7-2018

# Whole genome sequencing of 91 multiplex schizophrenia families reveals increased burden of rare, exonic copy number variation in schizophrenia probands and genetic heterogeneity

Fayeza F. Khan

Phillip E. Melton

Nina S. McCarthy

Bharti Morar

John Blangero

*See next page for additional authors*

## Authors

Fayeza F. Khan, Phillip E. Melton, Nina S. McCarthy, Bharti Morar, John Blangero, Eric K. Moses, and Assen Jablensky

# Title

Whole Genome Sequencing of 91 Multiplex Schizophrenia Families Reveals Increased Burden of Rare Copy Number Variation in Schizophrenia Probands and Genetic Heterogeneity

# Authors

Fayeza F. Khan, PhD[a]*, Phillip E. Melton, PhD (phillip.melton@uwa.edu.au)[a], Nina S. McCarthy, PhD (nina.mccarthy@uwa.edu.au)[a,b,c], Bharti Morar, PhD (bharti.morar@perkins.uwa.edu.au)[b,c,d], John Blangero, PhD (john.blangero@utrgv.edu)[e], Eric K. Moses, PhD (eric.moses@uwa.edu.au)[a], Assen Jablensky, MD, DMSc (assen.jablensky@uwa.edu.au)[b,c]

[a] Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, The University of Western Australia and School of Biomedical Sciences, Curtin University, MRF Building, 50 Murray St, Perth, WA 6000, Australia

[b] Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, University of Western Australia, MRF Building, 50 Murray Street, Perth 6000, Australia

[c] Cooperative Research Centre for Mental Health, Carlton, South Victoria, Australia

[d] Harry Perkins Institute of Medical Research, Centre for Medical Research, The University of Western Australia, 6 Verdun Street, Nedlands, WA 6009, Australia

[e] South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley, One West University Blvd, Brownsville, TX 78520, USA

* Corresponding Author

Contact details for corresponding author:

Postal Address: 50 Murray Street, MRF Building, Level 5, Perth, 6000, Western Australia

Telephone: +61 423 102259

Email: fayeza.khan@uwa.edu.au

## Abstract

The importance of genomic copy number variants (CNVs) has long been recognized in the etiology of neurodevelopmental diseases. We report here the results from the CNV analysis of whole-genome sequences from 91 multiplex schizophrenia families. Employing four algorithms (CNVnator, Cn.mops, DELLY and LUMPY) to identify CNVs, we find 1,231 rare deletions and 287 rare duplications in 300 individuals (77 with schizophrenia (SZ), 32 with schizoaffective disorder (SAD), 82 with another neuropsychiatric diagnosis and 109 unaffected). The size of the CNVs ranges from a few hundred base-pairs to about 1.3 Mb. The total burden of CNVs does not differ significantly between affected (SZ and SAD) and unaffected individuals. Parent-to-child transmission rate for rare CNVs affecting exonic regions is significantly higher for affected (SZ and SAD) probands as compared to their siblings, but rates for all CNVs is not. We observe heterogeneity between families in terms of genes involved in CNVs, and find several CNVs involving genes previously implicated in either schizophrenia or other neuropsychiatric disorders.

## Keywords

Schizophrenia; Family Study; Whole-genome Sequence; Copy Number Variation

# 1. Introduction

Schizophrenia (SZ) is a complex neuropsychiatric disorder with about one percent life-time risk and high heritability estimates of up to 80% (Sullivan et al., 2003). Despite the high heritability of SZ and years of genetic studies, only a small proportion of the genetic contribution to its causation has to date been accounted for.

Over the last decade, studies of genomic copy number variants (CNVs) have indicated that CNVs play an important role in the etiology of SZ (Stone et al., 2008) and generally have much higher penetrance than Single Nucleotide Variants (SNVs) (Kirov et al., 2014; Schizophrenia Working Group of the Psychiatric Genomics, 2014). Several SZ-associated CNVs have been found to also increase susceptibility to other neuropsychiatric disorders, such as autism and developmental delay. Interestingly, even when these CNVs do not result in any disorder, they still contribute to cognitive deficiencies in unaffected carriers (Stefansson et al., 2014).

To date CNV studies in SZ have relied on microarray technology for detection. This has meant that the sizes of CNVs detected and investigated has been larger than 10 kb (Malhotra et al., 2011). These studies have shown that both the global burden of rare CNVs and de novo CNVs are increased in SZ and related disorders (International Schizophrenia, 2008; Malhotra et al., 2011). Moreover, due to reduced fecundity and therefore negative selective pressure on causative variants, the highly penetrant CNVs discovered so far have been rare and recurrent. So far, most studies of SZ have been underpowered for individual rare variant detection, however 11 rare CNVs have been shown to increase risk for SZ in microarray studies (Kirov, 2015). All of these SZ-associated CNVs are large (100s of kb to several Mb in size) and very rare, such that their observed cumulative frequency is less than three percent in SZ cohorts and even less in matched controls (Kirov, 2015; Rees et al., 2014).

More recently, the CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (Cnv et al., 2017) called CNVs from the GWAS data of 41 321 subjects, and genome-wide significance was reached for eight CNV loci and suggestive support found for another eight loci. However, the top eight loci had been implicated previously (Rees et al., 2014). Unlike microarray and SNP datasets, whole genome sequence (WGS) data offers the potential to assay the full spectrum of CNVs across the genome, including smaller CNVs not robustly picked up by microarrays. There are no published studies to date using WGS data to call CNVs in SZ cohorts.

In the current study, we used WGS to detect CNVs (deletions and duplications) in multiplex families with SZ. This dataset provides a comprehensive picture of rare CNVs of all sizes, and the family structure allows the assessment of inherited and de novo CNVs. Having family data also improves detection of rare variants, as multiple family members may carry the same rare variant, increasing our confidence in the results. From our curated list of rare CNVs we test for statistically significant differences in the burden of CNVs between SZ probands and their family members. We also assess de novo CNVs in a smaller subset of families that have quality data from both parents. We compare all SZ probands to their siblings to compute statistical differences in the transmission rate of CNVs from their parents. Finally, we ascertain CNVs that involve previously determined candidate genes for SZ.

# 2. Methods

## 2.1 Samples

The Western Australia Family Study of Schizophrenia (WAFSS) has been described in detail elsewhere (Hallmayer et al., 2005). It was initiated in 1996 with the aim of comprehensively assessing families with ≥1 member affected with a disorder within the ICD-10 and DSM-IV schizophrenia spectrum. The majority of probands were recruited from consecutive admissions to a psychiatric hospital.

The present study used a subset of the WAFSS cohort. All families are of European ancestry. Affectedness status was divided into four categories: unaffected, schizophrenia diagnosis (SZ), schizoaffective disorder diagnosis (SAD), and other neuropsychiatric diagnosis (other). The 'other' category includes diagnoses of depressive episodes of any severity (21 individuals), adjustment disorder (11 individuals), recurrent depressive disorder (9 individuals), dysthymia (6 individuals), anxiety disorder (6 individuals), two cases each of agoraphobia, avoidant personality, mild cognitive disorder, mixed anxiety and depressive disorder, panic disorder, and bipolar disorder, and single cases of social phobia, Tourette's syndrome, and intermittent explosive disorder.

The average number of members from each family included in the study is 3.3 and the average number of SZ or SAD individuals is 1.2 per family. The breakdown of individuals in this study is as follows: 87 SZ/SAD probands, 93 siblings (13 SZ/SAD, 37 other and 43 unaffected), 117 parents (8 SZ/SAD, 45 other and 64 unaffected), and 3 other relatives (1 SAD and 2 unaffected). There were 13 families with data from one parent and proband, 37 families with data from one parent, proband and at least one sibling, 11 families with trio data, and 19 families with data from trios and at least one sibling. No parent data were available for 7 families and no proband data for four families (filtered in quality control). Complete family data is provided in Supplementary Tables 1 and 2.

## 2.2 Whole Genome Sequencing

Genomic DNA extracted from blood was used to perform WGS in 317 WAFSS participants through the commercial provider Macrogen (South Korea) using Illumina HiSeqX technology. Sequencing was to an average of 16x sequence depth (Supplementary Table 3). Resulting sequencing files were aligned to hg19 using the Isaac aligner (Raczy et al., 2013).

## 2.3 CNV Calling

Four algorithms were chosen to analyse the sorted and indexed BAM files, with duplicate reads marked and unmapped reads removed. Cn.mops (Klambauer et al., 2012) and

CNVnator (Abyzov et al., 2011) use read-depth to infer copy number states, while DELLY (Rausch et al., 2012) and LUMPY (Layer et al., 2014) use discordant paired-end and split reads to identify structural variants.

CNV calls from all tools for each sample were merged such that lists of CNVs called by any two tools were obtained for each sample. Further details of the CNV calling and quality control are provided in Supplementary Methods.

## 2.4 Selecting Rare or Novel Variants

Common variants were removed following comparison with the Database of Genomic Variants (DGV) Gold Standard Variants and 1000 Genomes CNV calls, and using a criterion of 50% or more reciprocal overlap with population CNVs with 1% or higher frequency. BEDTools (Quinlan and Hall, 2010) was used to identify called CNVs that overlapped with variants in databases.

## 2.5 Reducing False Calls

To minimize false calls, rare CNV calls (consensus of at least two tools) from each individual were first used to query single tool calls in family members and non-members for the same or similar breakpoints. If the CNVs were found to be called by only a single tool in members of the same family, they were added to the final list. If the same CNV was found to be called by any number of tools in members of more than two families and was not a known population CNV, it was excluded from further consideration.

## 2.6 Annotation of CNV Calls

ANNOVAR (Wang et al., 2010) was used for the gene-based annotation of all CNV calls using the hg19 refGene database, and transcription factor binding site scoring using the hg19 tfbsConsSites database. Figure 1 presents the workflow from sequence data to the final annotated list of CNVs.

## 2.7 DeNovo CNV Calling

In families with sequence data available from both parents, CNV calls in progeny were compared against those in the parents. Calls present in parents were ignored. Calls only made in progeny (putative de novo CNVs) were checked for having been called by a single tool in any other family member (parents or siblings) and removed from the de novo list if found to be shared.

## 2.8 PCR-based Validation of CNVs

To verify deletion calls, PCR primers in the flanking sequence and deleted sequence were designed. Amplified products were resolved by agarose gel electrophoresis. Real-time quantitative PCRs with commercially designed primers were used for duplication calls (Supplementary Methods).

## 2.9 Statistical Analysis

Comparisons were made between SZ/SAD cases and unaffected/other family members in terms of burden of total CNVs, exonic CNVs, CNVs larger than 50kb, CNVs containing a transcription factor binding site (TFBS) and parent-to-child transmission rates of all and just exonic CNVs. Student's t-tests were performed to compare the two groups. Family structure was accounted for using the R package 'kinship2' (Sinnwell et al., 2014) to generate a relationship matrix and the R package 'regress' to generate two regression models, one taking into account relationship data alone, and the other taking into account both relationship data and disease status. The significance of the difference between the likelihood of the two models was then calculated. Benjamini and Hochberg's FDR correction (Benjamini and Hochberg, 1995) was applied to account for the six groups of variants examined, with q values $<0.05$ considered significant.

# 3. Results

The final curated set of rare CNVs comprised 2,263 deletions and 527 duplications. On average, there were nine rare CNVs per individual. The rare CNV count per individual rangeds from 0 to 28, with a median of nine. Unique deletions (counted once when shared between individuals) numbered 1,231 and unique duplications numbered 287. As observed in previous studies, this is due to a lower detection of duplications over deletions, and increased likelihood of duplications to undergo further rearrangements to give rise to multiallelic sites (Sudmant et al., 2015). A total of 668 deletions (51%) and 202 duplications (~70%) did not overlap with CNVs from the population databases.

The distribution of CNVs between the genomic regions - namely, exonic, intergenic, intronic, and ncRNA (non-coding RNA) - is similar for deletions and duplications, except for a reversal of exonic to intronic CNV ratios (figure 2a). There are about three times as many intronic deletions as exonic deletions, and in contrast, about twice as many exonic duplications as intronic duplications. This shift is most likely attributable to the difference in average sizes of the deletions and duplications in this study (figure 2b). The frequency of duplication calls in the size range of 10kb< to <=1Mb is higher than the frequency of deletion calls in those size classes (figure 2b), in line with previous observations of the increased difficulty in detecting smaller duplications. Larger variants are less likely to be restricted to an intronic region which may explain why there are more exonic duplications than intronic ones.

## 3.1 CNV Counts and Size

The distribution of total CNV counts did not differ significantly between SZ/SAD and unaffected groups. For the same groups, the distribution of exonic CNV counts was significantly higher in the SZ/SAD group ($p=0.02$), but did not remain significant when family structure was accounted for ($p=0.22$).

The distribution of CNVs larger than 50kb also did not vary significantly between unaffected and SZ/SAD individuals.

### 3.2 Transcription Factor Binding Sites

CNVs were annotated for TFBSs using Annovar. About 57% of CNVs were assigned a TFBS scores between 683 and 1000, where 1000 was the maximum score signifying the highest consensus for a TF binding site. The distribution of CNVs with any TFBSS did not differ significantly between SZ/SAD and unaffected individuals.

### 3.3 Parent-to-Child Transmission Rates

For all parent-child pairs, 219 in total, the average rate of transmission of rare CNVs was 0.43. The difference in transmission rates between SZ/SAD and unaffected children was not significantly different after FDR correction ($p=0.049$, $q=0.016$). When only rare exonic CNVs were considered (168 transmissions in total), the difference in transmission rates between SZ/SAD probands and unaffected children was significant ($p=0.006$, $q=0.008$). The average rate of transmission of rare exonic CNVs was 0.54 for SZ/SAD probands, and 0.37 for all other children.

### 3.4 De novo CNVs

Quality data were available for both parents in eighteen families, allowing an evaluation of de novo CNVs. In a total of 37 progeny, 18 potentially de novo variants were observed. Of these 18 CNVs, 1 was exonic, 10 intronic, and 7 intergenic (Table 3).

The rate of de novo variants was 0.48 per child. Four individuals had two and ten individuals had one potential de novo variants. Similar non-uniform distribution of de novo CNVs has been reported before in both normal and diseased cohorts (Kloosterman et al., 2015; Malhotra et al., 2011).

The eighteen families could be split into two groups: those with at least one affected parent (SZ/SAD/Other), and those with both parents being unaffected with any neuropsychiatric disorder. We did not find any significant difference in the average number of inherited or *de novo* CNVs between the two groups.

## 3.5 Annotation of Genes within CNVs

### 3.5.1 CNVs involving genes previously associated with schizophrenia

Genes with exons or introns involved in CNVs were investigated for their function and potential role in disease. Six out of 127 exonic deletions, 10 out of 90 exonic duplications, 10 out of 384 intronic deletions and one out of 44 intronic duplications involved genes with a previous association with SZ based on a literature search. These genes, with details of the CNVs they are involved in, are listed in Table 1. Several other genes of known neuronal function were also found to be involved in CNVs (Supplementary Table 4).

### 3.5.2 Comparison with the Psychiatric Genomics Consortium GWASs

We surveyed CNVs found in the WAFSS sample for any overlap with the findings of Marshall et al (Cnv et al., 2017) and with the 108 SZ-associated loci reported by Ripke et al (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Five CNVs from the WAFSS fell within three of the genome-wide significant CNV loci from the former study and another 14 CNVs overlapped with 12 SZ-associated loci from the latter study (see Table 2). These 19 CNVs included 3 exonic, 8 intronic and 6 intergenic variants, and were found in 38 individuals (15 SZ/SAD, 14 Other, and 9 Unaffected).

## 3.6 Comparison with independent gene sets

In two independent studies from our group, 48 individuals from five multi-generational families with pre-eclampsia (PE) and 20 unrelated individuals with non-syndromic congenital heart disease (CHD) were also whole-genome sequenced by Macrogen, at the same read-depth and subjected to the same bioinformatics analysis for CNV detection as the WAFSS data. Genes with exonic CNVs in the WAFSS data were compared with genes with exonic CNVs in the PE and CHD datasets. Out of 348 genes in the WAFSS set, nine (2.7%) were also present in the PE set, and 11 (3.2%) were also present in the CHD set. 12 of these were from CNVs with the same breakpoints in both data sets, and were also present in population CNV databases at rare frequencies. Eight of the genes were involved in CNVs with different

breakpoints in the two data sets. The 20 shared genes were: *ANKRD12*, *PLEKHA5*, *CD36*, *U2SURP*, *LYG1*, *LYG2*, *GCA*, *NTSR1*, *PRSS35*, *SLC2A5*, *RAB32*, *SLC10A2*, *HFM1*, *AWAT1*, *FKBP14*, *SNX2*, *PLEKHA8*, *SNX24* and *HGFAC*. Thus, none of the genes identified in Table 1 and Table 2 were found involved in CNVs in the independent PE and CHD data sets.

## 3.7 Independent validation of CNV calls

### 3.7.1 PCR-based validation

To estimate the accuracy of the CNV calls, 18 deletions and nine duplications were selected for verification by PCR assays. The CNVs were selected to cover the observed size spectrum; the smallest CNV was 769 bp and the largest was about 600 kb. Four PCR reactions for deletions did not yield any products, including expected products from the undeleted or reference alleles. Thus, 14 out of 18 deletions and eight out of nine duplications were successfully verified with PCR assays (supplementary information). Three of the deletions not verified were identified by both read-depth and read-pair methods and one was only identified by read depth-based tools. Two of the verified deletions were only identified by read-pair methods despite being larger than 2 kb, and one was identified by read depth-based methods only. Thus, we had verification of CNVs that were identified by one type of method and not the other type, as well as of CNVs identified by at least one tool of each detection methodology.

### 3.7.2 Exome microarray-based validation

CNVs were called from Illumina exomecore genotype array data previously generated for the same WAFSS samples using PennCNV (McCarthy et al., 2016) (for methods see supplementary information). Of the CNV calls from the WGS data, 36.7% exonic deletions and 71.8% exonic duplications overlapped with CNV calls from PennCNV (any amount of overlap was considered). For intronic variants, a very small fraction overlapped: 16% duplications and 2.6% deletions. One duplication not verified by PCR, was also called from the array data with 100% of the array-based call overlapping 95% of the WGS-based call.

One deletion not verified by PCR was also called from the array data but with only 9% of the WGS-based call overlapping 100% of the array-based call. This deletion was intronic, hence in an area which was not comprehensively covered by the exome chip. The other three unverified deletions were also intronic. Thus, in combination, both independent methods validated 15 out of 18 (83%) deletions and nine out of nine (100%) duplications.

# 4 Discussion

In agreement with previous observations that CNVs associated with SZ have incomplete penetrance, many variants are shared between affected and unaffected individuals in the WAFSS sample. However, a greater burden of exonic variants is observed in schizophrenic probands when compared to their siblings, as shown by the significantly different transmission rates of exonic variants between the two groups. Given these are multiplex families, these results agree with the simplest explanation for inheritance of disease: functional variants are inherited at a higher rate by affected children as compared to their unaffected siblings. No particular gene is overrepresented in the CNVs. As de novo mutations of functional significance are more likely to occur in sporadic cases of the disease (Xu et al., 2008), it is unsurprising that we do not observe functionally significant de novo variants, albeit in a small subset of multiplex families.

When considering WGS data for CNV calling, there are various parameters that can affect sensitivity and accuracy. Ideally, for CNV detection DNA would be unamplified, uniformly fragmented and sequenced to a high read-depth (> 25x). WAFSS data did not meet all of these criteria. The DNA samples required amplification prior to sequencing and were sequenced to an average 16x read-depth. While there is a need to minimise false-negative calls, the primary concern remains a high number of false-positives. To improve our chances of detecting real variation, we used both read-depth and paired end/split read (RP/SR) methods. To minimize false-positives we decided to use two tools of each type and consider only those CNVs that are called by at least any two methods. The rationale was that some

CNVs will be in genomic regions amenable to read-depth methods but not RP/SR methods, and others vice-versa. Therefore, using both methodologies, but not considering variants if only called by a single algorithm, we aimed to maximise our accuracy. This approach has been recommended previously (Mohiyuddin et al., 2015). Similarly, most of the quality control steps (figure 1) were taken to minimize false-positive calls.

The 43% parent-to-child transmission rate of rare CNVs is evidence of high accuracy. The deviation from an ideal 50% transmission may be due to the small sample size (219 transmissions of an average of 9 rare CNVs per parent), false positive results in the parents and false negative results in the parents or children. Moreover, the small (<6%) overlap of genes involved in CNVs in the WAFSS cohort with two independent cohorts subjected to the same bioinformatics analysis offers proof that the CNVs are unique and are a property of the individual samples rather than a function of the analytical steps utilised.

To date what is understood of the contribution of CNVs to disease risk is that pathogenic CNVs are large, affect multiple genes and have a greater penetrance than SNVs. The picture that has emerged from our analysis indicates that smaller CNVs (thousand to tens of thousands of base pairs) are also likely to play an important role in disease pathogenesis. In our final curated set, 79% of deletions and 42% of duplications are less than 10kb in size (figure 2b). The penetrance of these variants, especially those within genes, is likely to be greater than SNVs but less than large CNVs. However, larger whole-genome sequencing studies will be required to arrive at more accurate estimates of effect sizes.

We did not find any of the large duplications and deletions previously reported in SZ, which is not unexpected given the very low frequency of those variants and our modest sample size.

The dosage effect of deletions overlapping genes is straightforward to interpret; but that of duplications remains largely unknown unless experimentally tested. If duplication breakpoints fall within genes, causing a disruption in the sequence, they are likely to lower

gene expression. If the breakpoints fall in intergenic regions, the interpretation can be more difficult based on sequence information alone. Having two intact copies of a gene on a haplotype may not necessarily increase gene dosage, and could even decrease dosage by interfering with transcriptional machinery.

One potential source of the missing heritability observed in common complex diseases could be lack of, or incomplete functional characterization of genomic sequences, for example non-coding RNAs. In our final set of CNVs, there are 16 duplications and 81 deletions overlapping non-coding RNAs, including long intergenic non-coding RNAs and micro RNAs.

In the present study we have not dealt with point mutations, indels, and structural variants other than deletions and duplications. An analysis of SNPs and indels in the WAFSS data is currently underway and will be published separately. Combining information from CNVs, SNPs and indels will allow a more comprehensive picture of the genetic variation underlying schizophrenia.

# 5 References


Abyzov, A., Urban, A.E., Snyder, M., Gerstein, M., 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 21(6), 974-984.

Amarillo, I.E., Li, W.L., Li, X., Vilain, E., Kantarci, S., 2014. De novo single exon deletion of AUTS2 in a patient with speech and language disorder: a review of disrupted AUTS2 and further evidence for its role in neurodevelopmental disorders. Am J Med Genet A 164A(4), 958-965.

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57(1), 289-300.

Betcheva, E.T., Yosifova, A.G., Mushiroda, T., Kubo, M., Takahashi, A., Karachanak, S.K., Zaharieva, I.T., Hadjidekova, S.P., Dimova, II, Vazharova, R.V., Stoyanov, D.S., Milanova, V.K., Tolev, T., Kirov, G., Kamatani, N., Toncheva, D.I., Nakamura, Y., 2013. Whole-genome-wide association study in the Bulgarian population reveals HHAT as schizophrenia susceptibility gene. Psychiatr Genet 23(1), 11-19.

Butler, M.G., McGuire, A.B., Masoud, H., Manzardo, A.M., 2016. Currently recognized genes for schizophrenia: High-resolution chromosome ideogram representation. Am J Med Genet B Neuropsychiatr Genet 171B(2), 181-202.

Cnv, Schizophrenia Working Groups of the Psychiatric Genomics, C., Psychosis Endophenotypes International, C., 2017. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet 49(1), 27-35.

Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsater, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M.C., de Mas, P., Bieth, E., Roge, B., Heron, D., Burglen, L., Gillberg, C., Leboyer, M., Bourgeron, T., 2007. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39(1), 25-27.

Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelmann, L., Davis, K.L., Silverman, J.M., Brunner, H.G., van Kessel, A.G., Wijmenga, C., Ophoff, R.A., Veltman, J.A., 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry 13(3), 261-266.

Hallmayer, J.F., Kalaydjieva, L., Badcock, J., Dragovic, M., Howell, S., Michie, P.T., Rock, D., Vile, D., Williams, R., Corder, E.H., Hollingsworth, K., Jablensky, A., 2005. Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. Am J Hum Genet 77(3), 468-476.

Heidari, A., Tongsook, C., Najafipour, R., Musante, L., Vasli, N., Garshasbi, M., Hu, H., Mittal, K., McNaughton, A.J., Sritharan, K., Hudson, M., Stehr, H., Talebi, S., Moradi, M., Darvish, H., Arshad Rafiq, M., Mozhdehipanah, H., Rashidinejad, A., Samiei, S., Ghadami, M., Windpassinger, C., Gillessen-Kaesbach, G., Tzschach, A., Ahmed, I., Mikhailov, A., Stavropoulos, D.J., Carter, M.T., Keshavarz, S., Ayub, M., Najmabadi, H., Liu, X., Ropers, H.H., Macheroux, P., Vincent, J.B., 2015. Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability. Hum Mol Genet 24(20), 5697-5710.

Ikeda, M., Tomita, Y., Mouri, A., Koga, M., Okochi, T., Yoshimura, R., Yamanouchi, Y., Kinoshita, Y., Hashimoto, R., Williams, H.J., Takeda, M., Nakamura, J., Nabeshima, T., Owen, M.J., O'Donovan, M.C., Honda, H., Arinami, T., Ozaki, N., Iwata, N., 2010. Identification of novel candidate genes for treatment response to risperidone and susceptibility for schizophrenia: integrated analysis among pharmacogenomics, mouse expression, and genetic case-control association approaches. Biol Psychiatry 67(3), 263-269.

International Schizophrenia, C., 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455(7210), 237-241.

Karayannis, T., Au, E., Patel, J.C., Kruglikov, I., Markx, S., Delorme, R., Heron, D., Salomon, D., Glessner, J., Restituito, S., Gordon, A., Rodriguez-Murillo, L., Roy, N.C., Gogos, J.A., Rudy, B., Rice, M.E., Karayiorgou, M., Hakonarson, H., Keren, B., Huguet, G., Bourgeron, T., Hoeffer, C., Tsien, R.W., Peles, E., Fishell, G., 2014. Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. Nature 511(7508), 236-240.

Kirov, G., 2015. CNVs in neuropsychiatric disorders. Hum Mol Genet 24(R1), R45-49.

Kirov, G., Rees, E., Walters, J.T., Escott-Price, V., Georgieva, L., Richards, A.L., Chambert, K.D., Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., 2014. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 75(5), 378-385.

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U., Hochreiter, S., 2012. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res 40(9), e69.

Kloosterman, W.P., Francioli, L.C., Hormozdiari, F., Marschall, T., Hehir-Kwa, J.Y., Abdellaoui, A., Lameijer, E.W., Moed, M.H., Koval, V., Renkens, I., van Roosmalen, M.J., Arp, P., Karssen, L.C., Coe, B.P., Handsaker, R.E., Suchiman, E.D., Cuppen, E., Thung, D.T., McVey, M., Wendl, M.C., Genome of Netherlands, C., Uitterlinden, A., van Duijn, C.M., Swertz, M.A., Wijmenga, C., van Ommen, G.B., Slagboom, P.E., Boomsma, D.I., Schonhuth, A., Eichler, E.E., de Bakker, P.I., Ye, K., Guryev, V., 2015. Characteristics of de novo structural changes in the human genome. Genome Res 25(6), 792-801.

Knight, H.M., Pickard, B.S., Maclean, A., Malloy, M.P., Soares, D.C., McRae, A.F., Condie, A., White, A., Hawkins, W., McGhee, K., van Beck, M., MacIntyre, D.J., Starr, J.M., Deary, I.J., Visscher, P.M., Porteous, D.J., Cannon, R.E., St Clair, D., Muir, W.J., Blackwood, D.H., 2009. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. Am J Hum Genet 85(6), 833-846.

Kwon, E., Wang, W., Tsai, L.H., 2013. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol Psychiatry 18(1), 11-12.

Lavedan, C., Licamele, L., Volpi, S., Hamilton, J., Heaton, C., Mack, K., Lannan, R., Thompson, A., Wolfgang, C.D., Polymeropoulos, M.H., 2009. Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. Mol Psychiatry 14(8), 804-819.

Layer, R.M., Chiang, C., Quinlan, A.R., Hall, I.M., 2014. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15(6), R84.

Li, Z., Xiang, Y., Chen, J., Li, Q., Shen, J., Liu, Y., Li, W., Xing, Q., Wang, Q., Wang, L., Feng, G., He, L., Zhao, X., Shi, Y., 2015. Loci with genome-wide associations with schizophrenia in the Han Chinese population. Br J Psychiatry 207(6), 490-494.

Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., Gill, M., Karayiorgou, M., Kelsoe, J.R., Krastoshevsky, O., Krause, V., Leibenluft, E., Levy, D.L., Makarov, V., Bhandari, A., Malhotra, A.K., McMahon, F.J., Nothen, M.M., Potash, J.B., Rietschel, M., Schulze, T.G., Sebat, J., 2011. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron 72(6), 951-963.

McCarthy, N.S., Melton, P.E., Ward, S.V., Allan, S.M., Dragovic, M., Clark, M.L., Morar, B., Rubio, J.P., Blangero, J., Badcock, J.C., Morgan, V.A., Moses, E.K., Jablensky, A., 2016. Exome array analysis suggests an increased variant burden in families with schizophrenia. Schizophr Res.

Mohiyuddin, M., Mu, J.C., Li, J., Bani Asadi, N., Gerstein, M.B., Abyzov, A., Wong, W.H., Lam, H.Y., 2015. MetaSV: an accurate and integrative structural-variant caller for next generation sequencing. Bioinformatics 31(16), 2741-2744.

Peykov, S., Berkel, S., Schoen, M., Weiss, K., Degenhardt, F., Strohmaier, J., Weiss, B., Proepper, C., Schratt, G., Nothen, M.M., Boeckers, T.M., Rietschel, M., Rappold, G.A., 2015. Identification and functional characterization of rare SHANK2 variants in schizophrenia. Mol Psychiatry 20(12), 1489-1498.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6), 841-842.

Raczy, C., Petrovski, R., Saunders, C.T., Chorny, I., Kruglyak, S., Margulies, E.H., Chuang, H.Y., Kallberg, M., Kumar, S.A., Liao, A., Little, K.M., Stromberg, M.P., Tanner, S.W., 2013. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 29(16), 2041-2043.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28(18), i333-i339.

Rees, E., Walters, J.T., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., Kirov, G., 2014. Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry 204(2), 108-114.

Schizophrenia Working Group of the Psychiatric Genomics, C., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511(7510), 421-427.

Sinnwell, J.P., Therneau, T.M., Schaid, D.J., 2014. The kinship2 R package for pedigree data. Hum Hered 78(2), 91-93.

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G.B., Jonsdottir, G.A., Doyle, O.M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S.R., Gudmundsson, L.J., Jonsson, G.F., Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B., Thygesen, J.H., Schwarz,

A.J., Didriksen, M., Stensbol, T.B., Brammer, M., Kapur, S., Halldorsson, J.G., Hreidarsson, S., Saemundsen, E., Sigurdsson, E., Stefansson, K., 2014. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 505(7483), 361-366.

Stoll, G., Pietilainen, O.P., Linder, B., Suvisaari, J., Brosi, C., Hennah, W., Leppa, V., Torniainen, M., Ripatti, S., Ala-Mello, S., Plottner, O., Rehnstrom, K., Tuulio-Henriksson, A., Varilo, T., Tallila, J., Kristiansson, K., Isohanni, M., Kaprio, J., Eriksson, J.G., Raitakari, O.T., Lehtimaki, T., Jarvelin, M.R., Salomaa, V., Hurles, M., Stefansson, H., Peltonen, L., Sullivan, P.F., Paunio, T., Lonnqvist, J., Daly, M.J., Fischer, U., Freimer, N.B., Palotie, A., 2013. Deletion of TOP3beta, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. Nat Neurosci 16(9), 1228-1237.

Stone, J.L., O'Donovan, M.C., Gurling, H., Kirov, G.K., Blackwood, D.H.R., Corvin, A., Craddock, N.J., Gill, M., Hultman, C.M., Lichtenstein, P., McQuillin, A., Pato, C.N., Ruderfer, D.M., Owen, M.J., St Clair, D., Sullivan, P.F., Sklar, P., Purcell, S.M., Scolnick, E.M., Holmans, P.A., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Williams, N.M., Toncheva, D., Milanova, V., Thelander, E.F., Morris, D.W., O'Dushlaine, C.T., Kenny, E., Waddington, J.L., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Curtis, D., Crombie, C., Fraser, G., Kwan, S.L., Muir, W.J., McGhee, K.A., Pickard, B., Malloy, P., Maclean, A.W., Van Beck, M., Visscher, P.M., Macgregor, S., Pato, M.T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J.A., Ferreira, C.P., Azevedo, M.H., McCarroll, S.A., Gates, C., Daly, M.J., Sklar, P., Consortium, I.S., 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455(7210), 237-241.

Sudmant, P.H., Mallick, S., Nelson, B.J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B.P., Baker, C., Nordenfelt, S., Bamshad, M., Jorde, L.B., Posukh, O.L., Sahakyan, H., Watkins, W.S., Yepiskoposyan, L., Abdullah, M.S., Bravi, C.M., Capelli, C., Hervig, T., Wee, J.T., Tyler-Smith, C., van Driem, G., Romero, I.G., Jha, A.R., Karachanak-Yankova, S., Toncheva, D., Comas, D., Henn, B., Kivisild, T., Ruiz-Linares, A., Sajantila, A., Metspalu, E., Parik, J., Villems, R., Starikovskaya, E.B., Ayodo, G., Beall, C.M., Di Rienzo, A., Hammer, M.F., Khusainova, R., Khusnutdinova, E., Klitz, W., Winkler, C., Labuda, D., Metspalu, M., Tishkoff, S.A., Dryomov, S., Sukernik, R., Patterson, N., Reich, D., Eichler, E.E., 2015. Global diversity, population stratification, and selection of human copy-number variation. Science 349(6253), aab3761.
Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60(12), 1187-1192.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16), e164.

Wang, L., Budolfson, K., Wang, F., 2011. Pik3c3 deletion in pyramidal neurons results in loss of synapses, extensive gliosis and progressive neurodegeneration. Neuroscience 172, 427-442.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., Karayiorgou, M., 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 40(7), 880-885.

Yang, J.Z., Si, T.M., Ruan, Y., Ling, Y.S., Han, Y.H., Wang, X.L., Zhou, M., Zhang, H.Y., Kong, Q.M., Liu, C., Zhang, D.R., Yu, Y.Q., Liu, S.Z., Ju, G.Z., Shu, L., Ma, D.L., Zhang, D., 2003. Association study of neuregulin 1 gene with schizophrenia. Mol Psychiatry 8(7), 706-709.

**Figure 1: CNV-calling pipeline.**

**Figure 2: Frequency distribution of deletions and duplications by (a) genomic region and (b) size.**

**Table 1: List of CNVs involving genes previously implicated in schizophrenia.**

| Breakpoints (hg19) | region | gene(s)[a] | CNV type | size (bp) | same breakpoints as population CNV | EUR AF[b] | Family ID[d], Relation ( Affectedness status[c]) | References |
|---|---|---|---|---|---|---|---|---|
| **Chr7:48 600 258-49 435 281** | exonic | *ABCA13* | duplication | 835 023 | NA | - | 38, Proband (SZ), Sister (Other) | (Knight et al., 2009) |
| **chr7:47 971 562-48 999 322** | exonic | *ABCA13, C7orf57, HUS1, PKD1L1, SUN3, UPP1* | deletion | 1 027 760 | NA | - | 67, Proband (SZ), Father (Other) | (Knight et al., 2009) |
| **Chr7:69 673 805-69 757 539** | exonic | *AUTS2* | deletion | 83 734 | NA | - | 10, Proband (SZ), Mother (Unaffected), Brother (Unaffected) | (Amarillo et al., 2014) |
| **Chr7:69 421 512-69 467 693** | intronic | *AUTS2* | deletion | 46 181 | NA | - | 4, Mother (SAD) | (Amarillo et al., 2014) |
| **Chr7:71 242 281-71 267 912** | exonic | *CALN1* | duplication | 25 631 | yes | 0.001 | 69, Mother (Unaffected), Brother (Unaffected) | (Li et al., 2015) |
| **Chr7:145 997 390-146 005 695** | intronic | *CNTNAP2* | deletion | 8 300 | yes | 0.001 | 41, Proband (SAD), 51, Proband (SZ)[e] | (Friedman et al., 2008) |
| **Chr7:146 358 859-146 360 069** | intronic | *CNTNAP2* | deletion | 1 210 | no | 0.0 | 71, Proband (SZ), Brother (Unaffected) | (Friedman et al., 2008) |
| **Chr7:147 478 397-147 485 678** | intronic | *CNTNAP2* | deletion | 7 281 | NA | - | 25, Brother (Unaffected) | (Friedman et al., 2008) |
| **Chr16:76 499 168-77 099 327** | exonic | *CNTNAP4* | duplication | 40 000 | NA | - | 61, Proband (SZ), Mother (Other), Sister (Other) | (Karayannis et al., 2014) |
| **Chr11:70 124 511-70 464030** | exonic | *CTTN, PPFIA1, SHANK2* | duplication | 339 519 | NA | - | 16, Proband (SZ), Father (Unaffected) | (Peykov et al., 2015) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Chr8:3 385 624-3 392144** | intronic | *CSMD1* | deletion | 6 520 | NA | - | 33, Proband (SZ), Sister (Other) | (Kwon et al., 2013) |
| **Chr8:3 590 734-3 592 031** | intronic | *CSMD1* | deletion | 1 297 | NA | - | 84, Proband (SZ), Father (Unaffected) | (Kwon et al., 2013) |
| **Chr8:3 619 046-3 661 461** | intronic | *CSMD1* | deletion | 42 415 | no | 0.0 | 88, Proband (SAD), Father (Unaffected) | (Kwon et al., 2013) |
| **Chr4:9 534 245-9 974 714** | exonic | *DRD5, SLC2A9* | duplication | 440 469 | NA | - | 63, Proband (SAD), Brother (Unaffected), Sister (Unaffected) | (Butler et al., 2016) |
| **Chr1:210 631 882-210 643 565** | exonic | *HHAT* | deletion | 11 683 | NA | - | 30, Niece (SAD) | (Betcheva et al., 2013) |
| **Chr2:138 672 001-138 789 000** | exonic | *HNMT* | duplication | 117 000 | NA | - | 60, Proband (SZ), Father (Unaffected), Sister (Other) | (Heidari et al., 2015) |
| **Chr7:110 871 645-111 032 341** | intronic | *IMMP2L* | deletion | 160 696 | no | 0.0 | 12, Father (Unaffected) | (Schizophrenia Working Group of the Psychiatric Genomics, 2014) |
| **Chr7:110 969 628-110 992 659** | intronic | *IMMP2L* | deletion | 23 031 | no | 0.0 | 33, Proband (SZ) | (Schizophrenia Working Group of the Psychiatric Genomics, 2014) |
| **Chr7:111 195 242-111 203 015** | UTR5 | *IMMP2L* | duplication | 7 773 | NA | - | 38, Sister (Other) | (Schizophrenia Working Group of the Psychiatric Genomics, 2014) |
| **Chr8:32 393 743-32 401 748** | intronic | *NRG1* | deletion | 8 005 | NA | - | 6, Proband (SAD), Mother (Unaffected), Brother (Unaffected), 46, Proband (SZ), Sister (SAD)[e] | (Yang et al., 2003) |
| **Chr8:31 718 997-31 741 430** | intronic | *NRG1* | duplication | 22 433 | NA | - | 89, Mother (Unaffected), Brother (Unaffected) | (Yang et al., 2003) |
| **Chr18:39 621 241-39 713 567** | exonic | *PIK3C3* | duplication | 92 326 | no | 0.0 | 14, Proband (SZ), Mother (Unaffected), Brother (Other) | (Wang et al., 2011) |
| **chr2: 200 069 472-200 489 914** | exonic | *SATB2* | duplication | 420 442 | NA | - | 6, Proband (SAD), Mother (Unaffected), Sister (Unaffected) | (Schizophrenia Working Group of the Psychiatric Genomics, 2014) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Chr22:51 136 406-51 140 599** | exonic | *SHANK3* | deletion | 4 193 | NA | - | 7, Proband (SZ), Father (Other) | (Durand et al., 2007) |
| **Chr2:107 418 324-107 420 740** | UTR3 | *ST6GAL2* | deletion | 2 416 | yes | 0.001 | 4, Mother (SAD) | (Ikeda et al., 2010) |
| **Chr1:175 430 122-175 761 840** | UTR5 | *TNR* | duplication | 331 178 | NA | - | 40, Proband (SZ), Mother (Other), Brother (Unaffected) | (Lavedan et al., 2009) |
| **Chr22: 22 312 001-22 579 000** | exonic | *TOP3B* | deletion | 267 000 | NA | - | 42, Proband (SAD), Mother (Other), Sister (Other) | (Stoll et al., 2013) |

[a] In case of more than one gene, the associated gene is underlined

[b]EUR AF: Allele frequency in European population from 1000 Genomes

[c]SZ: schizophrenia, Other: neuropsychiatric diagnosis other than SZ or SAD, SAD: schizoaffective disorder

[d]Familly ID's from Supplementary Table 1

[e] members of two families

**Table 2: WAFSS CNVs that lie within or overlap SZ-associated genomic loci from the PGC GWASs (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014, and CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium, 2016)**

| Bioinformatic summary data for genome-wide significant loci from Ripke et al. | | | | | WAFSS cohort CNVs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Rank** | p value | position (hg19) | SZ[a] | protein coding genes/locus | position (hg19) | region | genes (distance from start of gene in bp) | CNV | size (bp) | Family ID[b], relation (affectedness) |
| **1** | 3.48E-31 | chr6:28 303 247- 28 712 247 | Y | Locus too broad | chr6:28 669 001- 28 690 000 | intergenic | ZBED9(113 889), LOC401242(137 402) | duplication | 21 000 | 86, Proband (SZ) |
| **9** | 1.86E-14 | chr12:123 448 113-123 909 113 | Y | ABCB9 ARL6IP4 C12orf65 CDK2AP1 MPHOSPH9 OGFOD2 PITPNM2 RILPL2 SBNO1 SETD8 | chr12:123 872 001- 123 960 000 | exonic | RILPL1,RILPL2,SETD8, SNRNP35 | duplication | 88 000 | 86, Proband (SZ) |
| **15** | 3.03E-13 | chr7:11 0843 815-111 205 915 | N | IMMP2L | chr7:110 871 645- 111 032 341 | intronic | IMMP2L | deletion | 160 696 | 12, Father (unaffected) |
| | | | | | chr7:110 969 628- 110 992 659 | intronic | IMMP2L | deletion | 23 031 | 33, Proband (SZ) |
| | | | | | chr7:111 195 242- 111 203 015 | UTR5 | IMMP2L | duplication | 7 773 | 38, Sister (Other) |
| **25** | 1.26E-11 | chr11:46 342 943- 46 751 213 | Y | AMBRA1 ARHGAP1 ATG13 CHRM4 CKAP5 CREB3L1 DGKZ F2 HARBI1 MDK ZNF408 | chr11:46 645 121- 46 649 066 | intronic | ATG13 | deletion | 3 945 | 79, Proband (SZ) |
| **43** | 2.86E-10 | chr17:2 095 899- 2 220 799 | N | SGSM2 SMG6 SRR TSR1 | chr17:2 014 809- 2 016 092 | intronic | SMG6 | deletion | 1 283 | 65, Mother (Other) |
| **52** | 1.13E-09 | chr7:104 598 064-10 5063 064 | N | MLL5 PUS7 SRPK2 | chr7:104 710 555- 104 711 317 | intronic | KMT2E | deletion | 762 | 75, Proband (SZ), Father (Unaffected) |
| **59** | 2.24E-09 | chr11:57 386 294-57 682 294 | N | BTBD18 C11orf31 CLP1 CTNND1 MED19 SERPING1 TMX2 YPEL4 ZDHHC5 | chr11:57 502 365- 57 504 118 | ncRNA_intronic | TMX2-CTNND1 | deletion | 1 753 | 47, Proband (SZ), Sister (Other), Brother (Other), Mother (Unaffected), 41, Father (Unaffected)[c] |
| **64** | 3.73E-09 | chr1:243 503 719-244 002 945 | Y | AKT3 SDCCAG8 | chr1:243 836 806- 243 837 722 | intronic | AKT3 | deletion | 916 | 27, Mother (Unaffected) |
| **68** | 4.64E-09 | chr3:17 221 366- 17 888 266 | N | TBC1D5 | chr3:17 643 791- 17 660 779 | intronic | TBC1D5 | deletion | 16 988 | 76, Proband (SZ), Brother (SAD), Mother (Other) |
| **76** | 8.33E-09 | chr2:200 161 422-200 309 252 | N | SATB2 | chr2:200 069 472- 200 489 914 | exonic | SATB2 | duplication | 420 442 | 6, Proband (SAD), Mother (Unaffected), Sister (Unaffected) |
| **92** | 2.69E-08 | chr6:73 132 701- 73 171 901 | N | RIMS1 | chr6:72 694 617-72 700 596 | intronic | RIMS1 | deletion | 5 979 | 39, Mother (Other) |
| **99** | 3.91E-08 | chr12:29 905 265-29 940 365 | N | TMTC1 | chr12:29 956 416-29 960 398 | intergenic | TMTC1(18 724), IPO8(821 517) | deletion | 3 982 | 58, Proband (SZ), Mother (SAD), Brother (Other), 84, Father |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | (Unaffected)[c] |
| | | **Marshall et al (coordinates converted to hg19)** | | | | WAFSS cohort CNVs | | | |
| **7** | 5.52E-05 | chr16:28 822 500-29 052 499 | 16p11.2 (distal) | chr16:29 005 334-29 052 499 | Intergenic | LAT(3 230), RRN3P2 (76 571) | Deletion | 4 258 | 59, Mother (Other) |
| **8** | 1.68E-04 | Chr7:72 380 000-73 780 000 | 7q11.23 | chr7:73 141 947-73 146 252 | Intergenic | STX1A (8 000), ABHD11-AS1 (3 147) | Duplication | 4 305 | 62, Proband (SAD), Sister (Other) |
| | | | | chr7:72 831 630-72 833 142 | Intergenic | FKBP6 (58 984), FZD9 (14 967) | Deletion | 1 512 | 61, Proband (SZ), Sister (Other), 41, Proband (SAD)[c] |
| **17** | 5.79E-03 | Chr7:158 145 959-158 664 998 | 7p36.3 (VIPR2/WDR60) | chr7:158 644 240-158 645 343 | Intergenic | WDR60(3 933) | Deletion | 1 103 | 42, Proband (SAD), Mother (Other), Brother (Other), Sister (Other) |
| | | | | chr7:158 648 686-158 652 226 | Exonic | WDR60 | Deletion | 3 540 | 88, Mother (Other), Sister (Unaffected) |

[a]SZ: Whether the regions were previously found associated with schizophrenia (Y: yes, N: no)

[b] Family ID's from Supplementary Table 1

[c] Members of two families

CNVs: copy number variants, SZ: schizophrenia, Other: neuropsychiatric diagnosis other than SZ or SAD, SAD: schizoaffective disorder

**Table 3: Potential de novo CNVs**

| breakpoints (hg19) | region | genes (distance to gene) | size | type | Family ID[a], Relation, Affectedness |
|---|---|---|---|---|---|
| **chr2:236 981 798-236 985 995** | intronic | AGAP1 | 4 197 | deletion | 7, Proband, SZ |
| **chr4:97 082 599-97 083 209** | intergenic | PDHA2(dist=319 974), STPG2-AS1(dist=1 204 868) | 610 | deletion | 7, Proband, SZ |
| **chr17:16 258 981-16 264 624** | intergenic | CENPV(dist=2 169), UBB(dist=19 483) | 5 643 | duplication | 20, Proband, SAD |
| **chr1:156 016 016-156 017 425** | intronic | UBQLN4 | 1 409 | deletion | 23, Sister, Other |
| **chr22:41 708 180-41 710 931** | intronic | ZC3H7B | 2 751 | deletion | 23, Proband, SZ |
| **chr1:236 802 108-236 832 432** | intergenic | HEATR1(dist=34 267), ACTN2(dist=17 322) | 30 324 | duplication | 81, Sister, Unaffected |
| **chr5:107 407 409-107 407 982** | intronic | FBXL17 | 573 | deletion | 81, Sister, Unaffected |
| **chr11:63 906 001-63 915 000** | intronic | MACROD1 | 9 000 | deletion | 87, Proband, SZ |
| **chr14:105 002 123-105 005 283** | intergenic | KIF26A(dist=354 888), C14orf180(dist=40 738) | 3 160 | duplication | 87, Proband, SZ |
| **chr17:48 924 001-48 931 000** | intergenic | WFIKKN2(dist=4 292), TOB1(dist=8 587) | 7 000 | deletion | 88, Brother, Unaffected |
| **chr2:122 388 452-122 389 350** | intronic | CLASP1 | 898 | deletion | 91, Proband, SZ |
| **chr20:4 941 352-4 943 771** | intronic | SLC23A2 | 2 419 | deletion | 14, Proband, SZ |
| **chr5:14 878 611-14 879 349** | intergenic | ANKH(dist=6 724), LOC101929454(dist=312 406) | 738 | deletion | 16, Proband, SZ |
| **chr4:83 349 600-83 350 402** | exonic | HNRNPDL | 802 | deletion | 35, Proband, SAD |
| **chr9:123 318 744-123 320 055** | intronic | CDK5RAP2 | 1 311 | deletion | 60, Proband, SZ |
| **chr12:80 272 063-80 273 569** | intronic | PPP1R12A | 1 506 | deletion | 60, Proband, SZ |
| **chr1:116 486 572-116 493 144** | ncRNA_intronic | LOC101928977 | 6 572 | deletion | 60, Sister, Other |
| **chr13:59 453 830-59 457 813** | intergenic | LOC101926897(dist=670 213), DIAPH3(dist=781 908) | 3 983 | deletion | 84, Proband, SZ |

[a] Family IDs from Supplementary Table 1

Figure 1

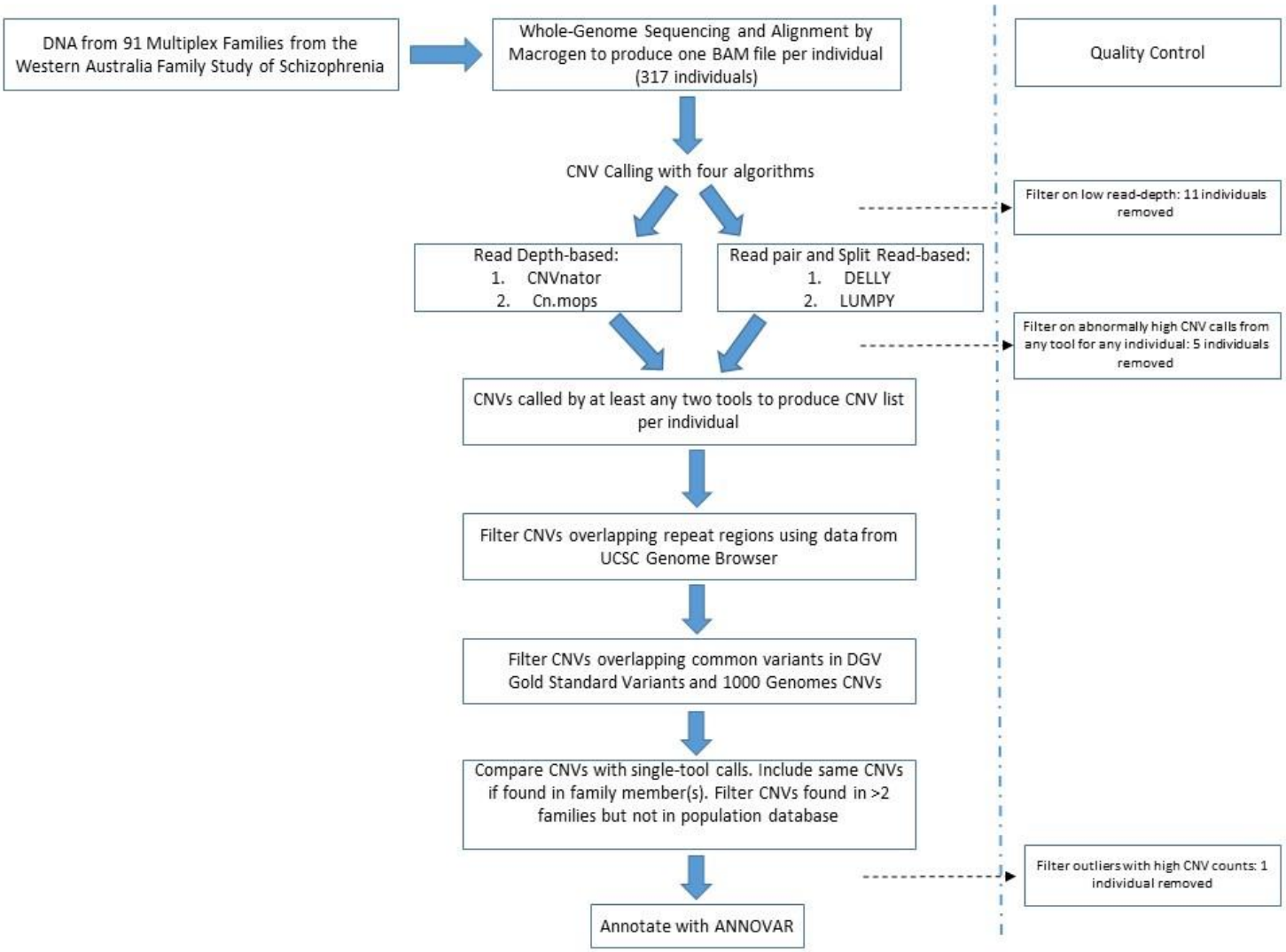
DNA from 91 Multiplex Families from the Western Australia Family Study of Schizophrenia
Whole-Genome Sequencing and Alignment by Macrogen to produce one BAM file per individual (317 individuals)
Quality Control
CNV Calling with four algorithms
Filter on low read-depth: 11 individuals removed
Read Depth-based:
1. CNVnator
2. Cn.mops
Read pair and Split Read-based:
1. DELLY
2. LUMPY
Filter on abnormally high CNV calls from any tool for any individual: 5 individuals removed
CNVs called by at least any two tools to produce CNV list per individual
Filter CNVs overlapping repeat regions using data from UCSC Genome Browser
Filter CNVs overlapping common variants in DGV Gold Standard Variants and 1000 Genomes CNVs
Compare CNVs with single-tool calls. Include same CNVs if found in family member(s). Filter CNVs found in >2 families but not in population database
Filter outliers with high CNV counts: 1 individual removed
Annotate with ANNOVAR

Figure 2(a)

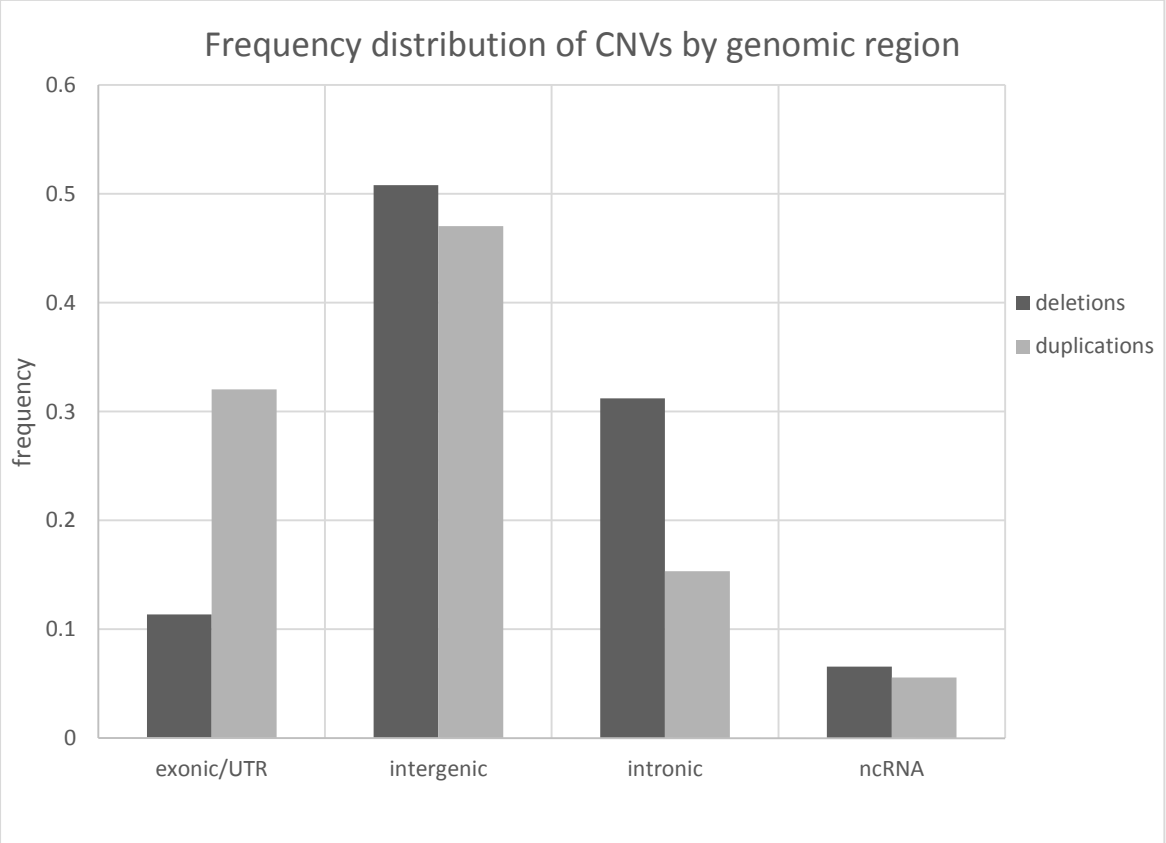


Figure 2(b)