Chapter 3
Familial CNV Study
in Schizophrenia

3.1 Whole-Genome CNV Screen in Families

3.1.1 Whole-genome Array CGH Screen in Three Familial Cases

The familial nature of schizophrenia is well-established. While acknowledging the presence of sporadic forms of the disorder, one strategy to understand schizophrenia genetics is to focus on familial cases, by assessing inheritance patterns of mutations within individual families and their segregation with disease. We applied this approach to study transmission of CNVs in families with schizophrenia and other neuropsychiatric conditions. Using the WGTP BAC array as a CNV detection platform, we screened 30 schizophrenia probands, and selected 3 familial cases for screening of CNVs predisposing to psychiatric diseases. In each case the proband was diagnosed with schizophrenia under DSM-IV criteria, and either of the parents was affected, together with at least one affected close relative in the family. The three pedigrees F-29, F-41 and F-192 are shown in Figure 3.1.

Each DNA sample was hybridized twice in dye swap experiments. The algorithm CNVFinder (Fiegler et al. 2006) was applied to detect copy number variants in affected members for each family. To detect potential disease-causing variants we applied a two-stage filtering: (a) to investigate CNVs segregating with disease, only BAC clones detected as CNVs in all affected members within each family were included in our analysis. This resulted in 13, 54 and 55 CNV clones in F-29, F-41 and F-192 respectively; and (b) to investigate rare variants not present in healthy controls, BAC clones frequently deleted/duplicated in the HapMap WGTP control dataset (Redon et al. 2006) (defined as CNVs in >10% of HapMap samples) were considered as benign copy number polymorphisms and therefore removed from analysis.

The analysis returned two CNV regions, the first was a duplication involving two clones (*Chr1tp-23A3* & *Chr1tp-1C7*) at chromosome 1p36 in pedigree F-29, and the second a deletion involving a single clone (*Chr8tp-17G4*) at chromosome 8q22 in pedigree F-41. Quantitative PCR confirmed the 1p36 duplication (Figure 3.2), but did not validate the 8q22 deletion in F-41. Further analysis revealed that log2ratio variations reported by the clone *Chr8tp-17G4* was most likely an experimental artefact.

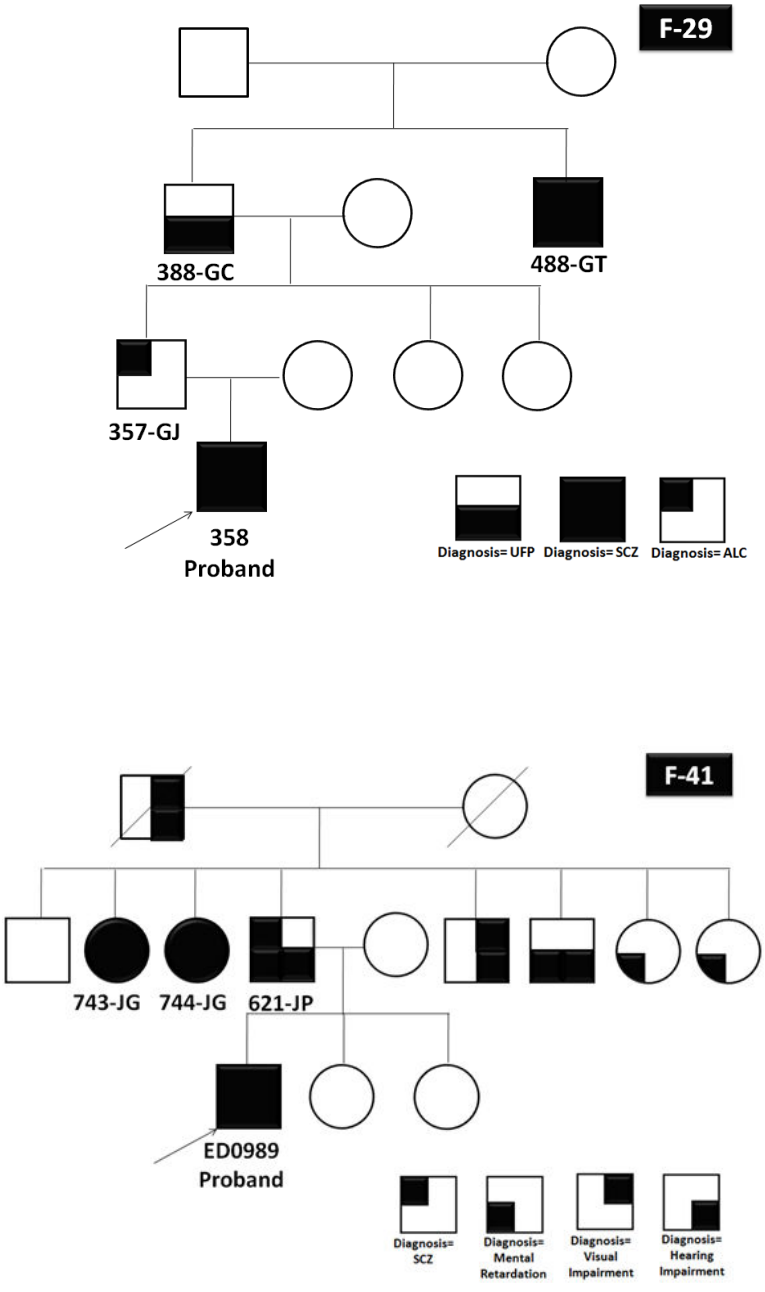


Figure 3.1 (to be continued)

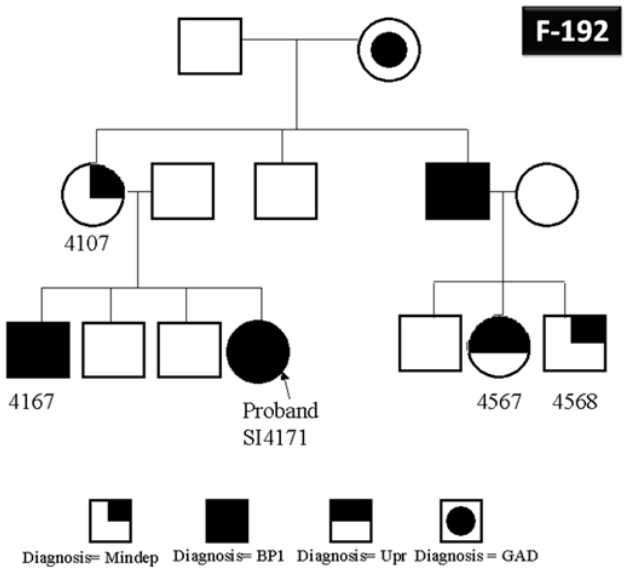


Figure 3.1 Three families analysed by whole-genome CNV screen.
(a) Pedigree F-29; **(b)** Pedigree F-41; **(c)** Pedigree F-192
UFP, unspecified functional psychosis'; SCZ. Schizophrenia; ALC, Alcoholism; Mindep, minor depression; BP1, bipolar affective disorder I; UPR, unipolar depression; GAD, General Anxiety Disorder

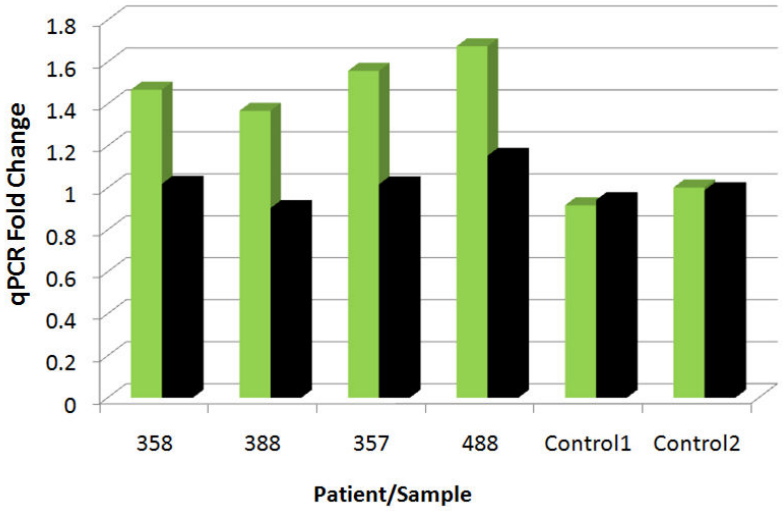


Figure 3.2 qPCR validation of duplication 1p36 in Pedigree F-29. Quantitative PCR using primers located at the 5' end of *SPSB1* within the duplication region (green) validated the duplication in all 4 patients (358, 388, 357, 488). The second set of primers (black) was at the 3' invariant end of *SPSB1*. Duplication was confirmed as detected by the 5' *SPSB1* primer.

3.1.2 Characterization of the Rare Familial Duplication at 1p36.22

At WGTP resolution, the duplication at 1p36.22 in Pedigree F-29 was located at chr1:9189407-9385015 (~200 kb in size). The CNV clones overlapped with two genes, namely hexose-6-phosphate dehydrogenase (*H6PD*) and splA/ryanodine receptor domain and SOCS box containing 1 (*SPSB1*). *H6PD* encodes an enzyme involved in glucose metabolism, and could play a role in unfolded protein response pathways leading to skeletal myopathy (Lavery et al. 2008).

SPSB1 (also known as SSB-1) possesses a SPRY domain (with repeats in splA and RyR) which binds to MET, a receptor protein-tyrosine kinase. It was shown that *SPSB1* could interact with MET to enhance Ras/ERK signalling (Wang et al. 2005). It has been demonstrated - by the use of a mouse model - that ERK (extracellular signal-regulated kinase) signalling plays an important role in cognition. Downstream of neurotransmitters and neuronal receptors, this signalling pathway plays a key function in transducing neuronal signals of downstream transcription machinery, regulating gene expression, and leading to synaptic changes (Brambilla 2003). For example, overexpression of the *Fyn* kinase in mice results in depletion of phosphorylated ERK and other key neuronal signalling components and causes impaired spatial memory in the mouse model of Alzheimer's Disease (Chin et al. 2005). SynGap, a Ras GTPase-activating protein with close association to NMDA receptor, by regulating ERK signalling, can lead to changes in synaptic activity including long term potentiation (LTP) (Komiyama et al. 2002).

We further characterized the structure of this rare duplication identified in psychiatric disease carriers, by performing fluorescence in situ hybridization on extended chromatin fibres (Fiber-FISH) prepared from a patient cell line. Two overlapping probes were hybridized, the first located entirely at the 5' duplicated region of *SPSB1* and the second

one covering both the 5' duplicated and the 3' invariant region of *SPSB1*. The CNV was demonstrated to be a tandem duplication in the same orientation (Figure 3.3).

To map the duplication breakpoints, we used a conventional long-range PCR method with primers tiling across 10 kb either side of the estimated breakpoints to PCR amplify the breakpoint junction. PCR fragments were subsequently sequenced, and the CNV was delineated as a duplicated segment of 86 kb in length at chr1:9240833-9326667 (Figure 3.4). The 3' breakpoint falls on the first large intron of the 5' end of *SPSB1*, within a SINE element (*AluY* repeat). The 5' breakpoint falls on the second last intron of the 3' end of *H6PD*, and is located near another *AluY* repeat. The region is also enriched with simple tandem repeats. These SINE and simple elements may act as chromosomal fragile sites for chromosomal rearrangements. For example, *Alu* repeats have been suggested to predispose to genome instability (Stankiewicz and Lupski 2002).

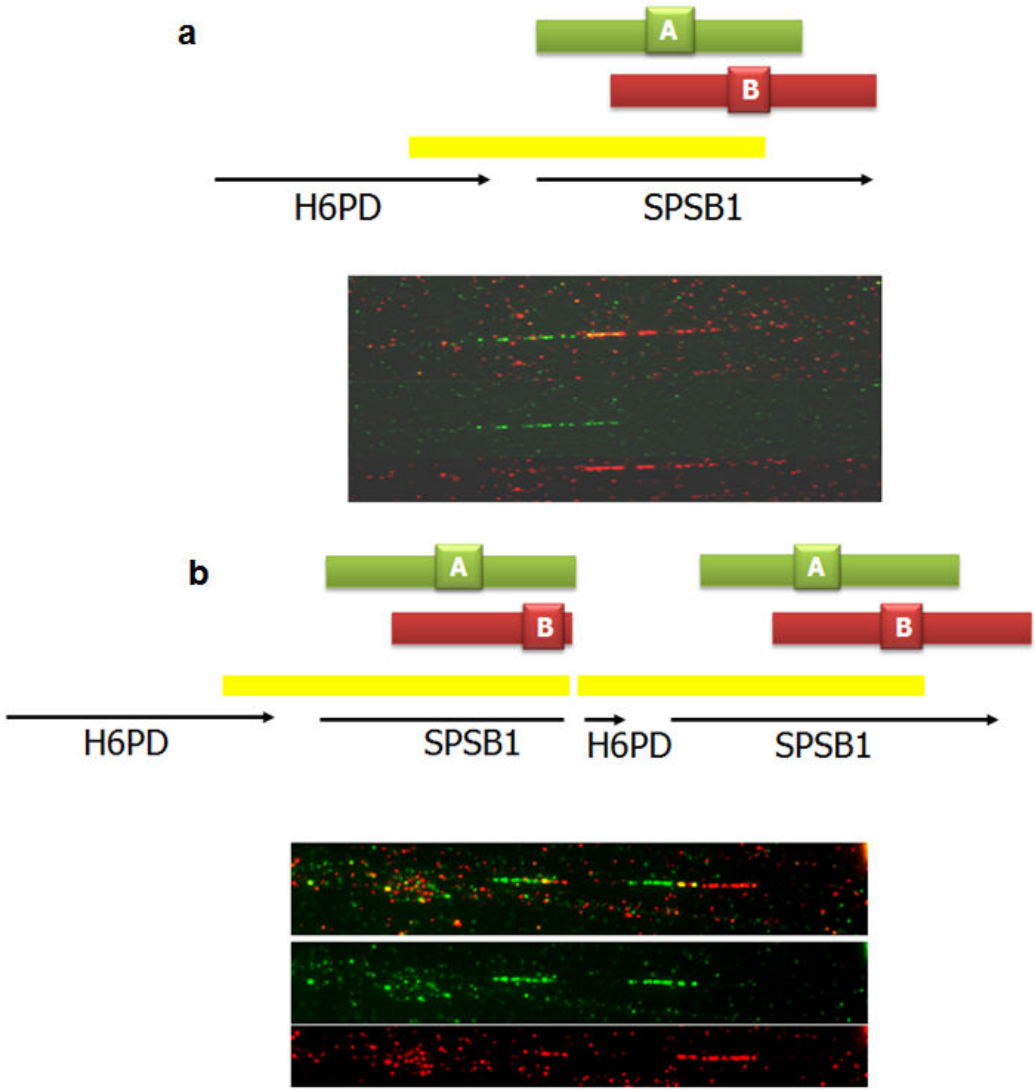


Figure 3.3 Delineation of 1p36.22 duplication structure by Fiber-FISH. Fiber-FISH was performed with two overlapping probes: Probe A (green), located at the 5' duplicated region of *SPSB1* and Probe B (red), partially overlapped with the 3' invariant region of *SPSB1*. FISH results for **a)** a normal control cell line **b)** a patient cell line (358) showing duplication. (yellow block: estimated duplication region)

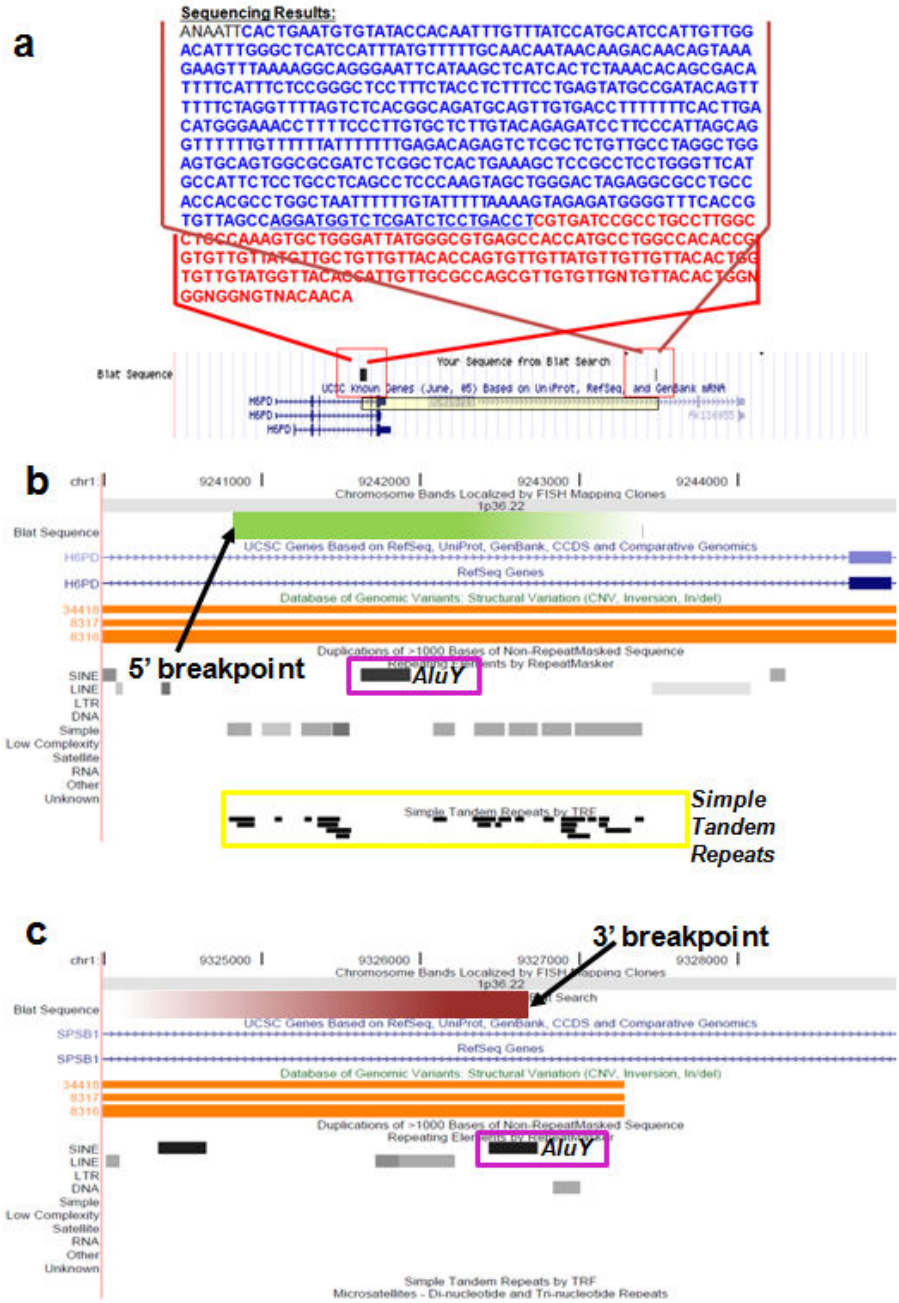


Figure 3.4 Sequencing breakpoints of 1p36.22 revealed repeat structures. a) CNV breakpoint junction was amplified using long range PCR. The amplified fragment was sequenced for breakpoint delineation. **b)** 5' breakpoint was mapped to chr1: 9240833 bp, the second last intron of *H6PD*. A block of SINE element (*AluY* repeat) was identified near the breakpoint, as well as a number of simple tandem repeats. **c)** 3' breakpoint was mapped to chr1: 9326667 bp, intron 1 of *SPSB1*. The breakpoint was located at block of *AluY* repeat sequences.

3.1.3 Known Genomic Rearrangements Near the 1p36.22 duplication

At the time we discovered the 1p36 duplication (in 2006) there was no reported chromosomal rearrangements at the region. The duplication was proposed to be a rare, private variant, potentially a disease predisposition factor within the affected family. As array CGH and other CNV detection techniques matured, more normal variants were detected and deposited into public databases, for example the Database of Genomic Variant (DGV). At the time of writing (in 2008), there are two reports in DGV of copy number variants in the 1p36.22 region (Pinto et al. 2007; Zogopoulos et al. 2007). Zogopoulos *et al.* reported 2 duplications in a control cohort of 1190 samples from Canada. Pinto et al. reported 4 duplications in a cohort of 776 controls (506 Germans & 270 HapMap samples). The 1p36.22 duplication is therefore recurrent in the population and is not unique to affected family members of F-29. We also noticed that the duplication as detected in these samples all have slightly different breakpoints, falling on different repeat sequences, even in the same study (Pinto et al. 2007). This example implicated potential genetic heterogeneity of a recurrent CNV at the sequence level. Alternatively, we cannot rule out that limited resolution of various detection methods compared to PCR sequencing could lead to imprecision in reported CNV breakpoints.

Table 3.1 Known genomic rearrangement at 1p36.22. Data and information gathered for the 1p36.22 region from the Database of Genomic Variant (DGV), compared with results from our current study.

Study	CNV type	# of samples	chr coodinates	method/platform	cohort info
Zogopoulos et al. (2007)	gain	2	chr1:9255290-9316595	Affymetrix 500K and 100K	1190 control samples (Ontario, Canada)
Pinto et al. (2007)	gain	1	chr1:9250000-9311805	Affymetrix 500K	776 control samples (506 Germans, 270 HapMap)
Pinto et al. (2007)	gain	3	chr1:9238434-9327280	Affymetrix 500K	776 control samples (506 Germans, 270 HapMap)
Redon et al. (2006)	NA	0	NA	WGTP BAC array	270 HapMap
Current Study	gain	4	chr1:9240833-9326667	WGTP BAC array and sequencing	3 schizophrenia families and 92 LBC controls

3.2 Deletion at *ABCA13* in an Extended Family with Schizophrenia

3.2.1 Evidence of Functional Mutations of *ABCA13* at 7p12.3

In collaboration with the Psychiatric Genetics group in University of Edinburgh, we investigated whether functional mutations of the ATP Binding Cassette Gene 13 (ABCA13) at chromosome 7p12.13 region could be associated with schizophrenia. A previous cytogenetic study had identified a translocation breakpoint at this locus in a patient with schizophrenia. Moreover, subsequent mutation screening within the ABCA13 gene detected an excess of rare, truncation SNP mutations in patients compared to control (personal communication, Professor Douglas Blackwood).

ABCA13 belongs to the ATP Binding Cassette (ABC) family of integral membrane proteins which regulates the transport of a wide variety of substrates across biological membranes (Dean et al. 2001). ABCA13 is an unusually large gene, spanning 450 kb across chromosome 7p12.13. It has a modular structure with two elements: a hydrophobic, transmembrane structure at the N-terminus, followed by a cytosolic domain which contains the ATP-binding domain (Prades et al. 2002). Although the potential function of ABCA13 in the brain or in central nervous system is not well known, a number of ABC transporters have been implicated in neurological diseases, due to their function in brain lipoprotein transport and their involvement in metabolism and homeostasis (Kim et al. 2008). ABCA1, for instance, was shown to be involved in the metabolism of cholesterol, phospholipids and apolipoprotein in the central nervous system (Hirsch-Reinshagen et al. 2004). Cholesterol homoeostasis at the neuronal membrane plays an important role in the processing of amyloid precursor protein (APP) and the regulation of beta-amyloid peptide levels. Several independent studies have demonstrated a role for ABC transporters in APP processing and their involvement in

Alzheimer's disease (Koldamova et al. 2003; Wahrle et al. 2005; Chan et al. 2008; Uehara et al. 2008).

In addition, ABC transporters are active drug efflux transporters across bio-membranes, and some members (especially those in the ABCB subfamily) are responsible for drug transport at the blood-brain barrier (Loscher and Potschka 2005). In line with this, polymorphisms and mutations in genes belonging to the ABC family have been implicated in drug metabolism (Cascorbi 2006; Hermann and Bassetti 2007): for instance, some ABCB1 genotypes were linked with responses to anti-depressant drug treatment (Takao et al. 2006; Uhr et al. 2008).

3.2.2 Oligo Array CNV Screen in an Extended Pedigree

To identify copy number changes in the region, a custom oligo array (Agilent Technologies) was designed with 34,036 probes covering 1.5 Mb of sequence spanning ABCA13 at chromosome 7p12. Validation experiments demonstrated the reliability and sensitivity of this custom array (see Appendix B).

Array CGH was then performed on DNA samples from 5 patients with possible CNV at ABCA13 (Family 10, 18, 178, 340 and a sporadic case) (Table 3.2). A deletion of approximately 11 kb in size was detected in Patient 4398 (with schizophrenia) from Family 340. The deletion is located in intron 53-54 of ABCA13 at chr7:48,552-48,563 kb (Figure 3.5). We subsequently performed CNV screening on custom arrays for all DNA samples from Family 340, presenting with varied psychiatric phenotypes (Figure 3.6). Out of 11 family members with DNA samples available, 6 have psychiatric disorders ranging from schizophrenia, depression (with variable degree) and anxiety disorder. 5 out of these 6 affected members carry the deletion.

In the first generation, the ABCA13 CNV was present in Sample 5688. Clinical details and DNA sample was not available from her deceased partner. The deletion was inherited by all members in the second generation, three of whom (4398 (proband), 5671 & 5725) had severe psychiatric disorders.

In the third generation, the deletion was present in two siblings (5687 & 5690) who had single episode of depression. However, another sibling with depression (5693) did not inherit the CNV. It is also noted that individual 5713, indicated as a normal individual in the pedigree, had a presumed single episode of depression which did not pass DSM-IV criteria for depression.

Table 3.2 Schizophrenia patients with DNA analysed on custom arrays.

	Schizophrenia Patient	Family	CNV
Proband 1	7808	178	no
Proband 2	7804	18	no
Proband 3	3341	10	no
Proband 4	4398	340	yes
Proband 5	7812	sporadic case	no

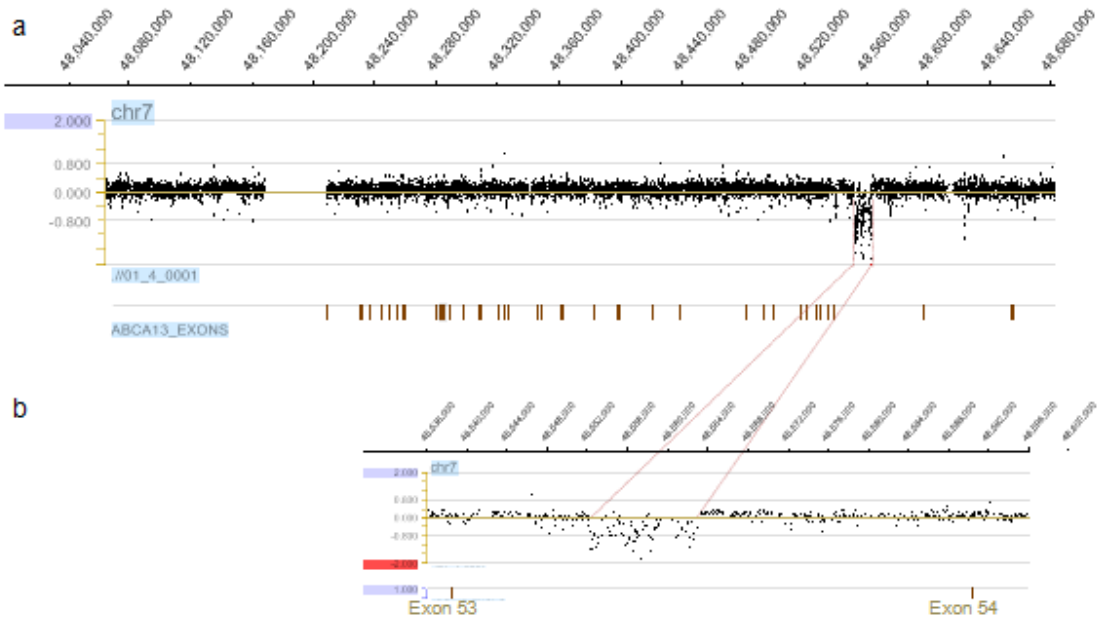
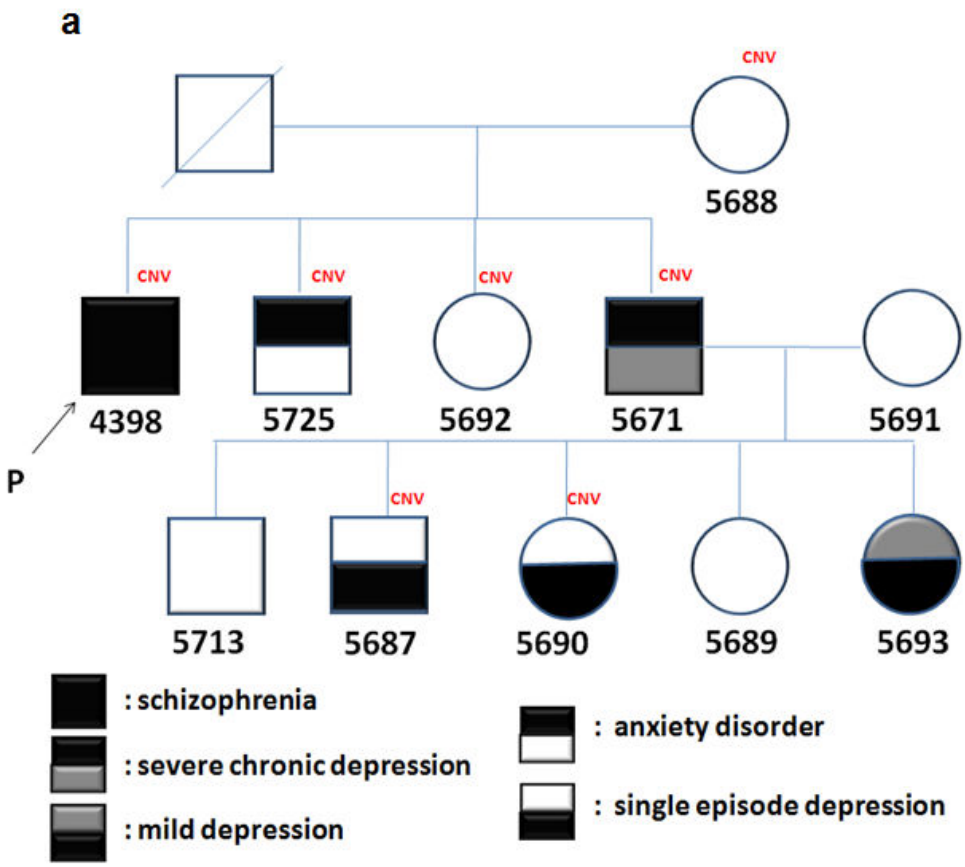


Figure 3.5 Array CGH detection of the 7p12 deletion for Patient 4398. Custom oligo array CGH was performed on 4398, the proband from Family 340 (against reference HapMap NA10851) **a)** Log2ratio profile at chr7:480,400,00-486,800,00. A deletion was detected at ~485.5 Mb on chromosome 7 (vertical bars in brown depicts the genomic locations of the ABCA13 exons). **b)** Detailed view of the deletion, located at ~48,552-48,563 kb, between exons 53 & 54 of *ABCA13*. The deletion size is ~11 kb.



(Figure 3.6 to be continued)

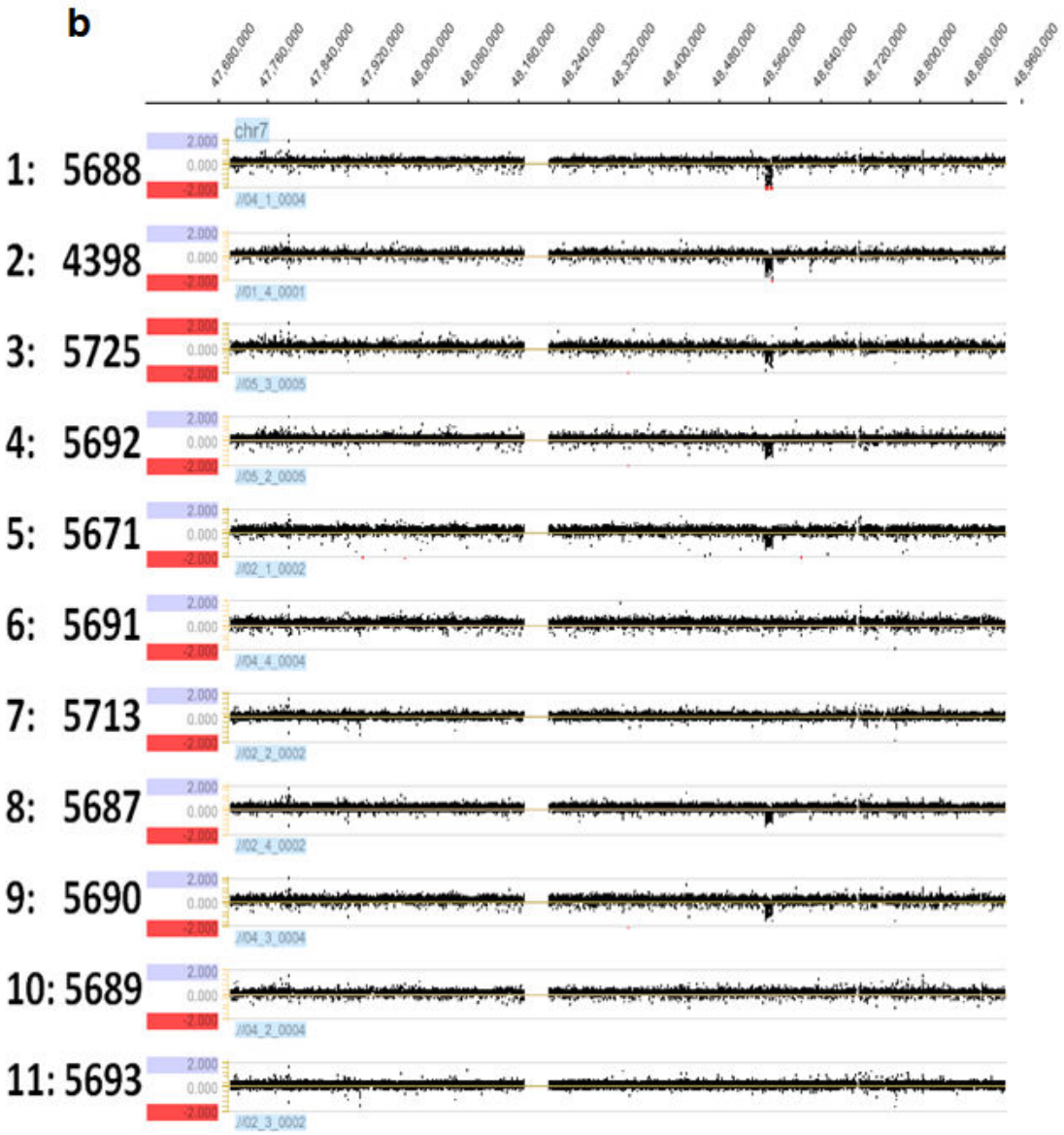


Figure 3.6 Array CGH profiles for all available members in Family 340. a) Pedigree of family 340. Individuals detected with the *ABCA13* deletion are marked "CNV" (in red). **b)** Oligo array CGH was performed on both affected and non-affected members of the family (against reference HapMap NA10851). *ABCA13* genomic region at chr7:47,680,000-48,960,000. A deletion at ~485.5 Mb was detected in samples 5688, 4398, 5725, 5692, 5671, 5687 and 5690.

3.3 Chapter Summary and Discussion

In this chapter, we described a CNV investigation in familial cases of schizophrenia and related neuropsychiatric conditions. As a first approach, we applied a whole-genome WGTP array CGH platform for 3 families. We detected a candidate CNV at 1p36 disrupting the gene *H6PD* and *SPSB1* in all affected members within F-29, but not among 270 normal HapMap individuals. The variant was further characterized by Fiber-FISH, long range PCR and breakpoint sequencing.

The second approach was focused on one candidate gene - *ABCA13* - on chromosome 7p12, a region previously linked to schizophrenia by cytogenetic studies (personal communication, Prof. Douglas Blackwood). With a high-resolution custom array we screened for cryptic CNVs not detected previously by other cytogenetic techniques. A ~11 kb deletion in intron 53-54 of the *ABCA13* was detected in one of the 5 families. The deletion segregated with the disease in all affected members except an individual with mild depression (5693), who could be explained as a phenocopy. On the other hand, a normal sibling (5692) was detected to inherit the deletion.

The inheritance of *ABCA13* CNV revealed the complexity in studying schizophrenia and related psychiatric disorders. Genetic heterogeneity and the complexity of the phenotype, characteristics of many human behavioural traits including schizophrenia, have challenged the study of psychiatric genetics for decades (Cantor and Geschwind 2008). In our study, we faced complicated genotype-phenotype correlation due to a number of reasons: First diagnostic criteria could only serve as guidelines in the complicated spectrum of psychiatric diseases. Overlapping clinical phenotypes in individuals exist, blurring the boundaries of normal to psychiatric patients of variable degrees. An individual in F-340, for example, was classified normal but in fact had single episode of

depression. Secondly, disease-causing variants may have incomplete penetrance and variable expressivity. This will lead to promising but incomplete segregation, as seen in our study and in several cases of well-established structural variations. The schizophrenia candidate gene *DISC1*, for example, was originally identified as an almost but not complete segregating translocation, in an extended Scottish pedigree with schizophrenia and multiple psychiatric conditions (St Clair et al. 1990; Blackwood et al. 2001).

Despite these complications, precedent studies focusing on causative factors in rare, inherited forms of the disease have shown this to be an effective strategy. *DISC1* was one such example. Another relevant example comes from the history of Alzheimer's disease genetics. Rare familial studies of Alzheimer's disease in the past were instrumental in the identification of the amyloid precursor protein (*APP*). These studies in turn provided important insights for investigation of the more common, sporadic form of disease. Furthermore, recent CNV studies in Alzheimer's, again focusing on autosomal dominant forms, identified the duplication of the *APP* locus as a predisposition factor in 8% of familial cases screened (Delabar et al. 1987; Rovelet-Lecrux et al. 2006; Sleegers et al. 2006).

Analogous to our familial CNV study in schizophrenia, the above approaches demonstrates how inherited forms of diseases could simplify the study of genetically complex neurological disorders. Studying schizophrenia as Mendelian disease trait has potentially two important advantages. First, schizophrenia probably comprise of some rare, virtually unique mutations (e.g. CNVs) as disease-causing variants (Abrahams and Geschwind 2008), only identifiable through extremely large case-control cohorts in population-based association. To circumvent the requirement for such large sample

sizes, looking for disease transmission in inherited forms could be more efficient and cost-effective.

Secondly, CNVs identified in rare familial cases could generate further candidates of disease association. CNVs could disrupt genes, for example, while the same gene might harbor more common variants with modest disease risk, and could be identified through targeted sequencing efforts. In addition, defining interaction partners and biological pathways of CNV gene candidates, one may discover novel disease candidates. For instance, *PDE4B* (Phosphodiesterase 4B) (Millar) and *NUDEL* (NudE-like) (Ozeki et al. 2003), both interacting factors of *DISC1*, were identified as schizophrenia candidates following the discovery of *DISC1* translocation.

Nevertheless, caution has to be taken in familial studies of such a complex psychiatric disease. First, genetic factors underlying schizophrenia and other psychiatric conditions may exhibit pleiotropic effects, with some CNVs demonstrating incomplete penetrance or variable expressivity. This will complicate segregation analysis in families, as in our case of *ABCA13* deletion. Secondly, our understanding of CNVs is still rudimentary. Evolving knowledge and views on this type of variant is likely to influence clinical decision on what is benign and what is causative. The 1p36 duplication we described, for example, suggested that some genomic loci may be under-ascertained for CNVs in the past or at present. Variants defined as “rare” or “unique” may in fact occur in the general population, although screening in large sample sizes is required for detection. This also alludes to the need for public resources and database for improved CNV entries, which will be discussed in Chapter 7.

The familial studies described in this chapter are limited in scale. Future work on larger-scale family-based schizophrenia studies, based on more pedigrees (as described for

the Alzheimer's study), or with larger kindred (such as the one used to identify the *DISC1* translocation) will be highly informative. Alternatively, families or trios may also be used to address the *de novo* nature of CNVs in schizophrenia. All these will be complemented with population-based association studies, another approach we employed which will be described in the next Chapter.