -stand_emit_conf 10.0 -dcov 200 -I INFO -A AlleleBalance -A

FisherStrand –log <path> Sample_uniqueAlignment_sort. rmdups.indelrealign.recal.SNP.log –o <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal.variants.SNP.vcf

$ java -Xmx8g -jar <path> /GenomeAnalysisTK- version no/GenomeAnalysisTK.jar –T UnifiedGenotyper -R <path> /ucsc.hg19.fasta –I <path> Sample_uniqueAlignment_ sort.rmdups. indelrealign.recal –D <path> /dbsnp_version no.hg19.vcf -L <path> /hg19_exome.interval_list -nt 8 -stand_call_conf 50.0 -stand_emit_conf 10.0 -dcov 200 -l INFO -A AlleleBalance -A FisherStrand –log <path> Sample_uniqueAlignment_sort. rmdups.indelrealign.recal.Indel.log –o <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal.variants.Indel.vcf -glm INDEL

1.9 Variant recalibration/ Hard filtering (GATK)

$ java -Xmx8g -jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T VariantRecalibrator -R <path> /ucsc.hg19.fasta -input Sample_uniqueAlignment_sort.rmdups. indelrealign. SNP.vcf resource:hapmap,VCF,known=false,training=true,truth=true,prior=15.0 <path> hapmap_3.3.hg19.vcf - resource:omni,VCF,known=false,training=true,truth=false,prior=12.0 <path> /1000G_omni2.5.hg19.vcf -resource:dbsnp,VCF,known=true,training=false,truth=false,prior=8.0 <path> DataFiles/dbsnp_version no.hg19.vcf -an QD -an HaplotypeScore -an MQRankSum -an ReadPosRankSum -an FS -an MQ --maxGaussians 6 -nt 8 -log <path> Sample_uniqueAlignment_sort. rmdups.indelrealign.recal.log -recalFile <path> Sample_uniqueAlignment_sort.rmdups.indelrealign. recal.recal -tranchesFile <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.tranches -rscriptFile <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.plot.R

$ java -Xmx8g -jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T ApplyRecalibration -R <path> /ucsc.hg19.fasta -input Sample_uniqueAlignment_sort.rmdups. indelrealign.SNP -recalFile <path> Sample_uniqueAlignment_sort.rmdups.indelrealign. recal.recal -tranchesFile <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.tranches -ts_filter_level 99.0 -log Sample_uniqueAlignment_sort. rmdups.indelrealign.recal.log -o Sample_uniqueAlignment_ sort.rmdups.indelrealign.recal.vcf

$ java –Xmx8g - jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T Variant Filtration -R <path> /human_g1k_v37.fasta -V <path> Sample_uniqueAlignment_sort.rmdups .indelrealign .recal.variants.SNP.vcf --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0" --

filterName "snp_hard_filter" -o <path> Sample_uniqueAlignment_sort.rmdups .indelrealign .recal.variants.filtered. SNP.vcf

$ java –Xmx8g - jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar -T VariantFiltration -R <path> /human_g1k_v37.fasta -V <path> Sample_uniqueAlignment_sort.rmdups .indelrealign .recal.variants.Indel.vcf --filterExpression "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0" --filterName "indel_hard_filter" -o <path> Sample_uniqueAlignment_sort.rmdups .indelrealign .recal.variants.filtered.Indel.vcf

1.10 Combine variant lists (SNP and Indel) (GATK)

$ java –Xmx8g - jar <path> /GenomeAnalysisTK- version no /GenomeAnalysisTK.jar –T Combine Variants -R <path> /ucsc.hg19.fasta (<path> /human_g1k_v37.fasta) --variant <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal.variants.SNP.vcf --variant <path> Sample_ unique Alignment_ sort.rmdups.indelrealign.recal.variants.Indel.vcf -o <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal.variants. combined.vcf.

1.11 Variant filtering

1.11.1 dbSNP

$ perl <path> /vcfhacks/annotateSnps.pl -d <path> /dbSnp version no.b37.vcf.gz <path>/ clinvar_ 20150330.vcf.gz -b 129 -f 1 -pathogenic –i <path> Sample_unique Alignment_ sort.rmdups.indelrealign .recal.variants. combined.vcf -o <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal. variants. Combined_1pc.vcf

1.11.2 EVS

$ perl <path> /vcfhacks/filterOnEvsMaf.pl -d <path> /evs/ -f 1 --progress –I <path> Sample_unique Alignment_ sort.rmdups.indelrealign.recal.variants.combined.vcf_notindbSNP version no or 1pc –o <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.variants.combined_1pc_evs.vcf

1.11.3 ExAC

$ perl <path> /vcfhacks/filterVcfOnVcf.pl -f <path> /ExAC.Version no /sites.vep.vcf.gz -w -y 0.01 -i <path> Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.variants.combined.vcf_1pc_evs –o <path>

339

Sample_uniqueAlignment_sort.rmdups.indelrealign.recal.variants.combined_1pc_evs_ExAC
.vcf

1.11.4 Control samples (3222 exomes of British Pakistani adults)

$ perl <path> /vcfhacks/filterVcfOnSample.pl -i <path> Sample_uniqueAlignment_sort.rmdups.
indelrealign.recal.variants.combined.vcf_1pc_evs_ExAC   -r   <path>   /bib/BUILD-2014-19-
05/8.BB.anno   .vcf.   gz   -o   <path>   Sample_uniqueAlignment_sort.rmdups.
indelrealign.recal.variants.combined_1pc_ evs_ExAC_CF.vcf

1.12 Variant annotation (ANNOVAR)

$ perl <path> annovar/convert2annovar.pl <path> Sample_uniqueAlignment_sort.rmdups.
indelrealign. recal.variants.combined.vcf_1pc_evs_ExAC_CF –outfile <path>
Sample_uniqueAlignment_sort.rmdups. indelrealign.
recal.variants.combined_1pc_evs_ExAC_CF.annovar.vcf

$ perl <path> annovar/table_annovar.pl <path> Sample_uniqueAlignment_sort.rmdups.
indelrealign. recal.variants.combined_1pc_evs_ExAC_CF.annovar.vcf <path>
/annovar/humandb -buildver hg19 -
protocolrefGene,phastConsElements46way,genomicSuperDups,esp6500si_all,1000gVersion
no apr_all ,snpversion no,ljb2_all -operation g,r,r,f,f,f,f -outfile <path>
Sample_uniqueAlignment_sort.rmdups. indelrealign.
recal.variants.combined_1pc_evs_ExAC_CF.annovar.vcf

B) Exome depth analysis

$ R
> library(ExomeDepth)
> data(exons.hg19)
> print(head(exons.hg19))
> Mohammed_bam_files <- c ("Sample_indelrealign_recal(File1).bam", "File2.bam","File3.bam",
etc.)
> Mohammed_counts<-getBamCounts(bed.frame=exons.hg19, bam.files=Mohammed_bams,
include.chr=FALSE, referenceFasta=" <path> /human_g1k_v37.fasta")

340

```
> ExomeCount.dafr <- as(my.counts[, colnames(my.counts)], 'data.frame')

> print(head(ExomeCount.dafr))

> Moh_counts.dafr<-as(Mohammed_counts[, colnames(Mohammed_counts)], 'data.frame')

> print(head(Moh_counts.dafr))

> Sample.test <- Moh.counts$ Sample_indelrealign_recal(File1).bam

> Moh.ref.samples <-c(File2.bam',File3.bam', etc. )

> Moh.reference.set<-as.matrix(Moh_Count.dafr[,Moh.ref.samples])

> Moh.choice<-select.reference.set
(test.counts=Sample.test,reference.counts=Moh.reference.set,bin.length=(Moh_Count.dafr$end
Moh_Count.dafr$start)/1000,n.bins.reduced=10000)

> print(Sample_choice[[1]])

> Moh.matrix <-as.matrix( Moh_Count.dafr[, Moh.choice$reference.choice, drop = FALSE])

> Moh.reference.selected<-apply(X=Moh.matrix,MAR=1,FUN=sum)

> Sample.all.exons <-new('ExomeDepth',test=Sample.test,
reference=Moh.reference.selected,formula ='cbind(test, reference)~1')

> Sample.all.exons<-CallCNVs(x=sample_all.exons,transition.probability=10^-4, chromosome
=Moh_
Count.dafr$space,start=Moh_Count.dafr$start,end=Moh_Count.dafr$end,name=moh_Count.daf
r$names)

> head(Sample.all.exons@CNV.calls)

> data(Conrad.hg19)

> head(Conrad.hg19.common.CNVs)

> Sample.all.exons<-AnnotateExtra(x=Sample.all.exons,
reference.annotation=Conrad.hg19.common. CNVs, min.overlap=0.5,
column.name='Conrad.hg19')

> print(head(Sample.all.exons@CNV.calls))

> exons.hg19.GRanges <-
GenomicRanges::GRanges(seqnames=exons.hg19$chromosome,IRanges::I
Ranges(start=exons.hg19$start,end=exons.hg19$end),names=exons.hg19$name)

> Sample.all.exons <- AnnotateExtra(x=Sample.all.exons,
reference.annotation=exons.hg19.GRanges, min.overlap =0.0001, column.name='exons.hg19')

> Sample.all.exons@CNV.calls[3:6,]
```

> output.file <- 'Sample_calls_File1CNVs.csv'

> write.csv(file=output.file,x=Sample.all.exons@CNV.calls,row.names=FALSE)

C) ANKRD2 and PDZD8 in UK10K analysis

-Download the specific data stets from the UK10K using FileZilla
-Decrypt the files Openssl enc -aes-256-cbc –d in
-Search for variants present in ANKRD2 or PDZD8 (vcfhacks)
perl vcfhacks/getVariantsByLocation.pl
–i UK10K_input_file.vcf
–o filtered_output_file.vcf
–r 10:99332256-99343641 10:119042606-119134937
(-i = input file, -o = output file, -r = regions to filter for)

-Annotate variants using variant effect predictor (VEP)
perl variant_effect_predictor.pl --offline --vcf --everything
--dir_cache /home/variant_effect_predictor/vep_cache
--dir_plugins /home/variant_effect_predictor/vep_cache/Plugins
--plugin Condel,/home/variant_effect_predictor/vep_cache/Plugins/config/Condel/config/
--plugin ExAC,/home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz
--plugin SpliceConsensus
--fasta
/home/variant_effect_predictor/fasta/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
-input filtered_output_file.vcf
-o filtered_output_file.vep.vcf
(-input = input file, -o = output file, all other options are telling VEP how to annotate your file)

-CADD score the variants (version 1.3) and filter out anything with CADD < 10.
perl vcfhacks/rankonCaddScore.pl -d
-c /data/shared/cadd/v1.3/*.gz
-i filtered_output_file.vep.vcf
-o filtered_output_file.vep.cadd1.3.vcf
(-c = location of files containing cadd scores, -i = input file, -o = output file)

-Convert to excel and remove anything that isn't functional (synonymous, intronic, etc.)
perl vcfhacks/annovcfToSimple.pl
-i filtered_output_file.vep.cadd1.3.vcf
-o filtered_output_file.vep.cadd1.3.xlsx
-v –f –u --canonical_only

(-i = input file, -o = output file, -v = file has been annotated with vep, -f = only show functional
variants, -u = summarise variant data, X = show who has non-reference genotype

Appendix 6- Sequence of pDEST-733 vector with DFNB31 insert

```
cttgtacaaagtggttcgatctagaatggctagcttgggatctttgtgaaggaaccttacttctgtggtg
tgacataattggacaaactacctacagagatttaaagctctaaggtaaatataaaatttttaagtgtata
atgtgttaaactagctgcatatgcttgctgcttgagagttttgcttactgagtatgatttatgaaaatat
tatacacaggagctagtgattctaattgtttgtgtattttagattcacagtcccaaggctcatttcaggc
ccctcagtcctcacagtctgttcatgatcataatcagccataccacatttgtagaggtttttacttgcttt
aaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgttt
attgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttttcac
tgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatcgatcctgcattaat
gaatcggccaacgcgcggggagaggcggtttgcgtattggctggcgtaatagcgaagaggcccgcaccga
tcgcccttcccaacagttgcgcagcctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgcg
gcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgcttt
tcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagg
gttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtggg
ccatcgccctgatagacggtttttcgcccttcgacgttggagtccacgttctttaatagtggactcttgt
tccaaactggaacaacactcaaccctatctcggtctattctttgatttataagggattttgccgatttc
ggcctattggttaaaaaatgagctgatttaacaaatatttaacgcgaattttaacaaaatattaacgttt
acaatttcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgcggatct
gcgcagcaccatggcctgaaataacctctgaaagaggaacttggttaggtaccttctgaggcggaaagaa
ccagctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaa
agcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgc
aaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactcc
gcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcc
tcggcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagcttg
attcttctgacacaacagtctcgaacttaaggctagagccaccatgattgaacaagatggattgcacgca
ggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctg
atgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgc
cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagct
gtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcc
tgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgct
tgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaa
gccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcca
ggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatat
catggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcag
gacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgc
tttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagc
gggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgatggccgcaataaaata
tctttattttcattacatctgtgtgttggtttttttgtgtgaatcgatagcgataaggatccgcgtatggt
gcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctga
cgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgc
atgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttt
tataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcgg
aacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataa
atgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttt
ttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatca
gttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcccc
gaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacg
ccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcac
agaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataac
actgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgg
gggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtga
caccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagct
tcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttc
cggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcact
ggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa
cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttact
catatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttga
taatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatc
```

343

aaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctac
cagcggtggtttgtttgccggatcaagagctaccaactcttttttccgaaggtaactggcttcagcagagc
gcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccg
cctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccg
ggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacaca
gcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccacg
cttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggg
agcttccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcg
atttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttc
ctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgta
ttaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcga
ggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagagct
tgcaattcgcgcgttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatat
ttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgt
ctaagaaaccattattatcatgacattaacctataaaaataggcgtagtacgaggccctttcactcatta
gatgcatgtcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattg
acgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagt
atttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgt
caatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcag
tacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggat
agcggtttgactcacggggatttccaagtctccacccccattgacgtcaatgggagtttgttttggcacca
aaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgta
cggtgggaggtctatataagcagagctctccctatcagtgatagagatctccctatcagtgatagagatc
gtcgacgagctcgtttagtgaaccgtcagatcgcctggagacgccatccacgctgttttgacctccatag
aagacaccgggaccgatccagcctccggactctagaggatccctaccggtgatatcctcgagaccatggc
ctcctccgaggacgtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgaacggccac
gagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtga
ccaagggcggccccctgcccttcgcctgggacatcctgtcccctcagttccagtacggctccaaggccta
cgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgc
gtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttca
tctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatggg
ctgggaggcctccaccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagatgaggctg
aagctgaaggacggcggccactacgacgccgaggtcaagaccacctacatggccaagaagcccgtgcagc
tgcccggcgcctacaagaccgacatcaagctggacatcacctcccacaacgaggactacaccatcgtgga
acagtacgagcgcgccgagggccgccactccaccggcgcctcgagcccatcaacaagtttgtacaaaaaa
gcaggcttcaacgcgccgctggacggcctgtcggtgagctcgtcctccaccggctcgctgggctcggcgg
ccggggcgggcggcggcgggggcgcggggctgcggttactgtctgccaacgtgcgccagctgcaccaagc
gctgaccgcgctgctgagcgaggcggagcgggagcagttcacccactgcctgaacgcttaccacgcgcgc
cgcaacgtcttcgacctggtgcgcaccctgcgcgtgctgctggacagtccggtcaagcggcgcctgctgc
ccatgcttcgtctggtcatcccgcgctccgaccagctgctcttcgaccaatacacggccgagggcctcta
cctgcccgccaccacccctacaggcagcccgcctggggcggcccccgacagcgcggggccaggggaggtg
cgcctggtgagtttgcggcgtgccaaggcccacgagggcttgggcttcagcatccgtgggggctcggagc
acggcgtgggcatctacgtgtctctggtggaaccaggctctctagctgagaaggaaggactgcgggtcgg
ggaccagattctgcgcgtcaacgacaaatccctggcccgggtgacccacgcggaggccgtcaaggctctg
aagggctccaagaagctggtgctgtctgtgtactcagcagggcgcatccctggggggctacgtcaccaacc
acatctacacctgggtggacccgcagggccgcagcatctccccaccctcgggcctgccccagccccacgg
tggtgccctgaggcagcaggagggtgaccggaggagcaccctgcacctcctgcaaggaggggatgagaaa
aaggtgaacctggtgctggggggacggccggtccctgggcctcacgatccgtgggggagctgagtacggcc
ttggcatttacatcactggcgtggacccaggctctgaagcagaaggcagcgggctcaaggttggggacca
gattctagaagtgaatgggcggagctttctcaacatcctacacgacgaggctgtcaggctgcttaagtca
tctcggcacctcatcctgacagtgaaggacgtcgggaggctgccccatgcccgcaccactgtggacgaga
ccaagtggatcgccagttcccggatcagggagaccatggcgaactcggcagggtttcttggcgatctcac
aacagaaggaataaacaagccaggattttacaagggcccagccggctcccaggtgaccctgagcagcctg
gggaaccagacacgagtgctgctggaggagcaggctcggcacctgctgaacgagcaggaacacgccacca
tggcctactacctggatgagtaccgtggcggcagcgtctctgtggaggccctcgtcatggccctgttcaa
gctgctcaacacccacgccaagttctcactcctctctgaggtgagaggcaccatttccccgcaagaccta
gaacgcttcgaccaccggtgctgaggcgtgagattgagtccatgaaggcgcggcagccccccaggccccg
gggctggggacacctactccatggtctcctacagtgacacgggttcatccacaggcagccacggcacctc
caccaccgtcagctcggccaggaacactctggacctggaggaaactggcgaggctgtccagggcaatatc
aacgccctcccagatgtgtccgtggatgatgtcagatccacctcccagggggctgtcaagcttcaagccac

```
tgcctcgcccaccacctctggcccaaggcaacgactcccactaggccagccaaggaagctggggagaga
ggacctccagccaccttcctccacgccttcctgctcgggcactgtcttctcggctccacagaaccgcagc
ccgccagcgggcaccgcacccaccccagggacctcctctgcacaggacttgccctcttcccccatctatg
cctccgtctcccctgccaaccccagctccaagaggccgctggacgcccatctggccctggtcaaccaaca
ccccatcggccccttcccacgggtccagtcaccccccgcacctgaaaagcccctctgcagaggccacagtg
gctgggggctgccttctgcccccatcaccctctggccacccagaccagacaggcacaaaccagcactttg
tcatggtggaggtccaccgccccgacagcgagccagacgtcaatgaagtgagggcgctgccccagacgcg
cacctctacgctctcccagctctcggacagcgggcagactctaagcgaggacagtggtgtggatgctggc
gaggcagaggccagcgccccaggccgaggaaggcagtcggcgtccaccaagagcaggagtagcaaggagc
tgcctcggaacgagaggcccacagatggggccaacaaaccgcctggacttctggagcccacgtccactct
ggtccgtgtgaagaaaagtgcggccaccctgggcatcgccatcgagggtggcgccaacacccgccagccc
ctgcctaggattgtcactattcagagaggcggctcagctcacaactgtgggcagctcaaggtgggccacg
tgattctggaagtgaatgggctgacgcttcggggcaaggagcaccgggaggccgcccgcattatcgccga
ggccttcaagactaaggaccgtgactacattgactttctggtcactgagttcaatgtgatgctctagcac
ccagcttt
```

red: mRFP tag

blue: whirlin cDNA

Appendix 7- Recipe of used buffers and reagents

Running buffer (X5 stock):
15.135g Tris-Base
71.315g glycine
0.5g SDS (0.1%)
Make up to 1L using dH_2O.
For x1, dilute 1:5 with dH_2O

Transfer buffer:
15.135g Tris-Base
71.315g glycine
Make up to 1L using dH_2O.
For X1, add 20% methanol:

PBST: add 500ul of Tween to 1L DPBS.

Stripping buffer:
3.8gTris-Base
10g SDS
3500ul β-mercaptoethanol
Make up to 500ml using dH_2O.
Correct pH to 6.8 using HCl.

Laemmli Buffer:
3.75 ml of 1M Tris-HCl pH 6.8
15ml of 100% glycerol
0.6ml of 1% bromophenol blue
6ml of 20% SDS
Make up to 30ml using dH_2O.
Add β-mercaptoethanol before use (200ul to 800ul sample buffer)

Appendix 8- List of X-linked genes known to be associated with ID.

This list is based on updated information from University of Colorado Denver, (http://gfuncpathdb.ucdenver.edu/iddrc/iddrc/data/IDgenelist_chr.html)

No	Gene Symbol	Gene Names
1	ABCD1	ATP-binding cassette, sub-family D (ALD), member 1
2	ACSL4	acyl-CoA synthetase long-chain family member 4
3	AFF2	AF4/FMR2 family, member 2
4	AGTR2	angiotensin II receptor, type 2
5	AP1S2	adaptor-related protein complex 1, sigma 2 subunit
6	ARHGEF6	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6
7	ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9
8	ARX	aristaless related homeobox
9	ATP6AP2	ATPase, H+ transporting, lysosomal accessory protein 2
10	ATP7A	ATPase, Cu++ transporting, alpha polypeptide
11	ATRX	alpha thalassemia/mental retardation syndrome X-linked
12	AVPR2	arginine vasopressin receptor 2
13	BCOR	BCL6 corepressor
14	BRWD3	bromodomain and WD repeat domain containing 3
15	CACNA1F	calcium channel, voltage-dependent, L type, alpha 1F subunit
16	CASK	calcium/calmodulin-dependent serine protein kinase (MAGUK family)
17	CDKL5	cyclin-dependent kinase-like 5
18	CUL4B	cullin 4B
19	DCX	doublecortin
20	DKC1	dyskeratosis congenita 1, dyskerin
21	DLG3	discs, large homolog 3 (Drosophila)
22	DMD	dystrophin
23	FANCB	Fanconi anemia, complementation group B
24	FGD1	FYVE, RhoGEF and PH domain containing 1
25	FLNA	filamin A, alpha
26	FMR1	fragile X mental retardation 1
27	FRAXE	fragile site, folic acid type, rare, fra(X)(q28) E
28	FTSJ1	FtsJ homolog 1 (E. coli)
29	G6PD	glucose-6-phosphate dehydrogenase
30	GDI1	GDP dissociation inhibitor 1
31	GK	glycerol kinase

347

No	Gene Symbol	Gene Names
32	GPC3	glypican 3
33	GRIA3	glutamate receptor, ionotrophic, AMPA 3
34	HCCS	holocytochrome c synthase
35	HPRT1	hypoxanthine phosphoribosyltransferase 1
36	HSD17B10	hydroxysteroid (17-beta) dehydrogenase 10
37	HUWE1	HECT, UBA and WWE domain containing 1
38	IDS	iduronate 2-sulfatase
39	IGBP1	immunoglobulin (CD79A) binding protein 1
40	IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells
41	IL1RAPL1	interleukin 1 receptor accessory protein-like 1
42	IQSEC2	IQ motif and Sec7 domain 2
43	KDM5C	lysine (K)-specific demethylase 5C
44	KIAA2022	KIAA2022
45	KLF8	Kruppel-like factor 8
46	L1CAM	L1 cell adhesion molecule
47	LAMP2	lysosomal-associated membrane protein 2
48	MAGT1	magnesium transporter 1
49	MAOA	monoamine oxidase A
50	MECP2	methyl CpG binding protein 2 (Rett syndrome)
51	MED12	mediator complex subunit 12
52	MID1	midline 1 (Opitz/BBB syndrome)
53	NDP	Norrie disease (pseudoglioma)
54	NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1
55	NHS	Nance-Horan syndrome
56	NLGN3	neuroligin 3
57	NLGN4X	neuroligin 4, X-linked
58	OCRL	oculocerebrorenal syndrome of Lowe
59	OFD1	oral-facial-digital syndrome 1
60	OPHN1	oligophrenin 1
61	OTC	ornithine carbamoyltransferase
62	PAK3	p21 protein (Cdc42/Rac)-activated kinase 3
63	PCDH19	protocadherin 19
64	PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1
65	PGK1	phosphoglycerate kinase 1
66	PHF6	PHD finger protein 6
67	PHF8	PHD finger protein 8

No	Gene Symbol	Gene Names
68	PLP1	proteolipid protein 1
69	PORCN	porcupine homolog (Drosophila)
70	PQBP1	polyglutamine binding protein 1
71	PRPS1	phosphoribosyl pyrophosphate synthetase 1
72	PTCHD1	patched domain containing 1
73	RAB39B	RAB39B, member RAS oncogene family
74	RAB40AL	RAB40A, member RAS oncogene family-like
75	RPS6KA3	ribosomal protein S6 kinase, 90kDa, polypeptide 3
76	SHROOM4	shroom family member 4
77	SLC16A2	solute carrier family 16, member 2 (monocarboxylic acid transporter 8)
78	SLC6A8	solute carrier family 6 (neurotransmitter transporter, creatine), member 8
79	SLC9A6	solute carrier family 9 (sodium/hydrogen exchanger), member 6
80	SMC1A	structural maintenance of chromosomes 1A
81	SMS	spermine synthase
82	SOX3	SRY (sex determining region Y)-box 3
83	SRPX2	sushi-repeat containing protein, X-linked 2
84	SYN1	synapsin I
85	SYP	synaptophysin
86	TIMM8A	translocase of inner mitochondrial membrane 8 homolog A (yeast)
87	TSPAN7	tetraspanin 7
88	UBE2A	ubiquitin-conjugating enzyme E2A
89	UPF3B	UPF3 regulator of nonsense transcripts homolog B (yeast)
90	ZDHHC15	zinc finger, DHHC-type containing 15
91	ZDHHC9	zinc finger, DHHC-type containing 9
92	ZNF41	zinc finger protein 41
93	ZNF674	zinc finger protein 674
94	ZNF711	zinc finger protein 711
95	ZNF81	zinc finger protein 81

Appendix 9- CG10362 expression following RNAi in males and females

These charts were generated from the qPCR that was done to measure the expression of CG10632. The expression was normalised by expression of eEF-1a (Eukaryotic elongation factor 1-alpha). The upper chart (Figure A) indicates the CG10632 expression in different geneotypes of *Drosophila*. Comparing with the RNAi responder and wild type (WT) flies, the progeny of RNAi responde and actin driver show an obvious CG10632 reduction. Figure B shows the compensation effect in CG10632 expression, at which the female progeny of RNAi responder and actin driver cross has an incresed expression of CG10632. This expreiment was only done once and no error bars could be generated.
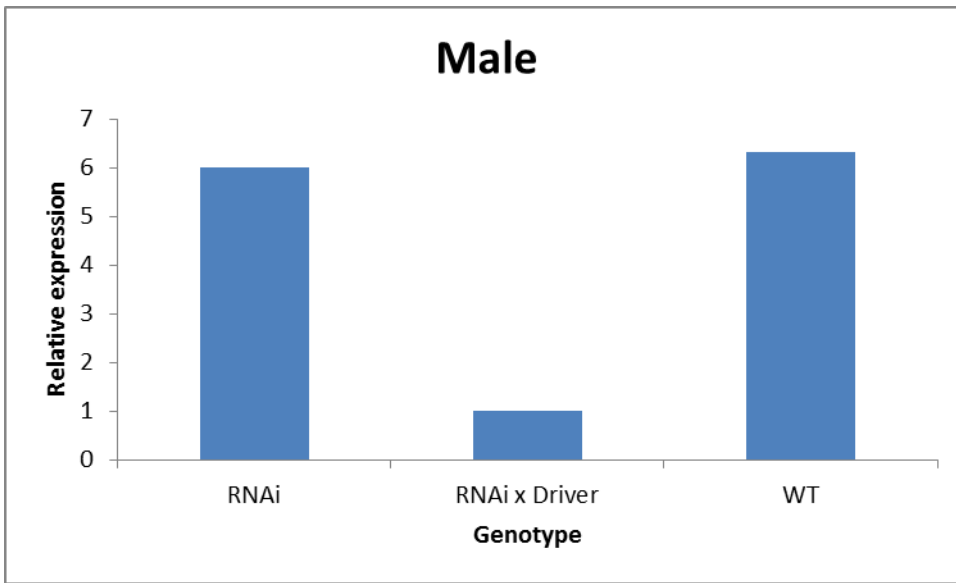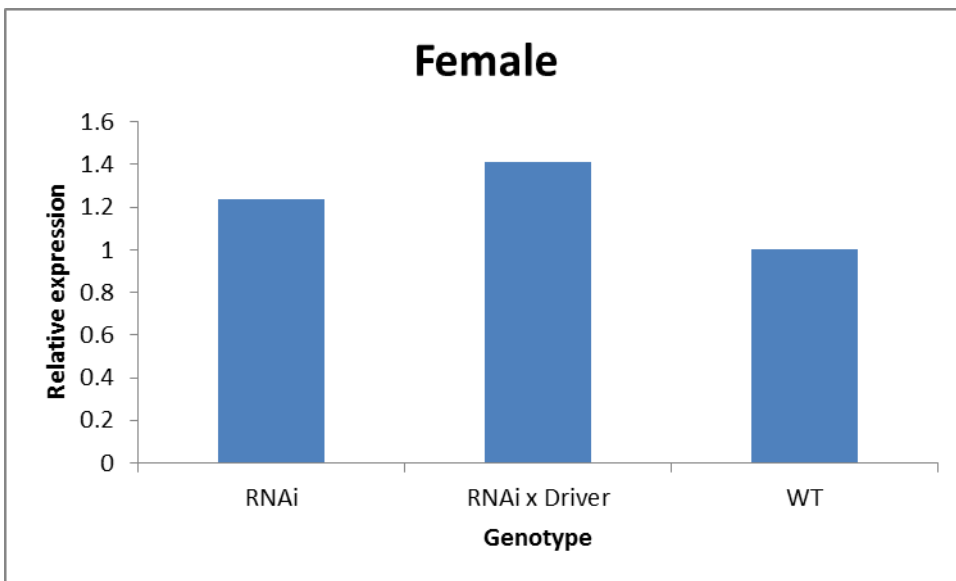

Figure A

Figure B

Appendix 10: List of the genes located in the 13q region (76,482,752..82,585,964)

start	stop	Symbol	Description
77454304	77460540	KCTD12	potassium channel tetramerization domain containing 12
77502585	77503224	BTF3P11	basic transcription factor 3 pseudogene 11
77522694	77532777	ACOD1	aconitate decarboxylase 1
77555247	77556072	RPL7P44	ribosomal protein L7 pseudogene 44
77558877	77563104	DHX9P1	DEAH-box helicase 9 pseudogene 1
77566059	77576652	CLN5	ceroid-lipofuscinosis, neuronal 5
77579389	77601331	FBXL3	F-box and leucine rich repeat protein 3
77618792	77901177	MYCBP2	MYC binding protein 2, E3 ubiquitin protein ligase
77649649	77661911	MYCBP2-AS1	MYCBP2 antisense RNA 1
78109809	78219398	SCEL	sciellin
78197507	78197607	RNY3P7	RNA, Ro-associated Y3 pseudogene 7
78233548	78233808	SPTLC1P5	serine palmitoyltransferase long chain base subunit 1 pseudogene 5
78235511	78236536	LOC100129307	putative UPF0607 protein ENSP00000383144
78271989	78338377	SLAIN1	SLAIN motif family member 1
78272147	78272251	MIR3665	microRNA 3665
78393072	78482577	EDNRB-AS1	EDNRB antisense RNA 1
78469616	78549664	EDNRB	endothelin receptor type B
78555981	78571963	LINC01069	long intergenic non-protein coding RNA 1069
78587018	78627730	LINC00446	long intergenic non-protein coding RNA 446
78628990	79191460	RNF219-AS1	RNF219 antisense RNA 1
78996958	78997309	RPL31P54	ribosomal protein L31 pseudogene 54
79062317	79062647	TCEB1P23	transcription elongation factor B subunit 1 pseudogene 23
79173227	79177695	POU4F1	POU class 4 homeobox 1
79188421	79233314	RNF219	ring finger protein 219
79335269	79335821	RPL21P111	ribosomal protein L21 pseudogene 111
79361454	79398991	LINC00331	long intergenic non-protein coding RNA 331
79423366	79425343	HSPD1P8	heat shock protein family D (Hsp60) member 1 pseudogene 8
79483065	79484682	CCT5P2	chaperonin containing TCP1 subunit 5 pseudogene 2
79484857	79484966	CCT5-2P	chaperonin containing TCP1, subunit 5 (epsilon) pseudogene 2
79664900	79665309	NIPA2P5	non imprinted in Prader-Willi/Angelman syndrome 2 pseudogene 5
79740362	79741024	BCAS2P3	breast carcinoma amplified sequence 2 pseudogene 3
79885962	79980393	RBM26	RNA binding motif protein 26
79963669	79963766	RNA5SP33	RNA, 5S ribosomal pseudogene 33
79980444	79998471	RBM26-AS1	RBM26 antisense RNA 1
80051499	80055366	NDFIP2-AS1	NDFIP2 antisense RNA 1
80055259	80130212	NDFIP2	Nedd4 family interacting protein 2

start	stop	Symbol	Description
80446721	80492171	LINC00382	long intergenic non-protein coding RNA 382
80616827	80624465	LOC101927216	uncharacterized LOC101927216
80774284	80785454	LOC101927238	uncharacterized LOC101927238
80910111	80915086	SPRY2	sprouty RTK signaling antagonist 2
81192905	81194261	HNRNPA1P31	heterogeneous nuclear ribonucleoprotein A1 pseudogene 31
81228890	81231137	PWWP2AP1	PWWP domain containing 2A pseudogene 1
81455698	81463020	ARF4P4	ADP ribosylation factor 4 pseudogene 4
81847005	81847302	HIGD1AP2	HIG1 hypoxia inducible domain family member 1A pseudogene 2
82264031	82265215	PTMAP5	prothymosin, alpha pseudogene 5