Copy number variation in co-morbid neurodevelopmental disorders

Kate Alexandra Wolfe

Division of Psychiatry,

University College London

Thesis submitted for the degree of

Doctor of Philosophy

UCL

I, **Kate Wolfe** confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Copy number variants (CNVs) have been implicated in the pathogenesis of clinically distinct neurodevelopmental disorders (NDDs), indicating common underlying pathophysiology. Yet, the frequency, genetic architecture, and phenotypic role of pathogenic CNVs in adults with co-morbid neurodevelopmental phenotypes has not yet been systematically investigated.

Adults with intellectual disability (ID) and psychiatric co-morbidities were recruited from ID psychiatry services across the UK (N=202). Using a genotype-first approach, chromosomal microarray analysis (CMA) was undertaken, and variants were categorised using the NHS regional genetics service (RGCs) clinical pipeline. Genetic and phenotypic data was combined with two independent samples to enable frequency analyses (N=599). Targeted recruitment of individuals with 2q13 CNVs was undertaken via a patient support group, RGCs and the online rare CNV database DECIPHER (N=25).

The frequency of pathogenic CNVs was 11%, rising to 13% in the replication cohort. Both novel and recurrent loci were found to harbour pathogenic CNVs, with 70% at established NDD risk loci. A significantly higher population frequency of CNVs was identified in NDD risk regions (10%), compared with schizophrenia (3.1%, $p<0.0001$) and ID/autism spectrum disorder (6.5%, $p<0.0008$) populations. Phenotypic characterisation of CNVs at the 2q13 region suggests an early-onset neuropsychiatric phenotype with a high incidence of attention deficit hyperactivity disorder (ADHD) and challenging behaviours.

There is a high yield of pathogenic CNVs in patients with co-morbid neurodevelopmental phenotypes. In the main part, distinct loci are not involved in co-morbid NDD risk, but risk arises from the same loci identified in single disorder cohorts. Detailed phenotypic investigation of the 2q13 locus indicates that pleiotropy exists, however there is a preferential psychiatric outcome – in this instance ADHD.

Understanding the factors which modulate a CNV region with a high general risk for NDDs to a preferential neuropathological pathway will be key to understanding the complex hierarchy of psychiatric nosology and developing successful therapeutic interventions.

Acknowledgements

> "All men dream: but not equally. Those who dream by night in the dusty recesses of their minds wake in the day to find that it was vanity: but the dreamers of the day are dangerous men, for they may act their dreams with open eyes, to make it possible." T.E. Lawrence

Undertaking a PhD once seemed like an unattainable dream to me. There have been many people who have supported me on this journey, without whom I would not be where I am today.

Firstly, I would like to thank my research supervisors. I worked with Dr Andrew McQuillin and Dr Nick Bass as a Research Assistant prior to starting my PhD, and it was mainly their support and encouragement which motivated me to apply for a PhD studentship. Andy always has his door open for you, and has offered a wealth of knowledge from helping me with a problem in the lab to an issue with computer coding. Nick's clinical expertise has been invaluable throughout my PhD and his comprehensive editing of my work has enabled me to significantly develop my skills in academic writing. Professor David Skuse also took on the role as supervisor for my PhD, following his supervision during a laboratory rotation project. It has been extremely valuable to have David as an additional supervisor, given his vast academic and clinical experience, and I appreciate how enthusiastic he has been to regularly meet and discuss my work. I would also like to thank Dr Andre Strydom, who has unofficially acted in a supervisory capacity many times throughout my PhD, and has also been fundamental to my development as an independent researcher. Finally, I would like to thank my mentor Dr Elvira Bramon, who has been a constant source of insight and encouragement. All of my supervisors have supportively challenged me, and pushed me to succeed – I owe a lot of my success to their mentorship.

Secondly, I would like to thank everyone who has made the research that I have undertaken physically possible. The Medical Research Council funded my PhD studentship, providing me with research and personal support costs throughout my PhD. The Baily Thomas Charitable Fund provided funding for all of the intellectual

disability psychiatry sample collection, and awarded an additional grant for the development of the European consortium project. The work could not have been achieved without this funding, and I am sincerely grateful for their support. None of this research would be possible without the patients and their families, who have dedicated their time to take part in these projects. Some of these individuals were facing extremely challenging circumstances, yet still took the time to contribute to research, and I am deeply admirable of them all.

I would also like to thank all past and current members of the Molecular Psychiatry Laboratory. It has been an enriching experience to be part of a group which has a diverse range of research techniques and interests. The support of the group has been invaluable throughout my PhD. I would particularly like to thank Niamh O'Brien, Mariam Al Eissa, and Johan Thygesen for their friendship, advice and support – I have been extremely lucky to share this experience with such great colleagues.

My family members have been a constant source of support and encouragement during my academic undertakings. My parents, Mike and Anne Wolfe, have been there for me from my undergraduate studies to the completion of this PhD thesis. I am lucky to have had the scientific mind of my father and artistic creativity of my mother as inspiration during in my early-life and studies. My success at University would not have been possible without their practical and financial support and regular drives up the M6 motorway. They have always encouraged me to achieve my dreams, and I am lucky to have them as parents. Special thanks to my sister, Fiona, who has looked after me – with regular walks and great food – as I stayed with her family in Canada to write my PhD thesis. Also my Grandma Kathleen, who encouraged my academic studies from an early age – bribing me with extra pocket money to learn my times tables by heart and with regular trips to the library. Sadly, she is no longer with us, but I hope that this would have made her proud.

Last, but not least, I would like to thank my fiancé Jack, my biggest critic and best friend. Ich bin außerordentlich glücklich, dich während meines Studiums getroffen zu haben. Du forderst mich immer heraus, du fragst mich immer schwierige Fragen

während meiner Präsentationen oder wenn du meine Forschungsideen gehört hast. Es macht mich eine bessere Forscherin. Du gibst immer deine Zeit für die Bearbeitung meiner Grafiken und Texte. Du glaubst in mich mehr als andere und du bist meine Inspiration für alles.

"Be the change that you want to see in the world" Mahatma Gandhi

Table of Contents

Table of Tables

Table of Figures

List of abbreviations

ACOXL	Acyl-CoA Oxidase Like
ADHD	Attention Deficit Hyperactivity Disorder
ANAPC1	Anaphase Promoting Complex Subunit 1
Array CGH	Microarray-based comparative genomic hybridization
ASD	Autism Spectrum Disorders
AS	Angelman syndrome
AUC	Area Under the ROC Curve
BAF	B Allele Frequency
BCL2L11	BCL2 Like 11
BPAD	Bipolar Affective Disorder
BPI-S	Behaviour Problems Inventory - Short Form
C4	Complement Component 4
CACNA1C	Calcium channel, voltage-dependent, L-type, alpha 1C subunit
CAMHS	Child And Adolescent Mental Health Services
ChA-PAS	Child and Adolescent Psychiatric Assessment Schedule
CHD2	Chromodomain Helicase DNA Binding Protein 2
CHD8	Chromodomain Helicase DNA Binding Protein 8
CMA	Chromosomal Microarray Analysis
CNTN4	Contactin 4
CNTN6	Contactin 6
CNVs	Copy number Variations
DAOA	D-amino acid oxidase activator
DD	Developmental Delay
DDD	Deciphering Developmental Disorders
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
DGV	Database of Genomic Variants
DNA	Deoxyribonucleic acid
DNMs	Damaging de Novo Mutations
DRD2	Dopamine Receptor D2
DSM	The Diagnostic and Statistical Manual of Mental Disorders
ExAC	Exome Aggregation Consortium
GENMID	GENetics of Mental disorders in Intellectual Disability
GP	General Practitioner
GRIA3	Glutamate Ionotropic Receptor AMPA Type Subunit 3
GRIN2B	Glutamate Ionotropic Receptor NMDA Type Subunit 2B
GWAS	Genome Wide Association Study
HGP	Human Genome Project
HPO	Human Phenotype Ontology
HWE	Hardy-Weinberg equilibrium
ICD-10	International Classification of Diseases 10th Revision
ID	Intellectual Disability
ID+	Intellectual disability and co-morbid psychiatric disorders
IQ	Intelligence Quotient

ISCA	International Standard Cytogenomic Array
LCRs	Low Copy Repeats
LRR	Log R Ratio
MCA	Multiple Congenital Abnormalities
MDD	Major Depressive Disorder
MERTK	MER Proto-Oncogene, Tyrosine Kinase
MHC	Major Histocompatibility Complex
MHRN	Mental Health Research Network
NAHR	Non-Allelic Homologous Recombination
NCBI	National Center for Biotechnology Information
NDDs	Neurodevelopmental Disorders
NF1	Neurofibromatosis type 1
NGS	Next-Generation Sequencing
NHS	National Health Service
NRXN1	Neurexin 1
NTRK2	Neurotrophic Receptor Tyrosine Kinase 2
OCD	Obsessive Compulsive Disorder
ODD	Oppositional Defiant Disorder
PAH	Phenylalanine Hydroxylase
PAS-ADD	Psychiatric Assessment Schedule for Adults with Developmental Disabilities
PCR	Polymerase Chain Reaction
PGC	Psychiatric Genomics Consortium
PI	Principle Investigator
PKU	Phenylketonuria
PWS	Prader–Willi syndrome
RGC	Regional Genetics Centre
SADS-L	Schizophrenia and Affective Disorders Schedule
SD	Standard deviation
SETD1A	SET Domain Containing 1A
SLC28A3	Solute Carrier Family 28 Member 3
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
STAG2	Stromal Antigen 2
SUPT16H	SPT16 Homolog, Facilitates Chromatin Remodeling Subunit
SYT1	Synaptotagmin-1
UBE3A	Ubiquitin–protein ligase E3A
UCSC	University of California Santa Cruz
UK	United Kingdom
VOUS	Variants Of Unknown Significance
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing

Chapter 1 Introduction

Parts of this chapter have been adapted from the following book chapter:

Wolfe, K., Strydom, A., Bass, N. (2017) Genetics of Intellectual Disability, In Kerr, M., Seminars in Psychiatry, Royal College of Psychiatrists, Cambridge University Press. (In press)

1.1 Terminologies

Two of the main terminologies that will be utilised throughout this thesis have differential usage in the published literature and are discussed here for clarification.

Neurodevelopmental disorders (NDDs) are a highly heterogeneous group of disorders characterised by perturbed cognition, communication, behaviours, and motor functioning, as a result of atypical brain development[1]. Intellectual disabilities (ID), autism spectrum disorders (ASD), attention deficit hyperactivity disorder (ADHD) and epilepsy are all considered to be NDDs. The category also extends to include neuropsychiatric disorders, such as schizophrenia[2]. There has been a shift in understanding in psychiatry, whereby many later-onset psychiatric disorders are now understood to have their origins in the developmental period. For example, cognitive deficits in schizophrenia have been shown to be present prior to the onset of the disorder[3]. However, the terms neurodevelopmental and neuropsychiatric disorder are still often used interchangeably in the published literature. Some authors have coined new terminology to describe this phenomenon, for example Moreno-De-Luca *et al.* suggested developmental brain dysfunction as a term to describe the group of neurodevelopmental and neuropsychiatric disorders that encompass various clinical diagnoses[1]. In the main discussion of this thesis I use the term NDDs for simplicity, however the exact phenotypic composition of each sample is provided in the method sections of the individual chapters.

Pathogenic is a term that is also often used variably in the published literature. Broadly speaking, a genetic variant can be said to be pathogenic if it causes disease. However, the usage of this term has been complicated by the discovery that a genetic variant, which is thought to cause disease in some individuals, can be present in other

individuals with no apparent effect. This is discussed in detail in the context of copy number variations (CNVs) in 1.3.2. Where relevant, the exact meaning of the term pathogenic has been discussed in the method sections of the individual chapters.

1.2 A brief history of genetic investigations for neurodevelopmental disorders

In the last century there has been a revolution in the genetic investigation of NDDs. In the 1950s, a technique for visualisation of the complete set of chromosomes (also known as the karyotype) under optical microscopes was developed[4]. This cytogenetic technique enabled the identification of abnormalities of chromosome number (aneuploidies) and large structural abnormalities, such as translocations. Translocations involve the movement of stretches of deoxyribonucleic acid (DNA) between chromosomes and can be balanced or unbalanced, depending on whether there is an overall loss or gain of genetic material. One example of a disorder arising from aneuploidy is Klinefelter syndrome, a disorder whereby there is an extra copy of the X chromosome causing ID and various psychiatric phenotypes[5].

The 1970s and 1980s saw the development of molecular genetic techniques to manipulate DNA – such as the polymerase chain reaction (PCR). PCR became a core technique in molecular genetics, as it can be used to generate millions of copies of specific segments of the genome for a variety of subsequent genetic assays. With the identification of polymorphic DNA markers and application of PCR, genetic linkage analysis became a powerful tool for gene discovery. Linkage analysis is based on the concept that polymorphic DNA markers of known chromosomal position can be used to approximately locate disease causing genetic variation through analysis of the co-segregation of specific alleles of the markers and the disease within families. This was predominantly successful for disorders with Mendelian patterns of inheritance. Mendelian disorders are caused by mutations of bases within a single gene, for autosomal dominant disorders only one copy of the mutated gene is required to give rise to the disease phenotype, whereas in autosomal recessive conditions two abnormal copies of the gene are required.

An example of a disorder which can be mapped by linkage analysis is phenylketonuria (PKU), an autosomal recessive disorder that can lead to the development of ID,

seizures, heart problems, and a range of psychiatric phenotypes[6]. PKU arises due to mutations in the phenylalanine hydroxylase (*PAH*) gene. The gene encodes an enzyme that catalyses the breakdown of the amino acid phenylalanine. Mutations result in enzyme deficiency and a resultant toxic build-up of phenylalanine in the brain and body. Delineation of the inheritance pattern and metabolic pathology of PKU occurred before the *PAH* gene was mapped to chromosome 12 in the 1980s[7]. The enzyme deficiency can be ameliorated by implementation of a low-phenylalanine diet, preventing the development of ID and other disease pathology. Newborn screening programmes for PKU have been widely implemented and have proven to be very successful[8].

Around the same time, deletions and duplications of sections of the chromosome were identified as being involved in the aetiology of NDDs. For example, in 1981 an interstitial deletion of the paternal chromosome was identified at the 15q11–q13 locus which causes Prader–Willi syndrome (PWS)[9]. PWS has a characteristic clinical phenotype, comprising: childhood-onset obesity with extreme hyperphagia, mild ID, a recognisable pattern of dysmorphism, hypogonadism and growth insufficiencies. Behavioural problems, including tantrums and compulsive traits, are estimated to affect 70–90% of individuals. ASD is also present in approximately 25% of cases, and psychosis in 5–10%[10]. PWS is a disorder of DNA methylation, an important mechanism by which genes are switched on or off, resulting in changes in gene activity. Normally, genes from both the maternal and paternal chromosomes are expressed, however, in a process called genomic imprinting, one of the parental genes is imprinted (epigenetically silenced by DNA methylation) and therefore only one active copy of the gene is present in the offspring. In PWS a set of genes in the 15q11-13 region are imprinted (turned off), on the maternal chromosome, so deletion of the equivalent region on the paternal chromosome results in the lack of any active copies of these genes.

Angelman syndrome (AS) is the reciprocal condition to PWS, arising through a deletion of the maternal chromosome in the imprinted 15q11-q13 region. A set of genes in this region are imprinted (turned off) on the paternal chromosome, so deletion of the region from the maternal chromosome results in lack of any active copies of these genes. AS presents with a very different phenotype – comprising: severe ID,

virtual absence of speech, seizure disorders and mild dysmorphisms. The behavioural phenotype is particularly distinct, with hyperactivity, frequent laughter, and motor stereotypies[11]. It is still unclear which imprinted gene(s) contribute to PWS, however, it is known that the gene encoding ubiquitin–protein ligase E3A (*UBE3A*) explains many of the manifestations of AS and single-gene mutations can also give rise to the disorder[12].

The Human Genome Initiative, more commonly known as the Human Genome Project (HGP), was conceived in 1986 and set out to sequence all 3 billion base pairs of the human genome. The sequence was made publically available and has enabled the development of bioinformatic resources, which provide powerful tools for genetic investigation and clinical genetics[13]. Furthermore, the HGP provided the driving force behind the development of high-throughput next-generation sequencing (NGS). NGS technologies have made it possible to rapidly sequence the whole protein-coding region of the genome – whole-exome sequencing (WES) – or even the whole genome – whole-genome sequencing (WGS), at a low cost. NGS is a very powerful tool for identifying pathological single-base changes in the DNA sequence, which are referred to as single nucleotide variants (SNVs). In particular, NGS has facilitated the identification of *de novo* mutations. By definition, *de novo* mutations are not present in the parents and have arisen for the first time during egg or sperm cell formation, or in early embryonic development. A combination of the ability to map changes in a genetic sequence to a specific gene and technological developments, enabling detection of submicroscopic structural variations, also gave rise to copy number variants (CNVs) being identified as an important class of genetic variation in NDDs.

1.3 Copy number variations

CNVs are structural variants of at least 1kb in size that are present at a variable copy number in comparison with a reference genome[14]. Studies of healthy control populations have revealed that large-scale copy number polymorphisms are common throughout the human genome[15,16], and approximately 12% of the genome comprises CNVs[17]. Whilst there are many CNVs that are neutral in function, if the deletion (loss) or duplication (gain) of genetic material affects a dosage-sensitive gene this can lead to changes in gene expression and protein function. These variants are under negative

genetic selection, with a general trend of reduced fecundity as the severity of the phenotype increases[18]. Thus, rare variants of recent origin are the primary contributors to genetic risk, and it is the rare CNVs (typically defined as occurring at a frequency of <1% in the general population) that are most frequently associated with disease[19].

Particular regions of the genome are more prone to CNV. Low copy repeats (LCRs), also known as segmental duplications, are regions which contain repetitive DNA sequences. Genomic regions rich in LCRs are predisposed to non-allelic homologous recombination (NAHR), whereby recombination errors occur between two sequences with high similarity. NAHR is one of the primary mechanisms giving rise to CNVs, and it has been estimated that there are approximately 130 hotspots in the human genome which are vulnerable to CNV caused by NAHR[20]. As a result of this mechanisms multiple CNVs, which are nearly identical to one another, can arise independently in different individuals. These recurrent *de novo* CNVs have a major role in the pathogenesis of NDDs. Other mechanisms of structural variation include: non-homologous end joining, fork stalling and template switching, and L1-mediated retrotransposition[21]. Equally, non-recurrent CNVs, rearrangements which arise in regions that don't contain LCRs and differ in size between patients, are another pathological mechanism[22].

1.3.1 Chromosomal microarray analysis

DNA microarrays, also known as nucleic acid arrays, are small slides to which thousands of nucleic acid probes are bound. This enables hundreds of thousands of genotyping reactions to be carried out simultaneously. Currently there are two main types of microarray used for chromosomal microarray analysis (CMA), microarray-based comparative genomic hybridization (array CGH) and single nucleotide polymorphism (SNP) microarrays. In array CGH, DNA from the patient and a reference control are differentially fluorescently labelled and changes in genomic copy number, using probes at varying intervals, are made visible by differences in florescence levels[23]. SNP platforms were primarily developed to detect changes to single bases of the DNA sequence. However, it is also possible to determine genomic copy number using SNP arrays by detecting changes in the intensity information of segments of DNA[24].

1.3.2 Clinical applications of chromosomal microarray analysis

In 2010, a consensus statement from the International Standard Cytogenomic Array Consortium recommended CMA as one of the first-line cytogenetic tests, replacing karyotyping, for postnatal investigation of idiopathic developmental delay (DD), ID, multiple congenital abnormalities (MCA) and ASD[25]. Historically medical genetics has adopted a phenotype-first approach, whereby a collection of patients with similar phenotypes are investigated with the aim of mapping the pathological genetic variant. The advantage of CMA is that it adopts a genotype-first model, whereby a reverse strategy is employed of identifying pathogenic genetic variants and analysing associated phenotypes[26]. Currently, clinical screening for pathogenic CNVs is not available for other psychiatric disorders, although there is ongoing debate about whether to introduce CMA testing for individuals with schizophrenia[27,28].

A major challenge in the clinical application of CMA has been the interpretation of CNV pathogenicity. Typically, variants are categorised into three main categories; pathogenic, variants of unknown significance (VOUS); and benign. Factors that influence variant categorisation include: the inheritance pattern (whether the CNV is inherited or arises *de novo* in the individual), the size and genetic content of the CNV (large CNVs affecting brain expressed genes are more likely to be pathogenic), the likely functional consequence of gene disruption, and whether the CNV is present in healthy control datasets[14].

One measure which attempts to quantify the functional consequence of CNV is the haploinsufficiency index. Haploinsufficiency is the inability of a gene to retain a normal function when only one copy of the gene is present, in other words a measure of the tolerance of particular genes to CNV deletions[29]. A comparable methodology predicting the functional consequence of CNV duplications is yet to be developed. Increasingly online datasets are being utilised to aid the categorisation of variants. The Database of Genomic Variants (DGV) catalogues variation seen in healthy control populations and therefore contains CNVs not thought to be implicated in disease[30]. Whereas the DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER) is an international online research portal which facilitates

anonymised data sharing for rare CNVs, primarily ascertained in children with developmental disorders of unknown cause[31].

A confounding factor in interpreting pathogenicity has been the variable penetrance and expressivity of many CNVs. Penetrance is the proportion of people with a particular genotype who exhibit the phenotype associated with that genotype. A disorder is said to have reduced (or incomplete) penetrance when the aberrant genotype is present in the absence of the associated phenotype. Expressivity refers to the severity of the associated phenotype[32]. Down syndrome (trisomy of chromosome 21) is an example of a genetic disorder with full penetrance, in that all individuals with trisomy 21 present with features of Down syndrome, including some degree of intellectual impairment. However, there is considerable variation in the associated phenotypes, for example, some individuals with Down syndrome may have an intelligence quotient (IQ) in the borderline range (IQ 70-85), whilst others are more severely affected with an IQ below 35[33]. The expressivity of the phenotype can be influenced by the degree of genetic pleiotropy, whereby the altered function of a gene can cause diverse phenotypic outcomes. For example, exonic deletions of the Neurexin 1 (*NRXN1*) gene have been identified in studies of ID, ASD, and schizophrenia[34]. A growing area of discovery is CNVs which are known as neurosusceptibility loci, whereby CNVs are present in the general population but are enriched in NDDs[35].

Whilst guidelines exist for the categorisation of CNVs there remains an element of subjectivity in clinical interpretation and designation of CNV pathogenicity continues to be a moveable field. For example, one study re-interpreted CNV results from 67 individuals with idiopathic ID two years after the initial analysis and found a statistically significant increase in potentially pathogenic CNVs[36]. I will now further discuss the genetic underpinnings of three primary disorders – ID, schizophrenia, and ASD – as a prelude to considering combined neuropathological mechanisms.

1.4 Intellectual Disabilities

1.4.1 Clinical characteristics and aetiology

ID is traditionally defined as significant impairments in intellectual and adaptive functioning with onset before the age of 18 years[36]. Before the age of five years

functioning is measured on a developmental, rather than intelligence coefficient, so it is more typical to talk about DD, although DD does not inevitability lead to ID in adulthood[37]. The International Classification of Diseases 10th Revision (ICD-10) criteria for ID refer to the historic term mental retardation, however this will be updated in the ICD-11 revision and I will use the term ID throughout the text[38]. The first axis of the ICD-10 diagnostic guidelines for ID describes the degree, or level of ID, segregated by scores on IQ scales[39]. ID is typically equated with a score of 2 or more standard deviations (less than 70) below the population mean on IQ tests[2]. Axes II-V of the ICD-10 criteria characterise the presence of associated physical, mental and psychosocial disorders[39]. ID is estimated to affect 2-3% of the general population and for approximately 50% of individuals with ID the cause is unknown[36].

The phenotype of ID can be highly variable in terms of severity and the domains of intellectual function that are affected. ID also often co-occurs with other medical and psychiatric phenotypes, such as congenital malformations, epilepsy and ASD. ID can, thus, be divided into two broad categories: syndromic ID, whereby there is a co-occurrence of particular clinical phenotypes, and non-syndromic ID. Typically, homogenous syndromic phenotypes are caused by large *de novo* events[40]. The presence of syndromic features can guide genetic investigations, for example testing for trisomy 21 in Down's syndrome or single-gene testing for PKU. Whereas, genome-wide CMA enables systematic investigation of the genome in the absence of syndromic features[37]. ID is frequently associated with co-morbid psychiatric disorders and/or behavioural problems. For example, the point prevalence of psychosis has been estimated as 10 times higher in ID[41]. Recent estimates from United Kingdom (UK) primary care records show that approximately 21% of individuals with ID have a psychiatric disorder and 25% have some record of challenging behaviours. More than two thirds of these individuals had a record of prescription of any psychotropic drug, with more than a quarter receiving an antipsychotic[42].

The extreme heterogeneity of ID has confounded the understanding of pathological mechanisms. ID has typically been considered to lie at the extreme end of the normal IQ distribution in the general population. However, recent research, evaluating 1,000,000 sibling pairs and 9,000 twin pairs, has revealed that the factors influencing severe ID differ from those influencing mild ID. Whereas, mild ID and IQ in the

normal range can be considered on the same spectrum, severe ID is thought to be a distinct disorder[43]. Therefore, the severity of ID, the presence of syndromic features and the range of associated medical and psychiatric phenotypes are important factors for consideration.

1.4.2 Genetic risk factors

Investigation of CNVs in DD/ID has predominantly occurred in paediatric cohorts. Analysis of 14 recurrent CNV regions in 15,749 cases referred for diagnostic testing (referrals for testing were due to a combination of DD/ID, MCA, dysmorphologies and ASD) and 10,118 controls identified 14 deletions and 7 duplications that were significantly overrepresented in cases[44]. Another large analysis of pathogenic CNVs in children with similar phenotypes was undertaken by Cooper *et al.* in 15,767 cases and 8,329 controls[45]. Over 8% of cases carried an imbalance at one of 45 previously documented genomic disorder loci. A CNV burden analysis was undertaken between the cases and controls, excluding common CNVs (>1% population frequency). At a threshold of >400kb ~25.7% (4,047) cases harbored an event of at least this size, as compared to 11.5% of the controls. The authors conclude from this that ~14.2% of disease in cases is caused by CNVs >400 kb, however this conclusion presumes that the percentage difference, or the excess between the cases and controls, is responsible for disease. Whereas, it may be that that some of these large CNVs do not contribute to the disease phenotype. Also, the authors did not consider genes disrupted by the CNVs in this analysis, and it may be that CNVs of differing size are more relevant to the phenotypic differences between cases and controls.

A refined CNV morbidity map, comprising data from 29,085 children with ID and MCA (some of which were included in the previous analysis) went on to identify 70 CNV regions significantly associated with DD[46]. It is through this mechanism of integrating data from large datasets that new CNV syndromes are being verified and characterised, for example the 17q21.31 microdeletion syndrome, a multisystem disorder characterised by ID, distinctive facial dysmorphisms, and hypotonia[47].

Investigation of rare SNVs has also served to elucidate the aetiology of ID, particularly for sporadic moderate–severe ID. An early exome sequencing study focusing on *de*

novo SNVs analysed ten patient–parent trios and identified an average of five candidate non-synonymous *de novo* mutations per affected individual[48]. Non-synonymous exonic mutations alter the normal amino acid sequence, which can result in changes in protein configuration and gene function. WGS is now also being applied to the study of ID. One of the first WGS studies comprised patient–parent trios in a cohort of 50 patients with severe idiopathic ID. The cohort had previously undergone extensive genetic testing, including single-gene testing, CMA and exome sequencing analysis. The diagnostic yield for WGS was 42%, in comparison with an average diagnostic yield of 12% for CMA and 27% for WES[49]. Interestingly, the variants identified were all in the coding regions of the genome and had been missed by limitations of the previous technologies, which highlights that intragenic variant interpretation remains challenging. The technological advances afforded by NGS have enabled rapid progress in our understanding of the genetics of ID over the last decade and over 700 ID relevant genes have now been identified[2].

Another major contribution to the literature on the genetic aetiology of ID has arisen from the UK-based Deciphering Developmental Disorders (DDD) study. The study aims to apply CMA and sequencing methods to families affected by severe developmental disorders, approximately 87% of whom had DD/ID, and the majority of which had no previous family history of the disorder. A combined exome and array based approach has enabled the detection of novel genes associated with developmental disorders[50]. The most recent results from an exome sequencing analysis of 4,293 families, meta-analysed with 3,287 cases from similar populations, yielded 94 genes which were enriched for damaging *de novo* mutations (DNMs). Overall 42% of the cohort carried pathogenic DNMs with an estimated prevalence of between 1 in 213 to 1 in 448 births, with prevalence increasing with parental age[51].

1.5 Schizophrenia

1.5.1 Clinical characteristics and aetiology

The core features of schizophrenia are positive symptoms (delusions and/or hallucinations – also known as psychotic symptoms), negative symptoms (lack of motivation and social withdrawal) and cognitive impairments[52]. Initial models of schizophrenia pathology centered around the dopamine hypothesis. The fact that

psychosis can be induced by drugs which activate the dopamine system, and that antipsychotic drugs target the dopamine D2/3 receptors, highlighted the system as a likely candidate for disease pathology[53]. However, the neurodevelopmental hypothesis gained traction following evidence from longitudinal studies showing that cognitive impairments preceeded the prodromal phase of psychosis and onset of clinical symptoms[3,54]. Furthermore, various prenatal and environmental risk factors were shown to increase risk for schizophrenia. More recently an integrated sociodevelopmental-cognitive model has been proposed, which incorporates the role of social risk factors for schizophrenia alongside the dopamine and neurodevelopmental hypotheses[53].

Support for a genetic component to schizophrenia susceptibility has been provided by numerous twin studies[55–57]. Latest estimates indicate that the concordance rate of schizophrenia is 33% in monozygotic twins and 7% in dizygotic twins, with an estimated heritability of 79%[57]. Heritability estimates of schizophrenia are second highest among major psychiatric disorders, with ASD being the highest[58]. However, the finding of 33% concordance in monozygotic twins also indicates a significant role for environmental factors in risk for developing schizophrenia[57].

1.5.2 Rare genetic risk factors

The strongest known genetic risk factors for schizophrenia are pathogenic CNVs, which typically have a low population frequency but confer significant risk for development of the disorder (ORs 2-60)[59]. In an early CNV study, Kirov *et al.* found 2 CNVs, a deletion at 2p16.3 and a duplication at 15q13.1, thought to be associated with schizophrenia[60]. The genes involved, *NRXN1* and *APBA2*, code for proteins involved in synaptic development and functioning. Another rare CNV study was conducted by Walsh *et al.*, comprising 150 patients with schizophrenia or schizoaffective disorder and 268 healthy controls. They identified that individuals with schizophrenia were more than three times likely to harbor CNVs that deleted or duplicate one or more genes ($P = 0.0008$). The significance of this finding increased when considering patients with an early age-of-onset (<18 years)[61].

Later, CNV analyses in 6,882 schizophrenia cases and 6,316 controls, identified 11 of 15 regions that were significantly associated with schizophrenia risk. The strongest support was for CNV duplications at the 16p11.2 locus and the AS/PWS critical region. Whereas, the strongest support for CNV deletions was at the 22q11.2, 1q21.1, 2p16.3 (*NRXN1*) and 15q11.2 loci. Overall 2.5% of schizophrenia patients and 0.9% of controls had CNVs at one or more of these 15 loci[62]. The latest CNV analysis from the Psychiatric Genomics Consortium (PGC), an international collaboration of psychiatric disorder researchers, comprised 21,094 cases and 20,227 controls. The study found a global enrichment of CNV burden in cases (OR=1.11) compared to controls. Gene set analysis revealed that most of the signal was driven by deletion CNVs in synaptic or other neuronal component gene sets[59].

The role of rare SNVs in risk for schizophrenia has also been explored with the increasing application of exome sequencing technologies. A *de novo* paradigm analysis in 623 schizophrenia trios and controls identified a significant enrichment of non-synonymous mutations in genes encoding synaptic proteins in cases. This enrichment was particularly observed in glutamatergic postsynaptic proteins and interaction proteins modulating synaptic strength[63]. Concurrent analyses of the exome sequences of 2,536 schizophrenia cases and 2,543 controls confirmed an enrichment of rare mutations (defined as less than 1 in 10,000) in similar synaptic gene sets[64]. Furthermore, exome sequencing of 12,332 unrelated individuals, 4,877 of whom were affected with schizophrenia, has highlighted the role of inherited rare variants in risk for schizophrenia. The excess rare variant burden identified in this study, ~0.25 per person, compared with results from *de novo* paradigms, suggests that this observed excess must arise from inherited variants[65].

Singh *et al.* carried out a comprehensive meta-analysis of three different types of rare variant data – WES data, *de novo* variants from family based trio data, and CNV data – in cohorts of individuals with schizophrenia and healthy controls. Combined analysis identified that schizophrenia cases have a significantly higher burden of rare damaging variants in 3,488 genes, which are typically depleted for loss-of-function variants. These loss-of-function intolerant genes were identified by analysing exomes from healthy controls, without a known psychiatric diagnosis, in the Exome Aggregation Consortium (ExAC) exome database[66]. The risk variants are concentrated in risk genes

for NDDs in patients with schizophrenia who also have ID, and these patients have a higher burden of rare, damaging variants. However, the significant enrichment persists in other genes when excluding these co-morbid cases, meaning rare variants also confer risk for individuals with schizophrenia without ID[67].

1.5.3 Common genetic risk factors

It is now understood that polygenic disorders are not caused by a single genetic event, but are the aggregate effect of multiple individual gene events and significant environmental contributions[26]. Schizophrenia has been shown to be polygenic in nature, involving thousands of common SNPs, with very small individual effect sizes that increase the risk for developing the disorder by 1.1- to 2-fold[27]. Study of case-control frequencies of common SNPs is traditionally undertaken using a genome wide association study (GWAS) method. Since the first GWAS in 2009 there have been concurrent increases in the sample sizes of GWAS studies and the number of loci associated with schizophrenia[68]. The latest PGC GWAS, comprising 36,989 cases and 113,075 controls, identified 108 loci achieving genome wide significance for association with schizophrenia[69]. Some of the GWAS hits tie in with existing hypotheses, such as involvement of the dopamine receptor D2 (*DRD2*) gene and abnormal dopamine signaling. Support for the involvement of acquired immunity is derived from the robust association of the major histocompatibility complex (MHC) region on chromosome 6[68]. Further work to delineate the signal at this locus has identified a large number of common, functionally distinct, structural variants affecting the complement component 4 (*C4*) genes. Increased risk of developing schizophrenia was particularly associated with variants that increase expression of the *C4A* gene, which potentially drives pathological synapse loss in schizophrenia[70].

The polygenic risk score method has been developed to aggregate the effect of multiple SNPs of small effect into a combined composite score. Thus, a schizophrenia polygenic risk score, provides a quantitative measure of genetic predisposition to developing schizophrenia. A discovery dataset, a GWAS with available effect sizes, is used to weight a target dataset that has genome-wide genotype data[71]. Although the discriminative accuracy of risk scores is not yet sufficient for screening in clinical populations[72], the score constitutes a powerful research tool. The schizophrenia

polygenic risk score was found to be the highest, out of ten complex traits, for predicting case-control status, with the best area under the receiver operating curve (AUC) 0.82[73].

The relationship between rare and common genetic variants and how these separable factors interact to contribute to risk of developing schizophrenia is an evolving area of research. Investigation of patients with schizophrenia who carry rare pathogenic CNVs has found that these patients still carry an excess of common risk alleles[74]. This provides support for a polygenic threshold model, whereby individuals with high penetrance schizophrenia variants possess many genetic risk factors, and it is the combination of these factors that is sufficient to surpass a threshold for clinical diagnosis.

1.6 Autism spectrum disorders

1.6.1 Clinical characteristics and aetiology

ASDs are characterised by a triad of impairments, comprising social and communication difficulties, stereotyped and repetitive behaviours and/or a restricted range of interests[75]. An interesting feature of ASDs are that they operate across the spectrum of IQ, with ASD being observed in individuals with extremely high and low IQ. Increased risk for ASD is well documented in ID and approximately one-third of individuals with chromosomal and genetic abnormalities have significant autistic traits[76]. Psychological investigation of ASDs has revealed cognitive deficits in theory of mind, the ability to understand other people's mental states, central coherence, and the ability to integrate information at different levels[77]. At a biological level abnormalities in dopamine signaling have been proposed as a model of ASD pathogenicity[78]. Finally, the observation of seizures and sensory hyperactivity being frequently associated, provides support for a cortical hyperexcitability model of ASD[79].

Evidence for a genetic component to the aetiology of ASD has been provided by family studies. The recurrence risk of ASD in siblings of ASD probands has been estimated at 10.9%, with approximately 20% of siblings displaying some phenotypic features of ASD[80]. Twin studies have revealed higher concordance rates, with

approximately 31% of dizygotic and 88% of monozygotic twins being concordant for ASD[81].

1.6.2 Rare genetic risk factors

There is a large body of research on the role of rare pathogenic CNVs in ASDs. Weiss *et al.* found a recurrent deletion and reciprocal duplication at 16p11.2 that confers risk for DD and ASD, accounting for approximately 1% of cases[82]. Many studies have focused on distinguishing between simplex ASD, whereby only one family member is affected, and multiplex ASD, whereby multiple family members are affected. Early evidence that these two classes (simplex vs multiplex) are indeed genetically distinct was provided by the observation that *de novo* CNVs were more frequent in simplex families[83]. CNV analysis in 1124 simplex families identified significant associations with rare recurrent *de novo* CNVs at numerous loci and concluded that large *de novo* CNV confer substantial risks (OR=5.6) for ASD[84]. Regional analysis of pathogenic CNVs has showed an overall enrichment of brain-expressed genes in probands[85], as compared to controls. Numerous functional gene networks have been implicated, including: the ubiquitination system, neuronal cell-adhesion molecules[75], cellular proliferation, projection and motility[86]. One of the largest *de novo* CNV studies in ASD to date, comprising 2,591 families, found a strong association for ASD risk at six loci (1q21.1, 3q29, 7q11.23, 16p11.2, 15q11.2-13, and 22q11.2). It was also established that small CNVs tend to encompass high risk ASD genes, whereas large CNVs encompass multiple ASD genes of modest effect size[87].

There are also converging lines of evidence for the role of rare SNVs in ASD risk. Exome sequencing of 175 autism trios found that whilst the overall rate of *de novo* SNV mutation was not significantly elevated from that expected by chance, the set of genes affected were highly biologically related to each other and/or had been previously identified as ASD/ID candidate genes[88]. The findings indicate the polygenic nature of ASD, whereby mutations in any of a large number of genes increases risk for ASD by 5- to 20-fold. Two concurrent exome sequencing publications further delineated the relationship between SNV and risk for ASD. Firstly, an integrated model, accounting for *de novo* and inherited variants, implicated a large number of genes in risk for ASD, particularly those involved in synaptic

formation, transcriptional regulation and chromatin-remodelling pathways[89]. Inclusion of inherited variants was thought to be critical, given the large number of ASD risk genes and the fact that many occur in loci with incomplete penetrance. Investigation of *de novo* mutation subtypes in 2,500 simplex families also revealed that ~43% of likely gene disrupting events (nonsense, frameshift and splice site mutations) in probands contribute towards risk of developing ASD[90].

1.6.3 Common genetic risk factors

Common genetic risk factors have also been shown to play an important role in the risk of developing ASD. Common genotyped SNPs are estimated to account for at least 20% of ASD liability[91], highlighting that a proportion of the genetic risk for ASD resides within common variants[92]. The ASD working group of the PGC, failed to find any genome-wide significant SNPs in a recently published discovery sample (7,387 ASD cases and 8,567 controls)[93]. Given the likelihood that these results reflect a lack of statistical power to detect an association, with the need for larger sample sizes, the authors also undertook a meta-analysis utilising two independent samples. This meta-analysis revealed a genome-wide significant association at the 10q24.32 locus and a significant concordance between the direction of effect of the top markers in the discovery sample and the two independent samples.

Recent work has further investigated the role of common polygenic variation in ASD and the relationship this has with rare variants. Firstly, data from 6,454 simplex families has revealed that polygenic risk for ASD is over-transmitted to probands, but not to unaffected siblings. This increase in parental polygenic risk transmission was still observed in the probands who had a contributing *de novo* variant[94]. Genome-wide common and rare variant genetic links between ASDs and typical variation in social and communication difficulties have also been identified in a large general population cohort[91]. This supports a continuum model, whereby multiple types of genetic risk for ASDs influence a continuum of traits, the severe tail of which could results in a diagnosis of ASD or another NDD.

1.7 Genetic overlap across neurodevelopmental disorders

Wide-scale genetic screening has revealed that many variants have broad – and often co-morbid – NDD phenotypes, which cross traditional diagnostic boundaries. For example, we have already seen that 16p11.2 duplications and mutations affecting the *NRXN1* gene are involved in the aetiology of DD/ID, schizophrenia and ASD. These cross-disorder findings are supported by the results of epidemiological studies, which have identified that increased general, rather than specific, risk for neuropsychiatric disorders is conferred by having a mother with schizophrenia, bipolar affective disorder (BPAD) or unipolar major depression[95]. Doherty and Owen have proposed that psychiatric disorders lie on a neurodevelopmental continuum, with ID being the most severe brain insult, followed by ASD, schizophrenia and mood disorders. The NDD gradient is indexed by the severity of the mutational load and cognitive impairment, which has implications for developing new methods of stratifying patients for research[96]. Cross-disorder research is a growing area of activity in psychiatric genetics and this section will provide an overview of key findings in the field.

1.7.1 Rare variant studies

The degree of phenotypic variability associated with pathogenic CNVs encompasses both early-onset neurodevelopmental disorders and adult-onset psychiatric disorders. Several CNVs, including those at the 1q21.1, 16p11.2, 17q12, and 22q11.2 loci, have been identified in individuals ascertained for different neurodevelopmental phenotypes[21]. All of the 11 robustly associated schizophrenia risk CNVs have been implicated in risk for other NDDs[97]. Likewise, all of the six main CNV risk loci for ASD have been shown to confer risk for ID[87]. I have selected two recurrent CNV loci, which are commonly associated with risk for multiple NDDs, for detailed discussion – 22q11.2 and 16p11.2 deletions and duplications.

Clinical presentation of the 22q11.2 deletion syndrome (also known as velocardiofacial syndrome or DiGeorge syndrome) is highly variable, with more than 180 clinical features described[98]. The 22q11.2 deletion syndrome is the strongest known risk factor for psychotic disorders, with prevalence rates as high as 30%[99]. It has been estimated that up to 60% of children with 22q11.2 deletions meet the criteria for at least one psychiatric diagnosis, notably ADHD and anxiety disorders[100]. The

22q11.2 duplication syndrome shares some features with the reciprocal deletion, however the phenotype is generally mild in comparison with the deletion and familial transmission is frequently observed[101]. Psychiatric and behavioural problems include ASD, which occurs in approximately 14-25% of carriers. Interestingly, psychosis phenotypes, which are common in the deletion carriers, are infrequently observed and the 22q11.2 duplication has been proposed as a protective variant for schizophrenia[102].

The 16p11.2 deletion syndrome is commonly associated with ID, macrocephaly, seizures and obesity. Zufferey *et al.* collected phenotypic data on 285 16p11.2 deletion carriers and found that their full scale IQ is typically two standard deviations lower than non-carrier relatives. Furthermore, more than 80% of 16p11.2 deletion carriers exhibit psychiatric disorders, of which 15% of children present with ASD[103]. Moreno-De-Luca *et al.* investigated the clinical variability of *de novo* 16p11.2 deletions in 56 individuals, undertaking a novel family-based study design. They identified significant parent-proband correlations on measures of cognition, social behaviour and neuromotor performance, such that the impairments in probands were most profound in domains where their parents are already showing lower quantitative performance[104]. This provides support for a model whereby family background has an effect on phenotypic variability. The clinical features of the 16p11.2 duplication syndrome exhibit a mirror phenotype to the deletion, for example, the duplication has been associated with microcephaly and a reduced body mass index. IQ testing in the duplication carriers revealed a higher variance than in deletion carriers[105]. A meta-analysis of 16p11.2 duplication studies found that the disorder confers a 14-fold increased risk of psychosis and a 16-fold increased risk of schizophrenia[106].

Kirov *et al.* assessed the penetrance of CNVs at previously established risk loci for NDDs. The penetrance for schizophrenia was compared with a group of early-onset developmental disorders – DD, ASD, and congenital malformations. Almost all CNVs had higher rates in the early-onset developmental disorders group. The average penetrance, for developing any of the associated disorders, was 41%. It was concluded that most of the CNVs are highly pathogenic and are therefore more likely to cause earlier-onset disorders rather than schizophrenia[107].

There is, however, evidence to suggest a degree of specificity in the NDD risk conferred by separate CNV loci. Moreno de Luca *et al.* investigated the prevalence of deletions and duplications at four commonly implicated NDD risk loci (7q11.23, 15q11.2-13.1, 16p11.2 and 22q11.2) in ID, ASD, schizophrenia and controls. They found that some CNVs show increased risk for all disorders (e.g. 16p11.2 duplications) whereas others showed a degree of specificity (e.g. 22q11.2 duplications are rarely observed in schizophrenia and 7q11.23 deletions show risk for ID but are rarely observed in ASD)[108]. This pattern of asymmetric risk provides support for both shared and distinct genetic aetiology, likely due to differing dosage effects and neuropathological mechanisms at different CNV loci.

Evidence has also emerged from exome sequencing studies for shared genetic aetiology between NDDs. Analysis of 57 schizophrenia trios identified *de novo* mutations in numerous genes that have been previously implicated in risk for ASD and ID, with an enrichment in genes involved in chromatin modification[109]. Singh *et al.* undertook a large WES in schizophrenia patients (4,264 cases and 1,077 patient-parent trios) and controls (n=9,343), identifying a significant enrichment of rare loss of function variants in the SET Domain Containing 1A (*SETD1A*) gene in schizophrenia patients. The *SETD1A* gene encodes a methyltransferase involved in the catalysis of lysine residues in histone H3. Interestingly, seven out of the ten schizophrenia patients identified with this variant also had learning difficulties and further variant carriers were identified by investigating severe developmental disorder cohorts[110]. Large exome sequencing studies in ASD have revealed similar findings, with 107 genes strongly enriched for ASD overlapping with 21 candidate genes for intellectual disability, 3 for epilepsy and 17 for schizophrenia[111].

1.7.2 Common variant studies

Results of GWAS studies have revealed that SNPs at the same genetic loci harbour variants which are associated with multiple seemingly clinically distinct traits, for example the calcium channel, voltage-dependent, L-type, alpha 1C subunit (*CACNA1C*) gene has been implicated in risk for both schizophrenia and BPAD[112]. A large schizophrenia GWAS study found an association between schizophrenia and BPAD samples, whereas no association was found with six non-psychiatric disorder

samples[113]. Results from the latest ASD GWAS also identified a strong genetic correlation between ASD and schizophrenia, and ASD GWAS results overlap with regions previously implicated in risk for schizophrenia[93].

Formation of the cross disorder group of the PGC has facilitated further analysis of GWAS data from five psychiatric disorders (schizophrenia, BPAD, ASD, major depressive disorder (MDD), and ADHD). Genetic correlation using common SNPs was high between schizophrenia and BPAD. Three disorders – schizophrenia, BPAD, ADHD – also showed moderate associations with MDD[114]. Also joint analysis, including data from all five psychiatric disorders, revealed four genome-wide significant loci, including the *CACNA1C* gene – which also showed a significant association when only considering the schizophrenia group[115].

1.7.3 Healthy control studies

The role of pathogenic CNVs in NDD aetiology has been further complicated by the finding that virtually every CNV that is associated with a psychiatric disorder is present at a low frequency in populations that are ascertained as healthy controls[21]. Stefansson *et al.* found that population controls who carry these pathogenic CNVs have a global assessment of functioning score 0.7 standard deviations (SDs) lower than population controls who don't carry the pathogenic CNV. Cognitive testing revealed that these controls perform at an intermediary level between schizophrenia patients and population controls, both of whom did not carry a pathogenic CNV. Additionally, pathogenic CNVs do not all affect the same cognitive domains, so factors influencing cognitive deficits vary from one CNV region to another. Structural MRI was also performed on carriers of 15q11.2 (BP1-BP2) CNVs and reciprocal changes in the same anatomical regions were identified for deletion and duplication carriers, showing – for the first time – a dosage dependent effect of the CNV on brain structure[116].

The role of CNVs in a general population cohort was also examined by Männik *et al.* using a random sample of individuals (N=7,877) from the biobank of Estonia. Phenotypic analysis of the 56 carriers of syndrome-associated CNVs revealed that these individuals had cognitive and psychiatric co-morbidities and low educational attainment[117]. Further insight into the cognitive impact of pathogenic CNVs in control

populations has been provided by other analyses in large population datasets. The UK Biobank is a large general population dataset of ~500,000 adults that includes genotyping, imaging and cognitive data[118]. Kendall *et al.* assessed the cognitive performance of carriers of 53 NDD CNVs (total N=1,571). The majority of CNV carriers had impaired performance on cognitive tests, as well as lower educational and occupational attainment. Similar to the findings from Stefansson *et al.*, the cognitive deficits were modest as compared to the deficits observed in schizophrenia patients[119].

1.7.4 Cross-disorder analyses aid gene discovery

The complexity with case-control association analyses is that huge sample sizes are required to detect very rare variants. One approach which has been adopted in cross-disorder research is to firstly identify clinically significant CNVs relevant to broader NDDs then test for these in single disorder cohorts. This approach has been successful in identifying new rare variants for ASD[120] and for schizophrenia[97]. Gonzalez-Mantilla *et al.* undertook joint analysis of genetic and phenotypic data encompassing six NDDs (DD/ID, ASD, ADHD, schizophrenia, BPAD, and/or epilepsy). They focused on pathogenic loss of function variants affecting at least two unrelated individuals (N=1,960). The cross-disorder approach enabled the addition of 33 genes to the knowledge base and increased the evidence level for 18 genes. The study found significant phenotypic heterogeneity, with the majority of genes (32.8) being associated with two disorders and two genes, *NRXN1* and *PARK2*, being associated with all six disorders. Although the picture was complex, with certain genes showing enrichment for particular disorders[121].

1.8 Thesis rationale and overview

In summary, pathogenic CNVs are rare at the individual locus level, but are collectively common risk factors for developing a range of NDDs. Almost all pathogenic CNVs identified thus far are most frequent in patient groups with severe early-onset developmental disorders[107]. They also present the greatest known genetic risk factors for psychiatric disorders to date[1]. Some of these CNVs show increased risk for various psychiatric disorders, whereas others show a degree of specificity (such as the 22q11.2 deletion CNV increasing risk for psychosis)[108].

Historically large-scale CNV research has predominantly taken place in paediatric cohorts referred for clinical genetic testing[45,46]. The limitations of this approach are that the available phenotype data is limited to the reason for the referral to genetic services. The cohort is both weakly phenotyped, with important phenotypic information missing, and encompasses a broad range of severe developmental phenotypes. Furthermore, the developmental nature of the cohort means that later-onset psychiatric phenotypes cannot be captured at the point of analysis. Other CNV research has taken place in cohorts with a single psychiatric diagnosis[59].

The discovery of CMA technologies has changed the face of clinical genetic testing for children presenting with various developmental disorders. CMA is now being applied as one of the first-line routine genetic investigations in many healthcare systems[25]. This genotype-first approach is particularly advantageous when the patient presents with a non-specific phenotype and there are no suspected syndromic forms of ID for targeted genetic testing. However, genetic investigations have not been a routine part of the assessment of adults with ID presenting to psychiatric services[122]. Clinical genetic testing is not routine for any other psychiatric disorders – although implementing CNV testing in schizophrenia is an item of recent debate[123].

Following the recognition that both rare and common variants implicated in risk for NDDs transcend clinical diagnostic boundaries, there has been a rise in cross disorder research – whereby different single disorder cohorts are tested for the same genetic variants. However, there remains a deficit of research on co-morbid phenotypes – whereby individuals present with more than one psychiatric diagnosis. Prior to the work described in this thesis no investigations of pathogenic CNVs in adults with co-morbid NDDs have taken place. Researching this population has the advantage that the typical age of onset for psychiatric disorders has passed. This has the potential to reveal novel CNVs implicated in co-morbid CNV risk and/or identify previously established pathogenic CNVs that have a higher frequency in co-morbid phenotypes.

The work presented in this thesis aims to research this novel population; with Chapter 2Chapter 1 taking a clinical perspective, with a survey of psychiatrists involved in their care; Chapter 3 and Chapter 4 undertaking CMA to investigate the type and architecture of pathogenic CNVs, as compared to other populations in the literature;

Chapter 5 undertaking follow-up in-depth phenotyping of individuals with 2q13 CNVs to inform genotype-phenotype correlations of individuals with these rare CNVs; and, finally, Chapter 6 investigating the relative CNV burden as compared to other cohorts.

Chapter 2 Survey of intellectual disability genetic testing practices in child and adult psychiatry

Parts of this chapter have been adapted from the following published journal article:

Wolfe, K., Stueber, K., McQuillin, A., Jichi, F., Patch, C., Flinter, F., Strydom, A. & Bass, N. (2017) Genetic testing in intellectual disability psychiatry: Opinions and practices of UK child and intellectual disability psychiatrists. *Journal of Applied Research in Intellectual Disabilities*. 1–12. PMID: 28833975.

I undertook all statistical analyses, drafted and revised the manuscript. Study design, survey development and data collection were conducted by: KS, AM, CP, FF, AS, and NB. FJ advised on the statistical analyses.

2.1 Introduction

A definition of ID and prevalence estimates have previously been described in section 1.4.1. The rate of psychiatric disorders among individuals with ID is approximately 4 to 5 times higher than in the general population[124]. Despite this, psychiatric disorders in ID are often underdiagnosed. One reason for this is diagnostic overshadowing, whereby symptoms which would normally be attributed to a psychiatric disorder are instead viewed as being a component of the existing ID diagnosis. The lack of appropriate diagnostic criteria and paucity of suitable assessment measures also serve to compound psychiatric diagnoses in ID. Individuals with ID have been shown to experience poorer health, compared to the general population, and are subject to health inequalities[125]. Increasingly, ID services are adopting a bio-psychosocial approach, considering biological, psychological and social causes, in the assessment and treatment of mental health problems in individuals with ID[126]. Personalised care and health action plans, including annual health checks, are also being put into place to help prevent health inequalities[127].

An overview of the developments in genetic testing for ID has been covered in 1.2-1.4.2. Investigation of the cause of DD/ID predominately occurs at onset in childhood and there is no formalised system of diagnostic review. In childhood patients may

initially present to a Paediatrician or a Child and Adolescent Psychiatrist. Whereas, in adulthood patients present to ID Psychiatrists, a subspecialty of psychiatry with specialist consultant accreditation, or other treating clinicians, such as neurologists – for instance if the cause of epilepsy is being investigated in an individual with ID. ID psychiatrists sit within learning disability services, which usually comprise an interdisciplinary community-based health and social care team. Services generally have a high threshold for eligibility (IQ<70 as well as significant impairment of functioning, which has been present from childhood). Referrals to services are typically made by general practitioners (GPs), who have a key role in managing co-morbid physical illnesses, although some services also accept referrals from patients, carers, or social services[126]. There are 23 National Health Service (NHS) Regional Genetics Centres (RGCs) across the UK, to which treating clinicians can make referrals for genetic testing services for patients and their families.

Inequalities in access to genetic testing have been reported across different medical specialisms and there is a shift towards mainstreaming of genetic practices, with clinicians being encouraged to order genetic tests directly to improve uptake and access[128]. It has previously been discussed that recommendations have been made for CMA to be one of the primary genetic tests for DD/ID, MCA and ASD[25]. However, evidence suggests that the uptake of testing for genomic disorders in routine clinical psychiatric practice has been slow[28]. However, little of known about the current practices of psychiatrists working with patients with intellectual disabilities.

2.2 Aims

This chapter aims to determine current knowledge and genetic testing practices of psychiatrists working with individuals with ID. Furthermore, potential differences between child and adult sub-specialties will be investigated.

2.3 Methods

Psychiatrists working in UK child and adolescent mental health services (CAMHS) and adult ID psychiatry services were surveyed as to their attitudes towards and current use of genetic investigations using an online survey.

2.3.1 Survey development and administration

The survey questions were developed through consultation with ID psychiatrists, a clinical geneticist, a genetic counsellor, a genetic researcher and a statistician. The primary themes under investigation were: attitudes towards genetic testing, ordering of genetic tests, confidence and training in the genetic testing process, concerns about genetic testing, feedback of genetic test results, and experiences of referring to genetic services. Following a pilot, a number of the questions were amended and the opportunity for open text responses was enabled. The 28-item self-administered survey was composed of yes/no responses, multiple choice Likert-scale questions, numeric outcomes and free text responses (see the associated publication for a copy of the survey questions). The survey was programmed not to force answers to questions and enable completion of the survey with missing responses. The survey was administered via the online service tool Survey Monkey (SurveyMonkey Inc. Palo Alto, California, USA).

2.3.2 Survey participants

The survey was distributed to members of the Faculty of Child and Adolescent Psychiatry and members of the Faculty of Psychiatry of Intellectual Disability via the Royal College of Psychiatrists mailing list. Psychiatrists were invited by email to participate in the survey. A participation reminder was sent after 1 week. Respondents were removed from the analysis if they were junior trainees or listed professions other than CAMHS psychiatry and adult ID psychiatry, if they lived outside the UK and if they had not seen any patients with DD/ID in the previous 12 months.

2.3.3 Statistical analysis

Quantitative statistical analyses were undertaken using IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp, Armonk, NY, USA). The analysis compared CAMHS psychiatrists (referred to henceforth as child psychiatrists) and adult ID psychiatrists (referred to henceforth as ID psychiatrists). Continuous outcome variables were analysed using a t-test where the data was normally distributed and Mann-Whitney U test for non-normally distributed data. The chi-squared test was utilised to test categorical outcome variables. Binary logistic regression was

undertaken to test univariable factors related to ordering a genetic test. For Likert scale responses the data was collapsed from 5 to 3 scale responses by merging 'strongly agree' with 'agree' and 'strongly disagree' with 'disagree', or 'very frequently' with 'frequently' and 'very rarely' with 'rarely'. The 5 category scale was retained for all statistical analyses and categories were only collapsed to simplify presentation of the descriptive findings. Participants were asked whether they felt confident in 8 aspects of the genetic testing process. To compare these confidence ratings a composite confidence score was generated by assigning the 5 point Likert scale responses a confidence value ranging from 1 for strongly disagree to 5 for strongly agree. These scores were then summed across the eight confidence measures to obtain an overall composite confidence score. Where analyses have been undertaken on a subset of the dataset due to missing values the number of respondents in the analysis has been indicated. A Bonferroni correction was applied and significance has been set at 0.006 to account for multiple testing.

2.3.4 Thematic analysis

Open text responses were thematically coded using Nvivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012). Three open text questions were included in the survey, focusing on the benefits and concerns of genetic testing in clinical practice. A word cloud was generated using all open text responses and word frequency analysis was undertaken using Nvivo. Word stemming was undertaken to combine variations of words from the same root (e.g. genetic and genetics). All words mentioned greater than 5 times, excluding common words, were inputed into Wordle[129] for the creation of the word cloud.

2.4 Results

Responses were received from 215 clinicians, comprising 121 child psychiatrists (56%) and 94 ID psychiatrists (44%); 56% were females (n=121) compared with males (n=94, 44%). The majority of respondents worked in England (n=170, 80%), followed by Scotland (n=29, 14%), Wales (n=9, 4%) and Northern Ireland (n=5, 2%). The majority of respondents worked in community teams (n=115, 57%) followed by specialist assessment inpatient units (n=23, 11%) and specialist referral centres (outpatient) (n=19, 9%). A further 46 respondents (23%) reported that they worked in

more than one of these settings. The median number of years working in the speciality was 10 (child psychiatrists 10 years, ID psychiatrists 11 years).

2.4.1 Sociodemographic factors

Available demographic factors, sex and place of work, were compared between the child and ID psychiatrist groups, see Table 2-1 Sociodemographic factors of child and ID psychiatristsTable 2-1. Chi-squared tests were undertaken and a significant difference between the groups was identified for sex (χ^2=17.06, p<0.000).

Table 2-1 Sociodemographic factors of child and ID psychiatrists

	Child psychiatrist	**ID psychiatrist**
Sex		
Female	83	38
Male	38	56
Place of work		
England	96	74
Scotland	15	14
Wales	5	4
Northern Ireland	4	1

2.4.2 Attitudes towards genetic testing

Respondents were asked to estimate the percentage of people with ID for whom genetic factors make a significant contribution towards the cause of their ID. Estimates from child psychiatrists (Mean=42%, SD=24.7, Range=2-100%) were comparable to those of ID psychiatrists (Mean=39.6%, SD=23.1, Range=3-90%) (n=206, Mean difference =2.4, 95% CI (-4.25, 8.1) p=0.48). The respondents were also asked to estimate the percentage of patients on their caseloads with an established genetic diagnosis. This serves as a proxy of the actual number of patients on the caseload, given that psychiatrists may misestimate this figure. Child and ID psychiatrists both estimated a low percentage of patients on their caseload as having an established genetic diagnosis. ID psychiatrists estimated a higher percentage of their own patients to have an established genetic diagnosis (Median=10%, Range=0-70%) compared to child psychiatrists (Median=5%, Range=0-100%), (n=205, U=3661.5, Mean rank = 120 vs Mean rank = 90, p=<0.001).

2.4.3 Ordering of genetic tests

More ID psychiatrists (77%), compared with child psychiatrists (56%), had ordered a genetic test in the last 10 years (n=162, χ^2=8.08, p=0.004). Respondent's estimates of the percentage of ID caused by genetic factors did not influence the likelihood of them ordering a genetic test (n=157, OR 1.01, 95% CI (0.99-1.03), p=0.19). The percentage of patients on respondents' caseloads with an established genetic diagnosis also did not affect the likelihood of ordering a genetic test (n= 156, OR 1.02, 95% CI (0.99-1.05), p=0.33).

2.4.4 Confidence in the genetic testing process

Respondents were asked how confident they felt in eight aspects of the genetic testing process, as presented in Table 2-2. Child psychiatrists had a lower average total confidence score (Mean=22.1, SD=6.8) in comparison with ID psychiatrists (Mean=27.4, SD=5.5). (n=186, Mean difference=5.3, 95% CI (3.42, 7.1), p=<0.001). In comparison with child psychiatrists, ID psychiatrists agreed that they were confident in: knowledge of genetic tests (69% vs 29%); assessing for dysmorphic features (63% vs 47%); ordering (47% vs 24%) and interpreting genetic tests (35% vs 12%); genetic counselling (22% vs 12%) and feeding back test results to patients (64% vs 32%) and their families (68% vs 34%).

Table 2-2: Self rated confidence scores of child and ID psychiatrists (n=186) in eight areas of the genetic testing process

		Disagree	Neither agree nor disagree	Agree
Knowledge of genetic tests	Child	52 (50%)	23 (22%)	30 (29%)
	ID	11 (14%)	14 (17%)	56 (69%)
Assessing for dysmorphic features	Child	34 (32%)	22 (21%)	49 (47%)
	ID	14 (17%)	16 (20%)	51 (63%)
Assessment of capacity to consent	Child	10 (10%)	12 (11%)	83 (79%)
	ID	2 (3%)	6 (7%)	73 (90%)
Ordering genetic tests	Child	55 (52%)	25 (24%)	25 (24%)
	ID	16 (20%)	27 (33%)	38 (47%)
Interpreting genetic test results	Child	70 (67%)	22 (21%)	13 (12%)
	ID	31 (38%)	22 (27%)	28 (35%)
Feedback to patients	Child	43 (41%)	28 (27%)	34 (32%)
	ID	13 (16%)	16 (20%)	52 (64%)
Feedback to family/carers	Child	41 (39%)	28 (27%)	36 (34%)
	ID	14 (17%)	12 (15%)	55 (68%)
Genetic counselling	Child	71 (68%)	21 (20%)	13 (12%)
	ID	36 (44%)	27 (33%)	18 (22%)

2.4.5 Concerns with the genetic testing process

Respondents were asked what their main concerns were in relation to the genetic testing process, see Table 2-3. Both child and ID psychiatrists agreed that lack of available treatment was one of the main concerns (58% vs 51% respectively). Another main concern was lack of resources, 54% of child and ID psychiatrists agreed that this was a concern. Implications for insurance were a greater concern for child psychiatrists in comparison to ID psychiatrists (50% vs 38%), whereas issues around counselling were a greater concern for ID psychiatrists (53% vs 43%).

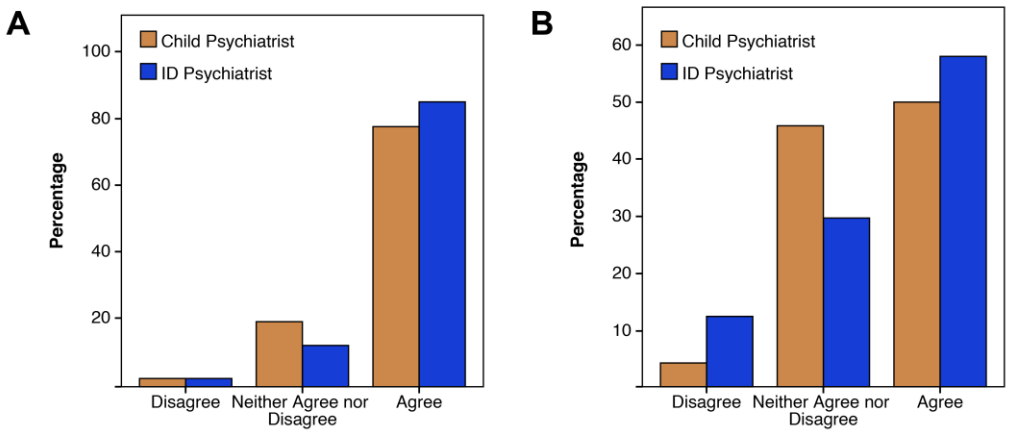
Table 2-3: Concerns child and ID psychiatrists (n=195) report in ten areas of the genetic testing process

		Disagree	Neither agree nor disagree	Agree
Stigma (-ve) of patients/families having a genetic diagnosis	Child	42 (39%)	39 (36%)	27 (25%)
	ID	39 (45%)	21 (24%)	27 (31%)
Lack of available treatment	Child	26 (24%)	19 (18%)	62 (58%)
	ID	27 (31%)	16 (18%)	44 (51%)
Lack of resources	Child	23 (22%)	26 (24%)	58 (54%)
	ID	21 (24%)	19 (22%)	46 (54%)
Implications for insurance	Child	23 (22%)	30 (28%)	53 (50%)
	ID	30 (35%)	24 (28%)	33 (38%)
Misuse of results	Child	22 (21%)	39 (37%)	45 (43%)
	ID	37 (43%)	17 (20%)	32 (37%)
Difficulty obtaining a family history	Child	40 (37%)	39 (36%)	28 (26%)
	ID	27 (31%)	19 (22%)	41 (47%)
Obtaining a sample	Child	38 (36%)	37 (71%)	31 (29%)
	ID	35 (40%)	23 (26%)	29 (33%)
Issues around counselling	Child	37 (35%)	24 (22%)	46 (43%)
	ID	22 (25%)	19 (22%)	46 (53%)
Issues around capacity to consent	Child	31 (29%)	32 (30%)	44 (41%)
	ID	36 (41%)	13 (15%)	38 (44%)

2.4.6 Feedback of results and clinical management

As seen in Figure 2-1 both child and ID psychiatrists agreed that a genetic diagnosis is more beneficial for family members than patients. In comparison with child psychiatrists, ID psychiatrists were more inclined to agree that a diagnosis is beneficial for family members (85% vs 78%) (Figure 2-1 A) and patients (58% vs 50%) (Figure 2-1 B).

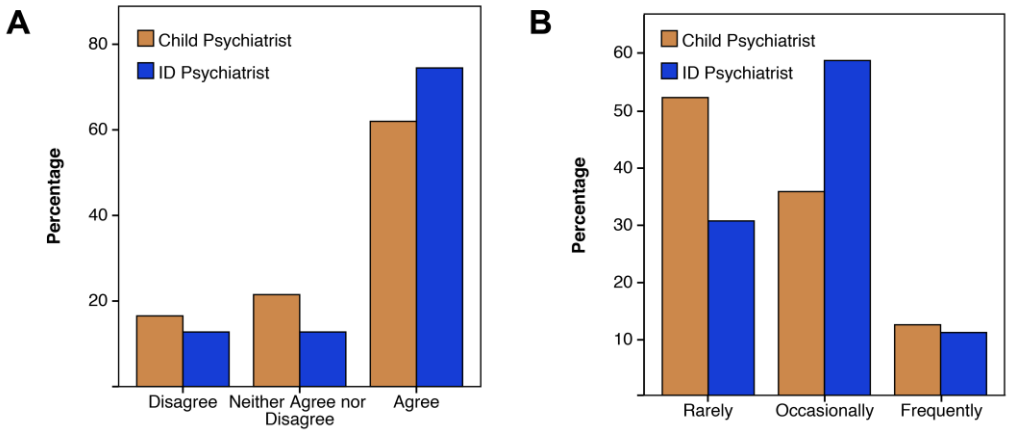
Figure 2-1: Percentage of child (n=72) and ID psychiatrists (n=81) who feel that a genetic diagnosis is helpful for family members (A) and patients with ID (B)



Respondents were also asked how they fed back results to their patients with ID, more than one response could be selected. Of the 146 respondents 8 (5%) had utilised videos, 20 (14%) had received input from speech and language therapists, 48 (33%) had used easy read materials, and 98 (67%) had used none of these aids. Responses were comparable for child and ID psychiatrists.

Figure 2-2 shows respondents' views and experiences of clinical management changes following genetic diagnoses. Respondents agreed that a genetic diagnosis would help with patient clinical management (75% ID vs 62% child) (Figure 2-2 A), however few agreed that they had seen frequent management changes in their patients (11% ID vs 12% child) (Figure 2-2 B).

Figure 2-2: Percentage of child and ID psychiatrists who feel that that a genetic diagnosis is helpful for patient management (A) (child n=121, ID n=94) and who report that genetic information has helped their patient management (B) (child n=73, ID n=82)



2.4.7 Referral to genetics services

Respondents were asked if they had ever ordered a genetic test or made a referral to a clinical genetics service. Those who had made a referral were also asked to estimate the number of referrals in last year. A significantly higher percentage of ID psychiatrists, compared with child psychiatrists had ordered a test or made a referral (90% vs 68%, n=214, χ^2=15.92, p =<0.001). ID psychiatrists also referred more patients per year to the genetics clinic compared with child psychiatrists (n=153, Range = ID 0-25, child 0-10, U= 2161.5, Mean rank = 87 vs Mean rank = 67, p=0.004).

Respondents were asked what the main reasons for referral to clinical genetics services were. Of the 155 respondents the most frequent reason for referral was presence of dysmorphic features (46% child, 57% ID) followed by intellectual disabilities (31% child, 38% ID). The least likely reason for referral was pharmacological treatment (2% both child and ID).

2.4.8 Service structure and training

Both ID and child psychiatrists agreed that closer links with regional genetics services would be helpful (83% vs 72%, n=197). Respondents were also in agreement that they would prefer to refer to a regional genetics service rather than order a genetic test

themselves (child 85%, 77% ID n=195). Finally there was a consensus that further training in genetics would be beneficial (child 71%, 66% ID, n=195).

2.4.9 Thematic analysis

Four main themes were identified from the 76 respondents who completed the open response questions comprising: family concerns, clinical management, and access to services and training.

Of the 23 respondents who reported family concerns, the most frequent benefits identified were relief from guilt and increased understanding of the patient's condition, followed by ability to access a support group and family planning. Respondents who discussed clinical management tended to mention the positive aspects, such as tailored medical and psychiatric interventions and clarification of syndrome specific behaviours. Only three respondents stated that they did not think a genetic diagnosis was helpful for clinical management. One respondent commented:

> "it is something of a paradox that the advances in the understanding of genetics and its potential impact upon our patient group has not translated into a significant increase in the use of genetic testing to help with diagnosis and care planning. I can only surmise that the social model of Disability as outlined in Valuing People has steered the diagnostic process away from genetic labelling".

Access to genetics services was mentioned by 22 respondents, who described problems with referring to genetics services and the variable levels of knowledge of professionals involved in the pathway. There was concern that psychiatrists, who have not specialised in genetics, do not have the skills to order genetic tests directly. Good working relationships with genetics services were said to be a valuable resource. Five child psychiatrists stated that they would defer to their paediatric colleagues to make decisions about genetic testing.

Several respondents felt that current training in genetics was insufficient and that training is not keeping abreast of technological advances. It was suggested that quick reference guides and screening tools would be valuable resources to support the decision making process. See Figure 2-3 for a word cloud of the most frequently used

words in the open text responses (results from both professional groups as responses were comparable for child and ID psychiatrists) and a summary of positive and negative opinions for each of the main themes.

Figure 2-3: Word cloud of words mentioned 5 or more times from open text responses with larger words mentioned more frequently. Positive and negative responses from the main themes are displayed in the text boxes

2.5 Discussion

The survey results reveal that the majority of child and ID psychiatrists working with patients with ID are already ordering genetic tests or making referrals to genetics services. However, there are several disparities in clinical genetic practices. In comparison with child psychiatrists, ID psychiatrists reported: a higher number of patients with genetic diagnoses, greater confidence in the genetic testing process, higher numbers of tests ordered and more patients referred per year to genetics services.

Respondents were asked to estimate the percentage of ID caused by genetic factors. The responses varied greatly, with some respondents estimating as low as 2% and others as high as 100%. Although both child and ID psychiatrists had similar mean

estimates (39.6% and 42%) of the percentage ID caused by genetic factors, these estimates were much higher than the estimated percentage of patients on caseloads with a known genetic diagnosis (median=10% ID and median=5% child). As we did not have access to the medical records, it is unclear whether psychiatrists accurately estimated the number of patients on their caseload with a genetic diagnosis. This would ideally be clarified in the first instance to test whether psychiatrists are misestimating the number of genetic diagnoses on their caseload, or if they do actually have a low proportion of patients with genetic diagnoses. Furthermore, it would be interesting to investigate whether genetic diagnoses are being reported to all professionals involved in the individual's care. For example, following genetic diagnosis what is the process to communicate this to the patient's GP and learning disability services, and for what proportion of cases is this being miscommunicated. This will be particularly important for individuals with ID and co-morbid diagnoses who are under the care of multiple medical professionals.

A high proportion of ID psychiatrists (77%) and just over half of child psychiatrists had directly ordered a genetic test. In comparison with child psychiatrists, ID psychiatrists were significantly more likely to order a genetic test and also referred more patients to the genetics clinic per year. This may have in part been a reflection of the ID psychiatrist's greater reported confidence in the genetic testing process. As evidenced in Table 2-2, ID psychiatrists were significantly more confident in all aspects of the testing process, apart from capacity testing which is likely to be more complex in adulthood.

The finding that adult ID psychiatrists are more confident in the genetic testing process is somewhat surprising, given that clinical genetic testing in DD/ID is only routine in childhood. In childhood clinical care plans are not yet in place, so it is more important to determine any underlying genetic diagnoses that may have implications for patient management. Furthermore, it is more pertinent at this stage for family members to receive information on recurrence risk to inform family planning decisions. Given these factors, one could assume that child psychiatrists have more experience, and therefore confidence, in genetic testing practices. However, one explanation for ID psychiatrists being more confident and ordering more genetic tests is due to the different structures of child and adult ID psychiatry services. Whereas, adult ID

psychiatrists are often the primary clinician involved in genetic referrals, in childhood pediatricians are often the first point of contact for patients and the child psychiatrists only become involved at a later stage. In support of this, a number of child psychiatrists reported in the qualitative analysis that they would defer to paediatric colleagues for opinions on genetic testing.

The survey highlighted a number of barriers to genetic testing in clinical ID services. Both child and ID psychiatrists reported that they were concerned about lack of available resources for genetic testing and the lack of available treatment options for patients. Interestingly child psychiatrists had specific concerns about implications for insurance. The Department of Health have released a moratorium extending until 2019 whereby the only genetic test required to be disclosed is for Huntington's disease on life insurance sums worth more than £500,000[130]. Therefore results from CMA should have no impact on insurance premiums and this misconception could be a barrier to clinicians ordering/referring for genetic testing. ID psychiatrists expressed concern about issues surrounding counselling. Feedback of genetic diagnoses to adults with ID could be more complex than feedback to parents of children with ID and this could be an important area for additional resources and research.

Although 77% of ID psychiatrists and 56% of child psychiatrists had directly ordered a genetic test, both child (85%) and ID (77%) psychiatrists agreed that they would prefer to refer to an RGC. However, links with NHS RGCs appeared to be variable. Some respondents reported good links with their local genetics services, whilst others felt that access to service was a barrier to referring for genetic testing. Both ID (83%) and child psychiatrists (72%) felt that better links with genetics services would be beneficial. Many of these clinicians felt that they do not have the knowledge or training to order genetic tests directly. This finding is supported by another survey, which found ID psychiatrists lacked adequate knowledge about genetic testing processes[131].

The majority of respondents expressed a wish for further training (71% child, 66% ID). Neither child and adolescent nor ID psychiatry curricula currently have learning objectives that specifically cover genetic disorders associated with ID[132]. The curricula also fail to cover the genetic work-up and basic genetic counselling skills that are required to take more of an active role in identifying and managing patients with

genetic disorders. However, there are several recent initiatives to improve the psychiatry curriculum. For example, the Gatsby-Wellcome initiative aims to ensure that training focuses more on scientific advances in basic and clinical neurosciences[133]. It is, therefore, hoped that future cohorts of psychiatrists will be more confident in utilising technological advancements in the assessment and management of their patients.

One of the reasons for undertaking genetic investigations is that a genetic diagnosis is likely to provide information about specific associated medical and psychiatric phenotypes and thus could improve treatment plans and clinical management for the patient. Whilst the majority of respondents felt that a genetic diagnosis would help with clinical management, fewer patients on their caseloads had a genetic diagnosis than they would expect and clinical management changes following genetic diagnoses were not frequently seen in practice. There are published medical guidelines available for several genetic disorders, for example via the Orphanet portal for rare diseases[134], and information guides on an extensive range of chromosomal disorders are available from the support group Unique[135]. It would be of interest for further research to investigate whether psychiatrists are aware of these guidelines when they receive a genetic diagnosis for their patient.

Another important consideration is that knowledge of behavioural phenotypes can place psychiatrists in a better position to deliver appropriate interventions and environmental adaptations. Whilst there is within syndrome variation it has been shown that certain behavioural features, such as repetitive and self-injurious behaviours, are more common in particular syndromes. There are also implications for health screening, for example gastro-intestinal problems are common in Cornelia de Lange syndrome and can exacerbate self-injurious behaviours[136]. A recent survey of ID professionals found that nine out of ten professionals interviewed felt that specific knowledge of a neurodevelopmental syndrome should play a key role in healthcare provision. A specific genetic diagnosis was particularly thought to prompt proactive screening for related physical and mental health problems, which is of particular benefit for patients with severe impairments[127]. One of the main challenges in practice is that individual syndromes are rare and psychiatrists are unlikely to care for many

individuals with the same disorder, although the overall burden of rare syndromic disorders is substantial.

Both child and ID psychiatrists agreed that receiving a genetic diagnosis was more beneficial for family members than for the patient. Research has shown that there is a benefit to mothers in receiving a diagnosis for a child with ID; however there is a lack of research as to the impact of a genetic diagnosis for adults with ID[137]. Several respondents reported that a diagnosis can help to alleviate guilt for family members, as well as increasing understanding of the patient's syndrome specific behaviours and enabling valuable access to support groups. It seems that respondents were able to report on a range of psychosocial benefits, which could indirectly improve patient management, however tangible changes in clinical decision making following a genetic diagnosis were less easy to define.

The limitations of this survey are that it was self-reported, which could have led to biases in estimations. There may have been an ascertainment bias in the clinicians who chose to respond to the survey. It may be that clinicians who already had an interest in genetics were more inclined to respond, or perhaps those clinicians who had more extreme views on genetics. If the respondents were self-selected as individuals who knew more about genetics, then the lack of confidence and knowledge identified in the survey would be even greater than initially indicated. There was a significant difference identified between the child and ID psychiatrist groups on the demographic of sex. It may be that there are sex differences in views and practices relating to genetic testing, which could inflate the differences between groups. No previous research on the impact of sex on clinician's genetic practices could be identified, however it is important to consider this as a potential factor in the interpretation of these findings. This survey specifically focused on psychiatrists, who are one of the medical specialists frequently in contact with patients with ID in the UK. These findings may not be generalisable to other countries where services are organised differently

2.6 Conclusion

Whilst a high number of child and ID psychiatrists appear to already be ordering genetic tests there remains a preference for referring directly to clinical genetics

services. Respondents highlighted several areas of the genetic testing process in which they particularly lack confidence, such as indications for testing, interpretation and feedback of genetic results. Child psychiatrists in particular felt less confident, ordered fewer genetic tests and referred fewer patients to genetic services.

Genetic investigations are continuing to advance at a very rapid pace, with WES and WGS beginning to enter clinical practice. For example, recent WES study identified 42% pathogenic DNMs in a cohort with severe developmental disorders[51]. In conjunction with other genetic investigations it is likely that a genetic diagnosis will be identifiable in a much higher proportion of patients with ID in the future. This should facilitate early diagnosis and tailored interventions for patients and their families. However, as the landscape of genetic investigations becomes more complex it is going to be a challenge for psychiatrists to keep pace of developments. Improvements in training and closer links with genetics services would appear to be key areas to address to meet this challenge. Phenotypic data from existing large DD/ID genetic investigations is limited, in that it is primarily focused on paediatric cohorts and is often restricted to the primary indication for CMA testing[138]. Thus, there is a need for research in adult ID psychiatry services, which offer a unique opportunity to delineate the neuropsychiatric phenotype of emerging NDD risk CNVs.

Chapter 3 The frequency and architecture of pathogenic CNVs in adults with idiopathic ID and co-morbid psychiatric disorders

Parts of this chapter have been adapted from the following published journal article:

Wolfe, K., Strydom, A., Morrogh, D., Carter, J., Cutajar, P., Eyeoyibo, M., Hassiotis, A., McCarthy, J., Mukherjee, R., Paschos, D., Perumal, N., Read, S., Shankar, R., Sharif, Saif., Thirulokachandran, S., Thygesen, J. H., Patch, C., Ogilvie, C., Flinter, F., McQuillin, A., Bass, N. (2017) Chromosomal microarray testing in adults with intellectual disability presenting with comorbid psychiatric disorders. *European Journal of Human Genetics*. 25: 66–72. PMID: 27650969.

I undertook and supervised participant recruitment (including psychiatric phenotyping), analysed the phenotypic data, and wrote and revised the manuscript for publication. Study design and analytical input from AS, AM, NB. CNV calling and pathogenicity ratings: DM and JC. Participant recruitment: PC, ME, AH, JMc, RM, DM, NP, SR, RS, SS, ST. Critical review of the manuscript was undertaken by all authors.

3.1 Introduction

An overview of the biological underpinnings of CNVs, CMA technologies, the clinical definition of ID, and the involvement of CNVs in psychiatric risk has been covered in Chapter 1.

Approximately 50% of adult ID is idiopathic or of unknown cause[36]. A large proportion of individuals with idiopathic ID have psychiatric co-morbidities[139]. There is evidence to suggest that there is an increased burden of CNVs as the severity of the neurodevelopmental phenotype increases[140], and that ID plus co-morbid mental disorders may contribute a higher CNV burden than ID alone[141]. Also, a recent investigation in a cohort of schizophrenia patients identified a significant increase in the yield of pathogenic CNVs as IQ decreased[142]. Despite this, the majority of the CNV research to date has focused on paediatric cohorts, whereby later-onset medical and psychiatric phenotypes cannot be reliably ascertained. Given that the investigation of

the cause of ID predominately occurs at diagnosis in childhood, there is a large cohort of adults, many with later onset psychiatric disorders, who have not had a diagnostic assessment utilising the latest genetic technologies[143].

3.2 Aims

The target population for this study was adults with ID (>18 years of age). Adults were selected, as phenotyping in children is limited by the fact that development is on-going and the typical age of onset for many psychiatric disorders is yet to be reached. We focused recruitment on adults with ID who also had challenging behaviours and/or one or more psychiatric diagnosis. This more severe co-morbid phenotype was selected as severity of phenotype has previously been associated with a greater CNV burden[140], and this population has never previously been investigated for pathogenic CNVs.

This study aimed to: (i) determine the yield of undiagnosed pathogenic CNVs in adults accessing ID psychiatry services in the UK, (ii) identify the architecture of CNVs involved in co-morbid NDD risk and the implicated genetic loci, (iii) describe the phenotypic presentation of adults with pathogenic CNVs, (iv) compare the psychiatric phenotype of patients with pathogenic and non-pathogenic CNVs.

3.3 Methods

3.3.1 Study design and participant recruitment

Ethical approval for the study was attained from the North Wales Research Ethics Committee West, reference 11/WA/0370. Recruitment was undertaken with the support of the Mental Health Research Network (MHRN), which is part of the National Institute for Health Research and aims to facilitate mental health research within the NHS. Recruitment took place at 32 NHS trusts and 1 non-NHS provider across England. Consultant Psychiatrists in Intellectual Disabilities acted as local principle investigators (PIs) at each site. PIs identified eligible participants from their caseloads based upon the study inclusion criteria. Inclusion criteria were: i) idiopathic ID; ii) one or more psychiatric diagnoses and/or significant challenging behaviours; and iii) over 18 years of age. Idiopathic ID was defined as no clear genetic or environmental cause of ID – to the best of the PI's knowledge.

Capacity to consent to the research project was assessed by the participant's Consultant Psychiatrist in accordance with the Mental Capacity Act 2005[144]. For eligible participants who were deemed to have capacity to consent to the study, information was provided in-person utilising Easy read information sheets and consent was provided via Easy read consent forms. This was undertaken by the patient's psychiatrist, by myself or by a Clinical Studies Officer or Research Nurse employed by the MHRN. All MHRN staff received training on the study protocol. In the absence of capacity consultees, primarily family members or carers, were identified to give advice as to the person's likely wishes regarding participation in accordance with the MCA guidance on nominating a consultee[145]. Participants, or their consultees, were given the option for the genetic results to be fed back to the treating psychiatrist, or another clinician involved in their care, such as the GP.

Clinical data including developmental, medical and psychiatric history (ICD-10 diagnoses)[146], was collected from an informant and/or medical records. General observations for dysmorphic features were made using a dysmorphology checklist, containing facial and bodily dysmorphisms. Measurements of height, head circumference and weight, where available, were also collected by the clinician or researcher. Photographs were taken (where consent was given) for corroboration by the study team.

Detailed psychiatric and behavioural phenotyping was undertaken using the Mini Psychiatric Assessment Schedule for Adults with Developmental Disabilities (Mini PAS-ADD) and Behaviour Problems Inventory – short form (BPI-S). The Mini PAS-ADD assesses psychiatric symptoms in seven diagnostic areas and provides threshold scores for symptoms that are likely to warrant a diagnosis in conjunction with a clinical assessment[147]. The BPI-S measures challenging behaviours over the previous six months and provides weekly frequency scores of behaviours on three domains, self-injurious behaviour, aggressive/destructive behaviour and stereotyped behaviour[148].

3.3.2 Genetic analysis and feedback

Participants were given the option to provide either a blood or saliva sample, 24% provided blood and 76% provided saliva. The saliva samples were collected using the

Oragene DNA collection kits (DNA Genotek Inc, Ontario, Canada). DNA extraction arrayCGH analysis was undertaken at the North East Thames Regional Genetics Service Laboratory on the Nimblegen 135K platform (Roche NimbleGen, Wisconsin, USA). Arrays were processed and CNVs were reported using in-house clinical diagnostic laboratory protocols, in keeping with the Association for Clinical Genetic Science (ACGS) best practice guidelines[149].

CNVs referred to as pathogenic include pathogenic causative CNVs and pathogenic susceptibility CNVs, both of which are thought to affect gene function in view of the associated phenotype[150]. As discussed in Chapter 1 there are many factors that influence variant categorisation. Some CNVs were classed as variants of unknown significance (VOUS) likely pathogenic. This categorisation typically arises when there is an existing single case report with similar CNV breakpoints and phenotype to the patient under investigation, or when a gene within the CNV interval has a compelling function relevant to the patient phenotype. In both instances there is insufficient evidence in published literature, or previous reported occurrences of the CNV, to warrant it being classified as pathogenic[14]. There is no uniform method of dealing with VOUS CNVs and some published manuscripts combine the VOUS likely pathogenic and pathogenic categories in their analyses[142,151]. However, VOUS likely pathogenic are not reported back to patients and their families. In order to take a conservative approach and focus on variants that are clinically reportable we did not include VOUS likely pathogenic in the pathogenic category. All pathogenic CNVs were added to the DECIPHER genome browser[152].

Pathogenic CNVs were fed back in writing to the participants' treating psychiatrist. The cytogenetic report and associated publications were provided alongside chromosomal disorder guides from the support group Unique where available[135]. There is a paucity of appropriate and accessible information for adults with ID receiving diagnoses. The study team developed easy to read materials to aid feedback for some of the clinically relevant CNVs. Psychiatrists also had the opportunity to speak with a member of the research team regarding the result prior to feeding this back to their patients and family members and/or carers. Referral to the RGCs for genetic counselling was recommended for all pathogenic results.

A feedback survey was sent to the treating psychiatrist of participant's with pathogenic CNVs one month after the genetic feedback was provided. The survey aimed to assess the clinician's experience of feeding back the genetic test result and to determine whether there had been any clinical management changes following the genetic diagnosis, see appendix for details of the questions included in the survey.

3.3.3 Statistical analysis

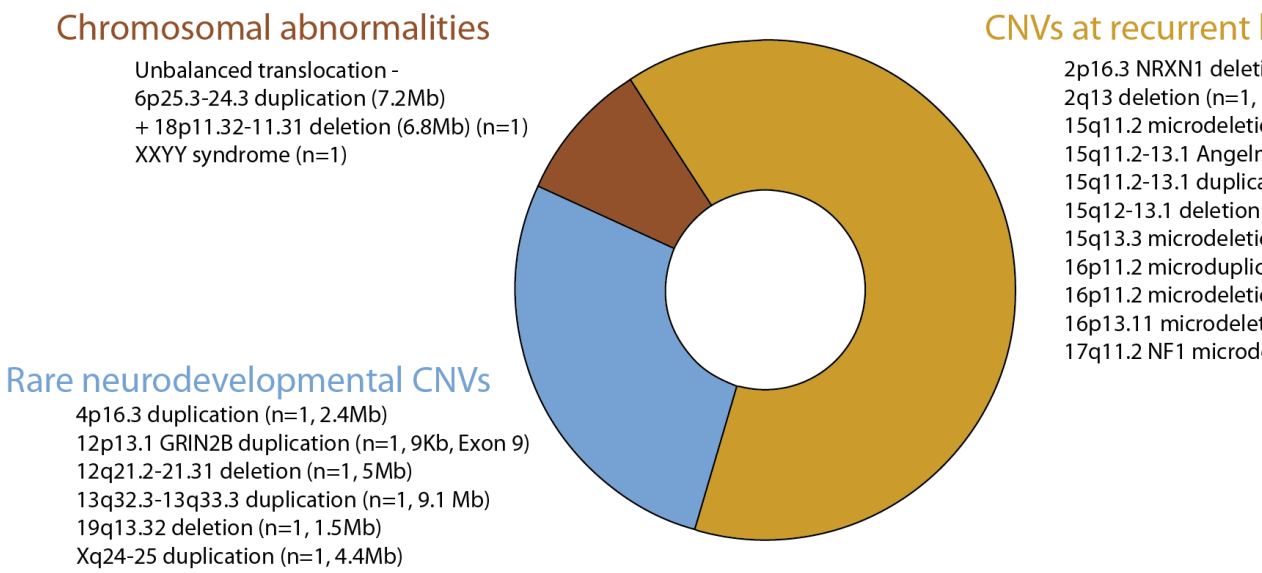
Statistical analyses were undertaken using IBM SPSS Statistics for Windows, Version 22.0. Univariate binary logistic regression was performed using Mini PAS-ADD thresholds and history of involuntary in-patient admission, including forensic in-patient section, as predictor variables. The binary outcome variable was presence or absence of a pathogenic CNV.

3.4 Results

3.4.1 Frequency of pathogenic CNVs

202 adults with idiopathic ID and co-morbid psychiatric disorders/challenging behaviours were recruited to the study (63% male; mean age 37 years, range 18-78 years; 74% White British). The yield of pathogenic CNVs, including chromosomal abnormalities, was 11% (22/202). A further 62% of participants had a least one CNV classed as VOUS (126/202) and 27% (54/202) had likely benign CNVs only. An overview of pathogenic CNVs is presented in Figure 3-1, with detailed genetic and phenotypic

data presented in Figure 3-1: Overview of pathogenic CNVs identified in a sample of 202 adults with idiopathic intellectual disabilities and co-morbid psychiatric disorders



Mb = megabases. Kb = kilobases, N= number of participants, BP = breakpoints

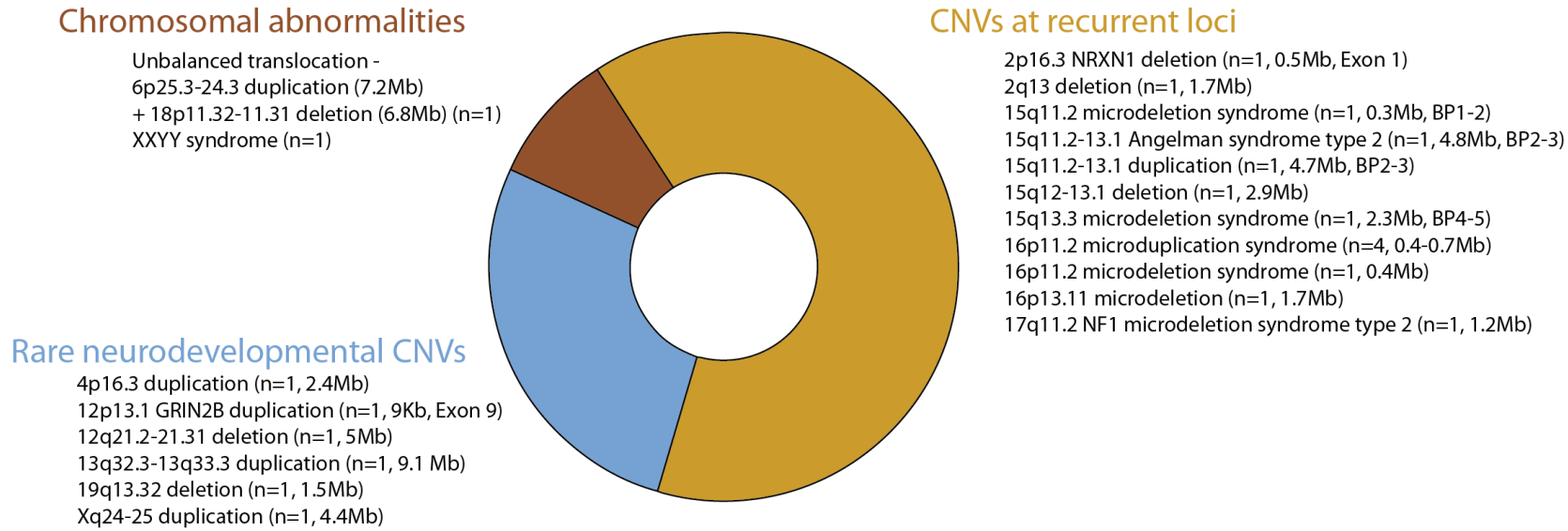
Table 3-1 and Table 3-2. A comparison of psychiatric diagnoses, subclinical symptoms and section history for likely pathogenic versus likely benign (including VOUS) CNVs is provided in Table 3-3. There were 21 participants on a forensic in-patient section (section 37 or section 37/41) at the time of recruitment and no other participants had a forensic section history. In all, 6/21 (28.6%) forensic in-patients carried pathogenic CNVs compared with 16/181 (8.8%) in participants not on a forensic in-patient section. Thus the proportion of pathogenic CNV carriers in forensic in-patients was higher with an OR of 4.1 (95% CI 1.40–12.04, P=0.01). However, if a Bonferroni correction was applied to account for multiple testing the level of significance would be set to p=0.007, therefore this finding would not survive a multiple testing correction.

3.4.2 Type of pathogenic CNVs

The majority of CNVs (64%, 14/22) were observed in regions of the genome prone to recurrent CNV, where pathogenic CNVs have previously been described for various

NDDs. The second largest group was very rare CNVs, not occurring at recurrent loci, which were either large in size (range 1.5-9.1Mb) or in known neurodevelopmental genes (6/22, 27%), finally two participants had large chromosomal abnormalities which are also detectable by aCGH (9%). For an overview see Figure 3-1.

Figure 3-1: Overview of pathogenic CNVs identified in a sample of 202 adults with idiopathic intellectual disabilities and co-morbid psychiatric disorders



Mb = megabases. Kb = kilobases, N= number of participants, BP = breakpoints

Table 3-1: Pathogenic CNVs and psychiatric phenotypes in adults with intellectual disabilities referred to psychiatric services

Decipher ID	Cytogenetic Location	Gain/ Loss	Chromosomal region	Gender	Age	Level of ID	Psychiatric history	Mini PAS-ADD	BPI-S			Ethnicity
327138	2p16.3	Loss	chr2:51,196,189-51,745,529	Female	21	Moderate	PD	PSY, UNS	-	-	+	White (British)
327136	2q13	Loss	chr2:111,391,616-113,103,446	Male	19	Mild	ASD	ANX, PSY	+	+	+	White (British)
327134	4p16.3	Gain	chr4:116,640-2,593,260	Male	33	Mild	ASD, ALC, FOR	N/A	N/A			White (British
327131	6p25.3-24.3[1]	Gain	chr6:195,429-7,392,549	Male	49	Mild	ANX, DEP	HYP, PSY, ASD	-	+	+	White (British)
327120	12p13.1	Gain	chr12:13,754,549-13,762,809	Female	68	Moderate	CB	None met	+	+	-	White (British)
327126	12q21.2-21.31	Loss	chr12:79,534,629-84,535,827	Male	31	Mild	SCZ, ALC, FOR	HYP, PSY	-	-	-	Black (African)
327125	13q32.3-33.3	Gain	chr13:100,465,759-109,578,071	Male	21	Mild	ASD, ADHD, FOR	None met	-	-	-	White (British)
327128	15q11.2	Loss	chr15:22,759,710-23,071,809	Female	42	Moderate	BP	DEP, ANX, HYP, OCD, PSY, UNS	-	-	+	White (British)
327127	15q11.2-13.1	Loss	chr15:23,643,100-28,519,140	Female	33	Severe	ANX	ANX	+	-	+	White (British)
327124	15q11.2-13.1	Gain	chr15:23,810,480-28,519,140	Male	22	Mild	ASD, BP, PD, ALC, FOR	N/A	-	+	-	White (British)
327123	15q12-13.1	Loss	chr15:26,587,699-29,576,869	Male	28	Severe	ASD, ADHD	OCD, ASD	+	+	+	White (Other)
327137	15q13.2-13.3	Loss	chr15:30,461,189-32,804,210	Male	25	Mild	ASD[2]	None met	-	-	-	White (British)
327122	16p11.2	Gain	chr16:29,443,979-30,192,560	Female	33	Mild	AFF	DEP, ANX, PSY	-	-	-	White (British)

327119	16p11.2	Gain	chr16:29,746,320-30,093,460	Female	27	Mild	ASD, DEP, OCD	DEP, OCD	-	-	-	White (Other)
327121	16p11.2	Gain	chr16:29,746,320-30,093,460	Female	62	Mild	DEP (psychotic)	PSY	-	-	-	White (British)
327133	16p11.2	Gain	chr16:29,746,320-30,192,560	Male	21	Mild	ASD, ADHD, FOR	None met	-	+	-	White (British)
327135	16p11.2	Loss	chr16:29,746,320-30,192,560	Male	19	Mild	None recorded	DEP, ANX, HYP, OCD, PSY	-	-	+	White (British)
327130	16p13.11	Gain	chr16:14,892,210-16,616,420	Female	45	Moderate	CB	N/A	N/A			White (British)
327132	17q11.2	Loss	chr17:29,068,140-30,281,850	Male	57	Severe	ASD, CB	ANX, HYP, OCD	+	+	+	White (British)
327131	18p11.32-11.31[1]	Loss	chr18:141,489-6,964,200	Male	49	Mild	ANX, DEP	HYP, PSY, ASD	-	+	+	White (British)
327129	19q13.32	Loss	chr19:45,741,741-47,268,131	Female	58	Mild	BP	DEP, PSY	-	+	-	White (British)
327139	Xq24-25	Gain	chrX:118,883,247-123,283,287	Male	57	Moderate	CB	None met	-	+	+	Black (Caribbean)
Not applicable	XXYY	Gain	Full chromosomal duplications[3]	Male	59	Mild	DEP, FOR	None met	-	+	-	White (British)

Chromosomal region coordinates in hg19 using the HGVS standard nomenclature; Size, size of CNV in kilobase pairs; Age, age at date of recruitment; Level of ID, taken from medical records in accordance with the UK ICD-10 diagnostic system: 50-69: Mild, 35-49: Moderate, 20-34: Severe, Mini PAS-ADD, Psychiatric Assessment Schedules for Adults with Developmental Disabilities thresholds met (in last 2 years); BPI-S, The Behaviour Problems Inventory-Short Form (Self-injurious behaviour, aggressive/destructive, stereotyped) items scored as + when behaviour occurs at least weekly; PD, Personality Disorder; PSY, Psychosis; UNS, Unspecified Disorder; ASD, Autistic Spectrum Disorder; ANX, Anxiety Disorder; ALC alcohol abuse; FOR on forensic in-patient section; N/A, Not Available; DEP, Depression; HYP, Hypomania/Mania; CB, Challenging Behaviour; SCZ, Schizophrenia; ADHD, Attention Deficit Hyperactivity Disorder; BP, Bipolar Disorder; OCD, Obsessive Compulsive Disorder; AFF, Schizoaffective Disorder; [1] CNVs in the same participant; [2] ASD traits only; [3] Confirmed by qPCR

Table 3-2: Medical phenotype and dysmorphic features in participants with pathogenic CNVs

Decipher ID	Cytogenetic Location	Age	Medical history	Head circumference (Cm)	Height (Cm)	Dysmorphic features
327138	2p16.3	21	Asthma	55.9	169	Abnormal facial shape, dental crowding
327136	2q13	19	Recurrent ear infections, urinary reflux, facial nerve palsy	N/A	N/A	Cranial abnormality
327134	4p16.3	33	Constipation	59	179	Upward slanting palpebral fissures, macrotia, prognathia
327131	6p25.3-24.3 + 18p11.32-11.31	49	Bilateral sensorineural hearing impairment, epilepsy, psychogenic polydipsia, hypogonadism, arthritis, osteoperosis, dysphagia	56	157	Facial asymmetry, abnormal facial shape, dental crowding, abnormality of the fingers, large ears, cranial abnormality
327120	12p13.1	68	Pes cavus (requiring callipers), gastric reflux	57	152	No gross dysmorphology
327126	12q21.2-21.31	31	None recorded	64	169	Hypertelorism, depressed nasal bridge, wide nasal bridge, low set ears, microtia
327125	13q32.3-33.3	21	Epilepsy, shuffling gait, bradykinesia	59	167	Low set ears, abnormality of the hand
327128	15q11.2	42	Recurrent urinary tract infections	55	N/A	Abnormality of external nose, abnormalities of the fingers
327127	15q11.2-13.1	33	Hypotonia (infant), epilepsy (infancy), probable diplopia, abnormality on neuroimaging	53	161	Upward slanting palpebral fissures, prognathia, protruding tongue, hypopigmentation of the skin
327124	15q11.2-13.1	22	None recorded	N/A	182	No gross dysmorphology
327123	15q12-13.1	28	Epilepsy (grand mal and absence seizures)	55.8	170	Cranial abnormality, facial asymmetry
327137	15q13.2-13.3	25	Type II diabetes	60	173	No gross dysmorphology
327122	16p11.2	33	Myopia	57.9	173.4	No gross dysmorphology

327119	16p11.2	27	Renal problems (childhood), menorrhagia, anaemia (severe), onychogryphosis	54.2	161.8	No gross dysmorphology
327121	16p11.2	62	Jaundice (childhood), epilepsy (childhood), type II diabetes, constipation, glaucoma	53.2	172.5	No gross dysmorphology
327133	16p11.2	21	Insulin dependent diabetes, hypercholesterolaemia	59	188	N/A
327135	16p11.2	19	Seizures (infancy), acne	59	162	Tapering fingers
327130	16p13.11	45	Hypertension, type II diabetes, constipation, asthma, obesity	54	147	Abnormal facial shape, microtia
327132	17q11.2	57	Pacemaker in situ (long Q-T syndrome), hypothyroidism, hypercholesterolaemia, cataracts	60	179	Abnormalities of the fingers
327131	18p11.32-11.31	49	Bilateral sensorineural hearing impairment, epilepsy, psychogenic polydipsia, hypogonadism, arthritis, osteoperosis, dysphagia	56	157	Facial asymmetry, abnormal facial shape, dental crowding, abnormality of the fingers, large ears, cranial abnormality
327129	19q13.32	58	Epilepsy, incontinence, Cataracts, broad based gait	55	148	Short upturned nose
327139	Xq24-25	57	Shuffling gait, bradykinesia	61	174	Abnormality of the skull, abnormality of the eyelid, abnormal nasal morphology, abnormalities of the fingers
Not applicable	XXYY	59	Hypertension, hypercholesterolaemia, type I diabetes, absent kidney, constipation, cataracts, anaemia, asthma, osteoarthritis	58.5cm	176cm	N/A

Age, age at date of recruitment ; Medical history was collected using a standard pro-forma from informants (family members and/or carers) and medical records where available. Dysmorphic features were recorded using a standard pro-forma by researchers carrying out the assessment. Only dysmorphic features obvious on general observation in normal clothing were documented no structured physical examination was undertaken. Dysmorphic features were corroborated by examination of accompanying photographs by the study team. Not all participants consented to be photographed; N/A, Not Available

Table 3-3: Psychiatric phenotype (ICD-10 diagnoses, Mini PAS-ADD thresholds and section history) for pathogenic and benign CNVs

	Total in sample (%) N=202	Pathogenic CNV group (%) n=22	Benign CNV group (%) n=180	*P*-Value[a]
ICD-10 Diagnosis				
Psychosis	49 (25%)	3 (14%)	46 (26%)	-
Bipolar disorder	23 (11%)	3 (14%)	20 (11%)	-
Depressive episode	62 (31%)	4 (18%)	58 (32%	-
Other anxiety disorders	45 (23%)	2 (9%)	43 (24%)	-
Hyperkinetic disorder	21 (10%)	3 (14%)	18 (10%)	-
Pervasive developmental disorder	68 (34%)	8 (36%)	60 (33%)	-
Mini PAS-ADD thresholds				
Psychosis	72 (36%)	9 (41%)	63 (35%)	0.63
Hypomania/Mania	33 (16%)	5 (23%)	28 (16%)	0.41
Depressive disorder	76 (38%)	5 (23%)	71 (39%)	0.12
Anxiety disorder	80 (40%)	6 (27%)	74 (41%)	0.84
Obsessive compulsive	55 (27%)	5 (23%)	50 (28%)	0.59
Mental Health Act Section History				
Previous history of involuntary admission	45 (22%)	7 (32%)	38 (21%)	0.27
Forensic section	21 (10%)	6 (27%)	15 (8%)	0.01

[a] *P*-Value from binary logistic regression analyses, ICD10 diagnoses - the psychosis group was amalgamated to comprise: F20 schizophrenia, F25 schizoaffective disorder and F29 unspecified nonorganic psychosis. Other ICD-10 diagnoses reported independently are: F31 Bipolar disorder, F32 Depressive episode, F41 Other anxiety disorders, F90 Hyperkinetic disorder, F84 Pervasive developmental disorder. Mini PAS-ADD thresholds - scores were calculated using standard guidelines, Mental Health Act (MHA) section – previous history of involuntary admission included previous and current MHA sections and forensic sections, Forensic section - all individuals were on a forensic section at the time of recruitment no history of being on a forensic section was identified in any of the other participants. N.B. Several individuals had co-morbid diagnoses and are included in more than one category.

3.4.3 Recurrent CNVs

Four of the recurrent pathogenic CNVs were identified at the 16p11.2 locus (4 duplications, 1 deletion). The 16p11.2 region is associated with increased risk for ASD, schizophrenia and MDD, in keeping with the phenotypes observed in this cohort[153]. One CNV duplication was identified at the 16p13.11 locus, which has previously been associated as a risk factor for cognitive impairments and behavioral abnormalities[154]. A further five CNVs were identified in the 15q11.2-13.3 region (15q11.2 deletion, Angelman syndrome type 2, 15q12-13.1 deletion, 15q11.2-13.1 duplication, and 15q13.3 deletion) with variable psychiatric phenotypes. The 15q11.2 deletion, 15q13.3 deletion, 16p11.2 deletion and duplication and 16p13.11 duplication can be considered as neurosusceptibility loci. These CNVs have incomplete penetrance in that they occur at significantly higher frequencies in disease cohorts, however do not inevitably result in a disease phenotype and can sometimes be observed in healthy controls[150].

Another region prone to pathogenic recurrent CNV is the 17q11.2 locus, which encompasses the neurofibromatosis type 1 (*NF1*) tumour suppressor gene. A participant with a *NF1* microdeletion was identified who presented with a clinical diagnosis of ASD and challenging behaviours. This supports previous evidence of ASD being associated with variants in the *NF1* gene mutations[155]. A deletion at 2q13 was also detected in a participant with a clinical diagnosis of ASD. Of the 29 patients now described with this CNV four are reported to have ASD, although assessment information was only available for six participants[156]. CNVs in this region have also been shown to be enriched in schizophrenia cohorts[157]. The participant also presented with sub-clinical features of psychosis in addition to anxiety and behavioural problems.

Finally, a recurrent CNV was identified in the *NRXN1* gene. The *NRXN1* gene is located at 2p16.3, it encodes a cell-surface receptor which is important for neurotransmission. Exonic *NRXN1* deletions have been associated with increased risk for schizophrenia and ASD[158]. The participant had a deletion of exon 1 of the *NRXN1*

gene, a clinical diagnosis of personality disorder and subclinical symptoms of psychosis and stereotyped behaviours.

3.4.4 Non-recurrent rare CNVs

One CNV was identified in a known neurodevelopmental gene, *GRIN2B*. The *GRIN2B* gene is located at 12p13.1 and encodes the NR2 subunit of a N-methyl-D-aspartate (NMDA) glutamate receptor heteromer, which mediates excitatory neurotransmission and is thought to play an important role in memory and learning. Variants in the *GRIN2B* gene have previously been associated with behavioural problems[159]. Three patients with moderate ID and facial dysmorphisms were initially described by Dimassi *et al.*, who had overlapping microdeletions encompassing all or part of the GRIN2B gene[160]. Further CNV deletions were subsequently described in patients with ID and macrocephaly[161], and DD, autistic features, dysmorphic features and congenital abnormalities[162]. Furthermore, three patients with duplication CNVs encompassing the GRIN2B gene were also described, all of whom had DD and dysmorphic features[163]. A duplication affecting exon 9 of the GRIN2B gene was identified in a participant displaying self-injurious and aggressive behaviours. This 9kb duplication only affected the GRIN2B gene, whereas the CNVs reported in previous literature were larger and encompassed multiple additional genes.

A duplication of the 4p16.3 loci, a region where deletions give rise to the better characterised Wolf-Hirschhorn syndrome, was also detected. The CNV identified partially overlaps with the CNV reported in a case study of a patient with ADHD[164]. The participant has a clinical diagnosis of ASD and was a forensic in-patient. A duplication at Xq24-25 was observed in a participant with aggressive and stereotyped behaviours. Abnormal behaviours, primarily hyperactivity, have previously been associated with CNVs in this region. This region encompasses the *GRIA3* gene, encoding glutamate receptor, ionotropic AMPA subunit 3 and the *STAG2* gene, which encodes a component of the cohesion complex and is essential for chromosome segregation in dividing cells. The CNV has been identified as X-linked intellectual disability syndrome[165]. Aggregation of patients with Xq24-25 duplications has enabled refinement of the shortest region of CNV overlap, implicating *STAG2* as the likely causative gene[166].

A CNV deletion was detected in one participant at 12q21.2-21.31 loci comprising 17 genes. This region contains the Synaptotagmin-1 (*SYT1*) gene, which encodes a calcium-binding synaptic vesicle membrane protein involved in triggering neurotransmitter release at the synapse. A variant in *SYT1*, with a dominant negative function, has recently been associated with profound cognitive impairment although no psychiatric phenotype is described[167]. Whilst this is a copy number loss and different phenotype a low haploinsufficiency score is suggestive of adverse functional consequences[29]. The participant has a clinical diagnosis of paranoid schizophrenia, a history of alcohol abuse and was a forensic in-patient.

Another participant had a duplication at 13q32.3-13q33.3 comprising 33 genes. This region contains the D-amino acid oxidase activator (*DAOA*) gene which indirectly affects glutamatergic transmission and dopamine turnover. The *DAOA* gene has been reported to be associated with both schizophrenia and BPAD[168]. The participant had a clinical diagnosis of ASD and ADHD and was a forensic in-patient. Finally a participant had a CNV deletion at 19q13.32 comprising 56 genes. This deletion partially overlaps with a case reported previously, but does not include any of the proposed candidate genes or a psychiatric phenotype[169]. The participant had a clinical diagnosis of depression and BPAD.

3.4.5 Clinician survey for patients with pathogenic CNVs

The survey was distributed to 22 psychiatrists, who were the responsible clinicians for the participants with pathogenic CNVs, and 13 psychiatrists responded to the survey. All psychiatrists reported that they fed back the genetic results to family members, whereas 11/13 psychiatrists fed back the genetic result to their patients. One of the patients died, so it was not possible to feedback the result, and no reason was provided by the other psychiatrist who did not feedback the result. Only 4/13 (31%) of psychiatrists referred their patients to clinical genetics for genetic counselling – despite this being recommended to all psychiatrists. Three psychiatrists said that the clinical genetics service offered helpful advice and support, whereas one psychiatrist received no response. The majority of psychiatrists reported the test result to the participant's GP (10/13, 77%). There were no adverse outcomes reported from feedback of the

genetic test results. None of the psychiatrists reported clinical management changes as a consequence of the genetic diagnosis. One of the psychiatrists commented:

> "A lot of what was known was clarified by result but a lot of work had been done already".

3.5 Discussion

The yield of pathogenic CNVs identified in adults accessing ID psychiatry services was 11%. Previously, investigations of paediatric DD/ID cohorts have identified a yield of pathogenic CNVs ranging between 14-20%[25,44]. As the cohort was sampled from a population of adults with idiopathic ID, it is likely to be depleted of clinically recognisable syndromic disorders, which were testable with pre-CMA technologies. In support of this, the majority of CNV identified are recurrent CNVs, which have been identified via the genotype-first approach and thus named after the chromosomal region involved[138]. Whereas, there were few occurrences of syndromes in which the aetiology was identified prior to the application of CMA, with one Angelman's syndrome, one NF1 microdeletion syndrome type 2 and two chromosomal abnormalities.

The yield identified was considerably higher than that found in cohorts recruited on the basis of a schizophrenia diagnosis. The yield of pathogenic CNVs in adults with schizophrenia is reported to be in the range of 2.5-5%[28,62]. There have been calls for increased clinical use of CMA in patients with schizophrenia, although there is also resistance to this given the low diagnostic yields. This study suggests that approximately 11% new genetic diagnoses could be made by testing adults accessing ID psychiatry services in the UK. The higher diagnostic yield in patients with ID and psychiatric co-morbidities argues that this patient group should be a priority for consideration of routine CMA in psychiatric practice.

A recent study by Lowther *et al*. also advocated for increased CMA testing in the adult DD/ID population, particularly for those with a dual diagnosis. The study undertook IQ phenotyping in a large community sample of adults with schizophrenia, identifying a yield of 24.1% pathogenic CNVs in individuals with co-morbid ID[142]. There are

several reasons why this study may have identified a higher yield of pathogenic CNVs. Firstly, the only co-morbid diagnosis included in this study is schizophrenia, and it may be that the yield is indeed higher for schizophrenia as compared to psychiatric disorders in general. Secondly, the higher yield may have arisen as the authors used a less stringent definition of pathogenicity, whereby they considered CNVs classed as VOUS likely pathogenic or pathogenic to be pathogenic. Further testing in adult ID psychiatry cohorts will be required to clarify the yield of pathogenic CNVs.

A broad range of psychiatric diagnoses/symptoms were observed across the cohort. The pattern of comorbidities, either defined by ICD-10 diagnoses or Mini PAS-ADD thresholds, was complex. Inclusion of Mini PAS-ADD thresholds indicated a burden of psychopathology not captured by ICD-10 diagnoses. It is of interest that 41% of participants with pathogenic CNVs met the Mini PAS-ADD threshold for psychosis, whereas based upon ICD-10 criteria this was only 14%, see Table 3-3. Given the challenges of diagnosing psychiatric disorders in individuals with ID, particularly as the assessments rely on self-reported symptoms, it may be that more individuals have a psychotic disorder than those with a formal diagnosis. Equally, it may be that as the screening assessments are more challenging in an ID population this figure is falsely inflated.

The frequency of particular psychiatric phenotypes was tested between the pathogenic and benign CNV groups. The only significant finding was that there was an excess of participants on a forensic section in the pathogenic CNV group in comparison with the benign (including VOUS) CNV group. One observation from participant recruitment, which may partly explain this finding, is that many of the forensic participants had a family history of mental health problems, often on both sides of the family. It may be that assortative mating in these families was contributing to a high psychiatric genetic loading. Additionally if many of these participant had inherited psychiatric risk CNVs the additional impact of the parent also having the CNV could also contribute to the severe phenotype. It was also observed that many forensic participants had disruptive childhoods and had been put into the care of social services. However, it was not possible to determine the significance of these factors as inheritance data was not readily available. Also, the significant finding does not survive a Bonferroni multiple

testing correction, thus may be a spurious result. No link can be made between specific CNVs and offending behaviour, as causality cannot be inferred. This finding warrants further investigation in much larger samples.

The majority of pathogenic CNVs – 64% - were found at recurrent CNV loci. Interestingly, the most frequently observed CNV in this study was the 16p11.2 duplication (4 individuals, 2%). This CNV has been widely reported in other studies, with a frequency of ~0.2% in DD/ID,[45] ~1% in ASD[82] and ~0.3% in schizophrenia[62]. Accepting the small sample size this may suggest a particular enrichment of this recurrent CNV in the adult population of ID and co-morbid psychiatric disorders.

Assessment of adults with ID and formulation of the psychiatric presentation can be challenging. The majority of pathogenic CNVs were found at recurrent CNV loci where, at least some, information on the associated phenotype is available. Such information may aid understanding of the patient's clinical presentation for both clinicians and family members. Furthermore knowledge of associated phenotypes may guide psychiatric evaluation. For example, the identification of a 16p11.2 duplication would be an indicator to screen for the presence of ASD, psychosis or affective disorders[153]. As part of the feedback process disorder support guides were provided to treating psychiatrists. However, survey results from psychiatrists of participants with pathogenic CNVs indicate that no clinical management changes were made as a result of the genetic diagnosis and one psychiatrist commented that a lot of the work had been done already. One potential explanation for this is that the mean age of the sample was 37 years, past the typical age for emergence of later-onset psychiatric and medical symptomatology. Thus, it is likely that clinical treatment plans have already been put in place. Further research in a larger sample, with a diverse range of ages, is required to further investigate this finding.

One of the main limitations of this study is that the recruitment strategy was susceptible to ascertainment bias. The recruitment strategy was focused on individuals with a more severe psychiatric phenotype, i.e. those presenting to psychiatric services, and PIs may have selected patients who they thought most likely to have a genetic disorder. However, those with the most severe phenotypes might have been under

sampled because of difficulties recruiting this population group to research studies. This is supported by the fact that the majority of the pathogenic CNV group had mild ID (15/22, 68%). We took steps to facilitate the recruitment of individuals with moderate and severe ID by producing Easy read information sheets and consent forms, making the study information accessible during the capacity assessment process. However, it is still more time consuming to explain the study information via Easy read documentation, which may have been a barrier for clinicians to recruiting patients with more severe ID. In addition the recruitment process required the availability of a family member or carer to provide informant information. It may be that informants felt less able to dedicate their time to research if they are already dealing with extreme challenging behaviours or psychiatric diagnoses, which require high levels of care. Another approach to participant recruitment might be to ask PIs to randomly sample participants from their caseload, although this would initially be more labour intensive and may have led to delays in meeting the tight recruitment targets.

Further limitations of this study are that the sample size was modest. Estimates of the penetrance of particular phenotypes would require epidemiological based studies. Technological limitations of the aCGH platform include; inability to detect balanced translocations, single gene disorders and low level mosaicism. Karyotyping enables the detection of chromosomal translocations and inversions, however it was not possible to undertake karyotyping in this analysis meaning that these genetic variants would be largely missed. As the array platform has not utilised in research studies of control populations comparisons with other studies is prone to technical confounds.

3.6 Conclusion

CNV screening using clinically available CMA offers over one in ten new aetiological diagnoses for adults with idiopathic ID presenting to psychiatric services in the UK. Clinical and research data on emerging CNV syndromes is strongly biased towards paediatric populations. However, the full extent of the phenotype associated with a particular CNV may only be realised in adulthood as psychiatric disorders emerge. Most pathogenic CNVs in co-morbid NDDs affect established risk loci, with the 16p11.2 duplication being particularly frequent in this understudied adult population. The addition of psychiatric phenotypic information to very rarely observed and novel

likely pathogenic CNVs could be beneficial for patient clinical management and management of children with new emerging CNVs.

Chapter 4 CNV analysis in a European ID psychiatry cohort identifies high rates of NDD risk CNVs

Parts of this chapter have been adapted from the following journal article, which is currently in press:

Thygesen, J.*, **Wolfe, K.***, McQuillin, A., Viñas-Jornet, M., Baena, N., Brison, N., D'Haenens, G., Esteba-Castillo, S., Gabau, E., Ribas-Vidal, N., Ruiz, A., Vermeesch, J., Weyts, E., Novell, R., Van Buggenhout, G., Strydom, A., Bass, N.*, Guitart Feliubadaló, M.*, Vogels, A.* (2018) High rates of neurodevelopmental risk CNVs in patients with intellectual disabilities and co-morbid psychiatric disorders. *The British Journal of Psychiatry*. In press. *Joint first-authorship

I undertook and supervised recruitment of the London participants (including psychiatric phenotyping), acted as project co-ordinator for the international consortium, undertook phenotypic data analysis, and wrote and revised the manuscript for publication. Study design and analytical input from: AM, NB, MGF, AV. CNV calling and analysis: NB, JV, GVB. Participant recruitment: MVJ, GAH SEC, EG, NRV, AR, RN. JT: led quality control, data amalgamation and statistical analyses. Critical review of the manuscript was undertaken by all authors.

4.1 Introduction

In Chapter 1 of this thesis it was established that a proportion of the risk for NDDs can be attributed to CNVs. As previously discussed, the majority of pathogenic CNVs are risk factors for multiple disorders. For example, all the CNVs identified as risk factors for schizophrenia are also associated with risk for ID[97], and the same is true for ASD risk CNVs[87]. Thus, collectively these CNVs can be referred to as NDD risk CNVs. Frequency estimates of these CNVs in different populations show a positive correlation between the severity of the phenotype and frequency of the CNV. Typically, there is an absence, or low frequency of CNVs in controls, whereas and the greatest frequency of CNVs is seen in early-onset NDDs[97,107]. These CNVs confer moderate to large risk factors for NDDs (Odds Ratio 1.5->50)[97], and therefore have important clinical implications for affected individuals and at risk family members.

Research to date has primarily focused on researching cohorts from traditional diagnostic categories, and there is a lack of research in adults who have multiple co-morbid NDD diagnoses.

Building on the work from Chapter 3, a consortium was formed to bring together three European samples of adults with idiopathic ID and co-morbid mental disorders with available genome-wide CNV and phenotypic data. The frequency of pathogenic CNVs and rate of NDD risk CNVs has never previously been investigated in a large cohort of adults with co-morbid NDD phenotypes.

4.2 Aims

This analysis aimed to determine; (i) the frequency of known NDD risk CNVs as compared to large NDD cohorts from the existing literature; (ii) the overall rate of pathogenic CNVs; (iii) the relationship between pathogenic CNVs, level of ID and co-morbid psychiatric diagnoses; and (iv) likely pathogenic CNVs affecting neurodevelopmental candidate genes.

4.3 Methods

4.3.1 Recruitment criteria and participant recruitment

The GENMID (GENetics of Mental disorders in Intellectual Disability) consortium comprises three primary research groups based in Catalonia, Spain; Leuven, Belgium; and England, United Kingdom. All sites recruited adults over the age of 18 years with idiopathic ID and one or more co-morbid psychiatric diagnoses and/or significant challenging behaviours. In Catalonia participants were identified from the Mental Health ID regional community Service Parc Hospitalari Martí i Julià, Girona. In Leuven, participants were recruited at the regional inpatient psychiatric unit for adults with ID in the St-Camillus Psychiatric Hospital, Bierbeek. Initially, only patients diagnosed with psychosis were recruited, but recruitment was later extended to other psychiatric phenotypes. In England, participants were recruited by consultant psychiatrists in intellectual disabilities in accordance with the methodology described in Chapter 3.

There were no significant differences in recruitment strategy between sites, all sites recruited adults with ID plus either challenging behaviour and/or one or more psychiatric diagnosis – as per the methodology in Chapter 3. Participants were screened for prior genetic diagnoses. Approximately 10 participants from the Leuven cohort were identified as having a diagnosis of Down's syndrome or fragile X syndrome and were excluded from further analyses.

4.3.2 Phenotypic assessments

For all sites the ID levels are in accordance with the ICD-10 ranges (<20 profound ID, 20-34 severe ID, 35-49 moderate ID, 50-69 mild ID, 70-84 borderline ID)[39]. As previous research has proposed that the factors influencing mild ID and IQ in the normal range are separate from those influencing severe ID[43], we combined the categories into two groups for further analyses. The <20-49 ranges were collapsed into a severe category and the 50-84 ranges were collapsed into a mild category. All sites identified psychiatric diagnoses from medical records and/or informants. Psychiatric diagnoses for Catalonia and Leuven were converted from Diagnostic and Statistical Manual of Mental Disorders IV[170] to ICD-10 criteria (with agreement between two psychiatrists).

4.3.3 Genetic analysis and CNV calling

DNA was extracted from blood and saliva samples. Samples from Catalonia were analysed using the 400K Agilent platform (Agilent Technologies, Santa Clara, California, USA) at the Genetics Laboratory, UDIAT-Centre Diagnòstic, Parc Taulí Hospital Universitari. Samples from Leuven were analysed on the CytoSure ISCA oligoarray set (OGT, Oxford, UK) at the Constitutional Cytogenetics Unit of the Center of Human Genetics, University of Leuven. Samples from England were analysed on the NimbleGen 135K platform (87%) (Roche NimbleGen, Madison, Wisconsin, USA) and the Cytoscan 750K platform (13%) (Affymetrics, Santa Clara, California, USA) at the North East Thames Regional Genetics Service Laboratory.

CNV calling took place at the respective clinical laboratories. There was a two-tier variant categorisation process. CNVs were initially reported by the independent

clinical laboratories and classified into three categories: pathogenic, VOUS and benign. This was in keeping with internal laboratory protocols based on the American College of Medical Genetics best guidelines[14] or the Association of Clinical Genetic Science Best Practice Guidelines[149]. The Database of Genomic Variants gold standards track was used as a control reference database for determining control CNV frequencies[30], and compared with internal laboratory databases for frequencies in patients with severe developmental disorders referred for genetic testing.

None of the pathogenic variants identified were required to be homozygous or compound heterozygous in order to be classified as pathogenic. Secondly, between site discrepancies in CNV pathogenicity were reclassified in accordance with Kearney *et al.*[14], see Table 4-1. CNVs designated as uncertain clinical significance were reclassified into likely benign or likely pathogenic using this methodology. The genome coordinates for all sites are reported according to the National Center for Biotechnology Information (NCBI) human genome build 37 (hg19, February 2009).

Table 4-1: Between site discrepancies in CNV classifications and reclassification rationale

CNV region	Initial classification	Re-classification	Reason for re-classification
2p16.3(50,882,091-50,949,412)x0	Vous likely pathogenic	Vous likely benign	Non-exonic NRXN1 CNV - no literature proof of pathogenicity
2p16.3(50,937,464-51,029,090)x0	Vous likely pathogenic	Vous likely benign	Non-exonic NRXN1 CNV - no literature proof of pathogenicity
15q11.2(22,698,579-23,249,693)x0	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 21359847, 25689425
15q11.2(22,753,658-23,084,392)x0	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 21359847, 25689425
15q11.2(22,753,658-23,187,967)x0	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 21359847, 25689425
15q13.2q13.3(30,419,801-32,861,612)x0	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 25077648, 26997942
16p11.2(29,652,360-30,199,696)x0	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 19914906, 25064419
16p11.2(29,652,360-30,199,696)x3	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 19914906, 26629640
16p13.12p12.3(14,622,055-17,409,257)x3	Vous likely pathogenic	Pathogenic	Literature evidence of pathogenicity. PMIDs: 21614007, 21150890

4.3.4 NDD CNV frequency methodology

A list of 63 NDD risk CNVs, that were identified from a sample of paediatric patients with severe developmental disorders (DD/ID, ASD and MCA)[45], and CNVs associated

with risk for schizophrenia were derived from Rees *et al.*[97], henceforth referred to as NDD CNVs. The NDD CNVs were called in accordance with the criteria outlined in Kendall *et al.*[119], also used by Rees *et al.*[97] (personal communication), see Table 4-2. CNVs fulfilling these calling criteria were classified as pathogenic and are included in the diagnostic yield. Duplications or deletions of the same chromosomal region were counted as separate loci (e.g. 22q11.2 deletion and duplication). The patient population rates in healthy controls, ID/ASD (the name given by Rees *et al.* to a severe developmental disorders cohort), and schizophrenia were derived from Rees *et al.*, where further information can be found about the respective samples[97]. Non-recurrent rare pathogenic CNVs and larger chromosomal abnormalities are not considered in this analysis. A rate percentage (rate in the sample divided by the sample size and multiplied by 100) was calculated to enable comparisons between different sample sizes and chi-square tests were used to determine the population differences. A Bonferroni correction was applied and significance was set at p=0.01 to account for multiple pairwise comparisons.

Table 4-2: Neurodevelopmental risk CNV critical regions and calling criteria

Locus Name	Critical Region	Calling Criteria
1p36 del/dup	chr1:0-2500000	Size >50% of critical region, affecting *GABRD*
TAR del/dup	chr1:145394955-145807817	Size >50% of critical region
1q21.1 del/dup	chr1:146527987-147394444	Size >50% of critical region
1q24 del	chr1:169680333-173303337	Size >50% of critical region
NRXN1 del	chr2:50145643-51259674	Exonic deletions
2p15-16.1 proximal dup	chr2:61245288-61414572	Size >50% of critical region
2q11.2 del	chr2:96742409-97677516	Size >50% of critical region, affecting both *LMAN2L* and *ARID5A*
2q13 del/dup	chr2:111394040-112012649	Size >50% of critical region
2q33.1 (*SATB2*) del	chr2:200134224-200325255	Size >50% of critical region
2q37 (*HDAC4*) del	chr2:239716679-243199373	Size >50% of critical region, affecting *HDAC4*
3p25.3 (*JAGN1* to *TATDN2*) dup	chr3:9932271-10322902	Size >50% of critical region
3p11.2 (*CHMP2B* to *POU1F1*) del	chr3:87267612-87531631	Size >50% of critical region
3q13 (*GAP43*) del	chr3:115332334-115504038	Size >50% of critical region
3q28-29 (*FGF12*) del	chr3:191859728-192126012	Size >50% of critical region
3q29 del	chr3:195720167-197354826	Size >50% of critical region
Wolf-Hirschhorn del/dup	chr4:1552030-2091303	Size >50% of critical region
4q21 (*BMP3*) del	chr4:81945477-81985327	Size >50% of critical region
5q14 (*MEF2C*) del	chr5:88011654-88200703	Size >50% of critical region
Sotos syndrome del	chr5:175720924-177052594	Size >50% of critical region
Williams-Beuren syndrome del	chr7:72744915-74142892	Size >50% of critical region
WBS dup	chr7:72744915-74142892	Size >50% of critical region
8p23.1 del/dup	chr8:8098990-11872558	Size >26.5% of critical region (equal to min 1Mb affected)
9p13 dup	chr9:32648800-38808255	Size >50% of critical region
9q34 dup	chr9:138460697-141036426	Size >38.8% of critical region (equal to min 1Mb affected)
10q11.21q11.23 dup	chr10:49390199-51058796	Size >50% of critical region
10q23 del	chr10:82045472-88931651	Size > 14.5% of critical region (equal to min 1Mb affected), including NRG3 and GRID1
Potocki-Shaffer syndrome del	chr11:43940000-46020000	Size >50% of critical region, including *EXT2*
12p13 dup	chr12:6471959-6825955	Size >50% of critical region
PWS/AS del/dup	chr15:22805313-28390339	Size > 71.6% of critical region (equal to min 4Mb affected)
15q11.2 BP1-BP2 del	chr15:22805313-23094530	Size >50% of critical region

15q13.3 del	chr15:31080645-32462776	Size >50% of critical region
15q24 del/dup	chr15:72900171-78151253	Size >50% of critical region
15q25 del	chr15:85139815-85716624	Size >50% of critical region
16p13.11 del/dup	chr16:15511655-16293689	Size >50% of critical region
16p12.1 del	chr16:21950135-22431889	Size >50% of critical region
16p11.2 distal del/dup	chr16:28823196-29046783	Size >50% of critical region
16p11.2 del/dup	chr16:29650840-30200773	Size >50% of critical region
17p13.3 del/dup	chr17:1247834-2588909	Exonic deletions; whole gene duplications
Smith-Magenis syndrome del	chr17:16812771-20211017	Size >50% of critical region
Potocki-Lupski syndrome dup	chr17:16812771-20211017	Size >50% of critical region
17q11.2 del/dup	chr17:29107491-30265075	Size >50% of critical region, affecting *NF1*
17q12 del/dup	chr17:34815904-36217432	Size >50% of critical region
17q21.31 del	chr17:43705356-44164691	Size >50% of critical region
22q11.2 del/dup	chr22:19037332-21466726	Size >50% of critical region
distal 22q11.2 del/dup	chr22:21920127-23653646	Size >50% of critical region
Phelan-McDermid syndrome del/dup	chr22:51113070-51171640	Size >50% of critical region

To determine the CMA yield each individual was grouped by the most pathogenic CNV detected. We also examined all likely pathogenic CNVs for recurrence in the cohort. Regions that have been previously implicated in NDD risk in the existing literature which reoccur in this cohort are further described. Finally, chi-square tests (or Fisher's exact tests were there were five or less individuals) were undertaken to examine the differences between psychiatric diagnoses, ID level and CNV pathogenicity. Since many of the co-morbid diagnoses are correlated and thus are non-independent, correction of p-values through Bonferroni or other methods was deemed too stringent. Thus, all p-values are presented uncorrected for multiple testing as recommended by several authors[171,172]. The statistical analyses were undertaken in R version 3.3.1[173].

4.4 Results

There were 599 adults (Catalonia (n=80), Leuven (n=272) and England (n=247)) with ID and one or more co-morbid psychiatric diagnoses/challenging behaviours recruited to the study (376 (62.8%) male, mean age 43.2). Just over half of the sample (50.8%) had severe ID and the remainder had mild ID. Each participant had, on average, 1.6 co-morbid psychiatric diagnoses, with pervasive developmental disorders being the most frequent diagnosis (25%), followed by unspecified non-organic psychosis (20%), see Table 4-3. The average number of CNVs per participant was 12.5 (7.4 deletions and 5.5 duplications).

Table 4-3: Full descriptive summary of the GENMID cohort

	GENMID	Catalonia	Leuven	England
Demographics				
N	599	80	272	247
Ratio (Male/Female)	1.7 (376/223)	0.9 (38/42)	2.1 (184/88)	1.7 (154/93)
Mean age (std.dev)	43.2 (14.1)	37.1 (9.8)	46.2 (14.5)	41.9 (14.1)
ID Level				
Mild	49.2%	63.7%	66.9%	25.1%
Severe	50.8%	36.2%	33.1%	74.9%
Psychiatric diagnoses				
Average number of co-morbid diagnoses (range)	1.6 (1-5)	1.8 (1-3)	1.4 (1-4)	1.7 (1-5)
F84 Pervasive developmental disorders	148 (25%)	9%	22%	32%
F29 Unspecified nonorganic psychosis	121 (20%)	12%	30%	12%
F61 Mixed and other personality disorders	108 (18%)	0%	36%	4%
Challenging behaviours	95 (16%)	62%	1%	17%
F32 Depressive episode	86 (14%)	4%	3%	30%
F41 Other anxiety disorders	60 (10%)	6%	1%	21%
F20 Schizophrenia	49 (8%)	4%	8%	9%
F31 Bipolar affective disorder	47 (8%)	2%	7%	11%
F90 Hyperkinetic disorders	41 (7%)	6%	4%	10%
F42 Obsessive-compulsive disorder	37 (6%)	16%	1%	9%
F43 Reaction to severe stress and adjustment disorders	27 (5%)	19%	4%	0%
F39 Unspecified mood disorder	25 (4%)	0%	9%	0%

N.B. Only psychiatric diagnoses found in ten or more individuals are listed in the table

4.4.1 NDD CNV frequency analysis

CNVs were identified in 23 of the 63 NDD loci described by Rees *et al.*[97]. At these 23 loci 58 CNV carriers were identified, with two subjects carrying two risk CNVs. The rate percentage (rate of participants with a NDD CNV) is 10.0%, while the rate percentage, determined from the data presented in Rees et al, is 6.5% in ID/ASD, 3.1% in schizophrenia and 1.2% in healthy control populations[97], see Table 4-4. The NDD loci frequencies are most comparable with the ID/ASD population, a sample which consisted mainly of children with DD/ID and/or ASD[46]. However, we still observe a significantly higher proportion of NDD CNVs in our ID and co-morbid psychiatric diagnosis sample, 3.5% higher (95% CI = 1-6, P = 0.00084).

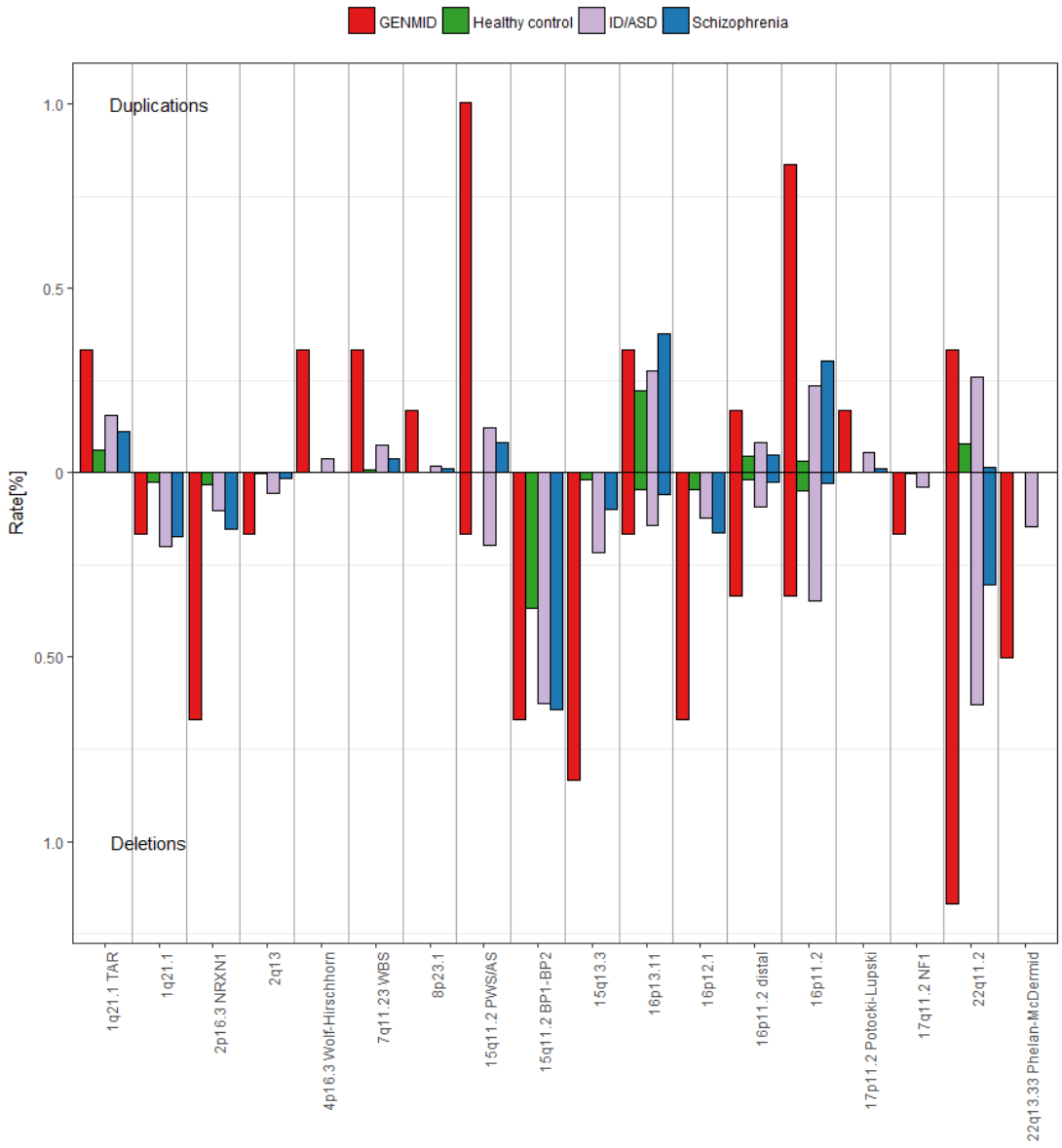
Table 4-4: Rate percentage of CNVs at 63 NDD risk loci compared with populations rates reported by Rees *et al.* (2016)

Sample	Sample size	Rate(%) of the 63 NDD-loci	Rate(%) difference (95%CI)	p-value
Healthy control	26628	1.2	8.8 (6.3-11)	2.8e-72
Schizophrenia	20403	3.1	7 (4.5-9.5)	9.7e-21
ID/ASD	29085	6.5	3.5 (1-6)	8.4e-04
GENMID	599	10.0	-	-

Rate percentage differences, 95% confidence intervals (CI) and p-values for rate comparisons are indicated.

The frequencies of the 23 NDD CNVs identified in the dataset are shown in Figure 4-1. The carrier frequency at each loci was the highest in our sample of ID and co-morbid mental disorders, with the exception of four loci for which we see comparable frequencies to the ID/ASD cohort. The five most frequent NDD CNVs in the GENMID cohort, in order of frequency, are: 22q11.2 deletion (n=7, 1.2%), 15q11.2 PWS/AS duplication (n=6, 1%), 16p11.2 duplication (n=5, 0.8%), 15q13.3 deletion (n=5, 0.8%) and 16p12.1 deletion (n=4, 0.7%). A description of all CNV loci and the carrier phenotypes can be found in Table 4-5.

Figure 4-1: NDD CNV frequencies in the GENMID sample compared to frequencies in healthy controls (N = 26628), ID/ASD (N = 29085) and schizophrenia (N = 20403) cohorts



Y axis: Rate percentage. This enables comparisons between different sample sizes, and is calculated by dividing the number of CNVs detected at each loci by the sample size and multiplying by 100, X axis: CNV region followed by the name of the associated syndrome and/or relevant genes. CNV deletions extend below the 0 and duplications extend above.

Table 4-5: Phenotypic data for carriers of NDD risk CNVs

Locus	N Observed	N Expected	GENMID rate(%)	ID/ASD rate(%)	ID Level	Other pathogenic CNVs	Diagnosis
22q11.2 del	7	3.8	1.2	0.629	Severe (4), Mild (3)	2p16.3 del (1)	F29 Unspecified nonorganic psychosis (3), F61 Mixed and other personality disorders (3), F31 Bipolar affective disorder (2), F39 Unspecified mood disorder (1)
15q11.2 PWS/AS dup	6	0.7	1.0	0.122	Mild (5), Severe (1)		F84 Pervasive developmental disorders (3), Challenging behaviours (2), F29 Unspecified nonorganic psychosis (1), F31 Bipolar affective disorder (1), F39 Unspecified mood disorder (1), F40 Phobic anxiety disorders (1), F61 Mixed and other personality disorders (1), F91 Conduct disorder (1)
15q13.3 del	5	1.3	0.8	0.218	Mild (4), Severe (1)		F32 Depressive episode (2), F41 Other anxiety disorders (2), Challenging behaviours (1), F29 Unspecified nonorganic psychosis (1), F31 Bipolar affective disorder (1), F84 Pervasive developmental disorders (1), F94 Disorders of social functioning with onset specific to childhood and adolescence (1)
16p11.2 dup	5	1.4	0.8	0.236	Severe (5)		F84 Pervasive developmental disorders (3), F32 Depressive episode (2), F25 Schizoaffective disorder (1), F29 Unspecified nonorganic psychosis (1), F42 Obsessive-compulsive disorder (1), F90 Hyperkinetic disorders (1)
2p16.3 NRXN1 del	4	0.6	0.7	0.103	Mild (2), Severe (2)	22q11.21 del (1)	F31 Bipolar affective disorder (2), Challenging behaviours (1), F22 Delusional disorder (1), F29 Unspecified nonorganic psychosis (1), F61 Mixed and other personality disorders (1)
15q11.2 BP1-BP2 del	4	3.7	0.7	0.625	Mild (2), Severe (2)	Xp22.31 del (1)	F29 Unspecified nonorganic psychosis (1), F31 Bipolar affective disorder (1), F41 Other anxiety disorders (1), F61 Mixed and other personality disorders (1)
16p12.1 del	4	0.7	0.7	0.124	Mild (3), Severe (1)		F20 Schizophrenia (1), F23 Acute and transient psychotic disorders (1), F29 Unspecified nonorganic psychosis (1), F43 Reaction to severe stress and adjustment disorders (1), F61 Mixed and other personality disorders (1), F90 Hyperkinetic disorders (1)
22q13.33 Phelan-McDermid del	3	0.9	0.5	0.148	Severe (3)		F84 Pervasive developmental disorders (2), Challenging behaviours (1), F29 Unspecified nonorganic psychosis (1), F31 Bipolar affective disorder (1)

1q21.1 TAR dup	2	0.9	0.3	0.155	Mild (1), Severe (1)		F20 Schizophrenia (1), F43 Reaction to severe stress and adjustment disorders (1), F84 Pervasive developmental disorders (1), F90 Hyperkinetic disorders (1)
4p16.3 Wolf-Hirschhorn dup	2	0.2	0.3	0.038	Mild (1), Severe (1)	12p13.33p13.32 del (1)	F29 Unspecified nonorganic psychosis (1), F84 Pervasive developmental disorders (1)
7q11.23 WBS dup	2	0.5	0.3	0.076	Mild (2)		F39 Unspecified mood disorder (1), F61 Mixed and other personality disorders (1), F90 Hyperkinetic disorders (1)
16p13.11 dup	2	1.6	0.3	0.275	Mild (1), Severe (1)		Challenging behaviours (1), F29 Unspecified nonorganic psychosis (1)
16p11.2 distal del	2	0.6	0.3	0.094	Mild (1), Severe (1)		F23 Acute and transient psychotic disorders (1), F29 Unspecified nonorganic psychosis (1)
16p11.2 del	2	2.1	0.3	0.347	Severe (2)		F29 Unspecified nonorganic psychosis (1), F41 Other anxiety disorders (1)
22q11.2 dup	2	1.6	0.3	0.260	Mild (1), Severe (1)		F20 Schizophrenia (1), F29 Unspecified nonorganic psychosis (1)
1q21.1 del	1	1.2	0.2	0.201	Mild (1)		F32 Depressive episode (1), F41 Other anxiety disorders (1)
2q13 del	1	0.3	0.2	0.057	Mild (1)		F84 Pervasive developmental disorders (1)
8p23.1 dup	1	0.1	0.2	0.017	Severe (1)		F39 Unspecified mood disorder (1), F61 Mixed and other personality disorders (1)
15q11.2 PWS/AS del	1	1.2	0.2	0.197	Severe (1)		F41 Other anxiety disorders (1)
16p13.11 del	1	0.9	0.2	0.142	Mild (1)		F61 Mixed and other personality disorders (1)
16p11.2 distal dup	1	0.5	0.2	0.083	Mild (1)		Challenging behaviours (1)
17p11.2 Potocki-Lupski dup	1	0.3	0.2	0.055	Mild (1)		F61 Mixed and other personality disorders (1)
17q11.2 NF1 del	1	0.2	0.2	0.039	Severe (1)		Challenging behaviours (1), F84 Pervasive developmental disorders (1)
Total	60	38.8	10.0	6.512			

N Expected, number of carriers we would expect based on ID/ASD frequencies (Rees et al. 2016) of the loci and our sample size of 599 participants, Rate(%) in GENMID and ID/ASD samples, ID level, other identified pathogenic CNVs and psychiatric diagnosis of the GENMID carriers.

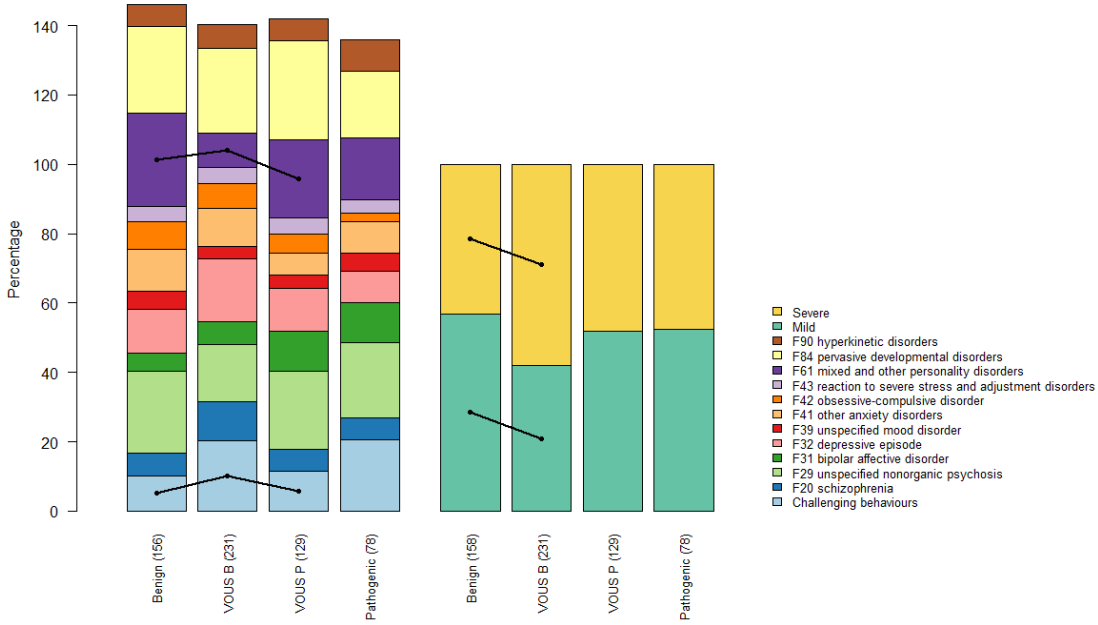
4.4.2 Yield of pathogenic CNVs

At least one pathogenic CNV was identified in 78 participants (13.0%, 95% CI 10.5-16.0), with similar yields found at all research sites (Catalonia: 13.8%, Leuven: 14.0%, and England: 11.7%). Pathogenic CNVs comprised those identified at the NDD loci previously described and a further 25 CNVs reported as pathogenic by the clinical laboratory services. The pathogenic CNVs were predominantly deletions (59.5%). A rate of 11% pathogenic CNVs has previously been was reported in a subset of 202 of the 247 participants from the England sample[174]. Removing these 202 individuals from the GENMID sample still provides a diagnostic yield of 13.9%, thus indicating that the yield of undiagnosed pathogenic CNVs in adults with co-morbid NDDs is in the range of 11-14%.

4.4.3 ID level, psychiatric diagnoses and CNV pathogenicity

Group differences between CNV pathogenicity, psychiatric diagnoses and level of ID were examined. There were some differences in the proportions of level of ID and psychiatric diagnoses between the CNV pathogenicity groups (pathogenic, likely pathogenic, likely benign and benign), see Figure 4-2. However, no simple unidirectional relationships were observed. Equally, minor differences in the severity of ID were found between CNV pathogenicity groups, but no overall unidirectional relationship was observed.

Figure 4-2: Association between CNV pathogenicity, diagnostic groups and ID severity



Bars 1-4 show the diagnostic rates of all diagnoses with 10 or more affected, for carriers with benign, variant of unknown significance likely benign (VOUS B), VOUS likely pathogenic (VOUS P) and pathogenic CNVs. Bars 5-8 show the frequency of mild and severe ID for individuals with benign, VOUS B, VOUS P and pathogenic CNVs. Lines indicate nominal statistical difference between groups P-value <0.05 only adjacent groups within bars 1-4 and 5-8 have been tested.

4.4.4 Likely pathogenic CNVs

The yield of likely pathogenic CNVs in the sample was 21.5% (95% CI 18.4-25.1). Investigation of recurrent likely pathogenic CNVs revealed 34 CNVs in 16 regions. Four recurrent CNVs identified corroborate existing evidence for the involvement of these regions in NDD risk, see Figure 4-3.

Figure 4-3: Chromosomal locations of four overlapping likely pathogenic CNVs

The top image shows the chromosomal location of the CNV, with the region highlighted by a red box, CNV deletions are shown in red and CNV duplications are shown in green, UCSC genes included in the CNV are shown, the image was exported from UCSC in chromosomal build GRCh37/hg19.

First, we identified two carriers of exonic duplications in the *CNTN6* gene at 3p26.3. The CNTN proteins belong to a immunoglobulin super family of cell adhesion molecules and have an important role in neurodevelopmental processes[175]. *CNTN6* duplications were first identified in patients with ASD[86,176] and later in a patient with ID and facial dysmorphisms[177]. A review of 14 patients with *CNTN6* CNVs revealed that both CNV deletions and duplications affecting *CNTN6* are thought to be involved in variable neuropsychiatric phenotypes[178]. The participants with *CNTN6* duplication CNVs both presented with mild ID. One had schizophrenia and personality disorder, and one had challenging behaviours and had been convicted of a serious criminal offence. Interestingly, the participant with schizophrenia and personality disorder also had a duplication in the *CNTN4* gene. CNVs affecting *CNTN4* are also thought to confer risk for various NDDs[179]. As CNTN4 is an important paralog of CNTN6, it may be that a double hit in both of these genes results in a more severe phenotype as multiple genes are affected in the same pathway.

Second, two participants with CNV duplications at the 9q21.32q21.33 locus were identified encompassing the *SLC28A3* and *NTRK2* genes. SLC28A3 is a nucleoside transporter involved in the regulation of multiple processes, including neurotransmission; however, there are no prior reports of its role in psychiatric risk. NTRK2 is a receptor tyrosine kinase with numerous neurodevelopmental functions, including synapse formation and plasticity. Altered *NTRK2* expression has been identified in the brains of patients with schizophrenia[180]. One participant had severe ID and BPAD, and the other had mild ID and unspecified non-organic psychosis.

Third, five participants with exonic CNVs in the *CHD* gene family were identified. The CHD proteins are involved in chromatin structure remodeling and the epigenetic regulation of transcription. Three of the participants had exonic CNVs (2 duplications, 1 deletion) in the *CHD8* gene at 14q11.2, which also encompass *SUPT16H*. The protein encoded by the *SUPT16H* gene is thought to be involved in DNA replication and repair. CNV deletions affecting *CHD8* and *SUPT16H* were initially described in children with DD and dysmorphic features[181]. Variants in the *CHD8* gene are thought to confer a phenotypic subtype of ASD, comprising macrocephaly, facial dysmorphologies and gastrointestinal abnormalities[182]. Both deletions[183] and

duplications[184,185] affecting *CHD8* and *SUPT16H* have been described with variable neurodevelopmental phenotypes. The two participants with CNV duplications both had severe ID, one was diagnosed with schizophrenia and the other with BPAD. The participant with the CNV deletion also had severe ID and ASD.

Finally, two participants with exonic CNVs in the *CHD2* gene at 15q26.1 (one deletion and one duplication) were identified. Several patients have been described with *CHD2* deletions; with a common phenotype of ID, epilepsy, and aggressive challenging behaviours[186,187]. To our knowledge, a CNV duplication in *CHD2* has not previously been described in the literature. The deletion carrier had severe ID and schizoaffective disorder, and the duplication carrier had challenging behaviours and BPAD. Both patients also had an epilepsy phenotype.

4.5 Discussion

Previous investigations in this novel patient group of adults with co-morbid ID and psychiatric disorders identified a diagnostic yield of 11% pathogenic CNVs[174]. Combining this with data from two additional European research sites, this finding is replicated with a higher yield of 13% pathogenic CNVs in 599 participants (or 13.9% with the previously reported cases removed). There were similar rates of pathogenic CNVs across the separate research sites, making this a relatively robust finding of a high diagnostic yield in this patient group.

In the study described in Chapter 3 the majority of pathogenic CNVs were found to occur at recurrent loci, which have been independently implicated in risk for various NDDs. This pattern was also seen in the GENMID consortium study, with 70% of pathogenic CNVs being identified at NDD risk loci. Out of 63 NDD risk loci, described by Rees *et al.*, carriers were identified at 23 of the loci. It is unsurprising that no carriers were identified in the remaining 40 loci, as these CNVs are very rare with reported frequencies in ID between 0.01-0.26% (mean = 0.06%)[97]. Presuming that there is an additive effect of having both ID and a co-morbid psychiatric disorder, then one would expect to see an increased frequency of the 63 NDD CNVs. Indeed, the cumulative frequency was significantly higher, as compared to both ID/ASD

populations not selected for psychiatric co-morbidity and individuals with schizophrenia.

One complication in interpreting these findings is that the ID/ASD population is poorly phenotyped. For example, the phenotypic information from Coe *et al.*[46], from which many of these participants were derived, states that 73% of the cases suffer from ID, DD and/or ASD, with the remaining cases either having congenital abnormalities or not being annotated. Firstly not all of the individuals had ID, and secondly it is unclear what proportion of the cohort have an ASD diagnosis. All participants in the GENMID cohort had ID plus challenging behaviour and/or one or more psychiatric diagnoses, so overall the participants will be more severely affected. However the exact differences between the cohort is unknown, particularly as the ID/ASD sample was derived from paediatric cases and many patients may have gone on to develop later-onset psychiatric diagnoses.

The phenotypic presentation of the majority of NDD CNV carriers is highly variable, both in terms of the level of ID and the psychiatric diagnoses. This indicates a broader role for genes within NDD CNV loci in conferring general, as opposed to disorder specific, psychiatric risk. Although, previous research has highlighted that NDD loci can show a complex pattern of both shared and distinct risk for NDDs[108]. Extremely large sample sizes will be required to further delineate these loci specific associations. Interestingly, at least one CNV carrier at each of the five most frequent loci has a psychosis phenotype. Of particular interest are the four carriers of the 16p12.1 deletion, which was significantly associated with risk for schizophrenia by Rees *et al.*[97]. The rate in the schizophrenia cohort was found to be 0.16% (33/20403), whereas we identified a higher rate of 0.67% (4/599) in the GENMID cohort. Three of the four carriers had a psychosis phenotype, offering further support for this locus as a risk factor for both ID and psychotic disorders. Previously, in Chapter 3, an enrichment of the 16p11.2 duplication (2%) was identified as compared to the frequencies reported in other studies. Analysis of this CNV in a larger cohort has identified a lower overall frequency (0.8%), although it still remains one of the most frequent CNVs identified in adults with co-morbid NDDs.

In addition to the CNVs identified at known NDD loci, a further 26 CNVs were reported as pathogenic by the clinical genetic services. The majority of these were large deletion CNVs (1.7Mb-13.2Mb), which overlapped CNVs described in single case reports in the existing literature. This group of CNVs are likely to be extremely rare and thus would not be observed at high enough frequencies in existing case-control studies. A clear unidirectional relationship between psychiatric diagnoses, level of ID and CNV pathogenicity could not be identified. It is possible that this partly reflects the difficulty of diagnosing psychiatric disorders and assessing ID severity in individuals with ID and delineation of this relationship requires larger sample sizes.

Following a literature review of likely pathogenic CNVs that recur, support is provided for neurodevelopmental candidate genes which have been implicated in previous literature. Unlike the pathogenic CNVs, the likely pathogenic CNVs supporting existing NDD candidate loci were small (<1Mb) and affected only a small number of genes. There is a growing body of literature for the role of the *CNTN* and *CHD* gene families in risk for ID and co-morbid psychiatric disorders. Again, there appears to be a highly variable phenotype associated with CNVs affecting these genes. Further research will be required to consider the clinical implications of these CNVs, which were not reported as pathogenic by the clinical genetics services.

One of the limitations of this study is that there were some differences between the recruitment strategies at the different sites, for example participants were recruited from in-patient psychiatric services in Leuven and primarily outpatient services in Catalonia and England. Most individuals lacked inheritance data, which is a valuable aid in categorisation of rare variants and may have led to an underestimate of the diagnostic yield. Different array CGH platforms were utilised to detect the CNVs at the different sites; however, as all the platforms used were high resolution this is unlikely to have major effects. For the phenotype analysis the ID levels were categorised into two broader groups, which resulted in participants with different levels of ID being classified together. Finally, further characterisation of the relationship between NDD risk CNVs and associated phenotypes would require much larger case-control samples or epidemiological based studies.

4.6 Conclusion

A 13% rate of undiagnosed pathogenic CNVs was detected in adults with idiopathic ID and co-morbid psychiatric disorders, which is much higher than in studies of schizophrenia alone. Consistent with the findings of Chapter 3, this suggests that if CMA is going to be offered more widely in psychiatric practice, ID psychiatry services should be a priority for increased testing. Replicating the findings of Chapter 3, the majority of CNVs were identified at recurrent loci – with 70% of pathogenic CNVs being identified at established NDD risk loci. The high rates of CNVs at established NDD loci support a model whereby the frequency of NDD CNVs increases with the severity of the phenotype. Studying this adult population also facilitates description of psychiatric and medical associations across the life course, for both pathogenic CNVs and likely pathogenic candidate loci.

Chapter 5 Delineating the psychiatric and behavioural phenotype of recurrent 2q13 deletions and duplications

Parts of this chapter have been adapted from the following published journal article:

Wolfe, K., McQuillin, A., Alesi, V., Boudry Labis, E., Cutajar, P., Dallapiccola, B., Dentici, M.L., Dieux-Coeslier, A., Duban-Bedu, B., Duelund Hjortshøj, T., Goel, H., Loddo, S., Morrogh, D., Mosca-Boidron, A.L., Novelli, A., Olivier-Faivre, L., Parker, J., Parker, M.J., Patch, C., Pelling, A.L., Smol, T., Tümer, Z., Vanakker, O., Haeringen, A.V., Vanlerberghe, C., Strydom, A., Skuse, D., Bass, N. (2018) Delineating the psychiatric and behavioural phenotype of recurrent 2q13 deletions and duplications. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. PMID 29603867.

I designed the study, undertook participant recruitment, data analysis and wrote the manuscript. Support with recruitment and participant phenotyping: AV, E. BL, PC, BD, M.L D, A D-C, B D-B, T. DH, HG, SL, DM, A.L M-B, AN, L O-F, JP, M.J P, CP, A.L P, TS, ZT, OV, A.V. H, CV, Support with study design and data analysis: AM, AS, DS, NB. Critical review of the manuscript was undertaken by all authors.

5.1 Introduction

In Chapter 3 an individual with a recurrent CNV at chromosome locus 2q13 was identified. Whilst this is classed as a recurrent CNV, in that it is flanked by LCRs and thus the same CNV has arisen at 2q13 in multiple independent individuals, very few patients with this CNV have been described in published studies. For example, the latest large study of 2q13 deletions presented a summary of 29 patients from all previous published studies[156], and even fewer published case series of 2q13 duplication patients exist. The addition of detailed phenotypic information to very rarely observed CNVs could be beneficial for patient clinical management. It is particularly pertinent to study psychiatric and behavioural phenotypes, which have been less well explored in the existing literature.

Investigation of chromosomal rearrangements in regions of LCRs, identified 2q13 CNVs in patients with developmental disorders[188]. However, the pathogenicity of these CNVs was initially described as of uncertain significance. This was due to a 2q13 duplication also being found in a healthy control[188] and the findings from a previous study that the same 2q13 deletion in two siblings with developmental problems had been inherited from an unaffected parent[189].

Analysis of larger samples revealed that 2q13 CNVs are associated with an increased risk of DD and ID. Cooper *et al.* reported 12 deletions and 9 duplications in cases (N=15,767) and observed 1 deletion and 0 duplications in controls (N=8,329). They found an enrichment of the deletion (P=0.032) and duplication (P=0.022) in cases, as compared to controls. The deletion was associated with cardiovascular disorders, whereas the duplication was associated with craniofacial features[45]. Yu *et al.* described the phenotype of five 2q13 patients alongside 14 additional cases from a literature review, concluding that 93% of individuals had impaired development and 63% had facial dysmorphisms[190]. Some of these patients had a diagnosis of ASD or ADHD, although many were too young for clinical assessment, or it was unclear whether assessments had taken place.

Costain *et al.* found 2q13 CNVs to be significantly associated with schizophrenia (P=0.0002) in a community-based schizophrenia cohort (N=459), as compared to a large population-based control sample (N=23,838)[157]. They identified three 2q13 CNV carriers (one deletion, two duplications) in cases and four CNV carriers (one deletion, three duplications) in controls. However, subsequent case-control studies in larger schizophrenia patient cohorts have failed to find a significant association at the 2q13 locus [59,97]. In a follow up study Costain *et al.* (2014) undertook detailed phenotyping with two unrelated 2q13 duplication carriers and their families, identified in the 2013 study. Four family members, from one patient pedigree, also carried the duplication and this co-segregated with a neuropsychiatric phenotype. There was a variable psychiatric phenotype, with one psychotic disorder, two major mood and/or anxiety disorders, and one mood and/or anxiety disorder and obsessive compulsive disorder (OCD). The original patient with schizophrenia also had OCD. None of these

individuals had significant DD, ASD or facial dysmorphisms, although three of the family members and one patient had learning difficulties[191].

Riley *et al.* identified three 2q13 deletion carriers and one 2q13 duplication carrier, and compared the phenotype with previous published cases. They concluded that congenital heart defects, hypotonia, dysmorphic features, and abnormal head size are common in deletion carriers and developmental delay, dysmorphic features and abnormal head size are common in duplication carriers. No ASD or psychiatric phenotype was described in these patients, likely because they were too young for clinical assessment[192]. Finally, Hladilkova *et al.* described two additional 2q13 deletion patients, one of whom has ASD and ADHD[156].

A large study of rare CNVs estimated the rate of occurrence of 2q13 deletions and duplications in healthy controls (0.004% deletions, 0.015% duplications), schizophrenia (0.015% deletions, 0.02% duplications) and a mixed developmental disorders (predominantly DD/ID and ASD) cohort (0.057% deletions, 0.022%, duplications)[97]. This suggests that 2q13 CNVs can be observed in healthy controls, but are more common in psychiatric and DD/ID cohorts. The 2q13 CNV is now understood to be a susceptibility locus, which describes a CNV that can be inherited from a healthy or mildly affected parent, but is enriched in individuals with various developmental disorders[35].

A limitation of current 2q13 CNV literature is that few studies have undertaken comprehensive behavioural and psychiatric phenotyping, so the full extent of the neuropsychiatric risk associated with these CNVs remains unclear.

5.2 Aims

This investigation aims to further delineate the 2q13 CNV profile by undertaking deep phenotyping comprising: developmental, medical, dysmorphic, behavioural and psychiatric features.

5.3 Methods

5.3.1 Participant recruitment

In order to maximise recruitment of patients with this rare CNV a multi-faceted approach to recruitment was employed. Unique is a UK-based support group, working internationally to inform and support anyone affected by a rare chromosome or single gene disorder and with professionals involved in their care[193]. The Unique Information Officer identified and emailed registered contacts of Unique members with 2q13 CNVs. Information was provided about the study, and contacts were encouraged to contact the study team if they wanted to participate. Patients with 2q13 deletions were also identified via two NHS RGCs – the North East and South East Thames RGCs (Great Ormond Street Hospital for Children NHS Trust and Guy's and St Thomas' NHS Foundation Trust respectively). Clinicians were approached in the first instance and where appropriate invitations to participate in the study were sent to the patient contact via letter. All participants recruited by KW either lacked capacity to consent to the research, or were deemed too young to provide consent, and consent was provided by a parent for participation in the study. Additionally, patients with 2q13 CNVs on the DECIPHER database were identified and further phenotypic information was sought from responsible clinicians[152]. One participant was included from a previous investigation of CNV in adults recruited from ID psychiatry services[174]. Ethical approval for the study was attained from the North Wales Research Ethics Committee West, reference 11/WA/0370.

5.3.2 Phenotyping and analysis protocol

All participants recruited through Unique and NHS RGCs underwent detailed phenotyping (n=10), whereby clinical data, including medical and psychiatric history, was collected from a parent in a face-to-face interview. This was conducted by KW and interviews were undertaken in person for UK recruits and via Skype for overseas recruits. Responsible DECIPHER contacts were contacted via email to provide further phenotypic data about their patients and anonymised data was collected via the UCL web-based survey tool Opinio (n=15).

All phenotypes were converted to Human Phenotype Ontology (HPO) terms for presentation in the manuscript[194]. The level of ID was taken from available medical records or reported by clinicians and was categorised in accordance with the HPO criteria: borderline intellectual disability (IQ 70-79); mild intellectual disability (IQ 50-69); moderate intellectual disability (IQ 35-49); severe intellectual disability (IQ 20-34).

Psychiatric phenotyping was undertaken using the Mini PAS-ADD for participants over 18 years of age, and the Child and Adolescent Psychiatric Assessment Schedule (ChA-PAS) for those under 18. These assessments provide threshold scores for psychiatric symptoms that are likely to warrant a diagnosis in conjunction with a clinical psychiatric assessment[147]. The Mini PAS-ADD includes ASD screening, but does not include an ADHD assessment. The ADHD section of the CHA-PAS requires a second informant, who is familiar with the individual in other contexts (typically a teacher). It was not possible to interview a second informant for the ChA-PAS, so both sections were completed by the primary informant.

Behavioural phenotyping was undertaken using the Behaviour Problems Inventory - Short Form (BPI-S). The BPI-S provides frequency scores of self-injurious and aggressive/destructive behaviours[195]. Behaviours were reported as present if they were scored at a minimum of a weekly frequency on the BPI-S measure or were documented in the medical history. General observations for dysmorphic features were also made and photographs taken where consent given. Dysmorphic features were independently verified by a second investigator (NB, Consultant Psychiatrist).

Analyses and data visualisation were undertaken using R version 3.4.2 and the ggplot2, Rcmdr and ontologyX packages[173,196–198]. For the breakdown of CNV carriers for each phenotype, deletion and duplication will be abbreviated to del and dup.

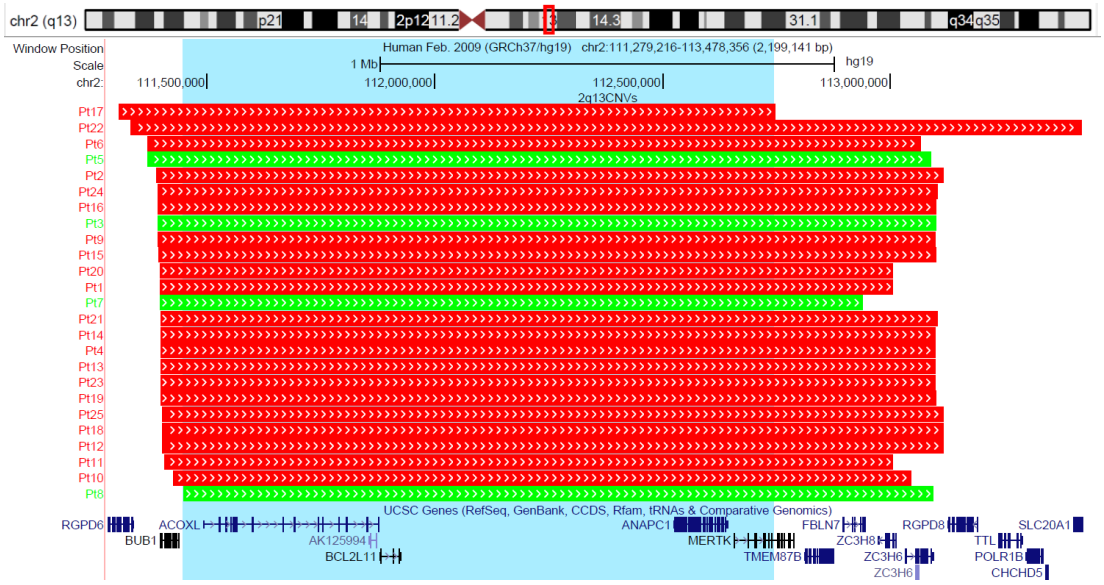
5.4 Results

5.4.1 Sample description

A total of 25 participants were recruited to the study, 10 from the Unique and NHS RGCs group and 15 from the Decipher group (64% male). The participants are

predominantly children (23/25 <18 years of age, median age 9 years, range 4-42 years). The dataset comprises 21 deletion and 4 duplication carriers. The CNVs ranged in size from 1.4Mb to 2.1Mb with a 1.3Mb region of overlap between all CNVs, see Figure 5-1. One family with an inherited 2q13 deletion is included in the case series, a father and two children, as removing the family did not change the results they are presented together with the rest of the cohort.

Figure 5-1: Chromosomal location of the CNV breakpoints for 2q13 CNV carriers



The top image shows the chromosomal location of the CNV, with the region highlighted by a red box, CNV deletions are shown in red and CNV duplications are shown in green, UCSC genes included in the CNV are shown, the blue highlighted region shows the 1.3Mb region of overlap between CNVs, the image was exported from UCSC in chromosomal build GRCh37/hg19.

5.4.2 Inheritance status

For 32% of participants the inheritance status was unknown (9 del, 1 dup). These were all participants from the DECIPHER group, where inheritance information was unavailable to responsible clinicians. A further 20% (5 del) had *de novo* CNVs, 12% (2 del, 1 dup) had a maternally inherited CNV, 28% (5 del, 2 dup) had a paternally inherited CNV, and finally 8% had inherited CNVs but the parental origin was unknown. Focusing on the 12 individuals with inherited 2q13 CNVs, 4 (34%, 2 del, 2 dup) had no family history of ID or mental health problems, 5 had a family history of ID and/or mental health problems (42%, 5 del), and 3 (25%, 2 del, 1 dup) had a family history of ID and/or mental health problems only on the side of the family from which the 2q13 CNV was not transmitted.

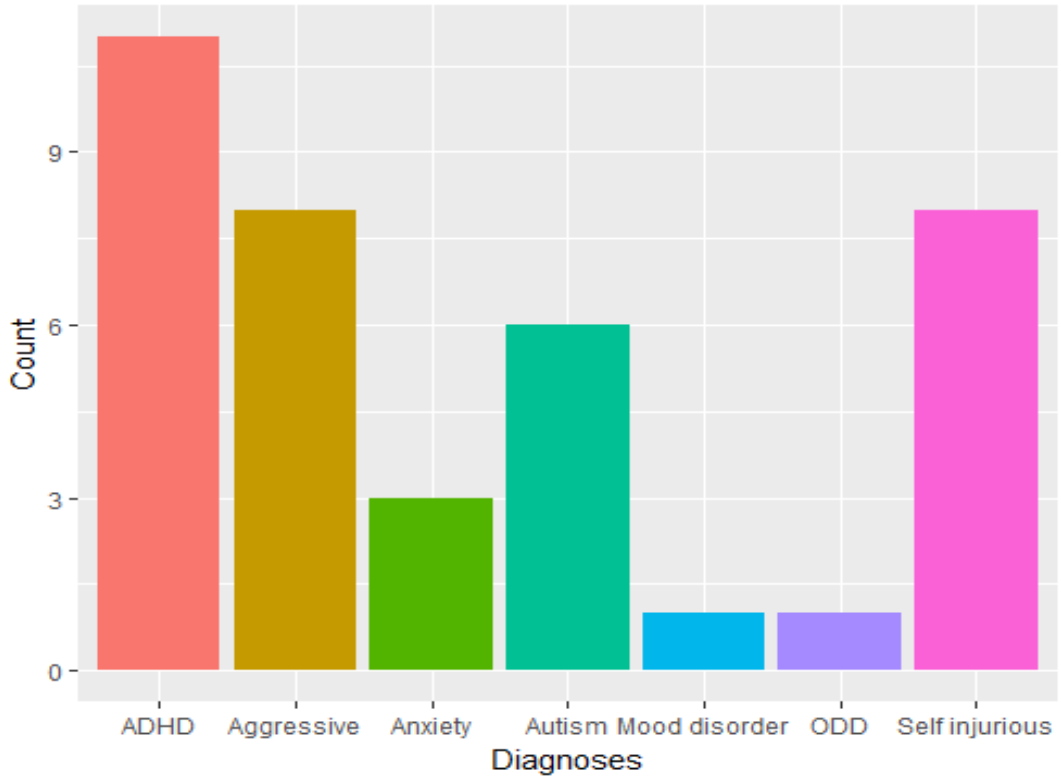
5.4.3 Intellectual and learning difficulties

Overall 76% of participants had DD (15 del, 4 dup). Just over half the participants had an IQ in the borderline or average range (52%, 10 del, 3 dup), and (32%, 8 del) had mild ID. There were no individuals with moderate ID and 12% had severe ID (2 del, 1 dup). We also asked informants or clinicians whether the participants had any other specific learning difficulties, 4 participants (16%) had dyslexia, 2 participants (8%) had dyscalculia, and 2 participants (8%) had an auditory processing disorder (all these were identified in del carriers only).

5.4.4 Psychiatric disorders and challenging behaviours

In total 64% of participants had a formal psychiatric diagnosis, amongst these 44% (9 del, 2 dup) had one diagnosis and 20% (4 del, 1 dup) had two. The most frequently diagnosed psychiatric disorder was ADHD (44%, 9 del, 2 dup), followed by ASD (24%, 5 del, 1 dup) and anxiety disorders (12%, 2 del, 1 dup). Both aggressive and self-injurious behaviours were also identified in the participants, 8 had aggressive behaviours (32%, all del) and 8 had self-injurious behaviours (32%, 7 del, 1 dup), for an overview see Figure 5-2.

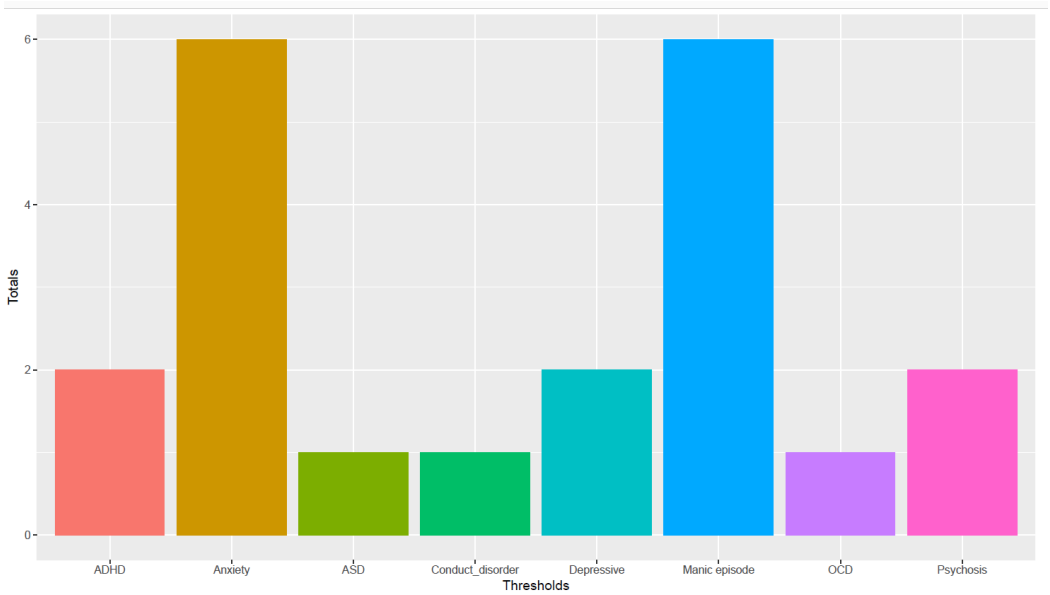
Figure 5-2: Clinically diagnosed psychiatric disorders and behavioural phenotype in 2q13 deletion (n=21) and duplication (n=4) carriers



Y axis: count - the number of participants with the diagnosis or behaviour; X axis: ADHD, attention deficit hyperactivity disorder, Aggressive, aggressive behaviours, Anxiety, anxiety disorder, Autism, autism spectrum disorder, Mood disorder, ODD, oppositional defiant disorder, Self-injurious, self-injurious behaviours.

Of the detailed phenotyping group (n=10), 5 had no clinical psychiatric diagnosis. For two of these participants, both aged 6, ADHD was suspected, but the families were awaiting formal clinical assessment. Additionally, ASD was suspected for one of these participants. Taking into account the PAS-ADD thresholds, 9/10 individuals reached one or more PAS-ADD thresholds. The most frequent thresholds met were anxiety disorder (60%, all del) and manic episode (60%, 5 del, 1 dup) followed by 20% each for ADHD, depressive disorder and psychosis (all del), see Figure 5-3.

Figure 5-3: PAS-ADD thresholds met in the 10 participants in the detailed phenotyping group (9 deletion and 1 duplication carrier)

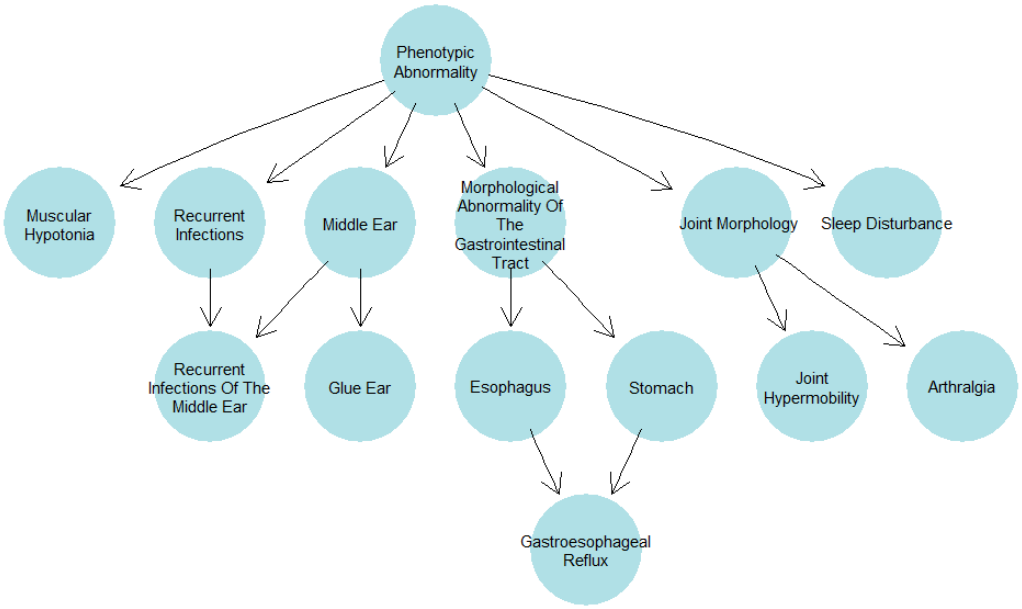


Y axis: totals - number of participants with the diagnosis or behaviour, X axis: Thresholds on the PAS-ADD or CHA-PAS assessments, ADHD, attention deficit hyperactivity disorder, ASD, autism spectrum disorder, OCD, obsessive compulsive disorder.

5.4.5 Medical phenotype

The most commonly observed phenotypes were; glue ear (40%, 9 del, 1 dup), followed by muscular hypotonia (32%, 7 del, 1 dup), sleep disturbances (28%, 6 del, 1 dup), arthralgia (24%, 6 del), recurrent infections of the middle ear (20%, 4 del, 1 dup), joint hypermobility (20%, 5 del), and gastroesophageal reflux (16%, 4 del). See Figure 5-4 for an overview of the systems affected.

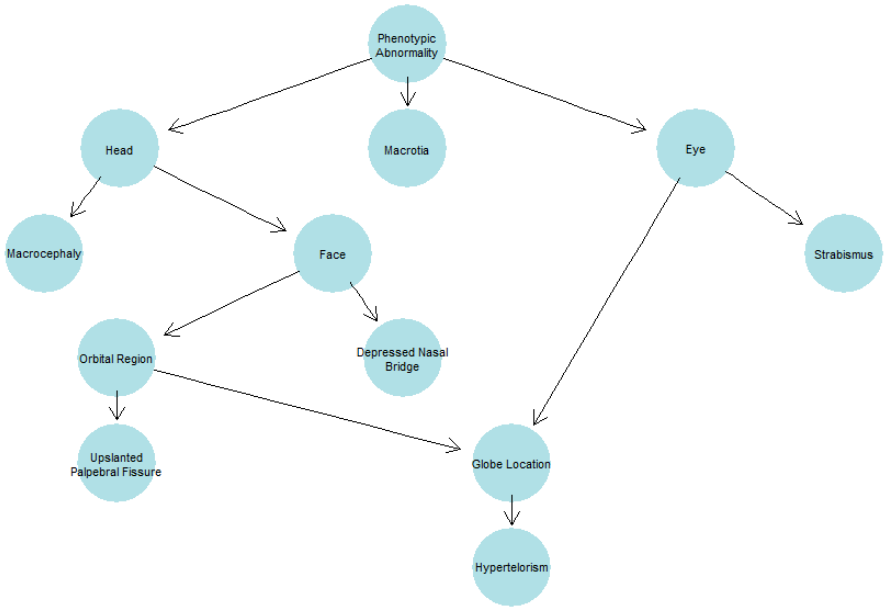
Figure 5-4: Human Phenotype Ontology tree plot with ancestral ontologies for the medical phenotypes occurring in more than three participants



5.4.6 Dysmorphology phenotype

The most commonly observed phenotypes were: macrotia (32%, 8 del), abnormality of the skull (28%, 4 del, 3 dup), macrocephaly (16%, 4 del), upslanted palpebral fissure (16%, 3 del, 1 dup), hypertelorism (16%, 4 del), strabismus (16%, 2 del, 2 dup), and depressed nasal bridge (16%, 4 del). See Figure 5-5 for an overview of the systems affected.

Figure 5-5: Human Phenotype Ontology tree plot with ancestral ontologies for the dysmorphology phenotypes occurring in more than three participants



5.4.7 Comparison with previous literature

To combine phenotypic data for all published 2q13 CNV carriers the dataset initially presented by Hladilkova *et al.* has been adapted (permission via personal correspondence)[156]. All the patients presented are derived from patient case studies, as equivalent phenotypic data was not available for healthy controls carrying these CNVs. For deletion carriers, the cases presented by Hladilkova *et al.* have been added together with new cases from subsequent literature and the cases presented in this study. An equivalent version of the dataset has also been compiled for 2q13 duplication carriers. Table 5-1 presents an overview of all known 2q13 deletion and duplications carriers to date and the phenotypes observed, note the denominator differs due to the varying availability of phenotypic information in published case studies.

Table 5-1: Summary of phenotypic observations in 2q13 participants in conjunction with patient phenotypes presented in previous published studies

	2q13 deletions	2q13 duplications
DD/ID	30/38 (79%)	14/20 (70%)
ASDs	9/27 (33%)	2/12 (17%)
ADHD/ADD	12/25 (48%)	3/5 (60%)
Dysmorphic features	34/40 (85%)	9/10 (90%)
Heart defect	9/34 (26%)	0/5 (0%)
Hypotonia	16/34 (47%)	3/7 (43%)
Seizures	9/31 (29%)	0/10 (0%)
Macrocephaly	10/34 (29%)	1/7 (14%)
Microcephaly	7/34 (21%)	2/7 (29%)

DD, developmental delay, ID, intellectual disabilities, ASDs, autism spectrum disorders, ADHD, attention deficit hyperactivity disorder.

5.5 Discussion

CNVs at the 2q13 locus are rare in the population, can be observed in healthy controls and transmitted from unaffected parents. Despite this, multiple studies have now shown that CNVs at 2q13 are risk factors for DD and dysmorphisms. This study represents the largest ever case series of 2q13 patients, comprising detailed phenotypic data for 25 new cases and combined analysis in 77 individuals, refining our understanding of the phenotypic associations of CNVs at the 2q13 locus.

DD was identified in 76% of participants in this study. Combined with all available data from existing literature 79% of deletion carriers and 70% of duplication carriers have DD/ID. This phenotype has been further delineated by investigating the level of ID, which revealed that the intellectual impairment is generally mild with 52% of participants having IQ in the borderline or average range. Only 12% of participants had severe ID, and data available for two of these participants in the detailed phenotyping group revealed that both were referred for further exome sequencing investigations, due to the 2q13 CNV not being thought to fully explain their phenotype.

Combined analysis reveals that 80% of 2q13 deletion carriers and 90% of 2q13 duplication carriers have dysmorphic features. Deep phenotyping in the new cases showed that macrotia, abnormalities of the skull, macrocephaly, upslanted palpebral fissures, hypertelorism, strabismus, and depressed nasal bridge were common in

deletion carriers. In duplication carriers, abnormalities of the skull and strabismus were observed. No other features achieved more than a single occurrence, however there were only 4 individuals who had 2q13 duplications. Combined analysis identified 29% of deletion carriers and 14% of duplication carriers as having macrocephaly and 23% of deletion carriers and 29% of duplication carriers as having microcephaly.

Previous 2q13 CNV literature has described congenital heart defects, hypotonia and seizures as associated medical phenotypes. Combined analysis found that 31% of deletion carriers had heart defects, and this phenotype was not observed in duplication carriers. Combined analysis identified 44% of deletion carriers and 43% of duplication carriers as having hypotonia, supporting previous results on the association of the 2q13 deletion with this feature and extending this to also affect duplication carriers. Seizures were only observed in deletion carriers at a frequency of 26%. Deep phenotyping in this study also associated novel medical phenotypes with 2q13 CNVs, including: glue ear, sleep disturbances and recurrent infections of the middle ear, both in deletion and duplication carriers, and arthralgia, joint hypermobility and gastroesophageal reflux in deletion carriers only.

A limitation of published 2q13 case reports is that many are in young children, who are below the typical assessment age for various psychiatric disorders. Also, it is unclear in some studies whether comprehensive behavioural and mental health assessments have taken place. The new cases presented in this study had a median age of 9 years, and 64% already had a clinical psychiatric diagnosis. Some of the remaining participant either had suspected psychiatric disorders, which had yet to be formally tested, or met PAS-ADD thresholds, indicating that this figure could be even higher. To our knowledge, challenging behaviours have never previously been assessed in 2q13 CNV carriers and we found both aggressive and self-injurious behaviours to be present in deletion carriers and self-injurious behaviours in one duplication carrier.

Despite the aforementioned limitations, previous case reports of individuals with 2q13 CNVs have reported both ASD and ADHD diagnoses. Combining analysis identified 48% of deletion carriers and 60% of duplication carriers as having an ADHD diagnosis, and 33% of deletion carriers and 17% of duplication carriers as having an ASD diagnosis. Both 2q13 deletions and duplications have also been identified in

schizophrenia patients. We did not identify any participants with schizophrenia, although only three individuals were over the age of 16, so the typical age of onset was not reached in most individuals. Our study identifies a strikingly high incidence of ADHD in 2q13 CNV carriers. A literature review of genes in the 2q13 region was undertaken and no prior association of genes in this region with risk for ADHD was identified, although postulations have been made about the involvement of genes in the region in other neuropsychiatric phenotypes.

A 1.3Mb common region of overlap was identified in CNV carriers, disrupting four genes: *ACOXL, BCL2L11, ANAPC1 and MERTK*. It has been suggested that disruption of the *ACOXL* and *BCL2L11* genes may contribute to neurodevelopmental and ASD phenotypes[190]. The *ACOXL* gene encodes a protein responsible for fatty acid oxidisation, alterations in fatty acid metabolism have been proposed to play a role in the pathogenesis of ASD[199]. *BCL2L11* encodes a neuronal apoptosis regulator and previous research has found decreased expression of this gene in the frontal cortex and cerebellum of autistic subjects. It has been hypothesised that an increase in apoptosis in these regions may contribute to the pathogenicity of autism[200]. *ANAPC1*, a neurodevelopmental facilitator, and *MERTK*, a TAM receptor and multiple sclerosis risk gene, have also been proposed as candidate genes for the psychosis phenotype of 2q13 CNV carriers[157]. All but one participant had CNVs which extend distally to include *FBLN7* and *TMEM87B.* Russell *et al.* undertook a functional analysis of candidate genes in the 2q13 region using zebrafish morpholino knockdowns. They found that depletion of *FBLN7* and *TMEM87B* orthologues resulted in cardiac hypoplasia and *FBLN7* additionally was associated with craniofacial abnormalities[201].

One theory as to why some CNVs show incomplete penetrance is that a second genetic hit is required to unmask the predisposition to a neuropsychiatric phenotype[202]. None of the participants in this study had another CNV that had been classified as pathogenic. However, as sequencing data was not available for analysis, it cannot be ruled out that the participants had another genetic variant contributing to their phenotype. Yu *et al.* recently identified a paternally inherited variant in the *TMEM87B* gene, one of the genes in the 2q13 region associated with the cardiac phenotype[201], in a patient with a severe cardiac phenotype who also had a maternal 2q13 deletion. It is

thought that the unmasking of this homozygous variant by the maternal deletion acted as a second genetic hit, resulting in the severe phenotype [203]. The inheritance pattern of CNVs was mixed, as was the family history of ID and mental health problems. It is of interest that 25% of participants had a family history of ID and mental health problems on the other side of the family from which the variant was transmitted. This could provide support to the second-hit hypothesis, but further genetic investigations would be required.

One of the benefits of receiving a genetic diagnosis for patients and their families is the ability to access diagnosis specific information, which could be used to facilitate early intervention screening for associated medical and psychiatric phenotypes. Disorder guides, written for both professionals and families, are available for 2q13 CNVs from the patient support group unique[135]. The findings from combined analyses in this study could also guide clinical management of individuals with newly diagnosed 2q13 CNVs, for example screening for ADHD might be considered. However, it must be acknowledged that whilst we find some phenotypes to occur more frequently, 2q13 CNV carriers still display variable phenotypic outcomes – posing challenges for genetic counselling of patients and their families.

Many studies of rare CNVs have been undertaken in paediatric cohorts, and comprehensive psychiatric and behavioural phenotyping has not been carried out. The degree to which neuropsychiatric phenotypes are common in rare CNV carriers has, therefore, yet to be established – making it difficult to find an appropriate comparison group for the frequency of psychiatric diagnoses. The medical and dysmorphology phenotypes described are also diverse, and it is difficult to ascertain 'core symptoms' of the disorder which could be an indicator for genetic testing for 2q13 imbalances. However, this study had a comprehensive phenotyping protocol and it may be that a similar pattern of diverse phenotypes would be observed if comparable phenotyping approaches were to be used in other studies investigating rare neurosusceptability CNVs. One unusual finding is that both micro and macrocephaly were present in both deletion and duplication carriers, whereas in other CNVs (such as 16p11.2) it is more typical to find an excess of one of the phenotypes segregating with the CNV type. There is natural variation in head size in the general population and the ideal method

of measurement would be deviation from parental means, however it was not possible to collect parental measurement in this analysis. It may be that the observations have arisen from normal variation, rather than being associated with the 2q13 CNVs. Increasing the depth of phenotyping in rare CNV studies is an important avenue for future research, as well as conducting ongoing mental health assessments in 2q13 carriers to elucidate associations with psychiatric disorders across the life course. Additionally, further studies of the unaffected parents and healthy controls with 2q13 CNVs will be important to elucidate potential protective factors.

The limitations of this study are that observations are being made in participants who have presented to clinical services. This may create an ascertainment bias, whereby the most severe cases are described. However, the accumulation of cases from a wide range of sources attempted to ensure as representative a sample as possible. The assessments of dysmorphology were not conducted by a clinical dysmorphology expert, although we utilised a second rater to improve the reliability of the observations. The PAS-ADD and ChA-PAS schedules were completed by a researcher, and clinical verification by a trained psychiatrist did not take place. Some of the participants were as young as four, meaning some of the later-onset phenotypes could not be accurately measured at this age. However, if anything this would have led to an under estimation of the phenotype frequencies.

5.6 Conclusion

In the largest study of 2q13 CNVs to date, we present detailed phenotypic data for 25 new 2q13 deletion and duplication carriers. Combining this with previous literature yields a total of 54 deletion and 23 duplication carriers, enabling a refined understanding of the phenotypic associations of CNVs at the 2q13 locus. Combined analysis predominantly supports existing literature on an increased rate of developmental, medical and dysmorphic phenotypes. Psychiatric investigations reveal that the majority of deletion and duplication carriers have been clinically diagnosed with a psychiatric disorder, with a particularly high incidence of ADHD. This could have important implications for psychiatric screening upon clinical diagnosis of 2q13 CNVs, and further investigation of this region may have some relevance to understanding the neurobiology of ADHD.

Chapter 6 Relative burden of rare CNVs in different neurodevelopmental cohorts

6.1 Introduction

The previous chapters have focused on the analysis of CNVs that have already been classified (as pathogenic, VOUS and benign) using clinical laboratory protocols for variant categorisation. As discussed previously, there are many factors involved in this variant classification process. For example, classification parameters include: the size and genic content of the CNV, the inheritance status, and population frequencies in reference disease and control cohorts. Another methodology used in the research literature to test the association of a class of genetic variants is a burden analysis. Burden analyses compare the collective frequency of variants in two groups, typically cases and controls. Typically, in CNV burden analyses, common CNVs are filtered out and the analysis is undertaken on the rare CNVs that are most likely to be implicated in disease pathology (<1% population frequency),.

There are some advantages to undertaking CNV burden analyses for testing the differences between groups. Many CNVs are individually rare and evolutionarily selected against. This means that extremely large sample sizes are required to detect statistically significant differences in CNV frequencies. Whereas, CNV burden analyses enable comparisons of rare CNVs beyond those that are currently deemed pathogenic, working under the general assumption that a greater CNV burden is likely to be associated with a greater propensity to disease[142].

One of the earliest schizophrenia CNV burden analyses was undertaken by the International Schizophrenia Consortium in 2008, comprising 3,391 patients and 3,181 controls. The analysis focused on rare CNVs (<1% frequency) >100kb in size. Controls on average had 0.99 CNVs per person and the rate was increased 1.15-fold in patients with schizophrenia to 1.14 CNVs per person. When stratifying by size of event, larger (>500kb) deletions were enriched, whilst the opposite was true for duplications – with shorter duplications showing a stronger association with disease[204]. Analysis of genome-wide data from case-control studies of DD/ID have also revealed an increased CNV burden in DD/ID cases. As discussed previously Coe *et al.* analysed

data from 15,767 children with developmental disorders and 8,329 unaffected adult controls. An excess of large CNVs was identified in cases and the effect was more pronounced with increasing CNV size. At a threshold of 400kb ~25.7% (4,047 cases), compared to 11.5% of the controls harboured a CNV event of this size[45].

Girirajan *et al.* assessed the relative contribution of CNV burden by undertaking analyses in three distinct NDDs – ID, ASD and dyslexia – comprising 1,227 cases and 337 controls. ID was most associated with a greater CNV burden. The three NDD groups were further characterised, as to their phenotypic severity, and it was identified that the most severe phenotype is correlated with a greater size of CNV and a greater gene density of genes affected by the CNV. The phenotype groups in order of severity are: ID with MCA, idiopathic ID, ID with ASD, ASD without ID, dyslexia, controls. No differences were identified when segregating by CNV type (deletion and duplication), although analysis of inheritance status revealed a trend of increased *de novo* CNVs with increased severity of the disorder[140].

CNV burden has also been investigated in patients with ID and co-morbid schizophrenia. Derks *et al*. studied patients with ID only (n=66), versus patients with ID and schizophrenia (n=64). No differences were found in the burden of CNV deletions and duplications >100kb in patients with ID only versus patients with ID and co-morbid schizophrenia. However, a higher burden of duplications larger than 1Mb was identified in patients with ID and schizophrenia. This was largely driven by duplications at the 15q11.2 region[141]. Lowther *et al*. also investigated CNV burden in 546 schizophrenia patients who were segregated into various IQ groups, ranging from low (< 85) to average (≥ 85) IQ. They identified a significantly (p=0.002) increased burden of rare genic duplications in individuals in the low IQ schizophrenia group. This higher burden persisted even after excluding individuals with a pathogenic CNV. [142].

CNV burden analyses have also been undertaken in patients with BPAD, however the results do not follow the expected trend of an increased CNV burden with increased phenotypic severity. In fact, multiple studies have shown that there is actually a reduced burden of rare large CNVs in BPAD patients as compared to controls[205–207]. One study by Malhotra *et al*. investigated parent proband trios with BPAD (n=185)

and healthy controls (n=426) and found an increased frequency of large de novo CNVs in BPAD patients. However, this effect only remained significant when considering BPAD cases with an age of onset below 18 years of age[208]. Noor et al. also found no increased burden of CNVs in BPAD cases as compared to controls[209].

Marshall *et al.* conducted the largest genome-wide CNV burden analysis for any psychiatric disorder to date, comprising 21,094 schizophrenia cases and 20,227 controls. Several parameters were significantly greater in schizophrenia cases, as compared to controls, including: total CNV distance (kb) covered, number of genes affected by CNVs and the number of CNVs. When split by CNV type, the effect size for CNV deletions was greater than for CNV duplications. Interestingly, this enrichment in CNV burden persisted even after exclusion of CNV loci implicated as schizophrenia risk factors in previous studies. Furthermore, CNV burden was enriched for genes associated with synaptic function and neurobehavioral phenotypes in mice[59].

There are now several lines of evidence to support the theory that there is an increased CNV burden as the severity of the NDD phenotype increases. ID is generally the most severe phenotype observed – although ID with MCA or ID with co-morbid schizophrenia are more severe phenotypes associated with a greater CNV burden. BPAD is the only psychiatric phenotype, that has been systematically investigated, which does not have an increased CNV burden as compared to controls. It is also towards the milder end of the spectrum of developmental and psychiatric disorders investigated. Some studies only find this increased CNV burden to be true for certain types of CNV and for CNVs of particular size. For example, in ID patients with co-morbid schizophrenia CNV duplications appear to be driving the significant differences in CNV burden. In one study, the significant difference was further limited duplications larger than 1Mb. To the best of my knowledge, no previously published study has compared the CNV burden between ID plus co-morbid mental disorders and schizophrenia.

6.2 Aims

The aim of this analysis is to compare CNV burden in a cohort of patients with ID plus co-morbid psychiatric disorders (ID+) with CNV burden in schizophrenia patients and healthy controls. The impact of CNV size and type will also be considered.

6.3 Methods

6.3.1 Sample collection and DNA preparation

The samples analysed in this section comprise participants from the DNA variation in adults with learning disability sample, described in Chapter 3 and Chapter 4, for which genomic DNA was available for analysis (n=228 out of N=248). Participants with large-scale chromosomal abnormalities were excluded from the analysis. Sample collection for this dataset has been described in Chapter 3. The UCL schizophrenia and control samples were derived from the DNA polymorphisms in Mental Illness (DPIM) study have been described elsewhere[210]. Briefly, patients with a clinical ICD-10 diagnosis of schizophrenia were recruited through NHS services. Diagnoses were confirmed and additional phenotypic data collected with the Schizophrenia and Affective Disorders Schedule (SADS-L)[211]. Controls were recruited with an absence of personal history of mental illness as well as an absence of mental illness in first-degree relatives. Data from a subset of the schizophrenia (N=1529), BPAD (N=1445) and control (N=1285) cohorts were used in this analysis. For the ID+ sample 17% of the sample (38/228) were of non-European ancestry, all of the DPIM participants were of European ancestry. DNA extraction for the ID+ study was undertaken at the North East Thames Regional Genetics Service Laboratory and for the DPIM sample DNA was extracted in-house. DNA quantifications were undertaken using the Qubit quantification protocol[212]. A number of students and Post-Doctoral researchers quantified the DNA samples from the DPIM study, I personally quantified all the samples for the ID+ study.

6.3.2 Genotyping and Quality Control

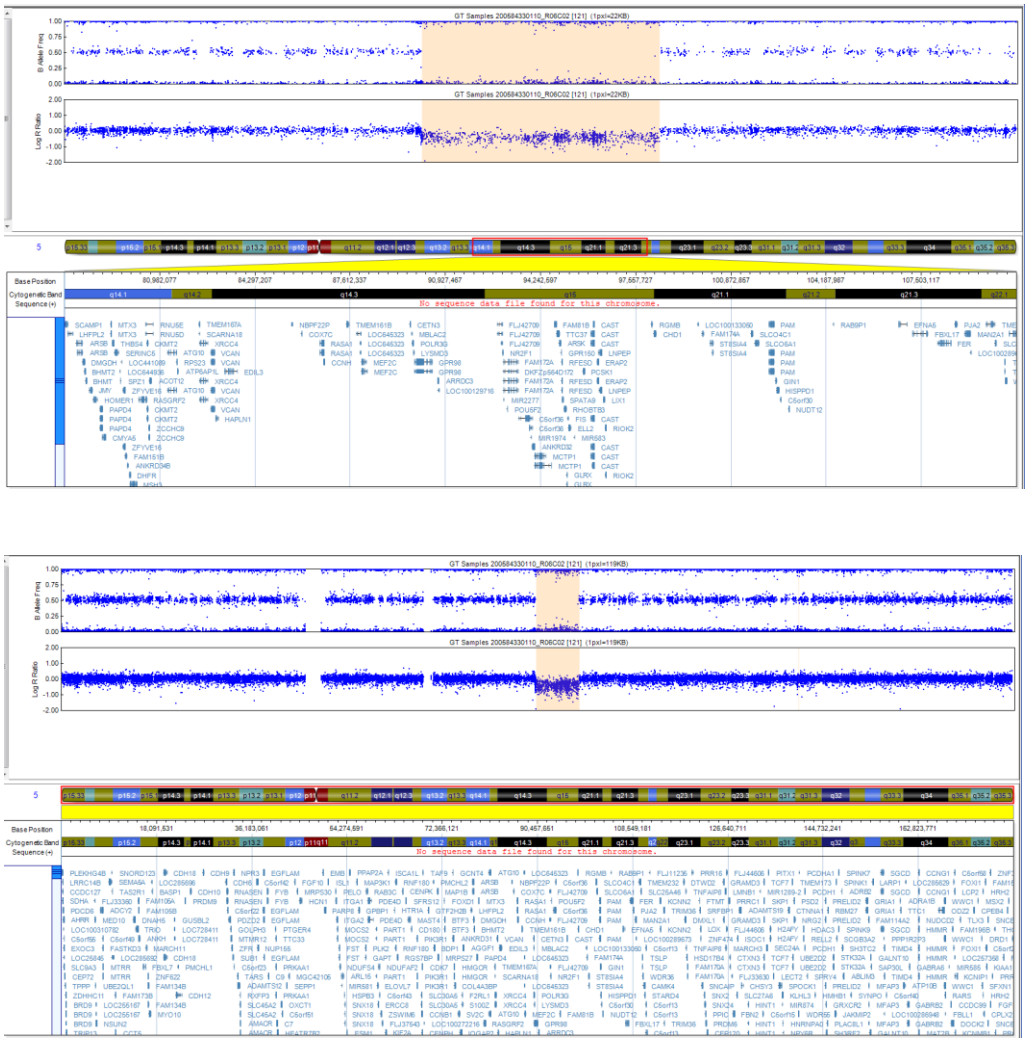
The PsychChip, also known as the PsychArray, is a high density (>500,000 markers) customised microarray chip containing 265,000 proven tag SNPs found on the

Infinium Core-24 BeadChip, 245,000 markers from the Infinium Exome-24 BeadChip, and 50,000 markers known to be associated with psychiatric disorders[213]. For the DPIM samples genotyping was performed at the Broad Institute of MIT and Harvard and the ID samples were genotyped at the University of Bonn.

Quality control of the data was undertaken in PLINK[214]. Samples were excluded if they: did not match the SNP sex, if they had excessive heterozygosity (more than 5 SD above the mean, measured using common SNPs, MAF>0.05), if more than 5% of the SNPs genotyped had missing information, and SNPs failing Hardy-Weinberg equilibrium (HWE), a test which measures if the observed allelic distribution fits within the expected distribution. This is in line with quality control protocols described in relevant literature[215].

For the ID+ dataset the PsychChip raw intensity files, (*.idat), were converted into *.gtc files via Illumina's GenomeStudio software[216], *.gtc files were supplied by the Broad Institute for the DPIM samples. The signal intensity file was exported from GenomeStudio containing, SNP information, B allele frequency (BAF) and log R ratio (LRR). The BAF is a normalised representation of how often the B allele is called. A normal BAF plot has three distinct bands, with homozygous calls at the top and bottom and heterozygous in the middle. Absence of the middle band could be indicative of a CNV deletion, whereas the presence of extra bands could be indicative of a CNV duplication. LRR is a metric that normalises signal intensity for CNV analysis, when the fluorescence values are above 0 this might indicate a CNV duplication and florescence below 0 could indicate a CNV deletion. See Figure 6-1 for an example of the BAF and LRR plots exported from GenomeStudio for an individual from the ID+ sample. Data was exported from GenomeStudio for the ID+ sample in accordance with the methods described on the PennCNV website[217].

Figure 6-1: CNV plots for a participant with a CNV deletion on chromosome 5



Plots exported from GenomeStudio, The top image shows the log R ratio (LRR) and B allele frequency (BAF) of each marker, with the CNV deletion region highlighted, below is the region of chromosome 5 that is affected and the implicated genes, the bottom image has the same content as the top but the perspective is zoomed out to show the whole of chromosome 5.

6.3.3 CNV calling in PennCNV

Postdoctoral Researcher Dr Johan Thygesen wrote the scripts 'prep_calling.sh', 'call_cnvs_array.sh', and 'post_calling.sh', to call the CNVs via PennCNV for the DPIM dataset. I ran these scripts to call the CNVs for the ID+ dataset to ensure methodological consistency. The script 'prep_calling.sh' modifies the signal intensity file to the format which is accepted by PennCNV. The script 'call_cnvs_array.sh' calls CNVs from the signal intensity data using the PennCNV algorithm. Finally, the 'post_calling.sh' script performs post calling quality control checks. These scripts were created in accordance with the CNV calling guidelines on the PennCNV website[218]. Samples with more than 300 CNVs, with a BAF drift bigger than 0.01 and

with a LRR SD bigger than 0.5 were removed. Copies of these scripts are available in the appendix. Both autosomal and X chromosome CNVs were called. The X chromosome is of particular interest to the ID phenotype, given that 5-10% of ID in males is caused by genetic variants on the X chromosome and over 150 X linked ID syndromes have been identified[219]. It is typical to undertake X chromosome burden analyses separately for males and females, given the sex differences in the number of X chromosomes. The participants in the ID+, schizophrenia, and healthy control groups with CNVs called on the X chromosome were separated by sex, see Table 6-1. It was not possible to undertake the X chromosome analysis segregated by sex given the low number of females in the ID+ group. The analysis protocol will therefore focus on autosomal CNVs.

Table 6-1 The number of males and females with X chromosome CNV calls in the three participants groups

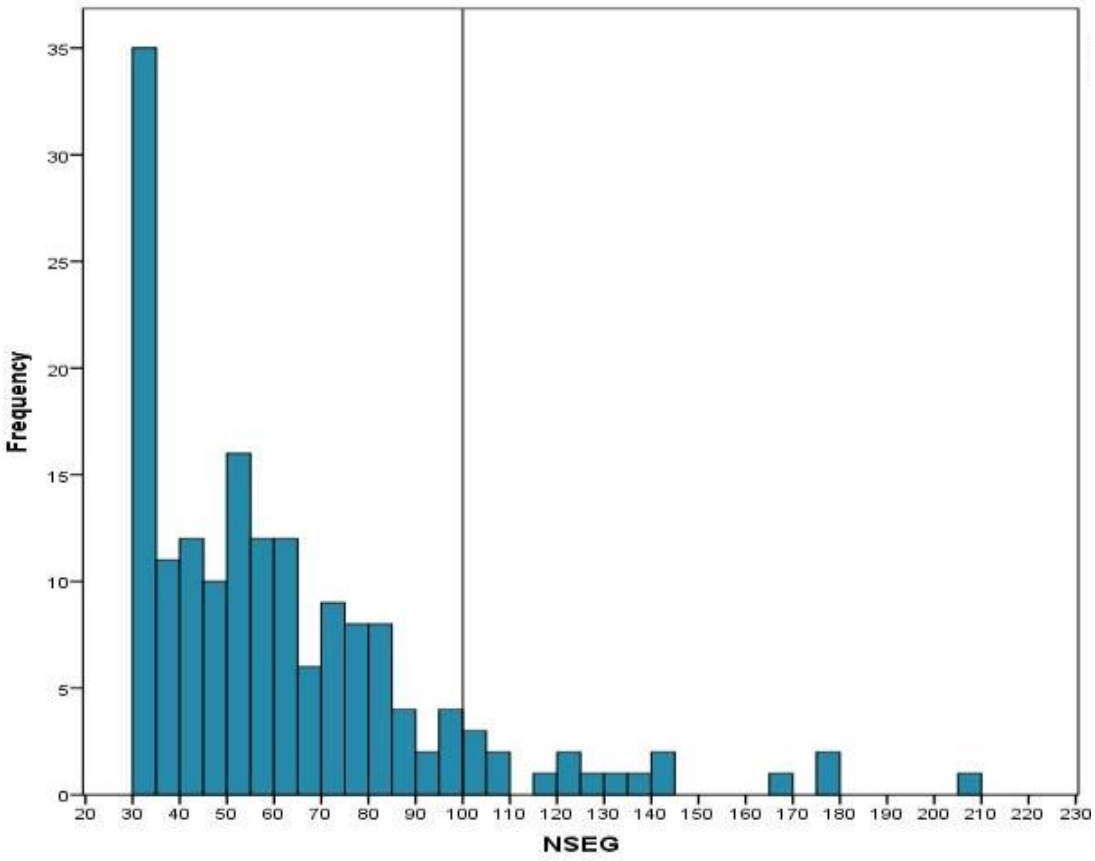
	Female (%)	**Male (%)**
Controls	266 (20.7)	347 (27.0)
Schizophrenia	65 (4.3)	765 (50)
ID+	20 (8.8)	89 (39.0)

6.3.4 CNV burden analysis – PLINK

The PennCNV output files were converted to the PLINK format. As the focus of the CNV burden analysis was rare autosomal CNVs, a new CNV file was created on a filtered subset of the data. It is typical to use a frequency cut off of <1% for rare CNVs[59,204], therefore CNVs were excluded that were present in >1% of the dataset. This frequency filter removed 74,112 CNVs from further analyses. CNV segments were considered to cover the same region if the minimum reciprocal breakpoint overlap was at least 50%. For the remaining CNVs the number of CNVs per individual was plotted to investigate individuals with a greater than expected number of CNVs. Taking a cut off of greater than 30 CNVs per individual, a histogram was plotted using IBM SPSS Statistics for Windows, Version 24.0 (IBM Corp, Armonk, NY, USA), see Figure 6-2. Furthermore, we decided to reduce the exclusion cut off of number of CNVs per individual from 300 to 100 to ensure a conservative cut off for possible falsely called CNVs. This resulted in 17 participants being excluded from further analyses. The BPAD cases were included for the frequency filter, to increase the sample size and reliability of the filter, however are not included in further analyses as

research has shown that there patients with BPAD have a lower burden of large rare CNVs than controls..

Figure 6-2: Histogram of number of CNV segments per person in individuals with 30 or more CNVs



Y axis: frequency, the number of individuals, X axis: NSEG, number of CNVs (or segments – the term used in the PLINK software), participants to the right of the vertical line (>100 CNVs per individual) were excluded from further analyses.

A MAP file was then created, which maps the start and stop of CNV segments facilitating subsequent CNV parsing and analysis. Burden analyses of segmental CNV data were undertaken, which compares the metrics of cases versus controls evaluated by permutation. A phenotype file was included to enable separate comparisons: ID+ versus controls, schizophrenia versus controls and ID+ versus schizophrenia. In accordance with previous literature CNVs >100kb were analysed[204], as calling of smaller CNVs can be less reliable and the primary interest is the role of rare and large CNVs in disease pathology. Additional size cut offs, >200kb and >400kb, were also considered. Furthermore, the role of CNV type, deletion or duplication, was also

considered. The script for this analysis was written by myself in accordance with the protocols described on the PLINK website[220]. An excerpt from this script for the ID versus controls analysis at a 100kb cut off for all CNVs, CNV deletions only, and CNV duplications only is detailed below.

6.3.5 CNV analysis script

```
#Creating a new CNV file on a filtered subset - excludes CNVs present more than 45 times (1% frequency in a 4,487 sample size) with a minimum reciprocal overlap of 50%
./plink1 --noweb --cnv-list pchip_comorbid_auto_over100rem.cnv --fam pchip_comorbid.fam --map plink.cnv.map --cnv-freq-exclude-above 45 --cnv-overlap 0.5 --cnv-write --out rarecnvless1per
./plink1 --noweb --cnv-list rarecnvless1per.cnv --cnv-make-map --out rarecnvless1per

#id versus controls all CNVs over 100kb in size
./plink1 --noweb --map rarecnvless1per.cnv.map --cnv-list rarecnvless1per.cnv --fam rarecnvless1per.fam \
--cnv-indiv-perm --mperm 10000 \
--cnv-kb 100 \
--pheno pchip_comorbid.pheno --pheno-name $pheno \
--out $pheno.allover100kb

#id versus controls CNV del over 100kb in size
./plink1 --noweb --map rarecnvless1per.cnv.map --cnv-list rarecnvless1per.cnv --fam rarecnvless1per.fam \
--cnv-indiv-perm --mperm 10000 \
--cnv-kb 100 \
--cnv-del \
--pheno pchip_comorbid.pheno --pheno-name $pheno \
--out $pheno.100kbdel

#id versus controls CNV dup over 100kb in size
./plink1 --noweb --map rarecnvless1per.cnv.map --cnv-list rarecnvless1per.cnv --fam rarecnvless1per.fam \
--cnv-indiv-perm --mperm 10000 \
--cnv-kb 100 \
--cnv-dup \
--pheno pchip_comorbid.pheno --pheno-name $pheno \
--out $pheno.100kbdup
```

6.4 Results

The PLINK burden analysis of segmental CNVs reports four comparisons between cases and controls. The rate test compares the rate (number of segments or CNVs) per

person, the prop test compares the proportion of cases/controls to have at least one event, Totkb compares the total distance spanned by segments or CNVs per person, and the Avgkb compares the average segments or CNV event size per person. Results are presented uncorrected for multiple testing, in accordance with equivalent literature, where multiple testing corrections are only applied to breakpoint and gene-based association tests[59]. As the ID+ sample comprised participants with non-European ancestry (38/228), post-hoc analyses were also undertaken with these individuals removed. Where this changed the significance of the results this has been indicated.

6.4.1 ID plus mental illness versus controls

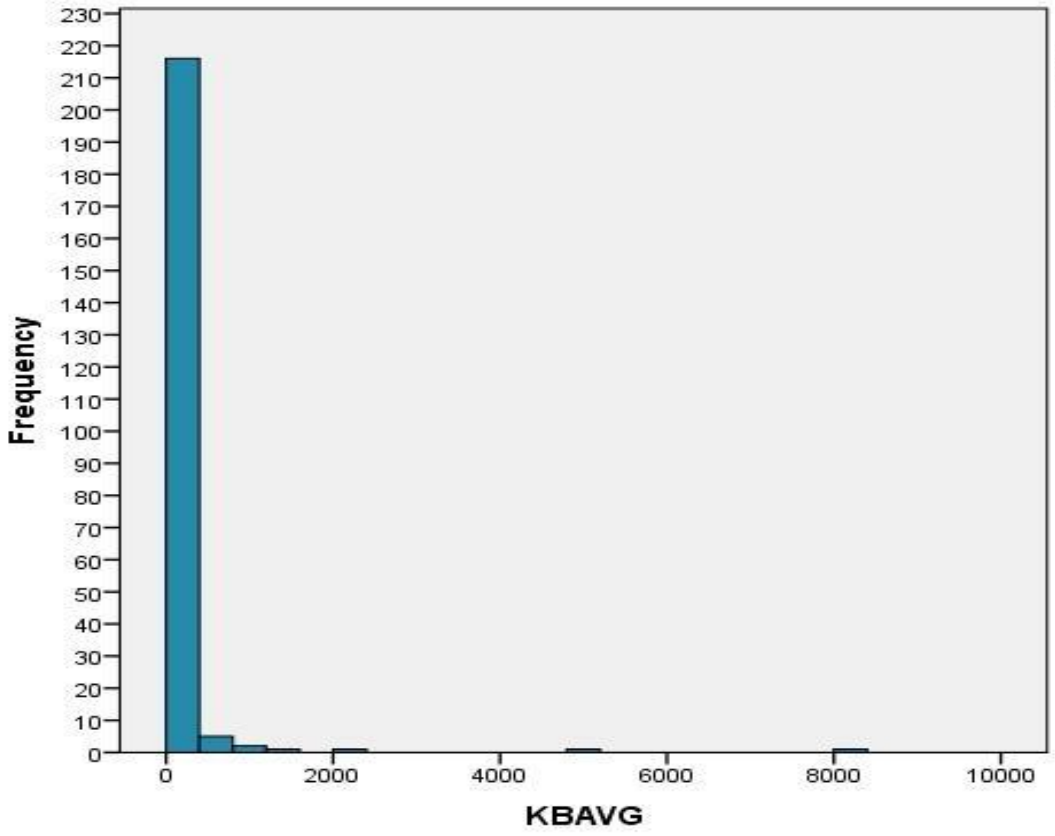
The results from the ID plus co-morbid mental disorders (ID+) versus control comparisons can be seen in Table 6-2. ID+ cases have a significantly increased rate of CNVs per person, as compared to controls. When filtering on CNV type this only remains significant for deletion CNVs and this remains significant at all size cut offs. The proportion of ID+ cases versus controls to have at least one event is only significant when considering CNVs >200 and >400kb, again filtering by CNV type this only remains significant for deletion CNVs. The total kb spanned by CNVs is significantly different between ID+ cases and controls except for at the 400kb cut off, again this only remains significant for deletion CNVs. For the average kb, or CNV event size per person, there is only a significant difference at the >100kb cut off and this is only significant when including both deletion and duplication CNVs. This result no longer remains significant when removing individuals with non-European ancestry. To further investigate the distribution of average CNV kb per individual a histogram was plotted, see Figure 6-3, and it appears that the significant difference between cases and controls is being driven by a small number of individuals who have high average CNV size in the ID+ group.

Table 6-2: CNV burden permutation test results for ID+ versus controls

ID+ (n=228) versus controls (n=1285)							
	Size	**CNVs cases**	**CNVs controls**	***P* Values**			
CNV type	**(kb)**	**(n)**	**(n)**	**Rate**	**Prop**	**TotKb**	**AvgKb**
Del+Dup	**>100**	818	1277	**1.00E-04**	0.082892	**1.00E-04**	**0.0046**
	>200	294	437	**1.00E-04**	**0.0009**	**1.00E-04**	0.067993
	>400	75	140	**1.00E-04**	**1.00E-04**	0.044296	0.209179
Del	**>100**	695	697	**1.00E-04**	0.017598	**1.00E-04**	0.013899
	>200	236	180	**1.00E-04**	**1.00E-04**	**0.0014**	0.190981
	>400	49	46	**1.00E-04**	**1.00E-04**	0.140486	0.294671
Dup	**>100**	123	580	0.19678	0.39836	0.112189	0.368063
	>200	58	257	0.159784	0.59644	0.082392	0.283572
	>400	26	94	0.094991	0.344966	0.10179	0.432957

Key: Rate = Number of segments, Prop = Proportion of sample with one or more segment, Totkb = Total kb length spanned, Avgkb = Average segment size, Del = deletion, Dup = duplication, Bold = results that are significant at a significance threshold of P <0.05

Figure 6-3 The average kb of CNVs for individuals from the ID+ group (N=228)



6.4.2 Schizophrenia versus controls

There were no significant differences identified in CNV burden between schizophrenia cases and controls, see results in Table 6-3.

Table 6-3: CNV burden permutation test results for schizophrenia versus controls

Schizophrenia (n=1529) versus controls (n=1285)							
	Size	**CNVs cases**	**CNVs controls**	***P* Values**			
CNV type	**(kb)**	**(n)**	**(n)**	**Rate**	**Prop**	**TotKb**	**AvgKb**
Del+Dup	**>100**	1241	1277	0.674633	0.40256	0.654835	0.09669
	>200	460	437	0.371663	0.446955	0.534847	0.560644
	>400	150	140	0.343666	0.359264	0.707929	0.687031
Del	**>100**	733	697	0.394561	0.418458	0.465653	0.290071
	>200	203	180	0.323468	0.231577	0.627637	0.735226
	>400	44	46	0.647135	0.70213	0.456454	0.487551
Dup	**>100**	508	580	0.906309	0.487951	0.79732	0.053495
	>200	257	257	0.565943	0.257374	0.811919	0.305869
	>400	106	94	0.171983	0.161484	0.534647	0.728327

Key: Rate = Number of segments, Prop = Proportion of sample with one or more segment, Totkb = Total kb length spanned, Avgkb = Average segment size, Del = deletion, Dup = duplication, Bold = results that are significant at a significance threshold of $P < 0.05$

6.4.3 ID+ versus schizophrenia

ID+ cases have a significantly increased rate of CNVs per person compared to schizophrenia only cases, see Table 6-4. When filtering on CNV type this only remains significant for deletion CNVs and this remains significant at all size cut offs. The proportion of ID+ cases versus schizophrenia to have at least one event is only significant when considering CNVs >200 and >400kb, again filtering by CNV type this only remains significant for deletion CNVs. The total kb spanned by CNVs is significantly different between ID+ and schizophrenia cases except for at the 400kb cut off, again this only remains significant for deletion CNVs.

Table 6-4: Permutation results for ID+ versus schizophrenia

ID+ (n=228) versus schizophrenia only (n=1529)							
	Size	**CNVs cases**	**CNVs controls**	***P* Values**			
CNV type	**(kb)**	**(n)**	**(n)**	**Rate**	**Prop**	**TotKb**	**AvgKb**
Del+Dup	**>100**	818	1241	**1.00E-04**	0.10159	**1.00E-04**	0.011499
	>200	294	460	**1.00E-04**	**0.0007**	**1.00E-04**	0.051395
	>400	75	150	**1.00E-04**	**1.00E-04**	0.023498	0.144386
Del	**>100**	695	733	**1.00E-04**	0.019898	**1.00E-04**	0.020398
	>200	236	203	**1.00E-04**	**1.00E-04**	**0.0004**	0.123188
	>400	49	44	**1.00E-04**	**1.00E-04**	0.142086	0.282772
Dup	**>100**	123	508	0.071693	0.39796	0.093291	0.674033
	>200	58	257	0.132287	0.742826	0.046795	0.380462
	>400	26	106	0.134087	0.573343	0.057694	0.283672

Key: Rate = Number of segments, Prop = Proportion of sample with one or more segment, Totkb = Total kb length spanned, Avgkb = Average segment size, Del = deletion, Dup = duplication, Bold = results that are significant at a significance threshold of $P < 0.05$

6.5 Discussion

A greater burden of large CNVs was identified in ID+ individuals, as compared to controls. Also a higher proportion of individuals with ID+ have at least one CNV event, as compared to controls, for CNVs >200 and >400kb. Whilst there has been no previous burden analysis in an ID+ cohort, analyses in paediatric cohorts with severe developmental disorders support this trend of greater CNV burden for individuals with ID and co-morbidities[45]. When segregating by CNV type, the association only remained significant for CNV deletions. Previous literature has been mixed as to the effect of CNV type on CNV burden. Research in severe developmental disorders found deletions to be twice as common as duplications[45], a study investigating a range of NDD phenotypes found no bias towards deletions or duplications[140], whereas a study of ID with and without schizophrenia only identified a significant difference between the groups for large CNV duplications[141]. It is unclear why these inconsistencies between studies exist. All of these studies have a slightly different ascertainment focus for their case samples, so it may be that the relationship with CNV type differs with the NDD under investigation.

For the size analyses the total kb of CNVs per person was found to be larger for ID+ cases versus controls, however this does not remain significant when filtering by

CNVs >400kb. This is surprising given that individuals with ID+ were already found to have a significantly greater number and proportion of CNVs >400kb as compared to controls. Potentially this is due to that fact that CNV >400kb are rarer and there are fewer CNVs contributing to the total kb metric (75 CNVs in cases and 140 CNVs in controls). There are only significant differences between the average kb in ID+ cases versus controls for both deletions and duplications at the >100kb cut off. This is the weakest association in the study and it no longer remains significant when cases are removed from the analysis to control for the effects of ancestry. As evidenced by Figure 6-3 this results is likely driven by a small number of individuals with CNVs with large average kbs. Therefore, there does not appear to be major differences in the average CNV size between ID+ cases and controls.

Surprisingly, no significant differences in CNV burden were identified between schizophrenia cases and controls. The rate, proportion, total kb, and average kb remained very similar between cases and controls at all size cut offs and for both deletions and duplications. A limitation of the analysis methodology was that the gene content of the CNV was not considered. The number of genes affected by CNVs has previously found to be the strongest signal of enrichment for schizophrenia burden analyses[59]. One explanation for the differences between this analysis and other published literature is that the study was underpowered to detect an effect, as the main schizophrenia CNV burden analyses to date have comprised much larger sample sizes[59,204]. Another factor may be that the schizophrenia samples recruited via the DPIM study are likely to be deselected for ID. Whilst having ID wasn't an exclusion criteria for entry to the study, the study did require that the participant has capacity to consent to the research project. This may have reduced the number of participants at the lower end of the IQ spectrum. It is unclear what this result means for the interpretation of other schizophrenia CNV burden analyses to date. Recent work has shown that having a low IQ, or ID phenotype, does increase CNV burden in schizophrenia patients[142]. Also findings from a WES study found that more than half of the patients with a significantly enriched variant (in the SETD1A gene) also had learning difficulties. Data on the intellectual functioning of patients is not provided in the published large CNV burden studies[59,204]. This will be an important consideration for future research in the field.

Finally, significant differences in CNV burden were found in a case-case analysis of ID+ and schizophrenia only cases. The significant results were nearly identical to those found in the ID+ versus control analyses, except in this instance none of the average kb tests were significant. Given that there were no significant results when comparing schizophrenia cases against controls, the results indicate that differences in CNV burden are more pronounced as the severity of the NDD phenotype increases. In other words, there are greater differences in CNV burden between ID+ and schizophrenia compared with schizophrenia versus controls.

This study has several potential limitations. Whilst all samples were analysed on the same platform, the PsychChip, the samples were processed at two different sites – The Broad Institute of MIT and Harvard and the University of Bonn. It could be that there were some differences in sample processing that are influencing the findings, for example differences in in-house genotype calling quality control measures. However, these results should still be more comparable than samples analysed on different platforms, whereby different densities of probes affect the calling sensitivity. CNV calling from SNP arrays is known to be less reliable than array CGH methodologies, which could have led to some false positive and negative CNV calls. Typically, there is greater noise – or a higher number of false calls – for small CNVs, which was attempted to be control for by only considering CNVs >100kb. However, it is unclear whether 100kb was the best size cut off to use as the previously referenced literature was published in 2008 and the Psychchip is a high-density array, which could potentially mean a smaller size cut off would have been more appropriate.

A further limitation is that the cases were included in the filter to remove CNVs occurring at a >1% frequency in the dataset. It may be that some CNVs were removed which were absent in controls and only observed in cases. As the focus of this analysis was rare CNVs, and rare CNVs typically have much lower population frequencies, this is unlikely to have a major impact on the results. However, a more conservative common variant exclusion method would have been to use a publically available dataset with controls only to exclude common CNVs. The common CNV exclusion process was the same for participants from ethnic minority groups, whereas a more robust approach would have been to filter out common CNVs using ancestrally

matched control cases. Furthermore, the CNVs were only called using one calling algorithm – PennCNV. It has been shown that the number of CNVs per individual varies according to the algorithm used to call the CNVs[221]. Ideally multiple algorithms would be used and only CNVs that are called by more than one algorithm included in the analysis. Inheritance data was not available for this sample and so the role of *de novo* CNV burden could not be considered. Finally, the role of sex was not considered and this analysis was only undertaken on autosomal chromosomes due to the small group sizes for called X chromosome CNVs when dividing the groups by sex. Also it was not within the remit of this analysis to consider the gene content of CNVs, although this would be a natural next step for this work.

6.6 Conclusion

It has been hypothesized that psychiatric disorders lie on a neurodevelopmental continuum, of which ID is the most severe brain insult, followed by ASD and schizophrenia[96]. This CNV burden analysis compared, for the first time, ID plus co-morbid psychiatric disorder cases, schizophrenia patients and controls with CNVs called from the same SNP array platform. One could hypothesize that ID with co-morbid mental disorders lies at the extreme end of the neurodevelopmental continuum. Interestingly, ID+ appears to be more different from schizophrenia than schizophrenia does from controls, suggesting that CNV burden becomes more pronounced at the severest end of the neurodevelopmental continuum.

Chapter 7 Future directions

Historically, research has followed psychiatric nosology systems, ascertaining patients for the presence of distinct NDDs of interest. It soon became clear that, for the main part, psychiatric disorders do not follow simple Mendelian patterns of inheritance and that they are polygenic in nature, with a range of rare and common variants contributing to risk of developing the disorder. This extreme genetic heterogeneity, for example with over 700 ID related genes identified, presents challenges for uncovering molecular mechanisms of disease pathology. Furthermore, studies of pathogenic CNVs and SNVs have revealed that the same variant can be involved in risk for multiple NDDs. The full extent of this genetic pleiotropy is still under investigation, however it appears that there is a complex pattern of risk conferred by different loci. Cross-disorder genetic investigations, whereby groups ascertained for different NDDs are analysed together, are becoming common-place in psychiatric genetics. However, there remains a paucity of research in individuals with co-morbid NDD phenotypes.

7.1 Summary of research findings

This thesis has focused on the investigation of CNVs in individuals with co-morbid NDDs. In Chapter 2, findings from clinical practice were investigated, with a survey of child and adolescent and adult ID psychiatrists. The survey aimed to determine the extent to which genetic investigations are being utilised by psychiatrists working with patients with ID, and psychiatrists' opinions around genetic testing practices. A need for increased training was identified, given the challenges clinicians face in keeping up to date with the genomic advances relevant to ID psychiatry. Currently, psychiatrist's training curriculum fail to cover genetic disorders and there is a wide geographical variability in links and access to genetic testing services and genetic counselling. It is important to address these issues to ensure that the advances in genetic testing reach patients, this is particularly important given the health inequalities already faced by individuals with NDDs. It would be of interest to undertake further research on clinical management changes following genetic diagnoses, which were rarely reported in clinical practice, despite there being a large number of disorder guides with clinical management guidelines.

In Chapter 3 adults with idiopathic ID and co-morbid psychiatric disorders (N=202) were recruited from ID psychiatry services with the aim of determining the frequency, type, and associated phenotype of pathogenic CNVs. An 11% frequency of pathogenic CNVs was identified, an important finding for clinical practice – that over 1 in 10 ID psychiatry patients have an undiagnosed aetiological CNV. This argues for more routine consideration of CMA in ID psychiatry services. The majority of pathogenic CNVs were found at recurrent loci, which have already been described in the literature. A higher proportion of pathogenic CNV carriers were forensic in-patients, as compared to non-pathogenic CNV carriers (OR 4.1). It would be an interesting avenue of future research to undertake CMA in forensic populations, however no cohorts could be identified to undertake this analysis within the remit of this thesis. No other significant differences in the frequencies of psychiatric symptomatology were identified between pathogenic and non-pathogenic CNV groups. This may mean that presence of a particular psychiatric disorder might not be a useful marker for prioritising patients for CMA testing. Again, further research on the phenotypic profile of pathogenic and non-pathogenic CNV carriers is needed to further investigate this.

In Chapter 4, the findings of Chapter 3 were replicated in a larger multi-population patient cohort (N=599). A similar frequency of pathogenic CNVs was identified in these patients with ID and co-morbid pscyhiatric disorders (13%). To further investigate the genetic architecture of the pathogenic CNVs the rate of established NDD risk CNVs, as compared to single-disorder cohorts, was analysed. Again, pathogenic CNVs were frequently found at recurrent CNV sites, with 70% of pathogenic CNVs at sites already known to be involved in NDD risk. Taking this sample and comparing it to large healthy control, schizophrenia, and ID/ASD cohorts from previous studies unveiled the highest sample rate of NDD CNVs. This large multi-centre cohort also enabled the detection of overlapping likely pathogenic CNVs, providing valuable contributions to the literature on the pathogenicity of the *CNTN* and *CHD* gene families and 16p12.1 deletions. Future research could focus on adding further samples to this existing dataset, which will likely lead to the emergence of other multiply occurring rare pathogenic CNVs.

In Chapter 5, deep phenotyping was undertaken in patients with CNVs at the 2q13 locus (N=25). Whilst developmental and psychiatric phenotypes had been described in previous published studies, these phenotypes had not been systematically investigated. A common phenotype of DD, mild intellectual impairment and a high frequency of psychiatric and behavioural disorders was identified. There appeared to be a global risk to developing a psychiatric disorder, with a wide-range of symptoms and diagnoses. However, the frequency of clinical ADHD diagnoses was particularly striking – with 44% of participants being diagnosed. Delineation of the phenoypes associated with rare CNVs is important to enable syndrome-specific phenotype information, which may guide therapeutic interventions. Furthermore, the identification of a high rate of ADHD diagnoses may guide functional follow up in this region in investigations of the neuropathological mechanisms of ADHD. This multi-faceted research approach – utilising RGCs, a patient support group and a rare CNV database, was particularly successful and enabled the largest case series of 2q13 patients reported to date. A similar approach could be undertaken in future research to facilitate phenotypic delineation of rare CNVs. Another further avenue of research would be to undertake WES or WGS in the 2q13 participants to test the second-hit model as a theory for the incomplete penetrance of this CNV.

Finally in Chapter 6, a burden analysis was undertaken to compare the individuals with co-morbid NDDs (ID+) to controls and schizophrenia only patients. Significant differences in rates of CNVs per individual, the proportion of cases with at least one CNV event and the total kb of CNVs spanned per person were found between the ID+ and controls. When filtering by CNV type and size this only remained true for CNV deletions at the larger size cut offs. No differences were found between schizophrenia cases and controls, although it may be that this analysis was underpowered. The ID+ cases showed significant differences in CNV burden, as compared to schizophrenia patients, indicating that the CNV burden is particularly high in the co-morbid ID+ cohort. There were several limitations of this analysis and further analysis is required to investigate this finding, in particular re-calling the CNVs using a different algorithm, investigating the impact of sex, and undertaking a gene-set analysis would be appropriate extensions of this analysis.

7.2 Insights from participant recruitment

Being personally involved in participant recruitment has afforded me unique insights into the challenges of undertaking psychiatric research. One observation, which recurred throughout my research visits, is that family members are hesitant to divulge mental health information. On many occasions a family member would answer 'no' when asked whether themselves or their child had a diagnosed mental health problem. However, it would emerge through later questioning in the interview that this was incorrect. For example, one family member said their child did not have a mental health problem, however following a discussion about medication history it emerged that the child was taking anti-depressants and did have a diagnosis of depression. It was far less frequent that a family member would misanswer questions relating to physical health. It is unclear why this phenomenon occurred, it may have arisen from family members' misunderstanding of the term 'mental health'. It could also be linked to that fact that mental health problems are more stigmatised than physical health problems, and thus the family finds it difficult to discuss or they are reluctant to share the information.

Another interesting observation was that during collection of family history information for patient pedigrees, family members would refer to their relatives using terms such as 'a bit odd' or 'a bit of a loner'. These relatives were never investigated for psychiatric illness, although family members often felt that they would meet diagnostic criteria. From a historical context awareness of, and screening for, mental health problems is much improved in current healthcare systems. There is an ongoing debate about whether mental health problems are increasing in prevalence or whether the increase is a consequence of improvements in psychiatric screening. For example, it is thought that raising awareness and screening for ASD may explain the rapidly increasing prevalence rates of the disorder. Research utilising US special education enrolment data has shown that as diagnoses of ASD have risen those of ID have decreased, suggesting that misdiagnosis of ASD as ID in the past may in part account for the increasing prevalence of the disorder[222]. This highlights the difficulties of ascertaining an accurate psychiatric family history from research participants. Ideally all family members would be re-screened using modern assessment procedures –

however the requirement for trained clinicians and lengthy screening processes are a barrier to implementing this in research practice.

Another element of the recruitment process required review of available medical records. This comprised records from a variety of practitioners involved in the care of the participant, including: psychologists, speech and language therapists, occupational therapists, psychiatrists, and clinical geneticists. Surprisingly, patient information was often inconsistent between reports. In some instances the wrong genetic or medical diagnosis was included on practitioner reports, or the reports failed to mention important information relating to the participant's diagnosis. This issue is particularly pertinent for individuals with ID and co-morbid mental illness, as there are often multiple medical professionals involved in the patient's care who have not necessarily communicated directly about the patient. This raises the important issue of whether medical records are a reliable source of information about the medical and psychiatric history of these individuals. Due to the study design, direct liaison with an informant and the treating psychiatrist made it feasible to resolve any discrepancies during the recruitment process. However, study designs which only utilise medical records are likely to be more error prone.

7.3 NDD risk CNVs and clinical heterogeneity

CNVs associated with risk for developing NDDs have been shown to have variable phenotypic outcomes. Taking the 2q13 CNV as an example, a wide range of medical, dysmorphic, behavioural and psychiatric features were identified in participants with 2q13 CNVs. However, nearly half of the individuals (44%) had an ADHD diagnosis. This work provides support for the asymmetric risk model, whereby rare CNVs confer both shared risk and distinct aetiology[108]. In this instance, shared risk for general psychiatric pathology, with a propensity for ADHD pathology.

It is important to consider the potential modifying factors that influence outcomes for these CNV carriers. Firstly, only one type of genetic variation – CNVs – was taken into account and many other forms of genetic variants affect phenotypic outcomes. For example, it is typical for every individual to have around 1-2 *de novo* exonic SNVs[223], and the rate is likely higher in severe NDD cohorts[51]. Furthermore, rare inherited

variants, identifiable by WES and WGS, also modify phenotypic outcomes. Research investigating the contribution of both rare and common variants to schizophrenia pathology has provided support for a polygenic threshold model, whereby a multitude of common and rare risk variants are involved in the disease phenotype[74]. Common variants were also not considered in this thesis, and it is likely that the burden of common risk variants acts as a modifier of phenotypic outcomes.

Currently, separate tests are required to reliably detect each type of genetic variant, making it challenging to consider all the contributing genetic factors. As WGS becomes cheaper and the methods for detection of other variants (such as CNVs) from sequencing data improve, it is likely that genetic testing in the future will be more comprehensive. Another challenge is that little is known about how modifier genes or gene-gene interactions modify phenotypic outcomes. Large well-phenotyped samples with comprehensive genetic data will be required to delineate these relationships. It is of interest to identify variants operating in the same molecular pathways, which likely exacerbate disease phenotypes, and compensatory or protective variants. For the 2q13 CNV carriers, 25% of participants had a family history of ID and mental health problems on the other side of the family from which the variant was transmitted. It may be that assortative mating and/or multiple hits in NDD gene pathways are contributing to the diverse phenotypic outcomes observed. A further challenge will be ascertaining genetic history for available family members, and teasing out the contribution of inherited variants.

It is not only genetic factors which contribute to phenotypic outcomes. Recent estimates of the concordance rates between monozygotic twins for schizophrenia are around 33%[57], highlighting that outcomes are often variable for genetically identical individuals. Non-genetic factors, including environmental factors – such as obstetric complications, drug abuse, and migration – have been shown to contribute to phenotypic outcomes in schizophrenia[224]. Stochastic, or chance factors, also play a role, given that the brain is a highly complex organ and neuronal development is error prone. Similar to the novel *de novo* mutations that arise by chance at the genetic level, there are comparable chance errors in the cellular processes involved in the developing brain.

7.4 Complexities with phenotyping and co-morbid phenotypes

One of the primary differences between the research presented in this thesis and other large cohort studies in psychiatric genetics is the degree of phenotypic information collected. GWAS analyses typically class disease as a categorical phenotype. For example, the presence or absence of schizophrenia defines the case and control groups respectively. Whereas, the term deep phenotyping is used for studies that comprise measurements on a number of clinical, behavioural, and neuropsychiatric assessments. The challenge for more comprehensive phenotyping is that there is no standardised method for deep phenotyping in psychiatry, meaning there is methodological inconsistency across the measures used. Deep phenotyping is also very time consuming, resulting in a limited number of subjects and often requiring collaboration with other researchers[225]. As previously discussed, family members do not necessarily answer research questions correctly and the information available from medical notes is not necessarily accurate. This poses questions about the validity of shallow phenotyping techniques, whereby limited information is collected.

A complexity with interpreting the findings from this thesis, with a more comprehensive phenotyping protocol, is that it is unclear how the research participants relate to those presented in other studies. As discussed in Chapter 4, the large paediatric severe developmental disorder cohorts are poorly phenotyped and it is unclear exactly how many had DD/ID and/or psychiatric disorders. Also the paediatric nature of the cohort means that the age of onset for many psychiatric disorders is yet to be reached. Furthermore, diagnosis of psychiatric illness in ID is challenging, and it is not routine practice for a standardised set of psychiatric assessments to be undertaken. Thus, diagnostic practices are variable, and it may be that individuals categorised as having ID only also have undetected mental health problems. This is particularly the case at the severe end of the ID spectrum, where non-verbal individuals are unable to self-report psychiatric symptoms.

The work in this thesis shows that patients with ID plus psychiatric co-morbidities have a greater frequency of CNVs at NDD risk loci and a higher burden of rare CNVs, as compared to other patient groups. The neurodevelopmental continuum model has already proposed that ID is the most severe early brain insult which has a higher burden

of CNVs and deleterious mutations[96,226]. It may be that ID plus co-morbid neurodevelopmental or neuropsychiatric phenotypes is located on the most severe tail of the neurodevelopmental continuum model, being a more severe phenotype than ID alone. However, the aforementioned limitations with a lack of suitable reference populations precludes confirmation of this hypothesis. Ideally, a comparable reference population of patients with ID and the absence of co-morbid mental illness would be required to fully test the difference between groups. However, this is a difficult population to ascertain.

The challenges in ascertaining an ID population with the absence of co-morbid psychiatric phenotypes is multifold. One could only study this in adulthood, after the age of onset for most psychiatric disorders. As previously discussed there is a historical context whereby many adults with ID will have not undergone psychiatric screening. It may be that there has been diagnostic overshadowing – whereby the ID is thought to be the cause of the phenotype, so appropriate screening for other disorders wasn't undertaken. Also previous versions of the psychiatric diagnostic manuals precluded some co-morbid diagnoses, for example only in DSM5 was it possible to diagnose ADHD and ASD in the same individual[226]. In order to rule out psychiatric co-morbidities, the most robust approach would be to re-screen everyone with ID using modern psychiatric screening schedules – although this would be complex and resource intensive.

Several lines of evidence suggest that diagostic categories, used in clinical psychiatric practice, map poorly onto the underlying biology. From a phenotypic perspective, there is also a wide variability within diagnostic categories and symtom overlap between diagnostic categories[226]. There is also a movement towards recognising the complex continuous nature of NDDs, and how their different patterns of impairments lie on a continuum marked by the severity of the brain insult[1]. However, it is difficult to determine where an individual with mild ID and schizophrenia lies in relation to an indiivudal with severe ID. Perhaps a better measure of impairment, rather than crude diagnostic status alone, would be to undertake a functional assessment of the degree of impairment to daily living skills. This might take into account some of the

complexities of studying co-morbid diagnoses, but would require an overhall of current research practices.

7.5 Further discussion of ascertainment bias

As discussed in previous chapters, one of the limitations of the findings of this thesis is that there is likely to be an ascertainment bias arising from the methodology used to recruit participants. Routine genetic testing for CNVs in psychiatric practice is currently only taking place in the context of paediatric DD/ID, MCA and ASD[25]. The next obvious group for routine genetic testing is adults with co-morbid psychiatric phenotypes, the target population for this body of research. This research aimed to sample participants directly from clinical services, with ID psychiatrist being the main point of contact for recruitment. It is hard to determine the extent to which ascertainment bias affected the results, as the bias could have operated in both directions – both in terms of psychiatrists selecting patients who they suspect have underlying genetic disorders, and de-selecting participants with more severe phenotypes who are harder to recruit.

Individuals with ID and co-morbid psychiatric disorders are a particularly hard-to-reach population. The recruitment strategy employed in this thesis was very successful, recruiting nearly 250 individuals in 18 months. It is challenging to think of alternative research methods to reduce ascertainment bias whilst maintaining recruitment levels. Indeed, all of the severe developmental disorders studies will have an element of recruitment bias, as many have captured patients referred for clinical genetic testing and a certain threshold of impairment must be passed to reach referral for clinical testing[44,45]. Ideally, a large population based longitudinal study, with available genetic data, would be the most representative sample, as it would enable detection of neurodevelopmental CNV carriers across the whole phenotypic spectrum – from healthy controls to patients with co-morbid phenotypes.

7.6 Clinical implications and utility

Several participants in this study received a genetic diagnosis, of previously undiagnosed pathogenic CNVs that were related to the individual's ID and/or

psychiatric condition. The implications of receiving a genetic diagnosis for patients and their families are widespread. For some families receiving a genetic diagnosis can help to alleviate feelings of self-blame, particularly for mothers who may falsely believe that they did something wrong during pregnancy. One study investigated self-reported quality of life scores in mothers whose child had received a genetic diagnosis from array CGH. They found that having an aetiological diagnosis for the child's DD/ID and/or MCA improved maternal quality of life. It is described as an 'emotional relief' to have a name for the disorder and understand the cause of the child's diability[137].

Confirmation of an aetiological diagnosis can also aid clinical symptom-based diagnosis. An aetiological diagnosis can facilitate screening for associated medical and psychiatric disorders, and in some instances provide information on likely responsiveness to treatments. Information leaflets with clinical guidelines are available for an increasing number of rare genetic syndromes[134,135]. More clear cut information about a syndrome can be helpful, particularly if it helps the patient understand phenotypic presentations. For example, a patient who has a syndrome which has a common phenotype of self-injurious behaviour[136].

A genetic diagnosis can also be empowering for patients and families, enabling them to access disorder-specific or general support groups, such as the Unique chromosomal disorder support group. For school age children a genetic diagnosis can be beneficial for a statement of special educational support, or access to disorder-specific support services. Furthermore, a genetic diagnosis may have broader implications for the family and indicate genetic counselling, genetic risk to subsequent offspring, and cascade testing via Regional Clinical Genetics Services. Genetic counselling in schizophrenia patients, who had been fed back aetiological CNV results, was found to improve understanding of the disorder, and significantly reduce internalized stigma and self-blame[227]. Recurrence risk information can be hugely beneficial to family planning, for example if a rare variant is found to be *de novo* in origin then there the recurrence risk for siblings is equal to the population prevalence of the disorder.

The full range of CNVs implicated in NDD risk are still being delineated. It is likely that many of the highest penetrance variants have already been identified, given their

stronger associations with specific phenotypic outcomes. However, more neurosusceptability CNVs (which are present at higher frequencies in NDD patients, but also present at low frequencies in healthy controls) may yet be identified. Neurosusceptability CNVs, at the 2q13 locus, were described in detail in this thesis. Further delineation of the phenotype, and it's association with childhood onset psychiatric disorders, could have important clinical implications for screening for psychiatric disorders. However, there were also a diverse range of medical and dysmorphology phenotypes observed. It is clear that the relationship between neurosusceptability CNVs and phenotypic outcomes will be even more challenging to delineate, and will pose greater challenging for genetic counselling of patients and their families. Description of the phenotype of participants with VOUS likely pathogenic CNVs in Chapter 4 may also help to guide the clinical interpretation of these CNVs in the future, given that reviewing published literature on the CNVs is an integral part of the categorisation process[14]. As evidenced in Chapter 2, there are barriers to translating the advances in genetic understanding of NDDs into clinical practice. Psychiatrists lack confidence in genetic testing practices and require further training and better links with genetic services. It is unclear how these findings link to other areas of medicine where genomic testing is on the rise. For example, many advances have been seen in genetic testing for various cancers and it would be interesting to research clinician's views in this branch of medicine. One of the major challenges in psychiatric research is that it still relies on self-reported symptom information to diagnose patients, whereas parallel tests in oncology have biological markers from blood tests and tumor biopsies.

In the broad field of psychiatry, ID psychiatry is likely to have the most utility for genetic testing in the clinic given the high rate of pathogenic variants. However, as it currently stands the onus is on psychiatrists to choose which patients, if any, to refer for clinical genetic testing. Implementing a routine genetic screening programme, at least for pathogenic CNVs, in ID psychiatry would enable comparable – or greater – success rates as in paediatric DD/ID. However, this would need to be supported with training for clinicians in the genetic testing process and the additional demand for genetic counselling would need to be supplemented by increased service provision in Regional Genetics Services. If the current scope of routine testing is expanded to

encompass this patient group this could also facilitate the discovery of further risk variants for NDDs.

There are currently limited therapeutic success stories that have arisen from discovery of underlying NDD genetic aetiologies. Drug discovery in psychiatry in general has stagnated, with no major developments in the previous 40 years and many pharmaceutical companies scaling down psychiatric drug research[226]. In cancer genetics the pathogenic effects of genetic variants arise at the cellular level, and drugs can be targeted to biochemical pathways. In psychiatry there is a complex myriad of genetic interactions, which has cascading effects on complex neural circuits and pathways. One study has reported a pharmacologically guided treatment in a patient with 15q13.3 deletion syndrome who had an aggressive phenotype. The CNV deletion was found to encompass the Cholinergic Receptor Nicotinic Alpha 7 Subunit (CHRNA7) gene. Administration of galatamine, a modulator of nicotinic cholinergic receptor function, led to a decline in the frequency and intensity of rage outbursts[228]. Further advances in pharmacogenomics will be dependent upon better characterisation of the region specific molecular consequences of pathogenic CNV.

7.7 Concluding remarks

One of the greatest challenges to the future of psychiatric research is how to reconcile these findings within the current system of psychiatric nosology and determine the type of patient stratification that is most appropriate for genetic analyses. More studies of rare pathogenic CNVs, with a focus on functional analysis of candidate genes and pathways, will be required to delineate the nature of asymmetric loci-specific risk. It will be imperative that this research cuts across cohort groups, traversing from healthy controls to individuals with co-morbid NDDs. Concurrently, phenotyping methods need to be developed – currently phenotyping is basic and the lack of standardised methods preclude comparisons between studies. Indeed, investigations of 'healthy controls' with pathogenic CNVs have revealed that many do show cognitive impairments and psychiatric symptomatology. Thus, better psychiatric and cognitive phenotyping, to encompass sub-clinical aetiology, is important for future genetic studies. Additionally, consideration of familial genetic background and the interaction

between genetic variants is likely to uncover protective mechanisms for the pentrance of pathogenic CNVs and expressivity of disease phenotypes.

Bibliography

1. Moreno-De-Luca, A. *et al.* Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol.* **12,** 406–14 (2013).
2. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Publ. Gr.* **17,** 9–18 (2015).
3. Nuechterlein, K. H., Ventura, J., Subotnik, K. L. & Bartzokis, G. The early longitudinal course of cognitive deficits in schizophrenia. *J. Clin. Psychiatry* **75,** 25–9 (2014).
4. Tjio, J. H. & Levan, A. The Chromosome Number of Man. *Hereditas* **42,** 1–6 (1956).
5. Verri, A., Cremante, A., Clerici, F., Destefani, V. & Radicioni, A. Klinefelter's syndrome and psychoneurologic function. *Mol. Hum. Reprod.* **16,** 425–433 (2010).
6. Blau, N., Van Spronsen, F. J. & Levy, H. L. Phenylketonuria. *Lancet* **376,** 1417–1427 (2010).
7. Scriver, C. R. The PAH gene, phenylketonuria, and a paradigm shift. *Hum. Mutat.* **28,** 831–45 (2007).
8. Cederbaum, S. Phenylketonuria: an update. *Curr. Opin. Pediatr.* **14,** 702–6 (2002).
9. Ledbetter, D. H. *et al.* Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. *N. Engl. J. Med.* **304,** 325–9 (1981).
10. Cassidy, S. B. & Driscoll, D. J. Prader–Willi syndrome. *Eur. J. Hum. Genet.* **17,** 3–13 (2009).
11. Pelc, K., Cheron, G. & Dan, B. Behavior and neuropsychiatry manifestations in Angelman syndrome. *Neuropsychiatr. Dis. Treat.* **4,** 577–584 (2008).
12. Van Buggenhout, G. & Fryns, J.-P. Angelman syndrome (AS, MIM 105830). *Eur. J. Hum. Genet.* **17,** 1367–1373 (2009).
13. Lehner, T., Senthil, G. & Addington, A. M. Convergence of Advances in Genomics, Team Science, and Repositories as Drivers of Progress in Psychiatric Genomics. *Biol. Psychiatry* **77,** 6–14 (2014).
14. Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F. & South, S. T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* **13,** 680–685 (2011).
15. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36,** 949–951 (2004).
16. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–8 (2004).
17. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
18. Grayton, H. M., Fernandes, C., Rujescu, D. & Collier, D. A. Copy number variations in neurodevelopmental disorders. *Prog. Neurobiol.* **99,** 81–91 (2012).
19. Hiroi, N. *et al.* Copy number variation at 22q11.2: from rare variants to common mechanisms of developmental neuropsychiatric disorders. *Mol. Psychiatry* **18,** 1153–65 (2013).
20. Iyer, J. & Girirajan, S. Gene discovery and functional assessment of rare copy-

number variants in neurodevelopmental disorders. *Brief. Funct. Genomics* 1–14 (2015).
21. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148,** 1223–41 (2012).
22. Gu, W., Zhang, F. & Lupski, J. R. Mechanisms for human genomic rearrangements. *Pathogenetics* **1,** 4 (2008).
23. Carter, N. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* **39,** S16–S21 (2007).
24. LaFramboise, T. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Res.* **37,** 4181–4193 (2009).
25. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86,** 749–64 (2010).
26. Stessman, H. a., Bernier, R. & Eichler, E. E. A genotype-first approach to defining the subtypes of a complex disease. *Cell* **156,** 872–877 (2014).
27. Collier, D. *et al.* Advances in the genetics of schizophrenia: will high-risk copy number variants be useful in clinical genetics or diagnostics? *F1000 Med. Rep.* **6,** 4–9 (2009).
28. Baker, K., Costain, G., Fung, W. L. A. & Bassett, A. S. Chromosomal microarray analysis—a routine clinical genetic test for patients with schizophrenia. *The Lancet Psychiatry* **1,** 329–331 (2014).
29. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6,** 1–11 (2010).
30. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42,** D986-92 (2014).
31. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385,** 1305–1314 (2015).
32. Vorstman, J. A. S. *et al.* Autism genetics: opportunities and challenges for clinical translation. *Nat. Rev. Genet.* **18,** 362–376 (2017).
33. Karmiloff-Smith, A. *et al.* The importance of understanding individual differences in Down syndrome. *F1000Research* **5,** 389 (2016).
34. O'Donovan, M. C. & Owen, M. J. The implications of the shared genetics of psychiatric disorders. *Nat. Med.* **22,** (2016).
35. De Wolf, V., Brison, N., Devriendt, K. & Peeters, H. Genetic counseling for susceptibility loci and neurodevelopmental disorders: The del15q11.2 as an example. *Am. J. Med. Genet. Part A* **161,** 2846–2854 (2013).
36. Palmer, E. *et al.* Changing interpretation of chromosomal microarray over time in a community cohort with intellectual disability. *Am. J. Med. Genet. A* **164A,** 377–85 (2014).
37. Miclea, D., Peca, L., Cuzmici, Z. & Pop, I. V. Genetic testing in patients with global developmental delay / intellectual disabilities. A review. *Clujul Med.* **88,** 288–92 (2015).
38. Salvador-Carulla, L. *et al.* Intellectual developmental disorders: towards a new name, definition and framework for 'mental retardation/intellectual disability' in ICD-11. *World Psychiatry* **10,** 175–80 (2011).

39. World Health Organisation. *ICD-10 guide for mental retardation*. (Division of mental health and prevention of substance abuse, 1996).
40. Hehir-Kwa, J. Y., Pfundt, R., Veltman, J. a & de Leeuw, N. Pathogenic or not? Assessing the clinical relevance of copy number variants. *Clin. Genet.* **84,** 415–21 (2013).
41. Cooper, S.-A. *et al.* Psychosis and adults with intellectual disabilities. Prevalence, incidence, and related factors. *Soc. Psychiatry Psychiatr. Epidemiol.* **42,** 530–536 (2007).
42. Sheehan, R. *et al.* Mental illness, challenging behaviour, and psychotropic drug prescribing in people with intellectual disability: UK population based cohort study. *BMJ* **351,** h4326 (2015).
43. Reichenberg, A. *et al.* Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proc Natl Acad Sci U S A* **113,** 1098–1103 (2016).
44. Kaminsky, E. B. *et al.* An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet. Med.* **13,** 777–84 (2011).
45. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43,** 838–46 (2011).
46. Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46,** (2014).
47. Koolen, D. a *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* **44,** 639–641 (2012).
48. Vissers, L. E. *et al.* A de novo paradigm for mental retardation. *Nat Genet* **42,** 1109–1112 (2010).
49. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511,** 344–347 (2014).
50. Fitzgerald, T. W. *et al.* Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519,** 223–228 (2014).
51. McRae, J. F. *et al.* Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542,** 433–438 (2017).
52. Owen, M. J., Sawa, A. & Mortensen, P. B. Schizophrenia. *Lancet* **388,** 86–97 (2016).
53. Howes, O. D. & Murray, R. M. Schizophrenia: An integrated sociodevelopmental-cognitive model. *Lancet* **383,** 1677–1687 (2014).
54. Bora, E. & Murray, R. M. Meta-analysis of cognitive deficits in ultra-high risk to psychosis and first-episode psychosis: Do the cognitive deficits progress over, or after, the onset of psychosis? *Schizophr. Bull.* **40,** 744–755 (2014).
55. Cardno, A. *et al.* Heritability Estimates for Psychotic Disorders. *Arch. Gen. Psychiatry* **56,** 162–168 (1999).
56. Cannon, T. D., Kaprio, J., Lönnqvist, J., Huttunen, M. & Koskenvuo, M. The genetic epidemiology of schizophrenia in a Finnish twin cohort: A population-based modeling study. *Arch. Gen. Psychiatry* **55,** 67–74 (1998).
57. Hilker, R. *et al.* Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol. Psychiatry* 1–7 (2017).
58. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13,** 537–51 (2012).
59. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia

from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49,** 27–35 (2017).
60. Kirov, G. *et al.* Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.* **17,** 458–465 (2008).
61. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320,** 539–43 (2008).
62. Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* **204,** 108–114 (2014).
63. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–84 (2014).
64. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506,** 185–190 (2014).
65. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19,** 1433–1441 (2016).
66. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nat. Publ. Gr.* **536,** 285–291 (2016).
67. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* 1–10 (2017). doi:10.1038/ng.3903
68. Flint, J. & Munafò, M. Schizophrenia: Genesis of a complex disease. *Nature* **511,** 412–413 (2014).
69. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
70. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530,** 177–183 (2016).
71. Wray, N. R. *et al.* Research Review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry Allied Discip.* **55,** 1068–1087 (2014).
72. Vassos, E. *et al.* An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biol. Psychiatry* **81,** 470–477 (2017).
73. So, H. C., Sham, P. C. & Valencia, A. Exploring the predictive power of polygenic scores derived from genome-wide association studies: A study of 10 complex traits. *Bioinformatics* **33,** 886–892 (2017).
74. Tansey, K. E. *et al.* Common alleles contribute to schizophrenia in CNV carriers. *Mol. Psychiatry* **21,** 1085–9 (2016).
75. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459,** 569–573 (2009).
76. Skuse, D. H. Rethinking the nature of genetic vulnerability to autistic spectrum disorders. *Trends Genet.* **23,** 387–395 (2007).
77. Levy, F. Theories of autism. *Aust. N. Z. J. Psychiatry* **41,** 859–68 (2007).
78. Pavăl, D. A Dopamine Hypothesis of Autism Spectrum Disorder. *Dev. Neurosci.* 355–360 (2017).
79. Takarae, Y. & Sweeney, J. Neural Hyperexcitability in Autism Spectrum Disorders. *Brain Sci.* **7,** (2017).
80. Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M. & Law, P. Sibling recurrance and the genetic epidemiology of autism. *Am. J. Psychiatry* **167,** 1349–1356 (2010).
81. Rosenberg, R. E. *et al.* Characteristics and concordance of autism spectrum disroders aming 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163,** 907–914

(2009).
82. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–75 (2008).
83. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316,** 445–9 (2007).
84. Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70,** 863–85 (2011).
85. Krumm, N. *et al.* Transmission disequilibrium of small CNVs in simplex autism. *Am. J. Hum. Genet.* **93,** 595–606 (2013).
86. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–72 (2010).
87. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87,** 1215–33 (2015).
88. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485,** 242–245 (2012).
89. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515,** 209–215 (2014).
90. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515,** 216–221 (2014).
91. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48,** 552–5 (2016).
92. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46,** 881–885 (2014).
93. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8,** 21 (2017).
94. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49,** 978–985 (2017).
95. Morgan, V. A. *et al.* Intellectual disability and other neuropsychiatric outcomes in high-risk children of mothers with schizophrenia, bipolar disorder and unipolar major depression. *Br. J. Psychiatry* **200,** 282–289 (2012).
96. Doherty, J. L. & Owen, M. J. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med.* **6,** 29 (2014).
97. Rees, E. *et al.* Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. *JAMA Psychiatry* **13,** 1–7 (2016).
98. Shprintzen, R. J. Velo-Cardio-Facial Syndrome: 30 Years of Study. *Dev. Disabilies Res. Rev.* **14,** 3–10 (2008).
99. Monks, S. *et al.* Further evidence for high rates of schizophrenia in 22q11.2 deletion syndrome. *Schizophr. Res.* **153,** 231–236 (2014).
100. Hooper, S. R. *et al.* A longitudinal examination of the psychoeducational, neurocognitive, and psychiatric functioning in children with 22q11.2 deletion syndrome. *Res. Dev. Disabil.* **34,** 1758–69 (2013).
101. Ou, Z. *et al.* Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genet. Med.* **10,** 267–77 (2008).

102. Rees, E. *et al.* Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol. Psychiatry* **19,** 37–40 (2014).
103. Zufferey, F. *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J. Med. Genet.* **49,** 660–668 (2012).
104. Moreno-De-Luca, A. *et al.* The role of parental cognitive, behavioral, and motor profiles in clinical variability in individuals with chromosome 16p11.2 deletions. *JAMA psychiatry* **72,** 119–26 (2015).
105. D'Angelo, D. *et al.* Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA psychiatry* **73,** 20–30 (2016).
106. Giaroli, G., Bass, N., Strydom, A., Rantell, K. & McQuillin, A. Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: A systematic review of literature and meta-analysis. *Schizophr. Res.* **159,** 340–6 (2014).
107. Kirov, G. *et al.* The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. *Biol. Psychiatry* **75,** 378–385 (2014).
108. Moreno-De-Luca, D., Moreno-De-Luca, A., Cubells, J. F. & Sanders, S. J. Cross-Disorder Comparison of Four Neuropsychiatric CNV Loci. *Curr. Genet. Med. Rep.* **2,** 151–161 (2014).
109. McCarthy, S. E. *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19,** 652–658 (2014).
110. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* **19,** 571–577 (2016).
111. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515,** 209–215 (2014).
112. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14,** 483–495 (2013).
113. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–52 (2009).
114. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45,** 984–94 (2013).
115. Smoller, J. W. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).
116. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505,** 361–6 (2014).
117. Männik, K. *et al.* Copy Number Variations and Cognitive Phenotypes in Unselected Populations. *Jama* **313,** 2044 (2015).
118. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One* **11,** e0154222 (2016).
119. Kendall, K. M. *et al.* Cognitive performance among carriers of pathogenic copy number variants: Analysis of 152,000 UK Biobank subjects. *Biol. Psychiatry* 0–27 (2016).
120. Moreno-De-Luca, D. *et al.* Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol. Psychiatry* **18,** 1090–1095 (2012).

121. Gonzalez-Mantilla, A. J., Moreno-De-Luca, A., Ledbetter, D. H. & Martin, C. L. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA psychiatry* **73,** 275–83 (2016).
122. de Villiers, J. & Porteous, M. Genetic testing of adults with intellectual disability. *Psychiatrist* **36,** 409–413 (2012).
123. Aas, M. *et al.* Summaries of plenary, symposia, and oral sessions at the XXII World Congress of Psychiatric Genetics, Copenhagen, Denmark, 12-16 October 2014. *Psychiatr. Genet.* **26,** 1–47 (2016).
124. Rush, K. S., Bowman, L. G., Eidman, S. L., Toole, L. M. & Mortenson, B. P. Assessing psychopathology in individuals with developmental disabilities. *Behav. Modif.* **28,** 621–637 (2004).
125. Emerson, E., Baines, S., Allerton, L. & Welch, V. Health Inequalities & People with Learning Disabilities in the UK. (2012).
126. Eng, B., Addison, P. & Ring, H. A guide to intellectual disability psychiatry assessments in the community. *Adv. Psychiatr. Treat.* **19,** 429–436 (2013).
127. Redley, M., Pannebakker, M. & Holland, A. Improving the Health and Well-Being of Adults With Conditions of a Genetic Origin: Views from Professionals, Syndrome Support Groups and Parents. *J. Appl. Res. Intellect. Disabil.* **21,** (2016).
128. PHG foundation. *Genetics and mainstream medicine*. *PHG Foundation* (2011).
129. Wordle. Available at: http://www.wordle.net/.
130. HMGovernment. Concordat and Moratorium on Genetics and Insurance. 1–11 (2014).
131. de Villiers, J. & Porteous, M. Genetic testing of adults with intellectual disability. *Psychiatrist* **36,** 409–413 (2012).
132. Royal College of Psychiatrists Curricula. Available at: http://www.rcpsych.ac.uk/traininpsychiatry/corespecialtytraining/curricula.aspx.
133. Neuroscience Project. Available at: http://www.rcpsych.ac.uk/traininpsychiatry/corespecialtytraining/neuroscienceproject.aspx.
134. Orphanet. No Title. Available at: http://www.orpha.net/consor/cgi-bin/Disease_ProEncyclo.php?lng=EN.
135. Unique The Rare Chromosome Disorder Support Group. Available at: http://www.rarechromo.org/html/DisorderGuides.asp.
136. Waite, J. *et al.* The importance of understanding the behavioural phenotypes of genetic syndromes associated with intellectual disability. *Paediatr. Child Heal. (United Kingdom)* **24,** 468–472 (2014).
137. Lingen, M. *et al.* Obtaining a genetic diagnosis in a child with disability: impact on parental quality of life. *Clin. Genet.* **89,** 258–66 (2016).
138. Lowther, C., Costain, G., Baribeau, D. A. & Bassett, A. S. Genomic Disorders in Psychiatry—What Does the Clinician Need to Know? *Curr. Psychiatry Rep.* **19,** 82 (2017).
139. Buszewicz, M. *et al.* Assessment of an incentivised scheme to provide annual health checks in primary care for adults with intellectual disability: a longitudinal cohort study. *Lancet Psychiatry* **1,** 522–530 (2014).
140. Girirajan, S. *et al.* Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet.* **7,** e1002334 (2011).
141. Derks, E. M. *et al.* A genome wide survey supports the involvement of large

copy number variants in schizophrenia with and without intellectual disability. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **162B,** 847–54 (2013).
142. Lowther, C. *et al.* Impact of IQ on the diagnostic yield of chromosomal microarray in a community sample of adults with schizophrenia. *Genome Med.* **9,** 1–13 (2017).
143. Baker, K., Raymond, F. L. & Bass, N. Genetic investigation for adults with intellectual disability: opportunities and challenges. *Curr. Opin. Neurol.* **25,** 150–8 (2012).
144. Department of Health. *Mental Capacity Act*. (HMSO, 2005).
145. Department of Health. *Guidance on nominating a consultee for research involving adults who lack capacity to consent*. (HMSO, 2008).
146. World Health Organisation. International Statistical Classification of Diseases and Related Health Problems 10th Revision. in *International Statistical Classification of Diseases and Related Health Problems 10th Revision* (World Health Organization, 2010).
147. Moss, S. *et al.* Reliability and validity of the PAS-ADD Checklist for detecting psychiatric disorders in adults with intellectual disability. *J. Intellect. Disabil. Res.* **42,** 173–183 (1998).
148. Rojahn, J., Matson, J. L., Lott, D., Esbensen, a J. & Smalls, Y. The Behavior Problems Inventory: an instrument for the assessment of self-injury, stereotyped behavior, and aggression/destruction in individuals with developmental disabilities. *J. Autism Dev. Disord.* **31,** 577–88 (2001).
149. Association for Clinical Cytogenetics. Professional Guidelines for Clinical Cytogenetics Postnatal Best Practice. 1–12 (2011).
150. Srebniak, M. I. *et al.* Types of array findings detectable in cytogenetic diagnosis: a proposal for a generic classification. *Eur. J. Hum. Genet.* **22,** 856–8 (2014).
151. Kushima, I. *et al.* High-resolution copy number variation analysis of schizophrenia in Japan. *Mol Psychiatry* 1–11 (2016). doi:10.1038/mp.2016.88
152. DECIPHER. Available at: https://decipher.sanger.ac.uk/.
153. Degenhardt, F. *et al.* Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **159 B,** 263–273 (2012).
154. Nagamani, S. C. S. *et al.* Phenotypic manifestations of copy number variation in chromosome 16p13.11. *Eur. J. Hum. Genet.* **19,** 280–286 (2011).
155. Garg, S. *et al.* Autism Spectrum Disorder Profile in Neurofibromatosis Type I. *J. Autism Dev. Disord.* (2014). doi:10.1007/s10803-014-2321-5
156. Hladilkova, E. *et al.* A recurrent deletion on chromosome 2q13 is associated with developmental delay and mild facial dysmorphisms. *Mol. Cytogenet.* **8,** 57 (2015).
157. Costain, G. *et al.* Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum. Mol. Genet.* **22,** 4485–4501 (2013).
158. Viñas-Jornet, M. *et al.* A common cognitive, psychiatric, and dysmorphic phenotype in carriers of NRXN1 deletion. *Mol. Genet. genomic Med.* **2,** 512–21 (2014).
159. Freunscht, I. *et al.* Behavioral phenotype in five individuals with de novo mutations within the GRIN2B gene. *Behav. Brain Funct.* **9,** 20 (2013).
160. Dimassi, S. *et al.* Interstitial 12p13.1 deletion involving GRIN2B in three

patients with intellectual disability. *Am. J. Med. Genet. Part A* **161,** 2564–2569 (2013).
161. Morisada, N. *et al.* A 12p13 GRIN2B deletion is associated with developmental delay and macrocephaly. *Hum. genome Var.* **3,** 16029 (2016).
162. Mishra, N., Kouzmitcheva, E., Orsino, A. & Minassian, B. A. Chromosome 12p Deletion Spanning the GRIN2B Gene Presenting With a Neurodevelopmental Phenotype: A Case Report and Review of Literature. *Child Neurol. Open* **3,** 1–6 (2016).
163. Poirsier, C. *et al.* Report on 3 patients with 12p duplication including GRIN2B. *Eur. J. Med. Genet.* **57,** 185–194 (2014).
164. Palumbo, O. *et al.* Report of a patient and further clinical and molecular characterization of interstitial 4p16.3 microduplication. *Mol. Cytogenet.* **8,** 15 (2015).
165. Philippe, A. *et al.* Xq25 duplications encompassing GRIA3 and STAG2 genes in two families convey recognizable X-linked intellectual disability with distinctive facial appearance. *Am. J. Med. Genet. A* **161A,** 1370–5 (2013).
166. Leroy, C. *et al.* Xq25 duplication: the crucial role of the STAG2 gene in this novel human cohesinopathy. *Clin. Genet.* **89,** 68–73 (2016).
167. Baker, K. *et al.* Identification of a human synaptotagmin-1 mutation that perturbs synaptic vesicle cycling. **125,** 1–9 (2015).
168. Bass, N. J. *et al.* Evidence for the association of the DAOA (G72) gene with schizophrenia and bipolar disorder but not for the association of the DAO gene with schizophrenia. *Behav. Brain Funct.* **5,** 28 (2009).
169. Castillo, A. *et al.* 19q13.32 microdeletion syndrome: Three new cases. *Eur. J. Med. Genet.* **57,** 654–658 (2014).
170. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5(5th ed.).* (American Psychiatric Association, 2013).
171. Perneger, T. V. What's wrong with Bonferroni adjustments. *BMJ* **316,** 1236–8 (1998).
172. Rothman, K. J. No Adjustments Are Needed for Multiple Comparisons. *Epidemiology* **1,** 43–46 (1990).
173. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (2017).
174. Wolfe, K. *et al.* Chromosomal microarray testing in adults with intellectual disability presenting with comorbid psychiatric disorders. *Eur. J. Hum. Genet.* **25,** 66–72 (2017).
175. Zuko, A. *et al.* Contactins in the neurobiology of autism. *Eur. J. Pharmacol.* **719,** 63–74 (2013).
176. van Daalen, E. *et al.* Social Responsiveness Scale-aided analysis of the clinical impact of copy number variations in autism. *Neurogenetics* **12,** 315–23 (2011).
177. Kashevarova, A. A. *et al.* Single gene microdeletions and microduplication of 3p26.3 in three unrelated families: CNTN6 as a new candidate gene for intellectual disability. *Mol. Cytogenet.* **7,** 97 (2014).
178. Hu, J. *et al.* CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *J. Neurodev. Disord.* **7,** 26 (2015).
179. Asami, O.-A., Zuko, A., Kleijer, K. T. E. & Burbach, J. P. H. A current view on contactin-4, -5, and -6: Implications in neurodevelopmental disorders. *Mol. Cell. Neurosci.* (2017).

180. Warnica, W. *et al.* Copy number variable micrornas in schizophrenia and their neurodevelopmental gene targets. *Biol. Psychiatry* **77,** 158–166 (2015).
181. Zahir, F. *et al.* Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children. *J. Med. Genet.* **44,** 556–561 (2007).
182. Bernier, R. *et al.* Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158,** 263–276 (2014).
183. Prontera, P. *et al.* Recurrent ~100 Kb microdeletion in the chromosomal region 14q11.2, involving CHD8 gene, is associated with autism and macrocephaly. *Am. J. Med. Genet. Part A* **164,** 3137–3141 (2014).
184. Smyk, M. *et al.* Novel 14q11.2 microduplication including the CHD8 and SUPT16H genes associated with developmental delay. *Am. J. Med. Genet. Part A* **170,** 1325–1329 (2016).
185. Breckpot, J. *et al.* Copy number variation analysis in adults with catatonia confirms haploinsufficiency of SHANK3 as a predisposing factor. *Eur. J. Med. Genet.* **59,** 436–443 (2016).
186. Courage, C., Houge, G., Gallati, S., Schjelderup, J. & Rieubland, C. 15q26.1 microdeletion encompassing only CHD2 and RGMA in two adults with moderate intellectual disability, epilepsy and truncal obesity. *Eur. J. Med. Genet.* **57,** 520–523 (2014).
187. Chénier, S. *et al.* CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems. *J. Neurodev. Disord.* **6,** 9 (2014).
188. Rudd, M. K. *et al.* Segmental duplications mediate novel, clinically relevant chromosome rearrangements. *Hum. Mol. Genet.* **18,** 2957–2962 (2009).
189. Bisgaard, A.-M. *et al.* Transmitted cytogenetic abnormalities in patients with mental retardation: Pathogenic or normal variants? *Eur. J. Med. Genet.* **50,** 243–255 (2007).
190. Yu *et al.* A recurrent 1.71 Mb genomic imbalance at 2q13 increases the risk of developmental delay and dysmorphism. *Clin. Genet.* **81,** 257–64 (2012).
191. Costain, G. *et al.* Adult neuropsychiatric expression and familial segregation of 2q13 duplications. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **165,** 337–344 (2014).
192. Riley, K. N. *et al.* Recurrent deletions and duplications of chromosome 2q11.2 and 2q13 are associated with variable outcomes. *Am. J. Med. Genet. Part A* **167,** 2664–2673 (2015).
193. UUnique The Rare Chromosomal Disorder Support Group. Available at: http://www.rarechromo.org.
194. Köhler, S. *et al.* The human phenotype ontology in 2017. *Nucleic Acids Res.* **45,** D865–D876 (2017).
195. Rojahn, J. *et al.* The Behavior Problems Inventory-Short Form for individuals with intellectual disabilities: part II: reliability and validity. *J. Intellect. Disabil. Res.* **56,** 546–65 (2012).
196. Wickham., H. *ggplot2: Elegant Graphics for Data Analysis.* (Springer-Verlag, 2009).
197. Fox, J. The R Commander: A Basic-Statistics Graphical User Interface to R. *J. Stat. Softw.* **14,** 1–42 (2005).
198. Greene, D., Richardson, S. & Turro, E. OntologyX: A suite of R packages for working with ontological data. *Bioinformatics* **33,** 1104–1106 (2017).

199. Wiest, M. M., German, J. B., Harvey, D. J., Watkins, S. M. & Hertz-Picciotto, I. Plasma fatty acid profiles in autism: a case-control study. *Prostaglandins. Leukot. Essent. Fatty Acids* **80,** 221–7 (2009).
200. Sheikh, a. M. *et al.* Cathepsin D and apoptosis related proteins are elevated in the brain of autistic subjects. *Neuroscience* **165,** 363–370 (2010).
201. Russell, M. W. *et al.* Functional analysis of candidate genes in 2q13 deletion syndrome implicates FBLN7 and TMEM87B deficiency in congenital heart defects and FBLN7 in craniofacial malformations. *Hum. Mol. Genet.* **23,** 4272–84 (2014).
202. Girirajan, S. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* **42,** 203–9 (2010).
203. Yu, H.-C. *et al.* Discovery of a potentially deleterious variant in TMEM87B in a patient with a hemizygous 2q13 microdeletion suggests a recessive condition characterized by congenital heart disease and restrictive cardiomyopathy. *Mol. Case Stud.* **2,** a000844 (2016).
204. Stone, J. L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).
205. Grozeva, D. *et al.* Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch. Gen. Psychiatry* **67,** 318–27 (2010).
206. Grozeva, D. *et al.* Reduced burden of very large and rare CNVs in bipolar affective disorder. *Bipolar Disord.* **15,** 893–898 (2013).
207. McQuillin, A. *et al.* Analysis of genetic deletions and duplications in the University College London Bipolar Disorder case control sample. *Eur. J. Hum. Genet.* **19,** 588–592 (2011).
208. Malhotra, D. *et al.* High frequencies of de novo cnvs in bipolar disorder and schizophrenia. *Neuron* **72,** 951–963 (2011).
209. Noor, A. *et al.* Copy number variant study of bipolar disorder in Canadian and UK populations implicates synaptic genes. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **165,** 303–313 (2014).
210. Al Eissa, M. M. *et al.* Exome sequence analysis and follow up genotyping implicates rare ULK1 variants to be involved in susceptibility to schizophrenia. *Ann. Hum. Genet.* 1–5 (2017).
211. Spitzer, R. L., Endicott, J. The schedule for affective disorders and schizophrenia, lifetime version (3rd ed.). (1977).
212. Qubit Quantification Protocol. Available at: http://www.ieg.uu.se/digitalAssets/176/c_176882-l_3-k_qubitquickrefcard.pdf.
213. Psych Array. Available at: https://emea.illumina.com/products/by-type/microarray-kits/infinium-psycharray.html?langsel=/gb/.
214. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
215. Lin, C.-F., Naj, A. C. & Wang, L.-S. Analyzing copy number variation using SNP array data: protocols for calling CNV and association tests. *Curr. Protoc. Hum. Genet.* **79,** Unit 1.27. (2013).
216. GenomeStudio Software. Available at: https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html?langsel=/gb/.
217. GenomeStudio output Guidelines. Available at:

http://penncnv.openbioinformatics.org/en/latest/user-guide/input/.
218. PennCNV CNV calling protocol. Available at: http://penncnv.openbioinformatics.org/en/latest/user-guide/test/.
219. Cristino, A. S. *et al.* Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol. Psychiatry* **19,** 294–301 (2014).
220. PLINK burden analysis guidelines. Available at: http://zzz.bwh.harvard.edu/plink/cnv.shtml.
221. Marenne, G. *et al.* Assessment of copy number variation using the Illumina Infinium 1M SNP-array: a comparison of methodological approaches in the Spanish Bladder Cancer/EPICURO study. *Hum. Mutat.* **32,** 240–8 (2011).
222. Polyak, A., Kubina, R. M. & Girirajan, S. Comorbidity of intellectual disability confounds ascertainment of autism: Implications for genetic diagnosis. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168,** 600–608 (2015).
223. Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **17,** 1–19 (2016).
224. Dean, K. & Murray, R. M. Environmental risk factors for psychosis. *Dialogues Clin. Neurosci.* **7,** 69–80 (2005).
225. Gershon, E. S. *et al.* Genetic analysis of deep phenotyping projects in common disorders. *Schizophr. Res.* (2017). doi:10.1016/j.schres.2017.09.031
226. Owen, M. J. New approaches to psychiatric diagnostic classification. *Neuron* **84,** 564–571 (2014).
227. Costain, G. *et al.* Evaluating genetic counseling for individuals with schizophrenia in the molecular age. *Schizophr. Bull.* **40,** 78–87 (2014).
228. Cubells, J. F. *et al.* Pharmaco-genetically guided treatment of recurrent rage outbursts in an adult male with 15q13.3 deletion syndrome. *Am. J. Med. Genet. Part A* **155,** 805–810 (2011).

Appendix

Clinically significant CNV questionnaire

Did you feel it was appropriate to feedback the genetic test result to your patient?

Yes ☐ No ☐

If YES what was your experience of feeding back the result?

If NO why did you not feel it was appropriate to feedback the result?

Did you feel it was appropriate to feedback the genetic test result to the patient's family members/carers?

Yes ☐ No ☐

If YES what was your experience of feeding back the result?

If NO why did you not feel it was appropriate to feedback the result?

Was the patient referred to clinical genetics?

Yes ☐ No ☐

If so what was your experience of the referral process?

Have family members of the patient undergone genetic testing?

Yes ☐ No ☐ Unknown ☐

If so have you been informed of the results?

What do you think was the psychological impact of receiving the diagnosis outcomes (for patient/family/carer)?

Has the patient been in contact with any support groups?

Yes ☐ No ☐Unknown ☐

Please provide details if known

Has your patient's GP been informed of the genetic test result?

Yes ☐ No ☐

Have any other professionals involved in the patient's care been informed of the genetic test result?

Yes ☐ No ☐

If so please provide details

Have there been any management changes (medical and social) for the patient related to their genetic diagnosis? Examples given below; please provide relevant details

Medical screening investigations (e.g. blood tests, echocardiogram, neuroimaging)

Specialist medical referral (e.g. cardiology, ophthalmology, endocrinology)

Discontinuation of previously recommended medical screening

Changes in medication

Changes to medical services input or eligibility

Changes to social circumstances or eligibility including care/housing/benefits

Other, please describe

Prep_calling.sh script

```
## Script to prepare for penncnv call of illumnia psychchip samples

projectdir=/home/rejujht/molpsych/projects/penncnv/psychchip

listdir=$projectdir/lists

# mkdir -p $listdir

## 1) Generate listfiles for array submission of penncnv jobs

# find $projectdir/psychchip_intensity-data/ -iname "*.gtc.txt" -type f > $listdir/listfile_all.txt

## 1b) (OPTIONAL) Remove very bad samples identified from first round from the list

### Script takes 3 arguments: 1) listfile to update, 2) very_bad_exclude list as generated by qc_cnvs_find_really_BAD.R, 3) sampleSheet cleened

# Rscript remove_very_bad_from_list.R $listdir/listfile_all.txt cnv-calls_130317/pchip.penn_really_BAD.excluded SampleSheet_cleaned.txt

# ## 1- continued) split listfile all into 10

# awk 'NR>=1&&NR<=500' $listdir/listfile_all.txt > $listdir/listfile1.txt

# awk 'NR>=501&&NR<=1000' $listdir/listfile_all.txt > $listdir/listfile2.txt

# awk 'NR>=1001&&NR<=1500' $listdir/listfile_all.txt > $listdir/listfile3.txt

# awk 'NR>=1501&&NR<=2000' $listdir/listfile_all.txt > $listdir/listfile4.txt

# awk 'NR>=2001&&NR<=2500' $listdir/listfile_all.txt > $listdir/listfile5.txt
```

```
# awk 'NR>=2501&&NR<=3000' $listdir/listfile_all.txt > $listdir/listfile6.txt

# awk 'NR>=3001&&NR<=3500' $listdir/listfile_all.txt > $listdir/listfile7.txt

# awk 'NR>=3501&&NR<=4000' $listdir/listfile_all.txt > $listdir/listfile8.txt

# awk 'NR>=4001&&NR<=4500' $listdir/listfile_all.txt > $listdir/listfile9.txt

# ## awk 'NR>=4501&&NR<=5500' $listdir/listfile_all.txt > $listdir/listfile10.txt

# ## 2) Strip header to file and change name of B-allel and R header to prepare for pbf file creation and CNV calling

# while read p;

# do echo $p

#     idcode=$(basename $p)

#     head -n12 $p > $p.header

#     sed -i '1,11d' $p

#     sed -i 's/bAllele Freq/${idcode}.B Allele Freq/g' $p

#     sed -i 's/Log R Ratio Illumina/${idcode}.Log R Ratio/g' $p

#     sed -i 's/SNP Name/Name/g' $p

#     sed -i 's/Chromosome/Chr/g' $p

# done < $projectdir/lists/listfile$SGE_TASK_ID.txt

## 2) Generate pbf file
```

/share/apps/genomics/PennCNV-1.0.3/compile_pfb.pl --listfile $listdir/listfile_all.txt --output $projectdir/pchip.pfb

3) Generate GCmodel

wget http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/gc5Base.txt.gz

gzip -fd gc5Base.txt.gz

sort -k 2,2 -k 3,3n gc5Base.txt > gc5Base.sorted.txt

use pennCNV function to calulate GC

/share/apps/genomics/PennCNV-1.0.3/cal_gc_snp.pl $projectdir/gc5Base.sorted.txt $projectdir/pchip.pfb -output $projectdir/pchip.gcmodel

clean up

rm $projectdir/gc5Base.sorted.txt

rm $projectdir/gc5Base.txt

4) Generate a sex file with updated info from the gwas data

Rscript generate_sex-file.R

Call_cnvs_array.sh script

```
## In .bashrc

# export PERL5LIB=/share/apps/genomics/PennCNV-1.0.3/kext/5.10.1/x86_64-linux:$PERL5LIB

projectdir=/home/rejujht/molpsych/projects/penncnv/psychchip

outdir=$projectdir/cnv-calls # output dir for raw cnv calls

penncnv=/share/apps/genomics/PennCNV-1.0.3 # Penncnv executables

outprefix=pchip.penn

mkdir -p $outdir

## Autosome detection

$penncnv/perl-5.10.1/bin/perl $penncnv/detect_cnv.pl -test \

  -hmm $projectdir/hhall.hmm \

  -pfb $projectdir/pchip.pfb \

  -gcmodel $projectdir/pchip.gcmodel \

  -list lists/listfile$SGE_TASK_ID.txt \

  -confidence \

  -log $outdir/$outprefix.$SGE_TASK_ID.log \

  -out $outdir/$outprefix.$SGE_TASK_ID.rawcnv \
```

```
# ## ChrX detection

$penncnv/perl-5.10.1/bin/perl $penncnv/detect_cnv.pl -test \

  -hmm $projectdir/hhall.hmm \

  -pfb $projectdir/pchip.pfb \

  -gcmodel $projectdir/pchip.gcmodel \

  -list lists/listfile$SGE_TASK_ID.txt \

  -chrx -sexfile $projectdir/pchip_penncnv_gender.txt \

  -confidence \

  -log $outdir/$outprefix.$SGE_TASK_ID.sex.log \

  -out $outdir/$outprefix.$SGE_TASK_ID.sex.rawcnv \
```

Post_calling.sh script

```
## Variables

projectdir=/home/rejujht/molpsych/projects/penncnv/psychchip

outdir=$projectdir/cnv-calls # output dir for raw cnv calls

prefix=pchip.penn

Rpath=/share/apps/R-3.1.1/bin/Rscript

## Convert output to coloumn style

cat $outdir/*.rawcnv > $outdir/$prefix.rawcnv

mkdir $outdir/sexlog

mv $outdir/*sex.log $outdir/sexlog

cat $outdir/*.log > $outdir/$prefix.log

# Clean up

rm $outdir/$prefix.*.rawcnv

rm $outdir/$prefix.*.log

# Arguments: 1) rawcnv file, 2) Sample_id file

$Rpath $projectdir/convert_cnv_output.R $outdir/$prefix.rawcnv $projectdir/SampleSheet_cleaned.txt

## Get values from log

pennlog=$outdir/$prefix.log
```

```
grep "NOTICE: quality summary for" $pennlog | sed 's/NOTICE: quality summary for //g' > ${pennlog/.log/.log.qc}

## Run QC

mkdir -p $outdir/plots

# Arguments: 1) cnv-call dir, 2) penncnv output prefix, 3) penncnv log, 4) listfiles, 5) Sample_id file

$Rpath $projectdir/qc_cnvs.R $outdir $prefix $pennlog $projectdir/lists/listfile_all.txt $projectdir/SampleSheet_cleaned.txt

## Find CNV regions of interest # Arguments: 1) cnv-call dir, 2) cnv-inputfile, 3) region of interest file path

$Rpath $projectdir/call_regions_of_interest.r $outdir $prefix.qc $projectdir/input/regions_of_interest_80.txt
```

Publications