A Molecular Analysis of 22q11.2 Deletion Syndrome

A thesis submitted for the Degree of Doctor of Philosophy at the School of Medicine, Cardiff University

By

Afnan Salah Salaka

MRC Centre for Neuropsychiatric Genetics and Genomics

Division of Psychological Medicine and Clinical Neuroscience

School of Medicine

Cardiff University

2017

Dedication

My father, who passed away while I was doing this PhD and he could not see this dream come true.

This thesis is dedicated to his soul.

Daddy you were, are, and will always be around for me!

Acknowledgements

First, I would like to thank Almighty Allah for the perseverance that he has bestowed upon me during this research project and indeed throughout my life. I believe that every single effort I've made, would not have happened without his mercy and blessing.

I wish to express my sincere gratitude to my supervisor, Prof Nigel Williams, for his invaluable guidance, unfailing patience, and encouragement throughout the PhD. I owe my utmost gratitude to Prof Williams for considering me worthy of this opportunity and for constantly motivating me when I felt down and out.
A special thank to my supervisory team, Prof Micelle Owen and Prof Marianne van den Bree for providing me with the permission to use the ECHO study samples to achieve this PhD.
I would also like to thank the ECHO study team with special thanks to Miss Hayley Moss, Dr Maria Niarchou, and Dr Samuel Chawner for their great help in obtaining the biological samples and the supporting data for the participants.
I am also extremely grateful to Dr Kiran Mantripragada and the lab core team in the MRC Centre for Neuropsychiatric Genetics and Genomics for their valuable support during my lab work.
I would also thank the bioinformatics team in the department, Dr Antonio Pardinas, Dr Elliott Rees, Dr Leon Hubbard, and Dr Valentina Escott-Price, for their tremendous help in the bioinformatics analyses.
Last but not least, I would like to thank the patients who provided the samples and without whose cooperation this project would not have been possible.

I would like to thank my sponsor Umm Al-Qura University with special thanks to Dr Ahmed Asshi, Dr Mohammed Baslamah and the members of the Faculty of Applied Medical Sciences for their continues support. I will not forget to thank the Ministry of the Higher Education of Saudi Arabia and the Saudi Cultural Bureau who sponsored me for this great opportunity to learn, experience, and gain qualifications.

I would like to express a special word of thanks with love to my friends especially Dr, Nada Almazrooa, Dr Rania Al-Otabi, Dr Amal Al Salamah, Dr Noa Carrera, Dr Isla Humphreys, and Dr Marisa Taylor, for your valuable friendship and your emotional support. You made my stay in Cardiff enjoyable and I hope to treasure your friendship forever.

I am most grateful to my mother, Aunty Faizah, my sisters, and brothers for having a constant faith in me and for your continuous support in all of my life's aspects. Your love and prayers are a big source of support during this PhD.
I cannot thank my husband, Homam, enough for supporting me throughout doing my PhD, and for always being there whenever I needed him, and for being patient during my busy years of the PhD.
I am greatly thankful to my father, who passed away while I was abroad for doing this PhD. I will always remember your supportive words of encouragement to continue my journey to learn and will always be praying for you. I love you daddy and I know that you are always around for me.

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ……..……...……..……Date …….

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ……..……...……..……Date …….

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ……..……...……..……Date …….

Statement 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……..……...……..……Date …….

Statement 4
PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed ……..……...……..……Date …….

Table of Contents

Contents

Abbreviations

Abbreviation	Meaning
µl	Microliter
22q11.2 DS	22q11.2 deletion syndrome
A	Absorbance
AAO	Age at onset
ADHD	Attention deficit and hyperactivity disorder
ADI-R	Autism diagnostic interview-revised algorithm
ASD	Autism spectrum disorder
BAF	B allele frequency
BAF SD	B allele frequency standard deviation
BBAG	Study of brain, behaviour and genetics in 22q11.2 deletion syndrome
BSS	Bernard-soulier syndrome
CAPA	Child and adolescent psychiatric assessment
CEDNIK	Cerebral dysgenesis, neuropathy, ichthyosis and keratoderma
CGH	Comparative genome hybridization
CHD	Congenital heart diseases
CNS	Central nervous system
CNVs	Copy number variants
cRNA	Complementary RNA
CTAF	Conotruncal anomaly face syndrome
DAT	Dopamine transporter
DAVID	The database for annotation, visualization and integrated discovery
DD	Developmental delay
ddNTP	Dideoxynucleotides
DE	Differentially expressed
Del	Deletion
DGS	DiGeorge syndrome
DNA	Deoxyribonucleic acid
DNP	Dinitrophenol
DSM-IV	The diagnostic and statistical manual, fifth edition of text revision
dup	Duplication
ECHO	Study of ExperienCes of people witH cOpy number variants
EOPD	Early age of onset
FDR	False discovery rate
FISH	Florescent *in situ* hybridization
FSIQ	Full scale IQ
GRR	Genotype risk ratio
GWA	Genome-wide association

GWAS	Genome-wide association study
HMM	Hidden Marcov model
HSA22	Human chromosome 22
HWE	Hardy Weinberg equilibrium
IAA	Interrupted aortic arch
IAC	Inter-array correlation
IBD	Identity by descent
ID	Intellectual disability
ID/DD	Intellectual disability/developmental delay
IPDGC	International Parkinson's disease genomic consortium
IQ	Intelligence quotient
IQR	Interquartile range
IVD	PAXgene blood RNA tube
IVT	In vitro transcription
LCRs	Low copy repeats
LD	Linkage disequilibrium
Log_2	Logarithm 2
LOPD	Late age of onset
LRR	Log R ratio
LRR SD	Log R ratio standard deviation
MAF	Minor allele frequency
MB-COMT	Membrane-bound COMT form
MCA	Multiple congenital anomalies
Met	Methionine
miRNA	Micro RNA
MLPA	Multiplex ligation dependent probe amplification
MMU16	Mouse chromosome 16
MZ	Monozygotic
NAHR	Non-allelic homologous recombination
ND	Neurodevelopmental delay
ng	Nano gram
NGS	Next generation sequencing
ODD	Oppositional defiant disorder
OR	Odds ratio
PA-VSD	Pulmonary atresia with ventricular septal defect
PCA	Principle component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDD-NOS	Pervasive developmental disorder-not otherwise specified
PFB	Population frequency B allele
PFC	Prefrontal cortex
PIQ	Performance IQ
PMCs	Peripheral blood mononuclear cells
PPI	Pre-pulse inhibition
PPMI	Parkinson's progression markers initiative
PTM	Post translation modification

QC	Quality control
qPCR	Quantitative PCR
r^2	Correlation coefficient
RI	Rank invariant
RIN	RNA integrity number
RNA	Ribonucleic acid
RS	Robust spline
RT	Reverse transcriptase
RTqPCR	Real time quantitative PCR
S-COMT	Soluble COMT form
SCQ	Social communication questionnaire
SD	Standard deviation
SE	Standard error
SMS	Smith Magenis syndrome
SNHL	Sensorineural hearing loss
SNP	Single nucleotide polymorphism
SoS	Soto syndrome
SPEM	Smooth pursuit eye movements
SS	Simple scaling
TA	Truncus arteriosus
TDR	Typically deleted region
TF	Tetralogy of Fallot
TRAP	Tremor at rest, rigidity, akinesia (or bradykinesia) and postural instability
UK-WTCCC2	UK Wellcome Trust case control consortium 2
Val	Valine
VCFS	Velocardiofacial syndrome
VIQ	Verbal IQ
VST	Variance stabilizing transformation
WASI	Wechsler abbreviated scale of intelligence
WBS	Williams-Beuren syndrome
WF	Waviness factor
WTCCC	Wellcome Trust case control consortium

Summary

22q11.2 Deletion Syndrome is a genetic syndrome that occurs in incidence of 1:4000 and is associated with variable phenotypic expression. It is caused by a deletion at chromosome 22q11.2. Individuals with 22q11.2DS have a greatly increased risk of developing neuropsychiatric disorders, in particular schizophrenia in adulthood, and ADHD and ASD in childhood.

This thesis sought to investigate the possible molecular mechanisms that underlie the psychiatric variability in 22q11.2DS by studying a well-characterized cohort of 76 children with 22q11.2DS. Four mechanisms were investigated: (a) dosage sensitivity of genes within 22q11.2 region, (b) a disruption of the expression of genes flanking the deletion i.e. positional effect and genome-wide, (c) the presence of additional genetic risk variants within the 22q11.2 region, and (d) the presence of additional CNVs outside of the 22q11.2 deleted region.

While this work revealed significant evidence for differential expression of 39 of the genes spanned by the deletion, there was no significant evidence for a positional effect at other genes on chromosome 22. The genome-wide differential gene expression analysis revealed four significantly enriched biological networks (FDR <0.05) that are involved in: 1) translation, protein synthesis machinery, and post-translation modifications; 2) apoptosis; 3) regulation of the immune system; and 4) intramembrane organelles.

The association analyses of genetic variants present on the non-deleted 22q11.2 chromosome did not identify any that were significantly associated with psychiatric phenotypes in 22q11.2DS.

The genome-wide screen for additional CNVs identified a non-significant trend that large, rare CNVs were enriched in 22q11.2DS patients with psychiatric phenotypes, however there was no evidence that these additional CNVs were enriched for CNVs that had been previously implicated in developmental delay and neuropsychiatric disease.

In addition, in this thesis also investigated the role of deletions at 22q11.2 in a large cohort of individuals with idiopathic Parkinson's disease. This provided significant evidence that deletions at 22q11.2 increase the risk of Parkinson's disease, particularly the early-onset form.

Taken together, the data presented in this PhD suggested that the mechanism by which haploinsufficiency at 22q11.2 increases risk to psychiatric illness is likely to be complex. To follow up this work, future studies should utilise larger numbers of samples, use neurologically relevant tissue and apply more sophisticated approaches to screen for changes in gene expression and additional genetic variants.

Associated Publications and Presentations

Publications:

1. Mok K., Sheerin U., Simón-Sánchez J., Salaka A., et al. Deletions at 22q11.2 in idiopathic Parkinson's disease: a combined analysis of genome-wide association data. Lancet Neurology, 15(6), pp.585-596.

Presentations:

1. Salaka A., Moss H., Niarchou M., Chawner S., van den Bree M., Owen M., and Williams N. Investigating the Two-Hit Hypothesis in 22q11.2 Deletion Syndrome. Presented as a poster at the 28th Annual Postgraduate Research Day, School of Medicine, Cardiff University; 15 November 2013.

2. Salaka A., Moss H., Niarchou M., Chawner S., van den Bree M., Owen M., and Williams N. Investigating the Two-Hit Hypothesis in 22q11.2 Deletion Syndrome. Presented as a poster at 9th Biennial International 22q11.2DS Conference, Mallorca, Spain; 19-22 June 2014.

3. Salaka A., Moss H., Niarchou M., Chawner S., van den Bree M., Owen M., and Williams N. Investigating the Two-Hit Hypothesis in 22q11.2 Deletion Syndrome. Presented as a poster at The 8th Saudi Students Conference, hosted by Imperial College London, United Kingdom; 31 January- 1 February 2015.

4. Salaka A., Moss H., Niarchou M., Chawner S., van den Bree M., Owen M., and Williams N. An investigation into the effect of 22q11.2 deletion on gene expression. Presented as a poster at The 9th Saudi Students Conference, hosted by University of Birmingham, United Kingdom; 13-14 February 2016.

5. Salaka A., Moss H., Niarchou M., Chawner S., van den Bree M., Owen M., and Williams N. An investigation into the effect of 22q11.2 deletion on gene expression. Published abstract at the American Society of Human Genetic Annual Meeting, Vancouver, Canada; 12-18 October 2016.

Chapter 1: General Introduction

1.1. 22q11.2 Deletion Syndrome

22q11.2 Deletion Syndrome (22q11.2DS) is a genetic disorder caused by a hemizygous microdeletion on the long arm (q) of chromosome 22 in an area designated by 11.2 (Scambler et al. 1992). It is the most common interstitial microdeletion disorder occuring in man and the second most common genetic syndrome after Down's Syndrome (Rauch et al. 2006). Moreover, 15% of congenital cardiac defects are represented by 22q11.2DS (Devriendt et al. 1998; Gothelf & Lombroso 2001).

The disease collectively describes multiple genetic conditions with overlapping features including DiGeorge Syndrome (DGS), Velocardiofacial Syndrome (VCFS), Shprintzen Syndrome, Conotruncal Anomaly Face Syndrome (CTAF), Cayler Syndrome, Sedlackova Syndrome, Cardiofacial Syndrome, Takao Syndrome, and CATCH-22 (Shprintzen 2005). The presence of multiple names for the same disorder is due to the markedly wide range of phenotype expression in the patients with the deletion (Shprintzen 2005). However, the term 22q11.2DS has been established relatively recently as a name that encompasses these congenital malformations (Shprintzen 2005). The disorder will therefore be referred to by 22q11.2DS throughout this PhD thesis.

1.1.1. Incidence of 22q11.2DS

The most frequently cited incidence of 22q11.2DS is 1 in 4000 live births, this estimation is based largely on infants with congenital anomalies leading to genetic testing (Goodship et al. 1998). Since 1993, 22q11.2DS patients are diagnosed based on a cytogenetic approach called Florescent *in situ* hybridization (FISH) (Tezenas Du Montcel et al. 1996; Goodship et al. 1998).

An equal distribution of the disease was found among males and females (Botto et al. 2003); however, a higher rate was found in Hispanics (~1:3800) than other ethnic groups (Botto et al. 2003; Kobrynski & Sullivan 2007).

From an etiologic perspective, 22q11.2 deletions occur as a *de novo* mutation in the vast majority of the patients (~90-95%) (McDonald-McGinn et al. 2001), while the deletion is inherited in the remaining of the 22q11.2DS patients (5-10%) (Scambler 2000).

1.1.2. Molecular genetic of 22q11.2DS

1.1.2.1. Causes of 22q11.2DS

Around 85% of 22q11.2DS patients carry a 3Mb microdeletion which is called the typically deleted region (TDR) (Driscoll et al. 1993; Kitsiou-Tzeli et al. 2005). The remaining cases (15%) carry a smaller deletion ~1.5Mb that is nested within the 3Mb deletion (Guna et al. 2015; Shaikh et al. 2000). There is no observable correlation between the deletion size and the severity of clinical symptoms (Carlson et al. 1997).

1.1.2.2. Complexity of 22q11.2 region

1.1.2.2.1. Low copy repeats of chromosome 22q11.2

The 22q11.2 region is a copy number variant (CNV) hotspot, where CNVs are four times more likely to occur than other regions of the genome (Shaikh et al. 2000). Deletions in chromosome 22q11.2 occur commonly due to the genomic architecture of this region, specifically the region contains eight low copy repeats (LCR22s A-H) which share a high level of sequence identity (Shaikh et al. 2000) (Figure 1- 1). Low copy repeats (LCRs) are DNA sequences that have been previously duplicated via evolution. They are typically 10-300Kb long and homologous by 95-97% (Colnaghi et al. 2011). The eight LCR22s within 22q11.2 region comprise around 11% of the region (Babcock et al. 2003; Bailey et al. 2002).

The 3Mb deletion spans the region of 18,658,219-21,865,185 (hg19) on chromosome 22 that contains four LCR22s A, B, C, and D. While the smaller deletion ~1.5Mb spans the region of 18,658,219-20,519,134 (hg19) on chromosome 22 spanning LCR22s A and B (Shaikh et al. 2000) (Figure 1- 1).

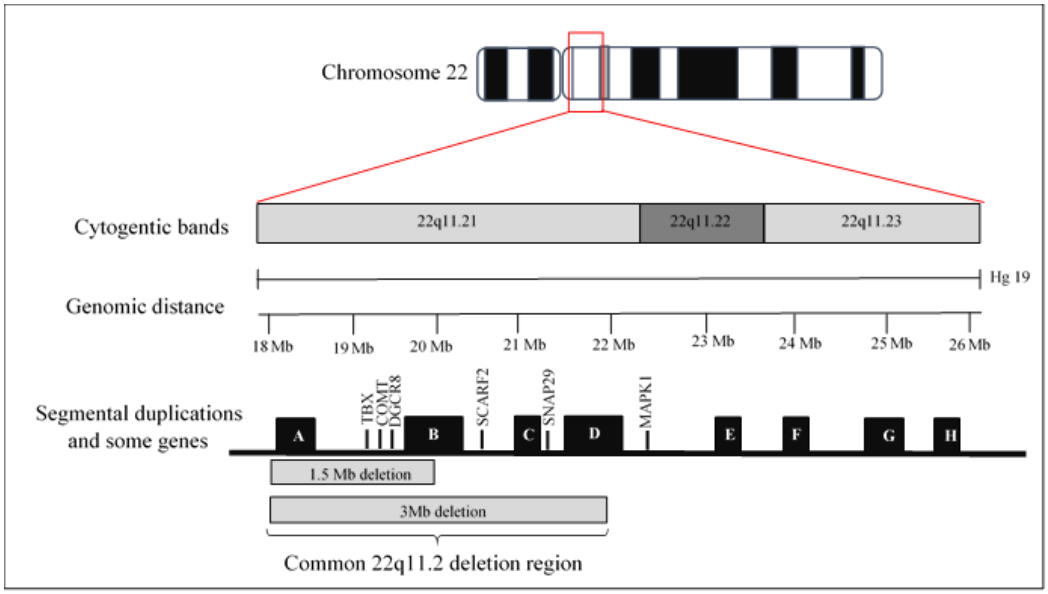


Figure 1- 1: **Schematic ideogram of chromosome 22.**

The figure indicates the position of LCR22s in 22q11.2 region (Black boxes). The common 3Mb deletion associated with 22q11.2DS spans LCR22s A-D, while the less common 1.5Mb deletion associated with 22q11.2DS spans LCR22s A-B, both deletions are indicated by the bottom grey bars (Original figure).

1.1.2.2.2. Non-allelic homologous recombination

Meiosis is a type of cell division in which number of chromosomes is reduced by half to produce gamete cells, eggs and sperm cells, for sexual reproduction (Alberts et al. 2002). During meiosis I, chromosomal crossover takes place in which genetic material is exchanged between two homologous chromosomes (Alberts et al. 2002). Crossover usually occurs when matching regions on homologous chromosomes break and then recombine to exchange genetic materials. This process is called allelic homologous recombination (Alberts et al. 2002) (Figure 1- 2).

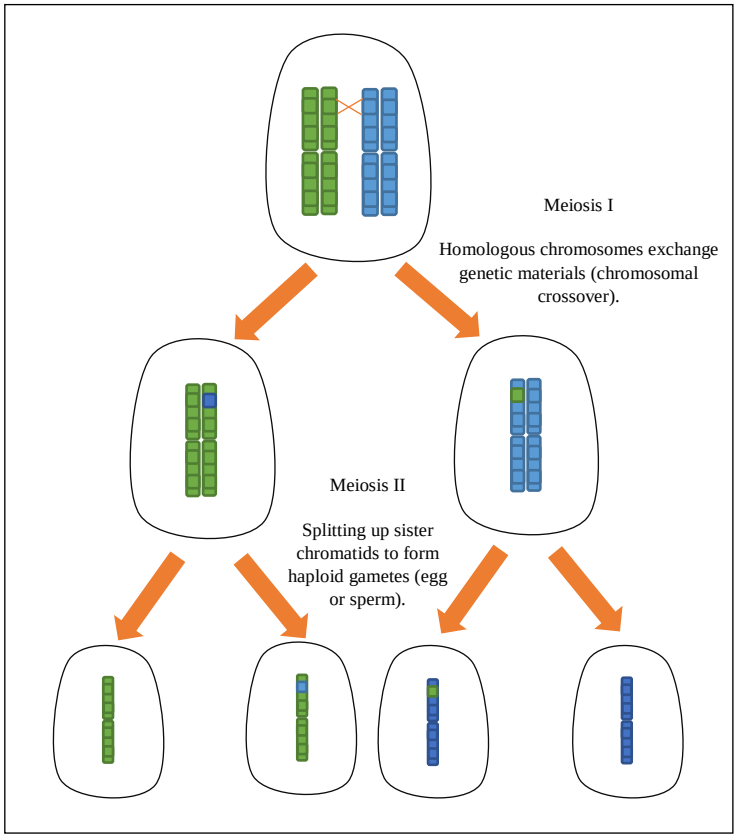


Figure 1- 2: **Normal process of meiosis and crossing over.**

The crossover takes place in meiosis I to exchange genetic materials between 2 homologous chromosomes (Red cross). In meiosis II, the sister chromatids are separated, carrying the exchanged regions, to form an egg cell or a sperm cell (Original figure).

The presence of low copy repeats at regions, such as 22q11.2, results in a high DNA sequence similarity that increases the chances of there being an unequal mispairing and non-allelic crossing over between homologous chromosomes (Shaikh et al. 2000). This abnormal mechanism is called non-allelic homologous recombination (NAHR). NAHR is a form of homologous recombination that occurs between two DNA regions with a high sequence similarity but are not alleles (Edelmann et al. 1999) (Figure 1- 3). This mechanism is a common source of aberrant inter-chromosomal exchanges and results in a high susceptibility for rearrangement (Shaikh et al. 2000).

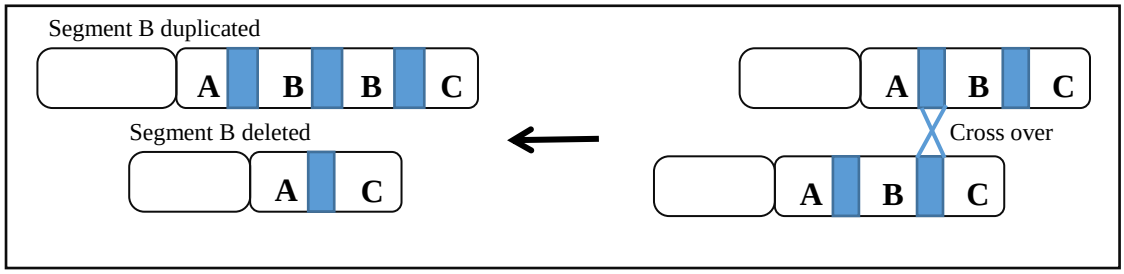


Figure 1- 3: **Schematic description of non-allelic homologous recombination and resulting inter-chromosomal deletion and duplication.**

Blue boxes represent LCRs. In NAHR, the process of crossover occurs between the identical LCR regions in the 2 homologous chromosomes but not between the identical allelic regions as in the normal recombination. Also, this abnormal NAHR unequally exchange the identical LCR regions which leads to a duplicated segment in one homologues chromosome and a deleted segment in the other homologues chromosome (Original figure).

At chromosome 22q11.2, the most common 3Mb deletion is flanked by LCR22A and LCR22D. These are the largest among the eight LCR22s with approximately 240Kb length (Babcock et al. 2003), and are composed of a 160Kb direct repeat that spanned by over 99% identical sequence to each other (McDonald-McGinn et al. 2015; Shaikh et al. 2000).

1.1.2.3. Genes within these deletions

The 3Mb deletion of chromosome 22q11.2 spans approximately 90 genes, and the smaller proximal ~1.5Mb deletion spans around 55 of these genes (Guna et al. 2015). Over half of the 90 genes are protein-coding (n= 46, 51.1%), and most of the 46 genes are expressed in the brain (n= 41, 89.1 %). There are 27 pseudogenes, one read-through transcript (classified as a non-coding RNA) within the TDR (Guna et al. 2015). Also the region contains 9 non-coding RNA genes and 7 microRNA (miRNA) genes, which include MIR185, MIR1306, MIR1286, MIR3618, MIR649, MIR4761, and MI6816 (Figure 1- 4) (Guna et al. 2015).

1.1.2.4. Testing of 22q11.2 deletions

1.1.2.4.1. Cytogenetic testing

The resolution of laboratory techniques to detect the deletion and define the breakpoints vary considerably. Since 1993, FISH has enabled diagnostic clinical laboratories to identify individuals with 22q11.2 deletions by using probes such as N25 or TUPLE/HIRA (Jalali et al. 2008) (Figure 1- 5). However, one main limitation in using FISH is that the probe used in FISH assays can detect the deletion, but it cannot distinguish the 1.5Mb deletion from the common 3Mb deletion (Ramachandran et al. 2015).

1.1.2.4.2. Molecular testing

FISH has been replaced in most laboratories by a range of more sophisticated techniques that are able to detect 22q11.2 deletions of any size such as comparative genome hybridization (CGH), genome-wide microarrays of single nucleotide polymorphisms (SNPs), multiplex ligation dependent probe amplification (MLPA), and quantitative PCR (qPCR) (Miller et al. 2010).

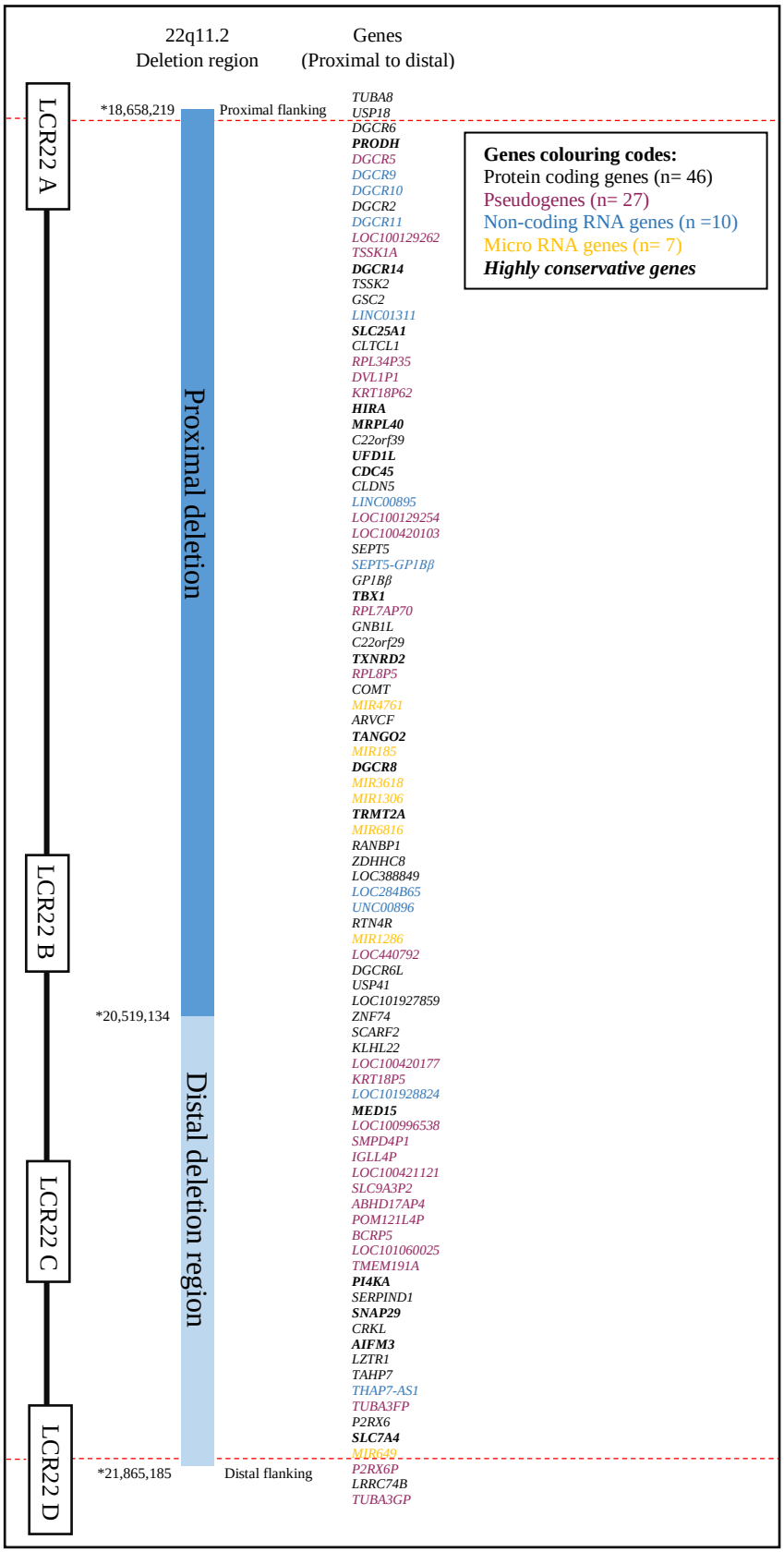


Figure 1- 4: **Genetic landscape of the human 22q11.2 region.**

The typical 3Mb 22q11.2DS deletion spans ~90 genes in the NCBI Reference Sequence Database (RefSeq genes). Region breakpoints are mediated by four chromosome specific LCRs (LCRA-D are indicated by white bxes; approximate locations shown). *Deletion breakpoints (hg19). Figure obtained from (Guna et al. 2015) with modifications.

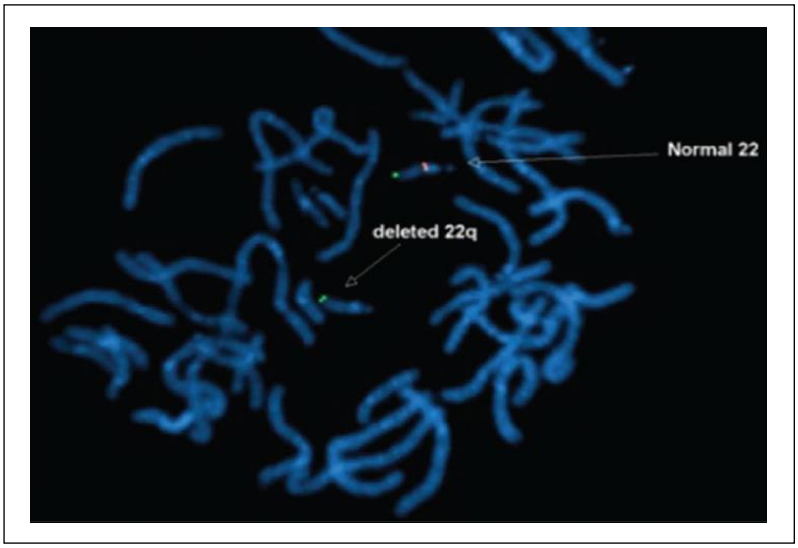


Figure 1- 5: **Dual-colour FISH for detection of chromosome 22q11.2 deletion.**

Interphase cells from patients were hybridized with a chromosome 22q11.2-specific probe, HIRA (red), and a control chromosome 22 probe, ARSA (green). The patient (a 4 months old infant) shows only one red signal with two green signals; this indicates a hemizygous deletion. Figure obtained from (Ramachandran et al. 2015).

1.2. Phenotypic diversity of 22q11.2DS

1.2.1. History of 22q11.2DS

Several incidents of 22q11.2DS had been reported in the medical literature during the 1950s; however, more attention was driven to the syndrome as a clinical diagnosis during 1990s.

The first description of speech abnormality in 22q11.2DS was observed in the earliest cases by Sedlackova in 1955. They were described as having a congenital short palate known as occult sub-mucous cleft palate (Sedláčková 1955). Later, more patients were also reported by Sedlackova in 1967 to have a similar congenital palatal abnormality (Sedláčková 1967). These reported children had hyper nasal speech with absence of overt palatal cleft and distinctive facial appearance (Sedláčková 1967). The occurrence of such abnormalities was thought to be attributed to failure in complete development of the branchial arch, which is the origin of the facial muscles and velums (Sedláčková 1967). Unfortunately, the findings of Sedlackova were published in a Czechoslovakian journal that was inaccessible for the scientific community outside the Czech Republic (Shprintzen 2005).

Further description of 22q11.2DS was observed in the form of immunological disorders and cardiac anomalies. A series of infants were reported with congenital absence of the thymus that leads to immunologic problems, congenital heart disease, and hypoparathyroidism (DiGeorge 1965). Facial dysmorphism was found later to be associated with the other phenotypes (Kretschmer et al. 1968). Accordingly, the term of DiGeorge Syndrome was established to describe patients with absent thymus, immunological disorder, congenital cardiac disease, and facial abnormality (Kretschmer et al. 1968).

In 1968, Strong reported 4 cases within a single family who had cardiac defects associated with a characteristic facial appearance and learning difficulties (Strong 1968). Children with cardiac anomaly with asymmetric crying faces were described to have Cardiofacial Syndrome as a clinical diagnosis for the co-occurrence of these symptoms (Cayler 1969). Cases with similar atypical facial features and heart problems (particularity conotruncal anomalies) were later seen by Kinouchi in Japan. Patients with these symptoms were referred to as having CTAF Syndrome (Kinouchi et al. 1976).

In 1978, Shprintzen reported a series of patients who had an association of cleft palate or velopharyngeal incompetence, heart defects, learning disability, and facial abnormality. These symptoms were collectively called VCFS (Shprintzen et al. 1978).

In 1980s, Shimizu in Japan noticed patients with CTAF had similar clinical features of DGS (Shimizu et al. 1984). A year later, Shprintzen reviewed cases diagnosed with CTAF and suggested they had clinical symptoms of VCFS (Shprintzen et al. 1985).

In 1981-1982, a microscopic deletion of chromosome 22q11.2 was found in patients with DGS (de la Chapelle et al. 1981; Kelley et al. 1982). Ten years later, Driscoll and colleagues identified that only 25% of DGS cases have a microscopic deletion in chromosome 22q11.2 while the remaining 75% of them remained questionable (Driscoll et al. 1992). In 1993, the FISH technique was introduced and Driscoll and colleagues used this method to detect a submicroscopic 22q11.2 deletion in the majority of DGS cases (Driscoll et al. 1993). They also demonstrated that the majority of VCFS patients carry a 22q11.2 deletion, which was similar to the deletion found in DGS patients. This finding explained the overlapping phenotypes in both syndromes and supported the fact that both diagnoses are for the same syndrome (Driscoll et al.

1993). Since then this group of clinical disorders, that share the same deletion have been collectively referred to as 22q11.2 Deletion Syndrome (22q11.2DS) (Bassett et al. 2011; Tang et al. 2015).

1.2.2. Symptoms of 22q11.2DS

Since the original description of the disease, the list of phenotypes associated with 22q11.2DS has been expanding and now it involves more than 180 physical and behavioural symptoms (Demily et al, 2015). Variability of number of associated features in 22q11.2DS is typically broad, ranging in severity from mild to severe (Tang et al. 2015).

1.2.2.1. Physical symptoms

The common physical abnormalities associated with the syndrome are often referred to as CATCH-22, which means **C**ardiac abnormality, **A**bnormal faces, **T**hymic aplasia, **C**left palate and **H**ypocalcemia/**H**ypothyroidism. The 22 indicates the chromosomal abnormality found on chromosome 22 (Wilson et al. 1993). The frequency of their detection in 22q11.2DS is described in Table 1- 1.

Table 1- 1: **Common 22q11.2DS psymptoms** (Kobrynski & Sullivan 2007).

Physical phenotype	Frequency
Cardiac abnormality	49-83%
Characteristic facial appearance	80-100%
Thymus aplasia related immunodeficiency	12-77%
Cleft palate	69-100%
Hypocalcemia	17-60%
Renal abnormality	36-37%
Ophthalmological abnormality	7-70%

1.2.2.1.1. Cardiovascular abnormalities

An average of 75% of 22q11.2DS patients have symptomatic congenital heart disease (CHD) and minor patients have asymptomatic aortic arch malformations (Momma 2010). CHDs are the main cause of mortality in 22q11.2DS and ~90% of deaths in the syndrome are caused by CHDs (Ryan et al. 1997). The most frequent characteristic symptomatic cardiac diseases in 22q11.2DS are conotruncal defects in which the cardiac outflow tract are affected. 22q11.2DS associated conotruncal defects include tetralogy of Fallot (TF), pulmonary atresia with ventricular septal defect (PA-VSD), truncus arteriosus (TA), and interrupted aortic arch (IAA) (Momma 2010). TF is the most prevalent cardiac disease, followed by TF with PA-VSD, VSD, interruption of the aortic arch (type B), and TA (Momma 2010). Age is correlated with type, severity, and prevalence of heart diseases in 22q11.2DS (Momma 2010). Infants presenting with full 22q11.2DS phenotypes tend to have more severe and recurrent CHDs, while those presenting in adolescence tend to have less severe CHDs (Momma 2010).

1.2.2.1.2. Craniofacial abnormalities

Facial abnormality is one of the most common manifestations in 22q11.2DS, which identifies individuals with 22q11.2DS from other symptoms with similar clinical presentation (Oskarsdóttir et al. 2008). However, the characteristic facial appearance is often subtle in infants with 22q11.2DS and becomes more prominent in older children. Therefore, it could not be used as a main sign for genetic testing (Digilio et al. 2003). These features are characterized by longer faces, large nasal tip with hypoplastic nares, small mouth with everted upper lip, cupped or overfolded lower jaw, small dysmorphic ears, and/or periorbital fullness with narrow up-slanted palpebral fissures (Shprintzen 2008) (Figure 1- 6).

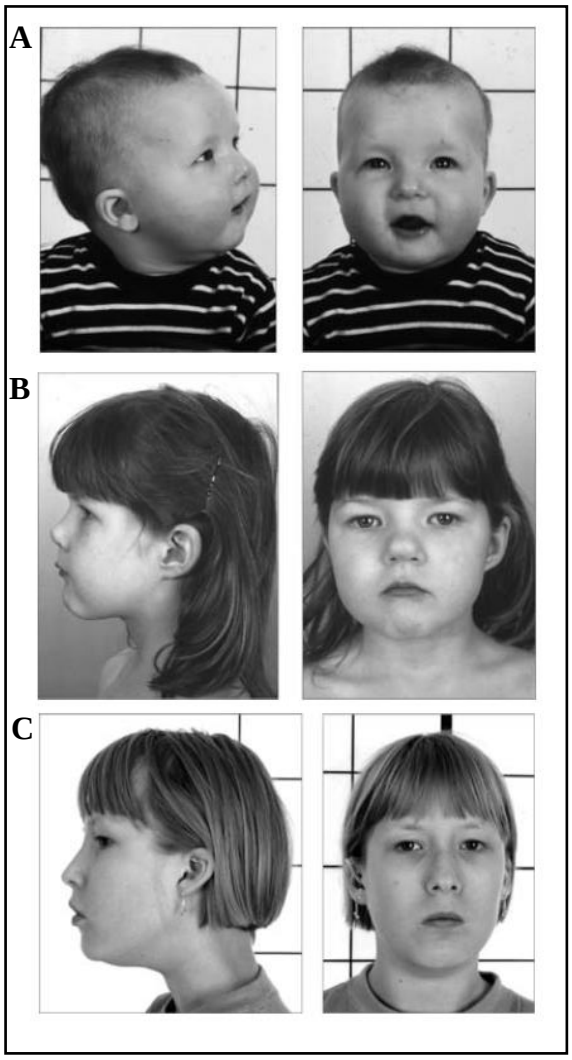


Figure 1- 6: **Some of the facial features associated with 22q11.2DS.**

A) A 10-month old boy with the 22q11DS with hooded eyelids (fullness of the eyelids), a small mouth, a retro/micrognathi, small round/broad ears and a thick helix.
B) A 5-year-old girl with the 22q11DS with a malar flatness, hooded eyelids, hypertelorism, a broad nasal bridge and a broad/round nasal tip, narrow nares, a small mouth and round/broad and slightly low set ears.
C) An 11-year-old girl with the 22q11DS with a malar flatness, hooded eyelids, short palpebral fissures, a broad nasal bridge, a small mouth, small, slightly low-set posterior rotated ears and a thick/overfolded helix.
Figures obtained from (Oskarsdóttir et al. 2008).

1.2.2.1.3. Thymic aplasia/hypoplasia and related immune disorders

Absent or underdeveloped thymus (thymic aplasia or hypoplasia) occurs in more than 80% of 22q11.2DS patients (Piliero et al. 2004). The thymus has a role in maturation of functional white blood cells, particularly T cells, that play a central role in cell-mediated immunity (Sullivan 2005). Thus, 22q11.2DS patients typically have impaired T cell production. Immunodeficiency occurs in 22q11.2DS patients as a consequence of thymic hypoplasia (Sullivan 2005). The degree of immunodeficiency in 22q11.2DS patients is extremely variable. Approximately 0.5-1% of 22q11.2DS patients have severe immunodeficiency with absent or nearly absent T cells, 20-30% have a normal T cells count, the remaining patients range from a mild to moderate defect in their T cell count (Sullivan 2005). In addition, 22q11.2DS is associated with autoimmune diseases such as autoimmune thyroiditis or juvenile rheumatoid arthritis, and also with haematological disorders such as thrombocytopenia purpura (DePiero et al. 1997; Hernández-Nieto et al. 2011).

1.2.2.1.4. Cleft palate and related velopharyngeal abnormalities

22q11.2DS is one of the most common syndromes associated with cleft palate (Scambler 2000). Approximately 5-8% of new-borns with cleft palate had a 22q11.2 deletion (Kirschner 2005). Velopharyngeal abnormalities are among the common clinical features of 22q11.2DS. On average 80% of the patients with the syndrome have some degree of velopharyngeal insufficiency. Also, the syndrome is associated with a wide range of palatal phenotypes including cleft palate, submucosal cleft palate, bifid uvula, absence or hypoplasia of the musculus uvula, or other palatal muscles (McDonald-McGinn et al., 1999). Children with 22q11.2DS have a higher risk of

developing speech disorders than other children with non-syndromic cleft palate or other cleft palate-associated syndromes (Golding-Kushner 2005).

1.2.2.1.5. Hypocalcemia/Hypoparathyroidism

Hypocalcaemia is a consequence of hypoparathyroidism, absent or undeveloped parathyroid glands, that leads to low blood calcium levels (Brauner et al. 2003). Neonatal hypocalcaemia is recognized in most 22q11.2DS children (Ryan et al. 1997). While congenital hypocalcaemia is one of the strongest predictors of 22q11.2DS in new-borns as it occurs in 17-60% of 22q11.2DS patients (Ryan et al. 1997), it can also develop later in childhood or adulthood (Kitsiou-Tzeli et al. 2005). Tetany or seizures can be developed in undiagnosed 22q11.2DS adults as a consequence of hypocalcaemia, which can also be manifested in adults with acute medical conditions or trauma (Perez & Sullivan 2002).

1.2.2.2. Cognitive and behavioural symptoms

Psychiatric disorders associated with 22q11.2DS are amongst the most prevalent manifestations of 22q11.2DS (Antshel et al. 2010; Fung et al. 2010). Like other components of 22q11.2DS, psychiatric, cognitive, and behavioural phenotypes are extremely variable between 22q11.2DS patients with the same underlying 22q11.2 deletion. In childhood, 26% and 41% of 22q11.2DS children develop ASD and ADHD respectively (Niarchou et al. 2014); whereas in adulthood, ~25% of 22q11.2DS adults develop schizophrenia (Murphy et al. 1999). Importantly, similar to the findings of the cardiac phenotype, no evidence showed that the length of the 22q11.2 deletion, or the

co-occurrence of other physical symptoms, contribute significantly to the variability of the cognitive or behavioural phenotype (Bassett et al. 2008; Gerdes et al. 2001).

1.2.2.2.1. Developmental trajectories of developmental, cognitive, and behavioural phenotypes

The brain is the least developed organ at birth, therefore phenotypic features of 22q11.2DS that are associated with abnormal development and function of the central nervous system (CNS) are expected to manifest gradually throughout the patient's life (Philip & Bassett 2011). Meechan and colleagues considered 22q11.2DS as a genetic disorder of neurodevelopment (Meechan et al. 2011). During prenatal or neonatal ages, a small number of CNS abnormalities, other than neonatal seizures related to hypocalcemia, are clinically detectable in 22q11.2DS (Bassett et al. 2005). Developmental delays are commonly observable during infancy (Roizen et al. 2007; Gerdes et al. 2001; Swillen, Devriendt, et al. 1999) including:

1) Motor delays associated with hypotonia (Gerdes et al. 2001; Swillen, Devriendt, et al. 1999).
2) Speech and language developmental delays (Gerdes et al. 2001; Swillen, Devriendt, et al. 1999).
3) Cognitive delays and learning difficulties which are common in 22q11.2DS (Antshel et al. 2010; Chow et al. 2006; van Amelsvoort et al. 2004; Swillen, Devriendt, et al. 1999).

These are identifiable at any stage from the age of pre-school to the age of secondary school and also might affect functioning during adulthood (Swillen, Vandeputte, et al. 1999). The intellectual level of the vast majority of patients with 22q11.2DS range

from an IQ of 70-84 (Antshel et al. 2010; Chow et al. 2006; van Amelsvoort et al. 2004; Swillen, Devriendt, et al. 1999). 33.3% of the patients have a mild intellectual disability, whereas more severe levels are uncommon in 22q11.2DS (Chow et al. 2006; Bassett et al. 2005; Swillen et al. 2000).

The behavioural symptoms in 22q11.2DS include high rates of communication, cognitive, and psychiatric disorders (Swillen, Vandeputte, et al. 1999). 22q11.2DS children and adults manifest poor communication skills, social withdrawal, other behavioural symptoms such as shyness withdrawn and internalizing behaviour, impulsivity, disinhibited behaviour, and attentional problems (Swillen, Vandeputte, et al. 1999). These behavioural problems are the consequences of other disorders which had been reported to occur at an increased rate in 22q11.2DS patients such as ADHD, ASD, and anxiety disorders (Antshel et al. 2006).

Neuropsychiatric phenotypes also change among the course of development in 22q11.2DS (Swillen, Devriendt, et al. 1999). In childhood, attention deficit and anxiety disorders have been reported to be common in 22q11.2DS (Niklasson et al. 2009; Antshel et al. 2010), considerably higher than in the general population (Niklasson et al. 2009). Symptoms of ASD are also common among 22q11.2DS children (Niklasson et al. 2009; Vorstman et al. 2006).

In adulthood, 60% of 22q11.2DS patients have been reported to have treatable psychiatric disorders with no psychotic phenotypes (Fung et al. 2010). Moreover, generalized anxiety disorder is also common in adult patients with a prevalence of two to three times higher than in the general population (Fung et al. 2010). However, the most important psychiatric illnesses with the greatest elevation in risk associated with 22q11.2DS in adulthood is schizophrenia. 20-25% of adult 22q11.2DS patients

developed schizophrenia or related psychotic disorders such as schizoaffective disorder (Murphy et al. 1999; Vorstman et al. 2006; Schneider et al. 2014), which is ~20 times increased compared to the general population (Fung et al. 2010). In addition, multiple studies have reported ~1% of schizophrenic patients in the general population have 22q11.2 deletions (Karayiorgou et al. 1995; Bassett & Chow 2008; Xu et al. 2008; The International Schizophrenia Consortium 2008; Stefansson et al. 2008; Kirov et al. 2009; Rees et al. 2014).

The International Consortium on Brain and Behaviour in 22q11.2 Deletion Syndrome have conducted the largest investigation of the lifetime psychiatric diagnoses in 22q11.2DS (Schneider et al. 2014). The study included 1,402 individuals with 22q11.2DS with different age categories; children (aged 6–12 years), adolescents (aged 13–17 years), emerging adults (aged 18–25 years), young adults (aged 26–35 years), and mature adults (36 years and older). ADHD, ASD, and anxiety disorders were the most common psychiatric phenotypes among 22q11.2DS children, whilst psychosis and mood disorders are more common in adolescence and young adulthood. Similar prevalence and developmental trends were observed accross the countries (Schneider et al. 2014) (Table 1- 2).

Table 1- 2: **Prevalence rates for common psychiatric disorders in 22q11.2DS** (Schneider et al. 2014).

Psychiatric disorders	**Prevalence rates in 22q11.2DS**				
	Children and Adolescents		**Adults**		
	6–12 years	**13–17 years**	**18–25 years**	**26–35 years**	**36^{+} years**
ASD	12.77%	26.54%	16.10%	-	-
ADHD	37.10%	23.86%	15.59%	-	-
Anxiety	35.63%	33.92%	24.07%	24.83%	27.56%
Mood disorders	3.29%	11.85%	18.27%	14.67%	20.47%
Schizophrenia spectrum disorder	1.97%	10.12%	23.53%	41.33%	41.73%

To date, relatively few studies have investigated the neurological symptoms in older 22q11.2DS patients. However, there are some initial indications that early onset Parkinson's disease (PD) is associated with 22q11.2DS (Booij et al. 2010; Zaleski et al. 2009). In a cohort of 68 adults with 22q11.2DS (> 35 years old), four were found to develop PD (Butcher et al. 2013). These patients had neither a family history of PD nor any identifiable known pathogenic risk mutations for PD. Post mortem analysis revealed that in addition to classic loss of midbrain dopaminergic neurons in the four cases, two of them also had α-synuclein-positive Lewy bodies. These findings possibly indicate that 22q11.2 deletions can be a risk factor for PD.

1.3. Identifying 22q11.2 candidate genes for neuropsychiatric phenotypes

The variability in the clinical features seen in 22q11.2DS is not explained by differences in gene content spanned by the deleted region (Philip & Bassett 2011). It is therefore likely that one or more of the genes spanned by the deletion increase risk for the behavioural symptoms and psychiatric disorders associated with in 22q11.2DS (Bassett & Chow 2008; Insel 2010; Meechan et al. 2011). The 22q11.2 region spans over 40 protein genes, and a number of which is expressed in mouse and human brain (Maynard et al. 2003; Meechan et al. 2009) and have roles in brain development, neurotransmitter levels, and myelination (Jungerius et al. 2008; Prasad et al. 2008). Some are therefore strong candidate genes for neuropsychiatric diseases and as such have been tested for association with these phenotypes.

Catechol-O-methyl transferase (*COMT*) encodes an enzyme that plays a role in the degradation of catecholamines including dopamine in the synapse. Two main well-characterized protein isoforms are encoded by this gene which display an altered structure, affinity, and capacity for their substrate (Tunbridge et al. 2006). The shorter cytoplasmic soluble form (S-COMT) predominates and accounts ~95% of total enzymatic activity in most assayed tissues. The other form is a longer membrane bound form (MB-COMT) that is more prevalent in brain tissue and responsible for dopamine inactivation in the surfaces of pre and post synaptic neurons (J. Chen et al. 2011). *COMT* has a nonsynonymous functional polymorphism (rs4680) with G allele substituted by an A allele and results an amino acid substitution (valine>methionine) at codon 108 in S-COMT and 158 in MB-COMT transcripts (Lachman et al. 1996). The outcome of the Val(108/158)Met polymorphism is an alteration in the activity of both COMT isoforms, with the Val COMT form (G variant) having higher activity

than the Met form (A variant). The *COMT* Met polymorphism decreases COMT enzymatic activity by 40% in prefrontal cortical (PFC) tissues (Chen et al. 2004). Although *COMT* is expressed in all brain regions, it has a minor role in dopamine clearance in comparison to the dopamine transporter and subsequent monoamine oxidase metabolism for neuronal synaptic uptake (Gogos et al. 1998). However, in the PFC the level of dopamine transporter expression is low, thus, the effect of active COMT is more relevant in this brain region (Gogos et al. 1998). Accordingly, individuals with a 22q11.2 deletion who carry the less active Met allele are expected to have higher dopamine levels in the PFC brain region which has been suggested could lead to an increased risk for developing psychosis (Egan et al. 2001). Therefore, *COMT* has been a strong 22q11.2 candidate gene for psychiatric disorders (Bearden et al. 2004).

The gene *PRODH* encodes for the proline oxidase or proline dehydrogenases 1, an enzyme that plays a role in degradation of the amino acid proline. Homozygous mutations in *PRODH* result in a rare neurological illness called by hyperprolinemia type 1 which is associated with a highly increased level of proline (Raux et al. 2007).

The genes *GNB1L* and *TBX1* are adjacent genes. Haploinsufficency of which has been correlated with the behavioural impairment observed in 22q11.2DS mice models (Paylor et al. 2006).

The gene *SEPT5* is a member of the septin family that is widely expressed in the brain. It is a negative regulator for an elevated release of neurotransmitter by binding to SNARE-protein (Suzuki et al. 2009).

The gene *RTN4R* encodes NOGO receptor 1 that regulates axonal growth and post-injury axon regeneration. Furthermore, it has been considered as a candidate gene for schizophrenia (Hsu et al. 2007).

ZDHHC8 encodes for an enzyme called palmitoyltransferase. Genetic variants in this gene have been reported to be associated with schizophrenia (Liu, Heath, et al. 2002; Xu et al. 2010).

The protein encoded by *DGCR8* gene has a critical role in the biogenesis of micro-RNAs that play a key function in regulating gene expression. This gene has been implicated to contribute to phenotypic variation in psychiatric and neurodevelopmental disorders (Stark et al. 2008).

The role of these genes in neuropsychiatric phenotypes associated with 22q11.DS have been investigated by either investigating genetic variants in 22q11.2DS patients or by directly assessing the genes function in murine models of 22q11.2DS (Jonas et al. 2014).

1.3.1. 22q11.2DS mice model

Using genetic linkage analysis and detailed physical mapping, Puech and colleagues identified the region on mouse chromosome 16 (MMU16) that is orthologous to the deleted region on human chromosome 22q11.2 (HSA22q11.2). Most of the genes in the 3Mb 22q11.2 region, from *IDD* (centromeric) to *IGL* (telomeric), are located in a single region of mouse chromosome 16, however, the organization of these genes is distinct from that in humans (Puech et al. 1997) (Figure 1- 7). Detailed physical mapping showed that some genes have the same order and orientation in the two species (*IDD*, *TSK*, *ES2*, *GSC*, and *CTP*), the position of some genes have changed

during evolution (*DGR6, LZTR-1,* and *ZNF74*), and while the relative order of some genes has been conserved and their overall orientation has been inverted (*HIRA, UFD1L, CDC45L, TMVCF, CDCREL1, GPIBβ, TBX1, T10, COMT, ARVCF,* and *RANBP1*) (Lindsay et al. 1999; Puech et al. 1997). *ATP6E, CLTD, GNAZ, BRC, GSTTII,* and *MIF* are genes present in HSA22q11.2 but not in its homologous region of MMU16qA13 (Puech et al. 1997). In addition, two genes (*DGCR6* and *DGCR6L*) are present in HSA22q11.2 while only a single gene (*Dgcr6*) is present at MMU16qA13 (Karayiorgou et al. 2010). Similar to the human 22q11.2 locus, the orthologous region on MMU16 is also spanned by LCRs making it highly susceptible to rearrangement (Puech et al. 1997). By targeting multiple contiguous genes, or single genes at the MMU16 orthologous region, a number of mice strains have been engineered (Figure 1- 8).

The first mouse model of 22q11.2DS has been created by deleting a ~1.2Mb region (*Df1*) of mouse chromosome 16 and were called *Df1*+/- mice model. This region spanned 22 functional genes that have human orthologues at 22q11.2 region (Lindsay et al. 1999). Similarly, the *Df(16)1*+/- mouse model has been generated with 23 deleted genes (Lindsay et al. 1999; Paylor & Lindsay 2006). Murine models with longer deletions have also been created: both the *Lgdel*+/- (Merscher et al. 2001) and *Df(16)A*+/− (Mukai et al. 2008; Stark et al. 2008) models harbour a 1.3 Mb hemizygous deletion which included 27 genes. Murine models with smaller deletions have also been engineered: *Smdel*+/- mice model, that carries a 550-kb deletion involved 16 genes without *Tbx1* gene (Puech et al. 1997), also another mice model with a deleted 150Kb portion of the proximal region of the deletion (Kimber et al. 1999).

Moreover, a range of mice models have been engineered by attenuating or deleting a specific candidate gene, which include *Tbx1* (Lindsay et al. 2001), *Comt, Prodh* (Paterlini et al. 2005), *Dgcr8,* and *Zdhhc8* (Mukai et al. 2004).

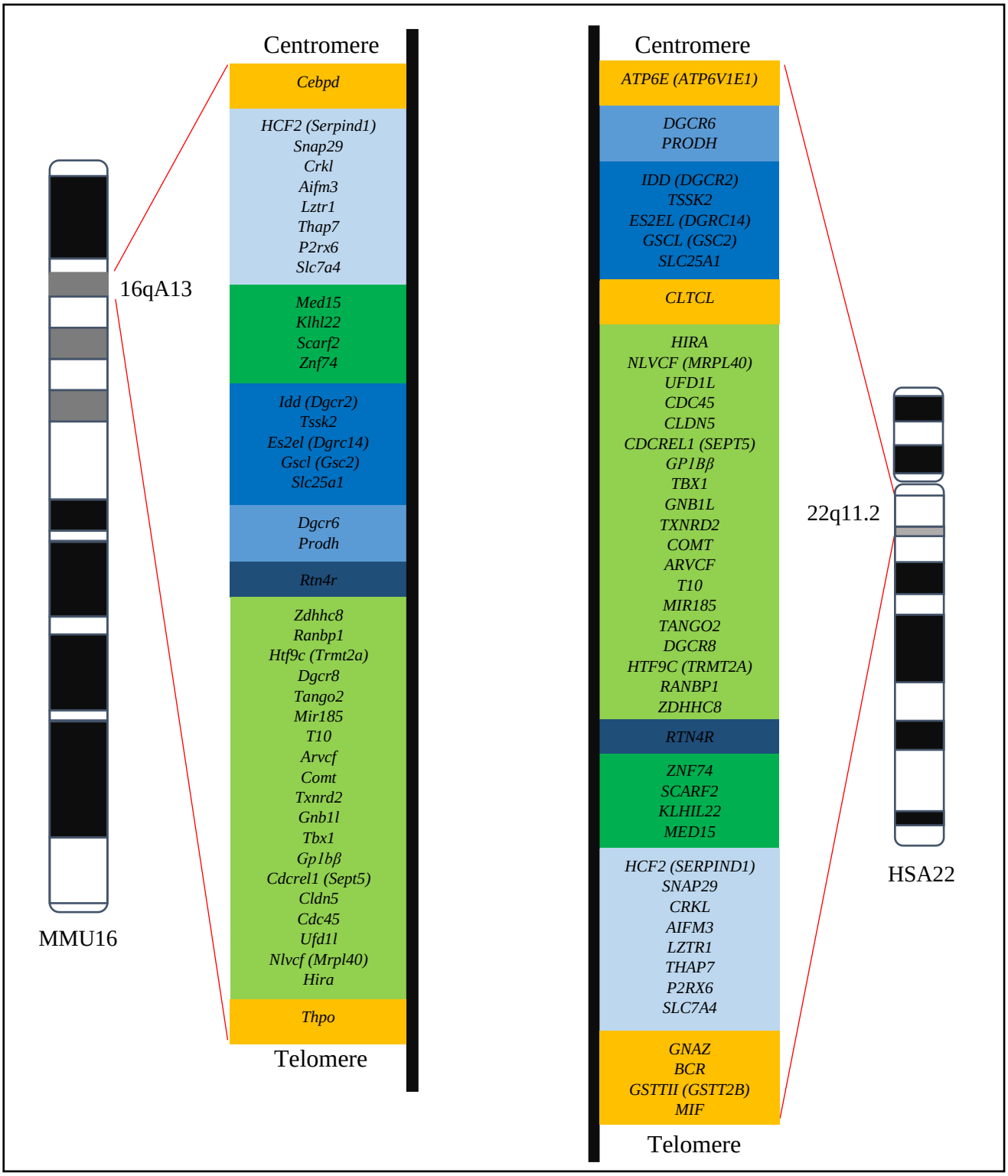


Figure 1- 7: **Human 22q11.2 region and the syntenic mouse 16qA13 genomic region.**

Human chromosome 22 (HSA22) and mouse chromosome (MMU16). The gene content of the human 22q11.2 and mouse 16qA13 is shown in centromeric-telomeric order. Genes in boxes with matching colours in both species indicating they are homologues. Genes in blue boxes (with different blue shadows) have the same orders and orientation but a different location in both species. Genes in green boxes (with different green shadows) have the same order but a reverse orientation and different position in both species. Genes in orange boxes are present in one species and absent in the other. Gene names between brackets are alternative gene names used in UCSC genome browser, The National Center for Biotechnology Information reference (Genome.ucsc.edu, n.d.). Original figure based on the data from UCSC genome browser for mouse and human and the information from (Lindsay et al. 1999; Puech et al. 1997; Paylor & Lindsay 2006).

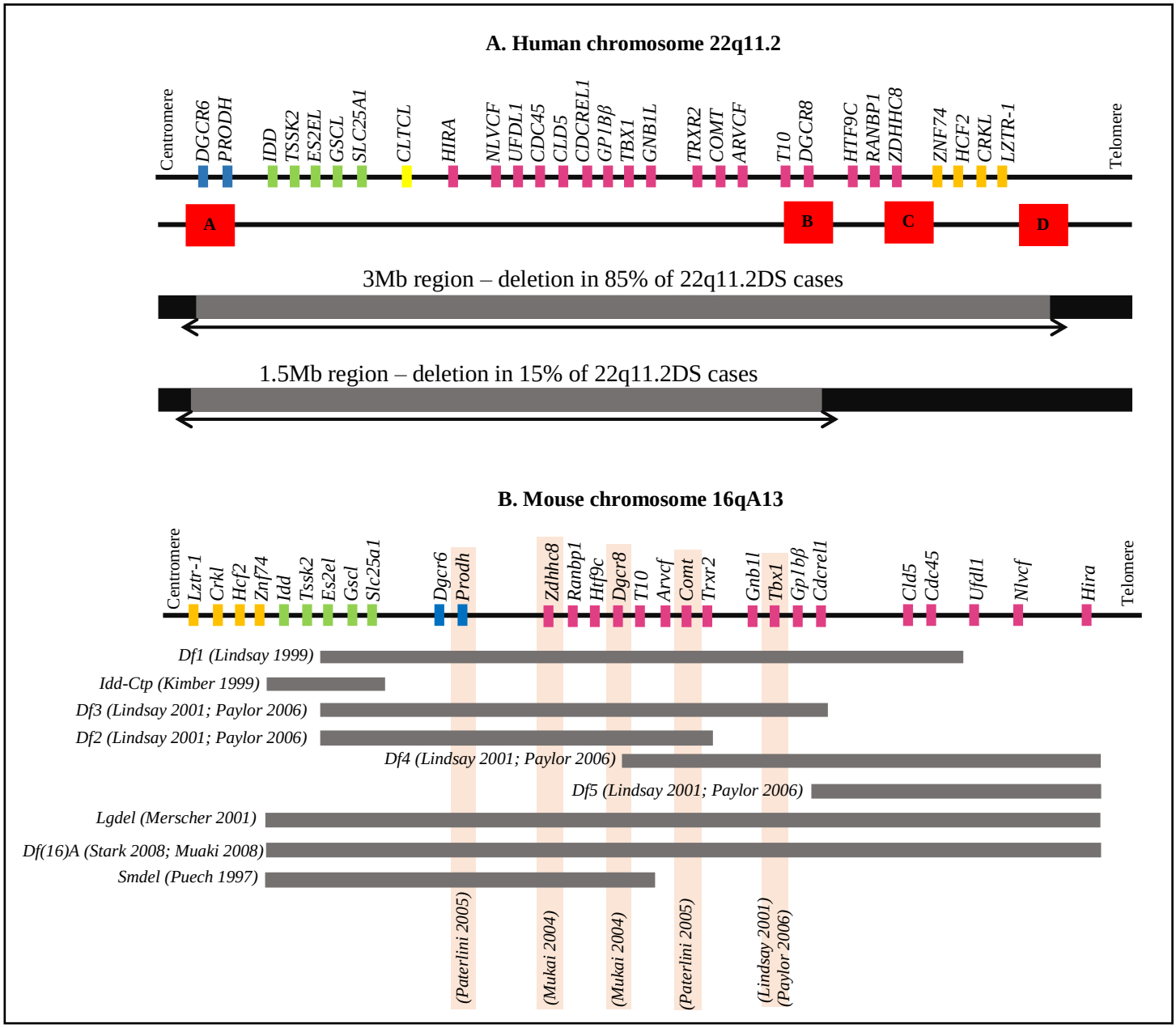


Figure 1- 8: **Human 22q11.2 deletions and 22q11.2DS engineered mice models.**

A. Human 22q11.2 gene content is shown in centromeric–telomeric order. Grey bars indicate the common ~3Mb deletion and the rarer ~1.5Mb deletions found in 22q11.2DS patients. Both deletions are mediated by aberrant homologous recombination between blocks of low copy repeat sequences (Red boxes) distributed through the 22q11.2 region.
B. Syntenic mouse 16qA13 gene content is shown in centromeric–telomeric order. In the mouse, there has been some reshuffling of gene order, orientation, and position with respect to the syntenic human chromosomal region (shown by genes matching bar colours in both species).
Primarily, the order of a large block of genes (*HIRA-ZDHHC8*) is inverted in the mouse genome. *CLTCL* gene has not been identified in the mouse (yellow bar). Grey bars indicate 22q11.2DS generated mice models with various sizes of deletions spanning the homologous 22q11.2 multi genes in mice. The gene content of the various deletions is depicted. Mice models with a single *Df1* gene knockdown are highlighted by light red vertical bars. Along with each mice model, references for studies investigated these 22q11.2 mice models. Original figure based on the information from (Paylor & Lindsay 2006).

1.3.1.1. Studies on *Df1* mice model

1.3.1.1.1. Behavioural and cognitive phenotypes in Df1 mice models

Mice models with homozygous deletions (*Df1-/-*) did not survive and had embryonic lethality (Lindsay et al. 1999). On the other hand, although those with heterozygous deletions *Df1+/-* were viable; however, they showed congenital heart defects similar to those seen in 22q11.2DS (Lindsay et al. 1999), In addition to the congenital abnormalities observed in these mice, they also expressed behavioural abnormalities (Lindsay et al. 1999) (Figure 1- 9).

Due to difficulties in studying 22q11.2DS behavioural deficit in mice models, cognitive functions that can be tested in 22q11.2DS patients as well as in *Df1* mice models were examined. These functions encompassing attention, executive function, working memory, and short-term verbal memory which can represent the activity of the prefrontal cortex and hippocampus (Sobin et al. 2006). Paylor and colleagues reported anomalies of sensorimotor gating in the *Df1+/-* mice, which was demonstrated by impaired pre-pulse inhibition (PPI) (Paylor et al. 2001). This phenotypic feature was also described in patients with schizophrenia and patients with 22q11.2DS (Sobin et al. 2006). Moreover, these mice also expressed memory and learning impairments (Paylor et al. 2001). *Lgdel+/-* mice also showed similar impaired PPI in addition to abnormality in grip strength and nociception (Long et al. 2006). Conversely, mice that are heterozygous for a deletion of proximal 150Kb region of *Df1* showed an increased PPI (Kimber et al. 1999).

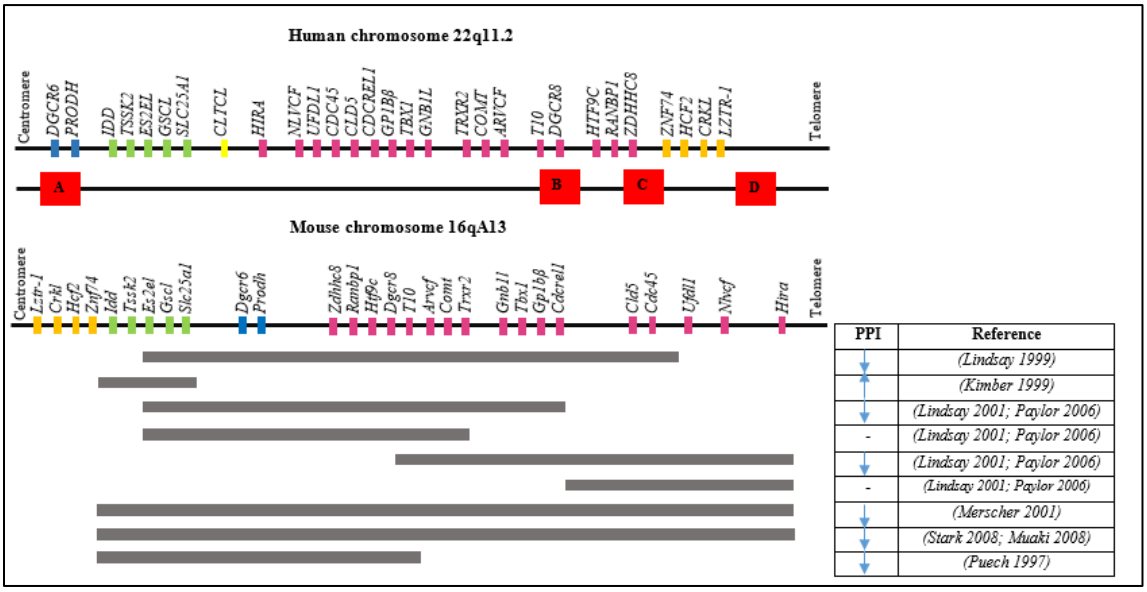


Figure 1- 9: **Pre-pulse inhibition phenotypes in 22q11.2DS mice models.**

Grey bars indicate generated 22q11.2DS mice models with various sizes of deletions spanning the homologous 22q11.2 multi genes in mice. The gene content of the various deletions is depicted. Along with each mice model, PPI phenotypes, either impaired indicated by a downward arrow or increased indicated by an upward arrow, and references for studies investigated these 22q11.2 mice models. Original figure based on the information from (Paylor & Lindsay 2006).

1.3.1.1.2. Gene expression in Df1 mice models

The gene expression studies on the *Df1* mice models showed that MMU16qA13 genes begin to express in mouse embryos and undergo several dynamic expression changes among different embryonic development stages (Amati et al. 2007).

The effect of haploinsufficiency on gene expression was thought to account for the cardiac defects in these *Df1*+/- mice, therefore, they had been studied to investigate gene expression by using microarray and Real Time Quantitative PCR (RTqPCR). Prescott and colleagues extracted RNA from dissected branchial arch region and heart of *Df1*+/- embryos at embryonic day (E) 10.5, during which the arch-artery phenotype is fully penetrant (Prescott et al. 2005). 12 of the genes mapping to the *Df1* deletion were covered by MG-U74a microarray used in this study. 75% (n= 9) of the 12 *Df1* genes were identified as significantly down regulated in *Df1*+/- mice (FDR <0.05).

33.3% of the 75% genes were validated by RTqPCR and showed reduced hemizygous expression. Genome-wide, the analysis showed some of the downregulated genes have a role in vasculogenesis and cardiogenesis including *Connexin 45* and *Dnajb9* (Prescott et al. 2005). These findings confirmed that the cardiac abnormality seen in by *Df1*+/- mice is resulted possibly by some dosage sensitive genes within the *Df1* deletion and some other genes that have a key role in the cardiac pathogenesis (Prescott et al. 2005).

Although the mice with a hemizygous *Df1* deletion have cardiac anomalies similar to those seen in 22q11.2DS patients; however, Lindsay and colleagues noticed that mice with a deletion in one 22q11.2 chromosome and a duplication in the other chromosome (*Df1/Dp1*) have no abnormal cardiac phenotypes (Lindsay et al. 1999). This finding indicated that cardiac defects seen in 22q11.2DS are possibly resulted by the haploinsufficiency of dosage sensitive genes in 22q11.2 region (Lindsay et al. 1999). To identify what particular genes in 22q11.2 region that is associated with cardiac defects, Lindsay and colleagues further investigated *Tbx*1 knockout mice. The results showed a defective development of the 4th pharyngeal arch artery observed in *Tbx1*+/- mice, which suggested that *Tbx1* gene is responsible for heart abnormality in 22q11.2DS (Lindsay et al. 2001). In humans, this finding was confirmed by observing a series of individuals with typical 22q11.2DS cardiac phenotypes; however, no 22q11.2 deletion in these individuals. However, these individuals were found to have a point mutation in the *TBX1* gene (Yagi et al. 2003).

A number of studies have investigated gene expression in the central nervous system of *Df1*+/- mice. Hippocampal sections of 10 week-old *Df1*+/- and wild-type mice were used for RNA extraction for microarray and RTqPCR analyses. The results showed that 57.1% of the analysed *Df1* genes (n= 12/21) have a highly significant

reduction in gene expression with an average expression reduction equal to 33%. The dosage sensitive genes included *Prodh* and *Comt* (Jurata et al. 2006).

Sivagnanasundaram and colleagues confirmed these findings by using RNA samples extracted from the brain of adult *Df1*+/- and wild-type mice (Sivagnanasundaram et al. 2007). 41.7% of the *Df1* genes expressed in the hippocampus (n= 5/12) and were significantly down regulated in the *Df1*+/- hippocampus (p-value <0.05). Globally, 159 other genes were identified to be differentially expressed in the hippocampus of these mice (p-value <0.05) (Sivagnanasundaram et al. 2007).

In another study, the expression of nine 22q11.2 orthologues genes (*Idd, Prodh, Zdhhc8, Ranbp1, T10, Comt, Tbx1, Ufd1l,* and *Hira*) were investigated in the brains or dissected brain regions from later fetal (E10.5 embryos), postnatal, adolescent, and adult *Df1*+/- mice. The results revealed a diminished expression of these genes in these models by 40-60% (Meechan et al. 2006).

1.3.1.2. Studies on a single 22q11.2 gene knocked out mice

Other models of mice, which were deficient for a single 22q11.2 gene, were investigated for association of specific 22q11.2 genes and behavioural phenotypes (Meechan et al. 2011; Paterlini et al. 2005; Gogos et al. 2009; Paylor et al. 2006; Gogos et al. 1998; Hsu et al. 2007; Suzuki et al. 2009).

Most of these models had impairments in PPI and/or learning and memory (Meechan et al. 2011; Paterlini et al. 2005). Mice with haploinsufficient *Dgcr8* and *Zdhhc8* genes showed impaired PPI (Mukai et al. 2004; Gogos et al. 1998). Impairment in PPI in *Dgcr8*+/- and *Zdhhc8*+/- mice was thought to be attributed to the impaired dendritic growth and spine development that were also observed in these mice (Stark et al. 2008;

Mukai et al. 2004). Defects in PPI was also found in mice with homozygous deficiency of *Prodh* gene (Gogos et al. 2009). Defect in sensorimotor gating was observed in mice with knocked down *Comt, Rtn4r*, and *Sept5* genes; however, they had normal PPI (Gogos et al. 1998; Hsu et al. 2007; Suzuki et al. 2009). In addition, it was observed that mice with either haploinsufficient *Tbx1* or adjacent *Gnb1l* have deficit in sensorimotor gating (Paylor et al. 2006).

As many of these mice models expressed defects in CNS functions, it is difficult to pinpoint a major gene being responsible for the behavioural phenotype in 22q11.2DS (Williams 2011). Additionally, genes interaction is highly suggested (Meechan et al. 2011; Paterlini et al. 2005). Studying mice with both haploinsufficient *Prodh* and *Comt* genes showed a deficit in working memory (Paterlini et al. 2005). Strikingly, mice with either haploinsufficient *Prodh* or *Comt* have no working memory deficit (Paterlini et al. 2005). Paylor and colleagues investigated many other overlapping deletions and determined a PPI critical region that compassing four genes including *Gnb1l, Tbx1, Gp1bβ,* and *Cdcrel1* (Paylor et al. 2006). A summary of phenotypes in these mice models are given in Table 1- 3.

Table 1- 3: **Behavioural and cognitive phenotypes of single22q11.2 gene knocked down mice.**

	Dgcr8 +/-	***Zdhh8*** +/-	***Prodh*** +/-	***Comt*** +/-	***Rtn4r*** +/-	***Sept5***		***Tbx1*** +/-	***Gnb1l*** +/-
						+/-	-/-		
PPI									
WM									
SG									
Reference	Mukai 2004 Gogos 1998		Gogos 2009	Gogos 1998	Hsu 2007	Suzuki 2009		Paylor 2006	

PPI: Pre-pulse inhibition
WM: Working memory.
SG: Sensory gating.
Red box indicates low performance.
Grey box indicates normal performance.
Green box indicates high performance.

1.3.2. Analysis of genotypes on the haploinsufficient chromosome in 22q11.2DS

1.3.2.1. *COMT*

Bearden and colleagues genotyped the *COMT* Val158Met polymorphism in 44 children with 22q11.2DS who had been subjected to neurocognitive testing (Bearden et al. 2004). Children with the Met allele revealed significantly better performance on executive functions composite score than those with Val allele (p-value <0.001). However, performance on verbal fluency and arithmetic tasks showed no difference between groups (p-value >0.05) (Bearden et al. 2004). Despite this finding being supported by 2 independent studies (Kates et al. 2006; Shashi et al. 2006), analyses of much larger sample sizes have either found evidence for association in the opposite direction (Gothelf et al. 2005; Bassett et al. 2007), or have failed to find evidence for association with neurocognitive measures (Baker et al. 2005; Glaser, Debbane, et al. 2006; van Amelsvoort et al. 2008).

A number of studies have investigated polymorphisms at *COMT* for association with psychiatric illness. In general, the findings have failed to identify significant evidence for association with psychiatric disease. In one of the earliest studies, Murphy and colleagues (Murphy et al. 1999) analysed 48 adults with 22q11.2DS, of which 12 individuals had schizophrenia, and did not identify any significant evidence for the *COMT* Val158Met polymorphism being associated with psychosis (p-value= 0.70). In addition, the study revealed no correlation between *COMT* allelic distribution and symptom severity measured by a schizotypy scale of all 22q11.2DS individuals (p-value = 0.50) (Murphy et al. 1999). Similarly, using an independent samples of 73 and 92 Caucasian 22q11.2DS adults, Bassett and colleagues (Bassett et al. 2007), and Raux and colleagues (Raux et al. 2007) failed to show any significant evidence for association with schizophrenia.

1.3.2.2. *PRODH*

Homozygous mutations in the gene *PRODH* result in hyperolinemia type 1, which is a rare neurologic disorder associated with variable phenotypes (Raux et al. 2007). The coding SNP rs4819756 has been associated with variable PRODH activity, with the A-allele resulting in a reduction in PRODH activity by ~45% (Bender et al. 2005). Recently, Radoeva and colleagues, investigated genotyped rs4819756 in a cohort of 22q11.2DS individuals with (n= 19) and without (n=67) ASD (Radoeva et al. 2014). The results revealed no significant association with ASD diagnosis (p-value >0.05) (Radoeva et al. 2014).

1.3.2.3. *PIK4CA*

Three SNPs (rs2072513, rs165862 and rs165793) located at *PIK4C4* were initially reported to be significantly associated with schizophrenia in a sample of 310 schizophrenic patients and 880 controls (Jungerius et al. 2008). As this gene is spanned by the common 22q11.2 deletion, a number of studies investigated their association with schizophrenia in patients with 22q11.2DS. The initial study of a Canadian adult cohort (n= 79) revealed significant evidence for a single marker (rs165793-G allele) being associated with schizophrenia in 22q11.2DS (p-value= 0.006), which also provided support to the hypothesis that the effect size would be increased in the 22q11.2DS cohort (OR= 9.47 (1.16-77.56)) (Vorstman, Chow, et al. 2009). However, despite this encouraging result, a subsequent independent study of 83 individuals with 22q11.2DS, of which 24 individuals had experienced psychosis, was not able to identify any significant evidence for association with schizophrenia for either rs165793 (p-value= 0.28) or any SNPs within *PIK4CA* (p-values >0.05) (Ikeda et al. 2010).

1.3.2.4. *GNB1L*

In an association study of 83 subjects with 22q11.2DS with and without psychosis, SNPs mapped to *GNB1L* (rs2269726 and rs5746832) were genotyped. The results revealed a significant allelic association of rs2269726 and rs5746832 with psychosis which was selective for males (p-value= 0.018) (Williams et al. 2008). By using allele specific expression analysis, rs2269726 was also correlated with significant differential allelic expression at *GNB1L* (p-value= $6.3x10^{-7}$); rs2269726 "T" and rs5746832 "G" alleles were correlated with high expression of *GNB1L* (Williams et al. 2008). These findings however require independent replication.

Overall, attempts to identify small critical segments or SNPs on individual genes within the 22q11.2 region have not been conclusive. In addition, it is hard to predict whether the weak association of a SNP with a psychiatric phenotype in 22q11.2DS reflected the weak impact of that SNP on the protein function, or the weak contribution of the protein in that psychiatric phenotype (Philip & Bassett 2011).

1.3.3. Evaluation of 22q11.2 genes as candidate susceptibly genes for common psychiatric disorders in the general population

Several studies have attempted to test common genetic variants mapped to 22q11.2 region for association with neuropsychiatric disorders in large case/control cohorts who do not carry deletions at 22q11.2. However, results of these studies showed inconsistent findings.

A large number of studies have tested the *COMT* Val(108/158)Met polymorphism for association with schizophrenia. Some studies have observed a slightly elevated *COMT*

Val allele frequency in schizophrenic patients (Glatt et al. 2003; Fan et al. 2005), however, this finding was not observed in other studies (Williams et al. 2007; Nieratschker et al. 2010; C. Chen et al. 2011). In addition, no significant association was identified in 3 meta-analyses of published schizophrenia case/control data (Munafò et al. 2005; Glatt et al. 2003; Fan et al. 2005). Furthermore, larger studies of cohorts that were not involved in the meta analyses also showed no evidence for *COMT* association in 22q11.2DS European (Williams et al. 2005), and Asian (Fan et al. 2005) populations. Studies investigating the association of the *COMT* Val(108/158)Met polymorphism with prefrontal cognitive functions have also showed inconsistent results; with some reporting an association between the *COMT* Met allele with better cognitive functions (Egan et al. 2001; Caldú et al. 2007) and others reporting an association with worse cognitive performance (Caldú et al. 2007; Wang et al. 2013).

Three SNPs (rs2072513, rs165862, and rs165793) at *PIK4CA* were initially found to be significantly associated with schizophrenia in a cohort of 310 Dutch schizophrenic patients and 880 Dutch controls (Jungerius et al. 2008). This finding has been partially supported by the International Schizophrenia Consortium where a single SNP (rs165862) was associated with schizophrenia with p-value< 0.0067 in a cohort of European individuals. However, this finding was not replicated in an independent cohort of individuals with a non-European background (Kanahara et al. 2009).

Association analysis of genetic variants at the *TBX1/GNB1L* locus identified a significant excess of homozygous genotypes in schizophrenic patients at markers rs5746832 and rs2269726 (an excess of both TT (OR= 1.28) and CC (OR=1.36) homozygotes). This association was stronger in male schizophrenic patients (Williams et al. 2008). Further investigation was conducted to study if this genetic association

could be link to gene expression showed that the "T" allele at rs2269726 was associated with a 20% increase in *GNB1L* expression when compared to the "C" allele. It was suggested that the increased risk of psychosis in these patients reflected dosage sensitivity of *GNB1L* gene (Williams et al. 2008).

Two studies have analysed genetic variants mapped to *TBX1* locus in two populations with different ethnic backgrounds (Funke et al. 2007; Ma et al. 2007). The results revealed no evidence for an association of *TBX1* variants with patients with psychiatric phenotypes compared to controls.

Significant genetic association has also been found between functional polymorphisms at *PRODH* and schizophrenia (Kempf et al. 2008; Li et al. 2004; Liu, Abecasis, et al. 2002). However, these findings have not been supported by independent replication studies (Fan et al. 2003; Glaser, Moskvina, et al. 2006; Glaser, Debbane, et al. 2006; Ohtsuki et al. 2004; Williams et al. 2003).Two independent studies initially reported significant association between variants at *ZDHHC8* and schizophrenia (Liu, Abecasis, et al. 2002; Mukai et al. 2004). However, these findings were not supported by further studies in Japanese and European populations (Glaser, Moskvina, et al. 2006; Otani et al. 2005; Saito et al. 2005). Xu and colleagues conducted a meta-analysis and identified that *ZDHHC8* variants were negatively associated with schizophrenia (Xu et al. 2010).

Only a single study has reported a significant association of schizophrenia and a functional polymorphism in *UFDIL* gene (De Luca et al. 2001).

Chen and colleagues failed to provide significant evidence for association between four SNPs at *ARVCF* and schizophrenia (Chen et al. 2005); however, Mas and

colleagues reported a significant association of rs165815, a functional variant in exon 15 of *ARVCF*, in a Caucasian sample from Catalonia (Mas et al. 2010).

1.4. Possible molecular mechanisms underlie the neuropsychiatric phenotypes variability in 22q11.2DS

There is strong evidence that deletions at 22q11.2 confer an increased risk to develop psychiatric disease. This thesis will consider 3 molecular mechanisms that have been suggested to potentially be involved in disease pathogenesis (Williams 2011):

1) Gene dosage sensitivity of the genes spanned by the deletion.
2) Presence of positional effects on nearby genes that are influenced by the deletion.
3) Presence of additional genetic risk variants that are independent to the primary 22q11.2 deletions that are located either i) on the non-deleted allele, or ii) elsewhere in the genome.

These molecular mechanisms will be described in details in the result chapters of this PhD study.

1.5. Thesis aims and objectives

The main aims of this study are: 1) to investigate the molecular mechanisms that possibly influence the increased psychiatric phenotypes risk in 22q11.2DS, 2) to investigate the involvement of 22q11.2 deletions in the risk of Parkinson's disease.

The work in chapter three aims to obtain a cohort for 22q11.2DS children, which are independent; ethnically homogenous; and true 22q11.2 deletion carriers, that can be used for the subsequent analyses conducted in this PhD thesis.

Chapter four aims to investigate gene expression in 22q11.2DS. The work firstly will assess the gene expression of haploinsufficient genes located within 22q11.2 to highlight dosage sensitive genes, which are potentially strong candidates for future investigations into the behavioural phenotype associated with 22q11.2DS. Secondly, expression of genes located outside 22q11.2 will be assessed to investigate positional effect that is possibly conferred by the deletion on adjacent genes. Thirdly, biological pathways that are potentially affected by the 22q11.2 deletion will be investigated by analysing genome-wide differential gene expression in 22q11.2DS. Lastly, changes in gene expression between 22q11.2 deletion carriers affected and unaffected with a psychiatric disorder will be investigated to identify genes that possibly play a direct role in the increased risk to psychiatric disease.

The work in chapter five aims to systematically screen genetic variants in the remaining, non-deleted chromosome of 22q11.2 to identify those that are associated with the increased risk of psychiatric disorders in 22q11.2DS.

In chapter six, we aim to investigate whether the presence of second large, rare CNVs in 22q11.2DS children is related to the increased risk of childhood neuropsychiatric phenotypes in these patients.

The work in chapter seven aims to investigate the association of 22q11.2 deletions with Parkinson's disease, particularly the early onset form of the disease.

Chapter 2: Materials and Methods

2.1. Participants

22q11.2DS participants analysed in this PhD study were recruited by two main studies: the Cardiff ECHO study (Study of **E**xperien**C**es of people wit**H** c**O**py number variants) and the London BBAG study (Study of **B**rain, **B**ehaviour **A**nd **G**enetics in 22q11.2 deletion syndrome).

The Cardiff ECHO study is an ongoing study at the MRC centre for neuropsychiatric genomics and genetics in Cardiff University. The principal investigators are Prof Marianne van den Bree and Prof Michael Owen. ECHO participants are referred from 11 genetics clinics within the UK, as well as a number of charities. The ECHO field team recruited ~92 22q11.2DS children and ~71 control siblings at the time of starting this PhD. The 22q11.2DS children had a mean age of 10.2 years (SD= 2.1), age range: 6–14 years. The psychopathology, cognition, and behavioural abnormalities were assessed in these individuals by the ECHO field team (Niarchou et al. 2014; Niarchou et al. 2015). Biological samples, for DNA and RNA extractions, were also collected during the field team visits.

The London BBAG study is ongoing at King's College London and is led by Dr Michael Craig. 22q11.2DS children and non-deleted controls were recruited mainly for brain imaging studies. Blood samples, for DNA and RNA extractions, were also collected from these individuals during their imaging sessions. Total 21 individuals were recruited including 3 22q11.2DS cases and 18 controls. 22q11.2DS children have a mean age of 13.6 years.

In total there were 166 samples available for DNA extraction (95 children with 22q11.2DS and 71 non-deleted controls), and 79 samples available for RNA extraction (36 children with 22q11.2DS and 43 non-deleted controls) from both studies.

2.1.1. Biological samples collection

For each recruited individual, the field teams aimed to collect two biological samples, one for DNA extraction and the other for RNA extraction. Blood or saliva samples were collected for DNA extraction. Blood samples were collected either in PAXgene tubes or sodium EDTA tubes and accordingly different DNA extraction kits were used based in the collected tubes. A blood sample was collected in a PAXgene tube for RNA extraction. A summary of the number of samples collected in both ECHO and BBAG studies for DNA and RNA extractions is given in Table 2- 1.

Table 2- 1: **Number of samples collected by ECHO and BBAG studies for DNA and RNA extractions.**

	For DNA extraction				**For RNA extraction**	
	Blood samples		**Saliva samples**		**Blood samples**	
	22q11.2DS	Controls	22q11.2DS	Controls	22q11.2DS	Controls
Cardiff ECHO study	79	62	13	9	33	25
London BBAG study	3	13	0	0	3	18
Total	82	75	13	9	36	43

2.1.2. Psychiatric illnesses assessment

Neuropsychiatric symptoms including ADHD and ASD, were examined in ECHO participants only. Psychiatric data was not available for BBAG participants.

2.1.2.1. ASD diagnosis

ASD phenotypes were assessed by the ECHO field team by using the Social Communication Questionnaire (SCQ) (Rutter, Bailey & Lord 2003), which was completed by the primary caregiver of the 22q11.2DS child.

The SCQ questionnaire consists of 40 questions obtained from the Autism Diagnostic Interview-Revised algorithm (ADI-R), which is an evaluation algorithm for several phenotypes appearing in autism patients such as reciprocal social interaction, communication, restricted, repetitive, and stereotyped patterns of behaviours (Kim & Lord 2002). The questions included in the SCQ questionnaire assess both current behaviour and developmental history between the ages of 4 to 5. The questions cover four domains of phenotypes seen in autistic patients as follow;

A: Qualitative Abnormalities in Reciprocal Social Interaction

B: Qualitative Abnormalities in Communication

C: Restricted, Repetitive, and Stereotyped Patterns of Behaviour

D: Abnormality of Development Evident at or Before 36 Months

ASD status of the ECHO 22q11.2DS children is mainly based on the SCQ. Total SCQ scores can range from 0 to 39. A cut-off of 15 suggests ASD and a cut-off of 22 is used to distinguish autism from other disorders on the spectrum (Berument et al. 1999). 22q11.2DS participants have to get a score of 15 or above to meet ASD criteria.

The average SCQ score among the recruited ECHO 22q11.2DS participants (n= 79) is 12.3 and the standard deviation equals to 6.8. Around 23.7% of the 22q11.2DS participants have an SCQ score $\geq$ 15 and diagnosed to have ASD; while ~61.8% are

healthy and have an SCQ score < 15. Summary statistics for SCQ scores in 22q11.2DS children affected and unaffected with ASD are given in Table 2- 2.

Table 2- 2: **Summary statistics of SCQ scores in 22q11.2DS children affected and unaffected with ASD.**

	22q11.2DS+ASD	22q11.2DS-ASD
SCQ mean	21.0	8.7
SCQ standard deviation	3.8	4.6

2.1.2.2. ADHD diagnosis

For ADHD diagnosis, the ECHO field team used the Child and Adolescent Psychiatric Assessment (CAPA) (Angold et al. 1995). It was used to obtain a DSM-IV-TR research diagnosis of ADHD for the 22q11.2DS participants. The CAPA is a semi-structured interview that provides categorical diagnoses in addition to symptom counts of psychiatric disorders. Symptom counts were derived from responses to questions about worries, depression, sleep, and oppositional defiant disorder (ODD) behaviour/conduct problems. All interviews were undertaken by trained psychologists, who were supervised by a consultant child and adolescent psychiatrist.

Also, the CAPA examines inattention, and hyperactivity and impulsivity sub-types of ADHD. All 22q11.2DS children who were diagnosed to have ADHD symptoms met DSM-IV criteria for ADHD.

~ 38.2% of the ECHO 222q11.2DS children (n= 79) were identified to have ADHD phenotypes, while ~ 55.3% of them are unaffected with ADHD.

2.1.3. Cognitive assessment

Neurocognitive assessments were conducted by trained raters in the ECHO filed team. The general intelligence (IQ) was estimated based on the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler 1999). WASI comprised of four subtests; two verbal tasks, vocabulary and similarities, and two non-verbal tasks, block design and matrix reasoning. From these tests, verbal IQ (VIQ), performance IQ (PIQ), and full scale IQ (FSIQ) were derived.

The children with 22q11.2DS children had a lower estimated total IQ (mean= 76.76; SD= 13) compared to their non-carrier siblings (mean= 108.56; SD= 15.6) (Niarchou et al 2014). Additionally, 22q11.2DS children performed worse than their normal siblings on all cognitive tests (Niarchou et al 2014). However, there was no IQ differences between 22q11.2DS children with ASD and those without ASD (Table 2-3). Also, no IQ difference was observed between 22q11.2DS children with and without ADHD (Table 2- 3).

Table 2- 3: **Summary statistics of IQ scores in 22q11.2DS children affected and unaffected with psychiatric illnesses (ASD or ADHD).**

	ASD		ADHD	
	22q11.2DS+ASD	**22q11.2DS-ASD**	**22q11.2DS+ADHD**	**22q11.2DS-ADHD**
IQ mean	77.4	77.5	78.08	76.8
Standard deviation	13.8	13.8	13.7	14.04

Therefore, as this thesis will compare the molecular changes between the 22q11.2DS children affected and unaffected with psychiatric phenotypes, ASD or/and ADHD, the observed molecular differences are more likely to be the responsible for the psychiatric phenotypes in these children rather than the reduced IQ.

2.2. DNA samples

2.2.1. DNA sample extraction and storage

Saliva samples were collected for DNA extraction by the Cardiff ECHO study used the Oragene DNA kit (DNA genoteck), while blood samples were collected in sodium EDTA tubes. All samples collected for DNA extraction by the London BBAG study were collected in the PAXgene DNA blood tubes. Kits for DNA extraction were selected based on the method of samples collection.

2.2.1.1. Illustra Nucleon BACC3 DNA Kit

All blood samples collected for DNA extraction by the Cardiff ECHO study were extracted by using Illustra Nucleon BACC3 Genomic DNA Extraction kit (GE Healthcare life sciences) according to the manufacturer's recommendation. Briefly, DNA extraction was initially established by cell lysis then de-proteinisation with sodium perchlorate. Following treatment with chloroform and nucleon resin the DNA was precipitated in 100% ethanol and then washed in 70% ethanol. DNA molecules were then re-suspended in a TE buffer and stored at -20-25^{o}C.

2.2.1.2. PAXgene DNA Kit

Blood samples collected by the London BBAG study for DNA extraction were collected, stored and transported to Cardiff in PAXgene tubes. DNA molecules were extracted from these samples by using the combined PAXgene Blood DNA System kit (QIAGEN) by following the manufacturers recommended protocol. Briefly, the DNA isolation procedure using this kit began with degrading red and white blood cells by cell lysis buffer followed by washing and a re-suspension in a digestion buffer. De-

proteinisation then was done by protease enzyme. DNA molecules were collected by isopropanol precipitation and washed by 70% ethanol. DNA molecules were then re-suspended in a re-suspension buffer provided by the manufacturer and stored at -20-25^{o}C.

2.2.1.3. Oragene® DNA Kit

All saliva samples collected in Oragene® DNA vial for DNA extraction by the Cardiff ECHO study were extracted by using Oragene® DNA Kit. DNA samples properly mixed with the Oragene® DNA solution followed by a heat-treatment step to maximise DNA yield and to inactivate nucleases. Then Oragene® DNA Purifier is added followed by incubating the sample at room temperature then centrifuging to precipitate impurities and inhibitors. The DNA-containing supernatant is transferred to a fresh tube to precipitate the DNA molecules by adding 95-100% ethanol. DNA sample is washed then by 70% ethanol and dried. Finally extracted DNA is rehydrated by TE buffer and stored at -20-25^{o}C.

2.2.2. DNA sample quantification and quality control

DNA samples were quantified and inspected for quality control using a Spectrophotometer (NanoDrop 8000 instrument) and Quant-iT PicoGreen® ds DNA assay kit.

2.2.2.1. Spectrophotometer DNA quantification

Extracted DNA samples were initially quantified using a NanoDrop 8000 instrument (Thermo Scientific) that measured the absorbance (A) of UV light at 260nm and 280nm wavelengths (λ). DNA concentration was calculated on the assumption that an A_{260} value of 1 was equivalent to 50µg/ml of DNA. A ratio of A_{260} to A_{280} equals to or larger than value of 1.8 is generally accepted as high quality DNA and absence of contaminating protein. If the ratio is lower than 1.8, it may indicate presence of contaminants such as protein, phenol or other contaminants that absorb strongly at or near 280nm.

2.2.2.2. PicoGreen® DNA quantification

Accurate DNA sample quantification was performed using a TECAN fluorometer (infinite F200 pro) and Quant-iT PicoGreen® ds DNA assay kit (Invitrogen).

Initially, all DNA samples were diluted to ~50ng/µl based on spectrophotometer readings in a total volume of 20µl. Sample aliqouts were then diluted in 1X TE buffer in a black 96 well cliniplate to get a final DNA concentration of 0.6-1.2ng/µl. In parallel, a Picogreen working solution was prepared by adding 5µl of Picogreen to 995µl of 1x TE buffer.

100µl of the Picogreen working dilution was added to each diluted DNA sample. The TECAN fluorometer measured the DNA concentration of a sample using an UV excitation wavelength of 485nm and an emission wavelength of 535nm. A standard curve that had been prepared for several dilutions of standard dsDNA, was then used to calculate the concentration of DNA for each sample. The original samples were then adjusted so each was at concentration of 400-800ng in a total volume of 8µl.

2.3. DNA sample genotyping

2.3.1. Genotyping

The microarray platform used for genotyping 22q11.2DS samples was the Illumina Infinium Human CoreExome-24 BeadChip. This genotypes a total 551,839 SNPs, which include 300,000 common and highly informative SNPs that cover the entire genome. The array also included an additional 240,000 SNPs, which are exome focused markers and 27,000 custom markers that were selected because of their relevance to neuropsychiatric and neurological disorders.

2.3.2. Pre-genotyping samples quality control

22q11.2DS participants were selected for genotyping based on the quality of their DNA samples. Only DNA samples with high quality, including high DNA purity (A_{260}/A_{280} ratio $\geq$1.8), and adequate DNA concentration, were selected for the analysis. Samples with a concentration lower than the required working concentration and volume for genotyping microarrays (200-400ng in 4µl) were excluded. Summary for DNA samples quality control is illustrated in Figure 2- 1.

2.3.3. DNA samples preparation

All DNA samples had a high quality with adequate concentration and volume, therefore, they were optimized to the working DNA concentration for the genotyping microarray (200-400ng). The manufacturer recommends using DNA samples at a concentration of 50-100ng/µl in total volume of 8µl, thus, total amount of DNA used was 400-800ng. Only 4µl (200-400ng) of each sample aliquot was processed for genotyping and loading on the chip.

2.3.4. Illumina genotyping protocol

The Illumina infinium assay protocol is based on direct hybridization of isothermally amplified and fragmented DNA samples to locus-specific probes. After hybridization, allelic specificity is conferred by enzymatic base extension, and the products are fluorescently stained and detected by the Illumina iScan instrument. Then, the genotyping module (version 1.9.4) on GenomeStudio (v2011.1) was used for genotype calling and signal intensity data normalization. An overview of this protocol is given in Figure 2- 2.

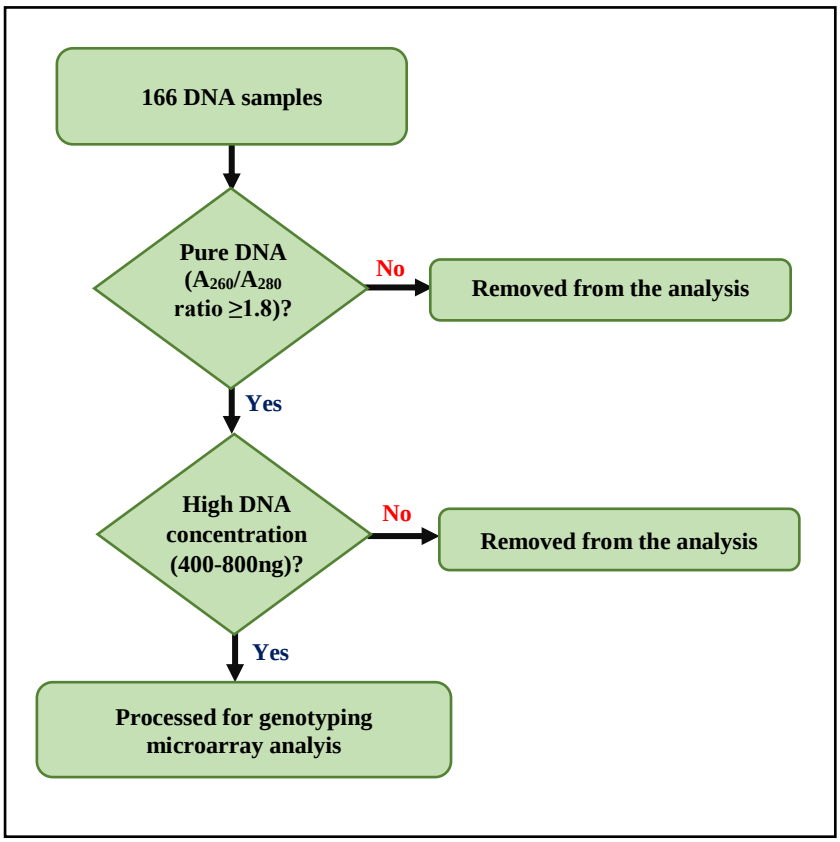
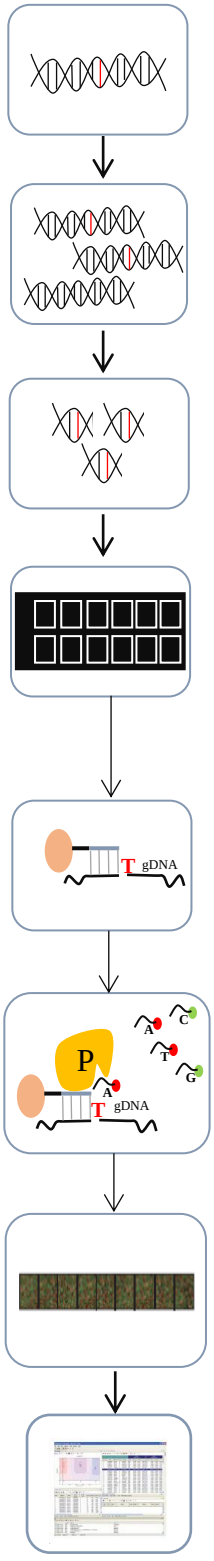


Figure 2- 1: **Flowchart for DNA samples quality control.**



A. 200-400ng of gDNA is required for Illumina HumanCore Exome BeadChips.

B. gDNA is amplified using PCR-free whole genome amplification.

C. Amplified DNA is fragmented using an enzymatic fragmentation approach.

D. Amplified fragmented DNA is hybridized to louci specific probes on the beadchip.

E. Two-steps allele detection:

1- Selectivity: Unlabelled DNA fragments hybridize to 50 mer probe on the array.

2- Specificity: Enzymatic single base extension with labelled nucleotide.

F. Hybridized arrays are scanned by iScan system which is a cutting-edge array scanner that supports rapid, sensitive, and accurate imaging of Illumina BeadChips for genetic analysis results.

G. GenomeStudio is employed for genotype calling, signal intensity data normalization and LRR and BAF establishing for each SNP.

Figure 2- 2: **Illumina Infinium HumanCoreExome BeadChip genotyping protocol.**

Original figure based on the information from (Fan et al. 2006).

2.3.4.1. Scanning genotyping chip

Hybridized microarrays were scanned by the iScan system, which is an array-scanner that supports accurate imaging of Illumina BeadChips for genetic analysis results. The iScan system has high performance lasers, optics and detection systems that enable the iScan to provide a sub-micron resolution with a high signal-to-noise ratio, and high sensitivity scanning.

The iScan system determined the signal intensity for each genotyped SNP by detecting the amount of the green signal of streptavidin and the red signal of dinitrophenol (DNP) attached to every bead (Figure 2- 3). Each bead chip had its own decode file (.dmap) provided by Illumina which contained the bead locations. This file was uploaded to the iScan and matched to the array by a barcode. Scanning results were reported in several output files including an intensity data file (.idat), that contained the mean signal intensity for red and green signals for each bead type, and a manifest file (.bpm or .opa), that listed the SNP ID and annotation for each bead type. These files were uploaded to GenomeStudio for data analysis.

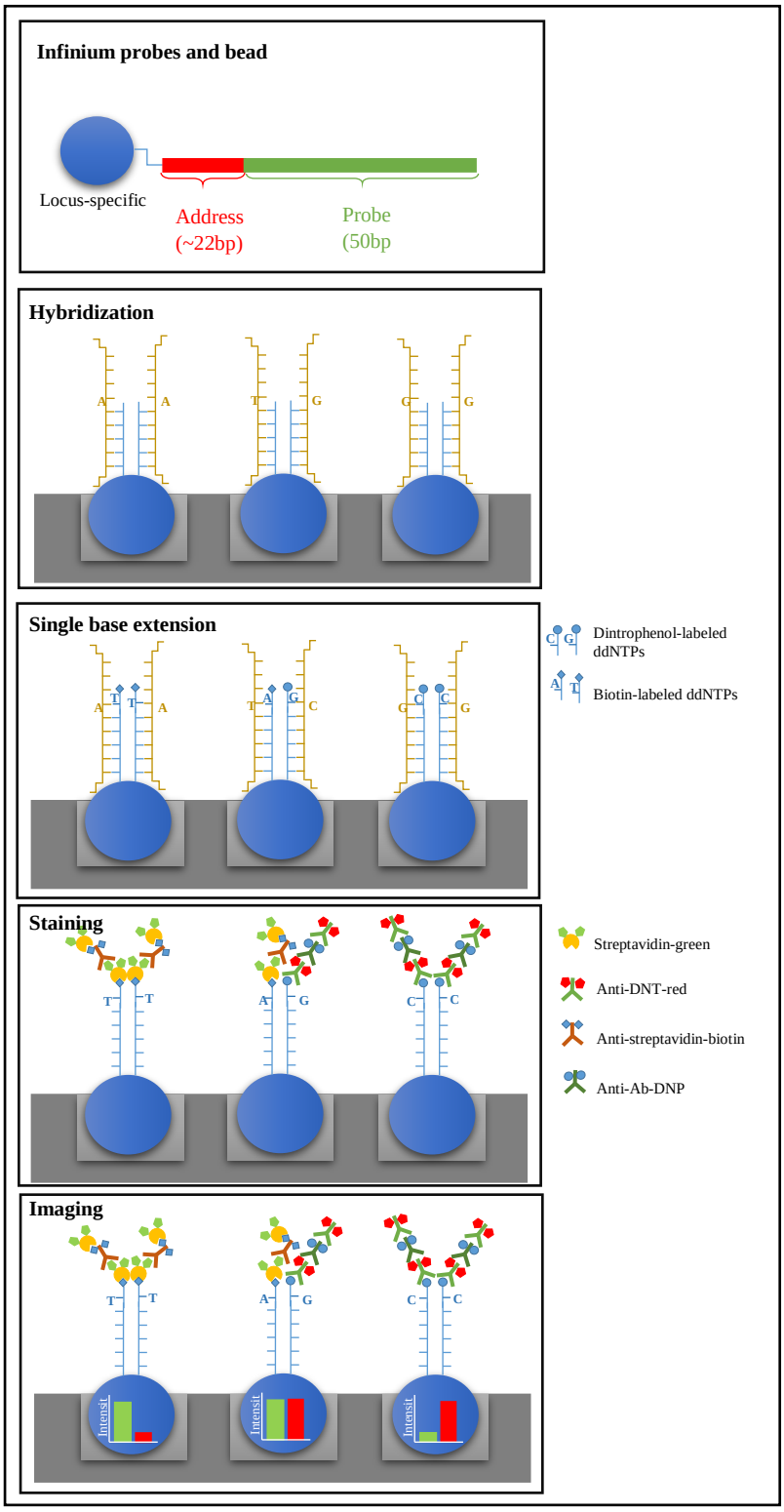


Figure 2- 3: **Beads signal intensity determination by iScan.**

Every single bead has a locus-specific probe and DNA sequence that matches probe sequence binds to the probe. Then a single base extension takes place to determine genotype in this marker. Based on the genotype either green streptavidin (For A and T genotypes) or red anti DNT (for G and C genotypes). Accordingly, anti-streptavidin-biotin and anti-Ab-DNP will form a complex. Based on the illuminated signal colour from each bead, green or red, which is detected by iScan, marker genotype is determined. Original figure based on the information from (Fan et al. 2006).

2.3.4.2. Genotypes calling (GenomeStudio)

The GenomeStudio software (Illumina) visualizes and analyses the data generated by Illumina microarray platforms. The Genotyping Module was used to conduct genotype data normalization, genotype calling, genotype clustering and data intensity analysis. It was also used to generate SNP statistics and to provide a graphical display of genotypes. For each genotyped marker, each sample was displayed as a dot according to its signal intensity and allele frequency. Genotypes were clustered based on A and B alleles composition with data points colour coded for the call (red = AA, purple = AB, blue = BB) (Figure 2- 4).

In addition, the Genotyping Module was used to estimate the Log R ratio (LRR) and B-allele frequency (BAF) for copy number analysis. Genotype data was exported for downstream analysis in the third party applications Plink (Purcell et al. 2007) and PennCNV (K. Wang et al. 2007).

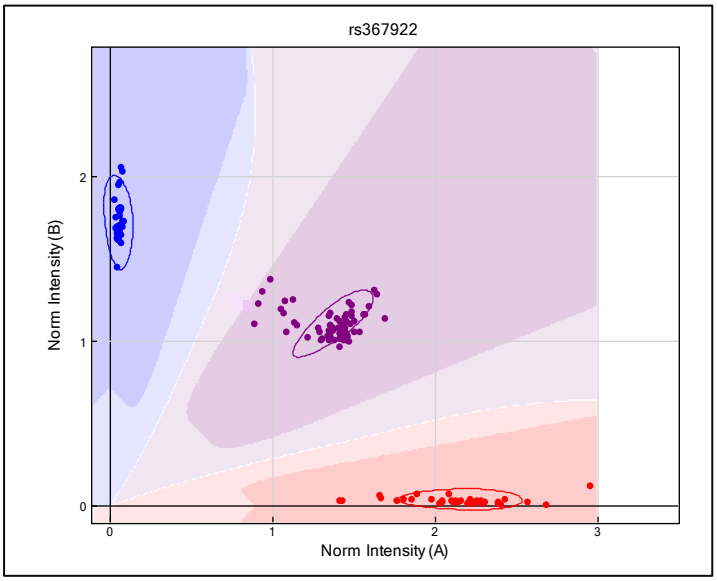


Figure 2- 4: **A screenshot from GenomeStudio for genotypes clustering plot in rs367922 marker.**

The red cluster represents homozygote genotypes for allele A where signal intensity for A allele is high and signal intensity for B allele is zero; the blue cluster represents homozygote genotypes for allele B where signal intensity for B allele is high and signal intensity for A allele is zero; and the purple cluster represents heterozygote genotypes where signal intensities for both alleles A and B are detected (GenomeStudio).

2.3.4.3. Extracting genotype data

Genotype data for each sample was exported after normalization and genotype calling in the form of PED and MAP files. The PED file was a space delimited file that contained six columns with information: family ID, individual ID, paternal ID, maternal ID, sex (1= male; 2= female; -9 or 0= unknown), and phenotype (-9 or 0= missing; 1= unaffected; 2= affected). The MAP file was also a space or tab delimited file, which contained information about genotyped markers. Each line of this file described a single marker and contained 4 columns: chromosome (1-22, X or Y), rs number or SNP identifier, genetic distance, and base-pair position (bp units).

Moreover, files including markers genotypes, LRR, and BAF values for all genotyped samples were extracted for PennCNV calling.

2.4. Genotype imputation

The genotypes for SNPs not on the array were imputed by the following protocol:

1) Aligning. The genotype data for all samples was first matched with the reference panel of haplotypes of the 1,000 Genome phase 3 data, using GRCh37/hg19 coordinates. Plink was used to flip all genotypes into the positive strand and SHAPEIT2 (v2 (r837)) was used to align the genotype data and to exclude any polymorphisms that could not be aligned (Delaneau et al. 2014; O'Connell et al. 2014).

2) Pre-phasing. SHAPEIT2 was used to estimate the haplotypes in the genotype data by applying statistical estimation algorithms such as the hidden Markov model (HMM).

3) Imputing: The FCGENE tool (1.0.7) was used to convert the genotype data into the format required for the program IMPUTE2. Genotype imputation was then performed using the pre-phase/imputation stepwise approach using IMPUTE version 2 (Howie et al. 2009). Imputation was based on a chunk size of 5Mb and was conducted only on autosomal chromosomes SNPs. Finally, FCGENE was used to convert the data generated by IMPUTE2 format to Plink format for association analyses.

2.5. CNV calling

2.5.1. CNV calling by PennCNV

CNV calling was performed using PennCNV (2009Aug27version) by following a well-established protocol (K. Wang et al. 2007). PennCNV required signal intensity values from genotype data in the form of LRR and BAF, which were exported from GenomeStudio. In addition to these values, population frequency of B alleles (PFB), SNP genome coordinates, and an appropriate HMM model data were also required for CNV calling. The PFB file was provided by Dr Elliott Rees, who worked on a study with a larger sample size that was genotyped by arrays similar to those used in this study. From this large cohort Dr Rees estimated the population frequency of B alleles for each genotyped marker. The HMM model was provided by the PennCNV package. PennCNV implemented a GC model for wave adjustment to reduce false positive calls.

CNV calls were subjected to an intensive quality control procedure to select reliable calls and this data was also used to conduct additional quality control of genotyped samples. Details about sample quality control analysis are explained in chapter 6 sections 6.4.1 and 6.5.1.

2.5.2. CNV quality control

The following quality control procedures were applied to all CNV calls in order to maximise the chances that the subsequent analysis included only true CNVs.

First, the called CNVs were restricted to only those spanning a minimum 10 consecutive SNPs, and larger than 100Kb with a minimum 20,000 SNPs density, which means a minimum number of 20,000 SNPs within 1Kb long. As large CNVs can be split by CNV calling algorithms, adjacent CNV calls larger than 100Kb were

merged if they existed in a single individual with an interval gap less than 50% of the entire length of the newly merged CNV (The International Schizophrenia Consortium 2008). The log R ratio and B allele frequency of SNPs spanning all CNVs, including the newly merged CNVs, were manually inspected before approval. At loci where it was not possible to manually exclude the existence of 2 independent CNVs, the adjacent CNVs were not merged (e.g. heterozygote genotype calls between adjacent hemizygous CNVs). A summary of CNVs quality control is given in Figure 2- 5.

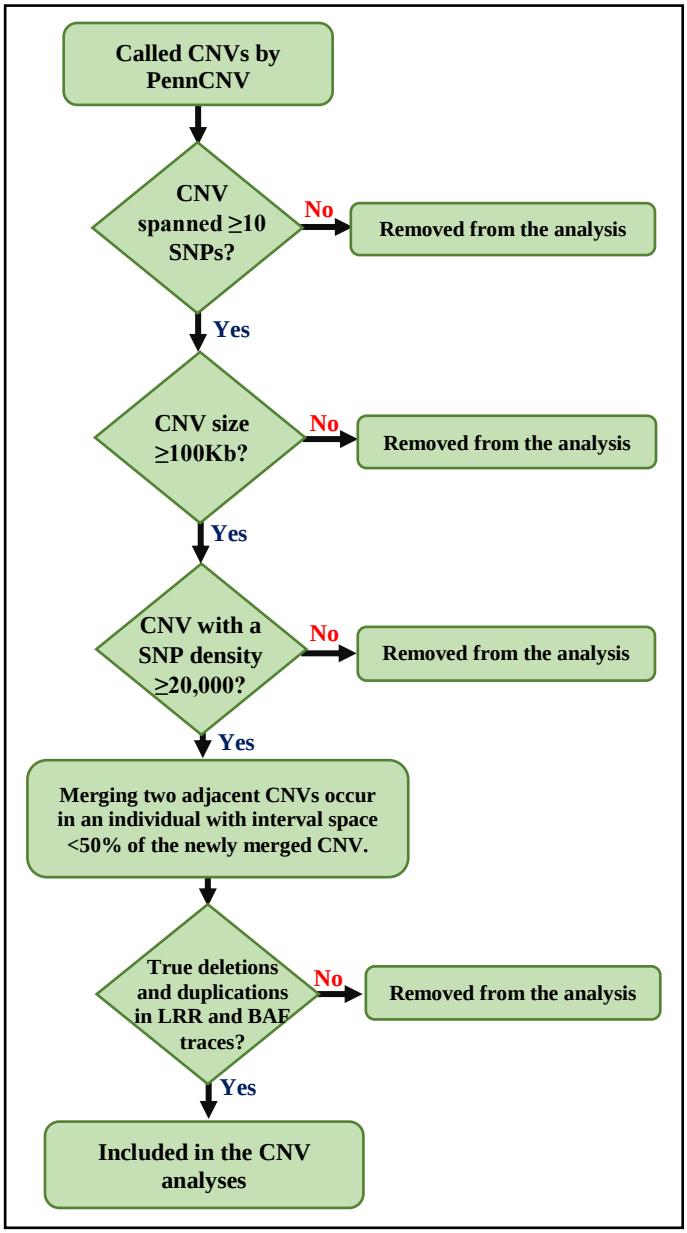


Figure 2- 5: **Flowchart for CNVs quality control.**

2.6. RNA samples

2.6.1. Blood samples for RNA extraction

Blood samples for RNA extraction were collected in PAXgene Blood RNA Tubes (IVD). These tubes are intended to be used for the collection, transport, and storage of blood samples to stabilize intracellular RNA for subsequent isolation and purification of intracellular RNA from whole blood. The tubes are combined with the PAXgene Blood RNA kit (QIAGEN).

58 ECHO blood samples were collected in PAXgene blood RNA tubes for RNA extraction. These samples included 33 samples from 22q11.2DS patients and 25 samples from healthy controls. An additional 21 samples were obtained from the London BBAG project, which were similarly collected in PAXgene RNA blood tubes, and these were composed of 22q11.2DS patients (n= 3) and controls (n= 18). Thus, collectively there were 79 available samples including, 36 deletion carriers and 43 non-deleted controls.

2.6.2. RNA extraction and storage

The procedure of PAXgene RNA extraction began with centrifuging the blood sample to pellet RNA molecules in PAXgene blood tube, followed by a washing and a re-suspension steps. Re-suspended RNA molecules were then incubated in optimized buffers with proteinase K for protein digestion. Then cell lysate and residual cell debris were removed, while lysate was transferred to a PAXgene RNA spin column for centrifugation during which the RNA was selectively bound to the PAXgene silica membrane. Contaminants that passed through the membrane were discarded, whereas the RNA-bound membrane was washed and treated with DNase-I to remove trace

amounts of bound DNA. Finally, RNA was eluted in an elution buffer and then heat-denatured. The quality of the extracted RNA samples was assessed and RNA molecules were quantified and finally stored at -80^{o}C.

2.6.3. RNA quantification and quality control

2.6.3.1. Bioanalyzer RNA quantification and quality assessment

The most critical step before gene expression analysis is determining RNA integrity to select only RNA with high quality for expensive cRNA synthesis and gene expression microarray hybridization. RNA integrity was assessed by determining RIN value using Agilent 2100 Bioanalyzer System and associated Agilent RNA 6000 Nano kit (Agilent Technologies). The Bioanalyzer also was used for RNA quantification. This method is based on using electrophoretic separation on micro-fabricated chips, where RNA samples were separated and detected by laser induced fluorescence detection.

2.6.3.2. Nanodrop for RNA quantification and RNA quality assessment

Additionally, extracted RNA samples were also quantified and assessed by the Nanodrop 8000 instrument (Thermo Scientific). A ratio of A_{260}nm to A_{280}nm equal to or larger than value of 2 was accepted for a high quality RNA sample.

Then based on Bioanalyzer and Nanodrop results RNA samples were subjected for quality control to select only pure, high integrity, and adequate concentration RNA for

gene expression analysis processing. Full details for RNA samples quality control are explained in chapter 4 sections 4.4.1 and 4.5.1.

2.7. Bioinformatic tools and statistical analyses

2.7.1. Bioinformatic tools

Various statistical and bioinformatic tools were used to handle and analyse the genotype and gene expression data.

Linux (version Fedora 16) was mainly used to handle large genotype data extracted from GenomeStudio in the form of PED and MAP files. Linux on the Cardiff University server (Raven), in which the data was stored. It facilitates modifying and sorting data for association analyses using Plink.

Plink (v1.07-10/Aug/2009) was used for all quality control analyses, for the association analyses of genotypes, and also for the CNV burden analyses.

For results visualization, R (version 3.2.2-2015-08-14), which is a language and an environment for statistical computing and graphics, was used for plotting results of principle component analysis (PCA). R and Bioconductor packages (Limma, Lumi, and Combat) were used for gene expression data quality control, gene expression data pre-processing, and differential gene expression analysis. The results for these processes were also visualised using R. The R package was also used for conducting association and correlation analyses.

For visualising genotypes association results, LocusZoom was used to plot Manhattan plots (Pruim et al. 2010).

For visualising pathway analysis results obtained from DAIVD analysis tool, Cytoscape (v.3.4.0) was used to categorize functional groups into functional networks (Shannon et al. 2003).

Power calculation for association studies was conducted by using the Genetic Power Calculator (Purcell, Cherny & Sham 2003), while it was performed by using the pwr package (version 1.2-0), implemented in R and G*Power tool v3.0.10 for the CNV burden study (Champely 2016; Rucker et al. 2016). PS software version 3.1.2 (Dupont & Plummer 1990) was used for power calculation for gene expression studies.

2.7.2. Statistical analyses

Statistical analyses applied for each study are explained in the result chapters.

Chapter 3: Characterization of the 22q11.2DS Cohort

3.1. Summary

The main aim of this introductory chapter is to obtain a well-characterized cohort of 22q11.2DS patients that have high quality genotype data for the subsequent analyses performed in this PhD thesis. To achieve that, pipelines for sample and genotype quality control were followed.

As a result of sample quality control analysis 19 samples were excluded from the original 95 that were ascertained. The remaining 76 individuals were confirmed to be independent, ethnically homogenous, and true 22q11.2 deletion carriers. These individuals will form the basis of the work presented in chapters 4, 5, and 6 of this thesis.

3.2. Introduction

Recently, association analyses of SNPs and CNVs have been conducted at increasing frequency using case-control samples to identify genomic loci that influence human traits (Turner et al. 2011). Particularly in the field of neuropsychiatric disorders genome-wide association studies (GWAS) have analysed common genetic variants spanning the whole genome (Ripke et al. 2014; Simón-sánchez et al. 2009; Weiss et al. 2009; Cross-Disorder Group of the Psychiatric Genomics Consortium 2013), and global burden analyses have analysed the structural variants genome-wide (Williams et al. 2010; Williams et al. 2012; Kirov et al. 2009; Girirajan et al. 2010; McQuillin et al. 2011).

The capability of association studies to identify true genetic associations depends upon the overall quality of the data (Turner et al. 2011). Samples included in association analyses that have not been properly cleaned based on the genotype data can potentially lead to false-negatives and false-positive associations (Turner et al. 2011). A typical genotype-based sample quality control (QC) procedure will investigate gender inconsistencies, sample relatedness, population substructure, and sample genotyping efficiency or call rate (Turner et al. 2011).

Testing for gender inconsistencies tests for potential sample identity problems that typically occur as a result of possible errors in sample handling (Turner et al. 2011). This analysis is performed by determining sex empirically based on X chromosome heterozygosity rates, then matching the molecular sex with the reported one (Turner et al. 2011).

Closely related or duplicated samples are identified by identity by descent (IBD) analysis (Turner et al. 2011). This approach calculates the proportion of loci where a pair of individuals share zero, one, or two alleles that are identical by descent. Either monozygotic twins, or a single sample processed twice share two alleles IBD at every locus. While unrelated individuals share zero alleles IBD at every locus. Parent-child pairs share one allele IBD at every locus. Whereas siblings, on average, share zero alleles IBD at 25%, one allele IBD at 50%, and two alleles IBD at 25% of the genome (Turner et al. 2011). The higher proportion of the IBD loci, the more alleles are shared between two pair of individuals (Turner et al. 2011).

Population stratification occurs in studies that include multiple groups of individuals who vary systematically in genetic ancestry or other main variables such as phenotypes (Cardon & Palmer 2003). Differences in allele frequency are introduced by systemic ancestry differences, which might interfere with the allelic associations and produce type 1 errors (Price et al. 2006). Thus, it is critical to check for population stratification within the samples involved in association studies. That can be achieved by applying principle component analysis (PCA), and using a global reference for human genetic variants database such as 1000 Genome project database (The 1000 Genomes Project Consortium 2015) to identify the samples ancestry and detect any systematic differences among the samples (Ringnér 2008).

3.3. Aims of this chapter

This chapter aims to conduct quality control of the 95 individuals that have been ascertained in order to establish a cohort of 22q11.2DS patients that can be used for the subsequent analyses conducted in this PhD thesis. This was achieved by conducting a series of assessments of genome-wide SNP array data in order to establish which individuals were true 22q11.2 deletion carriers, to identify a sample with minimal evidence for population stratification, and to ensure that they were all independent and uncontaminated samples.

3.4. Materials and methods

3.4.1. Samples quality control

The DNA for all 95 recruited participants was genotyped using an Illumina Infinium Human CoreExome-24 BeadChip array as described in chapter 2 section 2.3. The genotype data was subjected to a standard SNP-based quality control procedure by using Plink (v1.07-10/Aug/2009) (Purcell et al. 2007). Samples were excluded if they met one of the following criteria:

1) A SNP call rate lower than 0.98.
2) Were identical by descent (IBD) (PiHat>0.95) with another sample, indicating a sample duplicate, or monozygotic (MZ) twin.
3) Were identical by descent (IBD) (PiHat>0.15) with another sample, indicating a potentially contaminated sample, or two closely related individuals.
4) Had a reported gender that was different to that determined by molecular analysis.

To confirm that all 22q11.2DS samples had the expected deletion, CNVs spanning the 22q11.2 region were identified by using the 790 SNPs spanning chr22:18,658,219-21,865,185 using PennCNV (2009Aug27version) as fully described in chapter 2 section 2.5 (K. Wang et al. 2007). Only samples confirmed to carry deletions spanning the common 3Mb 22q11.2 region (chr22:18,658,219-21,865,185 on hg19) or less common 1.5Mb region (chr22:18,658,219-20,519,134 on hg19) were included.

Principal component analysis was performed on all SNPs for the 22q11.2DS patients combined with the individuals of 1000 Genome phase 2 (The 1000 Genomes Project Consortium 2015), who have known ethnic backgrounds by using R statistical

environment (version 3.2.2-2015-08-14) (R-project.org, n.d.). Outlying samples that deviate from the homogenous Caucasian 22q11.2DS samples cluster, located within 0.04< PC1 ≥0.03 and PC2 ≥0.04 region on PCA plot, possibly indicating a non-Caucasian ancestry were excluded. Summary for sample quality control is given in Figure 3- 1.

3.4.2. Psychiatric phenotypes in 22q11.2DS

The subsequent well-characterized 22q11.2DS patients, who passed the sample quality control, then were categorized into three psychiatric diagnosis groups based on the presence and absence of neuropsychiatric phenotypes in these patients, which were assessed by the Cardiff ECHO field team as described in chapter 2 section 2.1.2. 22q11.2 deletion carrier manifested ADHD phenotypes, either inattentive, hyperactive or mixed, were grouped into 22q11.2DS+ADHD. Whereas, carriers who showed phenotypes of ASD with high SCQ score (15-22), were grouped into 22q11.2+ASD. Collectively, carriers who manifested ADHD, ASD or both phenotypes were grouped in 22q11.2DS+PSYCH group.

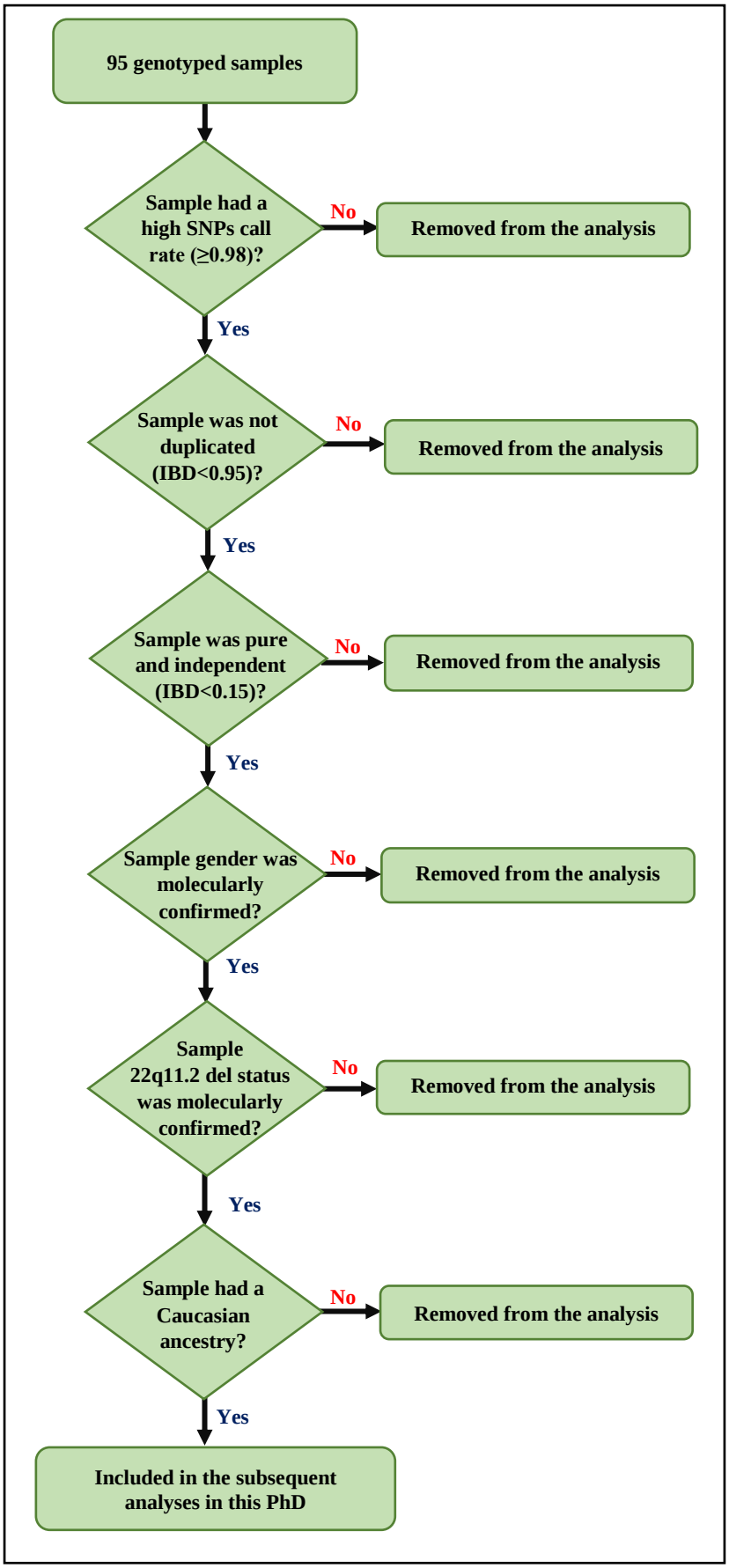


Figure 3- 1: **Flowchart for sample quality control based on genotype data.**

3.5. Results

3.5.1. Samples quality control

The reported gender of all 95 samples matched the sex-chromosome determined gender. The results of PennCNV calling for 22q11.2 deletions identified five of the individuals (5.2%) who did not carry either a 3Mb (spanning LCR22s A-D) or a 1.5Mb deletion (spanning LCR22s A-B) at 22q11.2 (Figure 3- 2). These 5 samples could not be classed as 22q11.2DS and were removed from the study.

As can be seen in Figure 3- 3, PCA analysis indicated that 78 individuals were homogenous with the individuals of known European ancestry, while the remaining 12 samples were clear outliers. The plot suggests that the outliers descend from a mixed ethnic background. Five outliers appear to have a mixed African background and located within $-0.02 < PC1 \leq 0.02$ and $0.04 > PC2 > 0.02$ region. While a single outlier is suspected to have a mixed Asian ancestry and located within $0.04 > PC1 > 0.02$ and $0.02 < PC2 \leq 0.0$ region. The remaining six outliers appear to have a mixed African, Asian, and European background with more dominant European genotypes as they are not fully homogenized with the European cluster. Accordingly, the 12 non-Caucasian samples were excluded from the study.

IBD analysis was sought to identify any related, duplicated or possibility contaminated samples. Two pairs were identified to have IBD proportion > 0.15 indicating relatedness. A single sample of each pair was removed.

A summary for samples QC results is given in Table 3- 1.

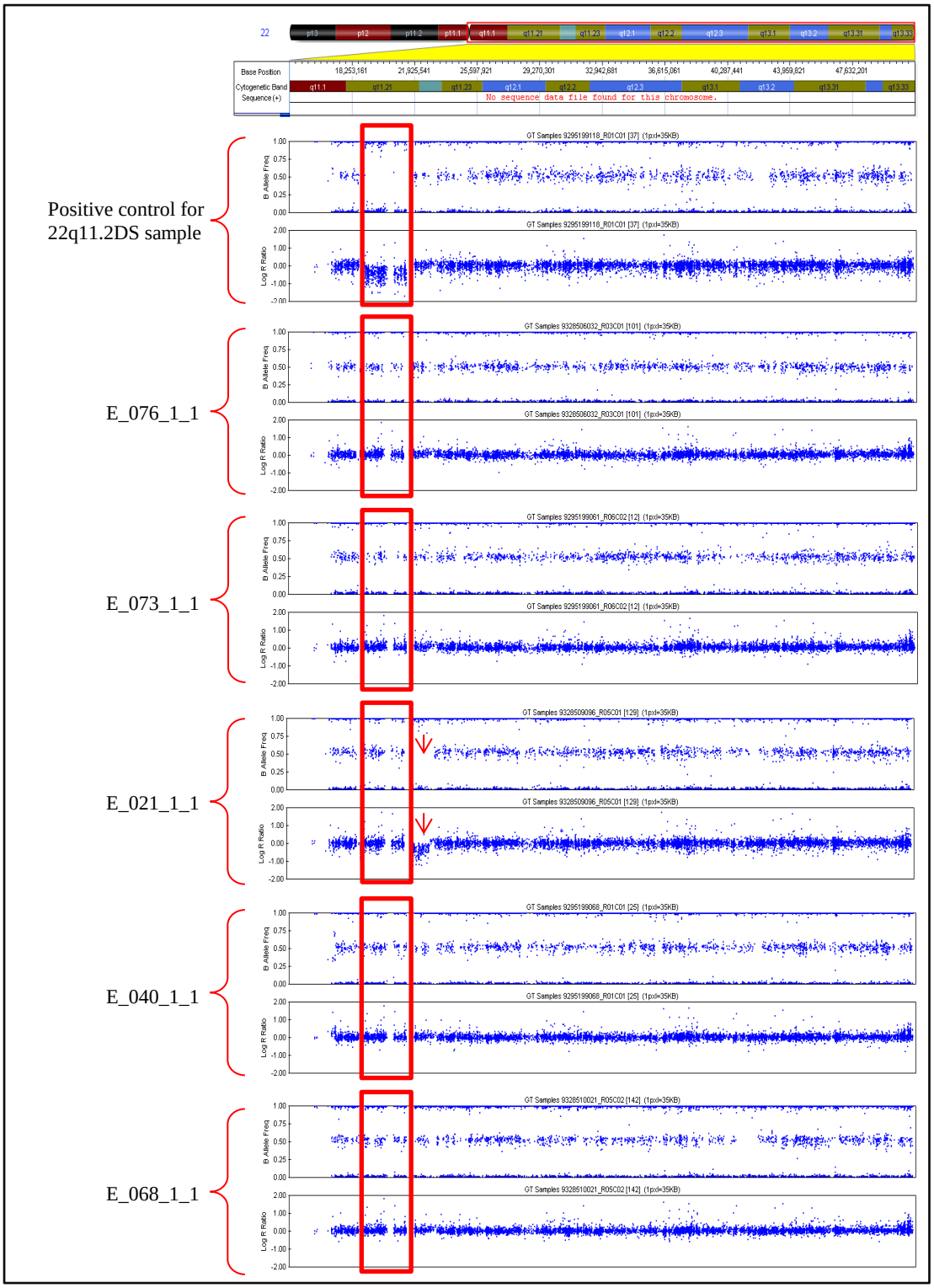


Figure 3- 2: **BAF and LRR traces for the five 22q11.2DS samples that are not carrying 22q11.2 deletions.**

In the top, the long arm of chromosome 22 with base location (on hg19) and the cytogenetic bands. The red boxes are the target 3Mb 22q11.2 region (chr22:18,658,219-21,865,185 on hg19). A control 22q11.2 deletion carrier sample shows all SNPs within the area are homozygous in BAF trace and the general signal intensity drops within the 22q11.2 region in LRR trace that indicate presence of a deletion in the region. The LRR and BAF traces of five samples show the absence of 22q11.2 deletion in comparison to the 22q11.2 deletion control sample. The red arrows indicate an irrelevant rare chromosome 22 deletion detected in E_021_1_1 sample (LRR and BAF traces are obtained from GenomeStudio).

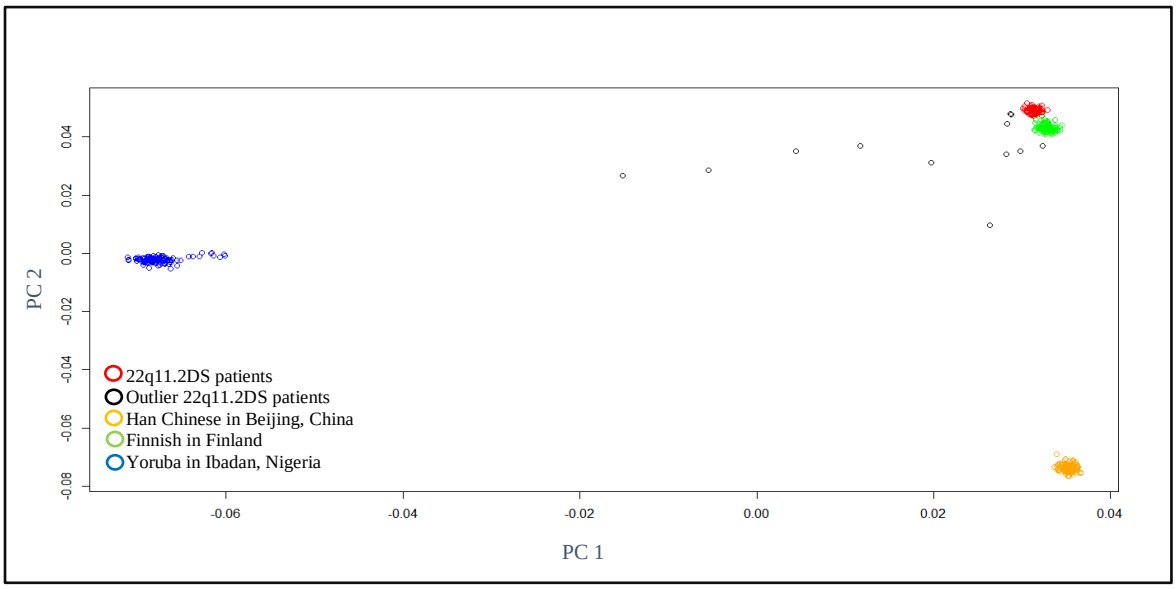


Figure 3- 3: **PCA plot for 22q11.2DS patients and 1000 Genomes participants.**

The plot shows three big clusters for the 1000 Genome participants with different ancestries. The blue cluster represents individuals with the African ethnicity, the orange cluster represents individuals with the Asian ethnicity and the green cluster represents individuals with the European ethnicity. The clusters distributed in the plot based on eigenvectors resulted from PCA analysis. Homogenous 22q11.2DS patients (n= 78) are in the red cluster which appears to descend from European ancestry. 22q11.2DS patients that are suggested to have a mixed ethnical background are highlighted with black as they deviated from the main cluster of 22q11.2DS patients. X-axis represents PC1 values and Y-axis represents PC2 values from PC analysis (PCA plot is plotted by R).

Table 3- 1: **Results of sample quality control.**

QC filters	*Threshold	**Initial total	***Removed	****Final total
Genotype call rate	0.98	95 cases	0 cases	95 cases
22q11.2 deletion	Presence of 22q11.2 deletion	95 cases	5 cases	90 cases
PCA	Caucasian	90 cases	12 outliers	78 cases
IBD	0.15	78 cases	2 cases	76 cases
Sex check	Reported gender is molecularly confirmed	76 cases	0 cases	76 cases

* Filtering threshold or criteria.
** Initial total: Initial total number of samples.
*** Removed: Number of removed samples due to failure in meeting filtering threshold or criteria.
**** Final total: Final total number of samples.

3.5.2. Characterization of 22q11.2DS cohort

The 76 individuals that survived this standard SNP-based quality control procedure formed the basis of the work presented in chapters 4, 5, and 6 of this thesis. The 76

individuals with 22q11.2DS had a mean age of 11.9 years. Three of the 22q11.2DS patients (~4%) were found to carry the ~1.5Mb deletion spanning LCR22s A-B, whereas 73 22q11.2DS patients (~96%) have the ~3Mb deletion spanning LCR22s A-D (Figure 3- 4).

Of the 76 22q11.2DS patients, 43 individuals are males (57%) while the remained 33 are females (43%).

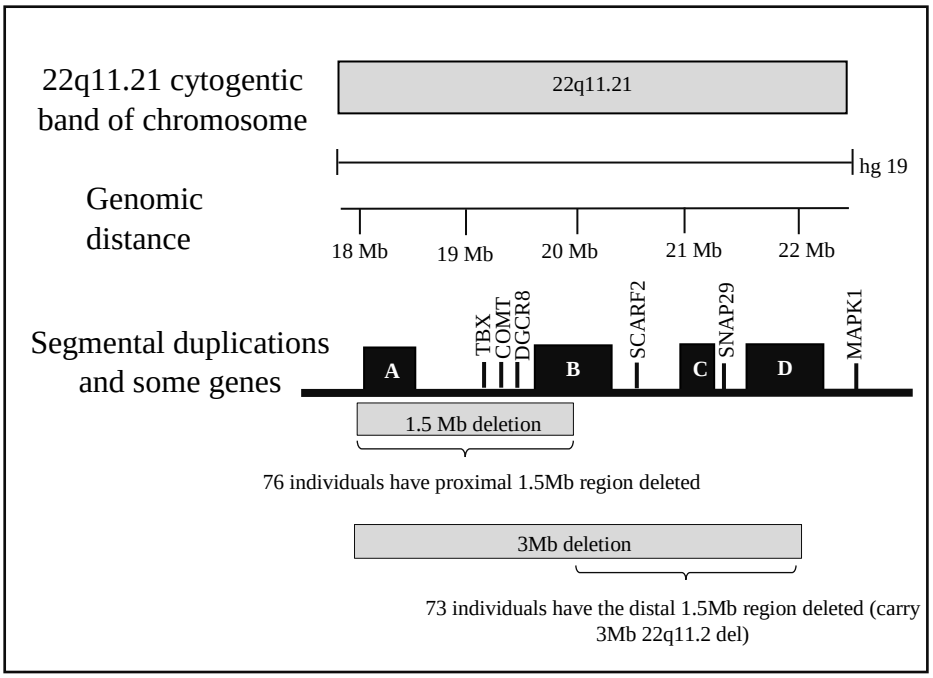


Figure 3- 4: **Schematic ideogram of 22q11.2 regions spanned by 1.5Mb and 3Mb deletions.**

LCR22s are indicated by black boxes and 22q11.2 deletions indicated by grey boxes. 73 individuals in our cohort carry 3Mb deletion (spanning LCR22A-D) while only 3 individuals carry the 1.5Mb deletion (spanning LCR22A-B). Therefore, total 76 individuals, including those who carry the small nested 1.5Mb deletion, share the deleted region spanning the proximal 1.5Mb (LCR22A-B), while only 73 individuals, who carry the larger deletions, shared the deleted region spanning the distal 1.5Mb region (LCR22B-D) (Original figure).

The 76 individuals were classified into 3 categories based on the psychiatric phenotypes present; 1) presence/absence of ADHD, 2) presence/absence of ASD, 3) presence/absence of psychiatric illnesses i.e. either ASD, ADHD or both (Table 3- 2).

Table 3- 2: **Psychiatric phenotypes in 22q11.2DS cohort.**

	76 samples carrying 1.5Mb 22q11.2 deletion (3 samples carry the 1.5Mb deletion and 73 carry the 3Mb deletion)		
	ADHD	**ASD**	**Psychiatric illness (ADHD, ASD or both)**
Affected	29	18	37
Unaffected	42	47	35
Missing information	5	11	4
Total	76	76	76
	73 samples carrying 3Mb 22q11.2 deletion		
	ADHD	**ASD**	**Psychiatric illness (ADHD, ASD or both)**
Affected	27	18	35
Unaffected	42	46	35
Missing information	4	9	3
Total	73	73	73

It is worth noting that 10 of 76 22q11.2 deletion carriers manifest both ADHD and ASD. Only 19 of the 76 have ADHD symptoms only, whereas only 8 showed ASD symptoms only. The 37 individuals with 22q11.2DS who have a psychiatric illness presented with the phenotypes of ADHD only, ASD only, or both ADHD and ASD. The 10 individuals who presented symptoms of both ADHD and ASD were included in both the 22q11.2DS+ADHD and 22q11.2DS+ASD groups, as well as in the combined group with psychiatric disease, 22q11.2DS+PSYCH. This was aimed at maximizing the power in each test. However, those unaffected with psychiatric illness have neither symptoms of ADHD nor for ASD.

3.6. Discussion

The main aim of this chapter is optimize our 22q11.2DS patients and the genotype dataset for the further analyses conducted in this PhD by following standard QC pipelines. Therefore, all the 95 recruited were subjected to an intensive QC procedure to exclude non-22q11.2 deletion carriers, non-Caucasian, related, duplicated, or possibly contaminated samples. The final number of 22q11.2DS samples after QC was 76 well-characterized 22q11.2 deletion carriers.

All the recruited 95 samples were selected based on genetic clinic diagnosis of 22q11.2DS. However, molecular analysis revealed that 5 did not carry a deletion at 22q11.2. Further information for these referred patients were obtained from their genetic clinics.

Sample 1 (E_076_1_1), was referred from a medical genetic clinic based on the manifestation of DiGeorge Syndrome phenotypes diagnosed at age of 2 years without confirming the presence of 22q11.2 deletion. FISH analysis results were obtained afterward and showed normal results in addition to an array analysis that did not identify a 22q11.2 deletion, nor a *TBX1* mutation, nor an array abnormality.

Sample 2 (E_073_1_1), the patient was described as having multiple congenital anomalies suggestive of DiGeorge Syndrome but no cardiac defects. FISH analysis was carried out three times after recruiting the case by the ECHO team and this failed to detect a 22q11.2 deletion.

Sample 3 (E_021_1_1), the laboratory results identified that this patient does not carry the standard 22q11.2 deletion, however, the patient instead carries a deletion that partly overlaps the distal deleted region in 22q11.2DS. This deletion mapped to chr22:21,796,000-22,916,000 on hg19 and spans LCR22s D-E. The deletion was

detected using Agilent's 4x44K array (version 2) and it was confirmed using the FISH probes LL22NC03-23C6 and LL22NC03-30E12. This finding was confirmed by the analysis of genotype data and CNVs calling in this chapter (Figure 3- 2). Thus, this sample was removed as there is no evidence that the deletion identified is relevant to 22q11.2DS.

Sample 4 (E_040_1_1), was found to be a mislabelled sample based on IBD analysis results. This patient was detected to be related to individuals from a different family with 50% shared genotypes. This percentage of genotype sharing between two individuals means they are first degree relatives. In this case, this sample is possiblely had been falsely labelled as no deletion was identified.

Sample 5 (E_068_1_1), was labelled as a 22q11.2 deletion carrier, however, while no 22q11.2 deletion was found, a deletion at 16p11.2 was identified instead. The ECHO field team confirmed that this samples was indeed a carrier of a 16p11.2 deletion and it had been mislabelled.

In conclusion, after intensive QC procedures, the final dataset composed of 76 well-characterized 22q11.2DS patients. The 76 22q11.2 carriers were then categorized into 22q11.2DS+ADHD, 22q11.2DS+ASD, and 22q11.2DS+PSYCH groups for the subsequent analyses described in chapter 4, 5, and 6 of this PhD thesis.

Chapter 4: Investigating Dosage Sensitive Genes in 22q11.2 Deletion Syndrome

4.1. Summary

Patients with 22q11.2DS have an increased risk of developing a diverse spectrum of behavioural phenotypes. As deletions at 22q11.2 are relatively homogeneous, it is possible that haploinsufficiency of dosage sensitive genes spanning the deleted region results in changes in gene expression that influence the pathogenesis of the resulting psychiatric phenotypes. Dosage sensitive genes that span 22q11.2 are therefore candidate loci for psychiatric phenotypes in 22q11.2DS patients. This chapter set out to identify which genes at 22q11.2 are dosage sensitive by conducting differential gene expression analysis of RNA obtained from the leukocytes of 22q11.2DS patients and non-deleted controls.

The results of differential gene expression analysis showed ~42.5% of 22q11.2 probes, representing 39 genes, are significantly differentially expressed (FDR <0.05) with down regulation effects in 22q11.2DS patients relative to healthy controls.

Secondly, the possibility that the deletion can influence the expression of genes located outside of the deletion was investigated by testing the correlation of differential gene expression with the probe's distance from the deletion. The results identified only a minor correlation (r^2 =0.0022) that was not statistically significant (p-value =0.13), which does not provide support for the 22q11.2 deletion causing a strong positional effect.

A pathway analysis for global differentially expressed probes revealed four significantly enriched biological networks (FDR <0.05) that are involved in: 1) translation, protein synthesis machinery, and post-translation modifications; 2) apoptosis; 3) regulation of the immune system; and 4) intramembrane organelles. These enriched biological networks could possibly suggest the involvement of dosage sensitive genes in some of the phenotypes associated with 22q11.2DS such as immune system abnormality.

Finally, in order to establish evidence whether any dosage sensitive genes could potentially influence the increased risk to psychiatric diseases, differential gene expression analysis was conducted between 22q11.2 deletion carriers who were affected and unaffected with a psychiatric disorder. The differential expression analysis results showed no significantly differentially expressed probes were found in 22q11.2DS patients affected by neuropsychiatric diseases relative to those unaffected (FDR >0.05).

4.2. Introduction

As microdeletions at 22q11.2 are considered to be the key genetic defect in the majority of 22q11.2DS cases, the resulting phenotypes for the syndrome are expected to be a result of haploinsufficiency of one or more genes within the deleted region (Williams 2011). In theory, as the hemizygous genes within the deleted 22q11.2 region have a reduced DNA copy number they are predicted to have a reduced level of expression when compared to the genes not spanned by a deletion (Williams 2011). In spite of that, it is well known that the expression level of many haploinsufficient genes can remain similar to those of disomic cells (Inoue & Lupski 2002; Stranger et al. 2007). This therefore indicates that not all the genes spanned by the deletion are necessarily sensitive to the DNA dosage imbalance. It is possible that the expression of these genes is compensated by homeostatic regulation at higher stages of the regulatory network that results in them retaining their normal expression level (Williams 2011; Stranger et al. 2007). Dosage sensitive genes at 22q11.2 have been shown to result in a number of phenotypes associated with 22q11DS. Highlighting dosage sensitive genes is a key role for identifying candidate genes for the neuropsychiatric phenotypes associated with 22q11.2DS (Williams 2011).

A number of studies have investigated the effect of the 22q11.2 deletion on gene expression in 22q11.2DS patients or in mice that have been engineered to carry a deletion of the syntenic region.

4.2.1. Gene expression studies in 22q11.2DS mice models

As described in chapter 1 section 1.3.1.1.2, a number of studies have investigated gene expression in tissues from the central nervous system of the *Df1*+/- mice models to

identify dosage sensitive genes that might be influenced by the 22q11.2 deletion and play a role in the behavioural phenotypes seen in these mice. Approximately 33-41.7% of *Df1* genes were identified to be differentially expressed in 22q11.2DS mice relative to wild-type mice (Jurata et al. 2006; Sivagnanasundaram et al. 2007).

4.2.2. Gene expression studies in 22q11.2DS patients

To date there are only two published studies of gene expression in humans with 22q11.2 deletions. First, Van Bervern and colleagues analysed the RNA from blood samples for 7 patients with 22q11.2DS and 7 non-carrier controls. Ten transcripts were identified to be differentially expressed, 7 of which were located within the 22q11.2 region and all were downregulated in 22q11.2DS. These genes were found to be enriched in signalling pathways relevant to the phenotypes seen in 22q11.2DS such as immunodeficiency and schizophrenia (van Beveren et al. 2012). The small sample size of this study is a major limitation of this study that is likely to result in an underpowered analysis. A second study has investigated gene expression in RNA extracted from the blood of 46 22q11.2DS patients and 66 healthy controls (Jalbrzikowski et al. 2015). This study was published while the study of gene expression analysis in 22q11.2DS described in this chapter was being conducted. Results of differential gene expression showed 45 probes that mapped to the 22q11.2 region were significantly downregulated in 22q11.2DS. Genome-wide analysis identified differentially expressed genes in 22q11.2DS patients that were involved in the regulation of neuronal action potentials, myelination, and axon ensheathment of neurons. Moreover, Jalbrzikowski and colleagues compared gene expression between 22q11DS patients with and without psychosis, and 22q11.2DS with and without ASD. No significantly differentially expressed probes mapped to the 22q11.2 region

(Jalbrzikowski et al. 2015). The 22q11.2DS analysed in the study of Jalbrzikowski and colleagues (Jalbrzikowski et al. 2015) were a mixture of children and adults (33 patients under 18 years and 13 patients between 18 and 61 years of age). The control cohort consisted of 24 and 42 individuals under and over 18 years of age respectively. This mixture of children and adults, together with their imbalance in the case/control cohorts are a potential limitation of this study.

4.2.3. CNVs positional effects on the expression of adjacent genes

It is possible that genomic structural variants have an impact not only on the expression of genes spanning them, but by removing or altering a regulatory locus, CNVs can directly influence the expression of nearby genes (Inoue & Lupski 2002). This phenomenon is called a 'position effect' in which genes within the same chromosome could be altered (in *CIS* position effect) or potentially on other chromosomes (in *TRANS* position effect) (Williams 2011). Stranger and colleagues (Stranger et al. 2007) have shown that there is significant long-distance disruption in gene expression for a large number of dosage sensitive genes that flank a CNV by more than 2Mb (Stranger et al. 2007).

Focussing on the 7q11.23 deletion that causes Williams-Beuren Syndrome (WBS), Merla and colleagues (Merla et al. 2006) investigated differential gene expression in RNA extracted from lymphoblastoid cell lines from 10 patients and from 40 healthy controls. The results showed evidence for a position effect for genes flanking the 7q11.23 deletion. For example, *HIP1*, *POR*, and *KCTD7* that are located at distances of 0.7, 1.2, and 6.5 Mb from the WBS region, respectively, showed a significant differential expression (Mann-Whitney test p-values= 0.025, 0.04 and $5.1x10^{-3}$

respectively). Overall differential expression appeared more pronounced for genes mapping closer to the breakpoint, suggesting the presence of distant long range in *CIS* regulatory elements within the deleted region (Merla et al. 2006).

In studying gene expression in the *Df1+/-* mice models, Prescott and colleagues demonstrated significant evidence for a *CIS* positional effect when they identified the reduced expression of the *Crkl* gene (that flanks the *Df1* deleted region) in *Df1+/-* mice compared to the wild-type mice (Prescott et al. 2005). In addition, as it plays a central role in miRNA production, haploinsufficiency of *DGCR8* gene could potentially lead to *CIS* and *TRANS* positional effects (Stark et al. 2008). This is supported by studies that have analysed *Dgcr8+/-* knockout mice that have identified a downregulation of a specific set of mature miRNA (Stark et al. 2008), and the subsequent upregulation of expression for a number of genes in the brain of *Dgcr8+/-* mice (Stark et al. 2008).

These findings suggest that *CIS* and *TRANS* positional effects could be relevant to the 22q11.2 deletion and if so could have an influence on the underlying biology in 22q11.2DS. Nevertheless, although mouse chromosome 16qA13 has synteny with human chromosome 22q11.2, there is a major divergence in gene organization and the sequences flanking the deletions. These therefore complicate applying the evidence of position effects observed in *Df1* mice models to humans. Studies of gene expression in humans are therefore required to investigate possible position effects in relation to the psychiatric phenotypes associated with 22q11.2DS.

4.2.4. Transcriptome profiling of peripheral blood in 22q11.2DS

Studies of transcriptome profiling aimed at investigating the neuropsychiatric disorders associated with 22q11.2DS should ideally conduct gene expression analysis in brain tissues of 22q11.2DS patients. However, the availability of post mortem human brain from 22q11.2DS patients is particularly limited. All studies have therefore conducted gene expression analysis on RNA extracted from peripheral blood mononuclear cells (PMCs) from 22q11.2DS patients. While this is not ideal, these cells have been demonstrated to express many brain-relevant genes (Sullivan et al. 2006). In this study, Sullivan and colleagues analysed transcriptional profiling of 33,698 genes in 79 human tissues which included whole blood and 16 samples from the CNS (amygdala, caudate nucleus, cerebellum, cerebellum peduncles, cingulate cortex, globus pallidus, hypothalamus, medulla oblongata, occipital lobe, parietal lobe, pituitary, pons, prefrontal cortex, subthalamic nucleus, temporal lobe, and whole brain). Gene expression analysis demonstrated that ~50% transcripts were expressed in both whole blood and CNS and half of a set of schizophrenia candidate genes were expressed in both whole blood and prefrontal cortex (Sullivan et al. 2006). Therefore, the authors indicated that gene expression in whole blood is neither perfectly correlated nor perfectly uncorrelated with gene expression in multiple brain tissues. Thus, compared to the more inaccessible CNS tissues, RNA from peripheral blood can offer a more accessible tissue for gene expression, however, extra caution and consideration are necessary to interpret the data from peripheral blood gene expression (Sullivan et al. 2006).

Encouragingly, a study of a functional gene expression analysis has revealed that 7 genes that have orthologues in 22q11.2 (*DGCR6, RANBP1, ZDHHC8, HTF9C,*

COMT, *CLDN5*, and *UFD1L*) have similar relative expression levels in: 1) hippocampus when compared in *Df1*+/- and wild-type mice (Sivagnanasundaram et al. 2007), 2) PBMC when compared in 22q11.2DS patients and controls (van Beveren et al. 2012). The relative expression levels of mice and humans are correlated strongly and significantly (r^2 =0.677, one-sided p-value =0.05), which showed reduced expression in *Df1*+/- mice and 22q11.2DS patients compared to wild-type mice and controls respectively by 40-60% (van Beveren et al. 2012). Similarly, in a study investigating the expression in the brain of developing and adult *Df1*+/- mice, 9 genes with orthologues at 22q11.2 had a similar reduction in expression levels in both brain and PBMC by 40-60% (Meechan et al. 2006).

Taken together, these findings indicate that while recognising that it is not ideal, when analysed with caution it is possible to study the gene expression profile of brain-relevant genes in peripheral blood.

4.3. Aims of this chapter

The study in this chapter will investigate the effect that 22q11.2 deletions have on gene expression of RNA extracted from leukocytes. Compared to previous studies, this will benefit from the increased power of analysing a larger cohort of 22q11.2 deletion children carriers and age-matching controls. The study will perform the following lines of investigation:

1) An assessment of gene expression of haploinsufficient genes located within 22q11.2. As not all genes are dosage sensitive the identification of genes that are differentially expressed between 22q11.2 deletion carriers and non-deleted controls will highlight those whose expression are influenced by the deletion. These dosage sensitive genes will be candidates for future investigations into the behavioural phenotype associated with 22q11.2DS.
2) An assessment of the expression of genes located outside 22q11.2. It has been reported that the expression of genes that are not spanned by a CNV can be directly influenced the presence of the CNV (Stranger et al. 2007). This phenomenon is called a 'position effect' in which the expression of genes can be effected within the same chromosome (in *CIS*). This will be investigated by performing gene expression analysis of RNA extracted from 22q11.2 deletion carriers and non-deleted controls for genes located on chromosome 22 but outside of the 3Mb deleted region. In addition, an analysis of genome-wide differential gene expression changes will allow the investigation of biological pathways that are potentially affected by the 22q11.2 deletion.
3) An investigation of differential gene expression between 22q11.2 deletion carriers with and without a psychiatric disorder. Differentially expressed genes

that are located at 22q11.2 will reveal those that could play a direct role in the increased risk to psychiatric disease.

4.4. Materials and methods

4.4.1. Samples

A total of 79 blood samples were recruited for RNA extraction from: 36 patients with a molecularly confirmed diagnosis of a 22q11.2 deletion, and 43 unaffected healthy controls. Specific full details for the recruited participants and the collected blood samples are given in chapter 2 sections 2.1.1 and 2.6.1. RNA were extracted from whole blood samples using the PAXgene Blood RNA extraction kit (QIAGEN). Detailed protocols of RNA extraction and RNA quantification are fully explained in chapter 2 sections 2.6.1, 2.6.2, and 2.6.3.

All 79 RNA samples were put through quality control to select high quality RNA samples for cRNA synthesis. RNA samples meeting one of the following criteria were excluded:

1) Poor RNA integrity with an RNA integrity number (RIN) below 7 (Schroeder et al. 2006).
2) RNA with a ratio of A_{260}/A_{280} below 2, potentially indicating contamination with residual phenol or other reagent associated with the extraction protocol (Manchester 1996).
3) Low total RNA concentration below 400ng (i.e. the working concentration for cRNA synthesis based on the manufacturer's recommendation).
4) A duplicated sample.

4.4.2. cRNA synthesis

400ng of RNA for each of the samples that passed QC was subjected to biotinylated cRNA synthesis using the Illumina® TotalPrepTM RNA Amplification kit (Ambion from Life Technologies). This involved an initial reverse transcription using an oligo (dT) primer including a T7 promoter and a reverse transcriptase (RT) called ArrayScript™ to produce first cDNA strand. *In vitro* transcription (IVT) of the cDNA with biotin-UTP produced a biotinylated, antisense RNA copy of each single mRNA in the sample, which is called cRNA (Figure 4- 1).

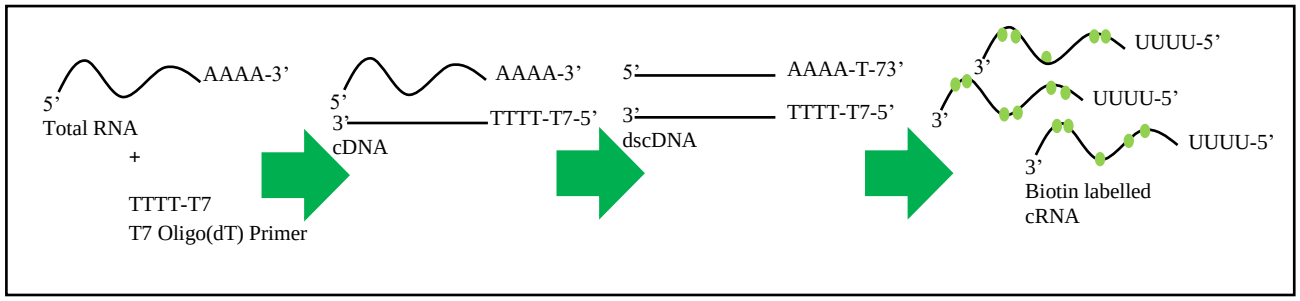


Figure 4- 1: **Method of biotinylated cRNA synthesis for gene expression microarray.**

The procedure starts with reverse transcription to synthesize the first strand of cDNA by using a T7 Oligo(dT) primer to synthesize cDNA containing a T7 promoter sequence. Then, the second strand cDNA is synthesized by converting the single-stranded cDNA into a double-stranded DNA (dsDNA) template for transcription. Then, *in vitro* transcription takes place to synthesize cRNA to generate multiple copies of biotinylated cRNA from the double-stranded cDNA templates (green dots in the cRNA represent biotin). Finally, cRNA is ready for use with Illumina's direct hybridization array kits. Obtained from (Illumina® TotalPrep™ RNA Amplification kit manual, n.d.) with modification.

4.4.3. Gene expression microarray

Whole-genome transcriptional profiling was performed using Illumina HumanHT-12 Expression microarrays (Illumina). The Illumina HumanHT-12 expression BeadChip targets 47,231 oligonucleotide probes that represent 34,602 genes.

750ng in a total volume of 5µl of biotinylated cRNA products were hybridized to Illumina Human V4.0-HT-12 BeadChips following the manufacturer's protocol. Loaded BeadChips were incubated for 14 hours for hybridization at 58°C, then washed

and processed for signal detection by introducing Cy3-Streptavidin (Thermo Fisher) to bind to the hybridized biotinylated cRNA probes to the BeadChip. This allows for differential detection of signals when the BeadChips were scanned by iScan system. Scanned gene expression data was analysed by gene expression module (version 1.9.0) in GenomeStudio software (v2011.1). GenomeStudio analyses intensity signal for each probe and converts the signal into an intensity value. Each probe has a signal value that represents the concentration of an RNA transcript in the sample.

4.4.4. Gene expression data quality control and pre-processing

4.4.4.1. Internal control probes quality control

Illumina have incorporated internal control features into the Gene Expression BeadChip, which includes 886 internal control probes to monitor data quality. These control features are either sample-independent metrics that use oligonucleotides in the hybridization solution, or sample-dependent metrics that use measurements from the actual loaded samples. The results of these controls were visualized by obtaining the Control Summary analysed by GenomeStudio (Gene Expression Microarray Data Quality Control, n.d.). Unprocessed gene expression data was initially checked for quality using the internal control probes in the expression microarrays. Internal control probes were also used for an initial background correction after which they were removed before the gene expression data under went pre-processing and batch effect correction.

4.4.4.2. Samples quality control

To detect possible sample handling errors, the gene expression data of probes within 22q11.2 were used to determine the deletion status of each carrier, and these results were compared to those previously determined from their DNA in chapter 3 section 3.5.1. Moreover, probes of two gender-specific genes *RPS4Y1* and *XIST*, which had been reported to be differentially expressed between males and females (Vawter et al. 2004), were analysed to confirm the reported gender molecularly.

Lastly, the quality of the microarray data for each of the remaining samples was analysed by calculating the proportion of high quality probes. Samples quality control filtering is explained in Figure 4- 2.

The remaining high quality samples then were categorized based on the manifested psychiatric symptoms as described in chapter 3 section 3.5.3.

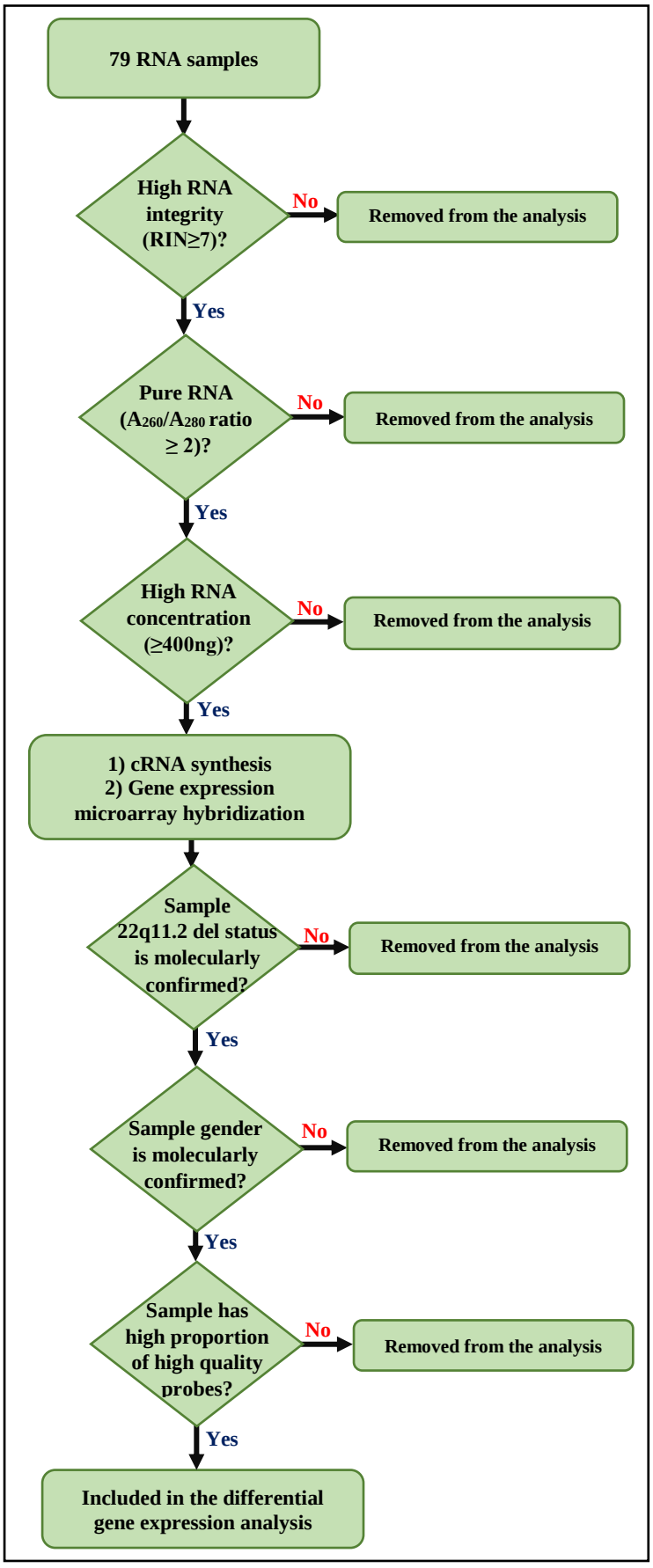


Figure 4- 2: **Flowchart for RNA sample quality control.**

4.4.4.3. Target probes quality control

In order to improve the power of detecting differential gene expression, probes with low expression were removed (Raw signal intensity threshold= 7000). In addition, probes with non-specific expression were identified by using annotation probe quality information obtained from gene expression platform specific annotation packages. ~12,741 probe sets were identified to have low or non-specific expression which then were removed from the subsequent analysis. A remaining of 34,490 target probes were selected for the differential gene expression analysis.

4.4.4.4. Gene expression data pre-processing

Gene expression data (34,490 probe sets) for each of the samples, that passed QC, were subjected for background correction by using the array negative control probes using Bioconductor (version 3.4) Lumi package version 2.26.4 in R statistical environment (Du et al. 2008). Raw intensities were normalized by the quantile normalization approach and transformed by the Log^2 transformation approach. Intensity plots for optimizing normalization and transformation approaches for our gene expression data are given in appendix 9.1 section 9.1.2.

4.4.5. Batch effect identification and correction

Batch effects are non-biological experimental variations that are commonly observed across multiple batches of microarray experiments. The main advantage of combining microarray data sets is increasing the study power by increasing sample size. However, it is unsuitable to combine gene expression data sets without correcting the experimental batch effect (Johnson et al. 2007).

Therefore, batch effects were identified in the gene expression data by using the approach of dendrogram clustering provided by the Bioconductor (version 3.4) Limma package, version 3.30.8, in R (Ritchie et al. 2015). The dendrogram clustered the samples, that passed QC, based on the most variable probes, which were determined by calculating the Interquartile range (IQR) for each probe among all samples. Probe IQR values then were ordered to identify the 500 (an arbitrary number) most variable probes. These probes then were used to categorize the samples into clusters (Dunning et al., 2016). Number of samples in the largest two clusters were determined based on their gender, phenotype, original study (ECHO or BBAG), and wave (A or B waves). The data then analysed by Fisher's exact test to test for significance. The identified batch effects were corrected by using Bioconductor (version 3.4) SVA package, version 3.22.0, including ComBat tool in R. ComBat applies parametric and nonparametric empirical Bayes approaches for correcting expression data for the identified batch effects (Johnson et al. 2007). Further illustrations about identifying batch effects are explained in appendix 9.1 section 9.1.3.1, and about the data after batch effects correction in appendix 9.1 section 9.1.3.2.

4.4.6. Differential gene expression analyses

4.4.6.1. Global differential gene expression analysis

Pre-processed expression data were analysed by using Bioconductor (version 3.4) Limma package version 3.30.8 in R (Ritchie et al. 2015) by fitting a linear model and applying empirical Bayes estimation approach to compare the signal intensity values between the 22q11.2DS samples and control samples for differential expression. Differential gene expression was calculated by mean of the $\log_2$ fold change (coefficients) and the resulting p-value for each probe were adjusted for multiple

testing by using Benjamini Hochberg False discovery rate (FDR) method (Benajmini & Hochberg 1995). FDR significance threshold used in this analysis was equal to 0.05 to accept 5% of significant results as false positive.

4.4.6.2. 22q11.2 region differential gene expression analysis

Differential expression analysis of probes within the 1.5Mb region (LCR22sA-B) (n =74) utilised all the 22q11.2DS samples (n =33); however, expression analysis for probes spanned by LCR22s B-D (n= 46) utilised only those that carry a 3Mb deletion at 22q11.2 (n =32).

4.4.6.3. In *CIS* and in *TRANS* positional effect analyses

To study in *CIS* position effect, adjacent probes to the 22q11.2 deletion in chromosome 22 (n =596) were selected. Expression level of these probes were compared between the 33 22q11.2DS samples and the 35 control samples. In order to investigate whether probes located within a short distance to the deletion are more prone to be differentially expressed in 22q11.2DS than those which are located within a long distance to the deletion, a linear regression analysis was performed to test the correlation between significance level (FDR) of each adjacent probe, as an independent variable, and probe distance from the deletion, as a dependent variable.

On the other hand, to investigate for the presence of a *TRANS* positional effect, differential gene expression analysis was performed on probes outside chromosome 22 (n =33,707) comparing all 22q11.2DS samples to control samples.

4.4.6.4. Differential expression analysis in 22q11.2DS patients with psychiatric illnesses

Also, similar genome-wide, in *CIS*, and 22q11.2 differential gene expression analyses were performed to compare 22q11.2DS patients affected and unaffected with psychiatric diseases. Only 26 of the total 33 22q11.2DS samples, which were recruited by the ECHO Cardiff study group, have available neuropsychiatric symptom data and were categorized based on their manifested psychiatric phenotypes. Then, we performed the following contrast analyses separately: 1) 22q11.2DS+ADHD vs 22q11.2DS-PSYCH, 2) 22q11.2DS+ASD vs 22q11.2-PSYCH, and 3) 22q11.2+PSYCH vs 22q11.2DS-PSYCH.

4.4.7. Pathway analysis

Global significantly differentially expressed probes (FDR <0.05) outside the 22q11.2 region (n =785) were selected for pathway analysis. The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 (Huang et al. 2009) was used to identify biological networks that are enriched with differentially expressed probes. DAVID is a bioinformatics enrichment tool that performs a functional interpretation of large lists of genes to identify biological processes that are most pertinent to our study. DAVID has an extent annotation content coverage to over 40 annotation categories including gene ontology terms, protein–protein interactions, protein functional domains, disease associations, bio-pathways, sequence general features, homologies, gene functional summaries, gene tissue expression, literatures, and more (Huang et al. 2007).

A minimum number of ten differentially expressed probes was used to identify an enriched pathway. Significance of enriched pathways were assessed by FDR with significance threshold equal to FDR= 0.05.

4.5. Results

4.5.1. Samples quality control

The gene expression data from 69 out of the initial 79 samples passed the initial quality control procedures. The deletion status determined from the RNA of all 22q11.2DS samples were identical to that determined from their DNA, however, a single control sample showed unmatched molecular gender which was removed from the subsequent analysis. This resulted in a total of 68 samples of which 33 belong to 22q11.2DS patients and 35 belong to age-matched non-deleted controls. Bioconductor analysis revealed two significant batch effects in the gene expression data of these samples (sample wave (p-value =0.0041) and sample original study (p-value =0.0019)) that were subsequently corrected. The resulting gene expression data of the 68 samples was of sufficient quality and was used for all gene expression work presented in this chapter. Breakdown of excluded samples is explained in Table 4- 1.

Table 4- 1: **Results of samples quality control analyses.**

Quality controls filtering	*Threshold	**Initial total	***Removed	****Final total
RIN score	7.0	79	3	76
RNA concentration (ng)	400	76	3	73
A_{260}/A_{280} ratio	2.0	73	0	73
Duplicated sample	-	73	4	69
Gender	Reported gender confirmed molecularly	69	1	68
22q11.2 deletion	Reported 22q11.2 deletion status confirmed molecularly	68	0	68
Quality of gene expression data	Large proportion of high quality probes	68	0	68

* Filtering threshold or criteria.
** Initial total: Initial total number of samples.
*** Removed: Number of removed samples due to failure in matching filtering threshold or criteria.
**** Final total: Final total number of samples.

The 33 22q11.2DS patients were categorized based on the presented psychiatric diseases into 3 groups: 22q11.2DS+ADHD, 22q11.2DS+ASD, and 22q11.2DS+PSYCH to determine comparison contrasts for the subsequent gene expression analyses (Table 4- 2).

Table 4- 2: **Number of 22q11.2 deletion carriers affected and unaffected by psychiatric disorders.**

Categories	Description	Number of individuals
22q11.2DS+ADHD	22q11.2 deletion carriers with ADHD phenotypes	13
22q11.2DS+ASD	22q11.2 deletion carriers with ASD phenotypes	5
22q11.2DS+PSYCH	22q11.2 deletion carriers with ADHD and/or ASD phenotypes	14
22q11.2DS-PSYCH	22q11.2 deletion carriers with no ADHD and ASD phenotypes	12

4.5.2. Estimating the background rate of differential gene expression in 22q11.2DS

To establish an estimation for the background rate of differential expression in 22q11.2DS, a global expression analysis was performed by comparing the 34,370 probes that were outside of the 3Mb deleted region at 22q11.2 between the 33 patients with 22q11.2DS and the 35 non-deleted controls. The results showed that only 2.3% probes are significantly differentially expressed (FDR <0.05). This finding was used as a comparison to the rate of differential gene expression seen for genes spanned by the 3Mb deletion at 22q11.2 (Table 4- 3).

Table 4- 3: **Results of differential gene expression analysis on genome-wide probes excluding those within 22q11.2.**

Global differential expression analysis						
Total number of probes	**Significantly differentially expressed probes (FDR <0.05)**		**Significantly downregulated probes (FDR <0.05)**		**Significantly upregulated probes (FDR <0.05)**	
	Number	Percentage	Number	Percentage	Number	Percentage
34,370	785	2.3%	477	1.4%	308	0.90%

4.5.3. Investigating dosage sensitive genes within 22q11.2 region

Differential gene expression analysis of the 120 probe sets spanning the 3Mb deletion at 22q11.2 between 22q11.2DS patients and controls was aimed at identifying dosage sensitive genes at 22q11.2. The results revealed that 51 probes (42.5%) were significantly differentially expressed in 22q11.2DS relative to controls (FDR <0.05), all of which were downregulated in 22q11.2DS patients compared to controls (Table 4- 4). The level of gene expression difference of these probes are variable with a $\log_2$ fold change ranges from -0.053 to -1.142 (Table 4- 5). The average level of expression difference of these probes was -0.568 with SD equal to 0.31, indicating there is a substantial variability in gene expression levels within 22q11.2DS patients.

The 51 differentially expressed probes implicate transcripts for 39 genes at 22q11.2, of which 29 have been reported to be brain-expressed (Guna et al. 2015) (Table 4- 5).

Table 4- 4: **Results of differential gene expression analysis on 3Mb 22q11.2 region probes.**

22q11.2 differential expression analysis						
Total number of probes	**Significantly differentially expressed probes (FDR <0.05)**		**Significantly downregulated probes (FDR <0.05)**		**Significantly upregulated probes (FDR <0.05)**	
	Number	Percentage	Number	Percentage	Number	Percentage
120	51	42.5%	51	42.5%	0	0.0%

Table 4- 5: **Results of differential gene expression analysis for probes within 3Mb 22q11.2 region.**

Probe ID	Target gene	CHR	Region	Start (hg19)	End (hg19)	Log_2 fold change 22q11.2 del carries vs non-22q11.2 del carriers	P-value	FDR
6900253	*DGCR6*	22	1.5Mb	18899136	18899185	-1.114	$5.05x10^{-20}$	$1.25x10^{-18}$
1010364	*DGCR6*	22	1.5Mb	18899542	18899591	-0.987	$4.98x10^{-13}$	$3.35x10^{-12}$
7200274	*DGCR6*	22	1.5Mb	18899545	18899591	-0.968	$8.73x10^{-13}$	$4.97x10^{-12}$
5670167	*DGCR2*	22	1.5Mb	19023794	19109967	-0.937	$4.72x10^{-19}$	$8.73x10^{-18}$
7560554	**DGCR14*	22	1.5Mb	19122576	19122625	-0.912	$3.16x10^{-20}$	$1.17x10^{-18}$
870020	*HS.572896*	22	1.5Mb	19159833	19159882	-0.157	0.0005	0.0012
6900440	*SLC25A1*	22	1.5Mb	19163460	19163509	-1.142	$1.74x10^{-20}$	$1.17x10^{-18}$
1090427	*CLTCL1*	22	1.5Mb	19175106	19175155	-0.340	$8.66x10^{-11}$	$4.27x10^{-10}$
5090414	*MRPL40*	22	1.5Mb	19423280	19423329	-0.410	$3.59x10^{-6}$	$1.02x10^{-5}$
5860156	*C22orf39*	22	1.5Mb	19428409	19435755	-0.538	$1.01x10^{-9}$	$4.67x10^{-9}$
730100	*UFD1L*	22	1.5Mb	19437950	19437999	-0.313	$1.16x10^{-7}$	$3.90x10^{-7}$
1260475	*UFD1L*	22	1.5Mb	19438089	19438138	-0.085	0.0009	0.0022
5130382	*CLDN5*	22	1.5Mb	19510627	19510676	-0.909	$2.75x10^{-7}$	$8.48x10^{-7}$
5890537	*SEPT5*	22	1.5Mb	19706119	19706168	-0.154	0.0036	0.0081
6960022	*SEPT5*	22	1.5Mb	19712239	19712288	-0.775	$1.42x10^{-5}$	$3.75x10^{-5}$
4230719	*GNB1L*	22	1.5Mb	19775933	19842462	-0.618	$7.91x10^{-8}$	$2.79x10^{-7}$
5860136	*LOC728139*	22	1.5Mb	19780042	19780091	-0.299	0.0011	0.0026
6110246	*C22orf29*	22	1.5Mb	19833660	19842371	-0.968	$2.44x10^{-18}$	$3.61x10^{-17}$
2230397	*TXNRD2*	22	1.5Mb	19863040	19929515	-0.738	$6.84x10^{-14}$	$5.62x10^{-13}$
520446	*COMT*	22	1.5Mb	19948774	19948812	-0.664	$4.65x10^{-8}$	$1.72x10^{-7}$
6940243	*COMT*	22	1.5Mb	19956318	19956367	-0.762	$1.34x10^{-17}$	$1.65x10^{-16}$
630446	*C22orf25*	22	1.5Mb	20053229	20053278	-0.824	$4.39x10^{-8}$	$1.71x10^{-7}$
290053	*DGCR8*	22	1.5Mb	20099142	20099191	-0.366	$5.92x10^{-7}$	$1.75x10^{-6}$
7150291	*TRMT2A*	22	1.5Mb	20099388	20104818	-0.262	$4.31x10^{-8}$	$1.71x10^{-7}$
1300402	*TRMT2A*	22	1.5Mb	20099388	20104818	-0.339	$1.71x10^{-7}$	$5.50x10^{-7}$
6220170	*TRMT2A*	22	1.5Mb	20099388	20104818	-0.116	0.0015	0.0034
7000735	*RANBP1*	22	1.5Mb	20109797	20109846	-0.961	$9.94x10^{-16}$	$9.19x10^{-15}$
4610047	*RANBP1*	22	1.5Mb	20114499	20114548	-0.790	$6.17x10^{-13}$	$3.80x10^{-12}$
4590154	*ZDHHC8*	22	1.5Mb	20133867	20133916	-0.847	$2.14x10^{-13}$	$1.58x10^{-12}$
5420014	*DGCR6L*	22	1.5Mb	20301760	20307628	-1.047	$5.24x10^{-17}$	$5.54x10^{-16}$
6270630	*TMEM191B*	22	1.5Mb	20379312	20379361	-0.691	$3.98x10^{-11}$	$2.10x10^{-10}$
3360762	*TMEM191B*	22	1.5Mb	20379507	20379556	-0.638	$4.50x10^{-9}$	$1.96x10^{-8}$
2360414	*PI4KAP1*	22	1.5Mb	20383730	20398695	-0.605	$1.32x10^{-5}$	$3.62x10^{-5}$
2710093	*ZNF74*	22	3Mb	20761879	20761928	-0.541	$2.17x10^{-9}$	$1.08x10^{-8}$
7570692	*SCARF2*	22	3Mb	20778873	20792146	-0.053	0.0015	0.004
5290259	*KLHL22*	22	3Mb	20795805	20850170	-0.839	$2.46x10^{-14}$	$2.66x10^{-13}$
6770672	*MED15*	22	3Mb	20929457	20929506	-0.619	$1.05x10^{-7}$	$4.03x10^{-7}$
5420243	*MED15*	22	3Mb	20941648	20941697	-0.971	$3.60x10^{-10}$	$2.04x10^{-9}$
4220195	*DKFZp434N035*	22	3Mb	21056978	21056997	-0.256	$2.37x10^{-6}$	$8.06x10^{-6}$
4120347	*DKFZp434N035*	22	3Mb	21057949	21057980	-0.365	$3.21x10^{-5}$	$9.55x10^{-5}$
450706	*PI4KA*	22	3Mb	21061978	21213100	-0.401	$3.49x10^{-6}$	$1.15x10^{-5}$
4920273	*PIK4CA*	22	3Mb	21062115	21062164	-0.626	$2.05x10^{-12}$	$1.36x10^{-11}$
7100437	*SNAP29*	22	3Mb	21243886	21243935	-0.617	$3.28x10^{-9}$	$1.56x10^{-8}$
5960767	*CRKL*	22	3Mb	21307655	21307704	-0.700	$6.50x10^{-10}$	$3.52x10^{-9}$
4260551	*AIFM3*	22	3Mb	21335320	21335320	-0.252	0.0001	0.0003
290086	*LZTR1*	22	3Mb	21353194	21353243	-0.738	$1.69x10^{-12}$	$1.18x10^{-11}$
2450132	***THAP7*	22	3Mb	21354060	21356404	-0.994	$2.44x10^{-16}$	$3.63x10^{-15}$
5560021	*THAP7*	22	3Mb	21354060	21356404	-0.065	0.0049	0.0117
5820672	*HIC2*	22	3Mb	21805532	21805581	-0.211	0.0029	0.0071
1110278	*TMEM191C*	22	3Mb	21822478	21822491	-0.171	0.0057	0.0133
4480753	*LOC220686*	22	3Mb	21828944	21828946	-0.335	0.0008	0.0023

* Lowest FDR for genes within 1.5Mb region.
** Lowest FDR for probes within 3Mb region.
Genes highlighted with red are expressed in the brain.

4.5.4. Investigating positional effects of the 22q11.2 deletion

Analysis of a *CIS* positional effect of 22q11.2 deletion sought to investigate whether the deletion has a direct effect on expression of nearby genes in chromosome 22.

The results of linear regression analysis showed a positive correlation between the probes differential expression significance level (FDR) and the probes distance from the deletion, however, this correlation was found very small with a correlation coefficient (r^2) of 0.0022 and was not statistically significant with p-value= 0.13. This indicates that there is a potential enrichment of differentially expressed probes closer to the deletion and it is potentially that closer probes are more likely to be differentially expressed than the ones in a remoter distance. However, no significant evidence for this potential in *CIS* positional effect conferred by the deletion. A Manhattan plot shows more significantly differentially expressed probes are present in nearby regions (Figure 4- 3).

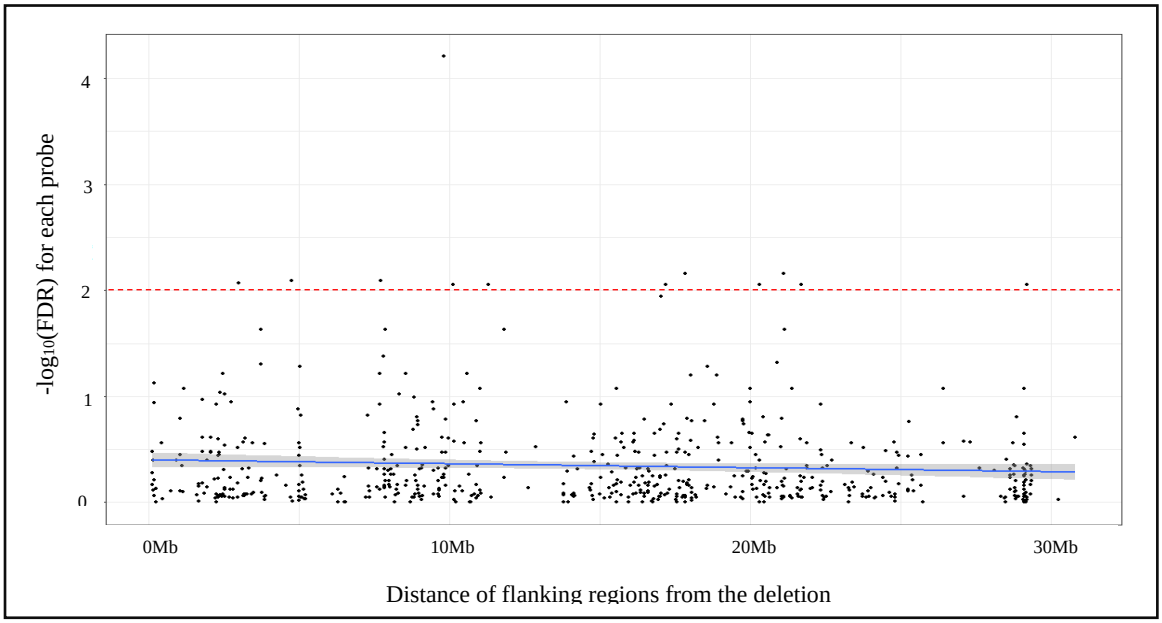


Figure 4- 3: **Manhattan plot for differential gene expression analysis for probes adjacent to 22q11.2 deletion.**

X-axis represents the flanking region distance from the 3Mb 22q11.2 deletion by Mb. Y-axis represents $\log_{10}$(FDR) for each probe in these regions. Each probe in the region is represented by a black dot. The red line represents FDR significance threshold (FDR= 0.05). The blue line represents a linear regression line showed a small positive correlation of 0.0022. As the probes located in longer distance from the deletion, they are unlikely to be differentially expressed. The plot was generated by R.

Analysis of a *TRANS* positional effect sought to investigate whether the 22q11.2 deletion influence probes outside chromosome 22. The results identified only 2.1% probes outside chromosome 22 are significantly differentially expressed (FDR <0.05). This finding is similar to the background rate of differential expression identified at this chapter section 4.5.2.

4.5.5. Investigating dosage sensitive genes in 22q11.2DS patients affected by neuropsychiatric disorders

This analysis sought to highlight candidate genes by identifying differentially expressed genes in 22q11.2 deletion carriers affected by neuropsychiatric phenotypes.

The results of differential gene expression analysis comparing either the 22q11.2DS+ADHD group (n=13) or the 22q11.2DS+ADHD group (n=5) to the 22q11.2DS-PSYCH group (n=12) revealed no 22q11.2 probes are significantly differentially expressed in 22q11.2DS patients affected by ADHD or ASD relative to those unaffected by any psychiatric symptoms (FDR threshold =0.05). To increase our sample size, the expression pattern was compared collectively between all 22q11.2 deletion carriers affected with ADHD or/and ASD (n=14) to the 12 deletion carriers who showed no psychiatric phenotype. Again, the results demonstrated no probes within 22q11.2, within nearby regions, and genome-wide reached the significance cut-off of multiple testing correction (FDR= 0.05).

4.5.6. Identifying functional pathways enriched for differentially-expressed genes

The 785 significantly differentially expressed probes (FDR <0.05) located outside the 3Mb 22q11.2 deleted region in 22q11.2DS patients were selected for building

biological networks and investigating relationships by pathway analysis. This analysis could give an insight about biological functions that possibly underlie the phenotypes manifested in 22q11.2DS.

Pathway analysis using DAVID revealed 25 significant functional groups (FDR <0.05) (Table 4- 6), that were categorised into four enriched functional networks which are involved in:

1) Translation machinery, protein synthesis by ribosomes, and post-translation modifications (PTM) including acetylation, phosphorylation, and ubiquitination.
2) Programmed cell death and apoptosis.
3) Immune system response.
4) Intramembrane organelles and cytosol.

The largest functional network consisted of 13 functional groups, that are involved in translation machinery, protein synthesis by ribosomes, and PTM (Figure 4- 4). 11 of the 13 groups were found in the most significant functional groups (minimum FDR= $8.88x10^{-6}$). In addition, the top functional group with the highest proportion of the differentially significant probes (48.5%) was associated with phosphoprotein (FDR =$1.1x10^{-4}$) included in this functional network (Table 4- 6). Proteins are synthesized by ribosomes that translate mRNA into polypeptide chains, which may then subject to PTMs to produce the mature protein product. PTMs are the modifications that occur covalently and generally via enzymes during or post protein synthesis, which are one of the key components in the process of cell signalling (Verdin & Ott 2014; Drazic et al. 2016). Therefore, we could indicate that a high number of the differentially

expressed probes in 22q11.2DS patients are involved in protein synthesis and PMTs processes.

The second largest functional network is apoptosis and programmed cell death including 6 functional groups (minimum FDR= 0.0015). This apoptosis network included ~3-6% of the differentially expressed probes (Figure 4- 4; Table 4- 6). Apoptosis is a process of programmed cell death which is a vital component of various cellular processes including normal cell turnover, proper development and functioning of the immune system, hormone-dependent atrophy, embryonic development and chemical-induced cell death (Elmore 2007).

A smaller functional network is involved in intramembrane organelles and cytosol. It composed of 4 functional groups with a proportion of differentially expressed probes ranged from 3-14% (minimum FDR= 8.44×10^{-5}) (Figure 4- 4; Table 4- 6).

The smallest functional network is involved in immune system response, particularly lymphocyte and leukocyte activation, that included two significant functional groups (FDR <0.05) enriched with ~4% differentially expressed probes in 22q11.2DS (Figure 4- 4; Table 4- 6).

Table 4- 6: **Significantly enriched functional groups by genome-wide differentially expressed probes in 22q11.2DS patients identified by using DAVID.**

Annotation categories	Functional group	* Total number of genes	** Number of DE genes	Percentage of DE genes	P-value	FDR
SP_PIR_KEYWORDS	Ribosome	563	16	2.7	$6.3x10^{-9}$	$8.9x10^{-6}$
SP_PIR_KEYWORDS	Ribonucleoprotein	563	30	5.1	$1.3x10^{-8}$	$1.9x10^{-5}$
GOTERM_CC_FAT	GO:0005840-Ribosome	415	26	4.4	$3.1x10^{-8}$	$4.2x10^{-5}$
GOTERM_BP_FAT	GO:0006414-Translational elongation	434	18	3.1	$4.1x10^{-8}$	$7.2x10^{-5}$
GOTERM_CC_FAT	GO:0005829-Cytosol	415	83	14.1	$6.1x10^{-8}$	$8.4x10^{-5}$
SP_PIR_KEYWORDS	Ribosomal protein	563	23	3.9	$6.3x10^{-8}$	$8.9x10^{-5}$
SP_PIR_KEYWORDS	Phosphoprotein	563	286	48.5	$7.7x10^{-8}$	$1.1x10^{-4}$
GOTERM_MF_FAT	GO:0003735-Structural constituent of ribosome	405	21	3.6	$3.0x10^{-7}$	$4.4x10^{-4}$
KEGG_PATHWAY	hsa03010: Ribosome	194	16	2.7	$1.1x10^{-6}$	0.0013
GOTERM_BP_FAT	GO:0008624-Induction of apoptosis by extracellular signals	434	17	2.9	$8.6x10^{-7}$	0.0015
GOTERM_BP_FAT	GO:0006412-Translation	434	30	5.9	$8.9x10^{-7}$	0.0016
SP_PIR_KEYWORDS	Protein biosynthesis	563	21	3.6	$1.5x10^{-6}$	0.0022
GOTERM_CC_FAT	GO:0033279-Ribosomal subunit	415	18	3.1	$1.7x10^{-6}$	0.0023
GOTERM_CC_FAT	GO:0030529-Ribonucleoprotein complex	415	39	6.6	$2.7x10^{-6}$	0.0038
GOTERM_CC_FAT	GO:0022626-Cytosolic ribosome	415	14	2.4	$2.9x10^{-6}$	0.0040
GOTERM_BP_FAT	GO:0045321-Leukocyte activation	434	24	4.7	$3.6x10^{-6}$	0.0063
GOTERM_BP_FAT	GO:0043065-Positive regulation of apoptosis	434	34	5.8	$5.8x10^{-6}$	0.0101
GOTERM_BP_FAT	GO:0046649-Lymphocyte activation	434	21	3.6	$5.9x10^{-6}$	0.0103
GOTERM_BP_FAT	GO:0001775-Cell activation	434	26	4.4	$6.0x10^{-6}$	0.0104
GOTERM_BP_FAT	GO:0043068-Positive regulation of programmed cell death	434	34	5.8	$6.7x10^{-6}$	0.0117
GOTERM_MF_FAT	GO:0005198-Structural molecule activity	405	42	7.1	$7.9x10^{-6}$	0.0119
GOTERM_BP_FAT	GO:0010942-Positive regulation of cell death	434	34	5.8	$7.5x10^{-6}$	0.0130
GOTERM_BP_FAT	GO:0006917-Induction of apoptosis	434	28	4.7	$8.6x10^{-6}$	0.0151
GOTERM_BP_FAT	GO:0012502-Induction of programmed cell death	434	28	4.7	$9.1x10^{-6}$	0.0160
GOTERM_CC_FAT	GO:0044445-Cytosolic part	415	18	3.1	$1.7x10^{-5}$	0.0240

* Total number of genes involved in this functional group.
** Number of significantly differentially expressed genes identified in this functional group.
Blue boxes are functional groups involved in translation, protein synthesis and PMTs.
Red boxes highlighted functional groups implicated in cell death and apoptosis.
Grey boxes are functional groups related to intracellular and cytosolic components.
Yellow boxes indicated functional groups associated with immune system.

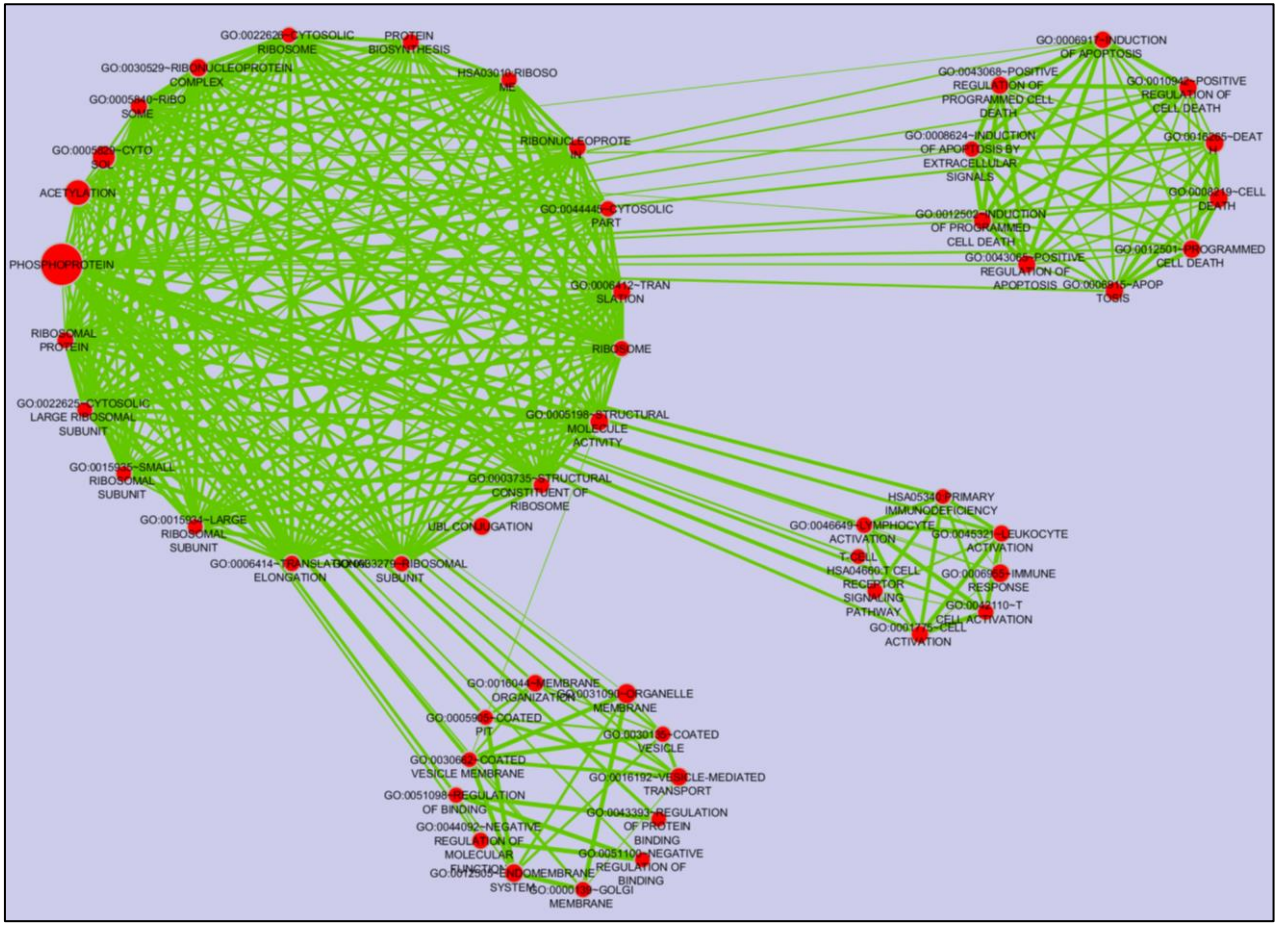


Figure 4- 4: **Significantly enriched functional networks by genome-wide differentially expressed probes in 22q11.2DS patients.**

Each red circle represents a functional group and as the group involved more probes the circle gets larger. The green lines represent there are shared probes between the two connected functional groups. The largest functional network is associated with protein synthesis process, followed by a network associated with apoptosis, then a smaller network associated with cell components, The smallest network is related to immune system. The figure is generated by DAVID and Cytoscape (v.3.4.0) (Shannon et al. 2003).

4.6. Discussion

It has been suggested that the underlying cause of the psychiatric phenotypes associated with 22q11.2DS is the haploinsufficiency of genes within the 22q11.2 region. In this chapter, gene expression was analysed and compared in 22q11.2DS patients and non-deleted controls to identify dosage sensitive genes that are differentially expressed in 22q11.2DS.

The results of differential gene expression analysis showed that 42.5% (n= 51) of the probes within 3Mb 22q11.2 deleted region are significantly differentially expressed in 22q11.2DS patients compared to healthy controls (FDR <0.05). This identified 39 dosage sensitive genes within the 3Mb deleted region that were found to be downregulated in 22q11.2DS patients.

Seven of the 39 genes were also reported to be differentially expressed in a smaller 22q11.2DS cohort (n= 7) (van Beveren et al. 2012), while 29 of them were also identified as dosage sensitive genes in a larger 22q11.2DS cohort (n= 46) (Jalbrzikowski et al. 2015). Six additional dosage sensitive genes identified by Jalbrzikowski and colleagues were not identified by this study, while an additional 8 differentially expressed genes were identified only our study (Table 4- 7). Comparing these findings to those of murine models of 22q11.2DS, 7 of the 39 genes were also found to be differentially expressed in the brain tissues of the *Df1*+/- mice compared to the wild-type mice (Sivagnanasundaram et al. 2007) (Table 4- 7). These observations indicate that there is a degree of consistency between the findings in all 2 gene expression studies conducted in 22q11.2DS (van Beveren et al. 2012; Jalbrzikowski et al. 2015).

Heterozygous knockout mouse models have been generated for a number of the genes that this study identified as being differentially expressed in 22q11.2, including *Dgcr8*, *Zdhh8*, *Comt*, *Sept5*, *Tbx1*, and *Gnb1l* (Gogos et al. 2009; Gogos et al. 1998; Hsu et al. 2007; Mukai et al. 2004; Paylor et al. 2006; Suzuki et al. 2009). Mouse models haploinsufficient for *Dgcr8* and *Zdhh8* manifested impaired PPI (Mukai et al. 2004; Gogos et al. 1998), while those with haploinsufficiency for *Comt* (Gogos et al. 1998), *Sept5* (Suzuki et al. 2009), *Tbx1*, and *Gnb1l* (Paylor et al. 2006) showed defects in sensory gating function. Both phenotypes are cognitive functions achieved by activity in the prefrontal cortex and the hippocampus and are considered as endophenotypes of schizophrenia and psychiatric symptoms in 22q11.2DS patients (Paylor et al. 2001).

Table 4- 7: **Dosage sensitive genes lists identified in differential genes expression studies in 22q11.2DS patients and 22q11.2DS mice models.**

Dosage sensitive genes identified by our study	Dosage sensitive genes identified by (Jalbrzikowski et al. 2015)	Dosage sensitive genes identified by (van Beveren et al. 2012)	Dosage sensitive genes identified by (Sivagnanasundaram et al. 2007)
COMT	*COMT*	*COMT*	*Comt*
CRKL	*CRKL*	*CRKL*	*Ranbp1*
DGCR14	*DGCR14*	*DGCR14*	*Ufd1l*
RANBP1	*RANBP1*	*RANBP1*	*Cldn5*
SLC25A1	*SLC25A1*	*SLC25A1*	*Dgcr6*
UFD1L	*UFD1L*	*UFD1L*	*Sept5*
MRPL40	*C22orf25*	*MRPL40*	*Zdhhc8*
C22orf25	*C22orf29*		*Arvcf*
C22orf29	*C22orf39*		*Prodh*
C22orf39	*CLDN5*		*Rtn4r*
CLDN5	*CLTCL1*		*Htf9c*
CLTCL1	*DGCR2*		*D16H22S680E*
DGCR2	*DGCR6*		
DGCR6	*DGCR8*		
DGCR8	*GNB1L*		
GNB1L	*KLHL22*		
KLHL22	*LOC220686*		
LOC220686	*LZTR1*		
LZTR1	*MED15*		
MED15	*PI4KAP1*		
PI4KAP1	*PIK4CA*		
PIK4CA	*SEPT5*		
SEPT5	*SNAP29*		
SNAP29	*THAP7*		
THAP7	*TMEM191B*		
TMEM191B	*TRMT2A*		
TRMT2A	*TXNRD2*		
TXNRD2	*ZDHHC8*		
ZDHHC8	*ZNF74*		
ZNF74	*DKFZp434-N035*		
DKFZp434-N035	*ATF4*		
AIFM3	*GRAP2*		
DGCR6L	*LOC150356*		
HIC2	*LOC400890*		
HS.572896	*PIK3IP1*		
LOC728139	*RFPL2*		
PI4KA			
SCARF2			
TMEM191C			

Grey labelled genes represent dosage sensitive genes that were identified in our study and the two studies in 22q11.2DS patients and the one study in 22q11.2DS mouse model.
Blue labelled genes represent dosage sensitive genes that were only identified in our study
Red labelled genes represent dosage sensitive genes that were identified in other studies but not in our study.
Yellow labelled genes represent dosage sensitive murine orthologues of human 22q11.2 genes that were identified in 22q11.2DS mouse model but not in 22q11.2DS patients.
Underlined genes represent murine orthologues of human 22q11.2 genes that have same expression levels in both microarray and RTqPCR data.

With reference to the gene expression profiles in (Jalbrzikowski et al. 2015; Guna et al. 2015), 29 of the 39 differentially expressed genes spanning the 3Mb 22q11.2 region were identified to be brain-expressed genes. This highlights these genes as potential candidate loci for the psychiatric phenotypes in 22q11.2DS. There was, however, a substantial variability in the gene expression levels in the 22q11.2DS patients.

An assessment of the gene expression of genes spanning the whole genome between 22q11.2 deletion carriers and non-deletion carriers revealed that 2.3% of probes were significantly differentially expressed (FDR <0.05). Pathway analyses of this data revealed there to be four functional biological networks that were significantly enriched (FDR <0.05) with differentially expressed genes (cell death programing; T-cells activation; protein synthesis, ribosome and translation regulation; and intermembrane organelles regulation). In comparison, the pathway analysis conducted by Jalbrzikowski and colleagues implicated a biological network associated with cellular development, growth, and proliferation which was not associated in this study. However, they also identified significant GOterms that were related to immune response, regulation phosphate metabolic process, and regulation of epithelial to mesenchymal transition (Jalbrzikowski et al. 2015), which are in line with the enriched biological pathways reported in this thesis. While the results of pathway analysis could indicate a role of the immune system abnormalities in 22q11.2DS, it is also possible that the results are being adversely influenced by the use of RNA from blood. As such these observations will require replication when sufficient post-mortem brain tissue from 22q11.2DS patients becomes available.

In analysing the differential expression of genes flanking the 22q11.2 deletion, this study failed to provide significant evidence that the deletion at 22q11.2 causes a large positional effect in *CIS*. A non-significant positive correlation between the probes

being differentially expressed and their distance to the deletion was observed, however, this correlation is very small (r^2= 0.0022) and statistically non-significant (p-value= 0.13). Failure in providing a *CIS* positional effect is most likely due to limited power of the study, if the effect is weak then it is possible that it could be identified by analysing larger numbers of samples. It is also possible that a position effect could be tissue specific and as such it could potentially be revealed by future investigations in other tissue/cell types from 22q11.2DS patients.

Comparing gene expression between 22q11.2DS patients with and without psychiatric illnesses was aimed to highlight candidate genes that could be specifically relevant to psychiatric disease. However, the results failed to provide an evidence for significantly differentially expressed probes in 22q11.2DS patients with psychiatric illness (FDR >0.05). Failure to identify differentially expressed genes is perhaps attributed to very small number of 22q11.2DS patients with available psychiatric phenotype data used for the analysis (n= 26). Jalbrzikowski and colleagues successfully identified 131 probes that were differentially expressed between 22q11.2DS patients with (n= 16) and without ASD (n= 24), however, none of these probes were mapped to the 22q11.2 region (Jalbrzikowski et al. 2015). More samples are therefore required for more fruitful study results.

Despite suggestions that studying gene expression in peripheral blood can be a useful representative for gene expression in the CNS if the relevant gene is expressed in both, its use in this study was nevertheless a major limitation. It is however neither feasible nor ethical to extract brain samples from living 22q11.2DS patients. This limitation

will be addressed when sufficiently sized samples of post-mortem brain tissue from 22q11DS patients are available.

It is also worth mentioning that some of the non-deleted controls used in this analysis were age-matched siblings of the 22q11.2 DS patients. As these cohorts were not fully independent it is possible that this might have diluted the power of the study. The study of Jalbrzikowski and colleagues faced the same problem, which they addressed by carrying out a secondary analysis in which 22q11.2DS patients were compared to a subset of age-matched, unrelated controls. The results of the secondary study were compatible to the results of overall analysis and showed a high correlation to the original study r^2 =0.91 (p-value <$2.2x10^{-16}$) (Jalbrzikowski et al. 2015). This indicates that it is unlikely that this issue adversely affected results presented in this thesis.

PS Power and Samples Size Calculation software version 3.1.2 was used to calculate the power of the sample size used in our gene expression study (Dupont & Plummer 1990). The sample size of 33 22q11.2DS patients vs 35 non-deleted controls had an increasing power to detect differentially expressed genes by fold differences of 2 (SD= 0.30) with false positive rate of 0.05. Also, the sample size for comparing gene expression in the 22q11.2DS patients who are affected with psychiatric disease to those unaffected (14 affected carriers vs 12 unaffected carriers) had a high power to detect differentially expressed genes by fold differences of 2 (SD= 0.11) with false positives rate of 0.05. This indicates that our differential gene expression analyses were enough powered to detect dosage sensitive genes that are influenced by the hemizygous deletion. Therefore, it is possible to say that as we had this statistical

power, in addition to the consistency of our findings with the previous studies, the identified differentially expressed probes in our analyses are highly possible to be realistic in 22q11.2DS patients.

In conclusion, 22q11.2 gene haploinsufficiency is one of the suggested molecular mechanisms behind the neuropsychiatric phonotypes heterogeneity in 22q11.2DS. The results of this analysis revealed that not all genes within the deleted 22q11.2 region are dosage sensitive. Only 42.5% of 22q11.2 probes are differentially expressed in 22q11.2DS patients. This represents 39 genes, of which 29 are brain expressed. Pathway analysis of genome-wide differentially expressed genes, excluding those within the 22q11.2 region, identified biological pathways that could possibly be relevant to the phenotypes seen in 22q11.2DS. Future studies with larger numbers of 22q11.2DS samples and controls are necessary to confirm these preliminary findings and to obtain more powerful study. Moreover, in order to identify genes that are related to psychiatric disorders future studies should analyse either post mortem brain tissue or neuronal cell lines derived from induced pluripotent cell lines obtained from 22q11.2DS deletion patients.

Chapter 5: Investigating the Role of a Second Independent Genetic Variant Located Within the 22q11.2 Region in Psychiatric Disease

5.1. Summary

There are several mechanisms that have been proposed in which deletions at 22q11.2 could potentially increase risk to psychiatric diseases (Williams 2011). In this chapter, one possible molecular mechanism was investigated. We hypothesized that if an individual was a heterozygous carrier of an additional risk variant, that is associated with the increased risk of psychiatric disease, located at 22q11.2 then the deletion could potentially unmask the deleterious effect of this allele that remains within the non-deleted chromosome. These variants might exist in the form of CNVs or SNPs on the non-deleted allele at 22q11.2 and potentially contribute to the diverse behavioural phenotypes.

Association analyses of 5,027 SNPs spanning the 3Mb 22q11.2 deleted region revealed no evidence for significant allelic association with psychiatric phenotypes in 22q11.2DS.

Three regions that potentially harboured a second deletion were identified. Association analyses of potential deletions failed to provide any evidence for a significant enrichment for secondary deletions in 22q11.2DS patients with psychiatric disorder.

The main limitation of this study is the small sample size (n= 76) used in the analyses, which limited the power of the study. Furthermore, a more precise method is required to reliably detect secondary CNVs occurring on the non-deleted 22q11.2 chromosome.

5.2. Introduction

The molecular mechanism that is most likely to contribute to the increased risk of psychiatric phenotype in 22q11.2DS patients is haploinsufficiency of the genes located within the deleted region. Haploinsufficient genes result from a single allele being removed by the hemizygous deletion and this can lead to a reduction in gene expression (Inoue & Lupski 2002). However, as not all individuals with 22q11.2DS develop a psychiatric phenotype, this alone is not likely to explain the clinical phenotypes.

Ordinarily, a recessive mutation will manifest its functional effect when it is present in a homozygous state. Alternatively, recessive mutations can also be deleterious when they co-occur with a second independent recessive mutation that affects the same gene and occurs on the normally wild-type chromosome. This compound heterozygosis typically occurs either when one recessive allele is transmitted from each healthy parent or when one allele is inherited while the other has arisen as a *de novo* mutation (Hochstenbach et al. 2012) (Figure 5- 1).

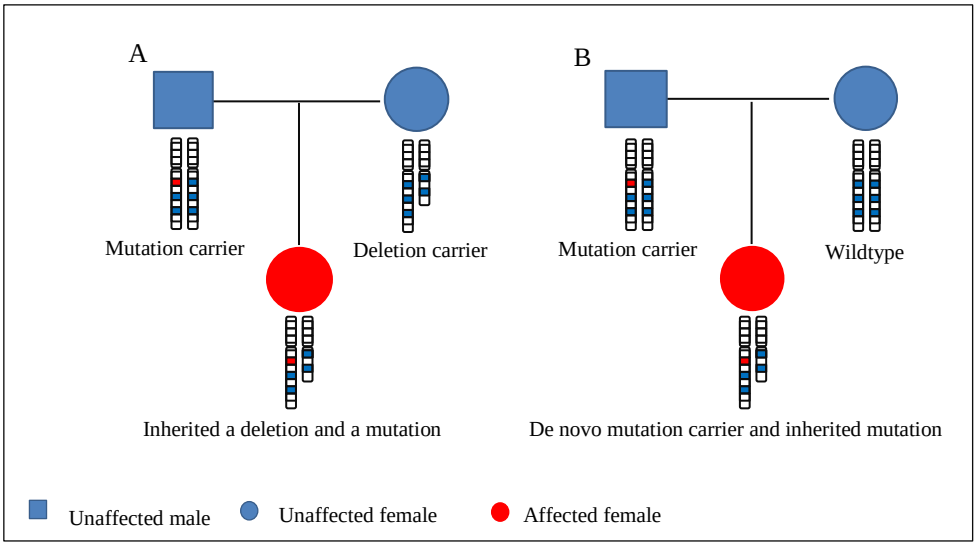


Figure 5- 1: **Unmasking of recessive allele and mode of inherence.**

This diagram explains the possible ways for independent loss of the two alleles in deletion syndromes. ***A.*** *The affected daughter* inherited the recessive mutated *allele from her* unaffected *father also she inherited the deletion from her* unaffected *mother which unmasked the paternal recessive mutation.* ***B.*** *The affected daughter* inherited the recessive mutated *allele from her* unaffected *father and a de novo deletion has arisen and unmasked the recessive paternal mutation (Original figure).*

It has been hypothesized that the incomplete penetrance of the psychiatric phenotype associated with 22q11.2DS could be the result of additional risk alleles that occur in compound heterozygous form on the non-deleted chromosome (Williams 2011). In this context, it is suggested that the primary 22q11.2 deletion could 'unmask' an additional risk allele that would then lead to the patient having an increased risk of psychiatric disease.

5.2.1. Investigating the mechanism of unmasking of recessive alleles in patients with disorders associated with hemizygous deletions and manifesting complex phenotypes

A number of examples have been reported that describe 'unmasking' additional risk alleles by hemizygous microdeletions and resulting in additional or more severe

phenotypes. These unmasked risk alleles could occur in variable forms such as a small structural variant, a single nucleotide variant, or a single nucleotide polymorphism.

An example for unmasking an additional single nucleotide variant by a primary hemizygous deletion was illustrated in a patient with Smith Magenis Syndrome (SMS) and a severe hearing impairment. SMS is caused by hemizygous deletions of chromosome 17p11.2 that span the *MYO15A* gene. The typical phenotype of this syndrome is characterised by multiple congenital anomalies, intellectual impairment, severe developmental disorders, and hearing impairment. ~68% of SMS patients showed variable degree of hearing loss. Conductive hearing impairment occurs more commonly than sensorineural hearing loss (SNHL) in these patients, whereas mixed hearing loss is also a feature of this disorder (Liburd et al. 2001). Also, it has been shown that mutations in *MYO15A* genes in consanguineous families with non-syndromic autosomal recessive profound hearing loss (DFNB3) are responsible for their deafness (Liburd et al. 2001). Therefore, it was hypothesized that a mutation located in *MYO15A* but present on the non-deleted chromosome 17p11.2 could explain the hearing loss variability in SMS patients. In attempting to test this hypothesis, eight SMS patients with SNHL were sequenced to screen for mutations of *MYO15A*. Seven of the eight SMS patients had mild SNHL, while one patient had moderately severe high frequency hearing loss. The latter single SMS patient has been identified to have a second missense mutation on the non-deleted *MYO15A* allele, whereas none of the remaining seven patients have a mutation in *MYO15A* (Liburd et al. 2001).

The second example is of a small CNV that is unmasked when spanned by a primary hemizygous deletion. A single case from patients with idiopathic mental retardation or developmental delay with or without multiple congenital abnormalities was identified with an inherited ~2.2Mb deletion at chromosome 16q23.3 (chr16:

82,660,578-84,826,825 on hg19). This 16q23.3 deletion was transmitted from a healthy hemizygous carrier mother. A second, paternally derived, smaller ~16.04Kb 16q23.3 deletion (chr16: 83,841,341-83,857,382 on hg19) was also detected in this patient, that specifically disrupted *HSBP1* gene which is located within the deleted region (Hochstenbach et al. 2012). This patient had a severe mental retardation, lack of speech, microcephaly, cheilognathopalatoschisis, and bilateral hearing loss. In a *Hsbp1* knockdown mice, a reduced neuroblast migration was observed which possibly indicated that this gene is involved in brain development in these mice (Khodosevich et al. 2009). This finding offers some support to the hypothesis that nullizygousity of the *HSBP1* gene confers a form of mental retardation in this patient (Hochstenbach et al. 2012).

Both examples demonstrate the mechanism of unmasking a second mutation by a primary large hemizygous deletion and the possibility of this mechanism to confer more complex and/or severe forms of the disease.

5.2.2. Investigating the mechanism of unmasking of recessive alleles in 22q11.2DS patients with atypical symptoms

A number of studies have investigated this mechanism in 22q11.2DS. A 22q11.2DS patient was identified to have phenotypic features of Bernard-Soulier Syndrome (BSS), a rare congenital bleeding disorder (Ludlow et al. 1996). BSS is characterized by a prolonged bleeding time, thrombocytopenia, and abnormally large platelets. This 22q11.2DS patient was identified with the BSS phenotype but Southern blotting analysis was not able to detect any platelet GpIbβ protein. Sequencing analysis of this patient revealed a mutation in the GATA binding site in the promoter of the remaining

allele of the *GPIBβ* gene. Therefore, the BSS phenotypes seen in this patient could be explained by a deleted copy of *GPIBβ* allele that unmasks a mutated GATA binding site in the promoter of the remaining allele (Ludlow et al. 1996).

Most recently, McDonald-McGinn and colleagues studied 17 carriers of 22q11.2 deletions who presented rare clinical phenotypes that atypically occurred in less than 10% of 22q11.2DS patients including laryngo-tracheal-oesophageal abnormalities, limb differences, polymicrogyria, myelomeningocele, cleft lip, and genitourinary abnormalities (McDonald-McGinn et al. 2013). This study sought to identify additional mutations that contribute to these atypical phenotypes. Whole exome sequencing analysis was performed on four patients presenting laryngo-tracheal-oesophageal and limb abnormalities. One patient (1) was identified with 2 variants present within the region of 22q11.2. One variant was a frameshift insertion within *SNAP29* gene, whereas the other was a non-synonymous variant within *CLTC1* gene. The frameshift insertion within *SNAP29*, c.388_389insGA, was identified to be paternally inherited and interfered by a *de novo* deletion on the 22q11.2 chromosome inherited from the mother. To support the presence of mutated *SNAP29* in the other patients with atypical phenotypes, the coding exons 5` and 3` splice sites of *SNAP29* were screened in additional 12 patients by Sanger sequencing. Two 22q11.2DS patients were identified to have additional *SNAP29* mutations; one patient (2) had a 5bp deletion in exon 1 (c.28_32delCCGTT), which was transmitted from the mother; the other patient (3) had a missense point mutation (c.265G>A). A point missense mutation in the coiled-coil domain (c.268C>T; p.R90C) was identified in a fourth patient (4) by Targeted exome sequencing and confirmed by Sanger sequencing.

Two patients (1 and 2) share common clinical features that are consistent with phenotypes of Cerebral Dysgenesis, Nephropathy, Icthyosis, and Keratoderma (CEDNIK) syndrome which is an autosomal recessive condition that is caused by truncating mutations in *SNAP29* and characterised by cerebral dysgenesis, neuropathy, ichthyosis, and keratoderma (Sprecher et al. 2005). The third patient (3) expressed atypical symptoms that overlap with the phenotypes described in the autosomal recessive Kousseff syndrome particularly sacral meningocele, conotruncal cardiac anomalies, and dysmorphic features (Kousseff 1984). The fourth patient (4) showed symptoms that have been described in individuals with the heterogeneous Opitz G/BBB syndrome including a combination of cleft lip and palate, hypertelorism, laryngotrachealesophageal anomalies and hypospadias (Robin et al. 1995).

The authors hypothesized that these recessive mutations in *SNAP29* gene are 'unmasked' by the 3Mb hemizygous deletion at 22q11.2 and potentially contributed to express atypical phenotypes caused by autosomal recessive diseases (McDonald-McGinn et al. 2013).

Given the complex clinical symptoms of 22q11.2DS and the large number of genes spanned by the deletions, the mechanism of unmasking additional risk alleles by the deletion could feasibly play a role in the clinical heterogeneity. As the haploinsufficiency of 22q11.2 genes is relevant to most of 22q11.2DS patients, it is logical to hypothesize that the background genetic variants that exist on the non-deleted 22q11.2 chromosome could have an important function in conferring additional risk to variable phenotypes such as neuropsychiatric disease. To date no study has investigated the presence of second rare functional variants in the remaining haploinsufficient chromosome in 22q11.2DS patients with psychiatric phenotypes. Alternatively, a wide range of studies have investigated the association of common

variants, minor allele frequency >0.1, with the expression of the neuropsychiatric symptoms in 22q11.2DS patients. These studies sought to test the hypothesis that 22q11.2DS patients with psychiatric phenotypes would be expected to carry common risk variants for neuropsychiatric disease on the remaining chromosome.

5.2.3. Investigating common variants in 22q11.2 region

The high susceptibility to develop psychiatric diseases in patients carrying a 22q11.2 deletion provides excellent opportunities for identifying candidate genes that might be implicated in cognitive, behavioural, and psychiatric phenotypes (Bassett & Chow 2008; Meechan et al. 2011). A large number of the genes located within the commonly deleted region are expressed in brain (Maynard et al. 2003; Meechan et al. 2009), therefore a large number of studies have tested the association of common variants present in some of these genes with neuropsychiatric phenotypes in 22q11.2DS patients. As discussed in detail in chapter 1 section 1.3.2, the promising results identified in some of these studies are not replicated by other studies; either by finding evidence for association in the opposite direction, or failure to provide a supportive replication for association with psychiatric phenotypes in 22q11.2DS.

5.3. Aims of this chapter

It has been suggested that, despite carrying a single, relatively homogeneous deletion, the reason why the 22q11.2 deletion confers an incomplete penetrance to develop psychiatric illnesses could be due to the presence of genetic variants that remain on the non-deleted chromosome. To date no study has systematically screened the entire non-deleted 22q11.2 chromosome and as such no genetic variant at 22q11.2 has been unambiguously shown to influence the increased risk of psychiatric disease.

Work described in this chapter will investigate a well-characterised cohort of 22q11.2 deletion carriers to identify genetic variants present on the remaining, non-deleted region of 22q11.2. The cohort will be investigated by two approaches:

1) An association analysis of 22q11.2 deletion carriers with and without psychiatric disorders for common SNPs spanned by the 22q11.2 deletion.
2) Secondary deletions spanned by the primary 22q11.2 deletion can be detected as regions carrying SNPs that have null genotypes in 22q11.2DS patients. Small duplications within the deleted region can be identified by CNV calling using PennCNV. The frequency of rare deletions and duplications identified by these approaches will be compared between 22q11.2 deletion carriers with and without a psychiatric disorder.

5.4. Materials and methods

5.4.1. Samples

DNA samples of the well-characterized 76 22q11.2DS patients were used for the analysis conducted in this chapter. The characterization of the 76 patients were fully explained in chapter 3 sections 3.5.1 and 3.5.2.

The patients were grouped into 3 psychiatric diagnosis groups based on the manifested phenotypes including: 22q11.2DS+ADHD, 22q11.2DS+ASD, and 22q11.2DS+PSYCH (Table 5- 1).

Specific details of the 22q11.2DS patients are described in chapter 2 sections 2.1. Protocols of DNA extraction from the collected biological samples from 22q11.2DS patients, DNA quantification and quality control are fully explained in chapter 2 section 2.2.

Table 5- 1: **Number of 22q11.2 deletion carriers affected and unaffected by psychiatric disorders.**

Categories	Description	Number of 22q11.2DS individuals
22q11.2DS+ADHD	22q11.2 deletion carriers with ADHD	29
22q11.2DS+ASD	22q11.2 deletion carriers with ASD	18
22q11.2DS+PSYCH	22q11.2 deletion carriers with ADHD and/or ASD	37
22q11.2DS-PSYCH	22q11.2 deletion carriers without ADHD or ASD	35

5.4.2. Genotyping SNPs within the non-deleted 22q11.2 chromosome

5.4.2.1. Common SNPs for association analysis

The DNA for all 22q11.2DS patients was genotyped using an Illumina Infinium Human CoreExome-24 BeadChip array as described in chapter 2 section 2.3.1.

Genotype data for samples passing the initial QC was subjected to stringent QC procedures which were aimed at excluding low quality genetic variants. All QC was performed using Plink (Purcell et al. 2007) . SNPs were removed if they had one of the following standard QC criteria:

1) Had a minor allele frequency (MAF) < 0.05.
2) Had a missing rate of > 0.01 over all samples.
3) The genotype frequencies significantly (p-value $<1.0x10^{-6}$) deviated from those expected under Hardy-Weinberg equilibrium (HWE).

The SNPs that survived these QC procedures were then used as a backbone to impute the genotypes of the majority of SNPs that were not included on the array. All imputation protocols are fully explained in chapter 2 section 2.4. Following imputation, the SNPs at 22q11.2 were further filtered to select those with good imputation quality (SNPs with MAF ≥0.01 and INFO score ≥0.8).

A summary of the genotypes quality control analyses is described in Figure 5- 2.

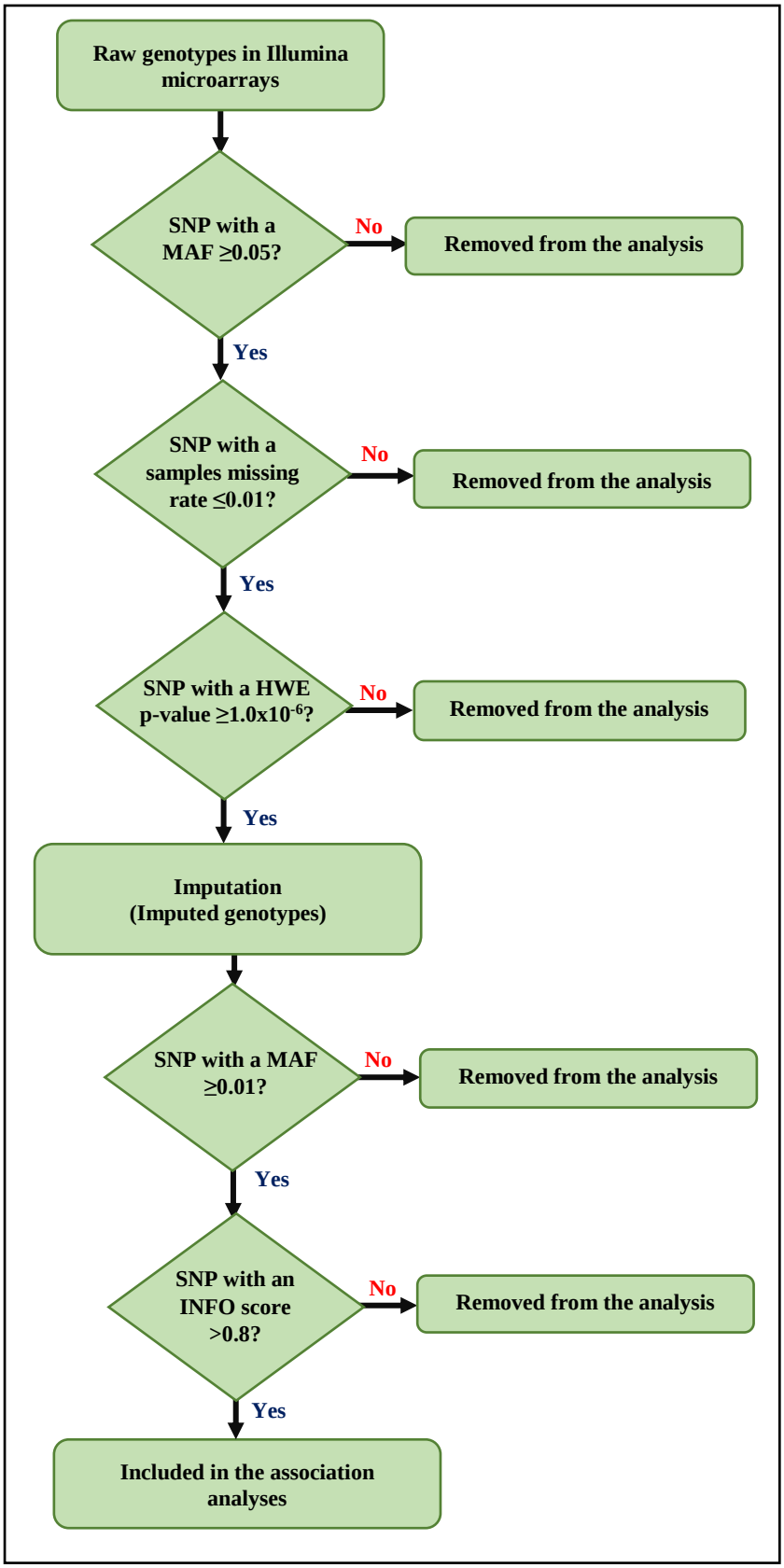


Figure 5- 2: **Flowchart for common raw and imputed genotypes quality control.**

5.4.2.2. Association analysis of common SNPs

Association analysis was performed for each SNP spanning 22q11.2 by comparing its allele frequency in the 22q11.2DS patients affected with ADHD (22q11.2DS+ADHD), ASD (22q11.2DS+ASD), and either ADHD or ASD (22q11.2DS+PSYCH) to its frequency in the deletion carriers with no psychiatric illness (22q11.2DS-PSYCH) (Table 5- 1).

Association analysis of the SNPs within the shared ~1.5Mb deletion spanning LCR22s A-B was conducted using all 22q11.2 deletion carriers, while the SNPs spanning LCR22s B-D were tested for association using only carriers of the ~3Mb deletion. The association analysis was carried out by logistic regression using Plink (Purcell et al. 2007). The resulting p-value of each marker was corrected for multiple testing using either Bonferroni correction (in which the significance threshold was calculated by dividing an alpha of 0.05 by number of tested SNPs) (Dunn 1961), or the less conservative FDR correction (FDR calculated for p-value for each marker using R using Benjamini–Hochberg approach) (Benajmini & Hochberg 1995).

5.4.2.3. Association analysis of common SNPs spanned by dosage sensitive genes

In chapter 4 section 4.5.3, the differential gene expression analysis results revealed 39 genes spanning the 3Mb 22q11.2 region that are dosage sensitive in 22q11.2DS patients compared to non-deleted controls (Table 5- 2). A targeted association analysis was performed, which focussed only on the SNPs that spanned the 39 dosage sensitive genes. Again, association analysis of SNPs spanned by dosage sensitive genes within the shared ~1.5Mb deletion (LCR22s A-B) was performed on all 22q11.2 deletion

carriers, whereas the SNPs spanned by dosage sensitive genes within LCR22s B-D were tested for association in only the carriers of the ~3Mb deletion.

Table 5- 2: **22q11.2 dosage sensitive genes identified in 22q11.2DS patients by differential gene expression analysis.**

Target gene	CHR	Region	Target gene	CHR	Region
DGCR6	22	1.5Mb	*DGCR6L*	22	1.5Mb
DGCR2	22	1.5Mb	*TMEM191B*	22	1.5Mb
DGCR14	22	1.5Mb	*PI4KAP1*	22	1.5Mb
HS.572896	22	1.5Mb	*ZNF74*	22	3Mb
SLC25A1	22	1.5Mb	*SCARF2*	22	3Mb
CLTCL1	22	1.5Mb	*KLHL22*	22	3Mb
MRPL40	22	1.5Mb	*MED15*	22	3Mb
C22orf39	22	1.5Mb	*DKFZp434N035*	22	3Mb
UFD1L	22	1.5Mb	*PI4KA*	22	3Mb
CLDN5	22	1.5Mb	*PIK4CA*	22	3Mb
SEPT5	22	1.5Mb	*SNAP29*	22	3Mb
GNB1L	22	1.5Mb	*CRKL*	22	3Mb
LOC728139	22	1.5Mb	*AIFM3*	22	3Mb
C22orf29	22	1.5Mb	*LZTR1*	22	3Mb
TXNRD2	22	1.5Mb	*THAP7*	22	3Mb
COMT	22	1.5Mb	*HIC2*	22	3Mb
C22orf25	22	1.5Mb	*TMEM191C*	22	3Mb
DGCR8	22	1.5Mb	*LOC220686*	22	3Mb
TRMT2A	22	1.5Mb			
RANBP1	22	1.5Mb			
ZDHHC8	22	1.5Mb			

5.4.3. Identification of secondary CNVs within the non-deleted 22q11.2 chromosome

5.4.3.1. Identifying secondary deletions

A second deletion that spanned the primary 22q11.2 deletion would result in a null genotype for any SNP spanned by both deletions. To identify 22q11.2DS patients carrying SNPs at 22q11.2 that had null genotypes the analysis was restricted to the array-based genotypes only.

Therefore, all 22q11.2 deletion carriers were screened for null genotypes at the SNPs spanning LCR22s A-B, while only the carriers of a ~3Mb 22q11.2 deletion were used to identify null genotypes at the SNPs spanning LCR22s B-D.

In order to exclude low quality SNPs that were more prone to failure, an additional 71 individuals who did not carry a deletion at 22q11.2 were also screened. SNPs that were also null in the non-deleted controls were unlikely to be pathogenic deletions, but more likely to indicate low quality genotyping. The genotype cluster plots of the remaining null SNPs were manually inspected for genotype calling quality. A flow diagram to illustrate the process of null genotypes SNPs screening in Figure 5- 3.

5.4.3.2. Identifying secondary duplications

Genome-wide CNVs were called by PennCNV using non-imputed genotype data as explained in chapter 2 section 2.5.1. All SNPs located in the 3Mb deleted region were used for PennCNV calling to identify secondary duplications that possibly occur in the non-deleted 22q11.2 chromosome. Duplication calling was set at a liberal criteria aimed at identifying CNVs that span a minimum 3 constitutive SNPs with a minimum size of 12bp.

5.4.3.3. Association analysis of secondary 22q11.2 CNVs

Association analysis was performed by comparing frequencies of null genotype SNPs at 22q11.2 between 22q11.2DS patients who were either affected or unaffected by psychiatric diseases, as defined in Table 5- 1 by Fisher's exact test using R. Significance of association findings was assessed by Bonferroni multiple testing correction to account for the identified Null SNPs (Bonferroni threshold= 0.017).

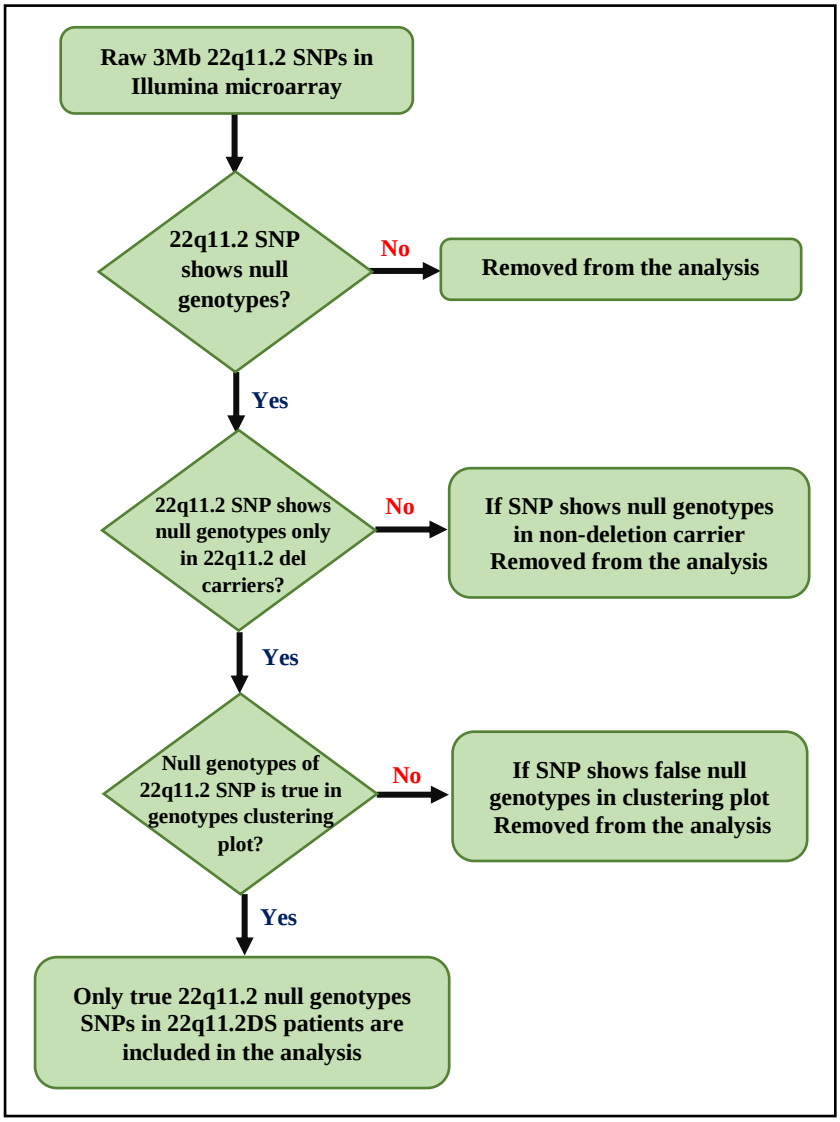


Figure 5- 3: **Flowchart for null genotypes SNPs screening.**

5.5. Results

5.5.1. SNPs quality control

Following the stringent genotypic QC procedures, of the 790 3Mb 22q11.2 SNPs, 177 high quality SNPs (MAF ≥0.05, sample missing rate ≤0.01 and HWE p-value ≥$1.0x10^{-6}$) remained and these were then used as a backbone to impute the genotypes of the majority of SNPs that were not included on the array.

Imputation resulted in 6,595, however, only 5,027 common SNPs with a high imputation quality (MAF ≥0.01 and INFO score ≥0.8) that spanned the 3Mb deletion at 22q11.2 were selected. Of the 5,027 22q11.2 SNPs, 3,327 SNPs spanned LCR22s A-B and 1,700 SNPs spanned LCR22s B-D.

5.5.2. Investigating whether SNPs within the non-deleted 22q11.2 chromosome are associated with psychiatric disease in 22q11.2DS patients

These association analyses sought to establish whether genetic variants within the non-deleted 22q11.2 chromosome could potentially have a role in the increased risk of 22q11.2DS to develop psychiatric illnesses.

5.5.2.1. ADHD in 22q11.2DS

The results of association analysis of all 5,027 SNPs spanning the 3Mb deleted region between 22q11.2DS+ADHD and 22q11.2DS-PSYCH patients identified 89 SNPs that were nominally associated with ADHD, with the strongest evidence for association at rs759406 (p-value = 0.0024) spanned by *PRODH* gene. A Manhattan plot of all SNPs is presented in Figure 5- 4, A1. However, after interpreting the results to account for the analysis of 5,027 SNPs, no genetic variant met the stringent significance threshold

set by Bonferroni correction (Bonferroni p-value= $9.9x10^{-6}$). Given that Bonferroni assumes that all SNPs are segregating independently, the results were also corrected using the less stringent FDR, which allows for the correlation between SNPs. This again revealed that no SNP exceeded the 5% threshold of FDR (most significant SNP was rs759406; FDR= 0.998). A Manhattan plot of all SNPs is presented in Figure 5- 4, A2.

5.5.2.2. ASD in 22q11.2DS

Association analysis that compared 22q11.2 common variants between patients in 22q11.2DS+ASD and 22q11.2DS-PSYCH groups identified 76 SNPs that were nominally associated with ASD, with rs759406 having the strongest evidence for association (p-value = 0.0020) which is spanned by *PRODH*. All SNPs with nominal p-value are presented in a Manhattan plot in Figure 5- 4, B1. However, after multiple testing correction, no genetic variant survived the conservative Bonferroni significance threshold (Bonferroni p-value= $9.9x10^{-6}$) and also no variant survived the less conservative FDR significant threshold (most significant SNP was rs759406; FDR= 0.968). A Manhattan plot of all SNPs after FDR correction is presented in Figure 5- 4, B2.

5.5.2.3. Psychiatric illnesses in 22q11.2DS

Results of association analysis that compared the 22q11.2 SNPs in deletion carriers with either ADHD and/or ASD (22q11.2DS+PSYCH) to those unaffected carriers (22q11.2DS-PSYCH) identified 43 SNPs that were nominally associated with psychiatric illnesses, with the strongest evidence for association at rs759406 (p-value

= 0.0024) spanned by *PRODH* gene. All SNPs are presented in a Manhattan plot in Figure 5- 4, C1. However, after Bonferroni multiple testing correction, no genetic variant reached the stringent significance level (Bonferroni p-value= $9.9x10^{-6}$) and also no genetic variant met the less stringent FDR significance level (most significant SNP was rs759406; FDR= 0.999). A Manhattan plot of all SNPs is presented in Figure 5- 4, C2.

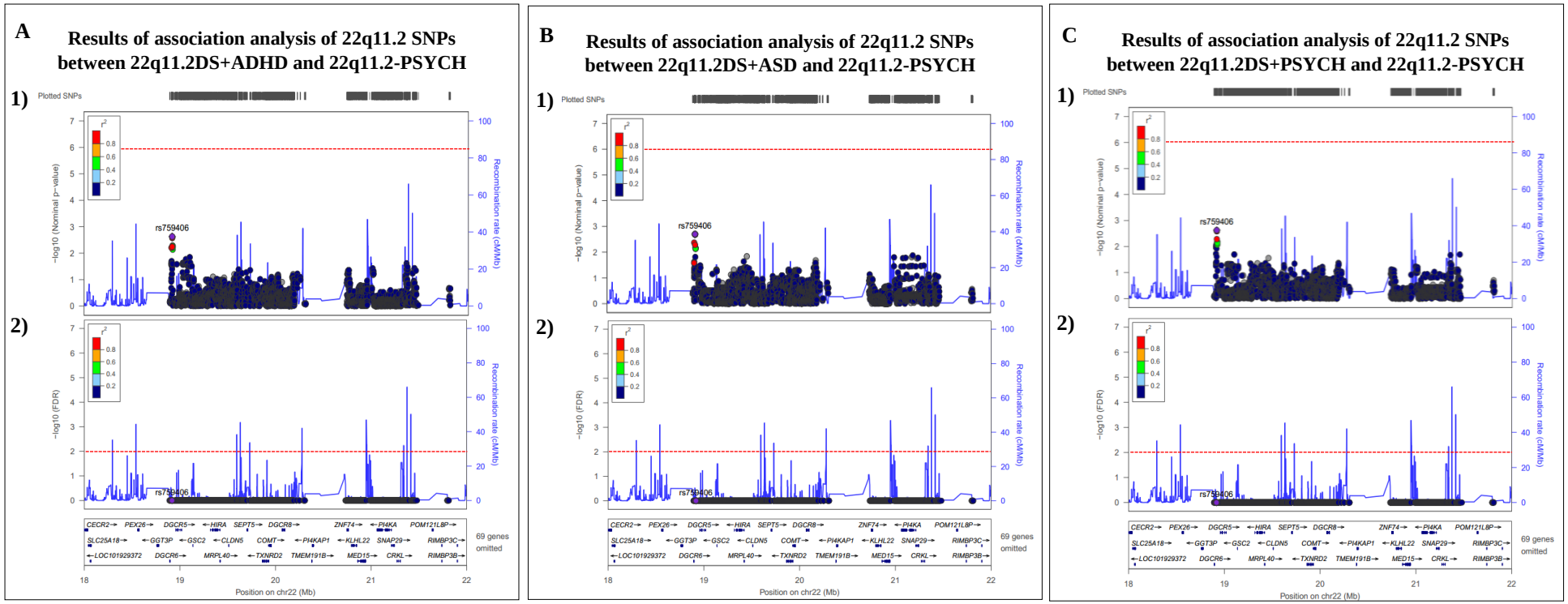


Figure 5- 4: **Manhattan plots for association analyses results for 3Mb 22q11.2 SNPs.**

Association analysis of 3,327 SNPs within the shared ~1.5Mb deletion spanning LCR22s A-B was conducted using all 76 deletion carriers, while the 1,700 SNPs spanning LCR22s B-D were tested for association using only 73 carriers of the ~3Mb deletion. P-values for each marker is plotted and represented by a single circle. The Y-axis shows the $-\log_{10}$ for the p-value of each marker and p-value thresholds are highlighted by red lines. X-axis shows some of spanning genes and genomic location (hg19). **Plots 1)** Uncorrected p-value for each marker is plotted. The corrected Bonferroni significance threshold (Bonferroni p-value= 9.9×10^{-6}) is highlighted by the red line. **Plots 2)** P-values for all SNPs are corrected by FDR multiple correction. The FDR threshold (FDR= 5.0×10^{-2}) is highlighted by red. In the plots, SNPs with the strongest association indicated by a purple circle and a black bold SNP name. Correlated SNPs are coloured based on the correlation coefficient scale in the plots. The plots are generated by Locus Zoom tool (Pruim et al. 2010).

A. Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+ADHD and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs759406) with nominal p-value= 0.0024 and FDR= 0.998.

B. Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+ASD and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs759406) with nominal p-value= 0.0020 and FDR= 0.968.

C. *Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+PSYCH and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs759406) with nominal p-value= 0.0024 and FDR= 0.999.*

5.5.3. Investigating whether SNPs within the non-deleted 22q11.2 chromosome and spanned by the dosage sensitive genes are associated with psychiatric diseases in 22q11.2DS patients

Of the 5,027 SNPs that passed the genotypic QC procedures, 2,103 SNPs spanned a dosage sensitive gene at 22q11.2. Association analyses comparing allele frequency of SNPs in the non-deleted chromosome and spanned by dosage sensitive genes between affected and unaffected 22q11.2 deletion carriers with psychiatric diseases sought to establish whether an additional risk allele occurs in the non-deleted haploinsufficient genes could potentially have additional role in the elevated risk of 22q11.2DS to develop psychiatric illnesses in addition to the low expression level of the spanning genes.

5.5.3.1. ADHD in 22q11.2DS

The results of association analysis that compared the allele frequency of the 2,103 SNPs spanned by 22q11.2 dosage sensitive genes between 22q11.2DS+ADHD and 22q11.2DS-PSYCH patients identified the 13 SNPs that were nominally associated with ADHD with the strongest evidence for association at rs2075269 (p-value = 0.024) spanned by *THAP7* gene. A Manhattan plot of all SNPs with nominal p-values is presented in Figure 5- 5, A1. However, after accounting for the analysed 2,103 SNPs, no genetic variant met the stringent significance threshold set by Bonferroni correction (Bonferroni p-value= 2.4×10^{-5}). We also corrected the results using the less stringent FDR which also revealed that no SNP exceeded the 5% threshold of FDR (most significant SNP was rs2075269; FDR= 0.996). A Manhattan plot of all SNPs with FDR corrected p-values is presented in Figure 5- 5, A2.

5.5.3.2. ASD in 22q11.2DS

Thirty-one SNPs that were nominally associated with ASD with rs165659 (spanned by *PI4KA*) having the strongest evidence for association (p-value = 0.013) the allele frequency of the 2,103 SNPs was compared between the individuals of 22q11.2DS+ASD and 22q11.2DS-PSYCH groups. A Manhattan plot of all SNPs is presented in Figure 5- 5, B1. However, after multiple testing correction by Bonferroni, no single SNP met the stringent Bonferroni significance threshold (Bonferroni p-value= $2.4x10^{-5}$). Also, no SNP reached the less stringent FDR significance threshold (most significant SNP was rs165659; FDR= 0.999). A Manhattan plot of all SNPs is presented in Figure 5- 5, B2.

5.5.3.3. Psychiatric illnesses in 22q11.2DS

The results of an association analysis that compared the common variants spanned by dosage sensitive genes in 22q11.2 between 22q11.2DS+PSYCH and 22q11.2DS-PSYCH patients identified only 3 SNPs nominally associated with psychiatric illness in 22q11.2DS. The strongest evidence for association identified at rs2075269 (p-value = 0.021) spanned by *THAP7*. A Manhattan plot of all SNPs is presented in Figure 5- 5, C1. However, following multiple testing correction, no genetic variant met the Bonferroni significance threshold (Bonferroni p-value= $2.4x10^{-5}$) and FDR significance threshold (FDR= 0.05) (most significant SNP was rs2075269; FDR= 0.995). A Manhattan plot of all SNPs is presented in Figure 5- 5, C2.

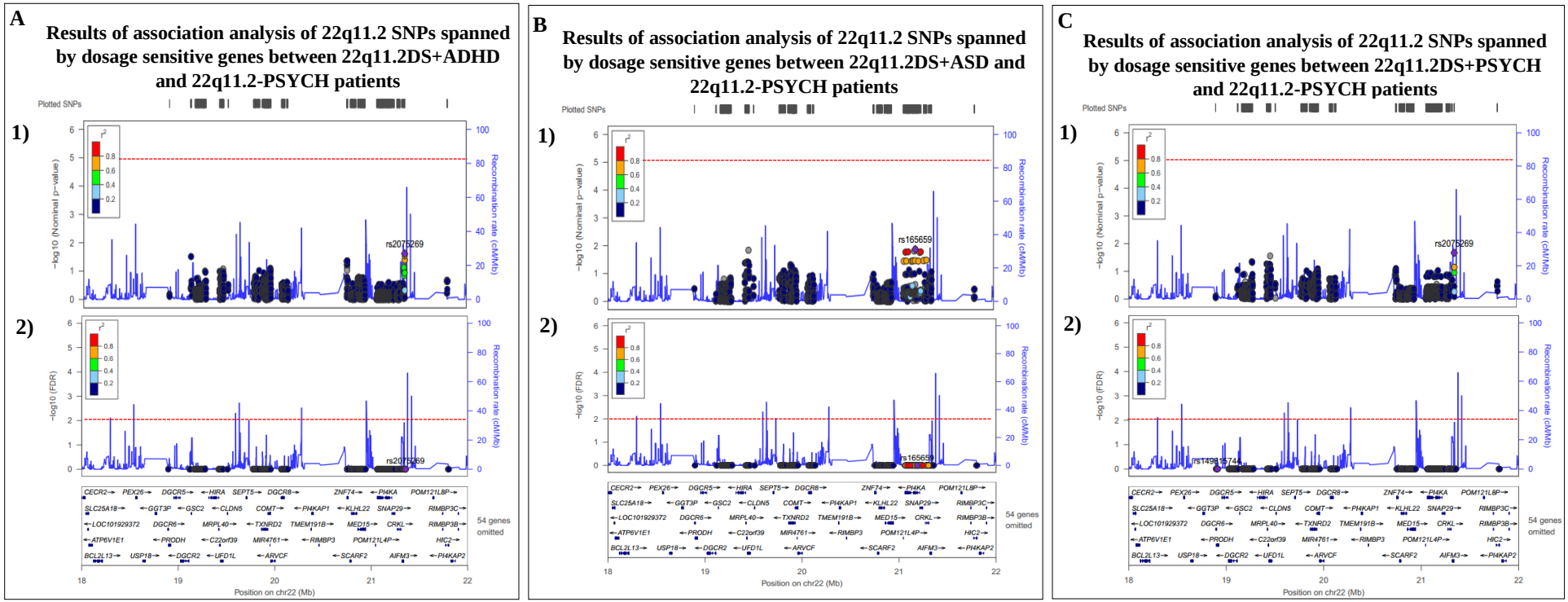


Figure 5- 5: **Manhattan plots for association analyses results for 3Mb 22q11.2 SNPs spanned by dosages sensitive genes.**

Association analysis of 1,158 SNPs within the shared ~1.5Mb deletion spanning LCR22s A-B was conducted using all 76 deletion carriers, while the 945 SNPs spanning LCR22s B-D were tested for association using only 73 carriers of the ~3Mb deletion. P-values for each marker are plotted and represented by a circle. The Y-axis shows the -$\log_{10}$ p-value for each marker and p-value thresholds are highlighted by red lines. X-axis shows some spanning genes and genomic location (hg19). **Plots 1)** Uncorrected p-value for each marker is plotted. The corrected Bonferroni significance threshold (Bonferroni p-value= 2.4×10^{-5}) is highlighted by the red line. **Plots 2)** P-values for all SNPs are corrected by FDR multiple correction. The FDR threshold (FDR= 5.0×10^{-2}) is highlighted by red. In the plots, SNPs with the strongest association indicated by a purple circle and a black bold SNP name. Correlated SNPs are coloured based on the correlation coefficient scale in the plots. The plots are generated by Locus Zoom tool (Pruim et al. 2010).

A. Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+ADHD and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs2075269) with nominal p-value= 0.024 and FDR= 0.996.

B. Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+ASD and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs165659) with nominal p-value= 0.013 and FDR= 0.999.

C. Showing Manhattan plots for association analysis comparing individuals in 22q11.2DS+PSYCH and 22q11.2DS-PSYCH groups. SNP with the strongest association to ADHD at (rs2075269) with nominal p-value= 0.021 and FDR= 0.995.

5.5.4. Investigating secondary CNVs within the 22q11.2 region

5.5.4.1. Identification of null genotype SNPs within the 22q11.2 region

Screening the 790 SNPs that spanned the ~3Mb deletion in the 76 patients with 22q11.2DS and 71 non-deleted controls revealed a total of 127 SNPs with null genotypes. 77 of the 127 SNPs were also null in the non-deleted controls and were therefore excluded. 47 of the remaining 50 SNPs were subsequently excluded as manual inspection of their genotype cluster plots indicated that they were due to low quality genotype calls. This resulted in the identification of 3 SNPs (exm1586223, exm1588273, and rs9604911) that had null genotypes in the 22q11.2DS individuals. The genotype cluster plots of each of these 3 SNPs are presented in Figure 5- 6.

The distance to the closest proximal and distal SNP that was successfully genotyped was used to estimate the maximum size of any potential deletion that each null genotype SNP could be indicating. As can be seen in Figure 5- 7 the 3 null genotype SNPs do not implicate a single gene. The SNP exm1588273 is located within exon 4 of *THAP7* gene at the distal part of the 3Mb deletion. The potential deletion implicated by its null genotype would have an estimated maximum size of ~640bp and would span chr22:18,918,666-18,936,993 (on hg19). This would potentially disrupt the gene *THAP7* (Figure 5- 7, B). The SNP exm1586223 is located in exon 10 of the gene *TANGO2*, which is located within the 1.5Mb deleted region (Figure 5- 7, C). The estimated maximum size of this deletion could be ~1.6Kb, spanning chr22:20,050,940- 20,052,566 (on hg19). Finally, the null genotype at rs9604911 could potentially implicate a deletion with an estimated maximum size of ~18.3Kb (chr22:21,354,936-21,355,576 on hg19). Although, the SNP is not itself located within

a gene, its implicated deletion would potentially disrupt the adjacent *PRODH* gene (Figure 5- 7, D).

5.5.4.2. Identification secondary duplications within the 22q11.2 region

Analysis of the 76 samples carrying 22q11.2 deletions using PennCNV did not identify any duplication that spanned at least 3 SNPs within the 3Mb region. Thus, no secondary duplications were included in the association analyses for potential secondary 22q11.2 CNVs in this study.

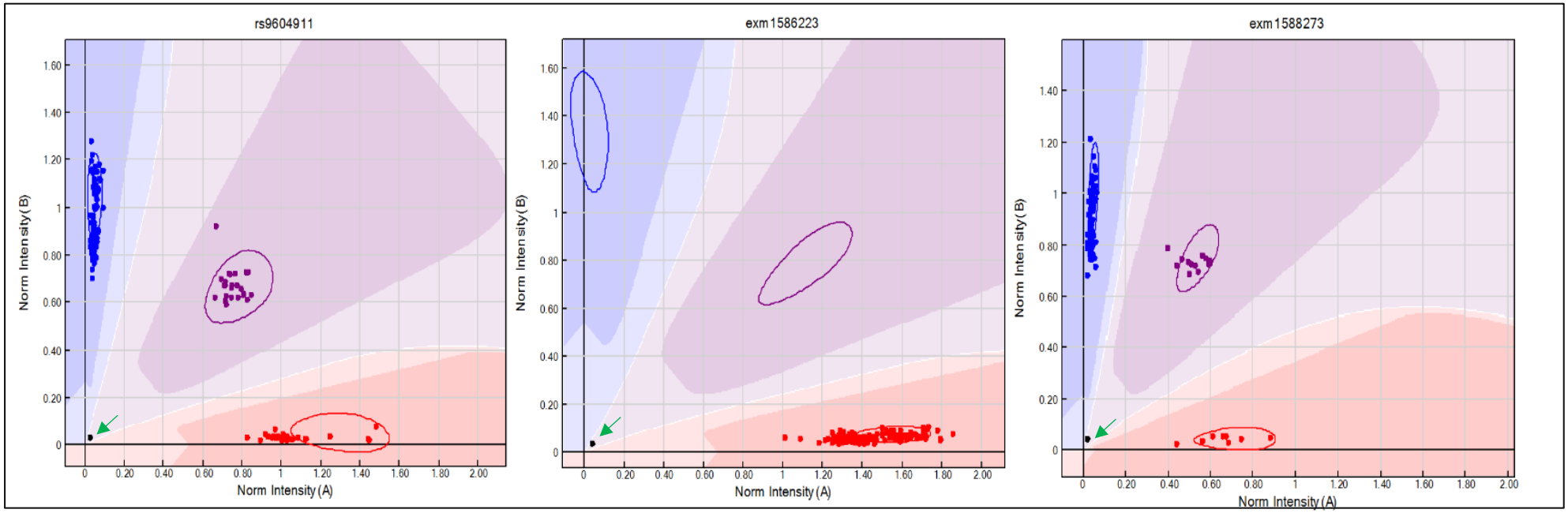


Figure 5- 6: **GenomeStudio SNPs clustering plots for the three SNPs with true null genotypes.**

SNPs for the 147 individuals are plotted, including 22q11.2 deletion carriers and non-deleted controls. Each plot represents a SNP with the SNP identifier indicated above the plot. Each circle represents an individual. The X-axis represents the intensity value of allele A and the Y-axis is the intensity value of allele B. Blue dots are samples with homozygous allele B, red dots are samples with homozygous allele A, and purple dots are samples heterozygous allele A and B. Samples with true null genotypes SNPs are represented by black dots with allele A and B signal intensities equal to 0 and located in the bottom left corners of the plot and indicated by green arrows (GenomeSudio).

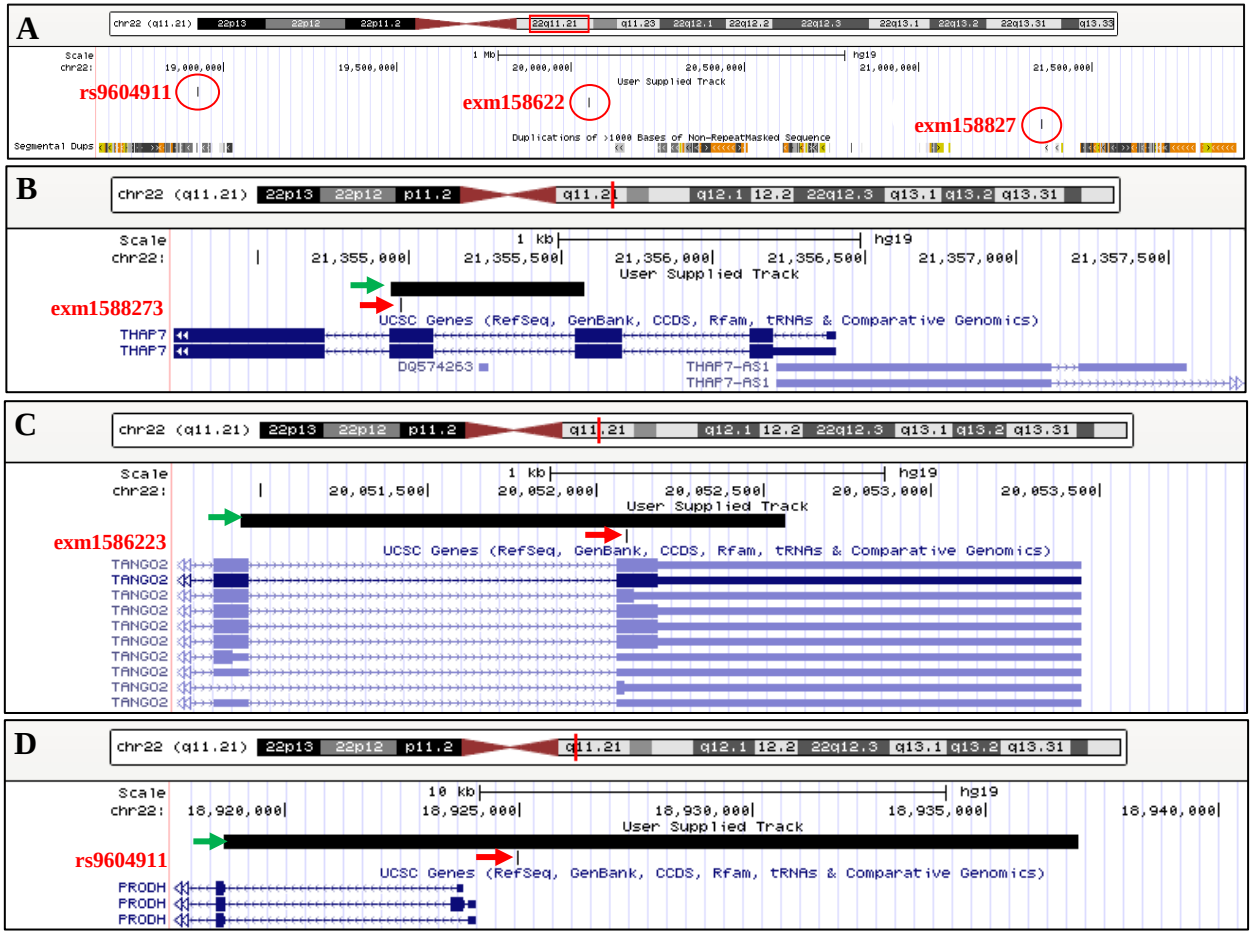


Figure 5- 7: **Snapshots for the three Null SNPs on the UCSC genome browser.**

A. Showing the 3Mb 22q11.2 deletion region (chr22:18,658,219-21,865,185 on hg19).The three Null SNPs are highlighted by red circles with markers names (on 1Mb scale) and showed the location of these SNPs in the 3Mb region. **B.** Null exm1588273 SNP, indicated by red arrow, with the potential ~640bp deletion, indicated by green arrow, spanned by *THAP7* gene (on 1Kb scale). **C.** Null exm1586223 SNP, indicated by red arrow, with the potential ~1.6Kb deletion, indicated by green arrow, spanned by *TANGO2* gene (on 1Kb scale). **D.** Null rs9604911 SNP, indicated by red arrow, with the potential ~18.3Kb deletion, indicated by green arrow, spanned by *PRODH* gene (on 10Kb scale) (Genome.ucsc.edu, n.d.).

5.5.4.3. Association analysis of potential secondary deletions at the 22q11.2 region

Comparing the rate of null genotype SNPs identified in 22q11.2DS patients with and without psychiatric disorders sought to establish whether additional secondary CNVs occurring within the non-deleted region of chromosome 22q11.2 could potentially play a role in the increased risk of psychiatric disorders.

Association analysis identified no evidence for significantly increased rate of null genotype SNPs, Bonferroni threshold= 0.017, in either the 29 22q11.2DS patients with ADHD (uncorrected Fishers exact test p-value= 0.56), the 18 patients with ASD (uncorrected Fishers exact test p-value= 1.0), or the 37 patients with a childhood psychiatric disorder (uncorrected Fishers exact test p-value= 1.0) when compared to those with no psychiatric disease (n=35) (Table 5- 3).

Table 5- 3: **Results of association analyses of secondary 22q11.2 deletions between 22q11.2DS patients affected and unaffected by psychiatric diseases.**

A									**Fisher exact two-tailed p-value**
22q11.2DS+ADHD					**22q11.2DS-PSYCH**				
Carrying a Null SNP		**Not carrying a Null SNP**		**Total**	**Carrying a Null SNP**		**Not carrying a Null SNP**		**Total**	
Number of 22q11.2DS patients	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	**Number of 22q11.2DS patients**	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	
2	6.9%	27	93.1%	29	1	2.6%	34	97.1%	35	0.59
B									**Fisher exact two-tailed p-value**
22q11.2DS+ASD					**22q11.2DS-PSYCH**				
Carrying a Null SNP		**Not carrying a Null SNP**		**Total**	**Carrying a Null SNP**		**Not carrying a Null SNP**		**Total**	
Number of 22q11.2DS patients	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	**Number of 22q11.2DS patients**	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	
0	0.0%	18	100%	18	1	2.6%	34	97.1%	35	1.0
C									**Fisher exact two-tailed p-value**
22q11.2DS+PSYCH					**22q11.2DS-PSYCH**				
Carrying a Null SNP		**Not carrying a Null SNP**		**Total**	**Carrying a Null SNP**		**Not carrying a Null SNP**		**Total**	
Number of 22q11.2DS patients	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	**Number of 22q11.2DS patients**	**Frequency**	**Number of 22q11.2DS patients**	**Frequency**	**-**	
2	5.7%	35	94.6%	37	1	2.6%	34	97.1%	35	1.0

5.6. Discussion

It is widely known that 22q11.2DS is associated with a wide spectrum of neuropsychiatric phenotypes. This chapter set out to identify whether additional genetic variants present on the non-deleted 22q11.2 chromosome influenced the increased risk to psychiatric disorders seen in 22q11.2DS patients. A number of studies have tested SNPs at 22q11.2 for association with psychiatric disease in 22q11.2DS patients, but these have been selected because of their relevance to candidate genes and have generated inconclusive results.

This study systematically tested 5,027 SNPs spanning the deletion at 22q11.2 for association with childhood psychiatric disorders in patients with 22q11.2DS. The results described in this chapter revealed no evidence for a significant association for common SNPs with neuropsychiatric symptoms in 22q11.2DS children (ADHD and/or ASD) (Bonferroni p-values $>9.9x10^{-6}$, FDR >0.05).

A reasonable explanation for the failure of this study to provide a positive finding is the small sample size involved in the analyses. Only 76 samples for well-characterized 22q11.2DS patients with available psychiatric phenotypes information were analysed. While this is comparable to previous candidate based studies, due to the large number of SNPs (n= 5,027) analysed, the corrected significance threshold of this study was more stringent. A larger sample size would be required to achieve adequate statistical power (Klein 2007). It is however worth noting that while more samples were originally ascertained for this study, the rigorous quality control procedures resulted ~19 recruited samples being excluded.

The Genetic Power Calculator (Purcell, Cherny & Sham 2003) was used to calculate the power of this study. The sample size (37 22q11.2DS affected vs 35 22q11.2DS

unaffected) had an 80% power to detect an association at p-value= 0.05 for a risk allele with a genotype relative risk (GRR) of 2.03 and allele frequency of 0.06. It is however possible that as 22q11.2DS patients already carry a primary insult (i.e. the large deletion) then the effect size of an additional risk allele at chromosome 22q11.2 could be larger in these patients than is present in studies of other psychiatric disease cohorts. For example, it has been shown that all 22q11.2DS children have a lower IQ when compared to their non-deleted siblings (Niarchou et al. 2014). It is therefore plausible that this compromised cognition could result in risk alleles for psychiatric disease having a greater relative risk in children carrying the deletion. While unlikely, it is potentially possible that the samples used in this thesis had sufficient power to detect an additional risk allele whose disease penetrance was amplified in this way. An estimate of the sample sizes required to obtain an 80% power to detect lower GRR are given in the Table 5- 4.

Table 5- 4: **The required sample sizes to obtain an 80% power to detect a common risk variant (allele frequency = 0.05) with a range of GRR.**

GRR	Sample size			
	*P-value of 0.05		**P-value of $9.9x10^{-6}$	
	Cases	Controls	Cases	Controls
1.05	13,807	13,117	48,689	46,255
1.1	3,483	3,309	12,284	11,670
1.2	1,022	971	3,607	3,427
1.3	461	438	1,626	1,545
1.4	263	250	927	881
1.5	170	162	601	571
1.6	120	114	432	410
1.7	89	85	315	300
1.8	69	66	245	233
1.9	55	52	196	186
2.0	45	43	161	153
2.01	44	42	158	150

* Uncorrected p-value threshold.
** Bonferroni corrected p-value threshold in our study.

From Table 5-, it can be seen that a larger sample size is required to detect risk alleles with decreased effect size.

While this study has focussed on testing risk alleles for association with psychiatric disease in patients with 22q11.2DS, any true risk allele present at 22q11.2 would also be identified (albeit with a reduced GRR) in association studies of non-deleted patients with psychiatric disease compared to controls. Several GWA studies of ASD have now suggested a number of common genetic variants associated with the risk of ASD. They yielded ~112 genome-wide significant loci spanned by 200 ASD candidate genes (Weiss et al. 2009; Wang et al. 2009; Anney et al. 2010; Salyakina et al. 2010). From all the reported findings for ASD and autism, none of these risk loci were found within the typical deleted region of 22q11.2 chromosome. Moreover, GWAS analysis was performed in a large case-control cohort for five psychiatric disorders; ASD, ADHD, bipolar disorder, major depressive disorder, and schizophrenia including 33,332 cases and 27,888 controls of European ancestry. The study identified SNPs at four significant risk loci at genome-wide significance threshold (p-value $<5x10^{-8}$). These regions are on chromosomes 3p21 and 10q24, and SNPs mapped to two L-type voltage-gated calcium channel subunits, *CACNA1C* and *CACNB2* (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). Again none of these reported psychiatric risk loci mapped to the 22q11.2 region. The results of the largest and very recent GWAS study conducted for schizophrenia identified 108 distinct genome-wide significant loci being associated with schizophrenia. Three loci of the 108 are found in chromosome 22 (chr22_39987017_D at 39,975,317-40,016,817; rs9607782 at 41,408,556-41,675,156; and rs1023500 at 42,315,744-42,689,41 on hg19) (Ripke et

al. 2014), however none of the three loci was located in the common 22q11.2 deleted region.

Taken as a whole, it is therefore possible that the findings of these GWA studies support those presented in this chapter, and that no psychiatric disease risk variants are present in the region of 22q11.2.

The second part of this chapter attempted to identify evidence for small potential deletions within the remaining 22q11.2 chromosome of 22q11.2DS patients. These were identified by detecting SNPs that had null genotypes and this revealed 3 potential secondary deletions present in three 22q11.2DS patients. Association analysis indicated that these potential deletions were not significantly enriched in the 22q11.2DS patients with psychiatric phenotypes (p-value >0.017). Nevertheless, it is of potential interest that the Null SNP (and a potential secondary deletion) carried by two 22q11.2DS patients with ADHD are located in the genes *THAP7* and *PRODH*, as previous studies have identified nominally significant evidence for association to ADHD with common variants spanned by both genes (Neale et al. 2010; Stergiakouli et al. 2012; Hinney et al. 2011; Yang et al. 2013; Mick et al. 2010; Neale et al. 2008). In addition, transcripts for *THAP7* were found to be significantly downregulated in 22q11.2DS patients compared to non-deleted controls.

This analysis of potential secondary deletions was intended as a preliminary investigation, however, the approach of identifying null SNPs was limited by the distribution of probes on the genotyping platform. In addition, a single marker with null genotypes is not a perfect representation of a small deletion, while this study did attempt to exclude them, null genotypes can also occur as a result of a failure in the

sample genotyping. A future study should utilise a next generation sequencing (NGS) in order to obtain a more detailed and precise understanding of the background variants in the non-deleted 22q11.2 chromosome in 22q11.2DS patients (McDonald-McGinn et al. 2013). This approach would allow a more systematic analysis and would not be dependant on the distribution of pre-selected SNPs that are present on the genotyping microarray. Future NGS based studies would also benefit from the increased power of studying larger numbers of individuals with 22q11.2DS in order to reduce both the false positive and negative rate of CNV detection.

Chapter 6: Investigating Additional Mutations Located Outside the 22q11.2 Region in 22q11.2DS Patients

6.1. Summary

A number of CNVs have now been implicated in different neurodevelopmental disorders and psychiatric diseases. It also has been shown that patients who carry more than one 'pathogenic' CNV are more likely to manifest a more severe or an atypical phenotype. This chapter will investigate the hypothesis that additional structural variants may contribute to the increased risk of neuropsychiatric illness in 22q11.2DS and that these 'second hit' CNVs may be enriched for CNVs previously associated with neurodevelopmental disorders and psychiatric diseases.

To address this issue, a genome-wide analysis of CNVs was performed on 73 children with 22q11.2DS and the rate of large, rare CNVs was compared between those who are affected with a psychiatric disorder to those who are unaffected.

The overall rate of CNVs ≥500Kb in 22q11.2DS patients with ADHD or ASD was 0.18 and 0.11 which was 3-times and 1.8-times higher than the rate seen in unaffected 22q11.2DS subjects (rate = 0.06). However, this increased rate of CNVs ≥500Kb was not significant in either the ADHD group (p-value= 0.22) or the ASD (p-value=0.44). Similarly, a non-significant (p-value= 0.29) enrichment of CNVs ≥500Kb was revealed in 22q11.2DS affected by any childhood psychiatric illnesses (rate= 0.14) compared to those unaffected (rate= 0.06).

As only a subset of CNVs are expected to be of 'pathogenic' relevance, the analyses focussed only on CNVs that overlapped with one of the 88 'pathogenic' CNVs that

previously have been reported to increase risk to developmental delay and neuropsychiatric disease. This analysis again revealed no significant evidence for an enrichment of 'pathogenic' CNVs in the 22q11.2DS patients with either ADHD (rate= 0.036; p-value= 0.85), ASD (rate= 0.056; p-value= 0.73), or any childhood psychiatric illnesses (rate= 0.028; p-value= 0.89).

The main limitation of this study is the small sample size (n= 73) used in the analyses, which limited the power of the study. Larger sample size is required for more powerful results.

6.2. Introduction

One of the molecular mechanisms that could explain the wide spectrum of neuropsychiatric phenotypes in 22q11.2DS is the presence of an additional pathogenic mutation. The additional mutation could occur in the form of a second risk allele within the non-deleted 22q11.2 chromosome. As illustrated in chapter 5, there was no evidence for a significant association of 22q11.2 SNPs occur in the 22q11.2 region with psychiatric phenotypes in 22q11.2DS.

However, it has also been hypothesized that additional mutations could occur outside the 22q11.2 deleted region. Such genomic variations may be enriched for CNVs previously identified in neurodevelopmental delay and neuropsychiatric diseases and increase the risk to psychiatric diseases in 22q11.2DS independently from the primary deletion. Such pathogenic mutations are called "second hits" (Williams et al. 2013).

6.2.1. Involvement of copy number variants in neuropsychiatric disorders

Rapid developments in the microarray technologies have provided opportunities to examine the human genome for copy number variations (CNVs) (Cook & Scherer 2008). These advances have resulted in a number of genomic rearrangements being associated with neurodevelopmental and neuropsychiatric disorders. For example, Kirov and colleagues studied CNVs in schizophrenia cases and found a 2.26-fold excess of large, rare CNVs in these cases compared to controls (p-value= 0.00027) (Kirov et al. 2009). The study also showed that structural variants identified in schizophrenia patients are enriched with CNVs previously implicated as susceptibility factors for neuropsychiatric diseases including 22q11.2 deletions, a susceptibility factor for schizophrenia; 14.0–15.4Mb 17p12 deletions, known to cause hereditary neuropathy with liability to pressure palsies; and 14.9–16.4Mb 16p13.1 duplications,

which has been previously implicated as a susceptibility factor for autism. Moreover, the study provided the first significant support for the association of 15q11.2 deletions and schizophrenia (p-value= 0.026) (Kirov et al. 2009).

Several other independent studies have also investigated the role of CNVs in other neuropsychiatric disorders (McQuillin et al. 2011; Costain et al. 2014; Georgieva et al. 2014; Rucker et al. 2016; Fry et al. 2016). Recent studies strengthen the hypothesis that the same CNVs might contribute to the susceptibility of different neuropsychiatric disorders such as schizophrenia, ADHD, intellectual disability (ID), and developmental delay (DD) (Rees et al. 2014; Williams et al. 2012; Williams et al. 2010; Elia et al. 2011; Jarick et al. 2014; Girirajan et al. 2010; Malhotra & Sebat 2012; Kaminsky et al. 2013).

A list of 15 implicated CNVs, that have been previously either reported to be associated or showed a trend of enrichment with psychiatric diseases, were studied in a large schizophrenia case-control cohort (n= 13,198) (Rees et al. 2014). The results of the study showed that 13 of the 15 previously implicated CNVs occur in excess in schizophrenia cases. For six of these, the difference in the rate of CNVs detected in cases and controls was nominally significant (p-value <0.05). However, when this dataset was combined with previous large data (~20,732 cases and ~81,821 controls), 11 CNVs showed a highly significant association with schizophrenia (p-values $<4.1x10^{-4}$) (Table 6- 1) (Rees et al. 2014).

Table 6- 1: **Results of meta-analysis of the 15 previously implicated CNVs on a large combined schizophrenia dataset** (Rees et al. 2014).

Locus	Cases frequency (%)	Controls frequency (%)	OR (95% CI)	P-values
1q21.1 deletion	0.17	0.021	8.35 (4.65–14.99)	$4.1x10^{-13}$
1q21.1 duplication	0.13	0.037	3.45 (1.92–6.20)	$9.9x10^{-5}$
NRXN1 deletion	0.18	0.20	9.01 (4.44–18.29)	$1.3x10^{-9}$
3q29 deletion	0.082	0.0014	57.65 (7.58–438.44)	$1.5x10^{-9}$
WBS duplication	0.066	0.0058	11.35 (2.58–49.93)	$6.9x10^{-5}$
VIPR2 duplication	0.11	0.069	1.54 (0.77–3.09)	0.27
15q11.2 deletion	0.59	0.028	2.15 (1.71–2.68)	$2.5x10^{-10}$
AS/PWS duplication	0.083	0.0063	13.20 (3.72–46.77)	$5.6x10^{-6}$
15q13.3 deletion	0.14	0.019	7.52 (3.98–14.19)	$4.0x10^{-10}$
16p13.11 duplication	0.31	0.13	2.30 (1.57–3.36)	$5.7x10^{-5}$
16p11.2 distal deletion	0.063	0.018	3.39 (1.21–9.52)	0.017
16p11.2 duplication	0.35	0.030	11.52 (6.86–19.34)	$2.9x10^{-24}$
17p12 deletion	0.094	0.026	3.62 (1.73–7.57)	0.0012
17q12 deletion	0.036	0.0054	6.64 (1.78–24.72)	0.0072
22q11.2 deletion	0.29	0.00	NA (28.27–∞)	$4.4x10^{-40}$

Regarding ADHD, eight association studies have investigated genome-wide CNVs in children with ADHD, and these have showed promising results (Elia et al. 2010; Elia et al. 2011; Jarick et al. 2014; Stergiakouli et al. 2012; Williams et al. 2012; Williams et al. 2010; Lionel et al. 2011; Lesch et al. 2011). All of the eight studies considered childhood ADHD patients; however, only three showed an increased CNV burden in the ADHD patients relative to controls (Williams et al. 2010; Stergiakouli et al. 2012; Williams et al. 2012). Williams and colleagues provided evidence supporting the enrichment of large, rare CNVs in ADHD children, mainly due to an increase rate of 16p13.11, or 15q13.13 duplications (Williams et al. 2010; Williams et al. 2012). In addition, an increased burden of rare CNVs affecting genes belong to metabotropic glutamate receptor gene networks was identified in multiple ADHD cohorts (Elia et al. 2011). Furthermore, an excess of deletions and duplications at chromosome 6q25.2-q27 at the *PARK2* locus was reported in ADHD children (Jarick et al. 2014). These studies have also demonstrated an enrichment of ADHD-related CNVs at loci

previously associated with different neurodevelopmental disorders, such as schizophrenia and/or autism was reported, and provided further supportive evidence of the hypothesis that genetic susceptibility factors might be shared by these disorders (Elia et al. 2010; Lionel et al. 2011; Williams et al. 2010). In total, these studies have resulted in a total of 15 CNVs being reported to be significantly implicated in childhood ADHD (Table 6- 2).

Table 6- 2: **CNV loci previously reported to be significantly associated with ADHD.**

Locus (hg19)	CNV type	OR (95% CI)	P-value	References
Chr1:56053497-56064495	Deletion	22.85 (3–190)	$1.54x10^{-3}$	(Elia et al. 2011)
Chr1:72317292-72328395	Duplication	Infinity	$3.9x10^{-4}$	(Elia et al. 2011)
Chr2:81419297-81446082	Duplication	5.07 (1–23)	$3.83x10^{-2}$	(Elia et al. 2011)
Chr3:1844168-1859889	Deletion	4.44 (1–13)	$8.81x10^{-3}$	(Elia et al. 2011)
Chr3:7183953-7197236	Deletion	Infinity	$8.14x10^{-5}$	(Elia et al. 2011)
Chr4:113772340-113788584	Duplication	5.07 (1–23)	$3.83x10^{-2}$	(Elia et al. 2011)
Chr5:65027976-65046520	Deletion	22.85 (3–190)	$4.68x10^{-4}$	(Elia et al. 2011)
Chr6:146657076-146694047	Duplication	15.24 (3–72)	$5.6x10^{-6}$	(Elia et al. 2011)
Chr6:162659756-162767019	Duplication	NA	$2.8x10^{-4}$	(Jarick et al. 2014)
Chr7:126525124-126536202	Deletion	Infinity	$8.14x10^{-5}$	(Elia et al. 2011)
Chr7:153495598-153564827	Duplication	15.24 (3–72)	$4.08x10^{-4}$	(Elia et al. 2011)
Chr11:88269449-88351661	Deletion	38.12 (5–298)	$1.36x10^{-6}$	(Elia et al. 2011)
Chr15:29811982-30232981	Duplication	2.22 (1.5-3.6)	0.000178	(Williams et al. 2012)
Chr16:15156431-18174650	Duplication	13.88 (2.3–82.2)	0·0008	(Williams et al. 2010)
Chr19:38427720-38444834	Deletion	5.33 (2–17)	$4.95x10^{-3}$	(Elia et al. 2011)

A number of genome-wide association studies have investigated genomic structural variations in large case-control cohorts for association with intellectual disability, developmental delay (ID/DD), multiple congenital abnormalities, and ASD (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013). Cooper and colleagues compared CNVs in 15,767 children with intellectual disability and various congenital defects to 8,329 adult controls. A significant excess of large CNVs was found among cases relative to controls. This enrichment was evident at 250Kb and becomes more obvious with larger CNV size. Fifty-nine pathogenic CNVs were

identified including 14 novel or previously weakly supported candidates (Cooper et al. 2012). Kaminsky and colleagues studied cases with various indications, including unexplained DD, ID, dysmorphic features, multiple congenital anomalies (MCA), ASD, or clinical features suggestive of a chromosomal syndrome (n= 15,749). Comparing CNVs in cases to those in 10,118 published controls, identified 14 deletions, and 7 duplications that were significantly enriched in cases (Kaminsky et al. 2013). Girirajan and colleagues on the other hand, studied children who had developmental delay with or without congenital malformations. The main finding reported a recurrent, 520Kbb 16p12.1 microdeletion associated with childhood developmental delay (Girirajan et al. 2010). Malhorta and colleagues reviewed findings of other studies that reported significant evidences for a number of large CNVs that are strongly associated with neuropsychiatric and neurodevelopmental disorders (Malhotra & Sebat 2012).

A total number of 72 large, rare CNVs were identified in these studies and showed overrepresentation in cases with neurodevelopmental phenotypes, or genomic disorders relative to controls (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013). Most of these CNVs showed a significant association with multiple diagnostic categories, with the most significant CNVs spanning 22q11.2; 3q29; 15q11.2; 15q11.2-13.1; 16p12.1; and 16p11.2 regions. All 72 CNVs were implicated as being potentially pathogenic CNVs as they occurred in higher frequency in cases compared to controls. Lists are given in this chapter section 6.4.2.2.

6.2.2. The second hit hypothesis

Girirajan and colleagues have established testing the second hit hypothesis in large case-control cohorts of ID/DD children. Initially they found an enrichment for a 520.8Kb 16p12.1 deletion (chr16: 21,942,499-22,432,499 on hg18) in ID/DD children relative to controls (p-value= 0.000086, OR= 3.78) (Table 6- 3) (Girirajan et al. 2010).

Table 6- 3: **Frequency of 16p12.1 microdeletion in ID/DD cases and controls** (Girirajan et al. 2010).

	ID/DD Cases		Controls		Significance	
	16p12.1 del	**Total**	**16p12.1 del**	**Total**	**P-value**	**OR**
Discovery set	20	11,873	2	8,540	0.0009	7.2
Replication set	22	9,254	6	6,299	0.028	2.5
Combined	42	21,127	8	15,199	0.000086	3.78

Following this finding, the group studied the relevance of additional CNVs on the phenotype of ID/DD patients carrying a 16p12.1 deletion. Interestingly, they found a higher frequency of additional CNVs in ID/DD patients with 16p12.1 microdeletion relative to population controls (Table 6- 4) (Girirajan et al. 2010). ID/DD patients with 16p12.1 microdeletion and a second large CNV were observed to manifest more severe or distinct phenotypes than the typical phenotypes of the syndrome (Girirajan et al. 2010).

Table 6- 4: **Enrichment for 'Second hit' CNVs among 16p12.1 microdeletion carriers** (Girirajan et al. 2010).

	ID/DD cases with 16p12.1 deletion			Controls			Significance	
	Two large CNVs	**Total**	**Percentage**	**Two large CNVs**	**Total**	**Percentage**	**P-value**	**OR**
Discovery set	6	20	30%	9	217	4.1%	0.0005	9.7
Replication set	4	22	18.2%	12	254	4.7%	0.029	4.48
Combined	10	42	23.8%	21	471	4.4%	0.000057	6.65

This study was then extended in order to investigate a cohort of 2,312 children who carry one of the 72 CNVs that have been previously implicated to be associated with neurodevelopmental phenotypes or genomic disorders including ID/DD, autism, cardiac abnormalities, speech deficits, craniofacial features, and other previously defined congenital malformations (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013). It was found that 8.7% of the 2,312 affected children to have at least one additional large CNV, of which 0.5% of those carrying a second CNV carried a third variant. The presence of a second CNV was found to be significantly enriched in children with a phenotypically variable genomic disorder (10.3%) in comparison to those children with a syndromic genomic disorder (5.1%) (p-value= 4.49×10^{-6}, OR= 2.13) (Girirajan et al. 2012). This possibly indicates that children with disorders associated with variable phenotypes are more likely to harbour a second hit than those with syndromes with typical phenotypes.

In order to understand the clinical outcomes of additional CNVs, a further investigation was performed by focusing on three CNVs (16p11.2 deletion and duplication, 1q21.1 deletion, and 16p12.1 deletion). A phenotypic scoring system including a checklist of clinical phenotypic features described for sub-telomeric and balanced *de novo* chromosomal imbalances was used to study the clinical outcomes in the individuals carrying one of the three CNVs. The scales of this approach ranged from 1 to 14 with lower scores indicate few features, whereas higher scores mean multiple features. The results showed that children with multiple large CNVs have higher phenotypic scores compared to those with a single large CNV, indicating that more heterogeneous phenotypes can be seen in the patients with a second large CNV. The presence of multiple large CNVs was significantly associated with more variable

phenotypes in patients carrying a primary 16p11.2 deletion (p-value= 0.008) and 1q21.1 deletion (p-value= 0.006) (Figure 6- 1) (Girirajan et al. 2012).

In addition, children with autism in this cohort were studied to assess the impact of the additional CNVs on their IQ. A strong association of median IQ with the number of genes spanned by rare CNVs was observed in autistic children. Lower median IQ (below the intellectual disability threshold IQ <70) was found in children with more affected genes by CNVs (≥18 genes) compared to children with less affected genes (<18 genes) (p-value= 0.002) (Figure 6- 2) (Girirajan et al. 2012).

These studies provided strong evidence for the second hit hypothesis in patients with neurodevelopmental delay. Those who carry an additional large CNV with a primary pathogenic CNV have been proved to have an increased risk to develop more severe phenotypes and/or additional clinical features than those with only a single large CNV.

Based on that, it is possible that the increased risk of neuropsychiatric disorders in 22q11.2DS is contributed to by the presence of an additional CNV elsewhere in the genome.

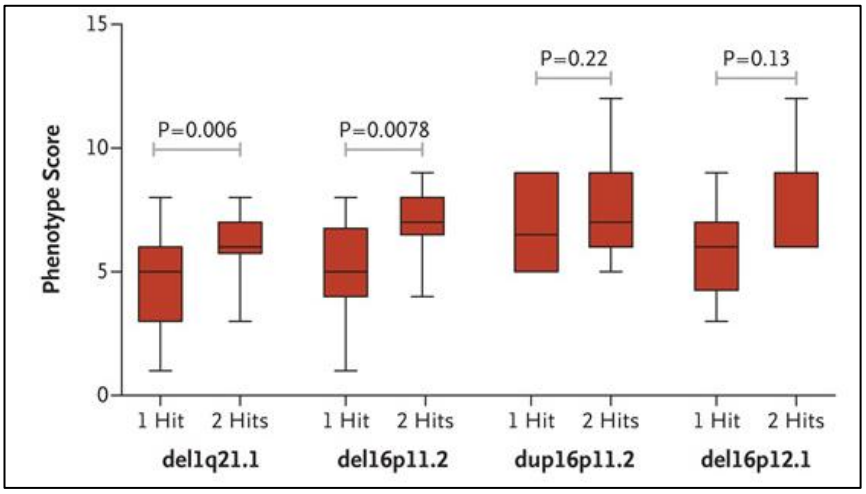


Figure 6- 1: **Phenotypic scores for children with one of the four CNVs that are associated with the most homogenous disorders.**

Children with one of these CNVs were grouped into either a group with a single additional CNV (1 hit) or a group with multiple additional CNVs (2 hits). Phenotypic scores ranged from 1-14 in which higher scores indicate more variable phenotypes. The median values of phenotypic scores for cases with 2 hits were equally higher than those cases with 1 hit among all four CNVs. Significant evidence was found in 1q21.1 deletion and 16p11.2 deletion and a trend was observed in both 16p11.2 duplication and 16p12.1 deletion (Girirajan et al. 2012).

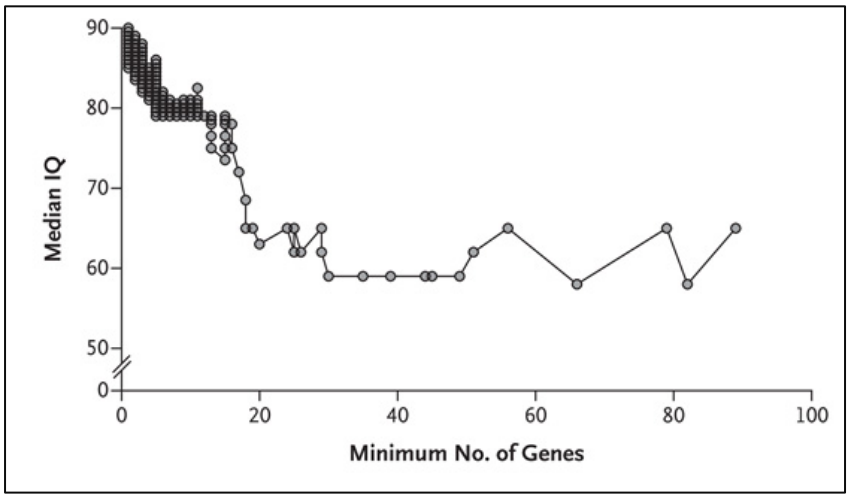


Figure 6- 2: **Association between median IQ and the number of genes spanning CNVs in children with autism.**

The plot shows that lower IQ level was observed in autistic children with larger number of disrupted genes by CNVs. Children with ≥18 affected genes have decreased median IQ (below the threshold for intellectual disability= 70 points). A significant association of IQ level with number of affected genes was observed when children with ≥18 affected genes were compared to those with <18 affected genes (p-value= 0.002 by the Wilcoxon rank-sum test) (Girirajan et al. 2012).

6.2.3. The second hit hypothesis in 22q11.2DS

There are a few other studies that have investigated the second hit hypothesis in 22q11.2DS (Bassett et al. 2008; Li et al. 2012; Williams et al. 2013; Girirajan et al. 2010). However, none of these studies reported significant evidence for the second hit hypothesis in 22q11.2DS and psychiatric phenotypes.

Li and colleagues studied only 15 samples with either 22q11.2 deletion or duplication, and identified two cases with additional CNVs. Both cases were for new-borns that manifested severe congenital heart defects. The additional CNVs identified in the two cases are: 1.56Mb Xp22.31 deletion (involving *STS* gene) carried by a 22q11.2 deletion carrier, 605Kb 15q13.3 duplication (involving *CHRNA7* gene), and 209Kb 16p13.2 deletion (involving *RBFOX1* gene) carried by a 22q11.2 duplication carrier. Only one CNV (15q13.3 duplication) of the three CNVs identified in this study have been previously implicated in neurodevelopmental delay and neuropsychiatric diseases (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013). The other two CNVs are not shown in the list of the previously identified pathogenic CNVs (Li et al. 2012).

In the study of developmental delay conducted by Girirajan and colleagues, that was discussed earlier, a subset of 22q11.2DS individuals (n= 113) was included (Girirajan et al. 2010). Testing specifically these individuals for the presence of a second hit CNV revealed a non-significant trend for additional CNV enrichment in 22q11.2DS individuals. 8% (9/113) of the 22q11.2DS patients had an additional large (>500Kb) CNV, in addition to 22q11.2, compared to only 4% (21/471) in controls (Girirajan et al. 2010).

Bassett and colleagues assessed the CNV content in 100 22q11.2DS adults, 44 of which experienced psychosis (Bassett et al. 2008). The study identified no excess of inherited CNVs or evidence of *de novo* CNVs in 22q11.2DS patients with schizophrenia (p-values =0.6). Also, the authors did not look particularly at pathogenic CNVs that may be second-hit loci (Bassett et al. 2008).

Williams and colleagues investigated the second hit hypothesis in 22q11.2DS adults with (n= 23) and without psychosis symptoms (n= 25). This study showed no significant evidence for increasing rate of second CNV hits in 22q11.2DS adults with psychosis when compared to the rate in non-psychotic adults (Fisher's p-value= 0.37) (Williams et al. 2013). The most interesting finding in this study is that a significant elevated CNVs average size (354Kb) was found in the psychotic 22q11.2DS group in comparison to average size (227Kb) in the non-psychotic 22q11.2DS group (p-value= 0.02). The authors identified four of the additional CNVs spanning genes that had been reported to be associated with neuropsychiatric disorders including 2q13, 3p26.3-p26.2, 7q31.1, and 10q21.1. None of these four CNVs overlapping one of the previously implicated CNVs that were enriched in neuropsychiatric diseases (Williams et al. 2013).

None of these studies have investigated whether the presence of additional large, rare CNVs is related to childhood neuropsychiatric illnesses, ADHD and ASD, in 22q11.2DS. Therefore, this chapter set out to investigate whether the second hit hypothesis is related to the childhood psychiatric diseases associated with 22q11.2DS in a well-characterized cohort of 22q11.2DS children.

6.3. Aims of this chapter

It is now established that large, rare copy number variants are associated with neurodevelopmental delay and neuropsychiatric disorders. It has also been reported that patients who carry additional CNVs with a primary pathogenic lesion can express more severe and/or additional symptoms than the typical syndromic phenotypes. The work described in this chapter will investigate whether the presence of additional large, rare CNVs are related to the increased risk to psychiatric disease seen in children with 22q11.2DS.

The second hit hypothesis in 22q11.2DS will be assessed in a well-characterised cohort of 22q11.2 deletion carriers by conducting a genome-wide analysis of additional large, rare CNVs between the 22q11.2DS patients with and without a psychiatric phenotype (ADHD and/or ASD). The sample will then be assessed for evidence of an increased rate of second hit CNVs in the 22q11.2DS patients with a psychiatric phenotype. As most CNVs are either benign or of unknown consequence, this analyses will also be conducted for the subset of 'pathogenic' CNVs, that intersect one of 88 CNVs that have been previously implicated to increase risk to developmental delay and neuropsychiatric disease.

6.4. Materials and methods

6.4.1. Samples

The full specific details of the samples analysed in this chapter are found in chapter 2 (DNA extraction protocol in section 2.2, and genotyping microarray in section 2.3).

CNV calling was conducted on the genotype data of the 76 well-characterised 22q11.2DS samples that passed the genotype-based QC used in chapter 3 section 3.5.1. However, as it is standard for CNV analysis (The International Schizophrenia Consortium 2008), the quality of the 76 samples were further assessed based on PennCNV QC statistics. Samples were therefore excluded on the basis of either a high SD in their genome-wide log R ratio (>0.36) and B allele frequency (>0.025), a genome-wide waviness factor out of -0.03 to 0.03 range, or because they carried more than 60 apparent CNVs larger than 100Kb. A summary of this additional quality control is described in Figure 6- 3.

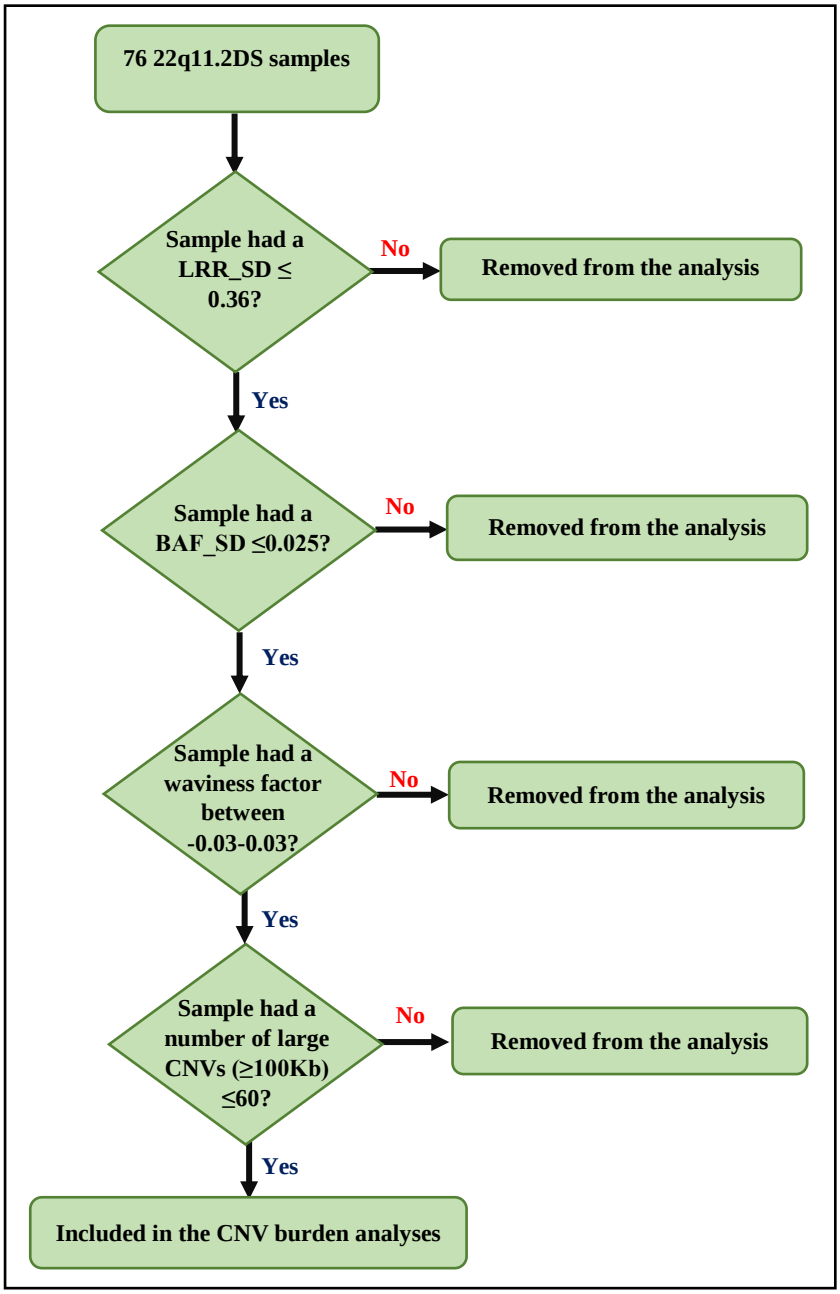


Figure 6- 3: **Flowchart for sample quality control based on PennCNV QC statistics.**

6.4.2. CNV calls for burden analyses

6.4.2.1. CNV calling and quality control

CNV analysis was limited to 523,228 autosomal non-imputed SNPs genotyped by the HumanCoreExome chip. CNVs spanning a minimum of 10 consecutive informative SNPs were defined by PennCNV (2009Aug27version) (K. Wang et al. 2007) using

the protocol described in chapter 2 section 2.5.1; with those copy number calls lower than two being classed as deletions, and greater than two as duplications. Rigorous quality control procedures were applied to all CNV calls in order to maximise the chances that our subsequent analysis included only true CNVs. Detailed CNVs quality control criteria were fully explained in chapter 2 section 2.5.2.

In addition, CNVs were excluded if more than 50% of their length spanned a known segmental duplications present in the February, 2009 human reference sequence at the National Centre for Biotechnology Information reference build 37, hg19 (Genome.ucsc.edu, n.d.). CNVs were excluded if they overlapped by >50% with one of the CNVs in a list of common CNVs (>1% in a large population control cohort of 11,255) that had been independently identified and provided by Dr Elliott Rees. It has been previously reported that there is a significant excess of large CNVs in ID and MCA cases relative to controls. This enrichment was evident at 250Kb and becomes more obvious with larger CNV size (Cooper et al. 2012). This study therefore focused on large CNVs ≥200Kb. Finally, as this study aimed to investigate additional CNVs, only CNVs occur outside the 3Mb region deleted in 22q11.2DS were selected. Summary of CNV QC is illustrated in Figure 6- 4.

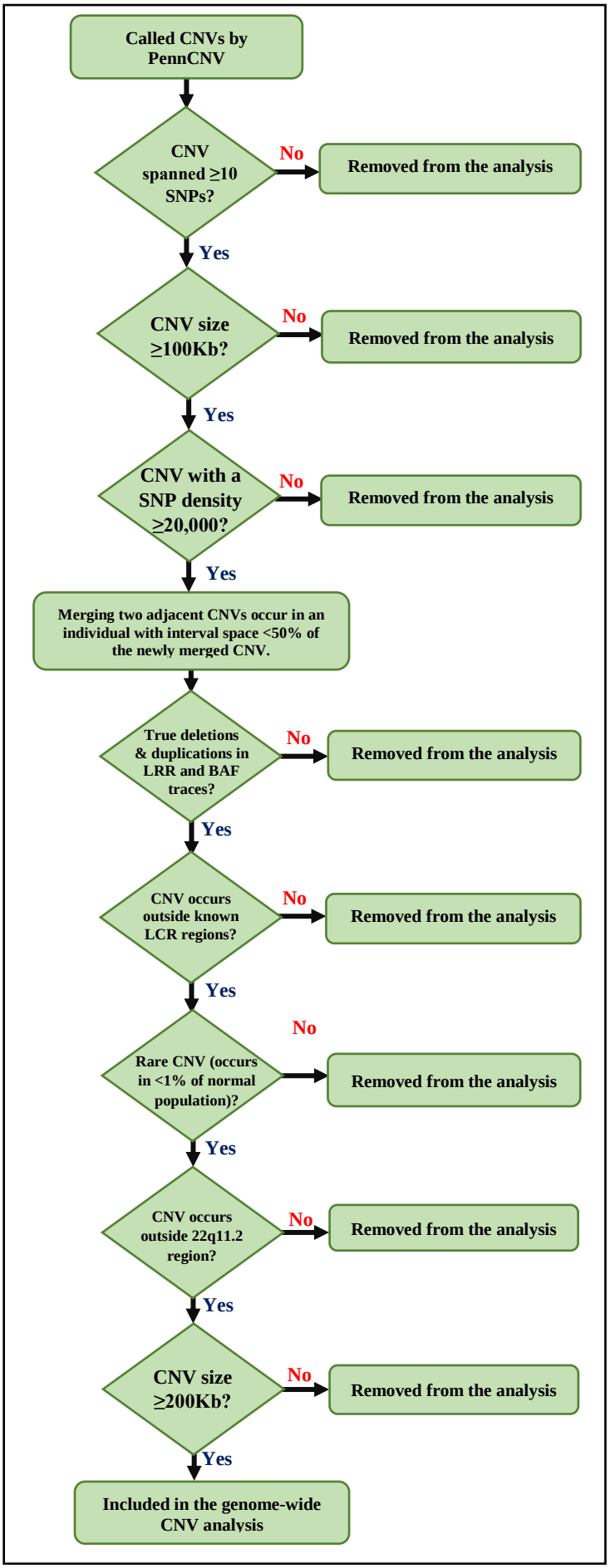


Figure 6- 4: **Flowchart for CNV quality control.**

6.4.2.2. 'Pathogenic' CNV loci

'Pathogenic' CNVs used in the analyses in this chapter were taken from published genome-wide CNV association studies using large case-control cohorts of neuropsychiatric diseases. These CNVs were identified to be significantly enriched in a large cohort of cases relative to a large cohort of population controls. These CNVs included:

1) 15 schizophrenia-associated CNVs studied by (Rees et al. 2014) (Table 6- 5).
2) 15 ADHD-associated CNVs reported by (Jarick et al. 2014; Williams et al. 2012; Elia et al. 2011; Williams et al. 2010) (Table 6- 6).
3) 72 potentially pathogenic CNVs implicated in neurodevelopmental diseases and genomic disorders reported by (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013) (Table 6- 7, Table 6- 8).

As 14 CNVs were reported repetitively in the above three lists, therefore collectively we had 88 'pathogenic' CNV loci used to identify candidate pathogenic CNVs in our cohort.

Table 6- 5: **Previously implicated copy number variation loci in schizophrenia** (Rees et al. 2014).

CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type
1	146,570,001	147,390,000	1q21.1 deletion	1	146,570,001	147,390,000	1q21.1 duplication
2	50,150,000	51,260,000	*NRXN1* deletion	7	72,740,000	74,140,000	WBS duplication
3	195,730,000	197,340,000	3q29 deletion	7	158,820,000	158,940,000	*VIPR2* duplication
15	22,800,000	230,90,000	15q11.2 deletion	15	24,820,000	28,430,000	AS/PWS duplication
15	31,130,000	32,480,000	15q13.3 deletion	16	15,510,000	16,300,000	16p13.11 duplication
16	28,820,000	29,050,000	16p11.2 distal deletion	16	29,640,000	30,200,000	16p11.2 duplication
17	14,160,000	15,430,000	17p12 deletion				
17	34,810,000	36,200,000	17q12 deletion				
22	19,020,000	20,260,000	22q11.2 deletion				

Table 6- 6: **Previously implicated copy number variation loci in ADHD** (Jarick et al. 2014; Williams et al. 2012; Elia et al. 2011; Williams et al. 2010).

CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type
1	56,280,909	56,291,907	Deletion	1	72,544,704	72,555,807	Duplication
3	1,869,168	1,884,889	Deletion	2	81,565,786	81,592,571	Duplication
3	7,208,953	7,222,236	Deletion	4	113,552,891	113,569,135	Duplication
5	64,992,220	65,010,764	Deletion	6	146,615,383	146,652,354	Duplication
7	126,737,888	126,748,966	Deletion	6	162,739,766	162,847,029	Duplication
11	88,629,801	88,712,013	Deletion	7	153,864,665	153,933,894	Duplication
19	33,735,880	33,752,994	Deletion	15	32,024,690	32,445,689	Duplication
				16	15,248,930	18,267,149	Duplication

Table 6- 7: **Previously implicated copy number variation loci in gnomic disorders, intellectual disability and developmental delay cases** (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013).

CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type
1	10001	10,077,413	1p36 deletion syndrome	16	21,942,499	22,432,499	16p12.1 deletion	2	239,705,243	242,471,327	2q37 deletion
1	146,573,376	147,393,376	1q21.1 deletion	16	28,822,499	29,052,499	16p11.2 (SH2B1) deletion	22	19,020,000	20,290,000	DiGeorge/VCFS deletion
10	81,960,020	88,800,020	10q23 deletion	16	29,652,499	30,202,499	16p11.2 deletion	22	21,910,000	23,670,000	22q11.2 distal deletion
15	22,798,636	23,088,559	15q11.2 deletion	17	1,053,250	2,633,250	17p13.3 deletion	22	43,000,056	51,163,134	Phelan-McDermid syndrome
15	24,818,907	28,426,405	Prader-Willi/Angelman deletion	17	1,203,250	1,323,250	17p13.3(YWHAE) deletion	3	195,745,603	197,355,603	3q29 deletion
15	31,132,708	32,482,708	15q13.3 deletion	17	2,423,250	2,593,250	17p13.3(PAFAH1B1) deletion	4	1,529,198	2,030,202	Wolf-Hirschhorn
15	32,012,708	32,452,708	15q13.3 deletion (chr15:32Mb-32.4Mb)	17	16,789,275	18,299,275	Smith Magenis syndrome	5	175,717,394	177,057,394	Soto syndrome
15	72,952,946	74,362,947	15q24 deletion	17	29,095,874	30,275,887	NF1 deletion syndrome	6	155,000	605,5001	6p25 deletion
15	75,972,945	78,202,945	15q24.2q24.3 deletion	17	34,815,887	36,225,887	17q12 deletion	6	100,813,279	100,943,279	6q16 deletion
15	83,182,945	84,738,996	15q25 deletion	17	43,704,217	44,164,182	17q21.31 deletion	7	72,742,064	74,142,064	William syndrome
16	3,749,999	3,949,999	Rubinstein-Taybi deletion	17	58,285,218	60,305,218	17q23 deletion	8	8,092,590	11,892,591	8p23.1 deletion
16	15,502,499	16,292,499	16p13.11 deletion	19	13,079,000	16,699,000	19p13.12 deletion	9	137,810,179	141,080,179	9q34 deletion
16	21,532,499	29,102,499	16p11.2p12.1 deletion	2	148,723,530	149,293,530	2q23.1 deletion	X	102,413,344	113,413,741	PLP1 deletion

Table 6- 8: **(Continue) Previously implicated copy number variation loci in gnomic disorders, intellectual disability and developmental delay cases** (Girirajan et al. 2012; Cooper et al. 2012; Malhotra & Sebat 2012; Kaminsky et al. 2013).

CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type	CHR	Start (hg19)	End (hg19)	CNV type
1	146,573,376	147,393,376	1q21.1 duplication	16	28,822,499	29,052,499	16p11.2(SH2B1) duplication	22	21,910,000	23,670,000	22q11.2 distal duplication
10	81,960,020	88,800,020	10q23 duplication	16	29,652,499	30,202,499	16p11.2 duplication	22	43,000,056	51,163,134	22q13 duplication
15	24,818,907	28,426,405	Prader-Willi/Angelman duplication	17	1,203,250	1,323,250	17p13.3(YWHAE) duplication	3	195,745,603	197,355,603	3q29 duplication
15	31,132,708	32,482,708	15q13.3 duplication	17	2,423,250	2,593,250	17p13.3(PAFAH1B1) duplication	4	1,529,198	2,030,202	WHS duplication
15	32,012,708	32,452,708	15q13.3 duplication (chr15:32Mb-32.4Mb)	17	16,789,275	18,299,275	Potocki Lupski syndrome	5	175,717,394	177,057,394	5q35 duplication
15	72,952,946	74,362,947	15q24 duplication	17	29,095,874	30,275,887	NF1 duplication	6	155,000	6,055,001	6p25 duplication
15	75,972,945	78,202,945	15q24.2q24.3 duplication	17	34,815,887	36,225,887	17q12 duplication	6	100,813,279	100,943,279	6q16 duplication
15	83,182,945	84,738,996	15q25 duplication	17	43,704,217	44,164,182	17q21.31 duplication	7	72,742,064	74,142,064	WBS duplication
16	15,502,499	16,292,499	16p13.11 deletion	17	58,285,218	60,305,218	17q23 duplication	8	8,092,590	11,892,591	8p23.1 duplication
16	21,532,499	29,102,499	16p11.2p12.1 duplication	2	239,705,243	242,471,327	2q37 duplication	9	137,810,179	141,080,179	9q34 duplication
16	21,942,499	22,432,499	16p12.1 duplication	22	19,020,000	20,290,000	22q11.2 duplication	X	102,413,344	113,413,741	PLP1 duplication

6.4.3. CNV burden analyses

6.4.3.1. Genome-wide CNV burden analysis

CNV calls were stratified according to their sizes into two categories (≥200Kb CNVs and ≥500Kb). To conduct the global burden analysis, the rate of additional CNVs was determined in each of the 3 categories of psychiatric diseases (22q11.2DS+ADHD, 22q11.2DS+ASD, and 22q11.2DS+PSYCH) and compared to the rate seen in the 22q11.2DS patients who had no psychiatric disease (22q11.2DS-PSYCH). This one-sided test of CNV burden analysis performed using Plink (Purcell et al. 2007) with significance being empirically assessed by 10,000 permutations.

6.4.3.2. 'Pathogenic' CNV burden analysis

The genomic coordinates were determined for each of the CNVs identified in the 22q11.2DS patients and also for the 88 CNVs that had been previously implicated in neurodevelopmental delay and neuropsychiatric disease (described in this chapter section 6.4.2.2). CNVs from the 22q11.2DS cohort were selected if they overlapped with any of the 88 'pathogenic' CNV by >50%. A burden analysis was then performed on this subset of CNVs in the same way that was described for all CNVs.

6.5. Results

6.5.1. Sample quality control

Three of the initial 76 samples did not pass our stringent quality control procedures. They were identified to carry more than 60 large CNVs (≥100Kb); also, the three samples were identified to have a WF value > 0.03. Therefore, they were removed from the analysis. The remaining 73 samples with 22q11.2DS were used for all CNV analyses described in this chapter and based on the presence or absence of psychiatric phenotypes were categorized into 3 groups: 22q11.2DS+ADHD, 22q11.2DS+ASD, and 22q11.2DS+PSYCH. Numbers of patients in each group are given in (Table 6- 9).

Table 6-9: **Number of 22q11.2 deletion carriers affected and unaffected by psychiatric disorders.**

Categories	Description	Number of individuals
22q11.2DS+ADHD	22q11.2 deletion carriers with ADHD phenotypes	28
22q11.2DS+ASD	22q11.2 deletion carriers with ASD phenotypes	18
22q11.2DS+PSYCH	22q11.2 deletion carriers with ADHD and/or ASD phenotypes	36
22q11.2DS-PSYCH	22q11.2 deletion carriers with no ADHD and ASD phenotypes	33

6.5.2. CNV detection

PennCNV analysis of the 73 individuals with 22q11.2DS identified a total of 18 large (≥200kb), rare (<1% population frequency) CNVs that survived the stringent QC criteria. Of the 18 CNVs, 5 were deletions (27.8%) and 13 were duplications (86.7%) mapped to 18 distinct genomic regions. Which means that none of the 18 CNVs were detected in more than one sample.

Of the 18 CNVs ≥200Kb, 7 CNVs are larger than 500Kb (1 deletion and 6 duplications) (Table 6- 10).

The 18 CNVs are carried by 14 22q11.2DS patients, 3 patients were found to carry multiple additional CNVs, while the remaining 11 patients carry a single additional CNV:

1) 1 patient has 3 second CNVs.
2) 2 patients have 2 second CNVs each.
3) 11 patients have a single second CNV each.

Of the 18 CNVs, 11 were found in the 22q11.2DS-PSYCH patients (n= 8). Two patients carried multiple CNVs in addition to 22q11.2 deletion (one patient has 3 CNVs and the other has 2 CNVs). While each of the remaining 6 patients has a single CNV (Table 6- 11).

Seven CNVs were found in the 22q11.2DS+PSYCH group where only 1 patient had 2 CNVs and 5 had a single additional CNV. None of the 7 CNVs were found in multiple samples (Table 6- 11).

Six CNVs were carried by five 22q11.2DS+ADHD patients, with only 1 patient having 2 CNVs and 4 patients having 1 additional CNV each. No single CNV was identified in more than one ADHD patient (Table 6- 11).

Four CNVs were identified in 4 individuals who had 22q11.2DS+ASD. Each patient carried a single additional CNV. Also there is no single additional CNV was detected in multiple ASD patients (Table 6- 11).

Investigating the intersection of these 18 CNVs with 88 ‘pathogenic’ CNVs, that had been previously implicated to neuropsychiatric disease, identified 3 that overlapped with a ‘pathogenic’ CNV. The 3 CNVs were carried by 3 22q11.2DS patients, one of which showed symptoms of both ADHD and ASD, while the other 2 had no psychiatric phenotype (Table 6- 12).

Table 6- 10: **Additional large, rare CNVs identified in 22q11.2DS cohort.**

CHR	Start (hg19)	End (hg19)	CNV type	Size (Kb)	Genomic loci	Sample ID	ADHD	ASD	PSYCH
1	248,072,026	248,688,317	Duplication	616.3	1	E_006_1_1			-
2	136,648,077	137,093,578	Duplication	445.5	3	E_015_1_1			-
2	138,600,801	138,865,871	Duplication	265.1	2	E_015_1_1			-
3	234,877	1,238,287	Deletion	1003.4	5	E_089_1_1			-
3	6,728,385	7,035,024	Duplication	306.6	4	E_116_1_1			-
3	141,845,000	142,085,310	Duplication	240.3	6	E_033_1_1			-
4	86,089	437,081	Deletion	351.0	7	E_116_1_1			-
9	11,578,233	11,890,045	Deletion	311.8	8	E_072_1_1		+	+
9	87,414,794	88,373,138	Duplication	958.3	9	E_064_1_1	+		+
9	116,958,266	117,360,729	Duplication	402.5	10	E_001_1_1	+	+	+
10	76,071,203	76,295,789	Deletion	224.6	11	E_077_1_1			-
12	37,927,114	38,891,067	Duplication	964.0	12	E_084_1_1	+		+
15	32,067,075	32,514,341	Duplication	447.3	13	E_034_1_1			-
16	7,048,405	7,356,303	Deletion	307.9	14	E_054_1_1			-
16	29,644,174	30,199,713	Duplication	555.5	15	E_070_1_1	+	+	+
20	52,288,724	53,267,627	Duplication	978.9	16	E_012_1_1	+	+	+
22	22,312,879	22,560,977	Duplication	248.1	17	E_015_1_1			-
22	23,700,907	25,010,874	Duplication	1310.0	18	E_064_1_1	+		+

CNVs highlighted by blue were carried by a single 22q11.2DS patient (E_015_1_1).
CNVs highlighted by grey were carried by a single 22q11.2DS patient (E_116_1_1).
CNVs highlighted by yellow were carried by a single 22q11.2DS patient (E_064_1_1).
(+) Affected with psychiatric disorder.
(-) Unaffected with psychiatric disorder.

Table 6- 11: **Number of patients carry second CNVs in psychiatric disease categories.**

22q11.2DS			22q11.2DS+ADHD			22q11.2DS+ASD			22q11.2DS+PSYCH			22q11.2DS-PSYCH		
*Number of CNVs	Number of patients carry second CNVs		Number of CNVs	Number of patients carry second CNVs		Number of CNVs	Number of patients carry second CNVs		Number of CNVs	Number of patients carry second CNVs		Number of CNVs	Number of patients carry second CNVs	
18	14		6	5		4	4		7	6		11	8	
	1 second CNV	≥2 second CNVs		1 second CNV	≥2 second CNVs		1 second CNV	≥2 second CNVs		1 second CNV	≥2 second CNVs		1 second CNV	≥2 second CNVs
	11	3		4	1		4	0		5	1		6	2

* Total number of second CNVs detected in patients in the psychiatric category.

Table 6- 12: **Second CNVs intersected one of the 88 previously reported implicated CNVs in neurodevelopmental and neuropsychiatric disorders.**

Called CNVs					Previously implicated CNVs				22q11.2DS carriers	
CHR	Start (hg19)	End (hg19)	CNV type	Size (Kb)	CHR	Start (hg19)	End (hg19)	CNV type	Samples ID	Psychiatric phenotype
16	29,644,174	30,199,713	Dup	555.54	16	29,652,499	30,202,499	16p11.2 duplication	E_034_1_1	22q11.2DS+PSYCH (ADHD and ASD)
15	32,067,075	32,514,341	Dup	447.27	15	32,012,708	32,452,708	15q13.3 duplication (chr15:32Mb-32.4Mb)	E_070_1_1	22q11.2DS-PSYCH
22	22,312,879	22,560,977	Dup	248.10	22	21,910,000	23,670,000	22q11.2 distal duplication	E_015_1_1	22q11.2DS-PSYCH

6.5.3. Investigating whether 22q11.2DS patients with ADHD or ASD are enriched for second hit CNVs

This analysis sought to assess whether additional large, rare CNVs could potentially have an impact on the increased risk of childhood psychiatric disease in 22q11.2DS children.

6.5.3.1. ADHD in 22q11.2DS

The 6 distinct CNVs detected in the 22q11.2DS+ADHD patients were used for burden analyses. Results of global CNV burden analysis revealed a 3-fold enrichment of CNVs ≥500Kb in the 22q11.2DS+ADHD patients (rate =0.18) compared to 22q11.2DS-PSYCH patients (rate =0.06) (Table 6- 13), however, this enrichment was not statistically significant (p-value= 0.20). A less pronounced difference in the rate of ≥200Kb CNVs between the two groups was observed (enrichment rate= 0.63; p-value= 0.83) (Table 6- 13).

The analysis was then restricted to the CNVs that overlapped with one of the 88 previously implicated CNVs with developmental delay or psychiatric diseases. Of the 6 additional CNVs identified in ADHD patients, only a single CNV was 'pathogenic' and included in the burden analysis. This revealed a 0.6-fold enrichment for 'pathogenic' CNVs in 22q11.2DS+ADHD patients (rate = 0.036) relative to those with no psychiatric symptoms (rate = 0.060), however, this finding was not statistically significant (p-value= 0.85) (Table 6- 13).

Table 6- 13: **Results of CNV burden analyses comparing 22q11.2DS+ADHD individuals (n= 28) and 22q11.2DS-PSYCH individuals (n= 33).**

Global CNVs								
	CNVs ≥ 200Kb				**CNVs ≥ 500Kb**			
	22q11.2DS +ADHD	**22q11.2DS -PSYCH**	***Ratio**	****EMP p-value**	**22q11.2DS +ADHD**	**22q11.2DS -PSYCH**	**Ratio**	**EMP p-value**
Number of CNVs	6	11	-	-	5	2	-	-
Total number of samples	28	33	-	-	28	33	-	-
***Rate**	0.21	0.33	0.63	0.83	0.18	0.06	3	0.20
****Proportion**	0.18	0.24	0.75	0.82	0.14	0.06	2.3	0.26
Total CNVs	17				7			

Pathogenic CNVs				
	22q11.2DS + ADHD	**22q11.2DS - PSYCH**	**Ratio**	**EMP P-value**
Number of CNVs	1	2	-	-
Pathogenic loci	16p11.2 duplication	15q13.3 duplication (chr15:32Mb-32.4Mb) 22q11.2 distal duplication	-	-
Total number of samples	28	33	-	-
Rate	0.036	0.060	0.6	0.85
Proportion	0.036	0.060	0.6	0.85
Total CNVs	3			

*Ratio: 22q11.2DS affected patients/22q11.2DS unaffected patients.
**EMP p-value: Empirical p-value, corrected for all tests.
***Rate: Number of CNVs (Number of CNVs/ Total number of samples).
****Proportion: Proportion of sample with one or more CNVs (Number of samples carry ≥1 CNVs/Total number of samples).

6.5.3.2. ASD in 22q11.2DS

The 4 distinct additional CNVs detected in the 22q11.2DS+ASD patients were tested for burden. Genome-wide CNV burden analysis revealed a 1.8-fold enrichment of CNVs ≥500Kb in the 22q11.2DS+ASD patients (rate =0.11) compared to 22q11.2DS-PSYCH patients (rate =0.06) (Table 6- 14), however, this enrichment was not statistically significant (p-value= 0.44). Similarly, the rate of ≥200Kb CNVs was not increased in the 22q11.2DS+ASD patients (rate =0.22) relative to the 22q11.2DS-PSYCH patients (rate =0.33), which again was not significant (p-value= 0.79) (Table 6- 14).

Restricting the analysis to the CNVs that overlapped with one of the 88 previously implicated CNVs with developmental delay or psychiatric disease identified only a single CNV of the 4 additional CNVs that was potentially 'pathogenic'. The rate of 'pathogenic' CNVs in 22q11.2DS+ASD patients was 0.056 compared to 0.060 in those with no psychiatric symptoms and this was not statistically significant (p-value= 0.73) (Table 6- 14).

Table 6- 14: **Results of CNV burden analyses comparing 22q11.2DS+ ASD individuals (n= 18) and 22q11.2DS-PSYCH individuals (n= 33).**

Global CNVs								
	CNVs ≥ 200Kb				CNVs ≥ 500Kb			
	22q11.2DS +ASD	22q11.2DS -PSYCH	*Ratio	**EMP p-value	22q11.2DS +ASD	22q11.2DS -PSYCH	Ratio	EMP p-value
Number of CNVs	4	11	-	-	2	2	-	-
Number of samples	18	33	-	-	18	33	-	-
***Rate**	0.22	0.33	0.67	0.79	0.11	0.06	1.8	0.44
****Proportion**	0.22	0.24	0.92	0.68	0.11	0.06	1.8	0.44
Total CNVs	15				4			

Pathogenic CNVs				
	22q11.2DS + ASD	22q11.2DS - PSYCH	Ratio	EMP P-value
Number of CNVs	1	2	-	-
Pathogenic loci	16p11.2 duplication	15q13.3 duplication (chr15:32Mb-32.4Mb) 22q11.2 distal duplication	-	-
Number of samples	18	33	-	-
Rate	0.056	0.060	0.93	0.73
Proportion	0.056	0.060	0.93	0.73
Total CNVs	3			

*Ratio: 22q11.2DS affected patients/22q11.2DS unaffected patients.
**EMP p-value: Empirical p-value, corrected for all tests.
***Rate: Number of CNVs (Number of CNVs/ Number of samples).
****Proportion: Proportion of sample with one or more CNVs (Number of samples carry ≥1 CNVs/Total number of samples).

6.5.4. Investigating whether 22q11.2DS patients with psychiatric phenotypes are enriched for second hit CNVs

In an attempt to increase the power of the study a final analysis was conducted which collectively looked at the rate of additional large, rare CNVs carried by 22q11.2 deletion carriers who had ADHD, ASD, or both.

The 7 distinct additional CNVs identified in the 22q11.2DS+PSYCH patients were included in the burden analyses. Results of global CNV burden analysis revealed a 2.29-fold enrichment of CNVs ≥500Kb in the 22q11.2DS+PSYCH samples (rate = 0.14) relative to those with no psychiatric symptoms (rate = 0.06) (Table 6- 15), however, this excess was not statistically significant (p-value= 0.29). The difference in the frequency of CNVs ≥200Kb between the two groups was less pronounced with the observed rate in the 22q11.2DS+PSYCH samples being 0.19 and 0.33 in those with no psychiatric symptoms (p-value= 0.88) (Table 6- 15).

The analysis was then limited to the CNVs that overlapped with one of the 88 CNVs previously reported to be associated with developmental delay or psychiatric diseases. A single CNV intersected one of the reported 'pathogenic' CNVs which was not significantly enriched in the 22q11.2DS+PSYCH patients (rate = 0.028) compared to those with no psychiatric symptoms (rate=0.060; p-value= 0.89) (Table 6- 15).

Table 6- 15: **Results of CNV burden analyses comparing 22q11.2DS+PSYCH individuals (n= 36) and 22q11.2DS-PSYCH individuals (n= 33).**

Global CNVs								
	CNVs ≥ 200Kb				**CNVs ≥ 500Kb**			
	22q11.2DS +PSYCH	**22q11.2DS - PSYCH**	***Ratio**	****EMP p-value**	**22q11.2DS +PSYCH**	**22q11.2DS - PSYCH**	**Ratio**	**EMP p-value**
Number of CNVs	7	11	-	-	5	2	-	-
Number of samples	36	33	-	-	36	33	-	-
***Rate**	0.19	0.33	0.58	0.88	0.14	0.06	2.29	0.29
****Proportion**	0.17	0.24	0.69	0.86	0.11	0.06	1.83	0.37
Total CNVs	18				4			

Pathogenic CNVs				
	22q11.2DS +PSYCH	**22q11.2DS -PSYCH**	**Ratio**	**EMP P-value**
Number of CNVs	1	2	-	-
Pathogenic loci	16p11.2 duplication	15q13.3 duplication (chr15:32Mb-32.4Mb) 22q11.2 distal duplication	-	-
Number of samples	36	33	-	-
Rate	0.028	0.060	0.47	0.89
Proportion	0.028	0.060	0.47	0.89
Total CNVs	3			

*Ratio: 22q11.2DS affected patients/22q11.2DS unaffected patients.
**EMP p-value: Empirical p-value, corrected for all tests.
***Rate: Number of CNVs (Number of CNVs/ Number of samples).
****Proportion: Proportion of sample with one or more CNVs (Number of samples carry ≥1 CNVs/Total number of samples).

6.6. Discussion

Recent studies have shown that a secondary CNV in patients carrying a primary pathogenic CNV can modify the phenotype in neurodevelopmental delay and other multiple syndromic disorders (Girirajan et al. 2010; Girirajan et al. 2012). The main aim of this chapter was to investigate whether the increased risk of psychiatric phenotypes in 22q11.2DS is conferred by additional genomic structural variants that manifest independently to the primary microdeletion.

Thus, CNV burden analyses were performed to test if there is additional CNVs being enriched in 22q11.2DS patients affected by neuropsychiatric diseases. Results of global CNVs burden analyses revealed a potential trend for enrichment of CNVs ≥500Kb in 22q11.2DS individuals with ADHD (by 3-fold), ASD (by 1.8-fold) and psychiatric illnesses (by 2.29-fold). However, none of these observations were statistically significant (p-values >0.05). By restricting the burden analysis on the previously reported 'pathogenic' CNVs, the study also failed to provide an evidence for a significant enrichment for 'pathogenic' CNVs in 22q11.2DS patients who showed neuropsychiatric diseases symptoms (p-values >0.05).

This potential enrichment observed in 22q11.2DS patients with psychiatric diseases is mainly due to five affected patients harbouring a single CNV ≥500Kb each, compared to only two unaffected individuals carrying one CNVs ≥500Kb each. Only one of the five affected patients carry a second CNV that overlap with one of the 88 reported pathogenic CNV. This individual has a ~555Kb duplication at 16p11.2 (chr16: 29,652,499-30,202,499 on hg19) and showed symptoms of ADHD and ASD.

Recurrent structural variations at 16p11.2 region (chr16: 29,652,499-30,202,499 on hg19) have been reported in a large genome-wide association study including 751

families and 512 children with autism (Weiss et al. 2008). An excess of 593Kb duplications spanning this region found in three autistic families, inherited in two families (6 cases and 1 control) and occur as *de novo* events in the third family (1 case and 1 control). The study provided significant evidence for an enrichment of 16p11.2 duplications in autistic subjects in the families compared to control subjects (p-value= $1.1x10^{-4}$ for both 16p11.2 deletions and duplications) (Weiss et al. 2008). In addition, presence of 16p11.2 duplications was also observed in four independent samples from 512 autistic children compared to none in 434 control children, an observation that was statistically significantly (p-value= 0.007 for both 16p11.2 deletions and duplications) (Weiss et al. 2008). Subsequent studies, reviewed by Malhotra and Sebat (Malhotra & Sebat 2012), have also reported significant evidence for involvement of duplications at 16p11.2 in ID/DD/Congenital malformations (p-value= $2.2x10^{-8}$; OR= 3.4) with a frequency of ~0.2%, in ASD (p-value= $6.2x10^{-11}$; OR= 11.8) with a frequency of ~1% in ASD, in schizophrenia (p-value= $3.2x10^{-14}$; OR= 9.4) with a frequency of ~0.3%, and in bipolar disease/recurrent depression (p-value= 0.0008; OR= 3.9). Additionally, in a large case-control cohort of unexplained ID, DD, dysmorphic features, multiple congenital anomalies, ASD, or clinical features suggestive of a chromosomal syndrome; a significant burden of 16p11.2 duplication was identified (1 in 404 in patients) relative to normal controls (p-value= $2.5x10^{-5}$; OR= 6.28) (Kaminsky et al. 2013). This duplication has been reported to be highly significantly associated with ID and MCA (p-value= 0.0004 and $1.0x10^{-6}$) (Cooper et al. 2012; Girirajan et al. 2012), in addition to the other significant evidence of this duplication being associated with to other multiple psychiatric diseases (Malhotra & Sebat 2012). Moreover, in a very recent genome-wide chromosomal microarray study conducted in 202 adults with idiopathic ID, an increased frequency of 16p11.2

duplication was observed which suggested a potential enrichment of this recurrent CNV in the adult population of ID and comorbid psychiatric disorder (Wolfe et al. 2017).

It is however of potential interest that when looking at the secondary CNVs ≥100Kb, a 22q11.2DS patient carries a ~124Kb deletion that spans the 16p11.2p12.1 pathogenic region (chr16: 21,532,499-29,102,499 on hg19) and has symptoms of only ADHD. Recurrent *de novo* deletions at 16p11.2p12.1 region (chr16: 21,532,499-29,102,499 on hg19) have been reported in five cases with developmental disabilities, distinct facial features, and multiple congenital abnormalities (Hernando et al. 2002; Ballif et al. 2007). Four of the five cases were identified among 8,789 patients with developmental disabilities who were screened for analysis with a targeted microarray. The precise breakpoints for the deletions in these individuals were determined. The proximal breakpoint is shared in the four subjects, while distal breakpoints were ~28.5Mb (subject 2), ~29.3 Mb (subjects 3 and 4), and ~30.1 Mb (subject 1) from the 16p telomere, resulting in overall deletion size ranges from ~7.1Mb, ~7.9Mb, and ~8.7Mb (Ballif et al. 2007) (Figure 6- 5). Although the size of the deletion identified in this study is smaller (~124Kb) compared to the deletion identified in the four reported individuals (~7.1Mb- ~8.7Mb), it does overlap with the proximal region of the reported deletions and shares the same proximal breakpoint (Figure 6- 5). Deletions at 16p11.2-12.1 have been nominally associated with ID/DD in one study (p-value= 0.011) (Girirajan et al. 2012), but this was not supported in another study (p-value= 0.43) (Cooper et al. 2012). The CNV spans 4 genes; *METT9, IGSF9, U6,* and *OTOA* (Figure 6-5). CNVs spanning *OTOA* genes have been reported to be the most common CNVs detected in 686 patients with non-syndromic hearing loss (13% of CNVs identified) (Shearer et al. 2014).

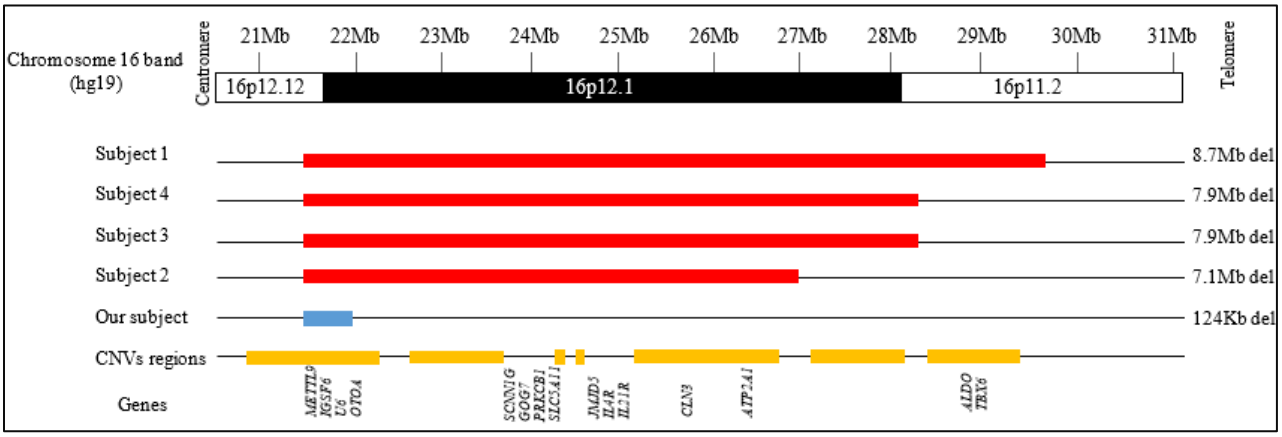


Figure 6- 5: **Schematic of the 16p11.2-p12.2 region (hg19) with a summary of the deletions identified in our case and the four reported subjects in** (Ballif et al. 2007).

Red bars indicate deleted regions for each reported subject of the four. Blue bar indicates deleted region for our subject. Orange bars indicate regions of copy number variation based on the Database of Genomic Variants. The locations of select genes from 4100 known genes in the region are shown. Figure obtained from (Ballif et al. 2007) with modification.

Comparing our findings to the previous studies, only one CNV (15q13.3 duplication) identified by Li and colleagues was also detected in one of our 22q11.2 deletion carriers, who has no symptoms of psychiatric phenotypes (Li et al. 2012), while none of the potentially pathogenic CNVs identified by Williams and colleagues overlapped with our detected CNVs (Williams et al. 2013).

Williams and colleagues found a significant increased CNVs average size (354Kb) in psychotic 22q11.2DS patients in comparison to average size (227Kb) in non-psychotic 22q11.2DS patients (p-value= 0.02) (Williams et al. 2013). This observation is consistent with the finding in our study that a trend of higher rate of larger CNVs (≥500Kb) was found in 22q11.2DS with additional psychiatric phenotypes risk compared to those without psychiatric phenotypes by 2.29-fold (p-value= 0.29), while rate of smaller CNVs (≥200Kb) has less pronounced difference (0.58-fold) between 22q11.2DS with and without psychiatric phenotype (p-value= 0.88). This possibly indicates larger CNVs are more likely to be seen in 22q11.2DS individuals with

psychiatric diseases. This is also in line with the finding of other large studies which reported that enrichment for additional CNVs in patients with ID and MCA was more evident in larger CNVs (≥250Kb) than smaller CNVs (Cooper et al. 2012).

Due to the small sample size used in this study and because no pathogenic loci were hit multiple times, statistical analyses failed to show a significant evidence for CNVs enrichment in 22q11.2DS affected with psychiatric diseases. Identification of recurrent CNVs in further datasets including larger sample size is required for proof of disease causality.

The Genetic Power Calculator (Purcell, Cherny & Sham 2003) was used to calculate the power of this study. Our sample size of 36 22q11.2DS patients affected with psychiatric disease and 33 unaffected patients has an 80% power to detect a variant with an allele frequency of 0.06 and an effect size of 2.08 at a p-value of 0.05. This indicates that our sample size is able to detect a single locus with large effect size.

The pwr package version 1.2-0 (Champely 2016), Power analysis functions in R, and G*Power v3.0.10 were used to calculate the power of CNV burden study (Rucker et al. 2016; Tansey et al. 2014), that compares proportions of large, rare CNVs between 22q11.2DS patients with (n= 36) and without (n= 33) psychiatric phenotypes. The study had an 80% power to detect a significant finding with an effect size of 0.68 at the alpha level of 0.05. That means by using this sample size, our study can detect an enrichment for CNVs that are common with an effect size of 0.68 and larger sample size is required to detect significant CNV burden for rare CNVs.

Also, in this study we could not assess the inheritance of most of our rare CNVs, this is due to no available samples from the parents of the patients included in the study. In addition, although we focused on CNVs that large enough to be detected at high sensitivity and specificity irrespective of SNP array, however, our results should be considered preliminary as CNVs were only determined using statistical algorithms and no validation via more precise methods, such as quantitative real time PCR, was performed.

In conclusion, the study in this chapter investigated whether the high risk of neuropsychiatric phenotypes in 22q11.2DS is conferred by presence of additional pathogenic CNVs. Despite 2.29-fold enrichment for additional large CNVs was observed in 22q11.2DS patients with a psychiatric phenotype compared to those with no psychiatric phenotype, the results in this chapter failed to provide significant evidence to support a role of additional second CNVs influencing the psychiatric phenotype of 22q11.2DS.

Chapter 7: Investigating 22q11.2 Deletions in Parkinson's Disease Patients

7.1. Summary

It is widely known that 22q11.2 deletions are associated with a diverse spectrum of multi systemic phenotypes. Recently idiopathic Parkinson's disease cases have been reported within 22q11.2 deletion cohorts (Krahn et al. 1998; Booij et al. 2010; Zaleski et al. 2009; Butcher et al. 2013). This finding suggests that 22q11.2 microdeletions increase the risk of PD. The study in this chapter aims to investigate the association of 22q11.2 deletions with PD by analysing deletions spanning chromosome 22q11.2 in an idiopathic PD cohort (n= 6,462) and healthy controls (n= 6,202) and comparing the rate of 22q11.2 deletions between the two groups.

Five independent PD cases (0.08%) were identified that carried a 3Mb deletion at 22q11.2, while none of the control cohort (0.0%) carried a deletion. Association analysis revealed that this enrichment was short of nominal significance (p-value = 0.069). Analysis of the 5,599 PD cases that had data available for Parkinson's disease age at onset revealed a significant enrichment for 22q11.2 deletions in the 660 EOPD cases (0.61%) compared to 4,939 LOPD cases (0.02%) (p-value= $9.0x10^{-4}$). Comparing the rate of 22q11.2 deletions between EOPD cases (0.61%) and unaffected controls (0%) showed highly significant evidence for an increased rate of 22q11.2 deletion in EOPD cases (p-value= $9.0x10^{-7}$). These findings suggest that patients with a 22q11.2 deletion have an increased risk to develop PD, particularly the early age at onset form of the disease. The analysis has been included in a meta-analysis of

multiple large PD cohorts that fully supported the findings of this chapter, which was conducted by the International Parkinson's Disease Genetics consortium (Mok et al. 2016).

7.2. Introduction

7.2.1. Parkinson's disease

Parkinson's disease (PD) is a neurodegenerative disorder of the central nervous system that specifically causes a defect in the motor system (Jankovic 2008). Abnormality in the motor system in PD patients is due to extensive degradation of dopaminergic neurons in the substantia nigra within the midbrain area in the brain (Gaig & Tolosa 2009). The substantia nigra has a central role in brain functions mainly in motor planning, eye movement, learning, and addiction (Gaig & Tolosa 2009). The reason for the death of dopamine-generating cells in the substantia nigra in PD patients is not fully understood (de Lau & Breteler 2006).

7.2.1.1. Prevalence of PD

PD is the second most common neurodegenerative disease after Alzheimer's disease. Prevalence of PD varies among people based on their age, sex, and ethnic background (de Lau & Breteler 2006) . The population prevalence is estimated at 0.3%, but increases with age up to 1–2% in people older than 60 years, and 3–4% in those older than 80 years (Nussbaum & Ellis 2003). The median age of PD onset is around 60 years (Lees et al. 2009). However, 5-10% of PD cases have an early age of onset PD which occurs between ages of 20-50 years old. Moreover, while rare, juvenile-onset forms of the disease show the symptoms before the age of 20 years (Samii et al. 2004). Regarding to ethnic background, PD appears to be more prevalent in Caucasians (Figure 7- 1) (de Lau & Breteler 2006), however, the rate among different ethnic backgrounds have been shown to be inconsistent. This is attributed to differences in

response rates, survival, and case-ascertainment in each epidemiological study (de Lau & Breteler 2006).

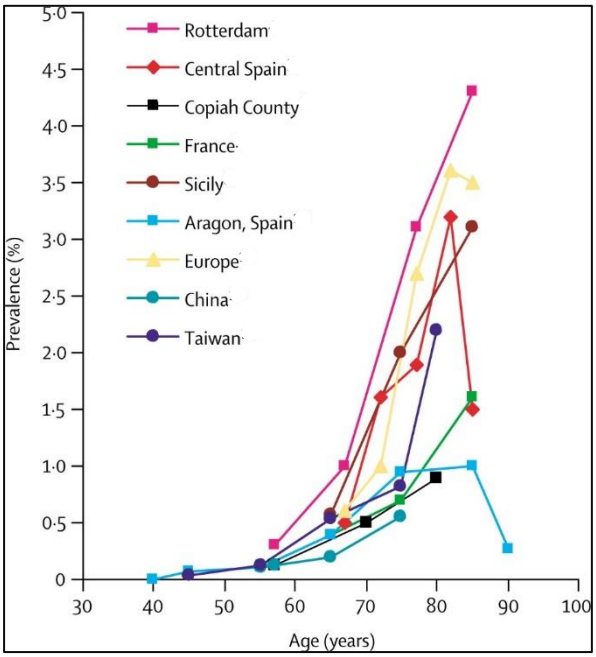


Figure 7- 1: **Population and age-based prevalence studies of PD**.

PD prevalence increases as age increases in all ethnical background. The figure indicated that Caucasian showed the highest PD rate compared to other ethnical groups (de Lau and Breteler 2006).

7.2.1.2. Symptoms of PD

PD is characterized by the presence of motor and non-motor phenotypic features. The most prominent and one of the earliest phenotypes in PD patients is movement abnormalities that include shaking, rigidity, slow movement, and walking difficulty (Jankovic 2008). Clinical features are considered as the crucial signs for PD diagnosis (Litvan et al. 2003). There are four phenotypic features used to distinguish PD; tremor at rest, rigidity, akinesia (or bradykinesia), and postural instability. These features are collectively grouped under the acronym TRAP and they represent the motor symptoms seen in PD (Jankovic 2008). Non-motor symptoms of PD are usually developed in the

late course of the disease. These symptoms include neuropsychiatric phenotypes such as disorders of speech, cognition, mood, behaviour, and thought (Jankovic 2008). Executive dysfunction is the most common cognitive deficit in PD patients. It is represented by defects in the processes of planning, cognitive flexibility, abstract thinking, rule acquisition, initiating appropriate actions and inhibiting inappropriate actions, and selecting relevant sensory information (Jankovic 2008). Non-motor features also include sleep behaviour disorder and sensory abnormalities such as olfactory dysfunction, pain, paresthesia, akathisia, oral pain, and genital pain. These symptoms occur frequently in PD patients, however, they are not used as diagnostic signs of the disease (Jankovic 2008).

7.2.2. Pathogenesis of PD

The main cause of the motor symptoms seen in PD patients is the loss of dopamine-generating cells in the substantia nigra. Due to the excessive neurons loss, the main pathological hallmark in PD appears in the brain of PD individuals as Lewy bodies, which are intracellular aggregations of α-synuclein protein (Dickson et al. 2009). However, the causes for the degradation of dopaminergic neurones, formation of Lewy bodies, and accumulations of α-synuclein are still unknown (Dickson et al. 2009).

The first genetic mutation to be associated with PD was discovered in the α-synuclein (*SNCA*) gene, where the p.A53T mutation was segregating in an Italian kindred and in three unrelated Greek families with autosomal dominant inheritance of PD (Polymeropoulos et al. 1997). This finding gave insights about the possibility of genetic mutations involvement in PD development. Several mutations have since been

identified in monogenic forms of PD. For example, mutations in *SNCA* were found to cause the early onset autosomal dominant inherited form of PD (Lesage & Brice 2009). Mutations in the *LRRK2* gene were identified in patients developing the late onset autosomal dominant inherited form of PD (Cookson 2010). In contrast, loss-of-function mutations in *PARK2, P1NK1,* and *DJ-1* genes were found to cause rare forms of autosomal recessive PD with early onset (Bonifati et al. 2003; Kitada et al. 1998; Valente et al. 2004). In addition, mutations in *ATP13A2, PLA2G6,* and *FBXO7* genes have been identified to cause atypical PD with juvenile onset (Paisan-Ruiz et al. 2009; Shojaee et al. 2008; Ramirez et al. 2006).

The idiopathic form of PD is thought to develop by a complex interplay of several genetic variants that interact with many other non-genetic, environmental risk factors (Polito et al. 2016). Several genome-wide association studies have now been performed and these have confirmed that the causative genes *SNCA, LRKK2,* and *MAPT* are also risk factors for sporadic PD (Simón-sánchez et al. 2009; Satake et al. 2009). In a large PD case-control cohort, 11 loci that met the threshold for genome-wide significance (p-value $<5\times10^{-8}$) were identified which were also reported in a larger replication cohort. Six were previously reported loci (*MAPT, SNCA, HLA-DRB5, BST1, GAK,* and *LRRK2*) and five were newly identified loci (*ACMSD, STK39, MCCC1/LAMP3, SYT11,* and *CCDC62/HIP1R*) (IPDGC 2011).

7.2.3. PD in 22q11.2DS cases

It is now widely known that 22q11.2 deletions are associated with a diverse spectrum of multi systemic phenotypes. Recently, unrelated idiopathic PD cases have been reported within 22q11.2 deletion cohorts. Krahn and colleagues first reported the case of a 30 years old man with a 22q11.2 deletion who also developed childhood onset of

schizophrenia with Parkinson's disease-like phenotype. The standard chromosomal analysis identified a normal 46,XY karyotype, however, FISH with DNA probe D22S75 revealed a microdeletion. Thus, the patient's karyotype was 46,XY.ish del (22)(q11.2q11.2)(D22S75) (Krahn et al. 1998). Later in 2009, Zaleski and colleagues reported 2 additional unrelated patients both with 22q11.2DS and early onset (age <45 years) PD (Zaleski et al. 2009). Booij and colleagues identified a case with 22q11.2DS whose diagnosis with early-onset PD had been confirmed using dopamine transporter (DAT) imaging. DAT is a protein that expressed exclusively in the membrane of presynaptic terminals of dopamine-generating cells (Booij et al. 2010). The largest single study was reported by Butcher and colleagues who identified that 5.9% of 22q11.2DS patients aged 35-64 years (4 out of a single cohort of 68 patients) developed PD (Butcher et al. 2013). The patients had no reported family history of Parkinson's disease. In addition, the genetic screening did not detect any known pathogenic mutations related to Parkinson's disease. A post- mortem analysis was performed on three of the four patients and revealed typical loss of midbrain dopaminergic neurons, with two patients showing α-synuclein-positive Lewy bodies. The authors estimated that 22q11.2 deletions increased the incidence of early-onset PD in their cohort by 0.4%, compared to that of 0.03% seen in the general population. Butcher and colleagues also compared the neuropathology between 22q11.2+PD cases (n= 3) to non-22q11.2DS+PD cases (n= 1), and 22q11.2–PD cases (n= 3) to non-deleted controls (n= 10). This revealed that similar pathological features, characterized by presence of Lewy bodies and dopaminergic degradation in brain tissues, were found in 22q11.2+PD cases and non-22q11.2+PD cases. In contrast, similar normal pathological results were seen in 22q11.2–PD cases and non-deleted controls (Butcher et al. 2013).

Taken together, these reports suggest that microdeletions at 22q11.2 might have a key role in increasing the risk of Parkinson's disease. However, these previous studies have focused on the identification of PD in 22q11.2DS patients, no previous study has investigated the rate of 22q11.2 deletions in a large cohort of PD patients.

7.3. Aims of the study

Based on the initial findings of 22q11.2DS patients developing PD, studies have focused on identifying PD patients among 22q11.2 deletion carriers. However, no previous studies have investigated deletions spanning the 22q11.2 locus in idiopathic PD patients. This study set out to estimate the frequency of 22q11.2 deletions in idiopathic PD in a large PD case-control cohort. The study will investigate association with PD by performing the following lines of investigations:

1) Detecting CNVs spanning the typical 3Mb deleted region in 22q11.2DS (chr22:18,658,219-21,865,185 on hg19) in PD patients and controls.
2) Investigating the association of 22q11.2 deletions with PD by comparing the proportion of 22q11.2 deletions in PD patients and controls.
3) Investigating the relationship of 22q11.2 deletions and PD age at onset by comparing their rate in patients with early PD onset age to those with late PD onset age.

7.4. Materials and methods

7.4.1. Participants

The work in this chapter used the genotype data from a large case-control PD dataset which is typically referred to as the IPDGC NeuroX dataset (Nalls et al. 2014). It was composed of 6,462 PD cases and 6,202 controls from the UK (804 cases, 684 controls), USA (2,069 cases, 2,652 controls), France (564 cases, 479 controls), Germany (1,298 cases, 883 controls), Greece (736 cases, 891 controls), Netherlands (316 cases, 447 controls), and the Parkinson's Progression Markers Initiative (PPMI) cohort (675 cases, 166 controls). All cases were recruited from neurological clinic visits and were diagnosed according to the standard UK Brain Bank criteria with a modification to allow the inclusion of cases that had a family history of PD. Cases had a ratio of males:females of 1.78:1 and had a mean age at onset of 60.2 years (SD =12.5). Controls were recruited from clinic visits, had a mean age of 63.2 (SD =15.6) and had a ratio of males:females of 1.21:1.

7.4.2. 22q11.2 CNV calls

All samples had been previously genotyped using a semi-custom genotyping array called NeuroX (Nalls et al. 2014). This meant that the genotypes of 267,607 SNPs were available for genome-wide CNV analysis. CNV calling was performed using PennCNV (K. Wang et al. 2007) using the same calling protocol that was explained in chapter 2 section 2.5.1.

As this study was specifically aiming to identify CNVs at 22q11.2DS then the CNV calling was restricted to only those that spanned either the 3Mb deleted region at 22q11.2 (chr22:18,658,219-21,865,185 on hg19) that is flanked by LCR22sA–D, or

the 1.5Mb deleted region (chr22:18,658,219-20,519,134 on hg19). Both deletions and duplications within these regions were identified.

7.4.3. Samples quality control

Derivatives of LRR and BAF, generated by PennCNV, were identified by using (LRR_SD and BAF_SD). Samples were excluded if they had a LRR_SD greater than 0.30; a BAF_SD greater than 0.15; or a number of CNVs 3 SDs above the mean CNV number of the case (27.7 CNVs) or control groups (24.1 CNVs). A summary of samples quality control is explained in Figure 7- 2.

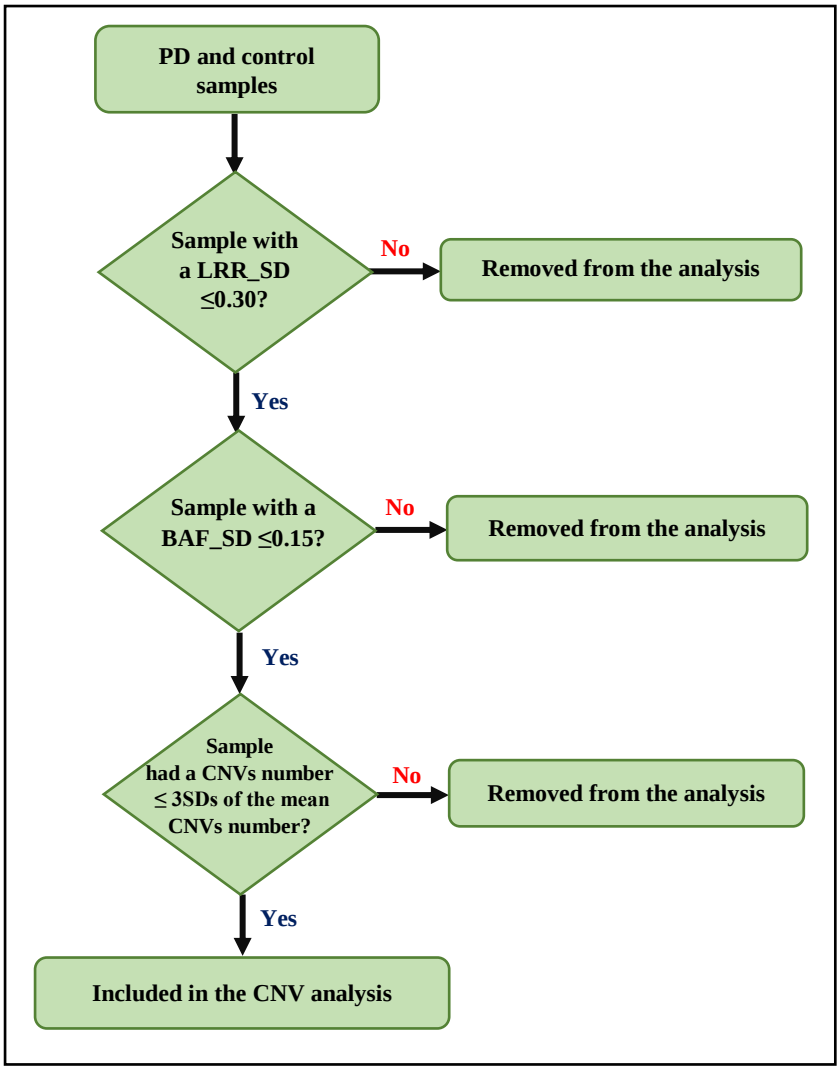


Figure 7- 2: **Flowchart for samples quality control based on PennCNV Statistics.**

7.4.4. CNV quality control

The CNV calls were subjected to the same quality control procedure that was described in chapter 2 section 2.5.2.

7.4.5. 22q11.2 deletions association analyses

For association analysis, the number of chromosome 22q11.2 deletions was compared between PD cases and controls using the Fisher's exact test (two-sided). Similarly, the number of 22q11.2 deletions found in PD patients with early and late PD onset was compared by Fisher's exact test (two-sided).

7.5. Results

7.5.1. Samples quality control

All samples of PD case and control met the quality control criteria. Therefore, all 6,462 patients with PD and 6,202 controls were included for the subsequent analyses.

7.5.2. Identifying 22q11.2 deletions in PD patients

Genome-wide analysis identified a total of 233,783 CNVs in this cohort, however, only 176 CNVs were located on chromosome 22. Following quality control filtering, 5 individuals were identified as carrying large deletions that spanned chr22:18,658,219-21,865,185 on hg19 (Table 7-1).

Table 7- 1: **CNV quality control filtering.**

Quality control filers	*Threshold	**Initial total	***Removed	****Final total
Number of constitutive SNPs spanned by a CNV	10 SNPs	176	0	176
Large CNV calls	100Kb	176	124	52
Genotypes dense CNVs	SNP density of 20,000	52	3	49
Merging adjacent large CNVs	2 adjacent large CNVs in an individual with the gap length <50% of the entire length of the newly merged CNV	49	1	48
Typical 22q11.2 deletion	Outside the common 3Mb 22q11.2 region or the rare 1.5Mb 22q11.2 region	48	43	5

* Filtering threshold or criteria.
** Initial total: Initial total number of samples.
*** Removed: Number of removed samples due to failure in matching filtering threshold criteria.
**** Final total: Final total number of samples.

The five deletion carriers were unrelated patients with PD and all carried the most common 3Mb deletion in 22q11.2DS spanning LCRs A-D (Figure 7- 3).

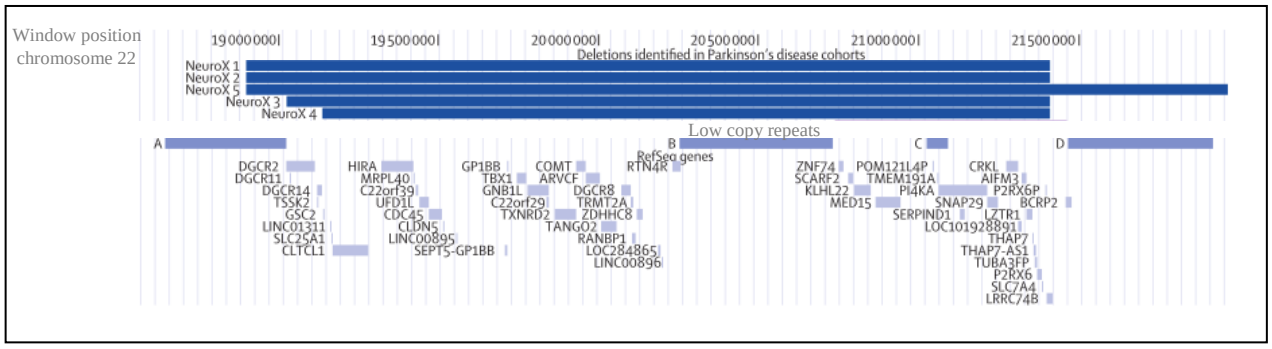


Figure 7- 3: **Genomic location of the 22q11.2 deletions found in the five patients with PD and location of the low copy repeat regions.**

7.5.3. Investigating the association of 22q11.2 deletions in PD patients

The number of 22q11.2 deletions in PD patients (n= 6,462) was compared to that in control cohort (n= 6,202). The results of this association analysis revealed a pattern of enrichment for 22q11.2 deletions in patients with Parkinson's disease, where cases had increased rate of 22q11.2 deletions (0.08%) compared to controls (0.0%) which fell just short of nominal significance (p-value= 0.063) (Table 7- 2).

Table 7- 2: **Results of association analyses of deletions in chromosome 22q11.2 in PD cases and controls**.

	Total	**22q11.2 deletion carriers**	**Frequency**
PD patients	6,462	5	0.08%
Controls	6,202	0	0.0%
Two-sided Fisher's exact test p-value	0.063		

7.5.4. Investigating the association of PD age of onset and 22q11.2 deletions

As the previous reports had indicated that patients with 22q11.2DS developed PD before 50 years of age, the data was further analysed to investigate whether the presence of 22q11.2 deletion is associated with PD age of onset. 5,599 of PD cases, which included all 5 PD patients that carried a 22q11.2 deletion, had available data for

PD age at onset. The average age of onset of PD in the 5 carriers of a 22q11.2 deletion was 43.9 years (SE= 4.37), while it was 60.3 years (SE= 0.14) in PD patients without a 22q11.2 deletion. This finding indicates PD cases with 22q11.2 deletion develop PD earlier than those without 22q11.2 deletion by ~16.4 years (SE= 4.0).

To formally test the relationship between deletions at 22q11.2 and PD age at onset the PD cases were considered as early onset PD (EOPD) if they presented with PD symptoms by <45 years, and late onset PD (LOPD) if they expressed PD phenotypes by >45 years of age (Wickremaratchi et al. 2009). In the NeuroX dataset, 660 PD patients had EOPD, whereas 4,939 PD patients had LOPD (Table 7- 3).

Table 7- 3: **Patients with Parkinson's disease entered into the age at onset analysis.**

	22q11.2 deletion carriers	Non-22q11.2 deletion carriers	Total number
EOPD patients (AAO <45 years)	4	656	660
LOPD patients (AAO >45 years)	1	4,938	4,939
Controls	0	6,202	6,202

AAO= Age at onset.
EOPD= Early onset of PD.
LOPD= Late onset of PD.

Association analysis revealed a significant enrichment for 22q11.2 deletions in EOPD cases (n=4; frequency=0.61%) compared to that seen in LOPD cases (n=1; frequency=0.02%) (p-value= 9.0×10^{-4}). Furthermore, when compared to the control cohort, the increased rate of 22q11.2 deletions detected in the EOPD patients was highly significant (p-value = 9.0×10^{-7}). On the other hand, no evidence of a significant enrichment of 22q11.2 deletions was identified in LOPD cases when they were compared to controls (p-value= 0.443) (Table 7- 4).

Table 7- 4: **Results of association analyses of deletions spanning 3Mb 22q11.2 region according to PD age at onset.**

	Number of 22q11.2 deletion carriers	Total number	Frequency of 22q11.2 deletion carriers	Association tests p-values
EPDO patients vs LOPD patients				
EOPD patients (AAO <45 years)	4	660	0.61%	$9.0x10^{-4}$
LOPD patients (AAO >45 years)	1	4,939	0.02%	
EPDO patients vs controls				
EOPD patients (AAO <45 years)	4	660	0.61%	$9.0x10^{-7}$
Controls	0	6,202	0%	
LPDO patients vs controls				
LOPD patients (AAO >45 years)	1	4,939	0.02%	0.443
Controls	0	6,202	0%	

7.6. Discussion

In the study presented in this chapter, the rate of 22q11.2 deletions was compared in a large idiopathic PD case-control cohort. Five unrelated PD patients were found to carry a ~3Mb deletion spanning LCR22s A-D at 22q11.2, while none of the controls carried any deletion at 22q11.2. Association analysis revealed that while the proportion 22q11.2 deletions was increased in PD patients (rate=0.08%) compared to controls (rate=0.0%), the difference fell just short of nominal significance (p-value= 0.063). PD cases with known PD loci mutations were excluded from this analysis. Also the selected cases had no reported psychiatric symptoms and other age-related disease, such as Alzheimer's disease, were excluded. The five PD 22q11.2 carriers fulfilled the UK Brain Bank for Parkinson's disease. A single patient had a valvular heart lesion with no additional information about the cardiac anomaly, while another was found with mitral valve prolapse identified after PD was diagnosed. Regarding to psychiatric illness, only a single patient had depression before PD diagnosis, while none of the 5 patients had schizophrenia before onset of PD. Two cases had a cleft palate or oropharyngeal malformation before the onset of PD, with one expressing bifid uvula during investigations for dysphagia after diagnosis of PD. One patient was identified to have hypocalaemia. These identified phenotypes are relevant to the clinical features seen in 22q11.2DS which confirm the influence of 22q11.2 deletions in the symptoms in these PD patients.

The results from this association analysis have been included in a larger meta-analysis that combined four large PD case-control cohorts (Total PD cases= 9,387, Total controls= 13,863) (Mok et al. 2016). In addition to the 5 PD patients carrying deletions at 22q11.2 that were identified in this chapter, 3 unrelated patients with PD were identified to carry the most common 3Mb deletion in 22q11.2DS from three

independent PD cohorts; UK-WTCCC2, Dutch Parkinson's Disease Genetics Consortium, and USA-NIA. Two PD patients were found to carry 22q11.2 deletions in the UK-WTCCC2 PD cohort (1,592 cases and 4,939 controls). Only a single PD patient was a 22q11.2 deletion carrier from the Dutch PD cohort (740 cases and 1,996 controls) (Figure 7- 4). No deletion carriers were identified in the smallest cohort from the USA-NIA (593 cases and 726 controls). The results of the meta-analysis conducted by Mok and colleagues (Mok et al. 2016) revealed that PD patients have higher rate of 22q11.2 deletions (0.09%) compared to controls (0%), supporting the findings in the NeuroX sample that were presented in this chapter. This finding was statistically highly significant (p-value= $5.6x10^{-4}$) (Table 7- 5).

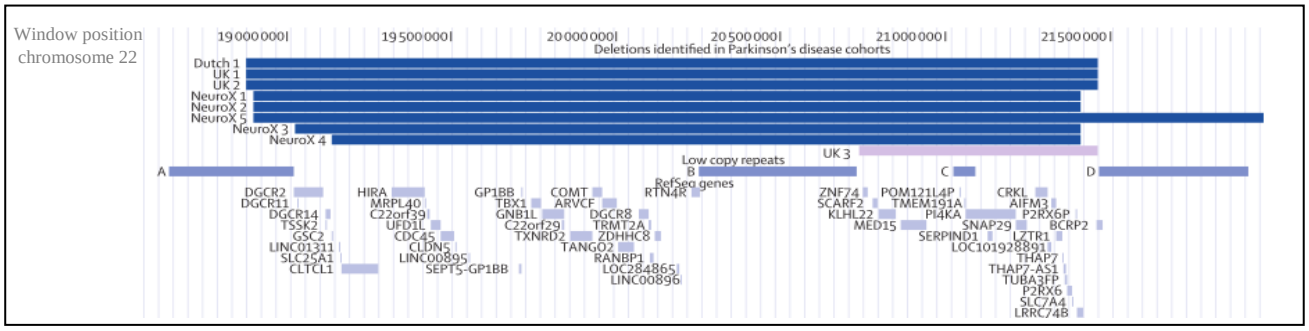


Figure 7- 4: **Genomic location of the 22q11.2 deletions found in the eight patients with PD from the four combined independent PD cohorts and location of the low copy repeat regions** (Mok et al. 2016).

Table 7- 5: **Results of meta-analysis of 22q11.2 deletions in PD cases and controls for the combined four independent PD case-control including NeuroX cohort.**

Studies		22q11.2 deletion carriers	Total	Frequency
Combined the four PD cohorts	PD patients	8	9,387	0.09%
	Controls	0	13,863	0.0%
	*Meta-analysis p-value	0.00056		

* Meta-analysis by Mantel-Haenszel test (Two-sided).

The initial clinical reports of PD occurring in subjects with 22q11.2DS typically reported the patients to have an early age at onset. As the age at onset was known for 5,599 of the PD patients analysed in this study, the data was further investigated to compare the rate of deletions in 22q11.2 region between PD cases with LOPD and those with EOPD. This revealed highly significant evidence (p-value= $9.0x10^{-4}$) that the EOPD cases (0.61%) carried more 22q11.2 deletions than the LOPD cases (0.02%). Comparing the frequency of 22q11.2 deletions in EOPD (0.61%) comparing to controls (0.0%) further demonstrated the significant over-representation of deletions in the EOPD patients (p-value= $9.0x10^{-7}$).

In the meta-analysis, conducted by Mok and colleagues (Mok et al. 2016), the age at onset of PD data was available for a total of 8,451 patients. Analysis of these samples allowed them to further investigate the role of 22q11.2 deletions and PD age of onset. The results showed a higher rate of 22q11.2 deletions in the EOPD patients (0.5%) compared to that in the LOPD patients (0.04%) and unaffected controls (0%). The results of Mantel-Haenszel exact test (two-sided) revealed a significant enrichment for 22q11.2 deletion in EOPD relative to LOPD (p-value= $7.0x10^{-4}$). Interestingly, highly significant evidence was found when rate of 22q11.2 deletions was compared between EOPD patients and unaffected controls (p-value= $1.47x10^{-11}$), while a nominally significant evidence was found when LOPD patients were compared to controls (p-value= 0.023) (Table 7- 6) (Mok et al. 2016). These support the association between EOPD and 22q11.2DS that were identified reported in this chapter and are in line with the previous 22q11.2DS cases that had been reported to show symptoms of PD-like motor abnormalities (Krahn et al. 1998; Zaleski et al. 2009; Booij et al. 2010; Butcher et al. 2013).

Table 7- 6: **Results of meta-analysis of 22q11.2 regions and PD age at onset in the combined large PD cohort including NeuroX cohort.**

	Frequency of 22q11 deletions (No. CNVs/No. samples)			Association Tests* (p-value)		
Studies	**EOPD subjects (AAO<45 years)**	**LOPD subjects (AAO<45 years)**	**Controls**	**EOPD vs LOPD**	**EOPD vs Controls**	**LOPD vs Controls**
Combined cohort	0.5% (5/1,014)	0.04% (3/7,437)	0% (0/13,863)	$7.0x10^{-4}$	$1.47x10^{-11}$	0.023

*Meta-analysis by Mantel-Haenszel test (Two-sided).

It has been reported that EOPD have been associated with rare mutations. 4-6% of EOPD cases have known genetic mutations in genes including *LRRK2*, *PARK2*, *SNCA*, *PARK7*, and *PINK1* (Houlden & Singleton 2012). EOPD patients with 22q11.2 deletions identified in this study share the same proximal deleted region (1.5Mb). This is also consistent with those 22q11.2DS cases with EOPD identified by Butcher and colleagues. This region involves approximately 30 genes, however, none of which overlap any known PD loci.

There are however, a number of candidate genes that might be implicated in PD-related pathways that are located in this region. These include microRNA miR-185, which is predicted to target *LRRK2* (Dweep et al. 2011; Heman-Ackah et al. 2013), and *DGCR8* that are key regulator genes in the biogenesis of brain microRNA (Stark et al. 2008). Altered gene expression in 22q11.2DS by disrupted microRNA-mediated posttranscriptional regulation could possibly directly or indirectly impact on the expression of PD risk genes elsewhere in the genome (Brzustowicz & Bassett 2012). Other possible 22q11.2 candidate genes are *SEPT5*, encoding a protein that functionally interacts with the product of *PARK2* (Zhang et al. 2000), *COMT* (Levodopa inhibition), and 6 mitochondrial genes (Maynard et al. 2008). All of these genes have been shown to be brain expressed in 22q11.2DS mice models (Meechan et

al. 2009; Maynard et al. 2003), as well as *DGCR8*, *SEPT5*, and *COMT* have been identified as dosage sensitive in our study in chapter 4.

In conclusion, in this chapter deletions at 22q11.2 were analysed in a large cohort of idiopathic PD and controls. A potential trend for an increasing 22q11.2 deletion rate in PD cases was revealed (p-value= 0.069), which was supported by a significant meta-analysis that including this PD case-control cohort and three independent PD cohort (p-value= $5.6x10^{-4}$). In addition, this study provided significant evidence for association of 22q11.2 deletions with early onset PD compared to late onset (p-value= $9.0x10^{-4}$), which was also supported in the meta-analysis of Mok and colleagues (Mok et al. 2016). This evidence can be used to make clinicians aware of the possibility that 22q11.2DS patients could develop PD symptoms at an early age. Also, screening for deletions in the 22q11.2 region is advised in patients with EOPD and psychiatric features. The biological mechanism behind the association of 22q11.2 deletions and EOPD should be investigated in future studies.

Chapter 8: General Discussion and Future Direction

8.1. Introduction

The 22q11.2 deletion is one of the most common pathogenic CNVs with an estimated incidence of 1 in 4000 births (Scambler 2000). It is widely known to be a strong risk factor for psychotic disorder and schizophrenia with about 25% of adult deletion carriers developing schizophrenia (Murphy et al. 1999). Children carrying the 22q11.2 deletion have an increasing risk to develop behavioural and psychiatric disorders such as ADHD (41%) and ASD (26%) (Niarchou et al. 2014). However, the syndrome is associated with incomplete penetrance of neuropsychiatric diseases as some patients express psychiatric phenotypes while others do not (Niarchou et al. 2014). 22q11.2 hemizygous microdeletions occur as a *de novo* mutation in 95% of 22q11.2DS patients and are inherited in the remaining 5% (Scambler 2000). The vast majority of the patients (85%) share a ~3Mb deletion and the remaining patients carry a smaller ~1.5Mb deletion (Scambler 2000).

The molecular mechanisms that underlie the increasing risk of psychiatric phenotypes from a single homogenous microdeletion on chromosome 22q11.2 are yet not clearly understood. This PhD thesis attempted to contribute to our understanding of how deletions at 22q11.2 are involved in the pathogenesis of neuropsychiatric disorders. This chapter discusses the general considerations drawn from the results and outlines the future studies that may be undertaken.

8.2. Implication of this PhD results to the 22q11.2DS knowledge

8.2.1. Investigating the molecular mechanisms underlie the psychiatric phenotype variability in 22q11.2DS

As a central working hypothesis, the increased risk of neuropsychiatric phenotypes in 22q11.2DS was hypothesized to be due to dosage sensitive genes within the haploinsufficient region on chromosome 22q11.2.

To conduct the work in this thesis it was first necessary to characterise a cohort of 76 children with 22q11.2DS who had high quality genotypic data and were ethnically homogenous. Gene dosage sensitivity in 22q11.2DS was then investigated by differential gene expression analysis for genome-wide genes to highlight differentially expressed genes in 22q11.2 deletion carriers relative to non-deleted controls. Thirty-nine genes within the 22q11.2 deleted region were found to be dosage sensitive in 22q11.2DS, 29 of which had been previously shown to be expressed in the brain. Most of these genes had been also supported to be differentially expressed in other expression studies in 22q11.2DS patients. 7 of the 39 dosage sensitive genes were also reported in a smaller 22q11.2DS cohort (n= 7) (van Beveren et al. 2012), while 29 of them were also identified in a larger 22q11.2DS cohort (n= 46) (Jalbrzikowski et al. 2015). Moreover, 7 of the 39 dosage sensitive genes were also found to be differentially expressed in brain tissues of the 22q11.2DS mice models compared to the wild-type mice (Sivagnanasundaram et al. 2007). Mouse models of heterozygous knockouts for some of these genes have been reported to express impaired cognitive functions that are processed in the prefrontal cortex and hippocampus, which are considered as endophenotypes of schizophrenia and psychiatric symptoms in

22q11.2DS patients (Gogos et al. 2009; Gogos et al. 1998; Hsu et al. 2007; Mukai et al. 2004; Paylor et al. 2006; Suzuki et al. 2009). It has been suggested that the deletion might regulate the expression of other nearby genes, causing positional effects. Differential expression analysis of genes flanking the deletion on 22q11.2 indicated that no significant evidence of there being a positional effect on the expression of genes located nearby the deletion.

The results of pathway analysis of differentially expressed genes located outside the deleted region highlighted significant (FDR <0.05) functional biological groups that could be involved in the pathogenesis of some phenotypes related to 22q11.2DS. These included functional groups related to immune system and protein synthesis, particularly functions in leukocyte activation and post translational modifications. However, it is difficult to associate the identified functional categories to the psychiatric phenotypes seen in 22q11.2DS as they are enriched for differentially expressed genes in the peripheral whole blood. Specific brain tissue are ideally required to accurately investigate psychiatric phenotypes in 22q11.2DS.

However, it has been shown that due to tight haemostatic regulation that can compensate for a gene dosage imbalance, not all genes spanned by deletions are dosage sensitive. This thesis therefore next investigated the hypothesis that risk variants remaining on the intact 22q11.2 chromosome can potentially be exposed by the primary deletion and act as modifying factors for neuropsychiatric illnesses seen in 22q11.2DS. To investigate this, the variants present on the remaining 22q11.2 chromosome were screened in 22q11.2 patients. After identifying common variants by genotyping then by imputing genotype data, association analyses were conducted

to test whether the identified common variants were associated with the psychiatric phenotypes in 22q11.2DS. The results of these analyses failed to provide significant evidence for either common SNPs (p-values $<9.9x10^{-6}$) or for putative CNVs (p-values <0.017) being associated with neuropsychiatric diseases in 22q11.2DS.

Despite this, a non-significant trend of secondary small 22q11.2 deletions was observed in 22q11.2DS patients with psychiatric phenotypes (rate= 5.7%) relative to patients with no psychiatric phenotypes (rate= 2.6%). This implicated 2 secondary small 22q11.2 deletions that were found in 2 patients with psychiatric disease, which are potentially disrupting the genes *PRODH* and *THAP7*.

Although there are a large number of studies that have analysed common variants in 22q11.2DS affected with psychiatric phenotypes (Gothelf et al. 2005; Bearden et al. 2004; Kates et al. 2006; Shashi et al. 2006; Shashi et al. 2010; Bassett et al. 2007; Baker et al. 2005; Glaser, Debbane, et al. 2006; van Amelsvoort et al. 2008; Vorstman, Turetsky, et al. 2009; Jungerius et al. 2008; Ikeda et al. 2010; Raux et al. 2007), they have typically focused on candidate genes spanned by the deletion. The findings of these studied were inconsistent and generally failed to replicate. To our knowledge, this is the first study that has systematically looked at the variants spanning the non-deleted 22q11.2 allele and studied their association to the psychiatric phenotypes. A few studies have investigated the risk of phenotypes associated with 22q11.2DS other than the psychiatric ones, such as Bernard-Soulier Syndrome and other atypical features (Budarf et al. 1995; Ludlow et al. 1996; McDonald-McGinn et al. 2013). These studies have identified additional mutations in genes that are relevant to the additional phenotypes. These promising results were achieved by applying more precise approaches to detect these mutations such as whole and targeted exome sequencing methods, which are recommended for future investigations.

Based on the evidence that a second CNV hit hypothesis was relevant in neurodevelopmental and psychiatric disorders, this thesis investigated whether CNVs that are additional to the primary 22q11.2 deletions could be relevant in psychiatric disease. The presence of a second CNV in addition to the main pathogenic 22q11.2 deletion was investigated by calling genome-wide CNVs using PennCNV. Then, the frequency of additional CNVs was compared in 22q11.2DS affected with childhood psychiatric disease to those unaffected. Again the results failed to provide significant evidence for additional large, rare CNVs in 22q11.2DS patients with psychiatric phenotypes. However, a non-significant potential trend for enrichment for large, rare CNVs was observed in patients with ADHD, ASD, and psychiatric illnesses (enrichment rates = 3, 1.8, and 2.29; p-values= 0.20, 0.44, and 0.29 respectively). Focusing on CNVs that span one of those previously implicated in a psychiatric phenotype, also failed to show an evidence for enrichment of pathogenic CNVs in 22q11.2DS patients affected with psychiatric disease. Despite not identifying a significant result, this study is the first to analyse the second CNV hits in 22q11.2DS with childhood psychiatric symptoms, ADHD and ASD. Three previous studies also attempted to investigate the presence of additional CNV phenotypic modifiers, however, in cohorts of 22q11.2DS adults (Bassett et al. 2008; Williams et al. 2013), and in a combined cohort of 22q11.2 deletion and duplication carriers (Li et al. 2012). These studies showed no significant evidence for either an increased burden of large, rare CNVs or enrichment for previously identified CNVs in 22q11.2DS with psychiatric disease.

8.2.2. Investigating 22q11.2 deletions in Parkinson's disease patients

This PhD thesis provided the opportunity to investigate the presence of deletions spanning the common 22q11.2 deleted region in the largest case-control cohort of idiopathic PD. This analysis was based on the previously reported cases of 22q11.2DS who manifested early onset of PD (Krahn et al. 1998; Zaleski et al. 2009; Booij et al. 2010; Butcher et al. 2013). However, no previous study had analysed 22q11.2 deletions in PD patients. Deletions within the commonly 22q11.2 deleted region were called using PennCNV and association analyses revealed a trend of an increased burden of 22q11.2 deletions in PD patients compared to controls (p-value =0.063). However, comparing the rate of deletions according to disease age at onset revealed significant evidence for a higher rate of 22q11.2 deletions in PD cases with an early onset compared to those with late onset, and healthy controls (p-values = $9.0x10^{-4}$, and $9.0x10^{-7}$ respectively). The data from this study was subsequently combined with other PD cohorts and included in a meta-analysis which identified highly significant evidence for an increased rate of 22q11.2 deletions in PD cases compared to controls (p-value =0.00056). Additionally, the meta-analysis also confirmed the highly significant increase of 22q11.2 deletions in EOPD relative to LOPD and controls (p-values = $7.0x10^{-4}$, and $1.47x10^{-11}$ respectively). These significant findings provide strong support that 22q11.2DS patients could develop PD symptoms in early age.

The possible molecular mechanisms, that have been studied in the rest of PhD thesis should be tested by future studies in a cohort of adult 22q11.2DS with and without PD to investigate whether they are relevant to the increased risk of EOPD in 22q11.2DS patients.

8.3. Limitations of this PhD study

8.3.1. Small sample size underpowered the study

The main limitation of this study was that it was conducted in a small sample size of 22q11.2DS patients (n= 76). Nevertheless, it is worth noting that at the time of starting these PhD studies, this was the largest available study 22q11.2DS for investigating the psychiatric phenotypes by gene expression analysis. Although more samples were collected by the ongoing ECHO study, I had to stop processing more samples at some point to start analysing the collected data from the microarrays experiments. Despite a number of genotyped samples that had been excluded (n= 19), however, the selected 76 samples were well-characterized and passed a rigorous protocol of QC that selected only true 22q11.2 deletion carriers that came from homogenous Caucasian populations. In addition, the identity of these samples were confirmed with the reported data by sex chromosome and IBD analyses to avoid related, duplicated, and potentially contaminated samples.

In association studies, particularly GWAS studies, an increased sample size is required as they typically assay more than half a million polymorphisms. Analysing this number of variants means only a p-value$< 5x10^{-8}$ (genome-wide significance level) can be confidently reported as an association signal (Pe'er et al. 2008). However, in order to achieve this level of significance a large number of samples are required (Pe'er et al. 2008). Accordingly, many research centres have collaborated to recruit more samples for GWAS analysis and to share control cohorts. This approach is necessary for 22q11.2DS and will yield better powered studies.

8.3.2. Limited efficiency in the ascertainment of 22q11.2DS patients with ASD

The method used to diagnose ASD phenotype in 22q11.2DS children is a potential limitation in this study. Unlike the diagnosis of ADHD in 22q11.2DS children, the ASD was diagnosed based on the SCQ score, which is not a categorical diagnosis that could certainly identify healthy 22q11.2DS children without ASD. However, this method is based on splitting the patients based on the SCQ score cut off of 15. Patients with an SCQ above 15 have ASD, while those with an SCQ below 15 have no ASD. This could artificially classify the patients into two groups as the SCQ scores are normally distributed among 22q11.2DS patients. Therefore, this might limit the power of the test by dichotomizing the continues variable (SCQ score) that have a normal distribution among our 22q11.2DS cohort.

However, the SCQ questionnaire is a conventional method that have been used to diagnose ASD by using the threshold of 15 that suggests ASD (Berument et al. 1999). In addition, the SCQ questionnaire is an understandable form of the ADI-R for the parents and can be completed as a written questionnaire rather than a semi-structured interview. It consists of questions that are derived from the ADI-R and focus on behaviours that can be identified by non-professionals (Charman & Baird 2002). Furthermore, the SCQ diagnosis method takes less time to be completed as it is a written questionnaire rather than an interview (Charman & Baird 2002). The SCQ cut-off of 15 for ASD diagnosis has a high sensitivity (0.96) and specificity (0.8) in identifying children with the disease (Charman & Baird 2002). Therefore, this method was selected over the ADI-R in the research protocols (Charman & Baird 2002).

An alternative method for ASD diagnosis is ADI-R (Rutter, Bailey & Lord 2003), which is a semi-structured interview conducted with the primary caregiver to assess

autism. The interview covers the same four phenotypic domains in the SCQ questionnaire mentioned in chapter 2 section 2.1.2.1.

However, SCQ score is highly correlated with ADI-R score (r^2= 0.71) and it is effective as the ADI-R in distinguishing children with autism from typically developing children; however, the ADI-R is better in identifying children with autism from those with ID (Intellectual disability) (Berument et al. 1999). Therefore, although the ADI-R interview is long, it is a better method for assessing ASD in 22q11.2DS children as the ID is also associated with the syndrome. Additionally, the SCQ is considered by the researchers who developed both the SCQ and ADI-R as a good screening method but inappropriate for determining a diagnosis in the level of the individual (Berument et al. 1999).

8.3.3. Limited efficiency in detecting secondary small 22q11.2 CNVs

The association study presented in this thesis aimed to detect common variants in the 22q11.2 region. Additional analysis is required to investigate rare variants. Nevertheless, we attempted to identify rare small CNVs that potentially occurred on the non-deleted chromosome at 22q11.2 by using the array-based genotypes only. The identification of rare second deletions was based on screening null genotypes SNPs, while identification of rare second duplications was performed by PennCNV calling. Both approaches are not very precise in detecting small CNVs as they rely only on the genotyped markers present on the array and it is possible that some regions do not have enough coverage. Even for the most up to date genome-wide microarrays, it is virtually impossible to include probes against every single nucleotide position (Hurd & Nelson 2009). In addition, genotyping microarrays utilize hybridization probes to

analyse the prevalence of known DNA sequences (Kahvejian et al. 2008) and as such cannot be used to identify unknown genetic changes (Kahvejian et al. 2008).

Other methods such as targeted exome or whole exome sequencing are more efficient. Contrary to microarrays, the high-throughput sequencing technologies that target exome sequencing directly provide nucleotide sequences at the thousands of exonic loci tested (Biesecker 2010). This means, that it could provide greater details of common and rare genetic variants in the coding genes, either within the non-deleted chromosome 22 by using target exome sequencing, or both 22q11.2 locus and other genomic loci by whole exome sequencing. However, one main limitation is that exome sequencing is more expensive than hybridization-microarray technologies on the basis of per-sample (Biesecker 2010). In addition, exome sequencing for many individuals generates a large quantity of data and sequence information, that requires a large amount of data analysis (Kahvejian et al. 2008). The statistical analysis of such volumes of sequencing data for multiple individuals is challenging (Kahvejian et al. 2008).

8.3.4. Transcriptome profiling of peripheral blood in 22q11.2DS

A major limitation of the expression analysis is the fact the cells originally came from blood. This is because it is the most accessible tissue for studies including children as it is neither feasible nor ethical to extract brain samples from living 22q11.DS patients. Sullivan and colleagues investigated the correlation of the expression profile in whole blood to different CNS regions in controls and, depending on CNS region, found correlations ranging from 0.44 to 0.58 (Sullivan et al. 2006). The authors also concluded “that gene expression in whole blood is neither perfectly correlated and

useful nor perfectly uncorrelated and useless with gene expression in multiple tissues" (Sullivan et al. 2006). Encouragingly, this thesis found that there was an overlap between genes differentially expressed in hippocampal tissue from *Df1*+/- murine models of 22q11.2DS (Sivagnanasundaram et al. 2007) and genes differentially expressed in PMBCs from 22q11.2DS patients. This was supported by the previous finding of van Beveren and colleagues, who also identified an overlap in dosage sensitive genes in 22q11.2DS patients and 22q11.2DS mice models (van Beveren et al. 2012).

Another consideration is that gene expression will change over time and the expression profiles relevant to psychiatric disorders might occur only during a specific critical development period. Amati and colleagues observed there are dynamic changes in expression profiles of the 22q11.2 genes and orthologous genes during mouse development (Amati et al. 2007). However, as previously indicated, phenotypes of ADHD and ASD mainly occur during childhood (Schneider et al. 2014). Therefore, the cohort analysed in this thesis were selected children with 22q11.2DS who had an average age 11.9 years.

Overall, this study was relatively small in scale, and a larger sample would provide more robust analysis. In addition, using more efficient and precise method to diagnose ASD in 22q11.2DS children than the SCQ score is highly recommended. Moreover, the study only utilised array-based genotypes for the detection of potential secondary CNVs at 22q11.2. Applying more robust approach, such as sequencing, will increase the efficiency to detect them. Furthermore, RNA was extracted from peripheral blood samples for differential gene expression analysis.

8.4. Conclusion and future direction

This study provided strong evidence for the involvement of 22q11.2 deletions in the early onset form of PD. Although the present study failed to provide significant evidence for some of the potential molecular mechanisms that underlie neuropsychiatric phenotypes, particularly the presence of additional risk genetic variants within the non-deleted 22q11.2 chromosome, it did identify a number of dosage sensitive genes in 22q11.2DS. At this point it is not possible to make definitive conclusions, but one can speculate that the dosage sensitive genes in 22q11.2DS could possibly play key roles in modifying the psychiatric phenotypes.

These preliminary findings although promising, need further investigation mainly due to the studies being underpowered. A collaboration between research groups is therefore required to obtain sufficient sample sizes to maximize the power of such studies.

Based on the evidence of the 22q11.2 deletion being a risk factor for early onset of PD, future studies should investigate an adult 22q11.2 deletion cohort with available PD data. This could investigate the biological mechanisms that underlie the increasing risk of PD in the deletion carriers. That can be achieved by using similar approaches to those used in this thesis to test the potential molecular mechanisms in 22q11.2DS children with ADHD and ASD phenotypes. Large cohorts of 22q11.2DS adults will also enable to investigations of other 22q11.2DS adulthood-related psychiatric phenotypes such as schizophrenia (Schneider et al. 2014).

The dosage sensitive genes identified in the present study are strong candidates for further investigations into the behavioural phenotype associated with 22q11.2DS. Among these genes, *DGCR8* was found to be dosage sensitive in 22q11.2DS, and is a

component of the "microprocessor" complex that is essential for microRNA production (Tomari & Zamore 2005). Mouse models with haploinsufficiency of *Dgcr8* gene have showed reduced *Dgcr8* expression and also a downregulation of a specific set of mature miRNA and an upregulation of a number of genes. This means that reduced *Dgcr8* expression can lead to reduced miRNA expression, which in turn resulted in the upregulated expression of other genes (Stark et al. 2008). It would therefore be interesting to look at *DGCR8*-target genes in 22q11.2DS patients and analyse their expression patterns to investigate the impact of haploinsufficiency of *DGCR8* on miRNA and other genome-wide genes. Thus, further investigations are necessary to study the interaction between the genes and the possibility of 22q11.2 dosage sensitive genes being regulator genes for adjacent genes. As this gene expression study has used RNA samples extracted from peripheral blood, future studies will be required to establish how they reflect gene expression in the human brain by using pluripotent stem cell (iPS cell) derived neuronal cells in 22q11.2DS. This work is needed to validate current findings and to investigate expression differences in genes not expressed in blood.

An additional molecular mechanism that can potentially influence the increased risk of psychiatric disease is the presence of differentially methylated genes. This hypothesis was raised on the basis of *DGCR8* being shown to have a reduced expression in 22q11.2 deletion carriers. *miR-185* is also another gene spanned by the 22q11.2 deletion which encodes miRNA-185 that is regulated by *DGCR8* (Y. Wang et al. 2007). A recent study has shown that the expression of *miR-185* is reduced to ~25% in *Df(16)A+/-* mice compared to the wild-type mice (Xu et al. 2013). As miRNA-185 regulates the expression of DNA methyl-transferase1 and somatic mutations at *miR-185* have been previously shown to result in genome-wide changes

in DNA methylation in tumour cells (Zhang et al. 2011), it could potentially lead to a disruption in DNA methylation in 22q11.2DS. This can be analysed by conducting studies that aim to quantify genome-wide patterns of DNA methylation in 22q11.2DS patients. Hannon and colleagues performed an example of such studies in schizophrenia and provided promising findings using epigenetic approach to understand the nature of complex psychiatric traits and diseases (Hannon et al. 2016).

Appendices

9.1. Gene expression analysis

9.1.1. Samples quality control

9.1.1.1. Confirming sample gender molecularly

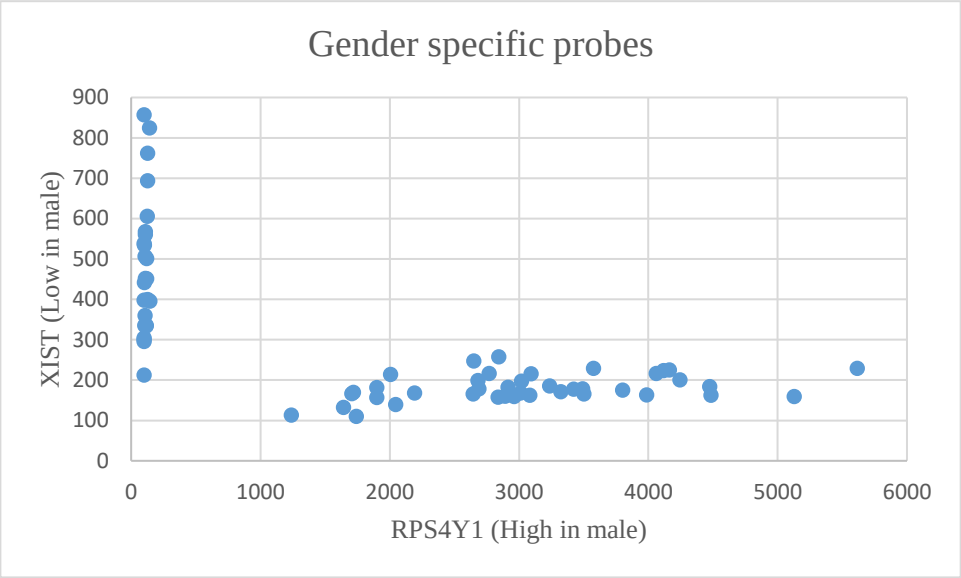


Figure 9- 1: **Gender-specific probes (*RPS4Y1* and *XIST* genes) in our expression data.**

Vawter and colleagues reported that the expression of *RPS4Y1* is high in males while the expression of *XIST* gene is low in males and the vice versa in females (Vawter et al. 2004). Expression values of both probes were plotted for each 22q11.2DS and control samples. Males are expected to be on the right side of the plot as they have high expression of *RPS4Y1* gene and low expression of *XIST* gene. On the other hand, females are expected to be on the left side of the plot as they have a high expression of *XIST* gene and low expression of *RPS4Y1* gene. All females in our cohort are aligned in Y-axis with high expression of *XIST* gene and low expression of *RPS4Y1* gene. On the other hand, males are aligned in X-axis with high expression *RPS4Y1* gene and low expression of *XIST* gene. Gender of a single control sample was not confirmed molecularly.

Table 9- 1: **Differential gene expression analysis on gender-specific probes.** Gender-specific probes are significantly differentially expressed when females compared to males. *RPS4YI* probe is downregulate, while *XIST* probe is upregulated in females compared to males.

Gender-specific probes	Fold change (Females vs males)	P-value
RPS4Y1	-5.264277509	$2.73x10^{-33}$
XIST	2.914531949	$1.01x10^{-19}$

.

9.1.1.2. Confirming the 22q11.2 deletion status by signal intensity of the 3Mb 22q11.2 probes

Figure 9- 2: **Heat map for the 3Mb 22q11.2 probes.**

In the heat map each column represents a sample and each raw represents a probe and the numbers are signal intensity values (expression values) for each probe in each sample. Expression of 22q11.2 probes is coloured based on the intensity values, probes with high expression are coloured by green shadows, while probes with low expression are coloured by red shadows. The map shows that the non-deleted controls have more green probes indicating more 22q11.2 probes have high expression which means no 22q11.2 deletion was found in these individuals. While the 22q11.2 deletion carriers have more red probes indicating more 22q11.2 probes have low expression due to the hemizygous 22q11.2 deletions in these individuals.
* The sample has a 1.5Mb 22q11.2 deletion which can be observed by more red probes within the proximal 1.5Mb region of 22q11.2 (top rows) and more green probes within the distal region of 22q11.2 (bottom rows).

9.1.2. Gene expression data pre-processing (Normalization and Transformation)

Data normalization and transformation are required to eliminate non-biological variance among samples and to make the data more appropriate for gene expression comparisons. In order to determine the most appropriate approaches of normalization and transformation for our data, different combinations of normalization and transformation methods were applied (Figure 9- 3). For each combined method, pre-processed expression data was plotted and data inter-array correlation (IAC) value was calculated. Intensity box plots and IAC scores were compared between all pre-processing approaches with different normalization and transformation methods. Then, the most suitable normalization and transformation methods for our data were selected, which show equal intensity box levels for all samples and a high IAC score for the normalized and transformed data.

The resulted plots showed quantile normalization is the most appropriate normalization method either combined with variance stabilizing transformation or Log_2 transformation. However, Log_2 transformed data has higher IAC score (IAC= 0.967) than VS transformed data (IAC= 0.965). Although there are other transformation methods have equal IAC value that is resulted by Log_2 transformation, however, unlike the transformed data by log_2, the boxplots showed that the data are variable and not suitable for ideal comparison after transforming the data with these methods.

Accordingly, our gene expression data was pre-processed by quantile normalization and Log_2 transformation based on this evidence. The data after pre-processing by the selected methods showed an improved intensity distribution and appeared more comparable.

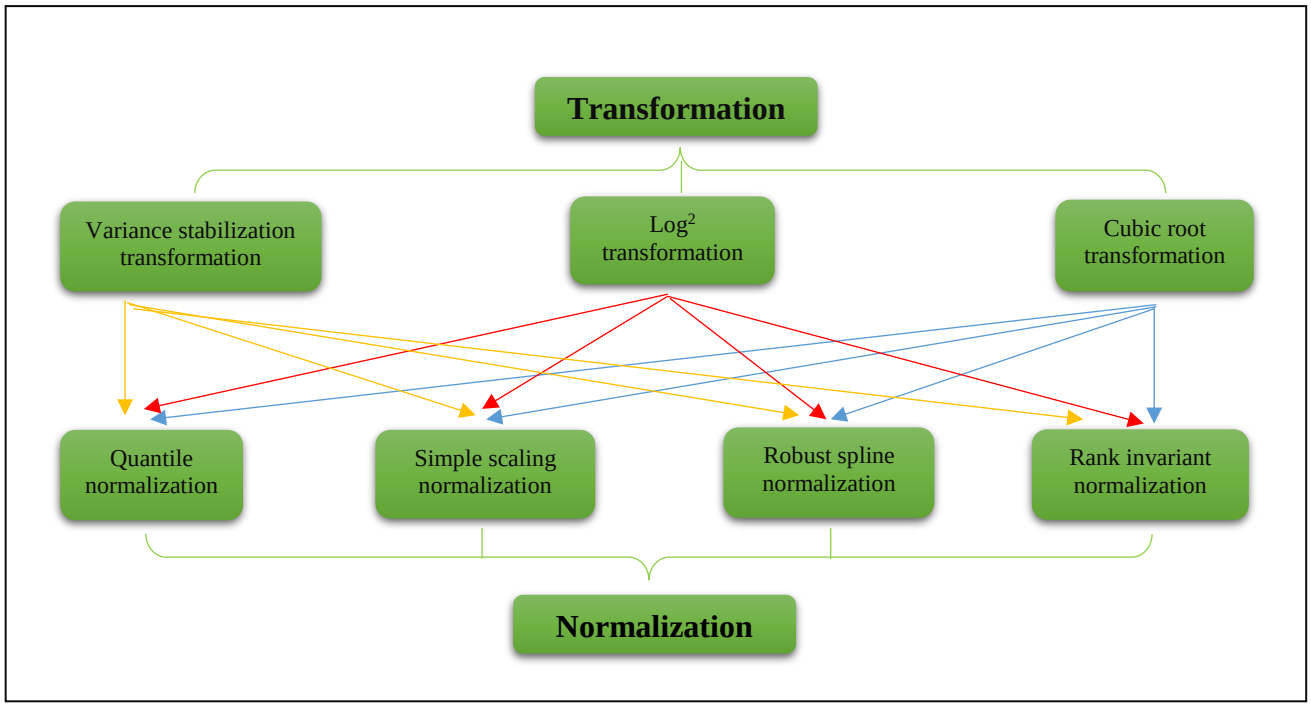


Figure 9- 3: **Schismatic flowchart for combined normalization and transformation methods applied on gene expression data to determine the best approaches to our data.**

Each transformation approach was applied with different normalization methods.

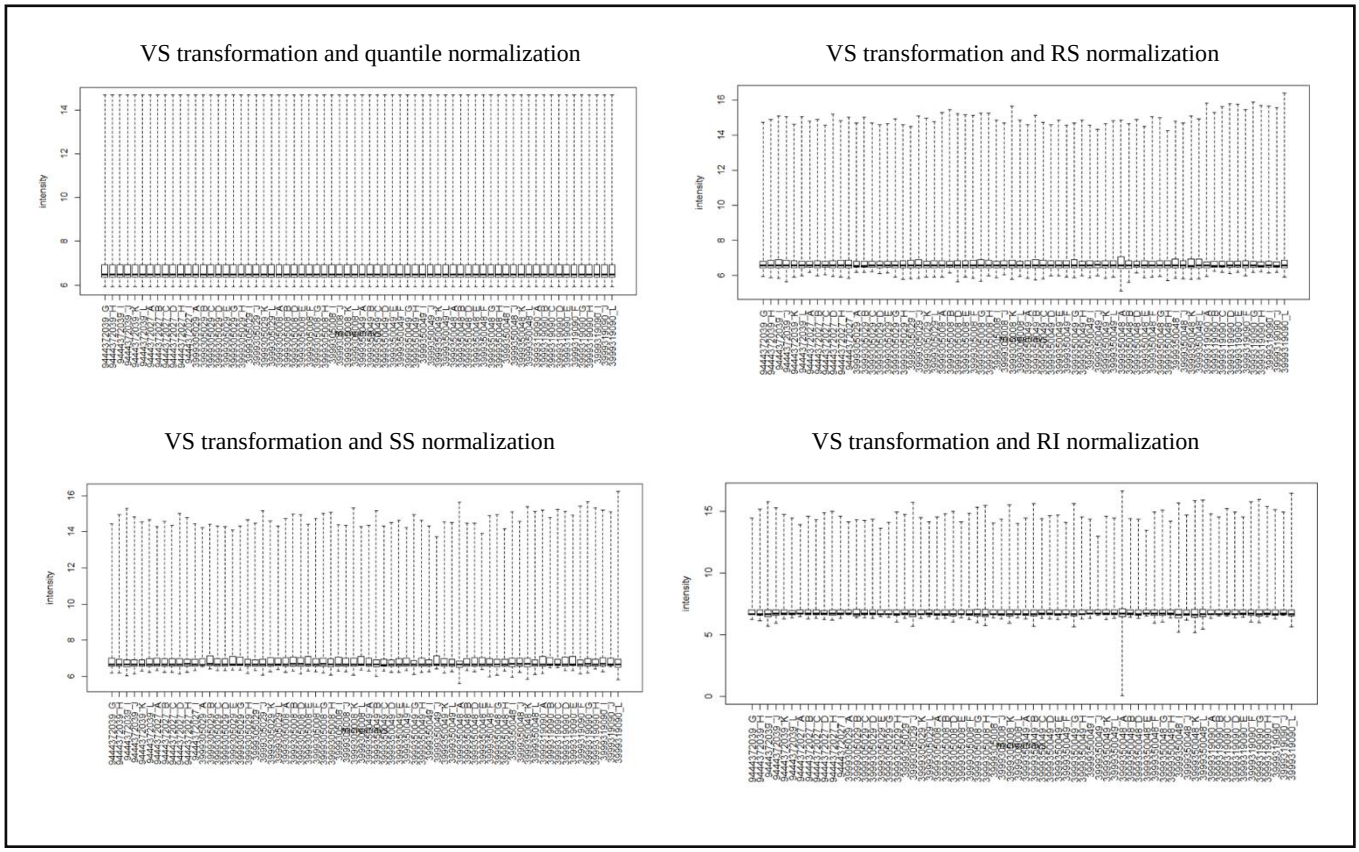


Figure 9- 4: **Intensity boxplots for regular probes after Variance Stabilizing Transformation (VST) combined with different methods of normalization.**

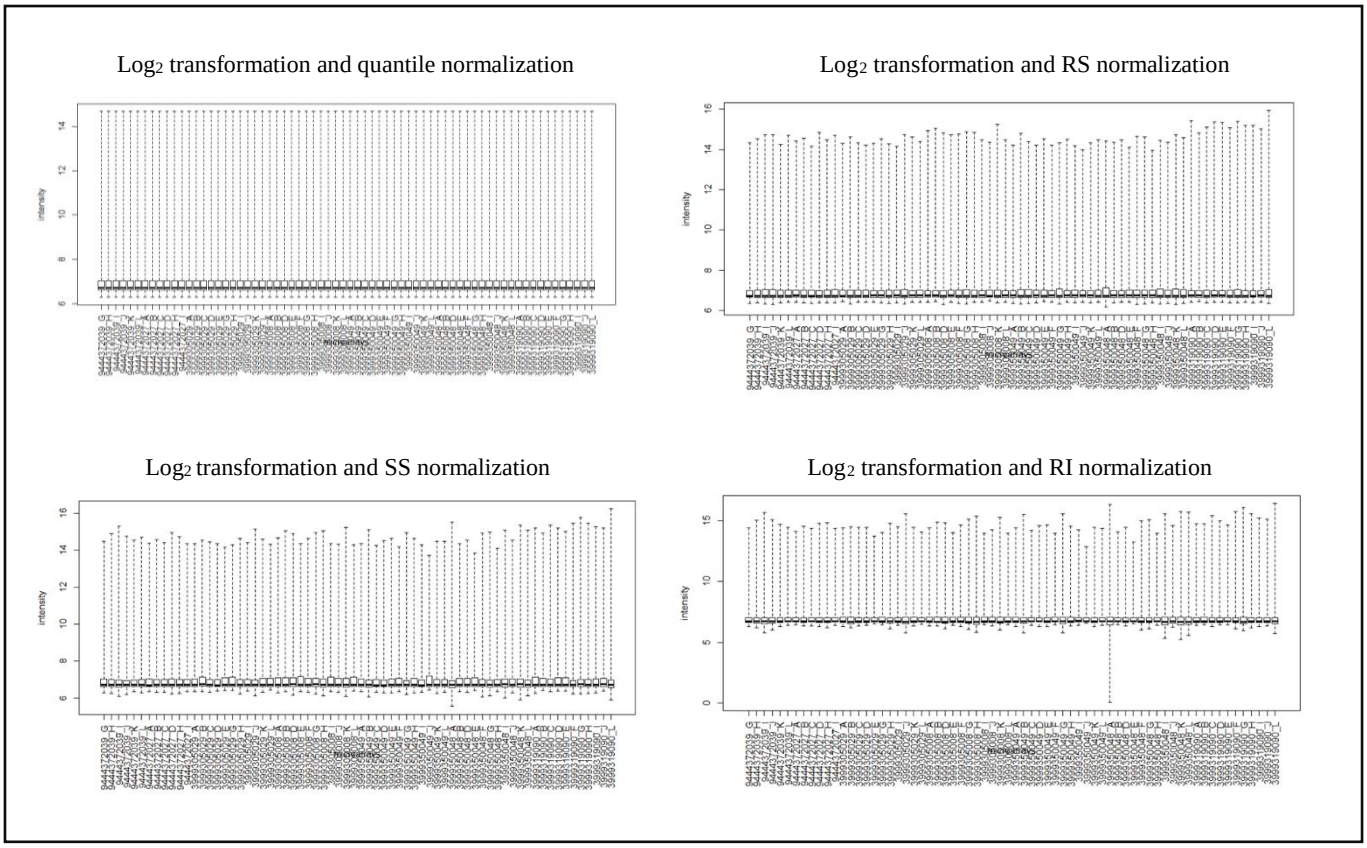


Figure 9- 5: **Intensity boxplots for regular probes after Log_2 transformation combined with different methods of normalization.**

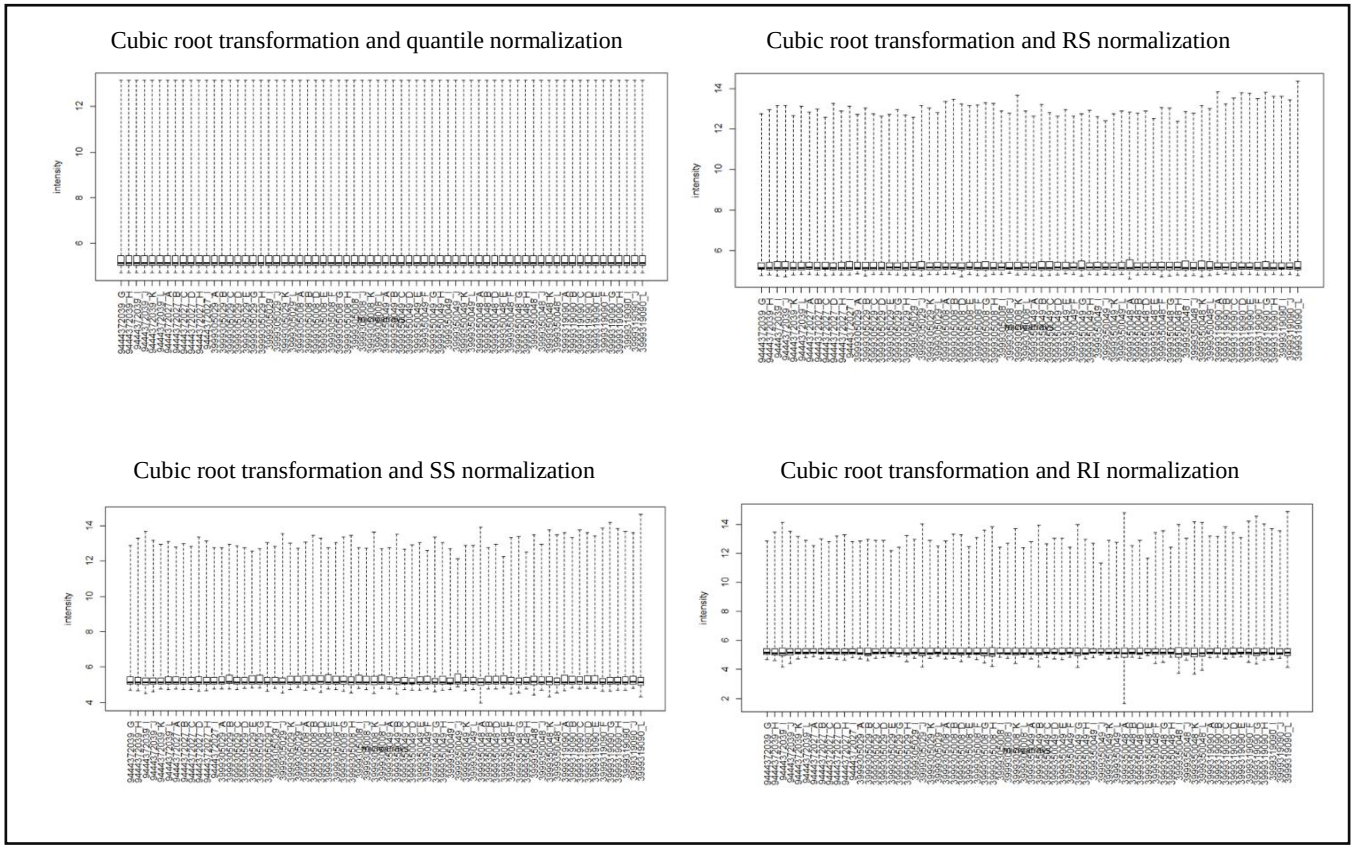


Figure 9- 6: **Intensity boxplots for regular probes after cubic root transformation combined with different methods of normalization.**

Table 9- 2: **Mean IAC values for gene expression data following normalization and transformation.**

		Normalization			
Transformation		Quantile	SSN	RSN	RI
	VS	0.965	0.964	0.964	0.965
	Log_2	0.967	0.965	0.967	0.965
	Cubic Root	0.95	0.939	0.942	0.967

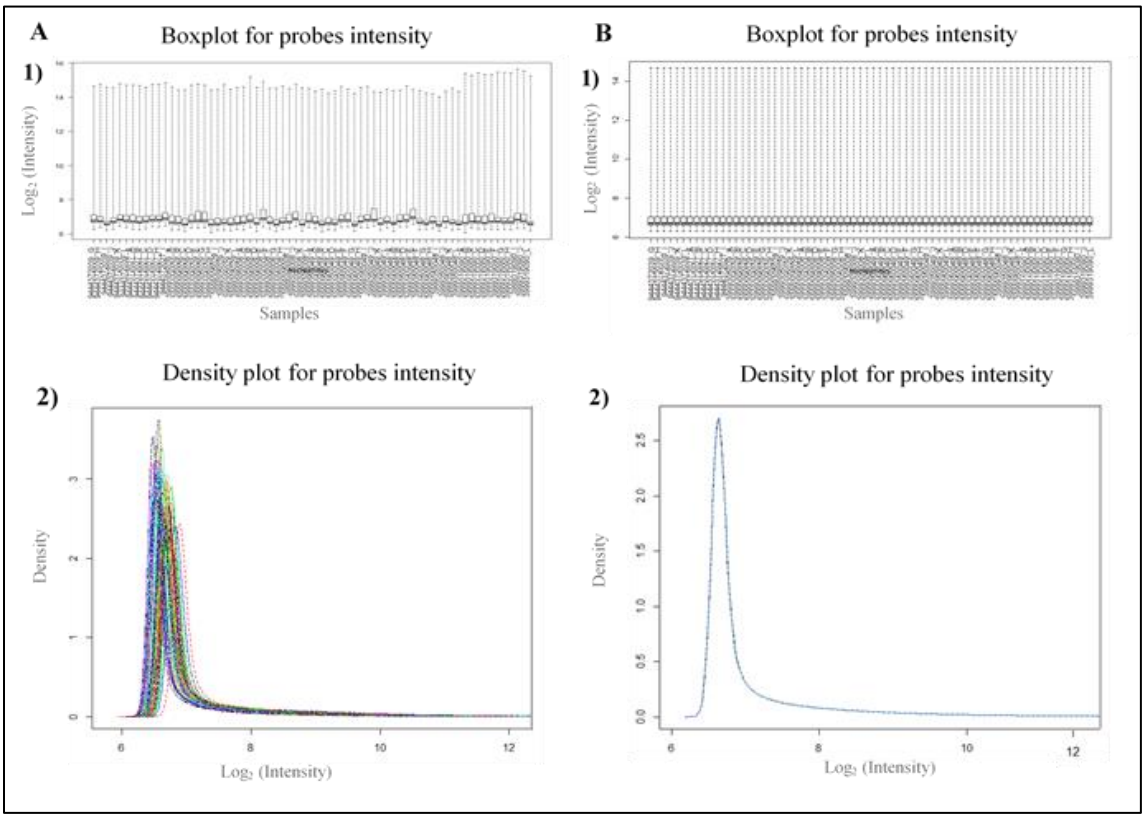


Figure 9- 7: **Plots for probes intensity before and after background correction, quantile normalization, and log_2 transformation.**

A. Intensity plots before background correction, quantile normalization, and log_2 transformation.
B. Intensity plots after background correction, quantile normalization, log_2 transformation.
1) Intensity boxplots for regular probes. Each box represents intensity signals range and average for each sample. Normalized and transformed data showed more comparable intensity levels among samples than unprocessed data.
2) Density curves illustrating the distribution of signal intensities in different samples. Each coloured line represents a sample. The density curves for all samples become identical after normalization and transformation.

9.1.3. Batch effects

9.1.3.1. Batch effects identification

Batch effects were identified by using dendrogram visualization approach on pre-processed data. Dendrogram cluster categorized the samples based on highly variable probes among all samples (Dunning et al., 2016).

The number of samples in the largest two clusters was determined based on their gender, phenotype status, origin, wave and batch. Thus, variances that play a significant role in sample clustering can be identified by applying categorical statistical analysis, such as a Fisher's exact test.

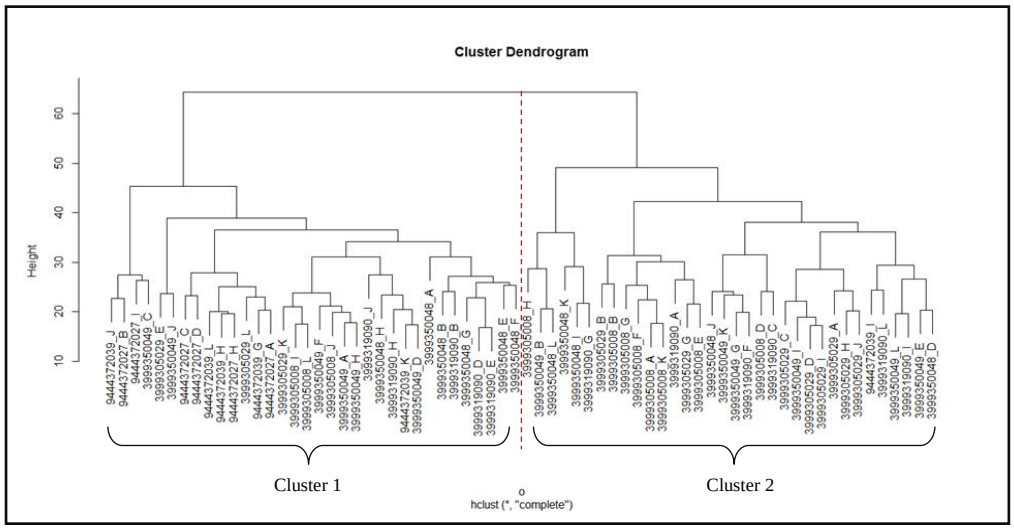


Figure 9- 8: **Dendrogram after background correction, normalization, and transformation.**

Dendrogram clusters were categorized based on highly variable probes among samples by calculating IQR for each probe among all samples. Probes then were ordered based on IQR values to identify the 500 most variable probes. These probes then were used to categorize the samples into clusters. Dendrogram identifies two main clusters.

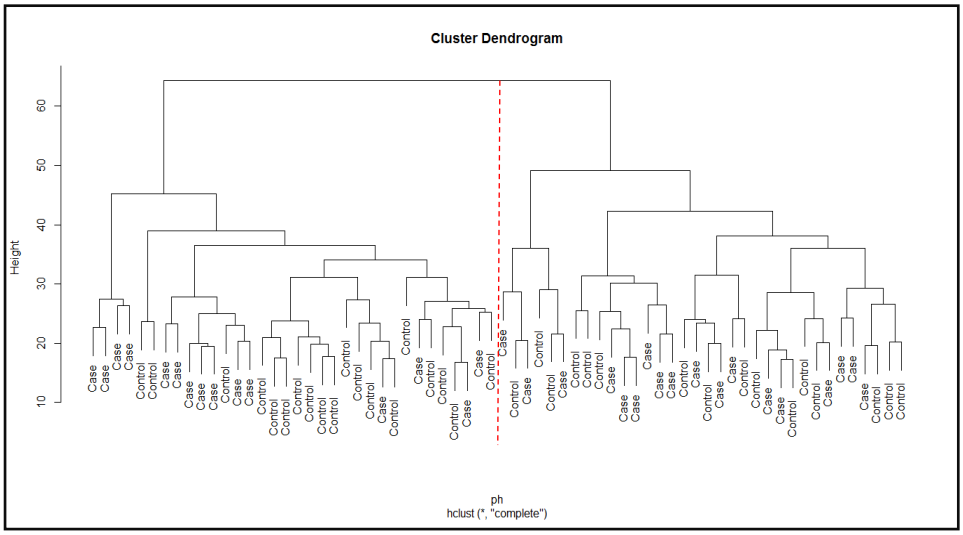


Figure 9- 9: **Dendrogram after background correction, normalization, and transformation.**

Samples are shown based on their phenotype status in both clusters.

Table 9- 3: **Results of Fisher's exact test based on number of cases and controls in the dendrogram two clusters.**

Sample phenotype	Case	Control	Total
Cluster 1	15	19	34
Cluster 2	18	16	34
Total	33	35	68
P-value	0.314		

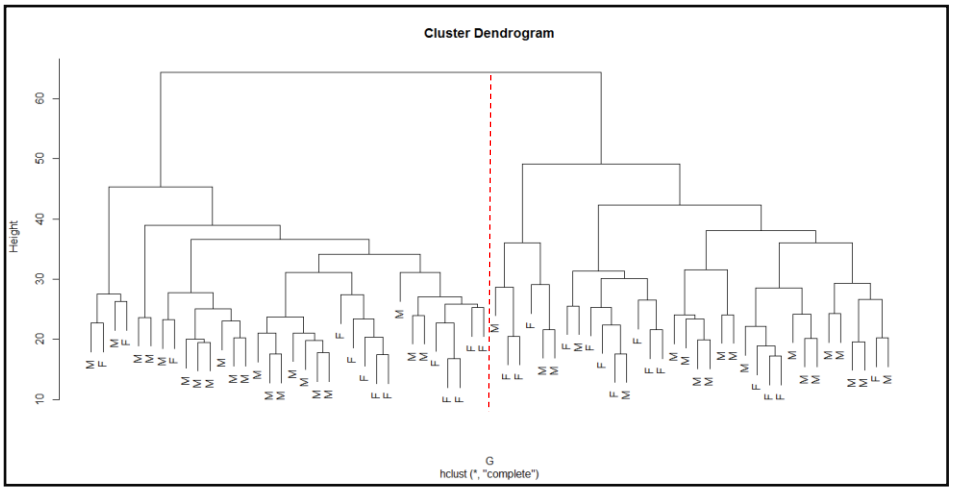


Figure 9- 10: **Dendrogram after background correction, normalization, and transformation.**

Samples are shown based on their gender in both clusters. M= male, F= female.

Table 9- 4: **Results of Fisher's exact test based on number of males and females in the dendrogram two clusters.**

Sample gender	Male	Female	Total
Cluster 1	21	13	34
Cluster 2	14	20	34
Total	35	33	68
Fisher's exact test P-value	0.0725		

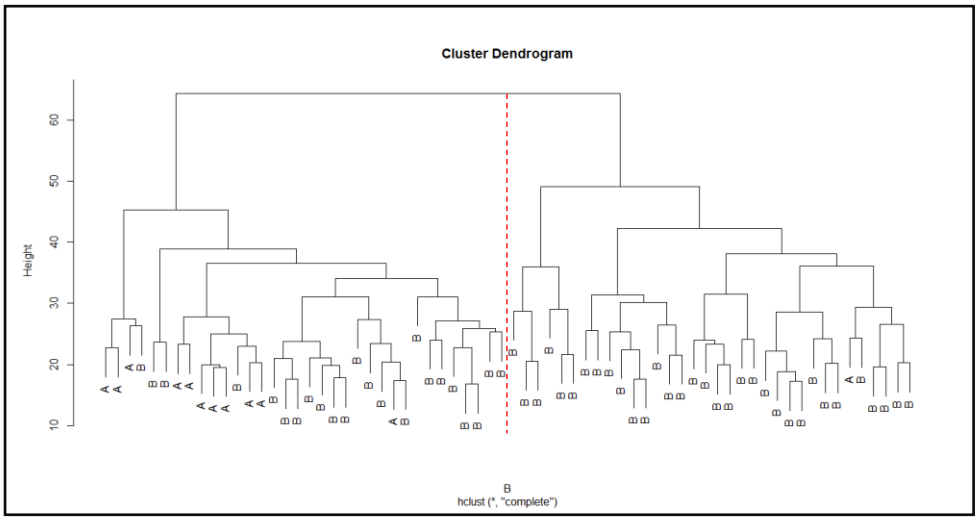


Figure 9- 11: **Dendrogram after background correction, normalization, and transformation.**

Samples are shown based on their wave in both clusters. A= wave A, B= wave B.

Table 9- 5: **Results of Fisher's exact test based on number of samples form wave A and wave B in the dendrogram two clusters.**

Sample wave	Wave A	Wave B	Total
Cluster 1	11	23	34
Cluster 2	1	33	34
Total	12	56	68
Fisher's exact test P-value	0.0041		

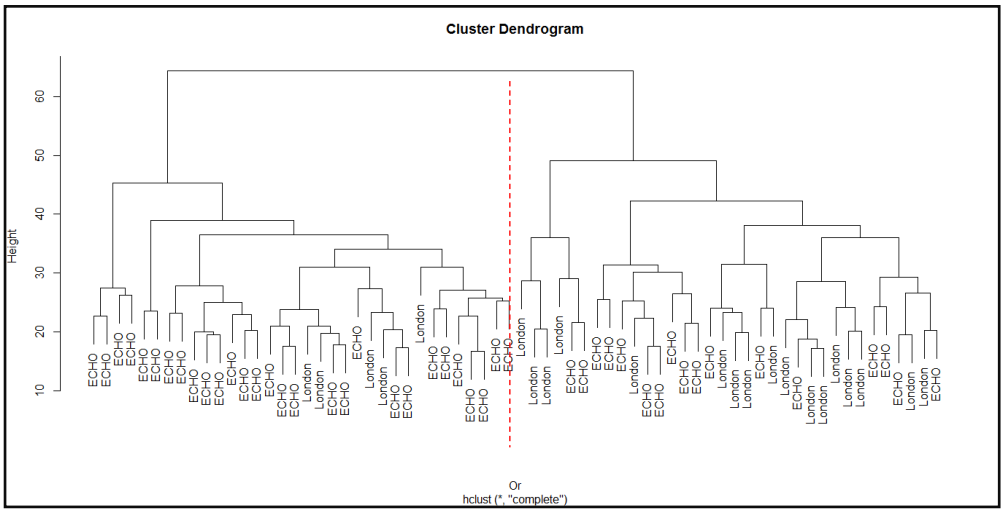


Figure 9- 12: **Dendrogram after background correction, normalization, and transformation.**

Samples are shown based on their origin in both clusters. ECHO= Cardiff ECHO study, London= London BBAG study.

Table 9- 6: **Results of Fisher's exact test based on number of samples form Cardiff ECHO study and London BBAG study in the dendrogram two clusters.**

Sample origin	London BBAG	Cardiff ECHO	Total
Cluster 1	5	29	34
Cluster 2	17	17	34
Total	22	46	68
Fisher's exact test P-value	0.0019		

The results revealed that there are two covariates are significant, sample wave (p-value= 0.0041) and sample origin (p-value = 0.0019). This finding is logic as the two waves were processed in two different seasons, samples of wave A were hybridized in April 2014 while those of wave B were processed in February 2015. Difference in the time may have an influence on sample hybridization and/or gene expression pattern potentially due to atmosphere temperature. Moreover, although samples from Cardiff research group and London research group were collected using the same

PAXgene approach, samples collecting, transferring, and storing procedures might differ in both groups and cause a specific effect.

9.1.3.2. Batch effects correction

Sample wave and origin were considered as batch effects. Adjusting batch effects was performed by using ComBat package on R (Johnson et al. 2007).

Then, expression data corrected for wave and origin was re-analysed by using the approach of dendrogram clustering and Fisher's exact test to confirm that all batch effects were corrected. The results showed there is no significant batch effect, which means both samples wave and origin were successfully adjusted. Therefore, the gene expression data is ready for differential gene expression analysis.

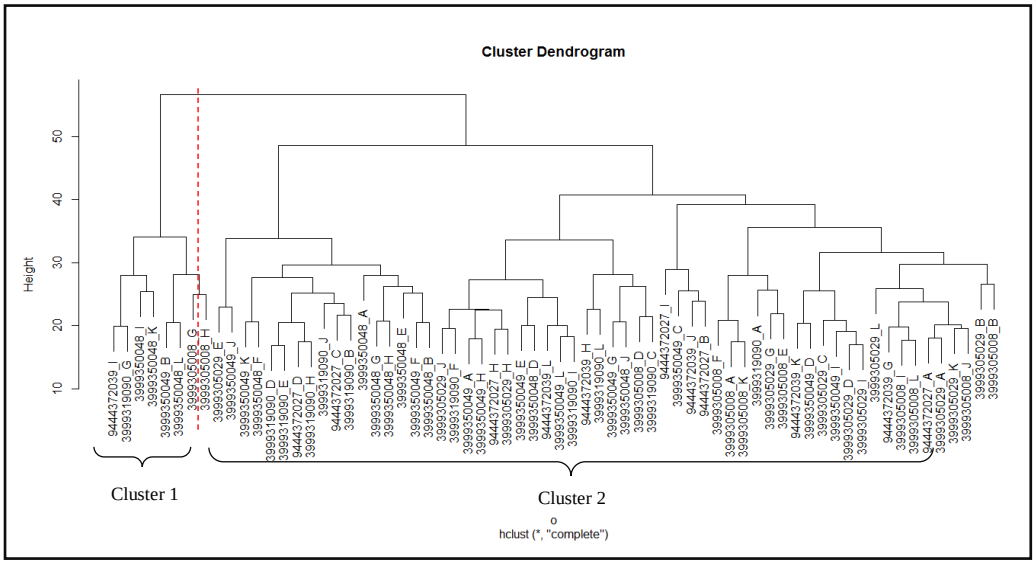


Figure 9- 13: **Dendrogram after background correction, normalization, transformation, and batch effect correction.**

Dendrogram identifies two clusters.

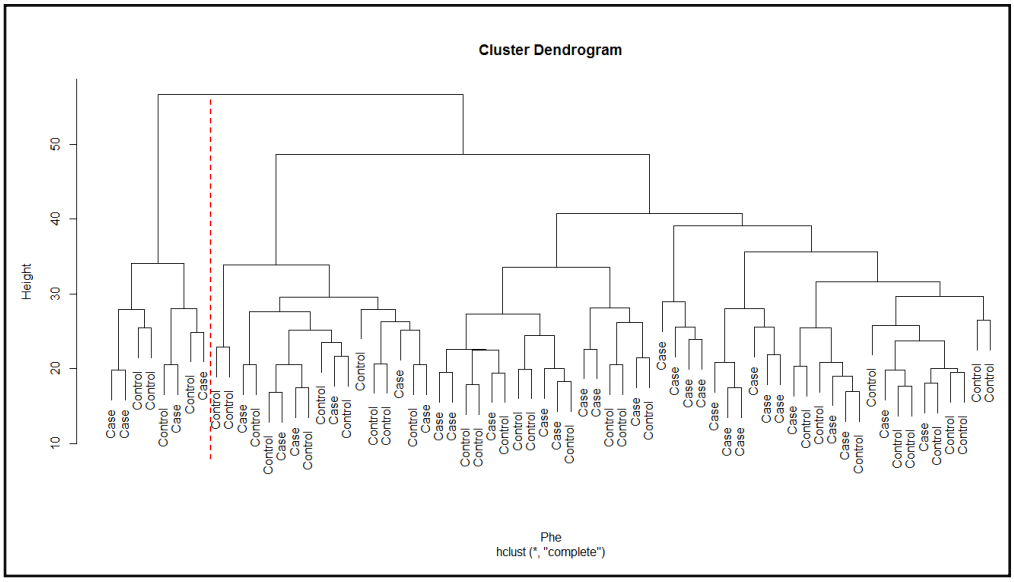


Figure 9- 14: **Dendrogram after background correction, normalization, transformation, and batch effect correction.**

Samples are shown based on their phenotype status in both clusters.

Table 9- 7: **Results of Fisher's exact test based on number of cases and controls in the dendrogram two clusters.**

Sample phenotype	Case	Control	Total
Cluster 1	4	4	8
Cluster 2	29	31	60
Total	33	35	68
Fisher's exact test P-value	0.6112		

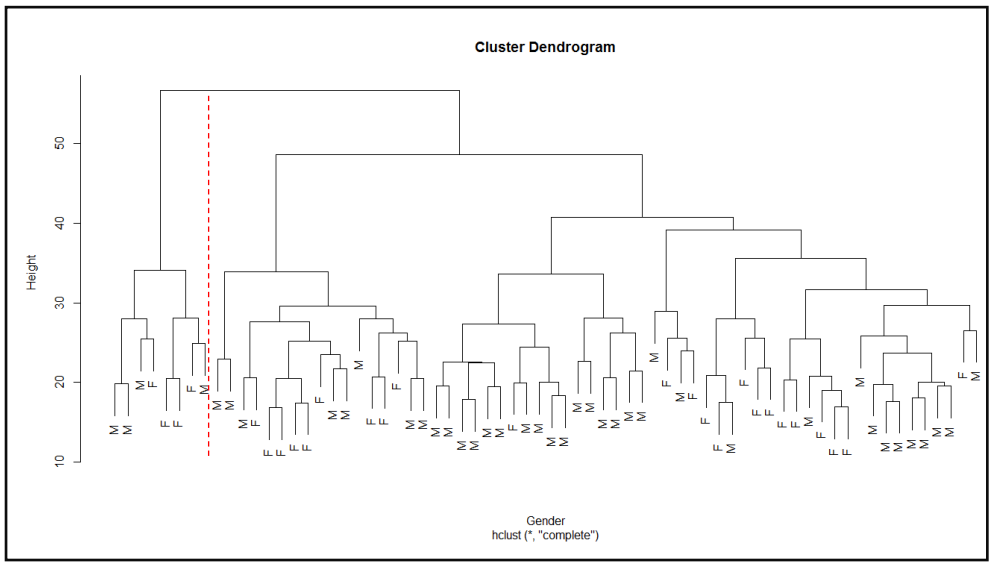


Figure 9- 15: **Dendrogram after background correction, normalization, transformation, and batch effect correction.**

Samples are shown based on their gender in both clusters. M= male and F= female.

Table 9- 8: **Results of Fisher's exact test based on number of males and females in the dendrogram two clusters.**

Sample gender	Male	Female	Total
Cluster 1	4	4	8
Cluster 2	23	37	60
Total	27	41	68
Fisher's exact test P-value	0.395		

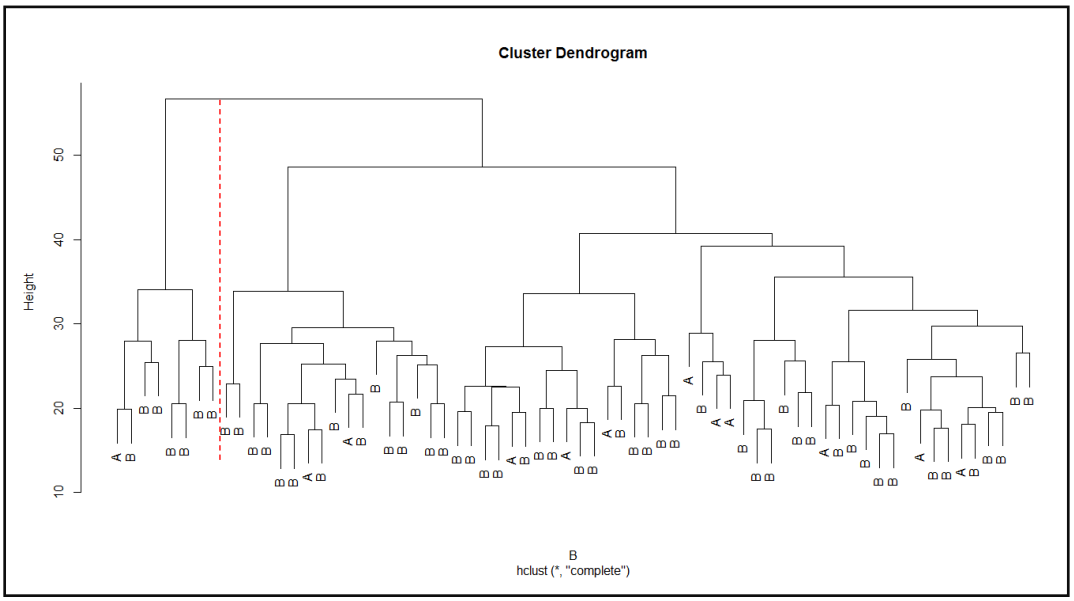


Figure 9- 16: **Dendrogram after background correction, normalization, transformation, and batch effect correction.**

Samples are shown based on their wave in both clusters. A= from wave A and B= from wave B.

Table 9- 9: **Results of Fisher's exact test based on number of samples from wave A and wave B in the dendrogram two clusters.**

Sample wave	Wave A	Wave B	Total
Cluster 1	1	7	8
Cluster 2	11	49	60
Total	12	56	68
Fisher's exact test P-value	0.569		

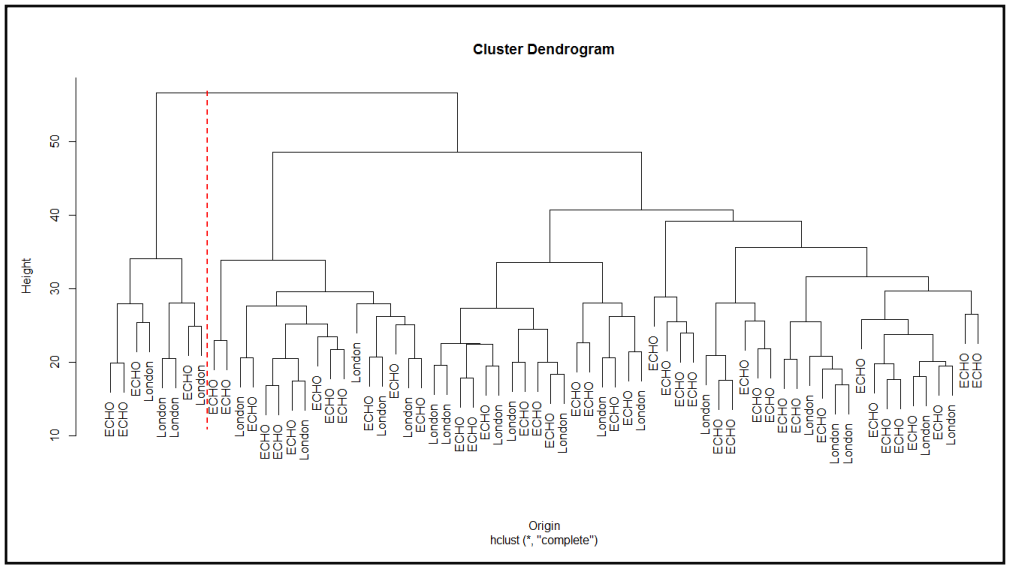


Figure 9- 17: **Dendrogram after background correction, normalization, transformation, and batch effect correction.**

Samples are shown based on their origin in both clusters. ECHO= from Cardiff ECHO study and London= from London BBAG study.

Table 9- 10: **Results of Fisher's exact test based on number of samples from London BBAG study and Cardiff ECHO study in the dendrogram two clusters.**

Sample origin	London BBAG	Cardiff ECHO	Total
Cluster 1	4	4	8
Cluster 2	18	42	60
Total	22	46	68
Fisher's exact test P-value	0.2271		

References

Alberts, B. et al., 2002. Meiosis. Molecular Biology of the Cell. 4th edition. New York: Garland Science Publishing, pp.846-854.

Amati, F. et al., 2007. Dynamic changes in gene expression profiles of 22q11 and related orthologous genes during mouse development. Gene, 391, pp.91–102.

Angold, A. et al., 1995. The Child and Adolescent Psychiatric Assessment (CAPA). Psychological Medicine, 25(4), pp.739-53.

Anney, R. et al., 2010. A genome-wide scan for common alleles affecting risk for autism. Human Molecular Genetics, 19(20), pp.4072–4082.

Antshel, K. et al., 2006. ADHD, Major Depressive Disorder, and Simple Phobias are prevalent psychiatric conditions in youth with Velocardiofacial Syndrome. Journal of the American Academy of Child & Adolescent Psychiatry, 45(5), pp.596–603.

Antshel, K. et al., 2010. Cognitive and psychiatric predictors to psychosis in Velocardiofacial Syndrome: A 3-year follow-up study. Journal of the American Academy of Child and Adolescent Psychiatry, 49(4), pp.333–344.

Babcock, M. et al., 2003. Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution. Genome Research, 13, pp.2519–2532.

Bailey, J. et al., 2002. Recent segmental duplications in the human genome. Science, 297, pp.1003–1007.

Baker, K. et al., 2005. COMT Val108/158Met modifies mismatch negativity and cognitive function in 22q11 deletion syndrome. Biological Psychiatry, 58, pp.23–31.

Ballif, B. et al., 2007. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. Nature Genetics, 39(9), pp.1071–1073.

Bassett, A. et al., 2005. Clinical features of 78 adults with 22q11 deletion syndrome. American Journal of Medical Genetics Part A, 138A, pp.307–313.

Bassett, A. et al., 2007. Catechol-o-methyl transferase and expression of schizophrenia in 73 adults with 22q11 Deletion Syndrome. Biological Psychiatry, 61, pp.1135–1140.

Bassett, A. & Chow, E., 2008. Schizophrenia and 22q11.2 deletion syndrome. Current Psychiatry Reports, 10(2), pp.148–157.

Bassett, A. et al., 2008. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. Human Molecular Genetics, 17(24), pp.4045–4053.

Bassett, A. et al., 2011. Practical guidelines for managing patients with 22q11.2 deletion syndrome. Journal of Pediatrics, 159(2), p.332–339.e1.

Bearden, C. et al., 2004. Effects of a functional COMT polymorphism on prefrontal cognitive function in patients with 22q11.2 deletion syndrome. American Journal of Psychiatry, 161(9), pp.1700–1702.

Benajmini, Y. & Hochberg, Y., 1995. Controlling the false discovery rate: A Practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), pp.289–300.

Bender, H. et al., 2005. Functional consequences of PRODH missense mutations. American Journal of Human Genetics, 76, pp.409–420.

Berument, S. et al., 1999. Autism screening questionnaire: diagnostic validity. British Journal of Psychiatry, 175, pp.431-444.

Biesecker, L., 2010. Exome sequencing makes medical genomics a reality. Nature Genetics, 42(1), pp.13–14.

Bonifati, V. et al., 2003. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science, 299, pp.256–259.

Booij, J., van Amelsvoort, T. & Boot, E., 2010. Co-occurrence of early-onset Parkinson disease and 22q11.2 deletion syndrome: Potential role for dopamine transporter imaging. American Journal of Medical Genetics Part A, 152A, pp.2937–2938.

Botto, L. et al., 2003. A Population-Based Study of the 22q11.2 Deletion: Phenotype, Incidence, and Contribution to Major Birth Defects in the Population. Pediatrics, 112(1), pp.101–107.

Brauner, R. et al., 2003. Parathyroid function and growth in 22q11.2 deletion syndrome. Journal of Pediatrics, 142(5), pp.504–508.

Brzustowicz, L. & Bassett, A., 2012. miRNA-mediated risk for schizophrenia in 22q11.2 deletion syndrome. Frontiers in Genetics, 3(291), pp.1–9.

Budarf, M. et al., 1995. Identification of a patient with bernard-soulier syndrome and a deletion in the digeorge/velo-cardio-facial chromosomal region in 22q11.2. Human Molecular Genetics, 4(4), pp.763–766.

Butcher, N. et al., 2013. Association Between Early-Onset Parkinson Disease and 22q11.2 Deletion Syndrome. JAMA Neurology, 70(11), pp.1359–1366.

Caldú, X. et al., 2007. Impact of the COMT Val108/158 Met and DAT genotypes on prefrontal function in healthy subjects. NeuroImage, 37(4), pp.1437–1444.

Cardon, L. & Palmer, L., 2003. Population stratification and spurious allelic association. Lancet, 361, pp.598–604.

Carlson, C. et al., 1997. Molecular definition of 22q11 deletions in 151 Velo-Cardio-Facial Syndrome Patients. American Journal of Human Genetics, 61, pp.620–629.

Cayler, G., 1969. Cardiofacial syndrome. Congenital heart disease and facial weakness, a hitherto unrecognized association. Archives of Disease in Childhood, 44(233), pp.69–75.

Champely, S. 2016. Basic Functions for Power Analysis [R package pwr version 1.2-0]. [online] Cran.r-project.org. Available at: https://CRAN.R-project.org/package=pwr [Accessed 26 Jan. 2017].

Charman, T. & Baird, G. 2002. Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. Journal of Child Psychology and Psychiatry, 43(3), pp.289-305.

Chen, C. et al., 2011. Association study of catechol-O-methyltransferase gene polymorphisms with schizophrenia and psychopathological symptoms in Han Chinese. Genes, Brain and Behavior, 10, pp.316–324.

Chen, H. et al., 2005. Mutation analysis of ARVCF gene on chromosome 22q11 as a candidate for a schizophrenia gene. Schizophrenia Research, 72, pp.275–277.

Chen, J. et al., 2004. Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. American Journal of Human Genetics, 75, pp.807–821.

Chen, J. et al., 2011. Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: Implications for drug development. Journal of Biological Chemistry, 286(40), pp.34752–34760.

Chow, E. et al., 2006. Neurocognitive profile in 22q11 deletion syndrome and schizophrenia. Schizophrenia Research, 87(1–3), pp.270–278.

Colnaghi, R. et al., 2011. The consequences of structural genomic alterations in humans: Genomic Disorders, genomic instability and cancer. Seminars in Cell and Developmental Biology, 22, pp.875–885.

Cook, E. & Scherer, S., 2008. Copy-number variations associated with neuropsychiatric conditions. Nature, 455, pp.919–923.

Cookson, M., 2010. The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease. Nature Reviews Neuroscience, 11(12), pp.791–797.

Cooper, G. et al., 2012. A Copy number variation morbidity map of developmental delay. Nature Genetics, 43(9), pp.838–846.

Costain, G. et al., 2014. Adult neuropsychiatric expression and familial segregation of 2q13 duplications. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 0(4), pp.337–344.

Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Identification of risk loci with shared eff ects on fi ve major psychiatric disorders: A genome-wide analysis. Lancet, 381, pp.1371–1379.

de la Chapelle, A., 1981. A deletion in chromosome 22 can cause DiGeorge syndrome. Human Genetics, 57(3), pp.253-256.

de Lau, L. & Breteler, M., 2006. Epidemiology of parkinson's disease. Lancet Neurology, 5, pp.525–535.

De Luca, A. et al., 2001. Association study of a promoter polymorphism of UFD1L gene with schizophrenia. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 105, pp.529–533.

Delaneau, O., Marchini, J. & The 1000 Genomes Project Consortium, 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nature Communications, 5(3934), pp.1–9.

Demily, C. et al., 2015. Neurocognitive and psychiatric management of the 22q11.2 deletion syndrome (in French). Le Encephale, 41(3), pp.266-273.

DePiero, A. et al., 1997. Recurrent immune cytopenias in two patients with DiGeorge/velocardiofacial syndrome. Journal of Pediatrics, 131(3), pp.484–486.

Devriendt, K., Fryns, J. & Mortier, G., 1998. The annual incidence of DiGeorge/velocardiofacial syndrome. Journal of Medical Genetics, 35(9), pp.789–790.

Dickson, D. et al., 2009. Neuropathological assessment of Parkinson's disease: refining the diagnostic criteria. Lancet Neurology, 8, pp.1150–1157.

DiGeorge, A., 1965. In discussion, Cooper, M., Peterson, R. & Good, R., A new concept of the cellular basis of immunity. Journal of Paediatrics, 67, pp.907-908.

Digilio, M. et al., 2003. Spectrum of clinical variability in familial deletion 22q11.2: from full manifestation to extremely mild clinical anomalies. Clinical Genetics, 63, pp.308–313.

Drazic, A. et al., 2016. The world of protein acetylation. Biochimica et Biophysica Acta, 1864, pp.1372–1401.

Driscoll, D. et al., 1992. Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. American Journal of Medical Genetics, 44, pp.261–268.

Driscoll, D. et al., 1993. Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: Implications for genetic counselling and prenatal diagnosis. Journal of Medical Genetics, 30, pp.813–817.

Du, P., Kibbe, W. & Lin, S., 2008. lumi: A pipeline for processing Illumina microarray. Bioinformatics, 24(13), pp.1547–1548.

Dunn, O., 1961. Multiple comparisons among means. Journal of the American Statistical Association, 56(293), pp.52–64.

Dunning, M., Shi, W., Lynch, A., Smith, M. and Ritchie, M. (2016). BeadArray expression analysis using Bioconductor. [PDF] p.28. Available at: https://www.bioconductor.org/packages/devel/data/experiment/vignettes/BeadArrayUseCases/inst/doc/BeadArrayUseCases.pdf [Accessed 26 Jan. 2017].

Dupont, W. & Plummer, W., 1990. Power and sample size calculations. A review and computer program. Controlled Clinical Trials, 11(2), pp.116–128.

Dweep, H. et al., 2011. MiRWalk - Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of Biomedical Informatics, 44(5), pp.839–847.

Edelmann, L., Pandita, R. & Morrow, B., 1999. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. American Journal of Human Genetics, 64, pp.1076–1086.

Egan, M. et al., 2001. Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 98(12), pp.6917–22.

Elia, J. et al., 2010. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Molecular Psychiatry, 15, pp.637–646.

Elia, J. et al., 2011. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. Nature Genetics, 44(1), pp.78–84.

Elmore, S., 2007. Apoptosis: A review of programmed cell death. Toxicol Pathology, 35(4), pp.495–516.

Fan, J. et al., 2003. A family-based association study of T1945C polymorphism in the proline dehydrogenase gene and schizophrenia in the Chinese population. Neuroscience Letters, 338, pp.252–254.

Fan, J. et al., 2005. Catechol-O-methyltransferase gene Val/Met functional polymorphism and risk of schizophrenia: A large-scale association study plus meta-analysis. Biological Psychiatry, 57, pp.139–144.

Fan, J. et al., 2006. Illumina universal bead arrays. Methods in Enzymology, 410, pp.57–73.

Fry, A. et al., 2016. Pathogenic copy number variants and SCN1A mutations in patients with intellectual disability and childhood-onset epilepsy. BMC medical genetics, 17(34), pp.1–9.

Fung, W. et al., 2010. Elevated prevalence of generalized anxiety disorder in adults with 22q11.2 deletion syndrome. American Journal of Psychiatry, 167(8), p.998.

Funke, B. et al., 2007. Analysis of variation in patients with psychotic and affective disorders. Molecular Medicine, 13(7–8), pp.407–413.

Gaig, C. & Tolosa, E., 2009. When does Parkinson's disease begin? Movement Disorders, 24(Suppl. 2), pp.656–664.

Gene Expression Microarray Data Quality Control. (n.d.). [PDF] pp.1-2. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_gene_expression_data_quality_control.pdf [Accessed 26 Jan. 2017].

Genome.ucsc.edu. (n.d.). UCSC Genome Browser Gateway. [online] Available at: https://genome.ucsc.edu/cgi-bin/hgGateway [Accessed 26 Jan. 2017].

Georgieva, L. et al., 2014. De novo CNVs in bipolar affective disorder and schizophrenia. Human Molecular Genetics, 23(24), pp.6677–6683.

Gerdes, M. et al., 2001. Taking advantage of early diagnosis: Preschool children with the 22q11.2 deletion. Genetics in Medicine, 3(1), pp.40–44.

Girirajan, S. et al., 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nature Genetics, 42(3), pp.203–209.

Girirajan, S. et al., 2012. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. New England Journal of Medicine, 367(14), pp.1321–1331.

Glaser, B., Debbane, M., et al., 2006. No evidence for an effect of COMT Val158Met genotype on executive function in patients with 22q11 deletion syndrome. American Journal of Psychiatry, 163, pp.537–539.

Glaser, B., Moskvina, V., et al., 2006. Analysis of ProDH, COMT and ZDHHC8 risk variants does not support individual or interactive effects on schizophrenia susceptibility. Schizophrenia Research, 87, pp.21–27.

Glatt, S., Faraone, S. & Tsuang, M., 2003. Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: Meta-analysis of case-control and family-based studies. American Journal of Psychiatry, 160, pp.469–476.

Gogos, A., van den Buuse, M. & Rossell, S., 2009. Gender differences in prepulse inhibition (PPI) in bipolar disorder: Men have reduced PPI, women have increased PPI. International Journal of Neuropsychopharmacology, 12, pp.1249–59.

Gogos, J. et al., 1998. Catechol-o-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. Proceedings of the National Academy of Sciences, 95, pp.9991–9996.

Golding-Kuchner, K., 2005. Speech and language disorders in velo-cardio-facial syndrome. In Murphy, K., Scambler, P., Velo-Cardio-Facial Syndrome: A Model for Understanding Microdeletion Disorders. Cambridge: Cambridge University Press, pp. 181-200.

Goodship, J. et al., 1998. A population study of chromosome 22q11 deletions in infancy. Archives of Disease in Childhood, 79, pp.348–351.
Gothelf, D. & Lombroso, P., 2001. Genetics of childhood disorders: XXV. Velocardiofacial syndrome. Journal of the American Academy of Child and Adolescent Psychiatry, 40(4), pp.489–491.
Gothelf, D. et al., 2005. COMT genotype predicts longitudinal cognitive decline and psychosis in 22q11.2 deletion syndrome. Nature Neuroscience, 8(11), pp.1500–1502.
Guna, A., Butcher, N. & Bassett, A., 2015. Comparative mapping of the 22q11.2 deletion region and the potential of simple model organisms. Journal of Neurodevelopmental Disorders, 7(18), pp.1–6.
Hannon, E. et al., 2016. An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and differential DNA methylation. Genome Biology, 17(176), pp.1–16.
Heman-Ackah, S. et al., 2013. RISC in PD: The impact of microRNAs in Parkinson's disease cellular and molecular pathogenesis. Frontiers in Molecular Neuroscience, 6(40), pp.1–17.
Hernández-Nieto, L. et al., 2011. Autoimmune thrombocytopenic purpura in partial digeorge syndrome: Case presentation. Journal of Pediatric Hematology/Oncology, 33(6), pp.465–466.
Hernando, C. et al., 2002. Comparative genomic hybridisation shows a partial de novo deletion 16p11.2 in a neonate with multiple congenital malformations. Journal of Medical Genetics, 39, p.E24.
Hinney, A. et al., 2011. Genome-wide association study in German patients with attention deficit/hyperactivity disorder. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 156B, pp.888–897.
Hochstenbach, R. et al., 2012. Discovery of variants unmasked by hemizygous deletions. European Journal of Human Genetics, 20, pp.748–753.
Houlden, H. & Singleton, A., 2012. The genetics and neuropathology of Parkinson's disease. Acta Neuropathologica, 124, pp.325–338.
Howie, B., Donnelly, P. & Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics, 5(6), p.e1000529.
Hsu, R. et al., 2007. Nogo Receptor 1 (RTN4R) as a candidate gene for schizophrenia: Analysis using human and mouse genetic approaches. PloS One, (11), p.e1234.
Huang, D. et al., 2007. DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Research, 35, pp.W169–W175.
Huang, D., Sherman, B. & Lempicki, R., 2009. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research, 37(1), pp.1–13.
Hurd, P. & Nelson, C., 2009. Advantages of next-generation sequencing versus the microarray in epigenetic research. Briefings in Functional Genomics and Proteomics, 8(3), pp.174–183.
Ikeda, M. et al., 2010. Failure to confirm association between PIK4CA and psychosis in 22q11.2 deletion syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 153, pp.980–982.
Illumina® TotalPrep™ RNA Amplification kit manual. (n.d.). [PDF] p.2. Available at: https://tools.thermofisher.com/content/sfs/manuals/IL1791ME.pdf [Accessed 26 Jan. 2017].
Inoue, K. & Lupski, J., 2002. Moclecular mechanisms fot gennomic disorders. Annual Review of Genomics and Human Genetics, 3, pp.199–242.
Insel, T., 2010. Rethinking schizophrenia. Nature, 468, pp.187–193.

IPDGC, 2011. Imputation of sequence variants for identifi cation of genetic risks for Parkinson's disease: A meta-analysis of genome-wide association studies. Lancet, 377, pp.641–649.
Jalali, G. et al., 2008. Detailed analysis of 22q11.2 with a high density MLPA probe set. Human Mutation, 29(3), pp.433–440.
Jalbrzikowski, M. et al., 2015. Transcriptome profiling of peripheral blood in 22q11.2 deletion syndrome reveals functional pathways related to psychosis and autism spectrum disorder. PLoS One, 10(7), p.e0132542.
Jankovic, J., 2008. Parkinson's disease: Clinical features and diagnosis. Journal of Neurology, Neurosurgery & Psychiatry, 79, pp.368–376.
Jarick, I. et al., 2014. Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. Molecular Psychiatry, 19, pp.115–121.
Johnson, W., Li, C. & Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8(1), pp.118–127.
Jonas, R., Montojo, C. & Bearden, C., 2014. The 22q11.2 Deletion Syndrome as a window into complex neuropsychiatric disorders over the lifespan. Biological Psychiatry, 75(5), pp.351–360.
Jungerius, B. et al., 2008. An association screen of myelin-related genes implicates the chromosome 22q11 PIK4CA gene in schizophrenia. Molecular Psychiatry, 13, pp.1060–1068.
Jurata, L. et al., 2006. Altered expression of hippocampal dentate granule neuron genes in a mouse model of human 22q11 deletion syndrome. Schizophrenia Research, 88, pp.251–259.
Kahvejian, A., Quackenbush, J. & Thompson, J., 2008. What would you do if you could sequence everything? Nature Biotechnology, 26(10), pp.1125–1133.
Kaminsky, E. et al., 2013. An evidence-based approach to establish the functional and clinical significance of CNVs in intellectual and developmental disabilities. Genetics in Medicine, 13(9), pp.777–784.
Kanahara, N., Iyo, M. & Hashimoto, K., 2009. Failure to confirm the association between the PIK4CA gene and schizophrenia in a japanese population. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150B, pp.450–452.
Karayiorgou, M. et al., 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proceedings of the National Academy of Sciences of the United States of America, 92, pp.7612–7616.
Karayiorgou, M., Simon, T. & Gogos, J., 2010. 22q11.2 microdeletions: Linking DNA structural variation to brain dysfunction and schizophrenia. Nature Reviews Neuroscience, 11, pp.402–416.
Kates, W. et al., 2006. A gender-moderated effect of a functional COMT polymorphism on prefrontal brain morphology and function in velo-cardio-facial syndrome (22q11.2 deletion syndrome). American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 141 B, pp.274–280.
Kelley, R., 1982. The association of the DiGeorge anomalad with partialmonosomy of chromosome 22. Journal of Pediatrics, 101, pp.197-200.
Kempf, L. et al., 2008. Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function. PLoS Genetics, 4(11), p.e1000252.
Khodosevich, K., Seeburg, P. & Monyer, H., 2009. Major signaling pathways in migrating neuroblasts. Frontiers in Molecular Neuroscience, 2(7), pp.1–18.
Kim, S., & Lord, C. (2012). New Autism Diagnostic Interview Revised algorithms for toddlers and young pre-schoolers from 12–47 months of age. Journal of Autism and Developmental Disorders, 42(1), pp.82–93.
Kimber, W. et al., 1999. Deletion of 150 Kb in the minimal DiGeorge/Velocardiofacial syndrome critical region in mouse. Human Molecular Genetics, 8(12), pp.2229–2237.

Kinouchi, A., 1976. Facial appearance of patients with conotruncal anomalies (in Japanese). Pediatrics Japan, 17, pp.84-89.

Kirov, G. et al., 2009. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Human Molecular Genetics, 18(8), pp.1497–1503.

Kirschner, R., 2005. Palatal anomalies and velopharyngeal dysfunction associated with Velo-cardio-facial syndrome. In Murphy, K., Scambler, P., Velo-Cardio-Facial Syndrome: A Model for Understanding Microdeletion Disorders. Cambridge: Cambridge University Press, pp. 83–104.

Kitada, T. et al., 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature, 392, pp.605–608.

Kitsiou-Tzeli, S., Kolialexi, A. & Mavrou, A., 2005. Endocrine manifestations in DiGeorge and other microdeletion syndromes related to 22q11.2. Hormones (Athens, Greece), 4(4), pp.200–209.

Klein, R., 2007. Power analysis for genome-wide association studies. BMC Genetics, 8(58), pp.1–8.

Kobrynski, L. & Sullivan, K., 2007. Velocardiofacial syndrome, DiGeorge syndrome: The chromosome 22q11.2 deletion syndromes. Lancet, 370, pp.1443–1452.

Kousseff, G., 1984. Sacral meningocele with conotruncal heart defects: A possible autosomal recessive trait. Pediatrics, 74(3), pp.395–398.

Krahn, L., Maraganore, D. & Michels, V., 1998. Childhood-onset schizophrenia associated with parkinsonism in a patient with a microdeletion of chromosome 22. Mayo Clinic Proceedings, 73, pp.956–959.

Kretschmer, R., 1968. Congenital aplasia of the thymus gland (DiGeorge's syndrome). The New England Journal of Medicine, 279(24), pp.1295-1301.

Lachman, H. et al., 1996. Human catechol-o-methyltransferase pharmacogenetics: Description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics, 6, pp.243–250.

Lees, A., Hardy, J. & Revesz, T., 2009. Parkinson’s disease. Lancet, 373, pp.2055–2066.

Lesage, S. & Brice, A., 2009. Parkinson’s disease: From monogenic forms to genetic susceptibility factors. Human Molecular Genetics, 18(1), pp.48–59.

Lesch, K. et al., 2011. Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: Association with neuropeptide Y gene dosage in an extended pedigree. Molecular Psychiatry, 16, pp.491–503.

Li, D. et al., 2012. Co-existence of other copy number variations with 22q11.2 deletion or duplication: A modifier for variable phenotypes of the syndrome? Molecular Cytogenetics, 5(18), pp.1–4.

Li, T. et al., 2004. Evidence for association between novel polymorphisms in the PRODH gene and schizophrenia in a Chinese population. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 129B, pp.13–15.

Liburd, N. et al., 2001. Novel mutations of MYO15A associated with profound deafness in consanguineous families and moderately severe hearing loss in a patient with Smith-Magenis syndrome. Human Genetics, 109, pp.535–541.

Lindsay, E. et al., 1999. Congenital heart disease in mice deficient for the DiGeorge syndrome region. Nature, 401, pp.379–383.

Lindsay, E. et al., 2001. Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. Nature, 410, pp.97–101.

Lionel, A. et al., 2011. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Science Translational Medicine, 3(95), pp.1–11.

Litvan, I. et al., 2003. SIC task force appraisal of clinical diagnostic criteria for Parkinsonian disorders. Movement Disorders, 18(5), pp.467–486.

Liu, H., Abecasis, G., et al., 2002. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 99(26), pp.16859–16864.

Liu, H., Heath, S., et al., 2002. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 99(6), pp.3717–22.

Long, J. et al., 2006. Behavior of mice with mutations in the conserved region deleted in Velocardiofacial/DiGeorge syndrome. Neurogenetics, 7, pp.247–257.

Ludlow, L. et al., 1996. Identification of a mutation in a GATA binding site of the platelet glyoprotein Ib B promoter resulting in the Bernard-Soulier syndrome. Journal of Biological Chemistry, 271(36), pp.22076–22080.

Ma, G. et al., 2007. An association study between the genetic polymorphisms within TBX1 and schizophrenia in the Chinese population. Neuroscience Letters, 425, pp.146–150.

Malhotra, D. & Sebat, J., 2012. CNVs: Harbingers of a rare variant revolution in psychiatric genetics. Cell, 148, pp.1223–1241.

Manchester, K., 1996. Use of UV methods for measurement of protein and nucleic acid concentrations. BioTechniques, 20(6), pp.968–970.

Mas, S. et al., 2010. A functional variant provided further evidence for the association of ARVCF with schizophrenia. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 0(5), pp.1052–1059.

Maynard, T. et al., 2003. A comprehensive analysis of 22q11 gene expression in the developing and adult brain. Proceedings of the National Academy of Sciences of the United States of America, 100(24), pp.14433–14438.

Maynard, T. et al., 2008. Mitochondrial localization and function of a subset of 22q11 Deletion syndrome candidate genes. Molecular and Cellular Neuroscience, 39, pp.439–451.

McDonald-McGinn, D. et al., 1999. The Philadelphia story: The 22q11.2 deletion: Report on 250 patients. Genetic Counselling. 10(1), pp. 11–24.

McDonald-McGinn, D. et al., 2001. Phenotype of the 22q11.2 deletion in individuals identified through an affected relative: Cast a wide FISHing net! Genetics in Medicine, 3(1), pp.23–29.

McDonald-McGinn, D. et al., 2013. Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.2DS. Journal of Medical Genetics, 50, pp.80–90.

McDonald-McGinn, D. et al., 2015. 22q11.2 Deletion syndrome. Nature Reviews Disease Primers, 1(15071), pp.1–45.

McQuillin, A. et al., 2011. Analysis of genetic deletions and duplications in the University College London Bipolar Disorder case control sample. European Journal of Human Genetics, 19, pp.588–592.

Meechan, D. et al., 2006. Gene dosage in the developing and adult brain in a mouse model of 22q11 Deletion syndrome. Molecular and Cellular Neuroscience, 33, pp.412–428.

Meechan, D. et al., 2009. Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 Deletion/DiGeorge syndrome. Proceedings of the National Academy of Sciences of the United States of America, 106(38), pp.16434–16445.

Meechan, D. et al., 2011. Three phases of DiGeorge/22q11 Deletion syndrome pathogenesis during brain development: Patterning, proliferation, and mitochondrial functions of 22q11 genes. International Journal of Developmental Neuroscience, 29, pp.283–294.

Merla, G. et al., 2006. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. American Journal of Human Genetics, 79, pp.332–341.

Merscher, S. et al., 2001. TBX1 is responsible for cardiovascular defects in Velo-cardio-facial/DiGeorge syndrome. Cell, 104, pp.619–629.

Mick, E. et al., 2010. Family-based genome-wide association scan of Attention-Deficit/Hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 49(9), pp.898–905.

Miller, D. et al., 2010. Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. American Journal of Human Genetics, 86, pp.749–764.

Mok, K. et al., 2016. Deletions at 22q11.2 in idiopathic Parkinson's disease: A combined analysis of genome-wide association data. Lancet Neurology, 15(6), pp.585–596.

Momma, K., 2010. Cardiovascular anomalies associated with chromosome 22q11.2 Deletion syndrome. American Journal of Cardiology, 105, pp.1617–1624.

Mueller, O., Lightfoot S. & Schroeder, A., RNA Integrity Number (RIN) – standardization of RNA quality control. Agilent Technologies. Available at: https://www.agilent.com/cs/library/applications/5989-1165EN.pdf

Mukai, J. et al., 2004. Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia. Nature Genetics, 36(7), pp.725–731.

Mukai, J. et al., 2008. Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. Nature Neuroscience, 11(11), pp.1302–1310.

Munafò, M. et al., 2005. Lack of association of the COMT (Val158/108 Met) gene and schizophrenia: A meta-analysis of case-control studies. Molecular Psychiatry, 10, pp.765–770.

Murphy, K., Jones, L. & Owen, M., 1999. High rates of schizophrenia in adults with Velo-cardio-facial syndrome. Archives of General Psychiatry, 56, pp.940–945.

Nalls, M. et al., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature Genetics, 46(9), pp.989–997.

Neale, B. et al., 2008. Genome-wide association scan of Attention Deficit Hyperactivity disorder. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147B, pp.1337–1344.

Neale, B. et al., 2010. Case-control genome-wide association study of Attention-Deficit/Hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry, 49(9), pp.906–920.

Niarchou, M. et al., 2014. Psychopathology and cognition in children with 22q11.2 Deletion syndrome. British Journal of Psychiatry, 204, pp.46–54.

Niarchou, M. et al., 2015. The clinical presentation of Attention Deficit-Hyperactivity disorder (ADHD) in children with 22q11.2 deletion syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 168B, pp.730–738.

Nieratschker, V. et al., 2010. The catechol-o-methyl transferase (COMT) gene and its potential association with schizophrenia: Findings from a large German case-control and family-based sample. Schizophrenia Research, 122, pp.24–30.

Niklasson, L. et al., 2009. Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. Research in Developmental Disabilities, 30, pp.763–773.

Nussbaum, R. & Ellis, C., 2003. Alzheimer's disease and Parkinson's disease. New England Journal of Medicine, 348(14), pp.1356–1364.

O'Connell, J. et al., 2014. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genetics, 10(4), p.e1004234.

Ohtsuki, T. et al., 2004. Failure to find association between PRODH deletion and schizophrenia. Schizophrenia Research, 67, pp.111–113.

Oskarsdóttir, S. et al., 2008. Facial features in children with the 22q11 Deletion syndrome. Acta Paediatrica (Oslo, Norway), 97, pp.1113–1117.

Otani, K. et al., 2005. The ZDHHC8 gene did not associate with Bipolar Disorder or schizophrenia. Neuroscience Letters, 390, pp.166–170.

Paisan-Ruiz, C. et al., 2009. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. Annals of Neurology, 65, pp.19–23.

Paterlini, M. et al., 2005. Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice. Nature neuroscience, 8(11), pp.1586–1594.

Paylor, R. et al., 2001. Mice deleted for the DiGeorge/Velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments. Human Molecular Genetics, 10(23), pp.2645–2650.
Paylor, R. & Lindsay, E., 2006. Mouse models of 22q11 Deletion syndrome. Biological Psychiatry, 59, pp.1172–1179.
Paylor, R. et al., 2006. Tbx1 haploinsufficiency is linked to behavioral disorders in mice and humans: Implications for 22q11 deletion syndrome. Proceedings of the National Academy of Sciences of the United States of America, 103(20), pp.7729–7734.
Pe'er, I. et al., 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiology, 32(4), pp.381–385.
Perez, E. & Sullivan, K., 2002. Chromosome 22q11.2 deletion syndrome (DiGeorge and Velocardiofacial syndromes). Current Opinion in Pediatrics, 14, pp.678–683.
Philip, N. & Bassett, A., 2011. Cognitive, behavioural and psychiatric phenotype in 22q11.2 Deletion syndrome. Behavior Genetics, 41(3), pp.403–412.
Piliero, L. et al., 2004. T-cell homeostasis in humans with thymic hypoplasia due to chromosome 22q11.2 deletion syndrome. Blood, 103(3), pp.1020–1025.
Polito, L., Greco, A. & Seripa, D., 2016. Genetic profile, environmental exposure, and their interaction in parkinson's disease. Parkinson's Disease, 2016(6465793), pp.1–9.
Polymeropoulos, M. et al., 1997. Mutation in the α-Synuclein gene identified in families with Parkinson's disease. Science, 276, pp.2045–2047.
Prasad, S., Howley, S. & Murphy, K., 2008. Candidate genes and the behavioral phenotype in 22q11.2 deletion syndrome. Developmental Disabilities Research Reviews, 14, pp.26–34.
Prescott, K. et al., 2005. Microarray analysis of the Df1 mouse model of the 22q11 deletion syndrome. Human Genetics, 116, pp.486–496.
Price, A. et al., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), pp.904–909.
Pruim, R. et al., 2010. LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics, 26(18), pp.2336–2337.
Puech, A. et al., 1997. Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization. Proceedings of the National Academy of Sciences of the United States of America, 94, pp.14608–146013.
Purcell, S., Cherny, S. & Sham, P., 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics, 19(1), pp.149-50.
Purcell, S. et al., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, pp.559–575.
Radoeva, P. et al., 2014. Association between Autism Spectrum Disorder (ASD) in individuals with Velo-cardio-facial (22q11.2 Deletion) syndrome and PRODH and COMT Genotypes. Psychiatric Genetics, 24(6), pp.269–272.
Ramachandran, R. et al., 2015. Role of imaging and cytogenetics in evaluation of DiGeorge syndrome - a rare entity in clinical practice. Journal of Clinical Imaging Science, 5(1), pp.1–4.
Ramirez, A. et al., 2006. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nature Genetics, 38(10), pp.1184–1191.
Rauch, A. et al., 2006. Diaphragmatic defects and limb deficiencies - taking sides. American Journal of Medical Genetics Part A, 140A, pp.2063–2074.
Raux, G. et al., 2007. Involvement of hyperprolinemia in cognitive and psychiatric features of the 22q11 deletion syndrome. Human Molecular Genetics, 16(1), pp.83–91.
Rees, E. et al., 2014. Analysis of copy number variations at 15 schizophrenia-associated loci. British Journal of Psychiatry, 204, pp.108–114.

Ringnér, M., 2008. What is principal component analysis? Nature Biotechnology, 26(3), pp.303–304.
Ripke, S. et al., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511, pp.421–427.
Ritchie, M. et al., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), p.e47.
Robin, N. et al., 1995. Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. Nature Genetics, 11, pp. 459 – 461.
Roizen, N. et al., 2007. 22q11.2DS deletion syndrome: Developmental milestones in infants and toddlers. Journal of Developmental and Behavioral Pediatrics, 28(2), pp.119–124.
R-project.org. (n.d.). R: The R Project for Statistical Computing. [online] Available at: http://www.R-project.org/ [Accessed 26 Jan. 2017].
Rucker, J. et al., 2016. Phenotypic association analyses with copy number variation in recurrent depressive disorder. Biological Psychiatry, 79, pp.329–336.
Rutter, M., Bailey, A. & Lord, C. (2003). Social Communication Questionnaire (SCQ). 1st ed. [PDF] pp.1-3. Available at: http://www.hogrefe.co.uk/scq.html [Accessed 17 Apr. 2017].
Ryan, A. et al., 1997. Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: A European collaborative study. Journal of Medical Genetics, 34(10), pp.798–804.
Saito, S. et al., 2005. No association was found between a functional SNP in ZDHHC8 and schizophrenia in a Japanese case-control population. Neuroscience Letters, 374, pp.21–24.
Salyakina, D. et al., 2010. Variants in several genomic regions associated with asperger disorder. Autism Research, 3, pp.303–310.
Samii, A., Nutt, J. & Ransom, B., 2004. Parkinson's disease. Lancet, 363, pp.1783–1793.
Satake, W. et al., 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nature Genetics, 41(12), pp.1303–1310.
Scambler, P. et al., 1992. Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus. Lancet, 339, pp.1138–1139.
Scambler, P., 2000. The 22q11 deletion syndromes. Human Molecular Genetics, 9(16), pp.2421–2426.
Schneider, M. et al., 2014. Psychiatric disorders from childhood to adulthood in 22q11.2 Deletion syndrome: Results from the international consortium on brain and behavior in 22q11.2 Deletion syndrome. The American Journal of Psychiatry, 171, pp.627–639.
Schroeder, A. et al., 2006. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology, 7(3), pp.1–14.
Sedláčková, E., 1955. Velopharyngeal insufficiency as developmental disturbance (in Czech). Čas Lék Čes, 94, pp.1304–1307.
Sedláčková, E., 1967 The syndrome of the congenitally shortened velum the dual innervation of the soft palate. Folia Phoniatrica, 19, pp.441–450.
Shaikh, T. et al., 2000. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: Genomic organization and deletion endpoint analysis. Human Molecular Genetics, 9(4), pp.489–501.
Shannon, P. et al., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13, pp.2498–2504.
Shashi, V. et al., 2006. Cognitive correlates of a functional COMT polymorphism in children with 22q11.2 deletion syndrome. Clinical Genetics, 69, pp.234–238.
Shashi, V. et al., 2010. Evidence of gray matter reduction and dysfunction in chromosome 22q11.2 deletion syndrome. Psychiatry Research - Neuroimaging, 181, pp.1–8.
Shearer, E. et al., 2014. Copy number variants are a common cause of non-syndromic hearing loss. Genome Medicine, 6(37), pp.1–10.

Shimizu, T. et al., 1984. Conotruncal anomaly face syndrome: Its heterogeneity and association with thymic involution. In: Nora, J., Takao, A., (eds). Congenital Heart Disease: Causes and Processes. Mount Kisco, New York: Futura Publishing, pp.29–41.

Shojaee, S. et al., 2008. Genome-wide linkage analysis of a Parkinsonian-Pyramidal syndrome pedigree by 500 K SNP arrays. American Journal of Human Genetics, 82, pp.1375–1384.

Shpritzen, R. et al., 1978. A new syndrome involving cleft palate, cardiac anomalies, typical faces, and learning. Cleft Palate-Carniofacial Journal, 5, pp.56-62.

Shprintzen, R. et al., 1985. Anomalies associated with cleft lip, cleft palate, or both. American Journal of Medical Genetics, 20(4), pp.585-595.

Shprintzen, R., 2005. Historical overview. In Murphy, K. & Scambler, P., Velo-Cardio-Facial Syndrome: A Model for Understanding Microdeletion Disorders. Cambridge: Cambridge University Press, pp.1–19.

Shprintzen, R., 2008. Velo-cardio-facial syndrome: 30 years of study. Developmental Disabilies Research Reviews, 14(1), pp.3–10.

Simón-sánchez, J. et al., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature Genetics, 41(12), pp.1308–1316.

Sivagnanasundaram, S. et al., 2007. Differential gene expression in the hippocampus of the Df1/+ mice: A model for 22q11.2 deletion syndrome and schizophrenia. Brain Research, 1139, pp.48–59.

Sobin, C. et al., 2006. Neuromotor deficits in children with the 22q11 Deletion syndrome. Movement Disorders, 21(12), pp.2082–2089.

Sprecher, E. et al., 2005. A mutation in SNAP29, coding for a SNARE protein involved in intracellular trafficking, causes a novel neurocutaneous syndrome characterized by cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma. American Journal of Human Genetics, 77(2), pp.242–251.

Stark, K. et al., 2008. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nature genetics, 40(6), pp.751–760.

Stefansson, H. et al., 2008. Large recurrent microdeletions associated with schizophrenia. Nature, 455, pp.232–236.

Stergiakouli, E. et al., 2012. Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. American Journal of Psychiatry, 169, pp.186–194.

Stranger, B. et al., 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science, 315(5813), pp.848–853.

Strong, W.B., 1968. Familial syndrome of right-sided aortic arch, mental deficiency, and facial dysmorphism. Journal of Pediatrics, 73(6), pp.882–888.

Sullivan K., 2005. Immunodeficiency in velo-cardio-facial syndrome. In Murphy K., Scambler, P., Velo-Cardio-Facial Syndrome: A Model for Understanding Microdeletion Disorders. Cambridge: Cambridge University Press; pp. 123–134.

Sullivan, P., Fan, C. & Perou, C., 2006. Evaluating the comparability of gene expression in blood and brain. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 141 B, pp.261–268.

Suzuki, G. et al., 2009. Sept5 deficiency exerts pleiotropic influence on affective behaviors and cognitive functions in mice. Human Molecular Genetics, 18(9), pp.1652–1660.

Swillen, A., Devriendt, K., et al., 1999. The behavioural phenotype in velo-cardio-facial syndrome (VCFS): From infancy to adolescence. Genetic Counseling, 10(1), pp.79–88.

Swillen, A., Vandeputte, L., et al., 1999. Neuropsychological, learning and psychosocial profile of primary school aged children with the velo-cardio-facial syndrome (22q11 deletion): Evidence for a nonverbal learning disability? Child Neuropsychology, 5(4), pp.230–241.

Swillen, A. et al., 2000. Chromosome 22q11 deletion syndrome: Update and review of the clinical features, cognitive-behavioral spectrum, and psychiatric complications. American Journal of Medical Genetics, 97, pp.128–135.

Tang, K. et al., 2015. Behavioral and psychiatric phenotypes in 22q11.2 Deletion syndrome. Journal of Developmental and Behavioral Pediatrics, 36(8), pp.639–650.

Tansey, K. et al., 2014. Copy number variants and therapeutic response to antidepressant medication in major depressive disorder. Pharmacogenetics, 14(4), pp.395–399.

Tezenas Du Montcel, S. et al., 1996. Prevalence of 22ql 1 microdeletion. Journal of Medical Genetics, 33, pp.473–474.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature, 526, pp.68–74.

The International Schizophrenia Consortium, 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature, 455, pp.237–241.

Tomari, Y. & Zamore, P., 2005. MicroRNA Biogenesis: Drosha Can't Cut It without a Partner. Current Biology, 15(2), pp.R61-64.

Tunbridge, E., Harrison, P. & Weinberger, D., 2006. Catechol-o-methyltransferase, cognition, and psychosis: Val158Met and beyond. Biological Psychiatry, 60, pp.141–151.

Turner, S. et al., 2011. Quality control procedures for genome wide association studies. Current Proceedings in Human Genetics, 68(1), pp.1–24.

Valente, E. et al., 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science, 304, pp.1158–1160.

van Amelsvoort, T. et al., 2004. Cognitive deficits associated with schizophrenia in velo-cardio-facial syndrome. Schizophrenia Research, 70, pp.223–232.

van Amelsvoort, T. et al., 2008. Effects of a functional COMT polymorphism on brain anatomy and cognitive function in adults with velo-cardio-facial syndrome. Psychological Medicine, 38, pp.89–100.

van Beveren, N. et al., 2012. Functional gene-expression analysis shows involvement of schizophrenia-relevant pathways in patients with 22q11 deletion syndrome. PLoS One, 7(3), p.e33473.

Vawter, M. et al., 2004. Gender-specific gene expression in post-mortem human brain: Localization to sex chromosomes. Neuropsychopharmacology, 29, pp.373–384.

Verdin, E. & Ott, M., 2014. 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and beyond. Nature Reviews Molecular Cell Biology, 16, pp.258–264.

Vorstman, J. et al., 2006. The 22q11.2 deletion in Children: High rate of autistic disorders and early onset of Ppsychotic symptoms. Journal of the American Academy of Child & Adolescent Psychiatry, 45(9), pp.1104–1113.

Vorstman, J., Chow, E., et al., 2009. Association of the PIK4CA schizophrenia-susceptibility gene in adults with the 22q11.2 Deletion syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150B(3), pp.430–433.

Vorstman, J., Turetsky, B., et al., 2009. Proline affects brain function in 22q11DS children with the low activity COMT 158 allele. Neuropsychopharmacology, 34, pp.739–746.

Wang, K. et al., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research, 17, pp.1665–1674.

Wang, K. et al., 2009. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature, 459(7246), pp.528–533.

Wang, Y. et al., 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nature Genetics, 39(3), pp.380–385.

Wang, Y. et al., 2013. COMT rs4680 Met is not always the "smart allele": Val allele is associated with better working memory and larger hippocampal volume in healthy Chinese. Genes, Brain and Behavior, 12(3), pp.323–329.

Wechsler, D. 1999. Wechsler Abbreviated Scale of Intelligence (WASI) Manual. Psychological Corporation.

Weiss, L. et al., 2008. Association between microdeletion and microduplication at 16p11.2 and autism. New England Journal of Medicine, 358(7), pp.667–675.
Weiss, L., Arking, D. & The Gene Discovery Project of John Hophins the Autism Consortium, 2009. A genome-wide linkage and association scan reveal novel loci for autism. Nature, 461(7265), pp.802–808.
Wickremaratchi, M., Ben-Shlomo, Y. & Morris, H., 2009. The effect of onset age on the clinical features of Parkinson's disease. European Journal of Neurology, 16, pp.450–456.
Williams, H. et al., 2003. Detailed analysis of PRODH and PsPRODH reveals no association with schizophrenia. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 120B, pp.42–46.
Williams, H. et al., 2005. No association between schizophrenia and polymorphisms in COMT in two large samples. American Journal of Psychiatry, 162, pp.1736–1738.
Williams, H., Owen, M. & O'Donovan, M., 2007. Is COMT a susceptibility gene for schizophrenia? Schizophrenia Bulletin, 33(3), pp.635–641.
Williams, H. et al., 2013. Schizophrenia two-hit hypothesis in velo-cardio facial syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 162B, pp.177–182.
Williams, N. et al., 2008. Strong evidence that GNB1L is associated with schizophrenia. Human Molecular Genetics, 17(4), pp.555–566.
Williams, N. et al., 2010. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: A genome-wide analysis. Lancet, 376, pp.1401–1408.
Williams, N. et al., 2012. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: The role of rare variants and duplications at 15q13.3. American Journal of Psychiatry, 169, pp.195–204.
Williams, N., 2011. Molecular mechanisms in 22q11 deletion syndrome. Schizophrenia Bulletin, 37(5), pp.882–889.
Wilson, D., Burn, J. & Scambler, P., 1993. DiGeorge syndrome: Part of CATCH 22. Journal of Medical Genetics, 30, pp.852–856.
Wolfe, K. et al., 2017. Chromosomal microarray testing in adults with intellectual disability presenting with comorbid psychiatric disorders. European Journal of Human Genetics, 25, pp.66–72.
Xu, B. et al., 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. Nature Genetics, 40(7), pp.880–885.
Xu, B. et al., 2013. Derepression of a novel neuronal inhibitor is a major outcome of miRNA dysregulation due to 22q11.2 microdeletion. Cell, 152(1–2), pp.262–275.
Xu, M., St Clair, D. & He, L., 2010. Testing for genetic association between the ZDHHC8 gene locus and susceptibility to schizophrenia: An integrated analysis of multiple datasets. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 153B, pp.1266–1275.
Yagi, H. et al., 2003. Role of TBX1 in human del22q11.2 syndrome. Lancet, 362, pp.1366–1373.
Yang, L. et al., 2013. Polygenic transmission and complex neuro developmental network for Attention Deficit Hyperactivity Disorder: Genome- wide association study of both common and rare variants. American Journal of Medical Genetic Part B: Neuropsychaitric Genetics, 0(5), pp.419–430.
Zaleski, C. et al., 2009. The co-occurrence of early onset Parkinson disease and 22q11.2 Deletion syndrome. American Journal of Medical Genetics Part A, 149A(3), pp.525–528.
Zhang, Y. et al., 2000. Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. Proceedings of the National Academy of Sciences of the United States of America, 97(24), pp.13354–13359.
Zhang, Z. et al., 2011. MiR-185 targets the DNA methyltransferases 1 and regulates global DNA methylation in human glioma. Molecular Cancer, 10(124), pp.1–16.