Spring 2014

Genomic copy number variation in schizophrenia

Danielle Song Rudd

*University of Iowa*

Copyright 2014 Danielle Song Rudd

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/4739

**Recommended Citation**

Rudd, Danielle Song. "Genomic copy number variation in schizophrenia." PhD (Doctor of Philosophy) thesis, University of Iowa, 2014.

https://doi.org/10.17077/etd.ezs0wnfu

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Genetics Commons
GENOMIC COPY NUMBER VARIATION IN SCHIZOPHRENIA

by

Danielle Song Rudd

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Genetics in the Graduate College of The University of Iowa

May 2014

Thesis Supervisor: Professor Thomas H. Wassink
This is to certify that the Ph.D. thesis of

Danielle Song Rudd

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Genetics at the May 2014 graduation.

Thesis Committee:

Thomas H. Wassink, Thesis Supervisor

Michael G. Anderson

Deborah V. Dawson

Pedro Gonzalez-Alegre

Toshihiro Kitamoto

Peggy C. Nopoulos
To dad, Grandma Nellie, Uncle Gary, Aunt Sharon – I miss you every day, and to my partner Jeffery who has supported me through each step of this amazing journey – thanks for always being there for me
ACKNOWLEDGMENTS

First and foremost, I would like to thank all of the study participants. None of the work in this thesis would have been possible without them and I am very grateful for their involvement. Second, I would like to thank my family, especially my parents, for their unconditional love and support. I would like to thank my thesis advisor, Dr. Thomas Wassink, who provided an open, supportive environment and let me pursue my scientific interests. To my fellow graduate students, Dr. Lea Davis and Dr. Kacie Meyer: I cannot ever stop thanking you for your inspiration, assistance and constant encouragement. Your guidance has been invaluable and I could not have made it though this without your love and support; I am forever grateful. I would also like to thank members of the Wassink Lab, past and present, Sarah Mugge, Amy Librant, Valerie Buffard and especially Dr. Eric Epping, Michael Axelsen, and Sarah Casey who have contributed to generating the data that has been reported in this thesis. Thanks to Dr. Nancy C. Andreasen for providing the research subjects and a wealth of resources for this thesis project. John Beck and Charles Searby of the Sheffield lab have also provided valuable technical insight and have assisted in the generation of data reported in this thesis. I would like to thank the Interdisciplinary Graduate Program in Genetics, especially Dr. Dan Eberl, Isabelle Hardy and Linda Hurst. Last, I would like to thank my committee members, Dr. Michael G. Anderson, Dr. Deborah Dawson, Dr. Pedro Gonzalez-Alegre, Dr. Toshihiro Kitamoto and Dr. Peggy Nopoulos for providing me with valuable insight, guidance and support throughout my graduate career.
ABSTRACT

Schizophrenia (OMIM 181500) is an incurable and severe psychiatric disorder comprised of three symptom domains (positive symptoms, negative symptoms and cognitive impairments) with a worldwide prevalence of approximately 1%. There is a substantial amount of evidence demonstrating that schizophrenia has a strong genetic component. Broad-sense heritability estimates range from 64-80% and first-degree relatives of schizophrenia patients have 10-fold increased risk of developing the disorder compared to the general population. It is thought that both single nucleotide polymorphisms and copy number variants (CNVs) contribute to the heritability of schizophrenia. This thesis focuses on the role of CNVs in the etiology of schizophrenia.

We performed a genome-wide CNV analysis of 166 schizophrenia patients and 52 psychiatrically healthy controls. In our overall CNV analysis we did not find any significant differences between cases and controls across a variety of CNV categories, nor did we find significant differences when CNVs were partitioned by size (small, medium or large). However, we were the first group to consider small CNVs (< 100-500 kb) in a multiple-hit model where we observed that a slightly higher proportion of case subjects had two-or-more conservative CNVs. We defined a CNV as conservative if it met any of the following three criteria: 1) a known deleterious CNV, 2) a CNV > 1 Mb that was novel to the Database of Genomic Variants (DGV) or 3) a CNV < 1 Mb that was novel to the DGV and that overlapped the coding region of a gene of interest. Genes of interest included genes with a previous association with a neuropsychiatric disorder, or genes with high or specific brain expression, or an association with any other neurocognitive or neuropsychiatric disorders. Two of our case subjects who harbored the highest amount of conservative CNVs also shared a 15q11.2 breakpoint 1-2 (BP1-2) deletion which is a compelling candidate risk locus for schizophrenia. We also found that a slightly higher proportion of case subjects harbored clinically significant CNVs.
(conservative CNVs > 1 Mb or clinically recognized as deleterious) when compared to controls. Additionally, we hypothesized that individuals with more severe CNVs would show more neurocognitive deficits and more pronounced abnormalities in brain structure volume, however, we had largely negative results. We also reported a case of childhood-onset schizophrenia who had three large chromosomal abnormalities including a paternally inherited 2.2 Mb deletion of chromosome 3p12.2-p12.1, a *de novo* 17.6 Mb duplication of chromosome 16q22.3-q24.3 and a *de novo* 43 Mb deletion of chromosome Xq23-q28.

We were able to confirm previous reports of CNV findings in schizophrenia such as the involvement of large, rare and *de novo* CNVs. In addition, the work in this thesis leads us to propose a multiple-hit CNV model which requires a shift in the way we currently approach schizophrenia genetics. First, we must identify all CNVs, especially those of smaller size (< 100 kb). Next, we require a more precise understanding of the impact that CNVs have on gene expression, especially in the brain. With all of the right tools in place, we can move towards a disease model for schizophrenia that considers the totality of CNVs in any given individual. We propose that the use of recurrent CNVs such as the 15q11.2 BP1-2 CNV is a good starting point for studying a multiple-hit CNV model.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... viii
LIST OF FIGURES .......................................................................................................... ix
LIST OF ABBREVIATIONS .............................................................................................. x

CHAPTER I INTRODUCTION .......................................................................................... 1

Clinical and Epidemiological Factors of Schizophrenia ...................................... 1
Schizophrenia Genetics Background .................................................................... 2
Heritability Estimates of Schizophrenia ................................................................. 3
Linkage Studies ........................................................................................................... 3
Genome Wide Association Studies ........................................................................ 7
Copy Number Variant Studies ............................................................................... 10
Background of Recurrent CNV at Chromosome 15q11.2 BP1-2 ......................... 14
Recurrent CNVs ........................................................................................................ 14
Association with Prader-Willi Syndrome and Angelman Syndrome ............. 15
Association with Neuropsychiatric Disorders ................................................. 17
15q11.2 BP1-2 Gene Function and Expression ..................................................... 20
Polygenic Model of Schizophrenia ......................................................................... 22

CHAPTER II A GENOME-WIDE CNV ANALYSIS OF SCHIZOPHRENIA
REVEALS A POTENTIAL ROLE FOR A MULTIPLE-HIT MODEL .................. 30

Introduction .............................................................................................................. 30
Materials and Methods .......................................................................................... 31
Study Sample ............................................................................................................ 31
CNV Detection ......................................................................................................... 32
15q11.2 BP1-2 Validation Assays ............................................................................ 33
Results ....................................................................................................................... 34
Analysis of All CNVs ......................................................................................... 34
15q11.2 BP1-2 Deletions ...................................................................................... 35
Discussion .................................................................................................................. 37

CHAPTER III COPY NUMBER VARIANTS EFFECTS ON COGNITIVE
ABILITIES AND BRAIN STRUCTURE VOLUMES ........................................... 60

Introduction .............................................................................................................. 60
Brain Morphology ................................................................................................. 61
Cognition ................................................................................................................... 63
Materials and Methods .......................................................................................... 64
Study Sample ............................................................................................................ 64
Brain Image Acquisition ....................................................................................... 66
Neurocognitive Assessment .................................................................................. 67
CNV Detection ........................................................................................................ 68
Analyzing Relationships between CNVs and Both Brain Structure
Volumes and Cognition ......................................................................................... 69
Results ....................................................................................................................... 70
Discussion .................................................................................................................. 71
CHAPTER IV CHILDHOOD-ONSET SCHIZOPHRENIA CASE WITH 2.2 MB DELETION AT CHROMOSOME 3P12.2-P12.1 AND TWO LARGE CHROMOSOMAL ABNORMALITIES AT 16Q22.3-Q24.3 AND XQ23-Q28 .................................................................80

Introduction .....................................................................................................80
Clinical Report ................................................................................................81
Methods and Results .......................................................................................82
  Chromosomal Microarray Study .............................................................82
Discussion .......................................................................................................83

CHAPTER V CONCLUSIONS AND FUTURE DIRECTIONS .....................................96

Conclusions .....................................................................................................96
  Summary of Results ................................................................................97
  Implications of Present Findings ...........................................................101
Future Directions ..........................................................................................106

REFERENCES ................................................................................................................108
LIST OF TABLES

Table 1-1. Summary of studies specifically studying 15q11.2 BP1-2 deletion.................26
Table 2-1. Summary of CNVs compared between cases and controls. .........................40
Table 2-2. Summary of novel (not found in DGV or in controls) CNVs compared between cases and controls.................................................................43
Table 2-3. Summary of all conservative CNVs identified in cases. Individuals are separated by a vertical space.................................................................46
Table 2-4. Summary of all conservative CNVs identified in controls. Individuals are separated by a vertical space.................................................................52
Table 2-5. Summary of the percentage of individuals with single or multiple conservative CNVs. ..........................................................................................55
Table 2-6. Summary of clinically significant conservative CNVs. 1Same individual ....................................................................................................................56
Table 2-7. Summary of additional conservative CNVs in case 1 and case 2. .................57
Table 3-1. Neurocognitive test battery. Five cognitive domains and their component tests........................................................................................................74
Table 3-2. Table provides number of subjects included in the analysis. .......................75
Table 3-3. Summary of results from ANCOVA between CNV state and intelligence or cognitive measures.............................................................................76
Table 3-4. Summary of results from ANCOVA between CNV category and intelligence or cognitive measures.................................................................77
Table 3-5. Summary of results from ANCOVA between CNV state and brain structure volume. ............................................................................................78
Table 3-6. Summary of results from ANCOVA between CNV category and brain structure volume. ....................................................................................79
Table 4-1. Summary of 20 genes located in the 16.7 Mb duplication at chr16q22.3-q24.3 that are of interest. .................................................................86
LIST OF FIGURES

Figure 1-1. Ideogram of chromosome 15q11.2 representing region between BP1-3........29

Figure 2-1. UCSC screen capture of 15q11.2 BP1-2 deletions with case 1 & 2 on the bottom in green and case 3 on top in blue. Deletion coordinates for case 1 (627 kb), as determined by Affymetrix Genotyping Console 4.0, are chr15: 20224751-20852202. Coordinates for case 2 (581 kb) are chr15: 20224751-20805409 and coordinates for case 3 (975 kb) are chr15: 20059085-21033790. .................................................................................................................58

Figure 2-2. Scatterplot where each dot represents the number of conservative CNVs in one individual. Overlapping dots appear darker..................................................59

Figure 4-1. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 2.2 Mb deletion at Chr3p12.2-p12.1. The individual represented in blue (top) is the patient and the individual represented in purple (bottom) is the father. The top of the panel shows the entire chromosome 3 and the bottom shows a zoomed in view of the deletion. ........................................................................................................90

Figure 4-2. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 16.7 MB duplication at chr16q22.3-q24.3. The patient is represented in blue. The top of the panel shows the entire chromosome 16 and the bottom shows a zoomed in view of the duplication. ..................................................................................91

Figure 4-3. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 43 Mb deletion at chXq23-q28. The patient is represented in blue. The top of the panel shows the entire X chromosome and the bottom shows a zoomed in view of the deletion...................................................................................................92

Figure 4-4. Chr3p12.2-p12.1 deletion coordinates in the patient and in the father. UCSC screen capture with patient and father both in red. The deletion, with coordinates determined by Affymetrix Genotyping Console at chr3:82554810-84777050, does not overlap any UCSC genes. ......................93

Figure 4-5. Chr16q22.3-q24.3 duplication coordinates in the patient. UCSC screen capture with patient in blue. The duplication coordinates determined by Affymetrix Genotyping Console are chr16: 71224799-88815024. The duplication overlaps 163 genes. For a summary of the most relevant genes in this region, see Table 4-3. ........................................................................................................94

Figure 4-6. ChrXq23-q28 deletion coordinates in the patient. UCSC screen capture with the patient in red. The deletion coordinates as determined by Affymetrix Genotyping Console are chrX: 111858197-154582680. .................................................................95
LIST OF ABBREVIATIONS

PGC – Psychiatric Genome Consortium
SNP – Single nucleotide polymorphism
LD – Linkage disequilibrium
GWAS – Genome-wide association study
ADHD – Attention deficit hyperactivity disorder
ASD – Autism spectrum disorder
CNV – Copy number variant
DGV – Database of Genomic Variants
ID – Intellectual disability
VCFS – Velo-cardio-facial syndrome
CDRV – Common-disease/rare-variant
NAHR – Non-allelic homologous recombination
Segdups – Segmental duplications
PWS – Prader-Willi syndrome
AS – Angelman syndrome
BP1-2 – Breakpoint 1 through breakpoint 2
DD – Developmental delay
CA – Congenital abnormalities
OCD – Obsessive compulsive disorder
FSIQ – Full-scale IQ
VIQ – Verbal IQ
PIQ – Performance IQ
CMA – Chromosomal microarray
CNS – Central nervous system
FMRP – Fragile-X mental retardation protein
CDCV – Common-disease/common-variant
MAF – Minor allele frequency
CASH – Comprehensive Assessment of Symptoms and History
CQC – Contrast quality control
MAPD – Median absolute pairwise difference
CGH – Comparative genomic hybridization
qPCR – Quantitative polymerase chain reaction
MRI – Magnetic resonance imaging
PSYCH – Psychiatric Symptoms You Currently Have
SANS – Scale for the Assessment of Negative Symptoms
SAPS – Scale for the Assessment of Positive Symptoms
AIMS – Abnormal involuntary movement scale
FHRDC – Family History-Research Diagnostic Criteria
PD – Proton density
BRAINS – Brain Research: Analysis of Images, Networks and Systems
ANN – Artificial neural network
COS – Childhood-onset schizophrenia
Clinical and Epidemiological Factors of Schizophrenia

Schizophrenia (OMIM 181500) is an incurable and severe psychiatric disorder with a worldwide prevalence of approximately 1% (Gottesman & Shields 1967). Schizophrenia comprises three symptom domains: positive symptoms, negative symptoms and cognitive impairments (American Psychiatric Association 2000). Positive symptoms include hallucinations, delusions and disorganized speech and/or thought. An example of a hallucination is hearing voices that converse with one another or with the patient. An example of a delusion is a paranoid belief that external forces are conspiring against the patient (Freedman 2003). Negative symptoms include flat affect, alogia (lack of speech) and avolition (lack of motivation) (Connor & Akbarian 2008). Other examples of negative symptoms are inability to pay attention, the loss of a sense of pleasure and social withdrawal (Freedman 2003). Positive and negative symptoms will vary in intensity throughout the course of the disease and one may predominate over the other at any particular time. The cognitive component of schizophrenia is a core feature of the disease and includes impairment in working memory, learning and executive function (Ross et al. 2006; Ho et al. 2008; Glessner et al. 2010).

Despite the fact that schizophrenia was first characterized over a hundred years ago, the underlying mechanisms remain elusive. Schizophrenia is a psychiatric illness with profound consequences for patients, their family members and society at large (van Dongen & Boomsma 2013). The disease is a life-long disorder with a relatively young onset; initial diagnosis typically takes place in the second or third decade of life (American Psychiatric Association 2000). Many patients require assistance in one or more aspects of independent living. There is also a high unemployment and homeless rate
for schizophrenia patients (Bond 2004; Kinoshita et al. 2013). In developed countries, schizophrenia is a leading cause of early death (Brown 1997; Lewis & Gonzalez-Burgos 2006; McGrath et al. 2008; Kilbourne et al. 2009) and the lifetime prevalence of suicide is 10% among patients (Siris 2001). Other factors that play a role in early death of patients are smoking/drug use, heart disease and diabetes (Saha et al. 2007; Osborn et al. 2008; Brown et al. 2010).

There are some environmental factors that are associated with increased incidence of schizophrenia and they include migrant status, urbanicity and economic status (Cantor-Graae & Selten 2005; Di Forti et al. 2007; McGrath et al. 2008). Other known associated environmental factors are maternal malnutrition during pregnancy (Susser & Lin 1992; St Clair et al. 2005), maternal infection in utero (Dalman et al. 2008) and obstetrical complications such as prematurity, low birth weight and complications of delivery (Cannon et al. 2002).

**Schizophrenia Genetics Background**

There is a substantial amount of evidence that schizophrenia has a strong genetic component. Monozygotic twins with schizophrenia have a concordance rate of ~50% while dizygotic twins have a concordance rate of ~17% (McGuffin et al. 1984; Cannon et al. 1998; Cardno & Gottesman 2000). Adoption studies reveal that first-degree relatives of schizophrenia patients have 10-fold increased risk of developing the disorder compared to the general population (Cardno & Gottesman 2000; Lichtenstein et al. 2009). Additionally, children born to mothers with schizophrenia, who are adopted out, still have a significantly elevated risk of developing the disorder (Heston 1966; Rosenthal et al. 1971; Wender et al. 1974; Kety et al. 1994; Ingraham & Kety 2000). When two
parents have schizophrenia, there is a 27.3% risk their offspring will develop the disorder (Gottesman et al. 2010).

Heritability Estimates of Schizophrenia

Heritability is the proportion of phenotypic variance that can be accounted for by the genotypic variance observed in a population. Traditional heritability, now known as broad-sense heritability, has historically been measured using familial data. Broad-sense heritability is based on a degree of resemblance between relatives and represents the total genetic variance (including additive, dominant and epistatic interactions) as a proportion of the phenotypic variance in the population sampled (Gottesman & Shields 1967). Broad-sense heritability estimates for schizophrenia range from 64% - 80% (Sullivan et al. 2003; Lichtenstein et al. 2009). Narrow sense heritability estimates only account for the additive effect of common variants, and ignore gene-gene, gene-environment and epistatic interactions (Gottesman & Shields 1967). Two recent estimates of narrow sense heritability for schizophrenia have come from the Psychiatric Genome Consortium (PGC), estimating the current amount of variation in liability that can be attributed to single nucleotide polymorphisms (SNPs) at 23% for Europeans of mixed ancestry (Lee et al. 2012; Lee et al. 2013) and 32% for subjects with Swedish ancestry (Ripke et al. 2013).

Linkage Studies

With all of the accumulated evidence that genetics plays a role in schizophrenia etiology, researchers have been trying to identify those components in a systematic manner. As genetic methodology advanced and certain technologies became readily
available, linkage studies became the first line of research for schizophrenia studies. Traditional linkage studies identify regions of the genome that harbor a mutation and that follow the same inheritance pattern of the disease or phenotype in question. This methodology generally assumes that the mutation being studied is highly penetrant, follows Mendelian inheritance patterns and requires large or multiple extended pedigrees with multiple affected individuals. One of the earliest linkage studies of schizophrenia was performed with five Icelandic and two British kindreds by Sherrington et al. (1988), and they reported linkage between schizophrenia and a group of markers on the long arm of chromosome 5. This study treated non-affected family members who exhibited major psychosis or schizophreniform disorder as affecteds and used models with varying the degrees of penetrance to achieve LOD scores ranging from 2.45 – 6.49 (Sherrington et al. 1988). However, linkage to the long arm of chromosome 5 was promptly rebuked by several studies (Kennedy et al. 1988; Detera-Wadleigh et al. 1989; St Clair et al. 1989; Aschauer et al. 1990; Crowe et al. 1990; McGuffin et al. 1990), highlighting the complexities and controversies in schizophrenia linkage studies that were to follow.

There has been a lack of linkage reports that are both highly significant and consistently reproducible. Instead there are four regions in particular that were initially identified as statistically significant, with follow-up studies that were suggestive of linkage to these regions: 1q21-q22 (Brzustowicz et al. 2000; Rosa et al. 2002; Zheng et al. 2006), 8p21 (Kendler et al. 1996; Blouin et al. 1998; Gurling et al. 2001), 13q32 (Blouin et al. 1998; Brzustowicz et al. 1999), and 6p24-p22 (Moises et al. 1995; Straub et al. 1995; 1996; Schwab et al. 2000; Maziade et al. 2001). It is important to note that linkage in these regions did not go completely undisputed. The fact that these four regions are the most robust reports highlights the questionable nature of linkage methodology for schizophrenia research.

In an attempt to address what was viewed as shortcomings in the field, low reproducibility and conflicting results, several methodological strategies were applied to
studies in an attempt to increase power of linkage detection. One major manipulation applied was to treat schizophrenia as a spectrum disorder and the subsequent widening or narrowing of the spectrum for study inclusion. Some examples of spectrum phenotypes used to define affected individuals include: schizophreniform disorder, schizoaffective disorder, schizotypal personality disorder and even bipolar disorder with psychotic features. Beyond this, investigators studied quantitative biological traits putatively related to schizophrenia such as paranoid personality traits; eye tracking dysfunction; attention impairment; allusive thinking; neuropsychological impairment and characteristic auditory evoked potentials (Faraone et al. 1995a). This manipulation of a schizophrenia spectrum was an attempt to address one of the largest confounders in schizophrenia research, phenotypic heterogeneity. Schizophrenia is based on a cluster of core symptoms rather than on specific biological components; therefore, some studies tried widening the schizophrenia phenotype for study inclusion in an attempt to capture more cases of “schizophrenia-like” individuals and thus improve power to detect linkage (Sherrington et al. 1988; Faraone et al. 1995a; Brzustowicz et al. 2000). An alternative approach was to narrow the diagnostic criteria for study inclusion based on the assumption that some of the underlying genetic architecture directly correlated with different aspects of varying phenotype. More stringent diagnostic criteria included patients with a longer duration of illness or defining a maximum age of onset for inclusion in linkage studies (Levinson & Mowry 1991; Rosa et al. 2002).

Another methodological approach to strengthen schizophrenia linkage detection is the meta-analysis, which combines data from a series of previously published independent linkage studies. The underlying hypothesis is that nominal linkage findings from several studies could reach statistical significance with the systematic pooling of data (Badner & Gershon 2002b). The most robust approach would be one where all of the original genotypic information is obtained and a new linkage analysis is performed. This is an ideal situation and does not represent the reality of published linkage studies;
therefore, different statistical strategies have been employed for meta-analyses (Badner & Goldin 1999; Wise et al. 1999). Meta-analysis of 18 linkage studies identified significant linkage at 8p, 13q, and 22q (Badner & Gershon 2002a). Another meta-analysis of 20 studies found significant linkage at 2p12-q22.1, 2q22.1-q23.3, 3p25.3-p22.1, 5q23.2-q34, 6pter-p21.1, 8p22-p21.1, 11q22.3-q24.1, 14pter-q13.1, 20p12.3-p11 and 22pter-q12.3 (Lewis et al. 2003).

Linkage studies have been successful for monogenic diseases that are caused by single variants of large effect, high penetrance and that follow Mendelian inheritance patterns. Schizophrenia is not a monogenic disease, and is likely caused by many genetic variants of varying effect sizes and levels of penetrance, so linkage studies have been less fruitful in this particular area of research. There are many contributing factors to what has been considered a shortcoming of schizophrenia linkage studies. First, varying populations were used in early linkage studies. Today there is a better understanding of how population stratification can confound a genetic study. Second, many of the models used for previous linkage studies were inconsistent with parameters, such as mode of inheritance and penetrance, being arbitrarily assigned. While researchers were making their best assumptions, it is clear that not knowing the true values of these parameters can produce negative consequences on linkage studies. Third, there has been an overlap in populations among linkage studies that were used for meta-analyses. Without all of the original data from each publication readily available, this was unavoidable in some cases. Finally, genetic and phenotypic heterogeneity were large contributing factors to the shortcomings of linkage studies. The complexity of schizophrenia does not make it an ideal candidate for linkage studies and a decade of research in this area has not proved otherwise.
Genome Wide Association Studies

Association studies are similar in scope to linkage studies in that the goal is to identify a genetic mutation that segregates with a disease, but different methodology is used to accomplish this goal. Genetic markers are identified in an affected case population and in a normal control population, based on the assumption that one of the markers will appear at a higher frequency in the cases when compared to the controls. The genetic markers found to be associated with the disease may be causative variants themselves, or will be in linkage disequilibrium (LD) with the causative variant. Association studies have an advantage over linkage studies in that they do not require large pedigrees, which can be problematic in diseases with reduced fecundity such as schizophrenia. Additionally, association studies can narrow down further the critical regions of the genome that were initially identified as candidate regions via linkage studies.

Initially, association studies were limited in scope to the available technology for genotyping markers in the genome. Over the last ten years, however, high throughput SNP genotyping methods have allowed researchers to interrogate hundreds of thousands of SNPs simultaneously. The ability to perform high throughput SNP genotyping led to an abundance of genome-wide association studies (GWAS). Initial results from GWAS of complex diseases were very promising. For example, GWAS in age-related macular degeneration identified a common Tyr401His polymorphism in the complement factor H gene associated with a three-fold increase in risk for developing the disease (Klein et al. 2005; Grassi et al. 2007). For Crohn’s disease, GWAS identified common polymorphisms in \textit{NOD2} on chromosome 16q12 in several independent patient groups (Ogura et al. 2001; Hugot et al. 2003; Barrett et al. 2008).
With encouraging results from initial GWAS studies in other complex disorders, it was hoped that GWAS in psychiatric disorders such as schizophrenia, autism, bipolar disorder, attention-deficit hyperactivity disorder (ADHD) and major depressive disorder would begin to fill in the many gaps left from linkage studies. The first GWAS on schizophrenia was published in 2007 and simultaneously examined ~500,000 markers and found a strong effect near the gene CSFR2A \((P = 3.7 \times 10^{-7})\) (Lencz et al. 2007). This study went on to sequence the CSFR2A and its neighboring gene IL3RA and identified common intronic haplotypes and several novel, rare missense variants (Lencz et al. 2007). Some examples of other GWAS findings with significant association with schizophrenia are APBA2, CCDC60, NRXN1, RELN (in females), ROBO2, ROBO1 and ZNF804A (O'Donovan et al. 2008; Shifman et al. 2008; Kirov et al. 2009c; Need et al. 2009; Potkin et al. 2009). To date, the National Human Genome Research Institute catalog of published GWAS contains 29 schizophrenia studies that report over 122 statistically significant associated SNPs \((p\text{-value} < 1.0 \times 10^{-5})\) (Hindorff 2009).

One of the criticisms of GWAS of schizophrenia is that sample sizes are inadequate. It is believed that larger sample sizes, in the range of hundreds of thousands of individuals or more, are needed for better power due to small effect sizes of the common variants being interrogated. The PGC has been formed to address this area of opportunity for the study of complex psychiatric disease. It is the largest group of investigators, with over 200 participants, with a goal of conducting meta-analyses of GWAS data for psychiatric disorders including schizophrenia. The PGC has made it possible to have an unprecedented sample size in its first publication with a stage 1 discovery sample of 21,856 individuals and a stage 2 replication sample of 29,839 independent subjects (Schizophrenia Psychiatric Genome-Wide Association Study 2011). After analyses of both stages genome-wide significant associations for schizophrenia were reported at seven loci. Five of the loci were new: 1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33; and two were previously reported loci: 6p21.32-p22.1 and 18q21.2
(Schizophrenia Psychiatric Genome-Wide Association Study 2011). The strongest association reported was a new variant that was located within an intron of a microRNA, MIR137. Interestingly, four other genome-wide significant schizophrenia loci identified were predicted to be targets of MIR137 (Schizophrenia Psychiatric Genome-Wide Association Study 2011).

Though the utility of GWAS studies in schizophrenia has not lessened over time, some criticisms exist. For example, why do we only have moderately significant results with an inconsistency among the top few associated regions? One approach under debate is the consideration of a wider threshold of significance for GWAS signals, meaning that true signals are likely buried amongst false-positive findings. It has also been proposed that better array technology will improve the power of GWAS. In the past, probe design has been based on which SNP probe works best, rather than careful selection of probes as not to choose SNPs that tag overlapping genomic regions. This is already being addressed as certain manufacturers of microarrays have started to become more careful and thoughtful when choosing probes with respect to LD block. That being true, it is still possible for multiple causal variants that are not identical by descent to lie within the same haplotype block, causing true variants to remain undetected with current array technology. Another relatively simple explanation is that, until now, we have primarily focused on SNPs in coding regions of the genome while the majority (> 80%) of associated variants identified in GWAS are located outside of coding regions (Hindorff et al. 2009). Finally, phenotypic heterogeneity remains a major limiting factor in our understanding of the pathophysiology of schizophrenia. Two recent GWAS reports have identified overlapping SNPs between schizophrenia, autism spectrum disorder (ASD), ADHD, bipolar disorder, and major depressive disorder (Lee et al. 2013; Psychiatric Genomics Consrotium 2013), pointing to the need for a more precise understanding of the phenotypic heterogeneity of schizophrenia.
Copy Number Variant Studies

The human genome is full of variation that varies in size from a single base-pair change to large chromosomal abnormalities. The advent of microarray technology made it possible to detect a set of intermediate size variants known as copy number variants (CNVs). Copy number variants were initially defined as chromosomal deletions or duplications ranging from 1 kb up to several Mb in size (Freeman et al. 2006). The definition of CNV size has evolved over time with the use of increasingly sensitive detection methodologies; deep sequencing technologies can currently detect CNVs as small as 31 bp (Nord et al. 2011).

It was once believed that CNVs were a rare source of variation, but high-throughput microarray technology has made it possible to interrogate the entire genome for them. Current estimates are that 13% of the human genome is copy number variable and that an individual harbors, on average, over 1000 CNVs covering approximately 4 Mb (Redon et al. 2006; Conrad et al. 2010; Mills et al. 2011), and a larger proportion of the human genome is affected by CNVs than it is by SNPs (Conrad et al. 2010). Results from CNV studies in control populations are all curated in the online Database of Genomic Variants (DGV), which is a valuable research tool. There has been a large-scale effort to identify and catalog CNVs in normal, healthy control populations (Iafrate et al. 2004; Sebat et al. 2004; Redon et al. 2006; Pinto et al. 2007; Kidd et al. 2008; Conrad et al. 2010). Pathogenic CNVs have been identified in neurodevelopmental disorders such as autism, intellectual disability (ID), epilepsy, ADHD, bipolar disorder and schizophrenia (Sebat et al. 2007; Mefford et al. 2008; Stefansson et al. 2008; Walsh et al. 2008; Helbig et al. 2009; Zhang et al. 2009; Malhotra et al. 2011; Priebe et al. 2012).

One of the earliest examples of a chromosomal deletion associated with schizophrenia risk, a deletion at 22q11, was first identified over ten years ago before the
routine use of microarrays to detect CNVs (Karayiorgou et al. 1995). A deletion at chromosome 22q11.2 is one of the strongest known risk factors for schizophrenia. This region is known by several names because many clinically distinct syndromes were identified and named before it became apparent that they all mapped to the same genomic location. Syndromes mapping to the 22q11 region include 22q11 deletion syndrome, Velo-Cardio-Facial syndrome (VCFS), DiGeorge syndrome, CATCH22 syndrome, and Spritzen syndrome (OMIM 611867). Specifically, up to one-third of deletion carriers go on to develop schizophrenia or schizoaffective disorder (Pulver et al. 1994; Murphy et al. 1999; Murphy & Owen 2001; Gothelf et al. 2007). There are several deletion sizes associated with increased schizophrenia risk, 90% of deletions are ~3 Mb and the remainder are ~1.5 or ~2 Mb in size (Tan et al. 2011; Hiroi et al. 2013). Over 40 genes are located within the deletion and there is still much uncertainty over which genes may be causative for schizophrenia and the exact role they would play. *TBX1, GNB1L, COMT, PRODH, DGCR8* and *ZDHHC8* are some of the most heavily studied genes in this region (Paylor et al. 2006; Williams et al. 2008; Meechan et al. 2009; Okochi et al. 2009). The 22q11.2 deletion is an example of the complexities underlying current CNV association studies. While we can readily identify genomic regions associated with an increased risk for schizophrenia, many times it is difficult to pinpoint which genes in the region are causative. This leaves open the possibility that disrupting the chromosomal architecture is disease causing. Another, more realistic theory is that not just one gene in the region is responsible for schizophrenia risk, but that the disturbance of multiple genes in this region act in concert to confer disease risk.

In the first published genome-wide CNV analysis of schizophrenia, Walsh et al. (2008) examined 150 case subjects and 268 controls for CNVs over 100 kb and reported that novel genic CNVs are present in 15% of cases and 20% of childhood-onset cases (diagnosis before age 18) compared to 5% of controls. They also reported that CNVs in cases disproportionately disrupted genes from signaling networks that control
neurodevelopment, including neuregulin and glutamate pathways (Walsh et al. 2008). The authors proposed that rare CNVs (> 100 kb, genic and novel to DGV) contribute to schizophrenia risk, making a case for the common-disease, rare-variant (CDRV) hypothesis.

Multiple recurrent CNVs have also been associated with schizophrenia. From a cohort of schizophrenia patients, 9878 parent-to-offspring transmissions were examined and 66 de novo CNVs were identified and tested for association (Stefansson et al. 2008). Three recurrent deletions at 1q21.1, 15q11.2 and 15q13.3 showed nominal association with cases that had schizophrenia and psychosis in a combined sample using 1433 cases and 33,250 controls as a discovery cohort and 3285 cases and 7951 controls as a replication cohort (Stefansson et al. 2008). The International Schizophrenia Consortium simultaneously released results from a genome-wide survey of rare CNVs in 3391 cases and 3181 controls. It was reported that CNVs with a sample frequency less than 1% and over 100 kb in size had a 1.15-fold increased burden in cases compared to controls and (International Schizophrenia 2008). An association was also replicated for recurrent deletions at chromosome 1q21.1, 15q13.3 and 22q11.2 (International Schizophrenia 2008).

In 2010 Glessner et al. performed a genome-wide CNV analysis using a discovery cohort of 977 schizophrenia cases and 2000 healthy controls, and validated findings in an independent cohort of 758 cases and 1485 controls (Glessner et al. 2010). They reported an enrichment of CNVs that contain genes belonging to the Gene Ontology family affecting synaptic transmission (Glessner et al. 2010). Deletions and duplications were reported for two of the candidate genes found in this study, CACNA1B and DOC2, which are calcium-signaling genes responsible for neural excitation. In addition, two ras-related genes that are important for neural crest development, RET and RIT2, were also reported as being significantly enriched for deletions (Glessner et al. 2010).
Some common themes have emerged from the many genome-wide CNV analyses of schizophrenia. First, it is now well-established that there is an enrichment of rare (<1% MAF), large (>100 kb) CNVs in schizophrenia (International Schizophrenia 2008; Walsh et al. 2008; Kirov et al. 2009a; Cooper et al. 2011). A second finding consistently reported in the literature is an enrichment of de novo CNVs in schizophrenia patients (Xu et al. 2008; Malhotra et al. 2011; Kirov et al. 2012). It has been more difficult to reach a consensus on other points, with some studies reporting contradictory information. For example, some studies report an excess of CNVs in schizophrenia patients compared to controls (International Schizophrenia 2008), while other studies do not observe a difference in CNV burden between cases and controls (Levinson et al. 2011). It has also been reported that schizophrenia patients harbor more genic CNVs (International Schizophrenia 2008; Walsh et al. 2008) or more deletions than controls (Buizer-Voskamp et al. 2011), however, these findings remain to be positively confirmed.

To date, no individual pathological variants stand out as an individual risk variant for schizophrenia. The 22q11.2 deletion is the strongest genetic predictor of schizophrenia risk, yet it only accounts for 1-2% of schizophrenia cases (Karayiorgou et al. 1995; International Schizophrenia 2008; Stefansson et al. 2008; Xu et al. 2008; Kirov et al. 2009a; Glessner et al. 2010; Buizer-Voskamp et al. 2011; Priebe et al. 2013). The wealth of CNV and GWAS data available has helped to identify potential pathways or specific systems that may be involved in schizophrenia etiology such as neurodevelopment, synaptic dysfunction and altered neural connectivity (International Schizophrenia 2008; Walsh et al. 2008; Tam et al. 2009; Kirov et al. 2012; Van Den Bossche et al. 2012). There has also been a set of CNVs identified that exhibit high degrees of statistical confidence as risk factors for schizophrenia such as deletions at 1q21.1, NRXN1, 3q29, 15q11.2, 15q13.3 and duplications at VIPR2, 16p11.2, 16p13.1 and 15p11-q13 (International Schizophrenia 2008; Stefansson et al. 2008; Kirov et al. 2009b; McCarthy et al. 2009; Sebat et al. 2009; Moreno-De-Luca et al. 2010; Mulle et al.
However, many of these CNVs exhibit high pleiotropy and are also implicated in other neurodevelopmental disorders, calling into question the exact role that they play in schizophrenia etiology.

Background of Recurrent CNV at Chromosome 15q11.2

BP1-2

Recurrent CNVs

Rare recurrent CNVs are most often formed through a mechanism known as non-allelic homologous recombination (NAHR). Small regions of homology cause misalignment during meiosis I that lead to unequal crossing over resulting in the deletion or duplication of intervening sequence (Christian et al. 1999). Segmental duplications (segdups), also known as low copy repeats or duplicons, are stretches of DNA with over 90% sequence homology that facilitate NAHR and comprise ~2-6% of the human genome (Stankiewicz & Lupski 2002; Egan et al. 2007).

Human chromosomes 7, 15, 16, 17 and 22 are enriched for segdups (Marques-Bonet et al. 2009) and recurrent CNVs at those loci (15q11.2 deletion, 16p11.2 deletion/duplication, 17q12 deletion and the 22q11.2 deletion) have consistently been associated with several psychiatric disorders including schizophrenia, autism, ID and epilepsy (Bassett et al. 2003; Stefansson et al. 2008; Walsh et al. 2008; Tam et al. 2009; Moreno-De-Luca et al. 2010). This raises the question, what is it about recurrent CNVs that predispose them to exhibit such high pleiotropy in neurodevelopmental disorders? One hypothesis is that the promiscuity of these recurrent CNVs can be attributed to the particular network of CNVs they exist with on an individual basis. The recurrent CNV at
15q11.2 will be used as an example to explore this possibility and is a highlight of this thesis.

Association with Prader-Willi Syndrome and Angelman Syndrome

Prader-Willi syndrome (PWS) (OMIM 176270) and Angelman syndrome (AS) (OMIM 105830) are clinical disorders caused by aberrant inheritance of the genomic region 15q11-q13 known as the PWS/AS critical region. Characteristics of PWS include: hypotonia, respiratory distress, failure to thrive in the postnatal period, hyperphagia in early adulthood resulting in obesity, short stature, small hands and feet, hypogonadism, mild-to-moderate ID, temper tantrums and obsessive-compulsive mannerisms (Butler 1990). Characteristics common to AS are: developmental delay, severe ID, lack of speech, movement ataxia, hyperactivity, seizures, aggressive behavior and excessive inappropriate laughter (Jiang et al. 1999). It is important to note that both PWS and AS exhibit a high comorbidity with autism (Nakatani et al. 2009; van der Zwaag et al. 2010). The maternal copy of the PWS critical region is imprinted and loss of the paternally derived alleles leads to PWS. Similarly, the paternal copy of the AS critical region is imprinted and loss of the maternally derived alleles results in AS (Knoll et al. 1989). Parent-of-origin deletions account for the majority (65-75%) of all PWS and AS cases (Nicholls & Knepper 2001). Other mechanisms leading to PWS and AS include uniparental disomy, which accounts for 20-30% of cases, and imprinting defects which occur in 2-5% of cases (Nicholls & Knepper 2001).

Chromosome 15, specifically the long arm, is one of seven chromosomes enriched for segdups (Bailey et al. 2002), which predispose the region to NAHR. Thus, segdups in the 15q11-q13 region coincide with recurrent CNV breakpoints. Specifically,
there are five recognized breakpoints in this region, 1-5, with the region between
breakpoint 1 and breakpoint 2 (BP1-2) not being imprinted (Chai et al. 2003). The 2 Mb
region between BP2-3 is imprinted and leads to the parent-of-origin effect observed with
deletions that result in PWS and AS (Buiting et al. 1995; Nicholls & Knepper 2001). As
previously mentioned, parent-of-origin deletions are responsible for PWS and AS 65-
75% of the time and can be divided into two sub-groups: type I deletions spanning BP1-
3, which occur ~40% of the time; and the smaller type II deletions spanning BP2-3,
which occur ~60% of the time (Figure 1-1) (Knoll et al. 1990; Christian et al. 1995; Hou
et al. 2011). Copy number variants of chromosome 15q11.2 BP1-2 first came to the
attention of researchers with the discovery that type I deletion patients have a more
severe phenotype than either type II deletion patients or uniparental disomy patients, for
both PWS and AS (Varela et al. 2004; Sahoo et al. 2006). A study comparing PWS
patients with type I deletions to those with type II deletions indicated that patients with
type I deletions generally exhibited poorer adaptive behavior, obsessive-compulsive
behavior, reading, math and visual-motor integration than patients with type II deletions
(or maternal uniparental disomy) (Butler et al. 2004). Another study reported that PWS
patients with type I deletions had significantly higher Reiss (physical) depression scores
than type II patients (Hartley et al. 2005). Additionally, PWS patients with the type I
deletion are more likely to meet autism criteria (Doornbos et al. 2009). Phenotypic
differences noted between type I and type II AS patients include a higher likelihood of
meeting autism criteria, lower cognitive scores, lower expressive language scores,
absence of vocalization and delay in ability to sit without support (Varela et al. 2004;
Sahoo et al. 2006). There are contradictory reports that did not find a significant
difference in phenotype between type I and type II PWS patients (Milner et al. 2005;
Varela et al. 2005; Dykens & Roof 2008). One of the studies that did not identify
behavioral differences between type I and type II deletion patients did, however, find that
type I PWS patients exhibit an association between age and lower adaptive skills, externalizing symptoms and more severe behavioral problems (Dykens & Roof 2008).

**Association with Neuropsychiatric Disorders**

There is growing evidence that, aside from the role it plays in PWS and AS, the BP1-2 region also plays a critical role in the etiology of other neurodevelopmental disorders such as schizophrenia, autism, epilepsy, ID and developmental delay (DD). The first report of the BP1-2 deletion, separate from PWS and AS, was a case report by Murthy et al. (2007). The patient was a 3.5-year-old boy suspected of having AS whose parents were second cousins. He exhibited delayed speech, delayed motor development, ID, cleft palate, ADHD and a constant unusual happy expression. A paternally inherited 253 kb deletion of 15q11.2 BP1-2 was identified by chromosomal analysis. His father, who also had an unusual happy expression, was slow to understand the details of the study and had a personal and familial history of DD (Murthy et al. 2007).

The second study to report the 15q11.2 BP1-2 deletion by Doornbos et al. (2009) investigated 1576 pediatric patients who were referred for ID and/or multiple congenital abnormalities (CA). Nine patients were reported to share the BP1-2 deletion (2 de novo, 3 maternal inheritance and 4 paternal inheritance) and exhibited a wide variety of phenotypes including speech delay, motor delay, cleft or narrow palate, ear abnormalities and slender fingers. Of note, a subset of the nine patients had ASD, ADHD, obsessive compulsive disorder (OCD) and/or an unusual happy expression (Doornbos et al. 2009).

In 2010 von der Lippe et al. reported seven new patients (two brothers) belonging to six different families that shared a 350 kb deletion of 15q11.2 BP1-2. All of the patients exhibited some degree of learning difficulties, DD, behavioral problems and all patients inherited the deletion from a mildly affected parent (von der Lippe et al. 2011).
Some of the dysmorphic features reported in patients were small stature, small face, irregular teeth, micrognathia, narrow nose, reduced distance between eyes and juvenile rheumatoid arthritis.

A report by van der Zwaag et al. (2010) highlighted a Dutch pedigree where a 471 kb duplication of 15q11.2 BP1-2 co-segregated with ASD in a family. The female proband met criteria for an autism diagnosis with a full-scale IQ (FSIQ) = 83, verbal IQ (VIQ) = 89, and performance IQ (PIQ) = 81, and inherited the duplication from her father, who has a diagnosis of Asperger syndrome. A previous case report about the proband was written by a neurologist and indicated that at the age of 9-years-old she began displaying general DD, limited speech production, clumsiness and a tendency to taste and lick many different objects. A half sister of the proband, who shared a father with the proband, also harbored the BP1-2 duplication and was highly suspected of having ASD, but passed away before the clinical evaluation could be completed. While this was a family-based study, the researchers used a control set of 945 ethnically matched individuals and identified 4 individuals with a 15q11.2 BP1-2 duplication (van der Zwaag et al. 2010). Due to the presence of the duplication in 4 control individuals, they were unable to report a statistical difference between the rate of the duplication in their family and in controls.

Burnside et al. (2011) published a case-only retrospective analysis of 17,000 pediatric patients referred for chromosomal microarray analysis (CMA). Of the 17,000 patients examined, 69 patients had a deletion and 77 had a duplication of the 15q11.2 BP1-2 region (Burnside et al. 2011). Clinical information was available for a subset of CNV patients (56 deletion individuals and 49 duplication individuals), and shared phenotypes reported were DD, motor delay, speech delay, and neurological and/or behavior issues, specifically ASD, ADHD and tantrums (Burnside et al. 2011).

Abdelmoity et al. (2012) also performed a case-only retrospective chart review on 1654 pediatric patients referred for CMA for DD, ID, neurological and/or
neuropsychiatric disorders, dysmorphism, microcephaly or multiple CA. They reported 26 individuals with a 15q11.2 BP1-2 CNV, 21 (1.27%) carried deletions and 12 (0.7%) carried duplications. Phenotypes associated with the deletion patients were DD, ID, ASD, learning disability, inattention and hyperactivity. Dysmorphic features ranging from mild alar flaring to severe microcephaly with cleft lip and palate were reported as well as other phenotypes such as hypotonia and skeletal abnormalities. Phenotypes associated with duplication patients were DD, autism, social withdrawal, aggressive behavior and focal epilepsy. Dysmorphic features reported ranged from abnormally shaped ears to craniosynostosis and long palpebral fissures (Abdelmoity et al. 2012). For a summary of all the previously mentioned 15q11.2 BP1-2 studies, please see Table 1-1.

There have been several association studies that report a nominal association of the 15q11.2 BP1-2 deletion with schizophrenia (Stefansson et al. 2008; Kirov et al. 2009a; Magri et al. 2010; Crespi & Crofts 2012; Kirov et al. 2012; Van Den Bossche et al. 2012). One study specifically examined the relationship between the 15q11.2 BP1-2 deletion and schizophrenia in a Chinese Han population and reported association with both CNVs and SNPs (Zhao et al. 2013). Deletions were identified in 12/2058 (0.58%) cases compared to 6/3275 (0.18%) controls as well as two common SNPs in the BP1-2 region.

The 15q11.2 CNV has also been identified in association studies of patients with autism, DD, ID or CA (Mefford et al. 2009; Cooper et al. 2011; De Wolf et al. 2013; Rosenfeld et al. 2013; Wisniowiecka-Kowalnik et al. 2013). Of note, the 15q11.2 BP1-2 deletion has also been associated with epilepsy (de Kovel et al. 2010; Mefford et al. 2010). In addition epilepsy has been reported in BP1-2 CNV patients or in their first-degree relatives (Doornbos et al. 2009; von der Lippe et al. 2011; Abdelmoity et al. 2012).

The 15q11.2 BP1-2 region is nested in the larger chromosome 15q11-q13 region which, as previously mentioned, has a long standing association with neuropsychiatric
disorders. The 15q11-q13 maternal duplication is one of the strongest recognized risk factors for autism (Cook et al. 1997; Moeschler et al. 2002; Szatmari et al. 2007; Marshall et al. 2008; Glessner et al. 2009). Duplication of chromosome 15q11-q13 is also recognized as a risk factor for schizophrenia (Ingason et al. 2011a; Boot et al. 2012; Liao et al. 2012), epilepsy (Michelson et al. 2011; Stewart et al. 2011), and general DD and abnormal behavior (Browne et al. 1997; Roberts et al. 2002; Thomas et al. 2003).

15q11.2 BP1-2 Gene Function and Expression

The BP1-2 region spans approximately 500 kb and there are four highly conserved genes located in this region: CYFIP1, NIPA1, NIPA2 and GCP5 (aka TUGCP5) (Burnside et al. 2011). The gene GCP5 is 47 kb in size and is the least documented of the four genes in the BP1-2 region. GCP5 encodes a member of the γ-tubulin complex, a protein that is thought to be involved in axonal migration during brain development (Murphy et al. 2001; Chai et al. 2003). GCP5 is expressed ubiquitously with the highest expression levels in the subthalamic nuclei of the brain (Nagase et al. 2001). The nuclei of the brain are implicated in ADHD and obsessive compulsive behavior (Francois et al. 2004).

NIPA1 and NIPA2 have previous associations with and specific expression patterns in the central nervous system (CNS) and brain (Chai et al. 2003; Zhao et al. 2008; Doornbos et al. 2009). NIPA1 and NIPA2 are paralogous gene members of an ancient gene family and are believed to have arisen phylogenetically 450-600 million years ago (Chai et al. 2003). The two genes have different structures and expression patterns, with NIPA1 exhibiting higher levels of brain and CNS-specific expression (Chai et al. 2003). NIPA1 and NIPA2 are 41kb and 29kb, respectively. Little is known about NIPA2 other than that it, along with NIPA1, is predicted to be a magnesium transporter.
that acts in a dose-dependent manner (Goytain et al. 2007; Zhao et al. 2008). More effort has gone into understanding the role of NIPA1, an inhibitor of bone morphogenic protein, which is important in axonal outgrowth (Tsang et al. 2009; Botzolakis et al. 2011). Nipa1 is expressed in the rodent (mouse and rat) frontal cortex, cerebellum and striatum of the brain in addition to the spinal cord (Zhao et al. 2008). There is strong evidence that a dominant negative mutation in NIPA1 is responsible for hereditary spastic paraplegia, which is a disorder characterized by progressive neural degeneration and proximal motor deficits (Rainier et al. 2003; Chen et al. 2005; Zhao et al. 2008).

CYFIP1 is highly expressed in the brain and there are reports of it playing a role in the stability of neuronal structures. First, it interacts with Rac1, a small GTPase that is implicated in the development and maintenance of neuronal structures (Schenck et al. 2003). CYFIP1 also interacts with FMRP (fragile-X mental retardation protein), a known regulator of neuronal development, and their combined interaction is implicated in neuronal growth, guidance, branching and cap-dependent translation repression (Schenck et al. 2001; Chai et al. 2003; Napoli et al. 2008). Much of the protein synthesis that takes place during neurogenesis is cap-dependent and requires assembly of the eIF4F translation initiation complex. One of the eIF4F complex proteins, eIF4E, is also a direct binding partner of CYFIP1, and the CYFIP1-FMRP-eIF4E complex is present at the neural synapse (Napoli et al. 2008). Thus, a pathway is predicted where CYFIP1 and FMRP are simultaneously recruited to capped mRNA and to eIF4E to block translation initiation. Upon neuronal stimulation, CYFIP1-FMRP is released from the mRNA, making eIF4E available for joining the eIF4F complex and translation initiation takes place (Napoli et al. 2008).

In the previously mentioned study of autism in a Dutch pedigree, where three family members harbored a 15q11.2 BP1-2 duplication, transcript levels in peripheral blood samples were investigated and all four genes were upregulated by more than 70% (van der Zwaag et al.). In the same study, the investigators performed RNA in situ
hybridization on mouse embryonic and adult brain sections and reported high expression levels of \textit{Cyfip1} and \textit{Nipa1} in the developing mouse brain (van der Zwaag et al.).

\textbf{Polygenic Model of Schizophrenia}

To date, research in schizophrenia has shed light on the fact that neurodevelopment, synaptic dysfunction and altered neural connectivity are important to pathogenesis of the disease. However, no specific pathophysiology has been discovered nor are we closer to developing specific clinical tests or targeted treatment options for patients. A polygenic model for schizophrenia was first proposed by Gottesman and Shields (1967). The hypothesis states that schizophrenia liability is based on the summation of multiple mutations with different genetic loads that, when combined, cross a disease threshold. At the time there was no way to definitively test this hypothesis and as genetic research began to make rapid advances after the completion of the Human Genome Project in 2003, optimism was high that a group of singular mutations would be discovered that underlie schizophrenia etiology. This was not to be the case and now the tides are turning back towards the original hypothesis made over 40 years ago.

Early during the genesis of GWAS studies, the common-disease/common-variant (CDCV) hypothesis that thousands of common variants, SNPs with minor allele frequency (MAF) > 1-5\%, of small to moderate effect size account for the underlying genetic liability of schizophrenia was popular (Chakravarti 1999; Reich & Lander 2001; Lohmueller et al. 2003; Peng & Kimmel 2007). Because SNPs are by definition, common variation, it was believed that much of the genetic burden of schizophrenia would be captured by SNP microarrays. However, this was not the case and the totality of significant SNP findings could not account for all of the heritability of schizophrenia. In answer to this “missing heritability” an alternative model was proposed, the CDRV
hypothesis which asserts that rare variants of large effect size underlie disease susceptibility (Pritchard & Cox 2002; McClellan et al. 2007). Detection of rare SNP variants is ongoing as deep sequencing technology becomes more accessible and affordable (Awadalla et al. 2010; Xu et al. 2011; Hu et al. 2013; Jouan et al. 2013; Timms et al. 2013). In addition to rare SNPs, the identification of genomic CNVs has also bolstered the CDRV hypothesis. Again, individual rare SNP and CNV findings are unable to account for the unresolved heritability estimates for schizophrenia. While it is possible that traditional broad-sense heritability estimates are inflated because of the inability to tease apart shared environmental factors in familial studies, it is unlikely that this will account for a large proportion of heritability.

While the CDCV and CDRV hypotheses are not mutually exclusive, they represent different models for study design to identify disease causing variants. In reality, it is likely that the underlying genetic architecture of schizophrenia is made up of a combination of both common and rare variants. The ability to search for both types of variation at the same time is complex and unprecedented. Therefore, as a starting point some researchers have begun to propose a two-hit or a multiple-hit CNV model where more than one CNV is proposed to be causative for neurodevelopmental disorder. A two-hit model has been proposed for several recurrent deletions, one of them the recurrent deletion at chromosome 16q12.1 where individuals with DD were more likely than controls to carry an additional large (> 500 kb) CNV (Girirajan et al. 2010; Kumar 2010).

The recurrent deletion at 22q11.2 has also been a variant of much speculation in schizophrenia risk, and it has long been questioned why only a subset of patients go on to develop schizophrenia. One report by Williams et al. (2013) investigated VCFS patients with and without psychosis for the presence of additional CNVs. Criteria for the “second-hit” was that the CNV be greater than 100 kb in size and overlap a gene having a previous association with a neuropsychiatric disorder. In the non-psychotic VCFS patients, only 1/10 had an additional genic CNVs that met the criteria for a second hit,
while 4/13 individuals in the psychotic group had second-hit CNVs (Williams et al. 2013). An independent sample of 17 VCFS patients with schizophrenia were used to replicate second-hit findings, and 16 individuals harbored additional CNVs, though only three met second-hit criteria. This brought the combined total of 7 individuals (27%) with potential second hits in the VCFS psychotic/schizophrenia group compared to 1 individual (10%) of the non-psychotic group (Williams et al. 2013). A study by Manolakos et al. (2011) reported a 6-year-old boy with growth retardation, dysmorphic features and learning difficulties who harbored a combined 3.2 Mb deletion of 22q11.1-q11.21 deletion and a 12 Mb duplication of 15q11.2-q13.3. The boy inherited both structural abnormalities from his mother through a balanced translocation between chromosome 15 and chromosome 22 (Manolakos et al. 2011).

Leblond et al. (2012) identified three ASD patients with de novo deletions of the candidate gene SHANK2 who also harbored inherited genetic imbalances at chromosome 15q11-q13 (two duplications of the BP5 region and one deletion of the 15q11.2 BP1-2) (Leblond et al. 2012). Another study by Madrigal et al. (2012) reported a family where two siblings were reported to have ID and dysmorphic features. One sibling, a boy, had a combined maternally inherited 15q11.2 BP1-2 deletion and an FMR1 premutation allele while his sister, from a different father, only carried the 15q11.2 BP1-2 deletion (Madrigal et al. 2012). The mother of the siblings carried the same variants as the son and was reported to have learning and behavioral problems as well. The 15q11.2 BP1-2 deletion was also identified in three other maternal relatives, two affected with learning disabilities and one healthy individual (Madrigal et al. 2012). In the milder affected relatives, as well as the sister, there was no further investigation as to whether or not they carried additional CNVs.

Recurrent CNVs are risk variants that have been under much speculation because they exhibit high pleiotropy, reduced penetrance and variable expressivity. Initial reports of a two-hit or multiple-hit CNV studies consistently implicate recurrent CNVs in this
type of model. Therefore, it is easy to speculate that discrepancies in pleiotropy, penetrance and expressivity can be contributed to a combination of CNVs that involve recurrent CNVs, rather than the recurrent CNVs themselves. The 15q11.2 BP1-2 CNV is ideal for pursuing this theme of investigation as it has been shown to be statistically enriched for co-occurring with other CNVs in a cohort of children with ID and CA (Girirajan et al. 2012). Additionally, many of the early reports of the 15q11.2 BP1-2 deletion mention the presence of additional CNVs in the BP1-2 subjects without further speculation about their presence (Doornbos et al. 2009; Burnside et al. 2011; von der Lippe et al. 2011; Abdelmoity et al. 2012).
Table 1-1. Summary of studies specifically studying 15q11.2 BP1-2 deletion

<table>
<thead>
<tr>
<th>First Author</th>
<th>Cases</th>
<th>Controls</th>
<th>Developmental/Psychiatric Phenotype</th>
<th>Dysmorphic or Other Phenotypes</th>
<th>Results</th>
<th>Additional CNVs Noted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murthy et al. 2007</td>
<td>n/a</td>
<td>n/a</td>
<td>Case report - 3.5 y/o boy, referred with suspicion of AS and presenting with ID, neurological disorder, DD, &amp; speech impairment</td>
<td>Muscular hypotonia &amp; brisk bilateral deep tendon reflexes, typical happy appearance w/ drooling and constant smiling</td>
<td>Deletion in proband and father (who presented with similar, but milder clinical features)</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Doornbos et al. 2009</td>
<td>1576 pediatric patients</td>
<td>350 healthy parents</td>
<td>Delayed motor development, delayed speech development, behavioral problems, ADHD, autism, obsessive-compulsive behavior, unusual happy expression</td>
<td>Hypertelorism, a cleft or narrow palate, ear abnormalities &amp; slender fingers</td>
<td>Deletion in 9 cases (0.57%) (2 \textit{de novo}, 3 maternal inheritance, 4 paternal inheritance)</td>
<td>3 patients</td>
</tr>
<tr>
<td>Study</td>
<td>Families/Subjects</td>
<td>Controls</td>
<td>Clinical Features</td>
<td>Molecular Features</td>
<td>Genetic Results</td>
<td>Other Notes</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>von der Lippe et al. 2010</td>
<td>7 families</td>
<td>n/a</td>
<td>MR, DD, behavioral problems, paranoid psychosis, &quot;phobic anxiety reactions&quot;</td>
<td>Small stature, small face, irregular teeth, micrognathia, narrow nose, reduced distance between eyes, and juvenile rheumatoid arthritis</td>
<td>Deletion in 7 patients (2 brothers) (6 inherited, 1 uncertain transmission)</td>
<td>1 patient</td>
</tr>
<tr>
<td>van der Zwaag et al. 2010</td>
<td>1 family w/ 7 members</td>
<td>945 adult healthy controls</td>
<td>Female proband diagnosed with ASD, father diagnosed with Asperger syndrome, half-sister of proband highly suspected to have ASD (deceased before testing possible)</td>
<td>None</td>
<td>Duplication in 3 individuals: female proband, her father (Asperger syndrome) and half-sister (w/ suspected ASD)</td>
<td>Proband who had total of 6 CNVs</td>
</tr>
<tr>
<td>Burnside et al. 2011</td>
<td>17,000 pediatric patients</td>
<td>Case only retrospective analysis</td>
<td>Developmental, motor &amp; speech delays and neurological and/or behavior issues. Autism/ASD, ADHD and tantrums. Gait or coordination problems reported, hypotonia and ID.</td>
<td>CA noted in 1/2 of subjects w/ microcephaly the most prevalent</td>
<td>Deletion in 69 patients (0.41%) (2 de novo, rest inherited) and duplications in 77 (0.45%) patients</td>
<td>17 deletion patients &amp; 16 duplication patients</td>
</tr>
</tbody>
</table>
Table 1-1 — Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Number</th>
<th>Group Description</th>
<th>Disease/Psychiatric Conditions</th>
<th>Clinical Findings</th>
<th>Deletion in 12 cases &amp; 6 controls (0.58%) &amp; 6 controls (0.18%)</th>
<th>Not mentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhao et al. 2012</td>
<td>2058</td>
<td>3275 healthy adults</td>
<td>Schizophrenia</td>
<td>None</td>
<td>Deletion in 12 cases &amp; 6 controls (0.58%) &amp; 6 controls (0.18%)</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Abdelmoity et al. 2012</td>
<td>1654</td>
<td>Case only retrospective analysis</td>
<td>DD, ID, neuropsychiatric disorders, ASD, learning disability, inattention, focal epilepsy, social &amp; aggressive behaviors</td>
<td>Hypotonia, skeletal abnormalities, cleft lip &amp; palate, microcephaly, abnormally shaped ears, craniosynostosis, long slender fingers, &amp; syndactyly</td>
<td>Deletion in 21 cases (1.27%) &amp; duplication in 12 cases (0.7%)</td>
<td>5 deletion patients &amp; 2 duplication patients</td>
</tr>
</tbody>
</table>
Figure 1-1. Ideogram of chromosome 15q11.2 representing region between BP1-3.
CHAPTER II
A GENOME-WIDE CNV ANALYSIS OF SCHIZOPHRENIA
REVEALS A POTENTIAL ROLE FOR A MULTIPLE-HIT MODEL

Introduction

Schizophrenia (OMIM 181500) is an incurable and severe psychiatric disorder with a worldwide prevalence of approximately 1% (Gottesman & Shields 1967). In developed countries, schizophrenia is one of the leading causes of early death and years of life lost due to disability (Lewis & Gonzalez-Burgos 2006; Cichon et al. 2009). Adoption and twin studies show that first degree relatives of schizophrenia patients have an elevated risk of developing the disorder themselves (Heston 1966; Connor & Akbarian 2008). Heritability estimates for schizophrenia range from 64%-80% (Sullivan et al. 2003; Lichtenstein et al. 2009).

The search for variants underlying this susceptibility has been guided by multiple genetic architecture theories. The CDCV hypothesis for complex diseases like schizophrenia is based on the theory that many common variants (population frequency > 1-5%) of small to moderate effect account for genetic liability (Reich & Lander 2001; Lohmueller et al. 2003; Peng & Kimmel 2007). To date, in the National Human Genome Research Institute catalog of published GWAS there are 29 published schizophrenia studies that report over 122 statistically significant associated SNPs (p-value < 1.0 x 10^{-5}) (Hindorff 2009). Two more recent reports from the PGC have estimated that the amount of variation in schizophrenia liability that can currently be attributed to SNPs is 23% for Europeans subjects of mixed ancestry (Lee et al. 2012; Lee et al. 2013) and 32% for subjects with Swedish ancestry (Ripke et al. 2013), both of which implicate a high proportion of common SNPs. Nonetheless, it remains that a substantial proportion of heritability must be accounted for by other types of variants.
The CDRV hypothesis is a second genetic architecture model that complements CDCV by asserting that rare variants of large effect underlie some proportion of disease susceptibility (Pritchard & Cox 2002; McClellan et al. 2007). In addition to rare SNPs, the identification of genomic CNVs has bolstered the CDRV hypothesis for schizophrenia research (Walsh et al. 2003; International Schizophrenia 2008; Tam et al. 2010; Van Den Bossche et al. 2012). A CNV is operationally defined as any contiguous segment of DNA where sequence is duplicated or deleted.

Due to their relatively recent identification, however, most studies of CNVs in schizophrenia have focused on detecting large CNVs that independently produce a substantial increase in disease risk. We sought to advance this line of research in two directions: examining CNVs across a broad size spectrum and searching for the presence of multiple disease-susceptibility CNVs within the individuals in our study sample. We performed a genome-wide CNV analysis of 166 schizophrenia patients and 52 psychiatrically healthy controls. We filtered down to a set of conservative CNVs most likely to be disease causing and examined differences in CNV patterns between cases and controls.

**Materials and Methods**

**Study Sample**

The Iowa sample set was recruited primarily from Eastern Iowa while also including patients from the rest of Iowa and neighboring states. Patients have provided written informed consent to a protocol approved by the University of Iowa IRB. This research has been conducted in accordance with the Declaration of Helsinki. Upon ascertainment, subject information was gathered using a semi-structured interview
instrument, the Comprehensive Assessment of Symptoms and History (CASH), created by our research group (Andreasen 1983, 1984; Andreasen 1985; Andreasen et al. 1992). This study included individuals with schizophrenia and psychiatrically normal controls who had undergone neuropsychological testing, brain imaging and clinical assessment, and who had provided blood samples for DNA extraction. Inclusion criteria for affected individuals were a Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) diagnosis of schizophrenia and age at intake of 16-45 years of age. Exclusion criteria included the presence of a neurological disorder, history of severe head trauma, or active substance dependence or abuse (except for nicotine). Control exclusion criteria consisted of a current or past psychiatric or neurological disorder, history of severe head trauma, active substance dependence or abuse (except for nicotine), ID or history of a psychotic disorder in a first-degree relative.

CNV Detection

Subject DNA was extracted from whole blood samples using a standard salt precipitation technique and run on Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA) according to recommendations from the manufacturer. Raw intensity data from the arrays were analyzed for copy number changes using the publicly available software program Affymetrix Genotyping Console 4.0 (Affymetrix). Data quality were assessed with both contrast quality control (CQC), which measures the quality of probe intensity data, and median absolute pairwise difference (MAPD), which provides a per-chip estimate of variability. Samples with a CQC < 0.4 (or a batch mean CQC < 1.7) were excluded from the study. Likewise if the MAPD value was over 0.34, samples were re-processed and hybridized to a new array or they were excluded. To be consistent with other published studies (Glessner et al. 2010; Van Den Bossche et al. 2013), we chose ten
probes as the minimum number required to make a CNV call. All CNVs were manually inspected using the UCSC Genome Browser (http://genome.ucsc.edu/) (Kent et al. 2002) and cross referenced with DGV (http://projects.tcag.ca/variation/) (Iafrate et al. 2004). First, CNVs were first categorized as deletions or duplications. Then CNVs were identified as genic if they included part or all of a gene and exonic if they overlapped the coding region of a gene; thus, a CNV that was completely within an intron would be genic but not exonic. CNVs were called novel if they had \( \leq 80\% \) overlap with DGV variants.

In order to identify CNVs most likely to be involved in schizophrenia risk, we defined a CNV as conservative if it met any of the following three criteria: 1) a known deleterious CNV, 2) a CNV > 1 Mb that was novel to DGV or 3) a CNV < 1 Mb that was novel to DGV and that overlapped the coding region of a gene of interest. Genes of interest included genes with a previous association autism or schizophrenia, genes with high or specific brain expression, or an association with other neurocognitive or neuropsychiatric disorders.

15q11.2 BP1-2 Validation Assays

The chromosome 15q11.2 BP1-2 deletions were validated with two additional molecular methods. These included NimbleGen 385K whole-genome-tiling comparative genomic hybridization (CGH) arrays and quantitative polymerase chain reaction (qPCR).

NimbleGen arrays required the use of 1 µg of high quality patient DNA and 1 µg of reference DNA (Promega, Madison, WI) that were fluorescently labeled in parallel, followed by co-hybridization to a NimbleGen 385K array (NimbleGen, Madison, WI). The array was then scanned using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA) and signal intensity data were analyzed using NimbleScan software.
NimbleScan uses the SegMNT algorithm to identify CNVs (Gokcumen & Lee 2009) CNV calling was performed using Nexus software (BioDisocvery, El Segundo, CA) (Darvishi 2010).

All qPCR primers were designed via Primer3 using genomic DNA sequence obtained from the UCSC Genome Browser and primer specificity was confirmed using the BLAT tool (Rozen & Skaletsky 2000; Kent 2002). The qPCR reaction contained 12.5 µl of 2x QuantiTect SYBR Green PCR Master Mix (QIAGEN, Germantown, MD), 12 µl genomic DNA (1ng/µl) and 0.25 µl of each primer (10pmol/µl) in a total volume of 25µl. The real-time PCR was run using an Applied Biosystems 7500 Real-Time PCR System (Applied Biosystems, Carlsbad, CA). Each sample was amplified in triplicate with primers designed for gene dosage controls using the genes GAPDH and G6PD and the putative CNV. Results from qPCR were analyzed using the ΔΔCt method, and the data was normalized by setting a pooled genomic DNA reference (Promega, Madison, WI) to a fold change of 1.0.

Results

Analysis of All CNVs

We analyzed DNA from 166 subjects with schizophrenia and 52 psychiatrically healthy controls using Affymetrix 6.0 SNP arrays. Consistent with the demographics of schizophrenia, the affected group had a greater proportion of male subjects (68.7% compared with 52.0% in the control group; $p = 0.0315$). Total deletions comprised a greater proportion of CNVs in cases (54.9%) than they did in controls (50.4%; $p = 3.32 \times 10^{-6}$). With a total of 9707 CNVs, case subjects had an average of 58 CNVs (10 CNVs > 100 kb) per person, while control subjects had a slightly higher average of 71 CNVs (12 CNVs > 100 kb) per person. For case subjects, the mean CNV size was 106.4 kb and the median size was 25 kb. Control subjects had similar results with a slightly higher mean
CNV size of 143.7 kb and a median size of 24 kb. Case subjects did not have a higher proportion of genic or exonic CNVs, or novel CNVs compared to controls (Table 2-1 & 2-2). Given previous data suggesting a size bias in affected versus unaffected individuals, CNVs were also divided into three different size categories: small (CNVs < 100 kb), medium (100 kb < CNVs < 1 Mb), and large (CNVs > 1 Mb). Cases were not different than controls in proportion of deletions, duplications, genic CNVs, exonic CNVs or novel CNVs in any of the size categories (Table 2-1 & 2-2).

After analyzing all CNVs, we focused on those that met the definition for conservative. Forty-nine case subjects harbored 88 conservative CNVs for an average of 0.53 per individual, while 20 controls harbored 31 conservative CNVs for an average of 0.60 per individual (for a summary of conservative CNVs see Tables 2-3 & 2-4). A slightly greater proportion of control subjects (28.8%) compared to cases (17.5%) harbored just one conservative CNV. However, case subjects have a small increase in individuals with two, three or four-or-more conservative CNVs (Table 2-5). Clinically significant CNVs (conservative CNVs > 1 Mb or known deleterious CNVs) were also compared between cases and controls. A slightly higher proportion of case subjects (6.0%, n = 10) harbored a clinically significant CNV compared to controls (1.9%, n = 1) (Table 2-6). However, none of these differences were statistically significant.

15q11.2 BP1-2 Deletions

Of particular interest, three subjects with schizophrenia were found to have a deletion of 15q11.2 known to be associated with a variety of neuropsychiatric phenotypes. Two of the deletions were maternally inherited, 627 kb and 581 kb in size, respectively, and spanned the BP1-2 interval of the PWS/AS critical region (Figure 2-1) (Doornbos et al. 2009; Burnside et al.). No parental DNA was available for the third
subject who had a larger deletion (975 kb) with breakpoints that extended beyond BP1-2 (Figure 2-1). The two cases with known inheritance patterns were also found to have a high number of additional conservative CNVs. All three 15q11.2 deletions were validated with qPCR and NimbleGen 385K array CGH. Neither this deletion nor its reciprocal duplication was found in any of the control samples.

**Case 1** is a 29-year-old female patient who was diagnosed with schizophrenia at age 20. She was born after an uneventful full-term pregnancy at a normal birth weight and did not exhibit motor or verbal delay. On examination at our institution, she had moderate psychotic and disorganized symptoms and more severe negative symptoms. On cognitive testing, the patient had a FSIQ = 85, VIQ = 90 and a PIQ = 74. The microarray study showed this patient to have a 627 kb deletion at 15q11.2 as well as six additional conservative CNVs. While most of these additional CNVs were small, the total of seven represents the highest number of conservative CNVs for any single case (Table 2-7).

**Case 2** is a 39-year-old male patient who was diagnosed with schizophrenia at age 15. This patient exhibited an extremely severe formal thought disorder with disorganized speech and behavior. He had moderate negative symptoms of avolition, apathy, and paucity of speech, and was relatively spared in the domain of psychotic symptoms. He was born with a normal birth weight after an uneventful full-term pregnancy to normal aged parents. This patient did not exhibit any motor or verbal delay, but did receive special education as a child. From cognitive testing, the patient was noted to have impaired memory, attention and concentration in the context of low average intellectual functioning, FSIQ = 85, VIQ = 84 and PIQ = 87. The patient was unemployed and being cared for by his parents at the time of assessment. This patient had 581 kb deletion at 15q11.2 as well as five additional conservative CNVs, comprising the second largest number of conservative CNVs of any individual in the study (Table 2-7).

**Case 3** is a 42-year-old white female who was diagnosed with schizophrenia at age 32. Her illness was characterized by prominent paranoid delusions, persistent
auditory hallucinations, disorganized behavior and negative symptoms such as low motivation and paucity of speech. She was born after an uneventful full-term pregnancy at a normal birth weight. She had normal developmental milestones. She graduated high school and took some college courses, but had been unemployed for most of her adult life. She had no significant past medical history. She had a sister who was diagnosed with paranoid schizophrenia and who died from suicide, and a mother with a history of depression that led to one psychiatric hospitalization. Cognitive testing showed FSIQ = 85, VIQ = 74 and PIQ = 85. This patient had a 975 kb deletion of 15q11.2 with no additional conservative CNVs.

Discussion

We compared the occurrence of genomic CNVs between individuals with schizophrenia and psychiatrically unaffected controls. In particular, we examined whether smaller CNVs than those traditionally examined or multiple CNVs were associated with schizophrenia. We performed these analyses looking first at all CNVs and then at “conservative” CNVs (those deemed most likely to be relevant to disease). When examining all CNVs, we found no significant differences between cases and controls in the rates of occurrence of genic or exonic CNVs, nor did we find significant differences when they were divided by size group (small, medium, or large) (Table 2-1). This also held true when we restricted analyses to CNVs that were novel, meaning that they were not reported in the DGV (Table 2-2). However, we did find that a slightly higher proportion of case subjects harbored clinically significant CNVs (conservative CNV > 1 Mb or known deleterious) compared to controls. Similarly we found that a slightly greater proportion of individuals with schizophrenia had two or more conservative CNVs (Figure 2-5). In this context, we did note with interest that two of the
individuals with the highest conservative CNV count were both individuals with schizophrenia who also both had 15q11.2 deletions as their primary deleterious chromosomal abnormality. It is also noteworthy that two of the three 15q11.2 individuals had abnormal MRI findings. Case 2 and 3 exhibit enlarged ventricles and case 3 has a large cavum septum pellucidum, all of which are MRI findings generally associated with schizophrenia (Weinberger et al. 1979; Reveley et al. 1982; Nopoulos et al. 1996; Nopoulos et al. 1997; Wright et al. 2000).

Our findings support the emerging association between the 15q11.2 BP1-2 deletion and schizophrenia (Stefansson et al. 2008; Kirov et al. 2009a; Zhao et al. 2013). This association with schizophrenia has been debated due to its association with other disorders such as autism, ID, ADHD, and epilepsy (Doornbos et al. 2009; de Kovel et al. 2010; Hogart et al. 2010; van der Zwaag et al. 2010; Van Den Bossche et al. 2012; Wisniowiecka-Kowalnik et al. 2013). 15q11.2 CNVs have one of the lowest odds ratios for disease-associated CNVs (Girirajan et al. 2012), and they are often transmitted from a mildly or unaffected parent (Murthy et al. 2007; Doornbos et al. 2009; van der Zwaag et al. 2010; von der Lippe et al. 2011; Abdelmoity et al. 2012). Furthermore, out of 72 CNVs previously known to be associated with neuropsychiatric disease, the 15q11.2 BP1-2 deletion was reported to have the highest likelihood of harboring additional CNVs in patients with ID and CA (Girirajan et al. 2012). Taken together, this means that it is highly likely that 15q11.2 deletion requires additional disease-susceptibility variants in order to directly increase the risk for schizophrenia in patients.

The 15q11.2 BP1-2 region harbors four genes CYFIP1, NIPA1, NIPA2 and GCP5 (Chai et al. 2003). CYFIP1 is highly expressed in the brain and interacts with FMRP to regulate activity dependent translation at the developing synapse during neurogenesis (Napoli et al. 2008). Expression of NIPA1 has been reported in the rodent (mouse and rat) frontal cortex, cerebellum and striatum of the brain in addition to the spinal cord (Zhao et al. 2008); both NIPA1 and NIPA2 are magnesium transporters that act in a dose-
dependent manner (Zhao et al. 2008). *GCP5* is expressed ubiquitously, but with highest levels in the subthalamic nuclei of the brain, a region involved in ADHD and obsessive compulsive behavior (Doornbos et al. 2009). *GCP5* encodes a member of the γ-tubulin complex, a protein that is thought to be involved in axonal migration during brain development (Chai et al. 2003). Patients with neuropsychiatric disorders that share the BP1-2 deletion exhibit lower intellectual ability, delayed speech and motor development, behavioral problems and psychosis (Chai et al. 2003; Stefánsson et al. 2008; Doornbos et al. 2009; Burnside et al. 2011; von der Lippe et al. 2011).

A small number of studies searching for multiple susceptibility CNVs have recently been reported. An examination of ID provided strong evidence for a multiple-hit CNV hypothesis, though only CNVs that were highly likely to be deleterious based on both size and known pathogenicity were studied, thereby excluding many novel and potentially real risk variants (Girirajan et al. 2012). Another study compared 22q11 deletion patients with and without psychosis and found those patients with psychosis had a higher burden of “second-hits” (genic CNVs > 100 kb with a prior involvement with neuropsychiatric disorders) (Williams et al. 2013). If a multiple-hit hypothesis is truly the underlying model for the majority of cases of schizophrenia, it is very likely that it is not just a combination of CNVs, but of SNPs and CNVs that are truly causative. A recent report has started to examine this possibility where both common SNPs and rare CNVs were implicated in the etiology of ADHD (Yang et al. 2013).
Table 2-1. Summary of CNVs compared between cases and controls.

<table>
<thead>
<tr>
<th>CNV Size Group</th>
<th>Cases (n=166)</th>
<th>% (n)</th>
<th>CNVs/person</th>
<th>% Deletion</th>
<th>Controls (n=52)</th>
<th>% (n)</th>
<th>CNVs/person</th>
<th>% Deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>9707</td>
<td>58.48</td>
<td>54.9% (5332)</td>
<td></td>
<td>3702</td>
<td>71.19</td>
<td>50.4% (1867)</td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>82.7% (8023)</td>
<td>48.33</td>
<td>47.3% (4594)</td>
<td></td>
<td>83.7% (3099)</td>
<td>59.60</td>
<td>44.8% (1657)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>16.0% (1551)</td>
<td>9.34</td>
<td>7.2% (697)</td>
<td></td>
<td>14.7% (544)</td>
<td>10.46</td>
<td>5.3% (198)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>1.4% (133)</td>
<td>0.80</td>
<td>0.4% (41)</td>
<td></td>
<td>1.6% (59)</td>
<td>1.13</td>
<td>0.3% (12)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-1 — Continued

<table>
<thead>
<tr>
<th>CNV Size Group</th>
<th>Genic CNVs</th>
<th>Controls (n=52)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases (n=166)</td>
<td>% (n)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>52.2% (5065)</td>
</tr>
<tr>
<td>Small</td>
<td></td>
<td>38.2% (3711)</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td>12.8% (1239)</td>
</tr>
<tr>
<td>Large</td>
<td></td>
<td>1.2% (115)</td>
</tr>
<tr>
<td>CNV Size Group</td>
<td>Cases (n=166)</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>% (n)</td>
<td>CNVs/person</td>
</tr>
<tr>
<td>Total</td>
<td>34.8% (3375)</td>
<td>20.33</td>
</tr>
<tr>
<td>Small</td>
<td>24.0% (2331)</td>
<td>14.04</td>
</tr>
<tr>
<td>Medium</td>
<td>9.8% (953)</td>
<td>5.74</td>
</tr>
<tr>
<td>Large</td>
<td>0.9% (91)</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Table 2-2. Summary of novel (not found in DGV or in controls) CNVs compared between cases and controls.

<table>
<thead>
<tr>
<th>CNV Size Group</th>
<th>All Novel CNVs</th>
<th>Controls (n=52)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases (n=166)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20.5% (1992)</td>
<td>23.0% (850)</td>
</tr>
<tr>
<td></td>
<td>12.00</td>
<td>16.35</td>
</tr>
<tr>
<td></td>
<td>10.3% (995)</td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>19.6% (1903)</td>
<td>22.0% (816)</td>
</tr>
<tr>
<td></td>
<td>11.46</td>
<td>15.69</td>
</tr>
<tr>
<td></td>
<td>10.0% (969)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>0.8% (75)</td>
<td>0.9% (32)</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>0.3% (25)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.1% (14)</td>
<td>0.1% (2)</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.01% (1)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-2 — Continued

<table>
<thead>
<tr>
<th>CNV Size Group</th>
<th>Cases (n=166)</th>
<th>Controls (n=52)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (n)</td>
<td>% (n)</td>
</tr>
<tr>
<td></td>
<td>CNVs/person</td>
<td>CNVs/person</td>
</tr>
<tr>
<td></td>
<td>% Deletion</td>
<td>% Deletion</td>
</tr>
<tr>
<td>Total</td>
<td>11.6% (1126)</td>
<td>12.7% (471)</td>
</tr>
<tr>
<td></td>
<td>6.78</td>
<td>9.06</td>
</tr>
<tr>
<td></td>
<td>5.8% (560)</td>
<td>6.3% (233)</td>
</tr>
<tr>
<td>Small</td>
<td>11.0% (1072)</td>
<td>12.2% (453)</td>
</tr>
<tr>
<td></td>
<td>6.46</td>
<td>8.71</td>
</tr>
<tr>
<td></td>
<td>5.7% (549)</td>
<td>6.1% (227)</td>
</tr>
<tr>
<td>Medium</td>
<td>0.5% (47)</td>
<td>0.5% (18)</td>
</tr>
<tr>
<td></td>
<td>0.28</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.1% (10)</td>
<td>0.2% (6)</td>
</tr>
<tr>
<td>Large</td>
<td>0.1% (7)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.01% (1)</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2-2 — Continued

<table>
<thead>
<tr>
<th>CNV Size Group</th>
<th>Cases (n=166)</th>
<th>% (n)</th>
<th>CNVs/person</th>
<th>% Deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7.0% (676)</td>
<td>4.07</td>
<td>3.2% (312)</td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>6.5% (629)</td>
<td>3.79</td>
<td>3.1% (305)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>0.4% (41)</td>
<td>0.25</td>
<td>0.4% (7)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.1% (6)</td>
<td>0.04</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Controls (n=52)</th>
<th>% (n)</th>
<th>CNVs/person</th>
<th>% Deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7.9% (293)</td>
<td>5.63</td>
<td>3.6% (132)</td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>7.5% (278)</td>
<td>5.35</td>
<td>3.5% (128)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>0.4% (15)</td>
<td>0.29</td>
<td>0.1% (4)</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-3. Summary of all conservative CNVs identified in cases. Individuals are separated by a vertical space.

<table>
<thead>
<tr>
<th>CNV type</th>
<th>Chr</th>
<th>Cyto start</th>
<th>Cyto end</th>
<th>Chr start</th>
<th>Chr end</th>
<th>Size (kb)</th>
<th>Genes of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>2</td>
<td>q37.1</td>
<td>q37.1</td>
<td>233439488</td>
<td>233468763</td>
<td>29</td>
<td>NGEF</td>
</tr>
<tr>
<td>Gain</td>
<td>4</td>
<td>q13.3</td>
<td>q13.3</td>
<td>75468691</td>
<td>75588421</td>
<td>120</td>
<td>AREG</td>
</tr>
<tr>
<td>Gain</td>
<td>5</td>
<td>q35.1</td>
<td>q35.1</td>
<td>170126890</td>
<td>170163266</td>
<td>36</td>
<td>GABRP</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81590188</td>
<td>81608923</td>
<td>19</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>q12.1</td>
<td>q12.1</td>
<td>61355704</td>
<td>61375209</td>
<td>20</td>
<td>CA8</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1549163</td>
<td>1571047</td>
<td>22</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q11.2</td>
<td>q11.2</td>
<td>20224751</td>
<td>20852202</td>
<td>627</td>
<td>CYFIP1, NIPA1, NIPA2, GCP5</td>
</tr>
<tr>
<td>Gain</td>
<td>22</td>
<td>q13.31</td>
<td>q13.31</td>
<td>45094635</td>
<td>45147712</td>
<td>53</td>
<td>CELSR1</td>
</tr>
<tr>
<td>Gain</td>
<td>2</td>
<td>p16.3</td>
<td>p16.3</td>
<td>50307256</td>
<td>50317992</td>
<td>11</td>
<td>NRXN1</td>
</tr>
<tr>
<td>Gain</td>
<td>3</td>
<td>p14.1</td>
<td>p14.1</td>
<td>68051761</td>
<td>68168065</td>
<td>116</td>
<td>FAM19A1</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81419261</td>
<td>81494501</td>
<td>75</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1531832</td>
<td>1571047</td>
<td>39</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q23.1</td>
<td>q23.1</td>
<td>84467878</td>
<td>84491234</td>
<td>23</td>
<td>NRG3</td>
</tr>
<tr>
<td>Gain</td>
<td>11</td>
<td>q22.1</td>
<td>q22.1</td>
<td>99199969</td>
<td>99261000</td>
<td>61</td>
<td>CNTN5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>1</td>
<td>q25.2</td>
<td>q25.2</td>
<td>175266563</td>
<td>175276043</td>
<td>9</td>
<td>ASTN1</td>
</tr>
<tr>
<td>Loss</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81442364</td>
<td>81455770</td>
<td>13</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.3</td>
<td>q21.3</td>
<td>97711618</td>
<td>97778868</td>
<td>67</td>
<td>BRI3</td>
</tr>
<tr>
<td>Loss</td>
<td>9</td>
<td>p13.2</td>
<td>p13.2</td>
<td>36947633</td>
<td>36963576</td>
<td>16</td>
<td>PAX5</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q21.2</td>
<td>q21.2</td>
<td>49843225</td>
<td>49852763</td>
<td>10</td>
<td>TMOD2</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q11.2</td>
<td>q11.2</td>
<td>20224751</td>
<td>20805409</td>
<td>581</td>
<td>CYFIP1, NIPA1, NIPA2, GCP5</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81421790</td>
<td>81494501</td>
<td>73</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>p15.1</td>
<td>p15.1</td>
<td>29173300</td>
<td>29502791</td>
<td>329</td>
<td>CHN2</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1531832</td>
<td>1571047</td>
<td>39</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Gain</td>
<td>11</td>
<td>q14.1</td>
<td>q14.1</td>
<td>83345395</td>
<td>8338264</td>
<td>43</td>
<td>DLG2</td>
</tr>
<tr>
<td>Loss</td>
<td>3</td>
<td>p21.31</td>
<td>p21.31</td>
<td>47550572</td>
<td>47579560</td>
<td>29</td>
<td>CSPG5</td>
</tr>
<tr>
<td>Loss</td>
<td>7</td>
<td>q22.1</td>
<td>q22.1</td>
<td>103257456</td>
<td>103271612</td>
<td>14</td>
<td>RELN</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123321541</td>
<td>123357137</td>
<td>36</td>
<td>FGFR2</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1524533</td>
<td>1571047</td>
<td>47</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Loss</td>
<td>10</td>
<td>p11.23</td>
<td>p11.23</td>
<td>28608921</td>
<td>28619042</td>
<td>10</td>
<td>MPP7</td>
</tr>
<tr>
<td>Loss</td>
<td>12</td>
<td>q13.2</td>
<td>q13.2</td>
<td>54771552</td>
<td>54777943</td>
<td>6</td>
<td>ERBB3</td>
</tr>
<tr>
<td>Loss</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1549163</td>
<td>1571047</td>
<td>22</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123316345</td>
<td>123367346</td>
<td>51</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>q21.1</td>
<td>q21.1</td>
<td>124176223</td>
<td>124216332</td>
<td>40</td>
<td>SEMA5B</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>18</td>
<td>q21.31</td>
<td>53845653</td>
<td>53907653</td>
<td>62</td>
<td>NEDD4L</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>2</td>
<td>q14.3</td>
<td>127521836</td>
<td>127578727</td>
<td>57</td>
<td>BIN1</td>
</tr>
<tr>
<td></td>
<td>Loss</td>
<td>22</td>
<td>q13.31</td>
<td>44481307</td>
<td>44521698</td>
<td>40</td>
<td>ATXN10</td>
</tr>
<tr>
<td></td>
<td>Loss</td>
<td>8</td>
<td>p23.3</td>
<td>1531832</td>
<td>1571047</td>
<td>39</td>
<td>DLGAP2</td>
</tr>
<tr>
<td></td>
<td>Loss</td>
<td>16</td>
<td>q23.3</td>
<td>82385302</td>
<td>82394233</td>
<td>9</td>
<td>CDH13</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>123300015</td>
<td>12357137</td>
<td>57</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>18</td>
<td>q21.31</td>
<td>53856404</td>
<td>53881779</td>
<td>25</td>
<td>NEDD4L</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>6</td>
<td>q22.1</td>
<td>116397042</td>
<td>11764218</td>
<td>1247</td>
<td>GPRC6A</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>22</td>
<td>q12.3</td>
<td>35418232</td>
<td>35430206</td>
<td>12</td>
<td>CACNG2</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>10</td>
<td>p11.23</td>
<td>28607813</td>
<td>28619147</td>
<td>11</td>
<td>MPP7</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>123266839</td>
<td>123342372</td>
<td>76</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>123336106</td>
<td>123345501</td>
<td>9</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td>Loss</td>
<td>17</td>
<td>p13.3</td>
<td>1217177</td>
<td>1231662</td>
<td>14</td>
<td>YWHAE</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>123288947</td>
<td>123357137</td>
<td>68</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>22</td>
<td>q12.2</td>
<td>28198855</td>
<td>28251846</td>
<td>53</td>
<td>NEFH</td>
</tr>
<tr>
<td>Type</td>
<td>Chromosome 1</td>
<td>Chromosome 2</td>
<td>Log2Ratio</td>
<td>Log2FoldChange</td>
<td>Genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>--------------</td>
<td>------------</td>
<td>----------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>3</td>
<td>p26.3</td>
<td>1.879700</td>
<td>2.853116</td>
<td>CNTN4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>16</td>
<td>q23.3</td>
<td>1.8753726</td>
<td>2.81766787</td>
<td>CDH13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>2</td>
<td>q34</td>
<td>2.09390838</td>
<td>2.210138000</td>
<td>MAP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q11.21</td>
<td>4.2923884</td>
<td>4.2940557</td>
<td>RET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>1.1549163</td>
<td>1.1571047</td>
<td>DLGAP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>22</td>
<td>q13.31</td>
<td>4.4485742</td>
<td>4.44505934</td>
<td>ATXN10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>p11.23</td>
<td>2.2860713</td>
<td>2.28619147</td>
<td>MPP7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>1.123266839</td>
<td>1.123289604</td>
<td>FGFR2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>1</td>
<td>p36.22</td>
<td>1.10372607</td>
<td>1.10531677</td>
<td>CORT, DFFA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>4</td>
<td>p16.2</td>
<td>1.3195187</td>
<td>1.3228119</td>
<td>HTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>2</td>
<td>p23.1</td>
<td>1.31278092</td>
<td>1.31320957</td>
<td>CAPN14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>8</td>
<td>q22.3</td>
<td>1.105053276</td>
<td>1.105087630</td>
<td>RIMS2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q11.2</td>
<td>1.20059085</td>
<td>1.21033790</td>
<td>CYFIP1, NIPA1, NIPA2, GCP5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>5</td>
<td>q21.2</td>
<td>1.104488059</td>
<td>1.106174981</td>
<td>none</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2-3 — Continued

<table>
<thead>
<tr>
<th>Gain</th>
<th>Chrm</th>
<th>Start</th>
<th>End</th>
<th>Fold</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>5</td>
<td>q35.2</td>
<td>q35.2</td>
<td>175242989</td>
<td>175353490</td>
</tr>
<tr>
<td>Gain</td>
<td>1</td>
<td>p31.3</td>
<td>p31.1</td>
<td>62555091</td>
<td>69789141</td>
</tr>
<tr>
<td>Gain</td>
<td>2</td>
<td>q14.2</td>
<td>q14.2</td>
<td>119907901</td>
<td>119949330</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.1</td>
<td>2181322</td>
<td>8027692</td>
</tr>
<tr>
<td>Gain</td>
<td>3</td>
<td>p26.3</td>
<td>p26.3</td>
<td>205010</td>
<td>2744392</td>
</tr>
<tr>
<td>Gain</td>
<td>12</td>
<td>q15</td>
<td>q15</td>
<td>69107611</td>
<td>69135504</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q21.3</td>
<td>q21.3</td>
<td>69002940</td>
<td>69125258</td>
</tr>
<tr>
<td>Loss</td>
<td>3</td>
<td>p21.1</td>
<td>p21.1</td>
<td>51713800</td>
<td>51745469</td>
</tr>
<tr>
<td>Loss</td>
<td>17</td>
<td>q12</td>
<td>q12</td>
<td>35102047</td>
<td>35108033</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1549163</td>
<td>1571047</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1549310</td>
<td>1571047</td>
</tr>
<tr>
<td>Gain</td>
<td>17</td>
<td>p11.2</td>
<td>p11.2</td>
<td>18693295</td>
<td>20166343</td>
</tr>
</tbody>
</table>
Table 2-3 — Continued

<table>
<thead>
<tr>
<th></th>
<th>3 q21.2</th>
<th>q21.2</th>
<th>125760858</th>
<th>125773010</th>
<th>12</th>
<th>KALRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss</td>
<td>14 q12</td>
<td>q12</td>
<td>24352180</td>
<td>24366394</td>
<td>14</td>
<td>STXBP6</td>
</tr>
<tr>
<td>Gain</td>
<td>16 q21</td>
<td>q21</td>
<td>60379710</td>
<td>60391854</td>
<td>12</td>
<td>CDH8</td>
</tr>
<tr>
<td>Gain</td>
<td>10 q26.13</td>
<td>q26.13</td>
<td>123336106</td>
<td>123357137</td>
<td>21</td>
<td>FGFR2</td>
</tr>
<tr>
<td>Loss</td>
<td>12 q21.1</td>
<td>q21.1</td>
<td>70650458</td>
<td>70675106</td>
<td>25</td>
<td>TPH2</td>
</tr>
<tr>
<td>Gain</td>
<td>10 q26.13</td>
<td>q26.13</td>
<td>123336106</td>
<td>123345501</td>
<td>9</td>
<td>FGFR2</td>
</tr>
<tr>
<td>Gain</td>
<td>8 p23.3</td>
<td>p23.3</td>
<td>15318324</td>
<td>1571047</td>
<td>39</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Loss</td>
<td>15 q25.2</td>
<td>q25.2</td>
<td>81348338</td>
<td>81361085</td>
<td>13</td>
<td>HOMER2</td>
</tr>
<tr>
<td>Gain</td>
<td>3 p21.31</td>
<td>p21.31</td>
<td>45238755</td>
<td>45265011</td>
<td>26</td>
<td>TMEM158</td>
</tr>
<tr>
<td>Loss</td>
<td>2 p23.1</td>
<td>p23.1</td>
<td>31278092</td>
<td>31324349</td>
<td>46</td>
<td>CAPN14</td>
</tr>
<tr>
<td>Gain</td>
<td>12 q24.31</td>
<td>q24.31</td>
<td>122663481</td>
<td>122723892</td>
<td>60</td>
<td>EIF2B1</td>
</tr>
</tbody>
</table>
Table 2-4. Summary of all conservative CNVs identified in controls. Individuals are separated by a vertical space.

<table>
<thead>
<tr>
<th>CNV type</th>
<th>Chr</th>
<th>Cyto start</th>
<th>Cyto end</th>
<th>Chr start</th>
<th>Chr end</th>
<th>Size (kb)</th>
<th>Genes of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss</td>
<td>2</td>
<td>p16.3</td>
<td>p16.3</td>
<td>50277111</td>
<td>50322145</td>
<td>45</td>
<td>NRXN1</td>
</tr>
<tr>
<td>Loss</td>
<td>4</td>
<td>p16.2</td>
<td>p16.2</td>
<td>3105354</td>
<td>3118622</td>
<td>13</td>
<td>HTT</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1621060</td>
<td>1641097</td>
<td>20</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Loss</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1523169</td>
<td>1583489</td>
<td>60</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>p11.23</td>
<td>p11.23</td>
<td>28587137</td>
<td>28619147</td>
<td>32</td>
<td>MPP7</td>
</tr>
<tr>
<td>Loss</td>
<td>20</td>
<td>q13.12</td>
<td>q13.12</td>
<td>44111204</td>
<td>44146247</td>
<td>35</td>
<td>SLC12A5</td>
</tr>
<tr>
<td>Gain</td>
<td>22</td>
<td>q13.32</td>
<td>q13.32</td>
<td>47402059</td>
<td>47432483</td>
<td>30</td>
<td>FAM19A5</td>
</tr>
<tr>
<td>Loss</td>
<td>2</td>
<td>p25.1</td>
<td>p25.1</td>
<td>10468932</td>
<td>10491011</td>
<td>22</td>
<td>HPcal1</td>
</tr>
<tr>
<td>Gain</td>
<td>11</td>
<td>p15.5</td>
<td>p15.5</td>
<td>1601016</td>
<td>1941703</td>
<td>341</td>
<td>CTSD</td>
</tr>
<tr>
<td>Loss</td>
<td>22</td>
<td>q13.32</td>
<td>q13.32</td>
<td>47482320</td>
<td>47494381</td>
<td>12</td>
<td>FAM19A5</td>
</tr>
<tr>
<td>Gain</td>
<td>3</td>
<td>q21.1</td>
<td>q21.1</td>
<td>124188528</td>
<td>12421632</td>
<td>28</td>
<td>SEMA5B</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123266839</td>
<td>123337714</td>
<td>71</td>
<td>FGFR2</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123272551</td>
<td>123367346</td>
<td>95</td>
<td>FGFR2</td>
</tr>
<tr>
<td>Loss</td>
<td>13</td>
<td>q12.12</td>
<td>q12.12</td>
<td>22417918</td>
<td>23814077</td>
<td>1396</td>
<td>SAC5</td>
</tr>
<tr>
<td>Loss</td>
<td>7</td>
<td>q31.32</td>
<td>q31.32</td>
<td>121426894</td>
<td>121440206</td>
<td>13</td>
<td>PTPRZ1</td>
</tr>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123336106</td>
<td>123357137</td>
<td>21</td>
<td>FGFR2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>---</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1524808</td>
<td>1571047</td>
<td>46</td>
<td><strong>DLGAP2</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1531832</td>
<td>1571047</td>
<td>39</td>
<td><strong>DLGAP2</strong></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>83589706</td>
<td>83603154</td>
<td>13</td>
<td><strong>SEMA3A</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>17</td>
<td>p13.3</td>
<td>p13.3</td>
<td>3347986</td>
<td>3359568</td>
<td>12</td>
<td><strong>ASPA</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>16</td>
<td>q23.3</td>
<td>q23.3</td>
<td>81750763</td>
<td>81766787</td>
<td>16</td>
<td><strong>CDH13</strong></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>9</td>
<td>q21.11</td>
<td>q21.11</td>
<td>71280626</td>
<td>71310085</td>
<td>29</td>
<td><strong>APBA1</strong></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>22</td>
<td>q13.31</td>
<td>q13.32</td>
<td>46847352</td>
<td>47607039</td>
<td>760</td>
<td><strong>FAM19A5</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>7</td>
<td>q31.32</td>
<td>q31.32</td>
<td>121428890</td>
<td>121452150</td>
<td>23</td>
<td><strong>PTPRZ1</strong></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1524808</td>
<td>1571047</td>
<td>46</td>
<td><strong>DLGAP2</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>8</td>
<td>p21.1</td>
<td>p21.1</td>
<td>28410213</td>
<td>28465072</td>
<td>55</td>
<td><strong>FZD3</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>5</td>
<td>p15.33</td>
<td>p15.33</td>
<td>1480803</td>
<td>1490408</td>
<td>10</td>
<td><strong>SLC6A3</strong></td>
</tr>
<tr>
<td><strong>Loss</strong></td>
<td>2</td>
<td>p16.3</td>
<td>p16.3</td>
<td>50307256</td>
<td>50323795</td>
<td>17</td>
<td><strong>NRXN1</strong></td>
</tr>
<tr>
<td><strong>Gain</strong></td>
<td>3</td>
<td>q21.1</td>
<td>q21.1</td>
<td>124191210</td>
<td>124216332</td>
<td>25</td>
<td><strong>SEMA5B</strong></td>
</tr>
</tbody>
</table>
Table 2-4 — Continued

<table>
<thead>
<tr>
<th>Loss</th>
<th>3</th>
<th>q13.31</th>
<th>q13.31</th>
<th>117220991</th>
<th>117231808</th>
<th>11</th>
<th>LSAMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>10</td>
<td>q26.13</td>
<td>q26.13</td>
<td>123266839</td>
<td>123328965</td>
<td>62</td>
<td>FGFR2</td>
</tr>
</tbody>
</table>
Table 2-5. Summary of the percentage of individuals with single or multiple conservative CNVs.

<table>
<thead>
<tr>
<th>Case Group</th>
<th>None</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four-or-more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>66.3% (n=110)</td>
<td>17.5% (n=29)</td>
<td>7.8% (n=13)</td>
<td>1.8% (n=3)</td>
<td>2.4% (n=4)</td>
</tr>
<tr>
<td>Controls</td>
<td>53.8% (n=28)</td>
<td>28.8% (n=15)</td>
<td>5.7% (n=3)</td>
<td>1.9% (n=1)</td>
<td>1.9% (n=1)</td>
</tr>
</tbody>
</table>
Table 2-6. Summary of clinically significant conservative CNVs. \(^1\)Same individual.

<table>
<thead>
<tr>
<th>Cases</th>
<th>CNV type</th>
<th>CNV size</th>
<th>CNV location</th>
<th>Genes of interest</th>
<th>Controls</th>
<th>CNV type</th>
<th>CNV size</th>
<th>CNV location</th>
<th>Genes of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Duplication</td>
<td>7.2 Mb</td>
<td>1p31.3-p31.1</td>
<td>PDE4B</td>
<td>Deletion</td>
<td>1.4 Mb</td>
<td>13q12.12</td>
<td>SACS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
<td>2.5 Mb</td>
<td>3p26.3</td>
<td>CHL1, CNTN6, CNTN4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
<td>1.6 Mb</td>
<td>5q21.2-q21.3</td>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
<td>1.2 Mb</td>
<td>6q22.1-q22.2</td>
<td>GPRC6A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
<td>5.8 Mb</td>
<td>8p23.3-p23.1</td>
<td>MCPH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deletion</td>
<td>627 kb</td>
<td>15q11.2</td>
<td>CYFIP1, NIPA1, NIPA2 &amp; GCP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deletion</td>
<td>581 kb</td>
<td>15q11.2</td>
<td>CYFIP1, NIPA1, NIPA2 &amp; GCP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deletion</td>
<td>975 kb</td>
<td>15q11.2</td>
<td>CYFIP1, NIPA1, NIPA2 &amp; GCP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
<td>1.5 Mb</td>
<td>17p11.2</td>
<td>RNF112, ULK2, SLC47A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^1)Deletion</td>
<td>2.2 Mb</td>
<td>3p12.2-12.1</td>
<td>none of interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^1)Duplication</td>
<td>17.6 Mb</td>
<td>16q22.3-q24.3</td>
<td>COTL1, JPH3, FAM38A, CDH15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2-7. Summary of additional conservative CNVs in case 1 and case 2.

<table>
<thead>
<tr>
<th>CNV type</th>
<th>Chr</th>
<th>Cyto start</th>
<th>Cyto end</th>
<th>Chr start</th>
<th>Chr end</th>
<th>Size (kb)</th>
<th>Genes of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>2</td>
<td>q37.1</td>
<td>q37.1</td>
<td>233439488</td>
<td>233468763</td>
<td>29</td>
<td>NGEF</td>
</tr>
<tr>
<td>Gain</td>
<td>4</td>
<td>q13.3</td>
<td>q13.3</td>
<td>75468691</td>
<td>75588421</td>
<td>120</td>
<td>AREG</td>
</tr>
<tr>
<td>Gain</td>
<td>5</td>
<td>q35.1</td>
<td>q35.1</td>
<td>170126890</td>
<td>170163266</td>
<td>36</td>
<td>GABRP</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81590188</td>
<td>81608923</td>
<td>19</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>q12.1</td>
<td>q12.1</td>
<td>61355704</td>
<td>61375209</td>
<td>20</td>
<td>CA8</td>
</tr>
<tr>
<td>Gain</td>
<td>8</td>
<td>p23.3</td>
<td>p23.3</td>
<td>1549163</td>
<td>1571047</td>
<td>22</td>
<td>DLGAP2</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q11.2</td>
<td>q11.2</td>
<td>20224751</td>
<td>20852202</td>
<td>627</td>
<td>CYFIP1, NIPA1, NIPA2 &amp; GCP5</td>
</tr>
<tr>
<td><strong>Case 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>1</td>
<td>q25.2</td>
<td>q25.2</td>
<td>175266563</td>
<td>175276043</td>
<td>9</td>
<td>ASTN1</td>
</tr>
<tr>
<td>Loss</td>
<td>7</td>
<td>q21.11</td>
<td>q21.11</td>
<td>81442364</td>
<td>81455770</td>
<td>13</td>
<td>CACNA2D1</td>
</tr>
<tr>
<td>Gain</td>
<td>7</td>
<td>q21.3</td>
<td>q21.3</td>
<td>97711618</td>
<td>97778868</td>
<td>67</td>
<td>BRI3</td>
</tr>
<tr>
<td>Loss</td>
<td>9</td>
<td>p13.2</td>
<td>p13.2</td>
<td>36947633</td>
<td>36963576</td>
<td>16</td>
<td>PAX5</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q11.2</td>
<td>q11.2</td>
<td>20224751</td>
<td>20805409</td>
<td>581</td>
<td>CYFIP1, NIPA1, NIPA2 &amp; GCP5</td>
</tr>
<tr>
<td>Loss</td>
<td>15</td>
<td>q21.2</td>
<td>q21.2</td>
<td>49843225</td>
<td>49852763</td>
<td>10</td>
<td>TMOD2</td>
</tr>
</tbody>
</table>
Figure 2-1. UCSC screen capture of 15q11.2 BP1-2 deletions with case 1 & 2 on the bottom in green and case 3 on top in blue. Deletion coordinates for case 1 (627 kb), as determined by Affymetrix Genotyping Console 4.0, are chr15: 20224751-20852202. Coordinates for case 2 (581 kb) are chr15: 20224751-20805409 and coordinates for case 3 (975 kb) are chr15: 20059085-21033790.
Figure 2-2. Scatterplot where each dot represents the number of conservative CNVs in one individual. Overlapping dots appear darker.
CHAPTER III
COPY NUMBER VARIANTS EFFECTS ON COGNITIVE ABILITIES
AND BRAIN STRUCTURE VOLUMES

Introduction

Over 30 years ago Gottesman and Shields (1973) posited that “measureable components unseen by the unaided eye [exist] along the pathway between disease and distal genotype, and thus [are] closer to the biological expression or pathophysiology of the illness than clinical symptoms”. These measurable components are known as endophenotypes and can range from single biological traits to complex sets of interrelated features. To be useful, endophenotypes should be 1) associated with the illness of interest, 2) heritable, 3) sometimes present in the absence of illness, 4) co-segregated with illness in families and 5) more stable and reliably measurable than the illness itself (Gottesman & Gould 2003). An endophenotype may be neurophysiological, biochemical, endocrinological, neuroanatomical, cognitive or neuropsychological (Thaker 2008). The utility of endophenotypes is that they may be less genetically complex and less influenced by non-genetic factors. Genetic factors that underlie endophenotypes may lead to a more straightforward identification of variants that underlie the disorder of interest. Endophenotypes that have been pursued in schizophrenia research include pre-pulse inhibition (Braff 1993), auditory P50 suppression (Freedman et al. 1997), smooth-pursuit eye tracking movements (Calkins & Iacono 2000) and working memory (Gasperoni et al. 2003).
Brain Morphology

A wealth of data emerging in recent years supports brain morphology as a valid schizophrenia endophenotypic trait. Genetic factors influence morphology of numerous brain structures (Thompson et al. 2001; Thompson et al. 2002; Wright et al. 2002; Scamvougeras et al. 2003; Winterer & Goldman 2003; Pfefferbaum et al. 2004; Styner et al. 2005; Rimol et al. 2010). In healthy subjects, the level of intellectual functioning has been positively associated with whole brain, gray and white-matter volumes (Thompson et al. 2001; Posthuma et al. 2002). Additionally, it has been recognized that the trajectory changes in cortical thickness during adolescence can be associated with levels of intelligence (Shaw et al. 2006). Knowing which genes are involved in brain structure can be very important, especially in regards to shared genes between brain structure and cognitive functioning. Studies in mice have led to the identification of specific genes that influence brain structure, such as Otx1, Cntf, Emx2, and Pax6 (Kimura et al. 2005; Beatty & Laughlin 2006). In humans, MCPH1 and ASPM appear to have influenced the evolution of human cortical enlargement and mutations in these genes cause microcephaly (Jackson et al. 1998; Pattison et al. 2000; Bond et al. 2002; Jackson et al. 2002). Additional genes that influence structural morphology in healthy controls include SLC6A4 (Pezawas et al. 2005), DAT, DRD4 (Durston et al. 2005) and DISC1 (Callicott et al. 2005).

Estimates of heritability of brain structure may be higher than any other schizophrenia intermediate trait. A powerful tool for measuring brain structure volume in the human brain is MRI which researchers have been using to calculate heritability estimates for global brain measures including intracranial volume (> 80%) (Carmelli et al. 1998; Pfefferbaum et al. 2000; Baare et al. 2001) and total brain volume (66-97%) (Bartley et al. 1997; Pennington et al. 2000; Baare et al. 2001; Wright et al. 2002).
addition, twin and sibling studies have led to heritability estimates of 82% for gray matter and 88% for white matter (Baare et al. 2001). In the first twin study of brain morphometry in children, additive genetic effects accounted for variability in nearly all brain regions including total cerebrum and lobar volumes, including gray and white matter sub-compartments (77-88%) (Wallace et al. 2006). Additionally, heritability for the caudate nucleus was estimated at 80% and the only structure which was inconsistent with findings in adults was the cerebellum which had the lowest additive genetic effect (Wallace et al. 2006).

Morphologies of numerous structures have been found to be associated with genetic risk for schizophrenia. The overall brain volume of patients with schizophrenia is 3% smaller than that of healthy controls (Ward et al. 1996; Wright et al. 2000; Narr et al. 2002; van Haren et al. 2012). A decrease in the volume of specific substructures has also been identified in: cortical gray matter (Wright et al. 2002; Jagannathan et al. 2010; van Haren et al. 2012); the hippocampus (Nelson et al. 1998); the thalamus (Konick & Friedman 2001); prefrontal and temporal lobes, particularly in the medial temporal lobe and superior temporal lobe (Lawrie et al. 2008); gray matter volumes of fronto-striato-thalamic and left lateral temporal regions; white matter volumes of left frontal and temporoparietal regions (McDonald et al. 2004). Additionally, increased ventricular size has been associated with schizophrenia (Raz & Raz 1990; Van Horn & McManus 1992; Wright et al. 2000; Sayo et al. 2012). Genes that affect brain structure in schizophrenia include COMT (Ohnishi et al. 2006), IL1-RN (Papiol et al. 2005), BDNF (Szeszko et al. 2005) and RGS4 (Prasad et al. 2005).
Cognition

Cognitive impairment is usually present early in schizophrenia (Cornblatt & Keilp 1994), neurocognitive deficits are stable across the course of illness and such deficits predict long-term outcome (Green 1996; Green et al. 2000). Cognitive dysfunction contributes substantially to the morbidity of schizophrenia and deserves careful scrutiny. Heritability for general IQ has been estimated at 50-70% (Bouchard et al. 1990; Devlin et al. 1997), while more specific cognitive traits such as novelty seeking (Heiman et al. 2004; Benyamin et al. 2005), verbal reasoning (Benyamin et al. 2005), task persistence (McCartney & Berry 2005) and others have heritability ranging from 35-80%. Cognitive phenotypes found to be associated with genetic risk for schizophrenia include central executive processing (Heydebrand 2006), spatial working memory (Pirkola et al. 2005), executive functioning (Goldberg et al. 1995) and impaired attention (Egan et al. 2000; Egan et al. 2001). Genes that have been shown to influence cognition in the normal population include \textit{DRD4} (Lusher et al. 2001), \textit{SLC6A4} (Lesch et al. 1996), \textit{DRD2} (Blasi et al. 2009) and \textit{DTNBP1} (Burdick et al. 2006). Genes shown to effect cognition in the context of schizophrenia include \textit{COMT} (Egan et al. 2001; Bertolino et al. 2005), \textit{G72} (Goldberg et al. 2006), \textit{GRM3} (Egan et al. 2004) and \textit{BDNF} (Egan et al. 2003).

Cognitive deficits also meet the criteria for being a useful endophenotype in schizophrenia because they are pervasive over a vast array of domains including attention, memory, psychomotor speed and executive functions (Bilder et al. 1992; Saykin et al. 1994; Heinrichs & Zakzanis 1998; Heydebrand et al. 2004; Galderisi et al. 2009). In general, unaffected relatives of schizophrenia patients also exhibit deficits in cognition, but usually to a lesser degree (Cannon et al. 1994; Faraone et al. 1995b; Park et al. 1995; Appels et al. 2003; Breton et al. 2011; Agnew-Blais & Seidman 2013). There is evidence that schizophrenia patients are impaired on tests assessing working memory
(Ucok et al. 2013), executive functions (Ucok et al. 2013), sustained attention (Chen & Faraone 2000), basic processing of sensory stimuli (Schwartz et al. 2001), verbal episodic memory (Cirillo & Seidman 2003) and smooth pursuit eye movements (Thaker et al. 1999).

For this study, we wanted to subdivide the CNV categories in an effort to capture a relationship between CNVs and IQ, cognitive domains and brain structure volume. In addition to dividing CNVs by state (deletion, duplication, no CNV), we also divided CNVs into three clinically relevant categories: abnormal, indeterminate and normal. For our analysis of cognitive function we grouped the large number of individual tests into cognitive domains: visuospatial skills, processing speed and attention, language ability, problem solving and verbal memory. We also considered intelligence derived from the Weschsler intelligence scale: FSIQ, VIQ and PIQ. For brain structure volume, we focused on the different tissue types in the cerebral cortical lobes: frontal lobe gray and white matter, temporal lobe gray and white matter, parietal lobe gray and white matter, occipital lobe gray and white matter, and the volume of cerebral spinal fluid in the cerebral ventricles. We hypothesized that more severe CNVs, either deletions or abnormal CNVs, compared to duplications or normal CNVs, would be associated with more severe neurocognitive deficits as well as more pronounced abnormalities in brain structure volume.

**Materials and Methods**

**Study Sample**

This study sample, the Iowa sample, includes individuals with schizophrenia who have been assessed with neuropsychological testing, extensively characterized for
phenotypic information and who have provided DNA. In addition, these individuals have been scanned with high-resolution brain MRI. Subjects have primarily been recruited from Eastern Iowa and neighboring states. Study inclusion requires a DSM-IV diagnosis of schizophrenia and age at intake of 16-45. Exclusion criteria include the presence of a neurological disorder, ID, history of severe head trauma, or active substance dependence or abuse (except for nicotine). The ethnic composition of the Iowa sample is reflective of the largely Caucasian population of Iowa: 94% Caucasian, 3.2% Hispanic, 1.3% African American, 1.1% Asian, and 1% other.

Individuals with schizophrenia receive an extensive assessment battery that records diagnosis, severity of symptoms and, for many, the course of the illness over time using two semi-structured interview instruments: the CASH and the Psychiatric Symptoms You Currently Have (PSYCH). The CASH is a structured interview which provides the basis for both DSM-criteria-based diagnosis and documentation of phenomenology (Andreasen 1985; Andreasen et al. 1992). It includes a comprehensive assessment of positive and negative symptoms, mood symptoms, substance abuse history, socio-demographic information and premorbid adjustment. The CASH includes the Scale for the Assessment of Negative Symptoms (SANS), the Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen 1983, 1984) and the Calgary Depression Inventory (Addington et al. 1990). The SANS, SAPS and Calgary provide “standard rating scale” measures of symptoms when the imaging and cognitive data are obtained. The PSYCH is an instrument that provides baseline information about psychosocial function and treatment during the time period prior to intake (Andreasen et al. 1986; Andreasen 1987). We also gather information describing exposure to nicotine; medication history; medication side effects using the Abnormal Involuntary Movement Scale (AIMS), Simpson-Angus, and Barnes akathisia scales; a neurological exam for soft signs; and family history of psychiatric illness with the Family History-Research Diagnostic Criteria (FHRDC) (Andreasen et al. 1986).
Brain Image Acquisition

In the Iowa sample, research subjects undergo state-of-the-art, multimodal MRI protocols with thin slices acquired across the whole brain. We have been gathering images (and DNA) from subjects for over 20 years, so our protocol has been updated once. Before December 1999 subjects were scanned using the PR1 protocol, while subsequent to January of 2000 we implemented the PR2 protocol. Images from the two protocols are typically analyzed together with the inclusion of protocol as a covariate.

**PR1 parameters:** Images are acquired on a 1.5 Tesla GE scanner. A T1 image is obtained as a 3D volume in the coronal plane using a spoiled GRASS sequence with the following parameters: 124 1.5 mm coronal slices, TE 5 ms, TR 24 ms, flip angle 40°, NEX 2, FOV 26 cm and matrix 256 x 192 with an echo train length = 8. Proton density (PD) and T2-weighted images are acquired with the following parameters: 3 mm coronal slices, TE 36 ms (for PD) or 96 ms (for T2), TR 3000 ms, NEX 1 FOV 26 cm, 256 x 192 matrix, and an echo train length = 8.

**PR2 parameters:** This protocol acquires T1 and T2 weighted sequences. The T1 sequence is obtained as a 3D volume in the coronal plane using a spoiled GRASS sequence with the following parameters: TE = 6 ms, TR = 20 ms, flip angle = 30°, FOV = 160x160x192 mm, matrix = 256x256x124 and NEX=2. The T2 images are acquired using a 2D fast spin-echo sequence in the coronal plane with the following parameters: TE = 85 ms, TR = 4800 ms, slice thickness/gap = 1.8/0.0 mm, FOV = 160x160 mm, matrix = 256x256, NEX = 3, number of echoes = 8 and number of slices = 124.

After acquisition, MR images are transferred to our Image Processing Lab where they are analyzed using the Brain Research: Analysis of Images, Networks and Systems (BRAINS) software package. The BRAINS package employs a multi-spectral discriminant analysis method to distinguish gray matter, white matter and cerebral spinal
fluid (Harris et al. 1999). To delineate larger regions of the brain such as cerebral lobes, BRAINS uses an automated Talairach-based structure identification algorithm. These regions include the frontal, temporal, parietal, and occipital lobes; ventricles; and cerebellum (Andreasen et al. 1996). This method relies on a piecewise linear transformation of the Talairach grid onto the subject of interest and produces measures that have been widely used (Andreasen et al. 1996; Swayze et al. 1996; Kates et al. 1999; Lancaster et al. 2000). Lastly, BRAINS uses an artificial neural network (ANN) to define several subcortical and cerebellar structures (Magnotta et al. 1999). Interrater and intrarater reliabilities were established both by technicians in Sweden and at the University of Iowa (Agartz et al. 2001)

**Neurocognitive Assessment**

All study subjects were administered a comprehensive neuropsychological assessment by psychometrists trained in standardized assessment and scoring procedures. The only neuropsychological assessment used in this study was the first assessment made at the time of intake into the study. Testing took place in a quiet room at times when the patients were most cooperative and alert and after staff determined that the severity of their symptoms would not interfere with testing. Patients with severe psychotic or disorganized symptoms were tested after being stabilized with antipsychotic medication. Testing generally took four hours to complete and, when necessary, occurred over several sessions. Subjects who were neuroleptic naive were tested before the administration of psychiatric medications. Those who were already receiving medication at the time of intake were tested while taking their medication at the prescribed dose.

In order to efficiently analyze the cognitive functioning of study subjects, neuropsychological test variables were grouped into five cognitive domains on the basis
of a priori theoretical considerations (Green et al. 2004; Hill et al. 2004; Milev et al. 2005). These domains were: verbal memory, processing speed and attention, language skills, visuospatial skills, and problem solving. By grouping our neuropsychological battery into cognitive domains, we diminished multiple testing in subsequent statistical analyses. These theoretical groupings were then tested for internal reliability using Cronbach’s alpha analyses. The neuropsychological tests within each of the five cognitive domains have good internal consistency (Cronbach’s alpha ≥ 0.75). The five domains and their component tests are shown in Table 3-1.

The raw test scores for each of the 27 neuropsychological test variables were converted to standardized scores on the basis of norms established by use of 546 healthy comparison subjects. To provide a consistent and uniform basis for establishing the norms, control subjects were recruited through newspaper advertisements from the same geographical area from which the schizophrenia subjects were ascertained using the criteria described above. The same psychometrists who administered the neuropsychological battery to our patients also tested these healthy comparison subjects. By definition, the healthy group had z scores with means set to zero and standard deviations set to one. Patient scores were then derived from comparison to these control z scores. Each domain score is the summed average of its component neuropsychological test variable standardized scores. A larger negative score always indicates poorer performance below the mean. Not all patients completed all tests and the domain scores are the means of all non-missing contributing variables.

CNV Detection

Subject DNA was extracted from whole blood samples using a standard salt precipitation technique, and run on Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara,
CA) according to recommendations from the manufacturer. Raw intensity data from the arrays were analyzed for copy number changes using the publicly available software program Affymetrix Genotyping Console 4.0 (Affymetrix). Data quality was assessed with both CQC, which measures the quality of probe intensity data, and MAPD which provides a per-chip estimate of variability. Samples with a CQC < 0.4 (or a batch mean CQC < 1.7) were excluded from the study. Likewise if the MAPD value was over 0.34, samples were re-processed and hybridized to a new array or they were excluded. To be consistent with other published studies (Glessner et al. 2010; Van Den Bossche et al. 2013), we chose ten probes as the minimum number required to make a CNV call. All CNVs were manually inspected using the UCSC Genome Browser (http://genome.ucsc.edu/) (Kent et al. 2002) and cross referenced with the DGV (http://projects.tcag.ca/variation/) (Iafrate et al. 2004). The CNVs were divided into three categories: abnormal if they have a previous association with disease or are novel, genic and are either a deletion > 500 kb or a duplication > 1 Mb.; indeterminate if they are novel, contain genes, but do not meet the size requirements for abnormal; or normal for all other CNVs.

Analyzing Relationships between CNVs and Both Brain Structure Volumes and Cognition

Copy number variants were assessed to determine whether they influence brain structure volumes and/or cognition; they were divided into three categories: abnormal, indeterminate and normal. Using ANCOVA analysis CNV genotype effects were tested on 1) the five cognitive domains (visuospatial abilities, processing speed/attention, language, problem solving and verbal memory), 2) gray and white matter volumes of the cerebellum and four cortical lobes and 3) CSF volume of the lateral ventricles. Cognitive
ability covariates included age and gender. Brain structure volume covariates included age, gender, diagnosis, scan protocol and total brain compartment volume. In addition, CNV state (duplication, deletion, no CNV) was also tested against the five cognitive domains.

Determining an appropriate significance level for tests performed in this study is complicated by a number of factors. For example the brain structure volumes that we test are correlated with each other, so treating them as independent, as in a Bonferroni correction, would be excessively conservative. Even after removing the variance due to the covariates, the gray matter measures are correlated with \( r \) values of 0.13–0.81, while the white matter volumes are inter-correlated with \( r \) values of 0.41–0.90, all of which are highly significant (except for frontal and occipital gray matter volumes which are not correlated). Similar relationships are observed with the cognitive measures. In light of this, we choose a p-value for significance of 0.01.

**Results**

Only subjects with cognitive or brain structure data were included in the analysis (Table 3-2). From the analyses of covariance, we found no difference in IQ or cognitive measures across CNV state (duplication, deletion or no CNV) (Table 3-3). Nor did we find a difference across the three clinical CNV categories (abnormal, indeterminate or normal) for the measures of IQ or cognitive abilities (Table 3-4). Our findings were similarly negative for brain structure volumes; we found no differences in any regional volume across CNV clinical state or category (Table 3-5 & 3-6).
Discussion

We have examined the role of genomic CNVs and whether they affect deeper aspects of the schizophrenia phenotype such as IQ, cognitive abilities and brain structure with largely negative results. With this approach we had hoped to learn more about how specific genetic defects such as CNVs work to produce disease by examining them against endophenotypic measures. Specifically, we had hypothesized that the identification of characteristic patterns of endophenotypic deficits that are associated with genetic defects could help resolve some of the uncertainty surrounding CNVs and their clinical classifications which remain imprecise. We had a very large category of indeterminate CNVs and our results would have benefited from a more precise understanding of the true pathogenicity of each CNV. Studies like ours would greatly benefit from more developed endophenotypic profiles that accurately distinguish between normal and truly abnormal CNVs.

There is very little in the literature regarding CNVs and endophenotypes. One study tested for associations between global CNV burden and total brain, white matter and gray matter volume in schizophrenia subjects; no relationship was found (Terwisscha van Scheltinga et al. 2012). However, the study only examined global CNV burden which has little basis for being clinically or biologically relevant. Similarly, they only examined global, rather than regional, measures of brain structure volume. Another study from the same group tested for associations between CNVs (total, deletion and duplications) and total IQ (van Scheltinga et al. 2013). One other report tested for associations between intelligence and total CNV load and also reported negative results (MacLeod et al. 2012). In contrast to these previous publications, we identified CNVs most likely to be clinically relevant and examined multiple brain regions and multiple measures of cognitive abilities.
In spite of our more detailed and clinically relevant approach, we found no significant associations between CNVs and lower IQ, poorer neurocognitive performance or more severe deficits in brain structure volume in individuals with schizophrenia. This was true whether we classified CNVs according to type (deletion, duplication or no CNV) or severity (abnormal, indeterminate or normal). There was no evidence for relationships if CNV category of gain/normal/loss was treated as a linear predictor (three copies/two copies/one copy) or if the quantitative measures were ranked to produce non-linear ANCOVAs. A number of factors might account for these inconclusive results. First, while we believe that is unlikely, there is the possibility that no association exists between CNVs and cognitive or brain structure measures. Second, patient selection excluded individuals with ID, known genetic defects or neurological disorders. Individuals with these features might be more likely to have CNVs and could certainly also have schizophrenia, but would not be included in our study. Third, our sample size, though larger than others for this type of study, nonetheless remains modest and may simply not be large enough to reveal relationships. Lastly, it may also be the case that any particular CNV will have an idiosyncratic effect and that our approach of grouping CNVs by type or severity conceals these effects.

Despite these non-significant results, we still advocate for the continued research into the effects of CNVs on quantitative endophenotypic traits in schizophrenia. The most obvious change in approach would be to increase sample sizes so that larger groups of particular categories and larger groups of individual CNVs can be examined. Another modification would be gene-based grouping of CNVs. It may be that one gene within any CNV produces its neuropsychiatric effect, and that classifying CNVs by related groups of genes could lead to a more biologically meaningful and homogeneous analysis. Finally, it would beneficial to the field if we could find utility for a middle ground between global and clinical classification of CNVs. There is growing evidence that a combination of
CNVs are associated with a phenotype of interest, therefore, an analysis accurately identifying all relevant CNVs could detect effects that are currently being missed.
Table 3-1. Neurocognitive test battery. Five cognitive domains and their component tests.

<table>
<thead>
<tr>
<th>Neuropsychological Function</th>
<th>Schizophrenia subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Domains</td>
<td></td>
</tr>
</tbody>
</table>
| Verbal Memory               | Rey Auditory Verbal Learning (RAVLT)  
Trials 1-5; Trial 7; Delayed recall  
Wechsler Memory-Revised (WMS-R)  
Logical Memory-immediate recall  
Logical Memory-delayed recall |
| Processing Speed and Attn   | WAIS-R  
Digit Span and Digit Symbol  
Trail Making A and B  
Stroop Color and Word, Trials 1, 2, 3  
Circle A letter cancellation, time to complete |
| Problem Solving             | Wisconsin Card Sorting Test  
Categories attained; perseverative errors  
Shipley Institute of Living Scale-Abstractions  
WAIS-R  
Comprehension, Similarities  
Picture completion, picture arrangement |
| Language Skills             | WAIS-R Vocabulary  
Shipley Vocabulary  
MAE Oral Word Association |
| Visuospatial Skills         | Rey-Osterrieth Complex Figure Test-copy  
WAIS-R Block Design, Object Assembly  
Benton Judgement of Line Orientation |
Table 3-2. Table provides number of subjects included in the analysis.

<table>
<thead>
<tr>
<th></th>
<th>FSIQ, VIQ, PIQ</th>
<th>Cognitive Domains</th>
<th>MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>13</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>18</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>Normal</td>
<td>99</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>Gain</td>
<td>21</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>Loss</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Normal</td>
<td>99</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>123</td>
<td>131</td>
</tr>
</tbody>
</table>
Table 3-3. Summary of results from ANCOVA between CNV state and intelligence or cognitive measures.

<table>
<thead>
<tr>
<th>CNV State</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-test</td>
</tr>
<tr>
<td>Loss</td>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wechsler Intelligence Rating</th>
<th></th>
<th></th>
<th>0.58</th>
<th>0.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal IQ</td>
<td>88.9</td>
<td>92.1</td>
<td>94.2</td>
<td></td>
</tr>
<tr>
<td>Performance IQ</td>
<td>88.0</td>
<td>91.2</td>
<td>92.2</td>
<td></td>
</tr>
<tr>
<td>Full-scale IQ</td>
<td>87.5</td>
<td>90.9</td>
<td>93.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive measures</th>
<th></th>
<th></th>
<th>0.58</th>
<th>0.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visuospatial skills (VSPAB)</td>
<td>-0.50</td>
<td>-0.91</td>
<td>-0.98</td>
<td></td>
</tr>
<tr>
<td>Speed/attention (SPATT)</td>
<td>-1.07</td>
<td>-1.21</td>
<td>-1.44</td>
<td></td>
</tr>
<tr>
<td>Language (Lang)</td>
<td>-1.28</td>
<td>-1.23</td>
<td>-1.24</td>
<td></td>
</tr>
<tr>
<td>Problem solving (FPROB)</td>
<td>-1.54</td>
<td>-1.07</td>
<td>-1.07</td>
<td></td>
</tr>
<tr>
<td>Verbal memory (FVM)</td>
<td>-1.40</td>
<td>-1.22</td>
<td>-1.42</td>
<td></td>
</tr>
</tbody>
</table>
Table 3-4. Summary of results from ANCOVA between CNV category and intelligence or cognitive measures.

<table>
<thead>
<tr>
<th>CNV Category</th>
<th>Normal</th>
<th>Indeterminate</th>
<th>Abnormal</th>
<th>Test Statistic</th>
<th>F-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wechsler Intelligence Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>92.1</td>
<td>93.6</td>
<td>91.0</td>
<td></td>
<td>0.15</td>
<td>0.86</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>91.3</td>
<td>94.9</td>
<td>85.5</td>
<td></td>
<td>1.72</td>
<td>0.18</td>
</tr>
<tr>
<td>Full-scale IQ</td>
<td>90.9</td>
<td>93.6</td>
<td>88.1</td>
<td></td>
<td>0.65</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>Cognitive measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visuospatial skills (VSPAB)</td>
<td>-0.91</td>
<td>-0.83</td>
<td>-0.82</td>
<td></td>
<td>0.05</td>
<td>0.95</td>
</tr>
<tr>
<td>Speed/attention (SPATT)</td>
<td>-1.20</td>
<td>-0.95</td>
<td>-1.80</td>
<td></td>
<td>2.9</td>
<td>0.06</td>
</tr>
<tr>
<td>Language (Lang)</td>
<td>-1.23</td>
<td>-1.26</td>
<td>-1.25</td>
<td></td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Problem solving (FPROB)</td>
<td>-1.07</td>
<td>-1.01</td>
<td>-1.50</td>
<td></td>
<td>0.88</td>
<td>0.42</td>
</tr>
<tr>
<td>Verbal memory (FVM)</td>
<td>-1.22</td>
<td>-1.27</td>
<td>-1.61</td>
<td></td>
<td>0.84</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Table 3-5. Summary of results from ANCOVA between CNV state and brain structure volume.

<table>
<thead>
<tr>
<th>MRI Measures</th>
<th>CNV State</th>
<th></th>
<th></th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Loss</td>
<td>None</td>
<td>Gain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-test</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>Frontal lobe gray matter</td>
<td>251.0</td>
<td>247.0</td>
<td>248.1</td>
<td>0.73</td>
</tr>
<tr>
<td>Temporal lobe gray matter</td>
<td>156.7</td>
<td>157.7</td>
<td>160.0</td>
<td>1.25</td>
</tr>
<tr>
<td>Parietal lobe gray matter</td>
<td>132.9</td>
<td>131.4</td>
<td>132.8</td>
<td>0.53</td>
</tr>
<tr>
<td>Occipital gray matter</td>
<td>72.5</td>
<td>73.0</td>
<td>74.9</td>
<td>1.45</td>
</tr>
<tr>
<td>Frontal lobe white matter</td>
<td>177.4</td>
<td>174.5</td>
<td>173.4</td>
<td>0.37</td>
</tr>
<tr>
<td>Temporal lobe white matter</td>
<td>72.2</td>
<td>70.9</td>
<td>71.3</td>
<td>0.37</td>
</tr>
<tr>
<td>Parietal lobe white matter</td>
<td>109.7</td>
<td>107.1</td>
<td>107.2</td>
<td>0.45</td>
</tr>
<tr>
<td>Occipital lobe white matter</td>
<td>40.8</td>
<td>41.4</td>
<td>39.5</td>
<td>0.81</td>
</tr>
<tr>
<td>Ventricles</td>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
<td>1.48</td>
</tr>
</tbody>
</table>
Table 3-6. Summary of results from ANCOVA between CNV category and brain structure volume.

<table>
<thead>
<tr>
<th>MRI Measures</th>
<th>CNV Category</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>Frontal lobe gray matter</td>
<td>248.1</td>
<td>248.1</td>
</tr>
<tr>
<td>Temporal lobe gray matter</td>
<td>157.8</td>
<td>159.6</td>
</tr>
<tr>
<td>Parietal lobe gray matter</td>
<td>131.7</td>
<td>132.1</td>
</tr>
<tr>
<td>Occipital gray matter</td>
<td>73.0</td>
<td>74.9</td>
</tr>
<tr>
<td>Frontal lobe white matter</td>
<td>175.2</td>
<td>176.1</td>
</tr>
<tr>
<td>Temporal lobe white matter</td>
<td>71.0</td>
<td>73.0</td>
</tr>
<tr>
<td>Parietal lobe white matter</td>
<td>107.2</td>
<td>109.4</td>
</tr>
<tr>
<td>Occipital lobe white matter</td>
<td>40.8</td>
<td>41.6</td>
</tr>
<tr>
<td>Ventricle</td>
<td>2.8</td>
<td>2.7</td>
</tr>
</tbody>
</table>
CHAPTER IV

CHILDHOOD-ONSET SCHIZOPHRENIA CASE WITH 2.2 MB DELETION AT CHROMOSOME 3P12.2-P12.1 AND TWO LARGE CHROMOSOMAL ABNORMALITIES AT 16Q22.3-Q24.3 AND XQ23-Q28

Introduction

Childhood-onset schizophrenia (COS) describes an individual meeting the adult DSM-IV-TR criteria for schizophrenia, but with onset of psychosis before the age of 13 (Gordon et al. 1994; Russell 1994; Ahn et al. 2013). Childhood-onset schizophrenia is rare, comprising less than 1% of all schizophrenia cases (Gordon et al. 1994; Spencer & Campbell 1994; Ahn et al. 2013; Schimmelmann et al. 2013). Compared to adult-onset schizophrenia, which often has an acute onset of disease, COS is typically characterized by an insidious onset of psychosis in the context of a history of language and motor delays that leads to a marked deterioration in functioning (Gordon et al. 1994; Russell 1994; Alaghband-Rad et al. 1995; Hollis 1995; Nicolson et al. 2000). Additional premorbid symptoms in COS can include social abnormalities, a variety of cognitive deficits and behavioral disturbances (Russell 1994; Alaghband-Rad et al. 1995; Nicolson et al. 1999).

Schizophrenia vulnerability appears to be largely determined by genetic factors, and COS conforms to this pattern. Studies of environment have found no evidence for an increased role of non-genetic factors in etiology (Nicolson et al. 2000); instead, COS cases have a higher frequency of large chromosomal abnormalities than is typically found in adult-onset cases (DeLisi et al. 1994; Nicolson et al. 1999; Nicolson et al. 2000; Walsh et al. 2008; International Schizophrenia et al. 2009). Here we describe the use of CMA technology to discover two large, de novo chromosomal duplications and an inherited 2.2
Mb deletion in a Caucasian female with COS. We hypothesize that these three large chromosomal aberrations act in concert to create genetic susceptibility for COS in this patient.

Clinical Report

At the time of examination by our research center, the patient was an 18-year-old female referred to us because of a clinical history consistent with COS. The patient was the second-born of six siblings to healthy nonconsanguineous parents. Both the pregnancy and delivery were normal with a birth weight of 6 lbs. Parents described the patient as walking at 14 months, and though they could not recall exactly when she began to speak, they reported that it was within a normal developmental time frame. Despite these normal milestones, the parents also reported that the patient had a difficult time with independent locomotion, falling frequently, being “clumsy” and having generally poor motor coordination. The parents also reported a long history of deficits in age-appropriate social skills. The patient had a limited range of social interaction compared to her siblings and developed few friendships with same-age peers. She seldom engaged in conversations, was considered for a diagnosis in the autism spectrum and in sixth grade was identified by school personnel as having “a behavioral impairment”. Academically the patient did well in early elementary school, scoring High Average in most performance areas on the Stanford Achievement Tests. She then experienced deterioration so that by sixth grade she began receiving special education services because she was no longer able to perform basic math or reading skills. At this time a sharp regression in motor skills was also noted.

At age 12 the patient was diagnosed with COS by a local psychiatrist. At age 16 the patient was referred to a regional medical center for a more extensive psychiatric evaluation which yielded the same diagnosis, and she was seen in our center at the age of
18. Neuropsychological testing at that time showed a FSIQ = 67, VIQ = 70 and PIQ = 65. The patient had global deficits including impairments in memory, language and visual spatial skills, with severe impairments in attention and concentration. The patient often appeared to be attending to internal stimuli. During assessment tasks, she would look away from test material then look back to the examiner and giggle or laugh. She would occasionally make inappropriate and bizarre comments and also exhibited inappropriate affect, exhibiting at various times giggling, laughter, crying and agitation, all without clear stimuli. She was distractible, impulsive and would stare into space and whisper to herself under her breath as if attending to voices. She exhibited avolition, paucity of speech and occasional echolalia. She had significant difficulties with abstract reasoning and problem solving, and her judgment was poor. Based on this history and examination, the patient was given a diagnosis of COS of the disorganized type, with an additional diagnosis of mild ID.

Methods and Results

Chromosomal Microarray Study

Peripheral blood samples from the patient, her mother and father were collected with informed consent through a protocol approved by the University of Iowa IRB. This research has been conducted in accordance with the Declaration of Helsinki. Subject DNA was extracted using a standard salt precipitation technique and run on Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA) according to the recommendations of the manufacturer. Raw intensity data from the Affymetrix 6.0 arrays was analyzed for copy number changes using the publicly available analysis software program Affymetrix Genotyping Console 4.0 (Affymetrix). Copy number variant parameters included a 10
probe minimum and a quality control MAPD value < 0.34. All CNVs were manually annotated using the UCSC Genome Browser (UCSC Mar. 2006 (NCBI36/hg18) assembly) (http://genome.ucsc.edu/). The CMA study showed three large chromosomal abnormalities: a 2.2 Mb deletion of chromosome 3p12.2-p12.1 that was inherited from the father (Figure 4-1), a de novo 17.6 Mb duplication of chromosome 16q22.3-q24.3 (Figure 4-2) and a de novo 43 Mb deletion of chromosome Xq23-q28 (Figure 4-3). None of the three chromosomal abnormalities were found in control individuals according to the DGV.

Follow-up molecular confirmation was performed with a NimbleGen 2.1 million probe whole-genome tiling array (Promega, Madison, WI), processed according to the instructions of the manufacturer. The NimbleGen arrays were scanned using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA), signal intensity data were analyzed using NimbleScan software (NimbleGen) and CNV calling was performed using Nexus software (BioDiscovery, El Segundo, CA). All three CNVs were identified on the NimbleGen array.

**Discussion**

The general course of this patient’s clinical history is consistent with COS. The patient had an insidious onset of illness with deterioration in early childhood across multiple domains of functioning including social interaction, behavior and general intellectual abilities. Psychotic and disorganized symptoms did not present until the age of 12, but have persisted into adulthood. Childhood onset schizophrenia often initially resembles a pervasive developmental disorder, but like this patient, the emergence in pre-teen years of psychotic symptoms that remain across the life span place the diagnosis in the psychotic spectrum (Ahn et al. 2013).
Cases of COS are noted for having a high rate of cytogenetic abnormalities. An early study examined differences between COS patients with and without cytogenetic abnormalities and found that those with aberrations had lower PIQs and more impairments of premorbid language, motor and social development (Nicolson et al. 1999). *De novo* CNVs are associated with schizophrenia (Xu et al. 2008) and specifically with COS (Malhotra et al. 2011). These findings are consistent with our patient who received special education and exhibited behavioral and social abnormalities. The patient described here had FSIQ = 67, VIQ = 70 and PIQ = 65, which are considered low (Russell 1994). This patient harbored three large chromosomal abnormalities: a paternally inherited 2.2 Mb deletion of chromosome 3p12.2-p12.1, a *de novo* 17.6 Mb duplication of chromosome 16q22.3-q24.3 and a *de novo* 43 Mb deletion of chromosome Xq23-q28.

The paternally inherited 2.2 Mb deletion at chromosome 3p12.2-p12.1 does not contain any genes (Figure 4-4). Approximately 400 kb distal is the gene CADM2 which is involved in the formation of presynaptic terminals and induces functional synapses in the CNS (Tanabe et al. 2013). Knock-out mice deficient for Cadm1 show behavioral abnormalities such as excessive aggression and anxiety (Tanabe et al. 2013).

Full trisomy of chromosome 16 is never seen in a live birth and accounts for 15% of spontaneous abortions due to chromosomal abnormalities (Hassold & Jacobs 1984). Partial trisomy 16 is rare, most likely because of lethality, and phenotype associations are complicated by the frequent presence, as in our case, of additional monosomies or trisomies (Lonardo et al. 2011). Patients with proximal 16q duplications are often characterized by a variable phenotype of speech delay, DD, learning difficulties and behavioral problems (Barber et al. 2006). More distal duplications, as in our patient, are associated with an even wider array of developmental and organ system abnormalities (Brisset et al. 2002). The 16.7 Mb duplicated region of chromosome 16q22.3-q24.3 in our patient contains 163 genes (Figure 4-5). Twenty of the genes in the region are of interest
due to previous associations with schizophrenia or autism, high or specific brain expression or an association with other neuropsychiatric disorders (Table 4-1).

The 42 Mb deletion of chromosome Xq23-q28 in this patient is categorized as one of two critical regions responsible for normal ovarian function (Figure 4-6) (Toniolo 2006; Mercer et al. 2013). Specifically, the Xq26.2-q28 region is proposed to be associated with premature ovarian failure (Mercer et al. 2013). There is no definitive association between Xq perturbations and cognitive impairment. There are a few reports of patients with large Xq disruption and an associated cognitive impairment, however, those patients had translocations that affected other autosomal chromosomes (Mercer et al. 2013). It is important to note that this patient does not have a diagnosis of Turner syndrome. A deletion on the long arm of the X chromosome is not sufficient for a clinical diagnosis of Turner syndrome, which requires deletion of the \textit{SHOX} gene on Xp22.3 (Skorka et al. 2012). Several case studies of female patients with deletions of the long arm of chromosome X reveal few features in common with Turner syndrome patients (Mercer et al. 2013).

In conclusion, only a small number of cases of COS attributable to chromosomal abnormalities have been reported. We contribute to this literature by describing a young female patient with COS and three large chromosomal abnormalities that we assert are acting in concert with each other to confer COS in this patient.
Table 4-1. Summary of 20 genes located in the 16.7 Mb duplication at chr16q22.3-q24.3 that are of interest.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP2C2</td>
<td>Associated with speech and language impairment.</td>
</tr>
<tr>
<td>JPH3</td>
<td>Brain specific expression and appears to have an active role in certain neurons involved in motor coordination and memory. Mutations in this gene are the cause of Huntington disease-like type 2.</td>
</tr>
<tr>
<td>CDH15</td>
<td>Defects in this gene are the cause of mental retardation autosomal dominant type 3.</td>
</tr>
<tr>
<td>FA2H</td>
<td>Associated with neurodegeneration with brain ion accumulation.</td>
</tr>
<tr>
<td>ZNRF1</td>
<td>Plays a role in the establishment and maintenance of neuronal transmission and plasticity. Regulates Schwann cells differentiation by mediating ubiquitination of GLUL. Promotes neurodegeneration by mediating 'Lys-48'-linked polyubiquitination and subsequent degradation of AKT1 in axons: degradation of AKT1 prevents AKT1-mediated phosphorylation of GSK3B, leading to GSK3B activation and phosphorylation of DPYSL2/CRMP2 followed by destabilization of microtubule assembly in axons.</td>
</tr>
</tbody>
</table>
Table 4-1 — Continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNTNAP4</td>
<td>This gene product belongs to the neurexin family, members of which function in the vertebrate nervous system as cell adhesion molecules and receptors.</td>
</tr>
<tr>
<td>DYNLRB2</td>
<td>High expression in brain.</td>
</tr>
<tr>
<td>GAN</td>
<td>Deletion of this gene associated with autism.</td>
</tr>
<tr>
<td>CMIP</td>
<td>Associated with speech and language impairment. Deletion of this gene associated with autism.</td>
</tr>
<tr>
<td>CDH13</td>
<td>This gene encodes a member of the cadherin superfamily. The protein lacks the cytoplasmic domain characteristic of other cadherins, and so is not thought to be a cell-cell adhesion glycoprotein. This protein acts as a negative regulator of axon growth during neural differentiation. Expressed at higher levels in adult brain than in developing brain. Associated with working memory and ADHD. Previous association with autism and schizophrenia.</td>
</tr>
<tr>
<td>NECAB2</td>
<td>Primary expression in the brain. Neuronal Ca (2+) binding protein. Previous association with autism.</td>
</tr>
<tr>
<td>KIAA0513</td>
<td>Widely expressed, highest levels in cerebellum, brain cortex, hippocampus, pons, putamen and amygdala. Highly expressed in neurons, but also present in glial cells. Slightly higher expression in the dorsolateral prefrontal cortex of schizophrenic patients compared to control individuals.</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ZDHHC7</td>
<td>Palmitoyltransferase with broad specificity. Palmitoylates SNAP25 and DLG4/PSD95. May palmitoylate GABA receptors on their gamma subunit (GABRG1, GABRG2 and GABRG3) and thereby regulate their synaptic clustering and/or cell surface stability.</td>
</tr>
<tr>
<td>SLC7A5</td>
<td>Involved in the transport of L-DOPA across the blood–brain barrier, and that of thyroid hormones triiodothyronine (T3) and thyroxine (T4) across the cell membrane in tissues such as placenta. Plays a role in neuronal cell proliferation (neurogenesis) in brain. Involved in the uptake of methylmercury (MeHg) when administered as the L-cysteine or D,L-homocysteine complexes, and hence plays a role in metal ion homeostasis and toxicity. May play an important role in high-grade gliomas. Mediates blood-to-retina L-leucine transport across the inner blood-retinal barrier which in turn may play a key role in maintaining large neutral amino acids as well as neurotransmitters in the neural retina.</td>
</tr>
<tr>
<td>ZFPM1</td>
<td>Mainly expressed in hematopoietic tissues. Also, expressed in adult cerebellum, stomach, lymph node, liver and pancreas.</td>
</tr>
<tr>
<td>TCF25</td>
<td>A member of the basic helix-loop-helix (bHLH) family of transcription factors that are important in embryonic development. In the embryo, widely expressed with highest levels in brain.</td>
</tr>
<tr>
<td><strong>Table 4-1 — Continued</strong></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td></td>
</tr>
</tbody>
</table>

| **TUBB3** | This gene encodes a class III member of the beta tubulin protein family. Beta tubulins are one of two core protein families (alpha and beta tubulins) that heterodimerize and assemble to form microtubules. This protein is primarily expressed in neurons and may be involved in neurogenesis and axon guidance and maintenance. |
| **ZNF778** | This gene is a candidate gene for autism and variable cognitive impairment in the 16q24.3 microdeletion syndrome. |
| **ANKRD11** | Mutations in this gene have been associated with KBG syndrome, which is characterized by macrodontia, distinctive craniofacial features, short stature, skeletal anomalies, global developmental delay, seizures and intellectual disability. Deletion has been associated with autism. |
| **ACSF3** | Defects in ACSF3 are the cause of combined malonic and methylmalonic aciduria (CMAMMA) [MIM:614265]. A metabolic disease characterized by malonic and methylmalonic aciduria, with urinary excretion of much larger amounts of methylmalonic acid than malonic acid, in the presence of normal malonyl-CoA decarboxylase activity. Clinical features include coma, ketoacidosis, hypoglycemia, failure to thrive, microcephaly, dystonia, axial hypotonia and/or developmental delay, and neurologic manifestations including seizures, psychiatric disease and/or cognitive decline. |
Figure 4-1. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 2.2 Mb deletion at Chr3p12.2-p12.1. The individual represented in blue (top) is the patient and the individual represented in purple (bottom) is the father. The top of the panel shows the entire chromosome 3 and the bottom shows a zoomed in view of the deletion.
Figure 4-2. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 16.7 MB duplication at chr16q22.3-q24.3. The patient is represented in blue. The top of the panel shows the entire chromosome 16 and the bottom shows a zoomed in view of the duplication.
Figure 4-3. Screen capture from Affymetrix Genotyping Console Browser software displaying graphical output of Affymetrix SNP 6.0 array data for 43 Mb deletion at chXq23-q28. The patient is represented in blue. The top of the panel shows the entire X chromosome and the bottom shows a zoomed in view of the deletion.
Figure 4-4. Chr3p12.2-p12.1 deletion coordinates in the patient and in the father. UCSC screen capture with patient and father both in red. The deletion, with coordinates determined by Affymetrix Genotyping Console at chr3:82554810-84777050, does not overlap any UCSC genes.
Figure 4-5. Chr16q22.3-q24.3 duplication coordinates in the patient. UCSC screen capture with patient in blue. The duplication coordinates determined by Affymetrix Genotyping Console are chr16: 71224799-88815024. The duplication overlaps 163 genes. For a summary of the most relevant genes in this region, see Table 4-3.
Figure 4-6. ChrXq23-q28 deletion coordinates in the patient. UCSC screen capture with the patient in red. The deletion coordinates as determined by Affymetrix Genotyping Console are chrX: 111858197-154582680.
CHAPTER V
CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

For more than 40 years, familial, twin and adoption studies have provided strong evidence that schizophrenia has genetic underpinnings. Although the role of genetics in schizophrenia is now widely accepted, there have been several shifts in conceptualization of genetic architecture as well as in the research methodology used to find the causative variants. Early on, linkage and association studies were at the forefront of schizophrenia genetic research, then a paucity of positive results combined with technological breakthroughs led to GWAS and CNV studies. Over the last decade, the total combined results of these studies in schizophrenia have only begun to truly scratch the surface of our understanding of the molecular genetic etiology of schizophrenia.

Copy number variants were discovered over a decade ago and initially associated with syndromic disorders such as Smith-Magenis syndrome, Williams-Beuren syndrome, DiGeorge syndrome/VCFS, and Sotos syndrome. The development and advancement of technology made it possible to detect CNVs on a genome-wide level with the surprising result that they are a common source of human genomic variation (Iafrate et al. 2004; Sebat et al. 2004; Redon et al. 2006; Pinto et al. 2007; Conrad et al. 2010). The availability and affordability of high throughput microarray technology also led to the genome-wide identification of CNVs in very large sample sizes. Thus, the research field exploded with CNV discovery studies and now CNV genotyping studies are in their infancy and remain an auspicious source of genetic information for complex diseases such as schizophrenia.

The work presented in this thesis has confirmed previous reports of CNV findings in schizophrenia such as the involvement of large, rare and \textit{de novo} CNVs. Additionally,
this work has made the case for yet another significant shift in the way we approach schizophrenia genetics with the proposal of the multiple-hit CNV model. A multiple-hit CNV model is gaining momentum in the autism literature, but still requires much deeper investigation in schizophrenia. It is our hope that the work presented here will contribute to the growing body of evidence that will ultimately lead to acceptance or rejection of this hypothesis for schizophrenia. The goal of this research is to move beyond mere recognition of CNVs and towards a more precise understanding of the role that they play in human disease.

Summary of Results

A Genome-Wide CNV Analysis of Schizophrenia Reveals a Potential Role for a Multiple-Hit Model Summary

In our genome-wide analysis of CNVs, we studied 166 patients with schizophrenia and 52 psychiatrically healthy controls. Our analysis was unique compared to other CNV analyses in that we considered CNVs of smaller size (< 100 kb) than those traditionally included in published reports. We did not find any significant differences between cases and controls in the rates of occurrence of genic or exonic CNVs, nor did we find significant differences when CNVs were partitioned by size (small, medium or large). This also held true when we restricted analyses to CNVs that were novel (CNVs not reported in controls or DGV).

We did, however, find that a slightly higher proportion of case subjects harbored clinically significant CNVs (conservative CNVs > 1 Mb or clinically recognized as deleterious) when compared to controls. Additionally, we were the first group to consider smaller CNVs (< 100-500 kb) in a multiple-hit model where we observed that a slightly
higher proportion of case subjects had two or more conservative CNVs. Although none of
the results for multiple conservative CNVs reached statistical significance, a trend begins
to emerge where case subjects tend to harbor more multiple conservative CNVs than
controls. We recognize that this analysis is likely to be underpowered given the sample
size and speculate that with a larger sample size the trend may become statistically
significant.

Interestingly, the two case subjects (1 and 2) with the greatest number of
conservative CNVs also share a 15q11.2 BP1-2 deletion. An additional case subject, case
3, who did not harbor additional conservative CNVs, also carried a 15q11.2 BP1-2
deletion. The 15q11.2 BP1-2 deletion has exhibited a nominal association with
schizophrenia (Stefansson et al. 2008; Kirov et al. 2009a; Zhao et al. 2013), and we were
able to contribute another report to the field detailing the presence of this deletion in
patients with schizophrenia. We are also the first group to include MRI data and to report
all other rare CNV findings present in conjunction with the 15q11.2 BP1-2 CNV. For
example, cases 2 and 3 exhibited abnormal MRI findings, including enlarged ventricles,
which are a common finding in individuals with schizophrenia (Weinberger et al. 1979;
Reveley et al. 1982; Wright et al. 2000). In addition, case 3 has a large cavum septum
pellucidum, a feature that is also associated with schizophrenia (Nopoulos et al. 1996;
Nopoulos et al. 1997). While these results alone remain anecdotal, they contribute to a
growing collection of endophenotypes found in patients with 15q11.2 CNVs.

Analysis of Covariance with Cognitive Abilities and Brain
Structure Volume Summary

In our quantitative analysis of endophenotypic traits, we looked for associations
between IQ, measures of cognitive functioning, brain structure volumes and CNVs in
schizophrenia subjects. Copy number variants were divided by CNV state (deletion, duplication or no CNV). We also divided CNVs into one of three categories: abnormal (CNVs that have a previous association with disease or CNVs that are novel, genic and are a deletion > 500 kb or a duplication > 1 Mb), indeterminate (CNVs that are novel, contain genes, but do not meet the size requirements of abnormal), and normal (all other CNVs). We hypothesized that more severe CNVs (i.e., abnormal or deletions) would be associated with lower IQ or poorer neurocognitive performance or more severe deficits in brain structure volume. However, we did not find a significant difference between CNV state or class with respect to IQ, cognitive capabilities or brain structure volume.

Our group is amongst the first to test CNVs against endophenotypic traits. To our knowledge, only three previous studies have tried to associate CNVs with deficits in IQ, cognitive abilities or brain structure and all three reported negative findings. Limitations of these studies are that they only tested global CNV burden (van Scheltinga et al. 2013), only divided CNVs by state (Terwisscha van Scheltinga et al. 2012), or only considered large (> 500 kb) rare CNVs (MacLeod et al. 2012), whereas we divided CNVs by state and by clinical relevance. Additionally, the previous studies only considered global IQ, global cognitive measures or global brain structure volume, while we examined multiple measures of cognitive abilities and multiple brain regions.

Despite the initial negative results for quantitative studies of CNVs and endophenotypic traits of schizophrenia, we believe that a variety of factors can account for this and that this type of research should continue. Although our study had a larger sample size than those studied previously it could still be considered small; larger sample sizes provide greater power for detecting moderate to small effects and may yield more promising results. There are also multiple ways that CNVs could be divided for such an analysis. We divided CNVs by clinical significance to the best of our knowledge; however, many CNVs were categorized as indeterminate. This means that we did not capture all the truly clinically relevant CNVs in our analysis and as research in
schizophrenia progresses, it is likely that more will be characterized. Finally, if a multiple-hit CNV model underlies neuropsychiatric disorders, then we require a more precise understanding of the underlying genetic architecture. In fact, testing for associations between different combinations of CNVs and quantitative endophenotypes such as IQ, neurocognition and brain structure volume could help elucidate underlying genetic interactions.

**Childhood-Onset Schizophrenia Case with 2.2 Mb Deletion at Chromosome 3p12.2-p12.1 and Two Large Chromosomal Abnormalities at 16q22.3-q24.3 and Xq23-q28 Summary**

In our case report we presented a patient with COS and three large chromosomal abnormalities. Childhood-onset schizophrenia is rare and comprises less than 1% of all schizophrenia cases (Gordon et al. 1994; Spencer & Campbell 1994; Ahn et al. 2013; Schimmelmann et al. 2013). The course of illness in our patient was consistent with previous reports of COS, which include insidious disease onset with deterioration in early childhood across multiple domains of functioning. Our patient suffered a decline in social interaction, behavior and general intellectual abilities (requiring special education services) culminating in psychotic and disorganized symptoms at the age of 12 that persisted into adulthood. The patient also exhibited low intelligence with a FSIQ = 67, VIQ = 70 and PIQ = 65.

Childhood-onset schizophrenia appears to be a disorder largely determined by genetic factors and there are many reports of a higher frequency of large chromosomal abnormalities than is typically found in adult-onset cases (DeLisi et al. 1994; Nicolson et al. 1999; Nicolson et al. 2000; Walsh et al. 2008). In addition, *de novo* CNVs are also
associated with COS (Malhotra et al. 2011). Our patient carried three large chromosomal aberrations, including a paternally inherited 2.2 Mb deletion of chromosome 3p12.2-p12.1, a \textit{de novo} 17.6 Mb duplication of chromosome 16q22.3-q24.3 and a \textit{de novo} 43 Mb deletion of chromosome Xq23-q28.

The paternally inherited 2.2 Mb deletion at chromosome 3p12.2-p12.1 did not contain any genes with a previous association with schizophrenia or autism, high or specific brain expression, or an association with other neuropsychiatric disorders. The 16.7 Mb duplicated region of chromosome 16q22.3-q24.3 in our patient contains 163 genes. The 16.7 Mb duplication is challenging to interpret because partial trisomy of the distal end of the long arm of chromosome 16 is rare. More proximal duplications of 16q are often characterized by a variable phenotype of developmental and speech delay, learning difficulties and behavioral problems (Barber et al. 2006), which is similar to the phenotype observed in our patient. However, many of these characteristics did not present themselves until later in childhood for our patient. The 43 Mb deletion of Xq23-q24 is similarly difficult to interpret as it is a region primarily associated with premature ovarian failure (Mercer et al. 2013). No cognitive or behavioral impairments are attributed to deletions in this region, nor is a diagnosis of Turner syndrome, which requires deletions of the \textit{SHOX} gene on Xp22.3 (Mercer et al. 2013). Our patient exhibited three large chromosomal abnormalities, none of which is conclusively individually causative. We hypothesize that, when combined, these chromosomal anomalies act in concert with each other to confer COS risk in this patient.

Implications of Present Findings

The discussion of the results presented in this thesis highlights limitations in current methodologies that need to be addressed in order for the field to move forward.
Taken together, these results underscore the importance for the CNV field to move away from mere CNV discovery and towards detailed characterization and functionality of CNVs themselves. For example, there is growing evidence that recurrent CNVs, such as 15q11.2 CNVs, are relatively more common than individually rare CNVs (those arising by a mechanism other than NAHR). Approximately 10% of the genome is composed of segdups, which are known mediators of NAHR, and 25 out of 130 recognized segdup regions in the genome are associated with known genomic disorders (Sharp et al. 2005). Regions of the genome that are heavily populated with segdups are now recognized as “CNV hotspots” and many recurrent CNVs mediated by segdups such as 15q11.2, 16p11.2, 17q12 and 22q11.2 are consistently implicated in multiple investigations of neuropsychiatric disorders, yet have no clear-cut role in the etiology of these disorders.

One of the most difficult interpretations to make in schizophrenia research is the functional and clinical significance of potentially pathogenic CNVs. Recurrent CNVs have been identified as risk variants for genetically related neurodevelopmental disorders and this has further confounded efforts to classify their true importance. Recent studies have estimated the penetrance of 15q11.2 BP 1-2 deletions for schizophrenia at 2% (Vassos et al. 2010; Kirov et al. 2013). Penetrance for the 15q11.2 BP1-2 deletion has also been estimated for CA, DD, ID and/or autism to fall between 10.4-11% (Kirov et al. 2013; Rosenfeld et al. 2013). However, there are limitations inherent in calculating the penetrance of the 15q11.2 CNV. First, penetrance calculations assume that the 15q11.2 CNV exhibits a monogenic effect on disease risk which is highly unlikely, given the arguments made in this thesis. Second, the penetrance calculations for DD, CA and/or autism are comparably high to estimates for schizophrenia because far more patients with these disorders receive CMA than schizophrenia patients. Currently, CMA has replaced G-banded karyotype in pediatric genetics clinics as the first-tier genetic test for the evaluation of patients displaying developmental disabilities including DD, ID, autism, epilepsy and/or CA (Manning & Hudgins 2010; Miller et al. 2010). In contrast,
schizophrenia patients are typically run on arrays for research rather than diagnostic purposes. The potential for inflated estimates of penetrance can be observed in the number of cases used for the three previously mentioned penetrance studies. Over 80,000 cases with ID, DD, CA and/or ASD were used for penetrance calculations (Kirov et al. 2013; Rosenfeld et al. 2013) compared to just over 15,000 schizophrenia cases (Vassos et al. 2010; Kirov et al. 2013).

The polygenic hypothesis states that schizophrenia liability is based on the summation of multiple variants with modest effect sizes that, when combined, cross a disease liability threshold (Gottesman & Shields 1967). There are many GWAS studies that have begun to study the additive effect of common SNPs in schizophrenia liability (Lee et al. 2012; Lee et al. 2013; Ripke et al. 2013). Studying the additive effects of rare variants such as CNVs will be much more complicated than the study of common SNPs; nonetheless, efforts are currently underway. Initial investigations focusing on the role for more than one CNV in an individual with a neurodevelopmental disorder have been very conservative in their approach and only consider large CNVs (100-500 kb minimum size) and only those with known pathogenicity (Girirajan et al. 2010; Girirajan et al. 2012; Williams et al. 2013). The conservative approach used by these studies likely excludes many novel and real risk variants in their analysis.

In particular, there is substantial evidence that the 15q11.2 BP1-2 CNV is a promising candidate for pursuing a multiple-hit CNV model. First, out of 72 CNVs previously known to be associated with neuropsychiatric disease, the 15q11.2 BP1-2 deletion was reported to have the highest likelihood of harboring additional CNVs in patients with ID and CA (Girirajan et al. 2012). A careful literature search provides further evidence for this, as several early reports of the 15q11.2 BP1-2 deletion also mention the presence of additional CNVs in affected individuals (Doornbos et al. 2009; van der Zwaag et al. 2010; Burnside et al. 2011; von der Lippe et al. 2011; Abdelmoity et al. 2012). However, in the studies citing the presence of additional CNVs, the CNVs are
not described in further detail or the patients were excluded from the study. Second, the 15q11.2 BP1-2 CNV has been identified in a large number of control individuals and has also been reported in DGV (Shaikh et al. 2009). It should be noted that DGV should be carefully considered in its use as a “control” database for adult-onset diseases. Many of the individuals included in it are initially screened for major psychiatric phenotypes. With no patient follow-up it is possible that some cases of adult-onset diseases like schizophrenia are missed. Additionally, the 15q11.2 BP1-2 CNV is often inherited from a mildly affected or non-affected parent (Murthy et al. 2007; Doornbos et al. 2009; van der Zwaag et al. 2010; von der Lippe et al. 2011; Abdelmoity et al. 2012). Finally, the 15q11.2 BP1-2 CNV appears to be a pleiotropic locus and has been implicated in several other neuropsychiatric disorders such as autism, epilepsy, ID, DD, ADHD and CA (Doornbos et al. 2009; de Kovel et al. 2010; Hogart et al. 2010; van der Zwaag et al. 2010; Van Den Bossche et al. 2012; Wisniowiecka-Kowalnik et al. 2013).

Our study is the first to present the possibility of a multiple-hit CNV model regarding the 15q11.2 BP1-2 deletion in schizophrenia. We have shown that two of the individuals with the 15q11.2 BP1-2 CNV have the greatest number of conservative CNVs. They also harbored the greatest number of non-conservative CNVs (CNVs of interest that were intronic or not novel to DGV). We are also the first to consider smaller CNVs for such an analysis, something that would be beneficial to the research community at large. While we tried to identify CNVs under 100 kb in size for our study, the reliability of such detection with the use of current array technology is low. Several studies have shown that CNV size versus CNV frequency follows an L-shaped curve with smaller CNVs (down to 30 bp in size) occurring far more frequently than do larger CNVs (Conrad et al. 2006; Levy et al. 2007; Bentley et al. 2008; Wang et al. 2008; Wheeler et al. 2008). The preponderance of CNV studies to date have focused on CNVs over 100 kb in size, thus the vast majority of CNVs have gone unrecognized. This is one possible explanation of why case 3, with the 15q11.2 BP1-2 deletion, did not harbor
additional CNVs of interest. Currently, there are several technologies available for
detecting small CNVs in a more robust and reliable manner, such as high throughput
paired-end mapping, which utilizes paired-end sequence data to identify CNVs as small
as ~3kb in size (Korbel et al. 2007). Next generation sequencing and PCR-based
approaches also allow for the detection of very small CNVs (Flores et al. 2007; Nord et
al. 2011). The best approaches use a combination of the three previously mentioned
approaches and current efforts are underway to catalog these results (Mills et al. 2011).

Our investigation for associations between CNV state or category and deficits in
IQ, neurocognitive abilities or brain structure volume was ultimately limited with respect
to the assumptions it made about CNVs including 1) deletions correspond to lowered
gene expression and duplications correspond to increased gene expression, 2) each gene
directly affected by CNVs is involved in disease risk and 3) genes that are not disrupted
structurally by CNVs are not involved in disease risk. Studies have shown that CNVs can
have an effect on flanking genes that are more than thousands of bases away (Merla et al.
2006; Stranger et al. 2007; Cahan et al. 2009; Henrichsen et al. 2009). One study in mice
showed that gene expression was generally correlated with copy number status. However,
20% of the genes were expressed in the opposite direction of copy number change and
15% of genes found in CNV regions had no change in expression (Orozco et al. 2009).
Furthermore, the brain has the lowest concordance rate between copy number and gene
expression when examining tissue specific gene expression in relation to copy number
(Henrichsen et al. 2009; Orozco et al. 2009). There is also evidence that the effect of
copy number on expression of brain related genes varies throughout the course of brain
development (Chaignat et al. 2011). Some of the discrepancies regarding CNV and gene
expression in the brain may be attributed to copy number mosaicism present in the brain
(McConnell et al. 2013), which only further confounds the true characterization of CNVs
and how they affect biology in the brain. Taken together, these studies suggest that we
need to rethink some of our methodological strategies regarding CNV related studies.
Future Directions

One of the largest unknowns in the underlying genetic architecture of schizophrenia is the way risk alleles interact with each other. Based on the work in this thesis, as well as previous work, there is overwhelming evidence that some iteration of the polygenic theory of schizophrenia is at play. In the face of the high population prevalence of schizophrenia, coupled with the reduced fecundity of patients we remain unable to accurately describe the molecular etiology of the disease after decades of research. It would benefit the schizophrenia research community to move forward in three ways. First, we must truly understand the exact role that CNVs play in biological processes. Second, we must identify and catalog the remaining CNV variants that have been ignored or not yet discovered with array-based studies. Finally, we must move beyond considering only one genetic insult at a time in individuals with schizophrenia. Instead, we should be able to consider the entire mutational load an individual may carry.

If a multiple-hit hypothesis is truly the underlying model of a subset of neuropsychiatric disorders then SNPs and CNVs, the most common types of variation in the human genome, should be considered simultaneously. O’Roak et al. (2011) reported a patient severely affected with autism, language delay, epilepsy, mild ID and possible regression that carried a combined 15q11.2 BP1-2 deletions with a de novo missense mutation in SCN1A. Another recent report has started to examine the possibility that both common SNPs and rare CNVs are involved in the etiology of ADHD (Yang et al. 2013). Although this report was not able to fully integrate the SNP and CNV data for analysis, it is a bold step in the direction of a fully integrated genomic analysis for neuropsychiatric disorders. One of the biggest challenges will be in the interpretation of the results from
studies of integrated mutational loads; we will need more models to be able to attribute significance to these types of findings.

There is already a mechanism in place for future study of the 15q11.2 BP1-2 CNV and how it relates to schizophrenia. As previously mentioned, there is a wealth of CMA data on child-based cohorts in pediatric clinics across the country. This abundance of data would be beneficial in the context of a prospective study because there are not yet any follow-up reports in the literature of the children that have ambiguous diagnoses. Having CMA data on a large cohort like this would allow a researcher to: 1) identify individuals with a 15q11.2 or other recurrent CNVs, 2) identify all other CNVs in those patients, 3) build a detailed clinical report on those patients and 4) determine if those individuals develop schizophrenia in the future. A prospective study like this would be highly impactful for three reasons. First, most premorbid schizophrenia symptoms are described in retrospect or are limited to individuals with a family history of the disorder. Second, previous studies identifying CNVs in schizophrenia have been very conservative in consideration of potential risk variants, thereby excluding all other CNV findings, especially CNVs of unknown significance. Finally, previous association studies of schizophrenia include only very limited clinical picture of the patient’s symptoms beyond a diagnosis of schizophrenia or not. A prospective study such as this would not only address many deficits of previous studies, but would allow us to gain specific insight into how a polygenic model for schizophrenia may work.
REFERENCES


Andreasen N.C. (1985) *Comprehensive Assessment of Symptoms and History (CASH)*. University of Iowa, Iowa City.


individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 86, 749-64.


and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 19, 1682-90.


