Check for updates

# Genome-wide copy number variant screening of Saudi schizophrenia patients reveals larger deletions in cases versus controls

Mahdi S. Abumadini[1], Kholoud S. Al Ghamdi[2], Abdullah H. Alqahtani[1], Dana K. Almedallah[1], Lauren Callans[3], Jumanah A. Jarad[1], Cyril Cyrus[4], Bobby P. C. Koeleman[5], Brendan J. Keating[3], Nathan Pankratz[6] and Amein K. Al-Ali[4]*

[1]Department of Psychiatry, King Fahd Hospital of the University, Al-Khobar and College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, [2]Department of Physiology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, [3]Department of Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA, United States, [4]Department of Clinical Biochemistry, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, [5]Department of Genetics, Division Lab, University Medical Center Utrecht, Utrecht, Netherlands, [6]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, United States

**Introduction:** Genome-wide association studies have discovered common polymorphisms in regions associated with schizophrenia. No genome-wide analyses have been performed in Saudi schizophrenia subjects.

**Methods:** Genome-wide genotyping data from 136 Saudi schizophrenia cases and 97 Saudi controls in addition to 4,625 American were examined for copy number variants (CNVs). A hidden Markov model approach was used to call CNVs.

**Results:** CNVs in schizophrenia cases were twice as large on average than CNVs in controls ($p=0.04$). The analyses focused on extremely large >250 kilobases CNVs or homozygous deletions of any size. One extremely large deletion was noted in a single case (16.5 megabases on chromosome 10). Two cases had an 814kb duplication of chromosome 7 spanning a cluster of genes, including circadian-related loci, and two other cases had 277kb deletions of chromosome 9 encompassing an olfactory receptors gene family. CNVs were also seen in loci previously associated with schizophrenia, namely a 16p11 proximal duplication and two 22q11.2 deletions.

**Discussion:** Runs of homozygosity (ROHs) were analyzed across the genome to investigate correlation with schizophrenia risk. While rates and sizes of these ROHs were similar in cases and controls, we identified 10 regions where multiple cases had ROHs and controls did not.

KEYWORDS

schizophrenia, CNV, genomewide, variant, Saudi

## 1. Introduction

Schizophrenia is a chronic mental disorder characterized by severe cognitive, behavioral, and emotional symptoms, with an estimated lifetime risk of approximately 1% in the general population. Despite pharmacological treatments for schizophrenia, associated disease progression remains high (Lieberman et al., 2005). Since the advent of antipsychotic medications, which target the type 2 dopaminergic receptor (DRD2) system, no new significant improvements in outcomes have been observed (Morgan et al., 2014). Elucidating the pathophysiology of schizophrenia is a necessary step in improving disease outcomes.

Schizophrenia is highly heritable. Several large genome-wide association (GWAS) and exome sequencing studies have unambiguously demonstrated that both common and rare genetic risk variation explain a considerable proportion of this heritability, and are linked to a range of disease-related factors, including susceptibility, progression, severity, and treatment response (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Howrigan et al., 2020; Kushima et al., 2022). In particular, rare copy number variations (CNVs) have been implicated as significant risk factors. While individually these CNVs are rare, they confer a significant increase risk for disease (ORs of more than 2 and as high as 60; Giaroli et al., 2014; Nakatochi et al., 2021). When combined, such rare structural variants show enrichment of heritable and *de novo* CNVs in cases versus controls. Because most CNVs are rare in the population and show a high degree of heterogeneity, it remains difficult for even the largest studies to have significant power to detect genomewide significance for single CNVs. The Psychiatric Genomics Consortium (PGC) performed the largest GWAS and CNV studies to date including more than 20,000 people with schizophrenia. The GWAS for identification of common inherited variation identified 108 genetic loci associations impacting schizophrenia etiology (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Replication of these 108 loci and identification of novel loci in smaller studies, however, suggests that associations at other genes may provide insight into new mechanistic hypotheses (Ma et al., 2013, 2021; Pardiñas et al., 2018; Lam et al., 2019). A CNV study comprising 41,321 individuals (21,094 cases) had sufficient power to detect significant enrichment of CNV burden and genome-wide significant association for eight loci: 1q21.1, 2p16.3 (*NRXN1*), 3q29, 7q11.2, 15q13.3, distal 16p11.2, proximal 16p11.2 and 22q11.2. The majority of GWAS and CNV studies have been conducted on European populations (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Studying European populations alone greatly limits the ability to discover the causal variants of varying frequencies that underpin mechanisms of debilitating diseases, including schizophrenia. A cross-disorder CNV study combined and compared the CNV analysis of schizophrenia, bipolar disorder, and autism spectrum disorder in 8708 Japanese sample. This study confirmed the enrichment of CNVs in schizophrenia and found interesting differences between the disorders. Similarly, Lam et al., conducted a GWAS in East Asian populations and identified 176 genetic loci, replicating 89 of the 108 previously reported SNPs (Lam et al., 2019). Compared to European populations, high genetic correlation was observed, suggesting that the genetic basis of disease is shared between populations.

Few studies have been performed that include individuals from the Kingdom of Saudi Arabia. Therefore, we aimed to investigate whether we could confirm the relevance of known common and rare genetic risk variants in 136 Saudi Arabian schizophrenia patients as compared to controls (97 Saudi and 4,625 American). More importantly, we aimed to investigate whether the rare risk CNVs detected in the European and Asian populations are also present in our cohort.

# 2. Materials and methods

## 2.1. Patient selection

This case–control study was conducted on 136 Saudi patients with clinically diagnosed schizophrenia recruited from the Department of Psychiatry, King Fahd Hospital of the University (KFHU), Al-Khobar

and Al-Amal Mental Health Complex, Dammam, both in the Eastern Province of Saudi Arabia. Schizophrenia was diagnosed according to the Classification of Mental and Behavioral Disorders Clinical Descriptions and Diagnostic Guidelines (10th revision) World Health Organization (WHO). This study was approved by the Ethical Committee of Imam Abdulrahman bin Faisal University and was conducted in accordance with the 1964 Helsinki Declaration and its later amendments. Informed written consent in English, with a verified translation in Arabic, was obtained from all participants in accordance with the Institutional Review Board (IRB # 2017-01-051). The control cohort consisted of 97 Saudi Arabian individuals and 4,625 American individuals. The United States controls consist of participants from iGeneTRAiN, which contain DNA from kidney allograft donors and kidney recipients and were genotyped using the same genotyping array in the same lab and processed together. The genome-wide genotyping array, quality control pipelines and cohorts have been described extensively before (International Genetics and Translational Research in Transplantation Network (iGeneTRAiN), 2015; Li et al., 2015).

## 2.2. DNA and genome-wide genotyping

Blood samples (5 ml) for DNA extraction were collected in EDTA tubes from Saudi schizophrenia patients and Saudi controls after obtaining signed informed consent. Genomic DNA was extracted from peripheral blood mononuclear cells using standard procedures and stored at −80°C (Qiagen, United States). DNA concentrations and purity were estimated by fluorometry using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Inc., Massachusetts, United States) and were diluted to 20 ng/μl for genome-wide genotyping. Genome-wide genotyping was performed using the Axiom SNP tx v1 array (Thermo Fisher, USA) which comprises over 780,000 genetic variant and CNV probes (Li et al., 2015). Raw intensity files (Affymetrix CEL files) were processed with the Affymetrix Power Tools (version 1.19.0) following the best practice workflow.

## 2.3. Quality control of samples

The raw intensity data used to call genotypes is repurposed to identify CNVs. If a drop in intensity is identified for markers that are physically close together in the genome, then this is indicative of a deletion in the DNA. The reciprocal increase in intensity indicates that a duplication is likely present. The intensity metric (log R ratio or LRR) is computed by taking the sum of the probe intensity for each of the two alleles (X and Y) and normalizing it to the population using standard formulas (Peiffer et al., 2006). The percentage of alleles that are the B allele (B allele frequency or BAF) is another useful metric for calling CNVs. However, if either value deviates too much due to low quality DNA or sample failure, then it can lead to spurious CNV calls. Therefore, any samples with values that were more than three standard deviations from the mean for LRR (only for values between −2 and +2, so as to exclude regions with real copy number variation) or for BAF (only for markers with putative heterozygous genotypes, i.e., values between 0.15 and 0.85) were excluded. All samples were required to have a call rate greater than 95%. Analysis was performed on samples that were verified to be unrelated based on PI_HAT<0.125 with all other individuals, as calculated by the --genome option implemented in PLINK (Purcell et al., 2007).

Intensity data for all high-quality samples were then used to validate sex and to search for individuals with sex aneuploidy. These can be identified by a split BAF pattern that extends for megabases at a time. Genvisis identifies these using a hidden Markov model similar to the CNV calling algorithm but that focuses exclusively on the BAF distribution. Chromosomal arms with large events can also be identified by plotting the standard deviation of the BAF for heterozygous markers (values 0.15 to 0.85) on one axis and the interquartile range (IQR) on the other. Finally, the SNPs that were used to call runs of homozygosity passed all marker QC, including genotype call rate > 95%, Hardy Weinberg $p$-value > 0.00001, PLINK mishap $p$-value > 0.00001, and differential missingness or frequency by sex ($p$ > 0.0001). In addition, markers were pruned to remove markers that were in strong linkage disequilibrium (LD; PLINK --indep-pairwise 50 5 0.2) to exclude common segments that appear as ROHs but that are more likely to have arisen from LD (Gibson et al., 2006; Howrigan et al., 2011). These same variants were used to perform a principal components analysis that captures ancestry using the Genvisis software version 0.1.16 (https://www.genvisis.org/).

## 2.4. CNV calling and filtering

Autosomal CNVs were called using Genvisis (genvisis.org; Skidmore et al., 2016), which is modeled after the PennCNV algorithm (Wang et al., 2007) with the following caveats. First, the LRR values for a set of probes that map uniquely to the genome and that do not commonly deviate from a copy number of two were used in a principal component analysis (PCA) to identify DNA quantity and quality effects (PCs 1 and 2, respectively) and to identify any batch effects that may have caused a shift in the intensity clusters (PCs 3–10). These effects were then regressed out of the LRR values to create more stable estimates of copy number for each probe. Data were then GC-wave adjusted and fed into the hidden Markov model (HMM) in Genvisis to call CNVs. CNV calls were then filtered for higher quality, with the minimum number of probes required for a CNV call being three for homozygous deletions, and 15 for heterozygous deletions or duplications. The confidence threshold for CNV calling was set to a Bayes factor of 10. Adjacent CNV calls were merged if they had the same copy number and if the distance between them was <20% of the smaller CNV call. If a CNV call spanned the centromere of a chromosome, it was split into two calls, with the first one ending at the last probe before the centromere and the second one beginning at the first probe after the centromere.

## 2.5. CNV selection and prioritization

We then used PLINK to perform segmental analysis to identify variants that were present in the Saudi schizophrenia cases but not in the Saudi controls using the --mperm 10,000 option. This analysis randomly permutes the case/control status of the individuals 10,000 times and determines how many times CNVs overlap a particular locus in each replicate. This generates an empirical value of p that relays how often the observed number of cases overlapping a particular locus might be by chance. We followed up on these variants to confirm that they were rare among the in-house controls. Permutation of the phenotype (1 million replicates) was used to determine statistical significance.

## 2.6. Runs of homozygosity

PLINK was also used to identify regions where 100+ consecutive markers have homozygous genotypes, more than by chance alone. Variants with an allele frequency less than 1% were excluded from the ROH analyses. Since linkage disequilibrium (LD) can similarly confound these analyses, markers were pruned such that no two markers had a pairwise $r^2$ value greater than 0.30. Regions were then identified (using the PLINK parameter in parentheses) that were (1) at least one megabase (Mb) in size (--homozyg-kb 1,000), (2) spanned at least 100 SNPs (--homozyg-snp 100), (3) had a density of at least one SNP per 50 kb (--homozyg-density 50), (4) had no more than one variant per megabase having a heterozygous genotype (--homozyg-window-het 1), and (5) had no gap larger than 1 Mb (e.g., overlapping at centromere; --homozyg-gap 1,000).

## 3. Results

Demographic and clinical characterization of patients and controls are presented in Table 1. High quality data were available for 136 Saudi cases (130 male) and 97 Saudi controls (57 male), and for 4,625 common controls genotyped on the same genome-wide genotyping array platform and subjected to the same genotype and CNV analyses. Principal components analysis of the genotypes confirmed the ancestry of the Saudi cases and controls to be from Southwest Asia (see Supplementary Figure S1). One male sample had a marked partial loss of chromosome Y, which is the most common genetic aberration in men (Danielsson et al., 2020). This control male (SZ-C[−100]) was 50 years of age at DNA collection and the CNV loss is illustrated in Supplementary Figure S2 in which the median intensities for the sex chromosomes are plotted. No other samples had aneuploidy of a sex chromosome (no Klinefelter, Turner or Triple X syndromes, mosaic or complete, were observed). One individual (SZ-P-129) had an autosomal mosaic chromosomal arm that extended for 85% of chr15q (Figure 1), evidenced by the modest split in the intermediate BAF band and no deviation of the corresponding LRR. Intriguingly, two stretches of homozygosity were proximal to this pattern. The plot of BAF standard deviation (SD) and interquartile range (IQR) did not identify any other samples with a mosaic event, although individuals with large runs of homozygosity were identified (Figure 2).

Analysis of all 233 Saudi case and control individuals resulted in a total of 1,590 autosomal CNV calls that passed inclusion filtering criteria. The patient cohort had similar numbers of CNVs as the control group, but the average length of each CNV in the patient cohort was

TABLE 1 Demographic and clinical characterization of patients and controls.

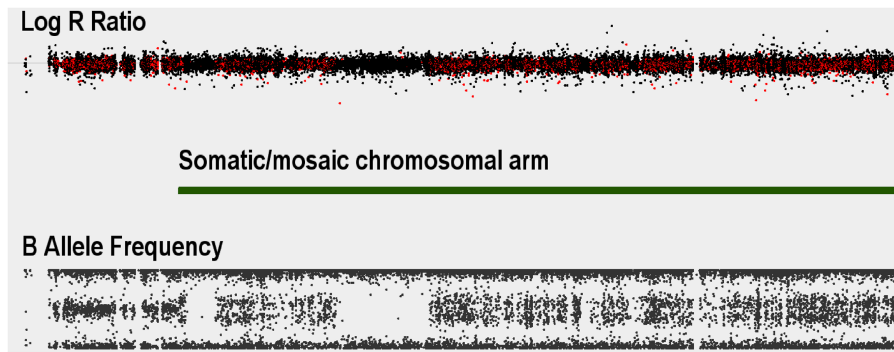| Parameter | Patients | Controls |
|---|---|---|
| Number of patients | 136 | 97 (+4,722 US) |
| Male (%) | 78.3 | 44.1 |
| Age ± SD (year) | 38 ± 10.6 | 50.9 ± 7.9 |
| Type of Schizophrenia (N) | | |
| Paranoid schizophrenia (%) | 105 (77.2) | |
| Residual schizophrenia (%) | 2 (1.5) | |
| Schizophrenia, unspecified (%) | 22 (16.2) | |
| Schizoaffective manic type (%) | 6 (4.5) | |

**FIGURE 1**
Autosomal somatic copy number variant example for a Saudi schizophrenia case: Log R Ratio (LRR) and B-allele frequency (BAF) values for a case with a somatic event spanning almost the entire length of chromosome 15q. The split in the BAF distribution starting at where Genvisis called the breakpoints for the mosaic event (green bar).
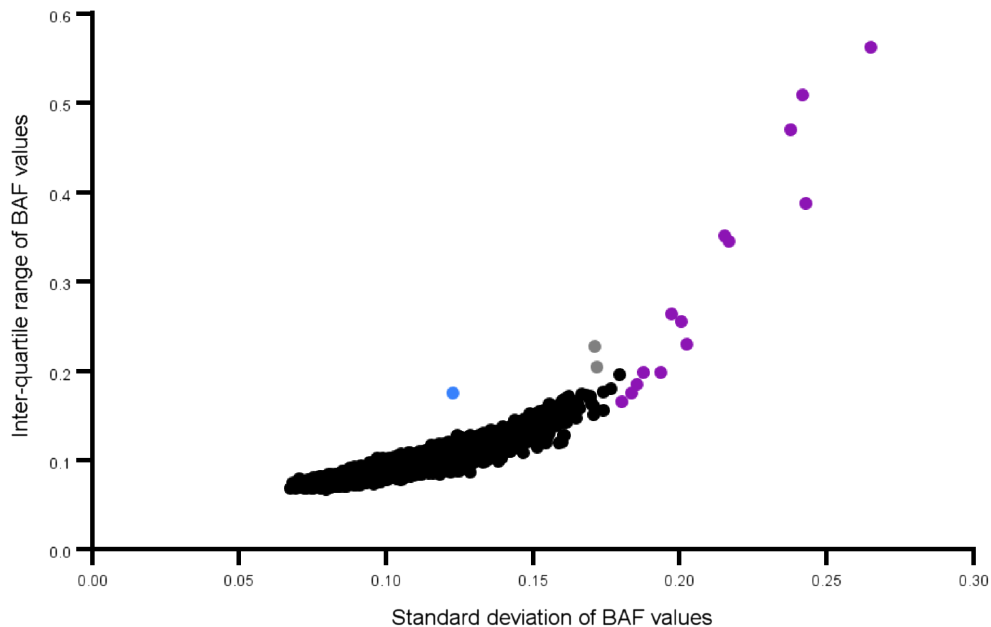


**FIGURE 2**
Mosaicism plot of all Saudi schizophrenia cases and controls: Each dot represents a chromosomal arm of an individual. Outliers for standard deviation and the interquartile range of the BAF values, are an effective way to screen for large mosaic events. The chr15q arm is shown in blue. The dots in purple are chromosomes where the entire arm or close to the entire arm is a long run of homozygosity. Only those B Allele Frequency (BAF) values between 0.15 and 0.85 are used in these calculations.

larger (125.7 kb compared to 86.15 kb), and therefore a larger proportion of the genome in the patient cohort was affected. However, these differences were not significantly different in our modestly sized cohort (empirically derived $p$-value = 0.32). This same pattern became significant when restricting the variants to those that were larger than 100 kb, where cases averaged 415.4 kb per variant and controls averaged 222.1 kb per variant ($p$ = 0.036) as shown on Table 2.

The segmental analysis did not reveal any variants at genome-wide significance (EMP2 < 0.05); however, two candidate regions were identified where a similarly sized CNV was present in more than one case, absent in the Saudi controls, and at least extremely rare in the in-house controls. These two regions on chromosome 7 and 9, respectively, and the affected genes are shown in Table 3. In addition, there was a particularly notable singleton CNV that was observed in

only one individual. This sample harbors a 16.5 megabase deletion on chromosome 10 (Figure 3). This deletion (chr10:52,931,142-69,418,051) spans at least 68 genes (listed in Supplemental Table S1), and consequently, it is likely pathogenic according to the ACMG recommendation. The next largest deletion observed is 1.4 Mb, which is similarly predicted to be pathogenic. Finally, 17 CNVs had previously been associated to schizophrenia in a large GWAS. In our relatively small cohort, we detected only one of these 17 CNVs, namely the 16p11, proximal duplication in a single case, and to cases with 22q11.2 deletions.

A large number of ROHs were identified in both cases and controls, and a skew was not found in either direction for number, rate, proportion, total affected or average size (Table 4). We identified five regions where 10 or more cases had an ROH overlapping a particular region. We then extended that region in both directions until fewer than
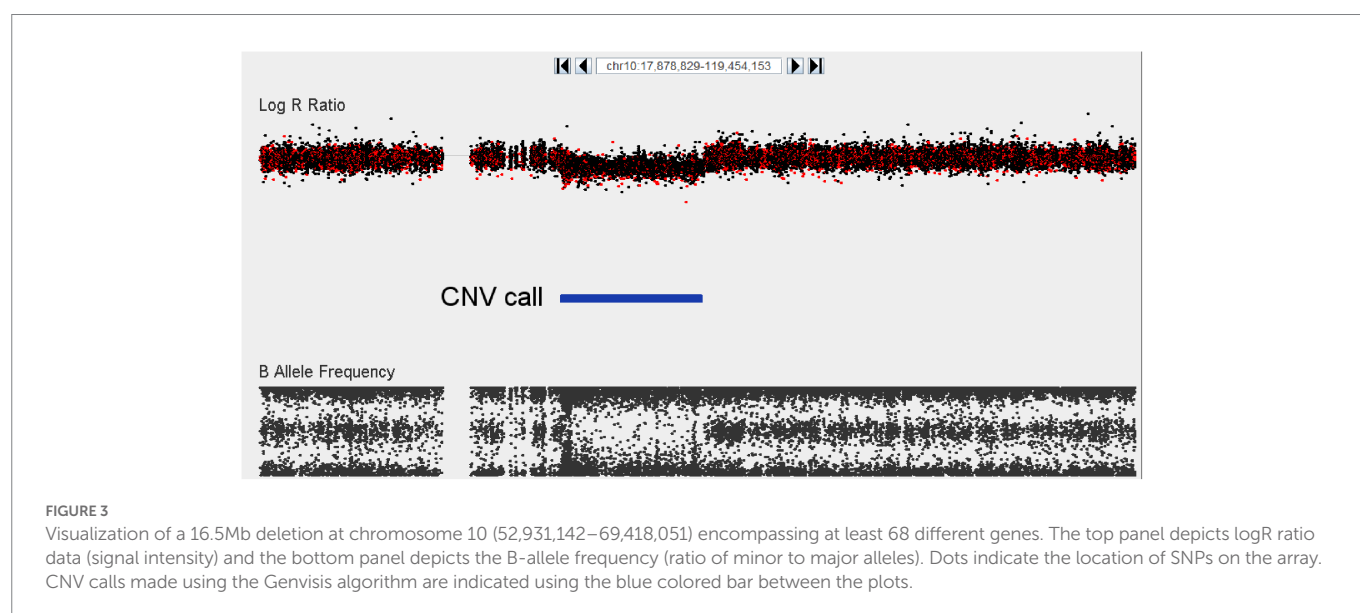
TABLE 2 Number of samples and events and a comparison of the rate of events.

| Parameter | Copy number variants (CNVs) that are >100kb | | | Runs of Homozygosity (ROHs) | | |
|---|---|---|---|---|---|---|
| | Cases | Controls | *p*-value | Cases | Controls | *p*-value |
| Number of individuals | 136 | 97 | | 136 | 97 | |
| Number of events | 197 | 164 | | 1884 | 1,357 | |
| Rate of events | 1.459 | 1.673 | 0.918 | 13.96 | 13.85 | 0.475 |
| Proportion of individuals with at least on event | 75.6% | 81.6% | 0.899 | 94.8% | 93.9% | 0.486 |
| Total amount of genome affected | 642.8 kb | 497.6 kb | 0.326 | 9.2 Mb | 9.1 Mb | 0.480 |
| Average size of event | 415.4 kb | 222.1 kb | 0.036 | 5.1 Mb | 5.0 Mb | 0.370 |

TABLE 3 Copy Number Variants present in more than one Saudi Schizophrenia case, zero matched controls, and rare in the American controls.

| Region | Type/Size | Number of cases (%) | Number of controls (%) | Number of in-house controls (%) | Genes affected |
|---|---|---|---|---|---|
| chr7:16,831,346-17,645,439 | Duplication 814 kb | 2 (0.8%) | 0 (0.0%) | 2 (0.04%) | AGR2, AGR3, AHR, LOC102659288, LOC101927630 |
| chr9:107,220,263-107,497,215 | Deletion 277 kb | 2 (0.8%) | 0 (0.0%) | 0 (0.0%) | OR13F1, OR13C4, OR13C3, OR13C8, OR13C5, OR13C2, OR13C9, OR13D1 |

Genes in these CNV regions are also listed.



FIGURE 3
Visualization of a 16.5Mb deletion at chromosome 10 (52,931,142−69,418,051) encompassing at least 68 different genes. The top panel depicts logR ratio data (signal intensity) and the bottom panel depicts the B-allele frequency (ratio of minor to major alleles). Dots indicate the location of SNPs on the array. CNV calls made using the Genvisis algorithm are indicated using the blue colored bar between the plots.

seven cases were overlapping the flanking regions, or a new control was seen to have an ROH. In addition to these five regions, we also identified five regions that had at least seven cases and zero controls with an ROH in that region. Each region and the number of individuals with an ROH in that region are delineated in Table 5.

## 4. Discussion

Several large association studies and clinical genetic studies have unambiguously shown the importance of CNVs, including both deletions and duplications, in understanding the pathology of schizophrenia. Furthermore, detection of a relevant pathogenic CNV can aid in diagnosis and counselling of patients. Recognizing the lack of data for schizophrenia cases with a specific Arabic ethnicity, we set out to assess the contribution of CNVs to the genetic risk of disease in the Saudi population. In our cohort of 136 schizophrenia cases, we detected one likely relevant variant and evidence for an increased average size of CNV compared to controls. These observations suggest that CNVs have putative biological relevance in the population of Saudi Arabia.

In our cohort of cases, there were three large CNVs encompassing loci previously associated with schizophrenia: one duplication of 16p11 (proximal) and two deletions in 22q11.2. Both of these are in "hotspot" where structural variants frequently form *via* non-allelic homologous recombination (NAHR) that is mediated by the repetitive DNA on either side of the breakpoints that are highly homologous (Kirov et al., 2009; Goldenberg, 2018). The 16p11 deletion has been robustly associated with psychiatric disorders, including intellectual disability, attention deficit

TABLE 4  Number of cases and controls with samples and events >100 kilobases in known regions.

| Locus | UCSC_hg19 | CNV type | Direction | #Case | #Cntrl | Reference |
|-------|-----------|----------|-----------|-------|--------|-----------|
| 1q21.1 | chr1:145,934,643-147,709,376 | Loss and gain | Risk | 0 | 8 | Levinson et al. (2011) |
| 2p16.3 NRXN1 | chr2:50,147,488-51,259,674 | Loss | Risk | 0 | 0 | Viñas-Jornet et al. (2014) |
| 3q29 | chr3:195,745,603-197,355,603 | Loss | Risk | 0 | 0 | Levinson et al. (2011) and Mulle et al. (2010) |
| 7q11.21 ZNF92 | chr7:64,838,768-64,865,998 | Loss and gain | Protective | 0 | 54 | Marshall et al. (2017) |
| 7q11.23 | chr7:72,742,064-74,142,064 | Gain | Risk | 0 | 0 | Mulle et al. (2014) |
| 7p36.3 VIPR2 WDR60 | chr7:15,8453,198-15,897,237 | Loss and gain | Risk | 0 | 0 | Levinson et al. (2011) and Vacic et al. (2011) |
| 8q22.2 VPS13B | chr8:100,025,494-100,889,808 | Loss | Risk | 0 | 0 | Marshall et al. (2017) |
| 9p24.3 DMRT1 | chr9:841,690-969,090 | Loss and gain | Risk | 0 | 2 | Marshall et al. (2017) |
| 13q12.11 ZMYM5 | chr13:20,411,593-20,437,773 | Hain | Protective | 0 | 3 | Marshall et al. (2017) |
| AS/PWS | chr15:2,482,000-2,843,000 | Gain | Risk | 0 | 0 | Rees et al. (2014a) |
| 15q11.2 | chr15:22,798,636-23,088,559 | Loss | Risk | 0 | 19 | Cox and Butler (2015) |
| 15q13.3 | chr15:31,132,708-32,482,708 | Loss | Risk | 0 | 1 | Sahoo et al. (2011) |
| 16p13.11 | chr16:1,551,000-1,630,000 | Gain | Risk | 0 | 0 | Ramalingam et al. (2011) |
| 16p11.2 (distal) | chr16:28,822,499-29,052,499 | Loss | Risk | 0 | 2 | Guha et al. (2013) |
| 16p11.2 (proximal) | chr16:29,652,499-30,202,499 | Gain | Risk | 1 | 1 | Dall'Olio et al. (2014) |
| 17p12 | chr17:1,416,000-1,543,000 | Loss | Risk | 0 | 3 | Kirov et al. (2009) |
| 17q12 | chr17:3,481,000-3,620,000 | Loss | Risk | 0 | 3 | Moreno-De-Luca et al. (2011) |
| 22q11.21 | chr22:19,020,000-21,420,000 | Loss | Risk | 2 | 0 | Bassett and Chow (2008) |
| 22q11.21 | chr22:19,020,000-21,420,000 | Gain | Protective | 0 | 3 | Rees et al. (2014b) |

TABLE 5  A list of regions with runs of homozygosity (ROH) in multiple Saudi Schizophrenia cases where few to zero are observed in controls.

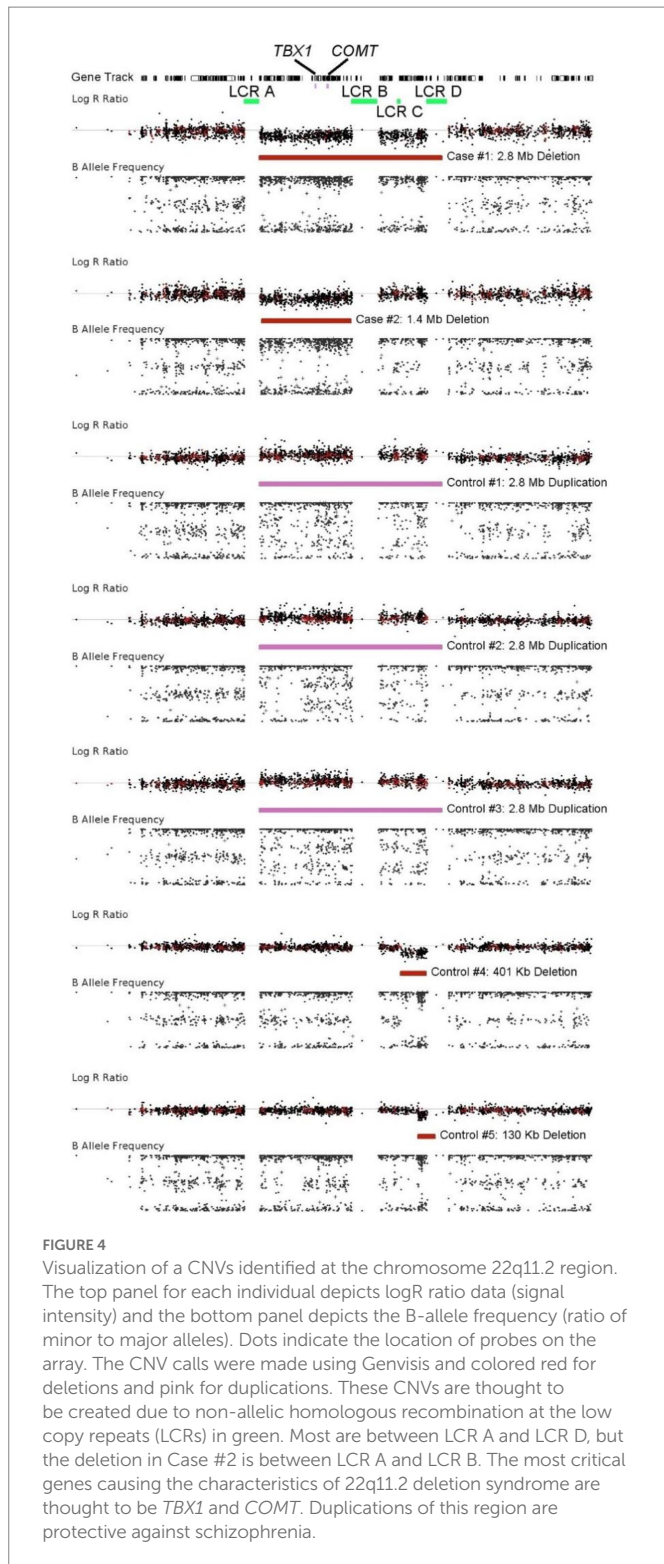| Region | minAFF | maxAFF | numUNAFF |
|--------|--------|--------|----------|
| chr11:41,633,193-43,738,673 | 8 | 9 | 0 |
| chr11:92,507,397-97,753,229 | 7 | 10 | 1 |
| chr11:100,189,226-103,910,105 | 8 | 13 | 2 |
| chr13:43,541,285-48,716,219 | 7 | 10 | 0 |
| chr14:25,142,208-48,716,219 | 7 | 8 | 0 |
| chr14:49,590,913-54,191,445 | 8 | 10 | 0 |
| chr14:61,894,646-62,670,326 | 7 | 7 | 0 |
| chr14:99,840,955-103,644,765 | 8 | 14 | 1 |
| chr15:100,328,822-101,076,889 | 8 | 8 | 0 |
| chr20:51,011,788-51,712,879 | 7 | 7 | 0 |

The maximum number of individuals with schizophrenia and an ROH (maxAFF) at any point in that region, as well as the minimum number of affected individuals at any given point in that region (minAFF) are indicated, as is the number of unaffected individuals with ROH events (numUNAFF).

hyperactivity disorder (ADHD), and autism spectrum disorder (Niarchou et al., 2019). 22q11.2 deletion syndrome (also known as DiGeorge syndrome) can lead to a variety of phenotypes including heart defects, cleft palate, hearing loss, intellectual disability, and delayed growth or speech development.

Which characteristics are present in a 22q11.2 deletion case may have to do with which genes are deleted. There are multiple low copy repeats (LCRs) in the region leading to multiple permutations of which genes are deleted. Most have a deletion of the full ~3 megabase region, but a "critical region" has been identified that includes the *TBX1* and *COMT* genes (Weksberg et al., 2007). Our own two cases corroborate this pattern, with both deletions spanning this region and Case #2 (Figure 4) having a far shorter deletion. Moreover, we see two smaller deletions in controls that are for only a small part of the distal region

(Figure 4) that do not overlap this critical region. In addition to deletions, duplications of this same region are noted in the general population and have been associated with protection against schizophrenia. In our study, we see three such duplications of the entire ~3 Mb region, all in controls (Figure 4).

Besides those at the known loci, only one likely pathogenic CNV was detected, the 16.5 Mb deletion on chromosome 10 that was detected in one patient that involved deletion of 68 genes. One patient showed evidence of a mosaic CNV involving 85% of the long arm of chromosome 15 (15q). This mosaic CNV overlaps with the recurrent 15q13.3 deletion that is robustly associated with schizophrenia. The signal from this case showed evidence for low-grade mosaicism, while it was unclear from the LRR whether this CNV is a duplication or a deletion, and therefore its clinical relevance remains unclear. Indeed,

**FIGURE 4**
Visualization of a CNVs identified at the chromosome 22q11.2 region. The top panel for each individual depicts logR ratio data (signal intensity) and the bottom panel depicts the B-allele frequency (ratio of minor to major alleles). Dots indicate the location of probes on the array. The CNV calls were made using Genvisis and colored red for deletions and pink for duplications. These CNVs are thought to be created due to non-allelic homologous recombination at the low copy repeats (LCRs) in green. Most are between LCR A and LCR D, but the deletion in Case #2 is between LCR A and LCR B. The most critical genes causing the characteristics of 22q11.2 deletion syndrome are thought to be *TBX1* and *COMT*. Duplications of this region are protective against schizophrenia.

schizophrenia. Therefore, the percentage of clinically relevant CNVs in our sample is slightly lower (~3%) as compared to that of previous studies that have mainly been performed in populations of European ancestry, which range between 6 and 8% (Costain et al., 2013; Daghsni et al., 2018). This observation may suggest that population specific factors, either genetic or non-genetic, may contribute to the disease risk in the Saudi population.

As can be expected for our relatively modest sample size, we could not detect an overall significant increase in CNV burden in cases versus controls. This is not surprising as it is well-known that although CNVs are robustly associated with high effect sizes with schizophrenia, they are present at low frequencies in cases. Indeed, the total number of CNVs observed in our case cohort was similar to that in controls. However, the average size of CNVs carried by cases was significantly larger than that of controls. Such an observation has been reported before and may relate to the probability that larger CNVs involve more genes than smaller CNVs, thus increasing the likelihood of pathogenicity. One limitation of our study is that most of our controls are from a slightly different ethnic background. It's possible that certain CNVs are more common in one population or another. That said, we had almost as many ethnically matched controls as we did cases, and the variants we did find were not present in those individuals, suggesting that they are not common in the Saudi population.

Finally, we observed a large number of regions showing a run of homozygosity in both cases and controls with no evidence of significant differences. Unlike a putatively causal CNV, a ROH is not thought to be causal in and of itself. It is merely an indication of the possibility that a rare recessive variant(s) may be found in the homozygous state in this region. For this reason, we favored regions where there were relatively few unaffected ROH events, but we did not require it when there were >10 cases.

Overall, our study shows evidence for an involvement of CNVs in the risk for schizophrenia in the population of Saudi Arabia. However, we cannot confirm that the rare, yet recurrent CNVs that are robustly associated in Western European populations, also contribute significantly to the risk of schizophrenia in this population. Further studies involving an increased sample size are needed to provide sufficient power to detect novel CNVs that may be unique to this population. Such a study would determine the benefit of genetic screening for clinically relevant CNVs to aid in diagnosis, prognosis, and management of schizophrenia. These data show a potential role of large CNVs as risk factors for schizophrenia in Saudi cases, and investigation into the genes that they affect may provide further insights into the etiology of schizophrenia in these populations.

## Data availability statement

The data presented in the study are deposited in the https://www.ebi.ac.uk/ena/browser/view/PRJEB59193 repository, accession number PRJEB59193.

## Ethics statement

The studies involving human participants were reviewed and approved by IRB Committee, Imam Abdulrahman Bin Faisal University.

large somatic/mosaic chromosomal aberrations (mCa), including somatic deletions and duplications that span almost an entire chromosome, are common in older individuals (~2% of individuals 85 years and older) but rare at younger ages.

These cases illustrate that known recurrent or other clinically relevant CNVs are also present in the Saudi population. One case presented a partial loss of the Y-chromosome that is observed frequently and associated with advanced age, which is likely unrelated to

The patients/participants provided their written informed consent to participate in this study.

## Author contributions

MA, AA, DA, and JJ were involved in the design of the work, critically revising the protocol, patient recruitment, data acquisition, and interpretation of data and drafting of the manuscript. KG, CC, and AA-A were involved in the design of the work, critically revising of protocol, laboratory work, data acquisition, and interpretation of data and drafting of the manuscript. NP, LC, BKo, and BKe were involved in the design of the work, analysis, and interpretation of data and drafting of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

The research project was financially supported by the Deanship of Research at Imam Abdulrahman bin Faisal University.

## Acknowledgments

The authors would like to acknowledge the financial support extended by the Deanship of Research at Imam Abdulrahman bin Faisal University. We are also grateful to the nurses and technical staff for their work and dedication.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnmol.2023.1069375/full#supplementary-material

## References

Bassett, A. S., and Chow, E. W. (2008). Schizophrenia and 22q11.2 deletion syndrome. *Curr. Psychiatry Rep.* 10, 148–157. doi: 10.1007/s11920-008-0026-1

Costain, G., Lionel, A. C., Merico, D., Forsythe, P., Russell, K., Lowther, C., et al. (2013). Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum. Mol. Genet.* 22, 4485–4501. doi: 10.1093/hmg/ddt297

Cox, D. M., and Butler, M. G. (2015). The 15q11.2 BP1-BP2 microdeletion syndrome: a review. *Int. J. Mol. Sci.* 16, 4068–4082. doi: 10.3390/ijms16024068

Daghsni, M., Lahbib, S., Fradj, M., Sayeb, M., Kelmemi, W., Kraoua, L., et al. (2018). TOP3B: a novel candidate gene in juvenile myoclonic epilepsy? *Cytogenet. Genome Res.* 154, 1–5. doi: 10.1159/000486945

Dall'Olio, G. M., Bertranpetit, J., Wagner, A., and Laayouni, H. (2014). Human genome variation and the concept of genotype networks. *PLoS One* 9:e99424. doi: 10.1371/journal.pone.0099424

Danielsson, M., Halvardson, J., Davies, H., Torabi Moghadam, B., Mattisson, J., Rychlicka-Buniowska, E., et al. (2020). Longitudinal changes in the frequency of mosaic chromosome Y loss in peripheral blood cells of aging men varies profoundly between individuals. *Eur. J. Hum. Genet.* 28, 349–357. doi: 10.1038/s41431-019-0533-z

Giaroli, G., Bass, N., Strydom, A., Rantell, K., and Mcquillin, A. (2014). Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: a systematic review of literature and meta-analysis. *Schizophr. Res.* 159, 340–346. doi: 10.1016/j.schres.2014.09.025

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Goldenberg, P. (2018). An update on common chromosome microdeletion and microduplication syndromes. *Pediatr. Ann.* 47, e198–e203. doi: 10.3928/19382359-20180419-01

Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S. E., et al. (2013). Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry* 70, 253–260. doi: 10.1001/2013.jamapsychiatry.71

Howrigan, D. P., Rose, S. A., Samocha, K. E., Fromer, M., Cerrato, F., Chen, W. J., et al. (2020). Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat. Neurosci.* 23, 185–193. doi: 10.1038/s41593-019-0564-3

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460

International Genetics and Translational Research in Transplantation Network (iGeneTRAiN) (2015). Design and implementation of the international Genetics and translational research in transplantation network. *Transplantation* 99, 2401–2412. doi: 10.1097/TP.0000000000000913

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., et al. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.* 18, 1497–1503. doi: 10.1093/hmg/ddp043

Kushima, I., Nakatochi, M., Aleksic, B., Okada, T., Kimura, H., Kato, H., et al. (2022). Cross-disorder analysis of genic and regulatory copy number variations in bipolar disorder, schizophrenia, and autism Spectrum disorder. *Biol. Psychiatry* 92, 362–374. doi: 10.1016/j.biopsych.2022.04.003

Lam, M., Chen, C. Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., et al. (2019). Comparative genetic architectures of schizophrenia in east Asian and European populations. *Nat. Genet.* 51, 1670–1678. doi: 10.1038/s41588-019-0512-x

Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., et al. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* 168, 302–316. doi: 10.1176/appi.ajp.2010.10060876

Li, Y. R., Van Setten, J., Verma, S. S., Lu, Y., Holmes, M. V., Gao, H., et al. (2015). Concept and design of a genome-wide association genotyping array tailored for transplantation-specific studies. *Genome Med.* 7:90. doi: 10.1186/s13073-015-0211-x

Lieberman, J. A., Stroup, T. S., Mcevoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., et al. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* 353, 1209–1223. doi: 10.1056/NEJMoa051688

Ma, L., Shcherbina, A., and Chetty, S. (2021). Variations and expression features of CYP2D6 contribute to schizophrenia risk. *Mol. Psychiatry* 26, 2605–2615. doi: 10.1038/s41380-020-0675-y

Ma, L., Tang, J., Wang, D., Zhang, W., Liu, W., Liu, X. H., et al. (2013). Evaluating risk loci for schizophrenia distilled from genome-wide association studies in Han Chinese from Central China. *Mol. Psychiatry* 18, 638–639. doi: 10.1038/mp.2012.63

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* 49, 27–35. doi: 10.1038/ng.3725

Moreno-De-Luca, D., SGENE CONSORTIUMMulle, J. G., Simons Simplex Collection Genetics ConsortiumKaminsky, E. B., Sanders, S. J., et al. (2011). Deletion 17q12 is a

recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.* 87, 618–630. doi: 10.1016/j.ajhg.2010.10.004

Morgan, C., Lappin, J., Heslin, M., Donoghue, K., Lomas, B., Reininghaus, U., et al. (2014). Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. *Psychol. Med.* 44, 2713–2726. doi: 10.1017/S0033291714000282

Mulle, J. G., Dodd, A. F., Mcgrath, J. A., Wolyniec, P. S., Mitchell, A. A., Shetty, A. C., et al. (2010). Microdeletions of 3q29 confer high risk for schizophrenia. *Am. J. Hum. Genet.* 87, 229–236. doi: 10.1016/j.ajhg.2010.07.013

Mulle, J. G., Pulver, A. E., Mcgrath, J. A., Wolyniec, P. S., Dodd, A. F., Cutler, D. J., et al. (2014). Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. *Biol. Psychiatry* 75, 371–377. doi: 10.1016/j.biopsych.2013.05.040

Nakatochi, M., Kushima, I., and Ozaki, N. (2021). Implications of germline copy-number variations in psychiatric disorders: review of large-scale genetic studies. *J. Hum. Genet.* 66, 25–37. doi: 10.1038/s10038-020-00838-1

Niarchou, M., Chawner, S. J. R., Doherty, J. L., Maillard, A. M., Jacquemont, S., Chung, W. K., et al. (2019). Psychiatric disorders in children with 16p11.2 deletion and duplication. *Transl. Psychiatry* 9:8. doi: 10.1038/s41398-018-0339-8

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* 50, 381–389. doi: 10.1038/s41588-018-0059-2

Peiffer, D. A., Le, J. M., Steemers, F. J., Chang, W., Jenniges, T., Garcia, F., et al. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 16, 1136–1148. doi: 10.1101/gr.5402306

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ramalingam, A., Zhou, X. G., Fiedler, S. D., Brawner, S. J., Joyce, J. M., Liu, H. Y., et al. (2011). 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. *J. Hum. Genet.* 56, 541–544. doi: 10.1038/jhg.2011.42

Rees, E., Kirov, G., Sanders, A., Walters, J. T. R., Chambert, K. D., Shi, J., et al. (2014b). Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol. Psychiatry* 19, 37–40. doi: 10.1038/mp.2013.156

Rees, E., Walters, J. T., Georgieva, L., Isles, A. R., Chambert, K. D., and Richards, A. L. (2014a). Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* 204, 108–114. doi: 10.1192/bjp.bp.113.131052

Sahoo, T., Theisen, A., Rosenfeld, J. A., Lamb, A. N., Ravnan, J. B., Schultz, R. A., et al. (2011). Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet. Med.* 13, 868–880. doi: 10.1097/GIM.0b013e3182217a06

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Skidmore, Z. L., Wagner, A. H., Lesurf, R., Campbell, K. M., Kunisaki, J., Griffith, O. L., et al. (2016). GenVisR: genomic visualizations in R. *Bioinformatics* 32, 3012–3014. doi: 10.1093/bioinformatics/btw325

Vacic, V., Mccarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., et al. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 471, 499–503. doi: 10.1038/nature09884

Viñas-Jornet, M., Esteba-Castillo, S., Gabau, E., Ribas-Vidal, N., Baena, N., San, J., et al. (2014). A common cognitive, psychiatric, and dysmorphic phenotype in carriers of NRXN1 deletion. *Mol. Genet. Genomic Med.* 2, 512–521. doi: 10.1002/mgg3.105

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., et al. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665–1674. doi: 10.1101/gr.6861907

Weksberg, R., Stachon, A. C., Squire, J. A., Moldovan, L., Bayani, J., Meyn, S., et al. (2007). Molecular characterization of deletion breakpoints in adults with 22q11 deletion syndrome. *Hum. Genet.*, 120, 837–845. doi: 10.1007/s00439-006-0242-x