Chapter 7
General Discussion

7.1 Design of CNV Discovery and Association Study

My PhD thesis describes a survey of copy number variation associated with schizophrenia. Deciphering the genetic basis of schizophrenia has always been a major challenge in human genetics. In the past two years, CNV discovery and association studies have generated some promising results in the field of neuropsychiatric genetics, and more investigations on copy number variation in psychiatric diseases can be expected in the near future. Below I discuss the current and future trends of such CNV studies.

7.1.1 Multiple Approaches of CNV Study

Multiple genetics approaches and experimental designs are necessary to explore complex genetic traits such as schizophrenia. Our study focussed on CNV investigation with two components: a) a familial study of CNVs associated with schizophrenia in pedigrees with multiple affected individuals; and b) a population-based CNV investigation. There are many possible approaches for identifying CNVs in disease, including (i) looking for concordance in affected siblings or twins; (ii) contrasting CNVs in affected versus non-affected siblings of schizophrenia; (iii) identifying transmitting and non-transmitting CNVs in families or trios (to investigate the role of inherited or *de novo* CNVs) (Xu et al. 2008); iv) identifying common CNVs in families with high density of illness; (v) large-scale population-based case-control cohort CNV associations (ISC 2008; Stefansson et al. 2008); (vi) and investigating the less explored somatic CNVs in diseases. Furthermore, early discovery-based CNV studies revealing novel CNV loci are paving ways to more comprehensive case-control association studies, either at the genome-wide level or targeting particular genes of interest (eg. Rejuscu *et al.* reported a study of CNVs in Neurexin genes) (Rujescu et al. 2008). Combinations of these

approaches will reveal different aspects about the genetic architecture of the disease and the candidate genetic pathways underlying schizophrenia pathogenesis.

7.1.2 Future Large-Scale CNV Studies

Our CNV study on schizophrenia was based on modest sample sizes of 4 familial pedigrees and a case-control cohort of 91 patients and 92 controls. To achieve more definitive disease association, substantially larger sample sizes would be necessary. Recently published schizophrenia CNV studies demonstrated the importance of large-scale studies to identify novel rare recurrent CNV loci (ISC 2008; Stefansson et al. 2008) and to detect association. These risk loci have low frequency in the population, probably due to reduced fecundity as conferred by disease (Stefansson et al. 2008), and would require at least thousands of samples for sufficient statistical detection power. In contrast, common alleles are frequent but may only have marginal increase risks between case-controls, therefore considerable sample sizes would again be required to detect any association.

Collaboration among clinical and research centres for sample collections will largely facilitate schizophrenia genetic research. The most recent CNV studies are results of such large collaborative efforts, such as ISC (International Schizophrenia Consortium) (http://pngu.mgh.harvard.edu/isc/), SGENE (http://www.sgene.eu/) and GROUP (Genetic Risk and Outcome in Psychosis) (https://www.group-project.nl/). Combined, these 3 studies involved close to 8,000 psychosis samples (ISC 2008; Stefansson et al. 2008; Vrijenhoek et al. 2008). One limitation of such multi-centre studies would be the genetic heterogeneity and diagnostic differences between samples collected from different sources (Burmeister et al. 2008). Nevertheless, continued collection of patient and control cohorts, together with improved genotype-phenotype characterization, will be instrumental in rapid progress towards the understanding of schizophrenia.

7.1.3 Enrichment of Schizophrenia Subtypes

Another decision to make in CNV study design is the choice of suitable patient types or subtypes. One criterion is the inclusion of sporadic versus familial cases. The study of Xu *et al.*, for example, compared these two patient types in trios (parent-child sets) and demonstrated that *de novo* CNVs increased disease risk in sporadic cases (a 8 times increase was reported for cases versus control), but not in familial cases (Xu et al. 2008). The study suggested variable genetic architecture among the sporadic and the familial subtypes of patients, namely the involvement of *de novo* CNVs versus inherited CNVs. Although there were examples of *de novo* CNV involved in familial cases, for example in an affected sib pair in autistic spectrum disorder (as explained by post-zygotic mosiacism) (Cook and Scherer 2008), these were mostly exceptions rather than the rule. Consequently, by enriching sample cohorts to either sporadic or familial cases, CNV study may steer towards identification of different types and characteristics of CNVs.

Analogous to conventional linkage and association studies with SNPs, the use of endophenotypes in CNV studies would be another debatable strategy. Endophenotypes represent quantifiable trait markers for the inheritance of a given genotype (see section 1.5) (Gottesman and Gould 2003). Schizophrenia as a complex disease may comprise many diverse behavioural and cognitive traits, each corresponding to different underlying genetic variants. One rationale behind using endophenotypes in schizophrenia study was the advantage of bypassing traditional disease classification to obtain a more phenotypic homogenous group of patients. By sub-typing samples according to endophenotypes, one might increase detection power for the underlying genetic risk variants (Abrahams and Geschwind 2008). For example, the phenotype of P50 sensory gating deficit demonstrated association with a *CHRNA7A* polymorphism with a LOD score of 5.3, while association with schizophrenia disease as a phenotype resulted in a

much lower LOD score of 1.3 (Freedman et al. 1997). Magnetic Resonance Imaging (MRI) has also been used to determine association between functional variants of a serotonin transport gene and anxiety (Hariri et al. 2002). The disadvantage of using such endophenotypes would be the high cost involved, for example in performing imaging studies on every single patient. Other quantifiable traits, for instance executive function and visual memory, could also be incorporated as patient subtype enrichment (ISC 2008).

Finally, patients with different schizophrenia disease onset could be another subtype for enrichment in CNV study, and could alter the likelihood of detecting disease-causing variants. Evidence came from a study from Walsh *et al.*, who reported that patients with early or childhood onset were particularly enriched with rare copy number variants (Walsh et al. 2008). These early onset cases, or other schizophrenia subtypes representing more “severe” forms of the disease, may provide a good subset of patients for CNV detection (Vrijenhoek et al. 2008), while subsequent association studies could be performed in a larger, more general schizophrenia cohort. In addition, atypical clinical cases such as those co-occurring with mental retardation and dysmorphology may also be considered for patient enrichment, although CNVs detected in such cohorts may be linked to other phenotypes rather than schizophrenia *per se*.

7.2 Understanding the Genetic Model of Schizophrenia

7.2.1 Rare Variants Versus Common Variants

CNV studies have provided us with important insights into the genetic model of schizophrenia. One controversy surrounding schizophrenia genetics and many other complex disease traits is the debate between common disease-common variant (CD-CV) (Lohmueller et al. 2003) versus common disease-rare variant (CD-RV) models (Pritchard 2001; McClellan et al. 2007). The CD-CV model proposes that common alleles with small to moderate disease risks may have additive or multiplicative effect on schizophrenia (Cook and Scherer 2008), and have inspired many SNP association studies investigating common polymorphisms in schizophrenia. Recent large-scale association studies have successfully identified such risk variants, one of which being the polymorphism at *ZNF804A* associated with schizophrenia (O'Donovan et al. 2008).

The CD-RV model suggests that heterogeneity of schizophrenia comes from multiple rare variants, a theory which has received increasing attention in schizophrenia genetics during the past decade (Craddock et al. 2007; McClellan et al. 2007). Recent CNV studies further demonstrated that rare risk loci, individually or collectively, could predispose to schizophrenia (ISC 2008; Stefansson et al. 2008), supporting the CD-RV model. Our study replicated the findings of one of these rare loci (the 15q11.2 deletion), and provided a number of other putative disease-specific genetic variants. Apart from locus heterogeneity, CNV studies showed that even within the same disease-associated genomic locus there could be allelic heterogeneity (for example in the case of the Neurexin1 gene, section 1.6) (Rujescu et al. 2008). Some of these rare alleles could be recurrent in the population (either inherited or sporadic) but would remain at low

frequency due to the selective disadvantage they confer (Pritchard 2001), while others may be private events found only in single individuals or families.

On the other hand, current CNV discoveries, including our results, may be skewed towards the lower end of the minor allelic frequency spectrum (Ionita-Laza et al. 2008b), possibly due to the bias of CNV platforms and algorithms detecting relatively large mutations. As advances in CNV discovery techniques allow accurate detection of smaller, more frequent copy number polymorphisms, we may be able to detect common CNV alleles associated with schizophrenia. As presented in chapter 6 of this thesis, we attempted to test disease association of two common copy number polymorphisms, using targeted, PCR-based strategies, one of which was a small CNV of ~5.5 kb demonstrating correlation to gene expression.

Genetic studies of other complex diseases (e.g. in Alzheimer's disease) have demonstrated that the CD-CV and CD-RV theories need not be mutually exclusive in disease populations (McClellan et al. 2007). CNV studies on schizophrenia combined with SNP association have further suggested that these two models coexist, leading to the genetic heterogeneity in schizophrenia.

7.2.2 Incomplete Penetrance and Expressivity

Genetic investigations on schizophrenia have usually supported a complex mode of inheritance. Incomplete penetrance was demonstrated in a number of schizophrenia-associated CNV loci. The three novel recurrent deletions at 1q21, 15q11, 15q13, for example, were all detected in normal, healthy controls, suggesting penetrance of these genetic variants was not complete. In our familial CNV study of the *ABCA13* deletion (section 3.3), incomplete penetrance could explain the presence of deletions in one normal family member within the pedigree. This is consistent with the incomplete segregation in previous familial CNV studies, such as the *DISC1* translocation in an extended Scottish pedigree in which only 70% of translocation carriers were affected (Millar et al. 2003). Instead of a Mendelian mode of transmission, many structural variants demonstrate complex inheritance patterns with a variable degree of penetrance, as well as variable expressivity depending on interactions with other genetic and environmental factors.

7.2.3 Aetiological Overlap of Schizophrenia with Other Psychiatric Diseases

The well-documented link between the 22q11 deletion syndrome and schizophrenia demonstrated that the same structural variant can be associated with numerous psychiatric phenotypes, from schizophrenia to attention-deficit hyperactivity disorder, bipolar disorders and autism spectrum disorders (ASDs) (Antshel et al. 2007; Burmeister et al. 2008; Cook and Scherer 2008). These and other genetic loci argued for pleiotropic genetic effects among psychiatric disorders.

Our study, together with a number of CNV reports on schizophrenia in the past year, has reinforced this notion of aetiology overlap among neuropsychiatric diseases. The

15q11.2 recurrent deletion as detected in our patients and reported in Steffanson's study, for instance, is a region involved in Prader-Willi/ Angelman Syndrome with known cognitive and neurological impairment (Cassidy et al. 2000), and a subset of patients also exhibited autistic features (Hogart et al. 2008). The 15q13.3 recurrent deletion was previously associated with mental retardation and seizures (Sharp et al. 2008), as well as with ASDs (Veltman et al. 2005). Another recurrent schizophrenia CNV at neurexin1 (Kirov et al. 2008; Rujescu et al. 2008; Walsh et al. 2008) was also detected in autism (Autism Genome Project Consortium 2007) and mental retardation patients (Friedman et al. 2006). Furthermore, we detected multiple schizophrenia cohort-specific rare variants with candidate genes for multiple psychiatric diseases such as bipolar disorder (*PIK3C3*) (Stopkova et al. 2004), Myoclonus-Dystonia with obsessive compulsive disorder (*SGCE*) (Marechal et al. 2003; Misbahuddin et al. 2007), and mental retardation and Alzheimer's Disease (*RCAN1*) (Porta et al. 2007).

The CNV study in schizophrenia provided evidence that some genetic risk factors could give rise to diverse psychopathology falling across the traditional diagnostic boundaries. Alternatively, they may also reveal misdiagnosis of psychiatric diseases, or of variable expressivity of the genes involved (Cook and Scherer 2008).

7.3 Clinical Relevance of CNVs in Schizophrenia

7.3.1 CNV Findings Translating into Disease Classification

Investigating CNVs in psychiatric genetics has strengthened the view that the same genetic factors could predispose to diverse psychiatric diseases across traditional diagnostic categories. Schizophrenia and bipolar disorder, for instance, share symptoms of psychosis and mood disorders, with substantial overlap in genetic susceptibility, and are conventionally dichotomized into two psychiatric disorders (Craddock et al. 2006). In contrast, a single psychiatric disorder may comprise subtypes grouped into one disease category for diagnostic purposes. This concept of clinical diversity was recognized as early as the time of Kraeplin and Bleuler, who introduced subtypes or categories of schizophrenia (Bleuler 1911; Kraepelin 1919).

Future characterization of CNVs and other genetic variants in a wider range of psychiatric patients have the potential to revolutionize disease diagnosis. Traditional diagnostic criteria heavily rely on clinical signs and symptoms. These guidelines have presented a major challenge in studying psychiatric disorders (WHO 1992; American Psychiatric Association. 1994), since such diseases present no tangible lesion or phenotype, but rather complex behavioural and cognitive phenotypes. With information from CNV and other molecular genetics, future disease diagnosis could become more genetic-based, giving precise genotype-phenotype definition and more personalized treatment. Of equal importance, the information from genetic research could guide clinicians to a better understanding of phenotypes of the various psychiatric disorders. These will potentially have a substantial impact on clinical practice.

The more precise characterization of genotype-phenotype correlations through CNV studies will eventually translate back to advancement in molecular genetic research.

With high certainty of disease classification and refined phenotype descriptions, disease cohorts could be stratified to homogenous subtypes for genetic studies. This will in turn benefit the investigation of genetic variation associated with psychiatric phenotypes, leading to better understanding of the biology of schizophrenia.

7.3.2 Genetic Counselling and Therapeutic Potential

The translation of CNV findings into genetic testing and therapeutics remain a more distant target. In particular, the issue of genetic testing in schizophrenia is confounded by the complex genetic architecture and heterogeneity of the disease. Multiple rare variants - each having small effect size (as identified by current CNV studies) - are not particularly informative for genetic testing, nor are common variants with small to modest disease risks. Nevertheless, the accumulation of CNV data on pathogenic versus benign variants will be beneficial for clinical and research geneticists screening for variants in patient cohorts, for example by routine karyotyping and CGH experiments. Ultimately, high-throughput and reduced cost in sequencing techniques for individual genome sequencing could provide more complete knowledge of the catalogue of genetic risk factors in schizophrenia. Such techniques will also generate comprehensive genetic information on an individual basis for genetic testing for disease diagnosis and prevention.

Regarding CNV-based treatment of schizophrenia, multiple biological pathways, such as those involved in neuroligin-neurexin or neuregulin-ErRb signalling in the synapse, have started to emerge from recent CNV studies (reviewed in section 1.6). Some components of these signalling pathways are also known drug targets (Marchionni et al. 1996; Wang et al. 2008). Integrating current and future findings of CNVs into pharmacology could

rapidly progress the development of novel therapeutic strategies to treat psychiatric disorders.

7.4 Thesis Summary

This dissertation describes a multi-faceted study of copy number variants in schizophrenia, mainly based on array CGH platforms. In chapter 3 we presented a number of familial cases with schizophrenia and other psychiatric illnesses, and illustrated the search for CNVs carried by all affected members within family. Two rare CNV candidates were described, one on 1p36 with partial duplication of *H6PD* and *SPSB1*, and the other an intronic deletion on 7p12 within *ABCA13*. These two rare CNVs segregate with disease and could possibly play a role in disease pathogenesis. Extensive characterization of the 1p36 duplication revealed an interesting genetic architecture at the CNV breakpoints.

In chapter 4, we described a case-control population-based CNV study in schizophrenia. 91 patients and 92 ethnically matched controls were inspected for CNVs using WGTP BAC array. Subsets of CNVs were validated by qPCR and SNP genotyping CNV data. Based on previous literature establishing a role of rare variants in schizophrenia, we identified a number of rare variants specific to the disease cohorts overlapping plausible disease candidate genes. We also replicated the finding of the 15q11.2 novel schizophrenia-associated deletion as described by Steffanson *et al.*. Extending the findings into a larger CNV dataset, we explored the presence of recurrent CNVs as disease-causing candidates in schizophrenia. Of particular interest was the observation that multiple schizophrenia-specific rare CNVs overlap with candidate genes for various psychiatric disorders.

In the second half of Chapter 4 we presented an investigation of common copy number polymorphisms and schizophrenia. 31 out of 577 “genotypable” CNVs reported significant differences in genotype distributions between cases and controls. We further

genotyped two of these CNPs, a deletion 5' of *CHL1*, and a deletion spanning *CHRFAM7A*, by PCR-based strategies. Both candidates have significant links to schizophrenia with evidence from the literature. Neither candidate revealed a statistically significant bias in the extended case/control cohort of ~300+300. The two candidates and the genotyping experiments were described in Chapter 5.

Finally in Chapter 6, we targeted a set of functionally important molecules, the NMDA-receptor complex (NRC/MASC) in the post-synaptic density of neurons. Among the 183 NRC/MASC genes, we identified CNVs in 20 of these cognitive and behaviour-related genes in normal HapMap individuals. These copy number variants may confer a phenotypic effect on normal individuals by modulating synaptic functions through the NRC complex. Copy number variants near the *NSF* gene in 17q21 were further characterized. The two CNV blocks showed unusual population bias, and we identified a new European-specific H1 haplotypes with potential evolutionary significance.

7.5 Future Direction

The identification of CNVs as risk loci has provided a glimpse of the biological basis of schizophrenia. Looking forward, CNVs are just in the initial step of our understanding of schizophrenia genetics. Other independent risk factors in schizophrenia, including those that are genetic (SNPs, INDELs, and other structural variants such as inversions and translocations), epigenetic, environmental or stochastic, are all part of the comprehensive picture of this complex psychiatric illness. In addition, the same genetic locus may harbour multiple allelic variants of different types, further complicating the issue. The genetic architecture of schizophrenia, taking into account the number of variants at a given disease locus, their frequencies, and the penetrance of the genotype combinations, will add another layer of complexity to our comprehension of this disease (Pritchard and Cox 2002) (Figure 7.1).

Improvement of genomic technologies – in particular more accurate, higher-resolution and higher-throughput CNV detection platforms as well as advancement in sequencing technologies - will facilitate the search of such schizophrenia risk factors in a near future. New bioinformatic tools for CNV analysis will also be indispensable. Better algorithms for CNV detection and genotyping are needed (McCarroll and Altshuler 2007; Barnes et al. 2008; Ionita-Laza et al. 2008a), together with improved statistical methods to look for disease association (McCarroll and Altshuler 2007). Databases for CNV information will also be important in providing baseline comparison as the field is awash in CNV discoveries. For instance, the Database of Genomic Variants (DGV; http://projects.tcag.ca/variation/) and DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources; https://decipher.sanger.ac.uk/) are two public repositories collating CNV discoveries in normal and affected individuals

respectively. Databases dedicated to schizophrenia CNVs, such as one analogous to the Autism Chromosome Rearrangement Database (http://projects.tcag.ca/autism/), will also be valuable to the schizophrenia research and clinical community.

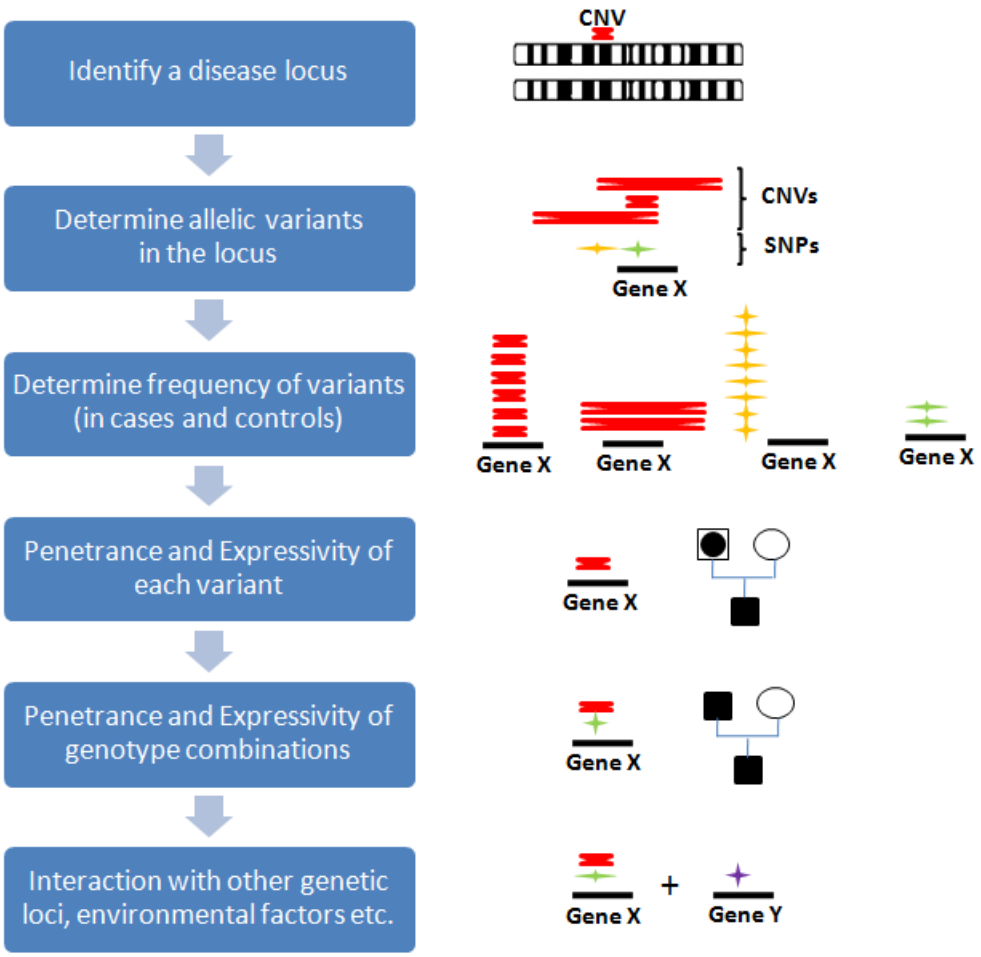

Figure 7.1 Identification of CNVs as disease risk loci for further characterization.

Finally, genetic studies alone are not sufficient to unravel such a complex human trait as schizophrenia. Multiple parallel approaches, including system biology, neuroscience and animal models, will benefit from the results of CNV investigations to deliver more valuable insights into schizophrenia. Candidate genes identified in CNV studies described in this thesis or in literature are prime candidates to be tested in animal models to assess behaviour and cognitive phenotypes, for example, or in neuronal cell cultures to confirm electrophysiological and neurological relevance. As such, the work described in this dissertation is part of the Genes2Cognition Program (http://www.genes2cognition.org/), and CNV findings described herein will be further investigated within the integrated, multi-faceted research program (Grant 2003; Croning et al. 2008).