Analysis of copy number variations at 15 schizophrenia-associated loci

Elliott Rees, James T. R. Walters, Lyudmila Georgieva, Anthony R. Isles, Kimberly D. Chambert, Alexander L. Richards, Gerwyn Mahoney-Davies, Sophie E. Legge, Jennifer L. Moran, Steven A. McCarroll, Michael C. O’Donovan, Michael J. Owen and George Kirov

Background
A number of copy number variants (CNVs) have been suggested as susceptibility factors for schizophrenia. For some of these the data remain equivocal, and the frequency in individuals with schizophrenia is uncertain.

Aims
To determine the contribution of CNVs at 15 schizophrenia-associated loci (a) using a large new data-set of patients with schizophrenia (n = 6882) and controls (n = 6316), and (b) combining our results with those from previous studies.

Method
We used Illumina microarrays to analyse our data. Analyses were restricted to 520766 probes common to all arrays used in the different data-sets.

Results
We found higher rates in participants with schizophrenia than in controls for 13 of the 15 previously implicated CNVs.

Conclusions
We strengthen the support for the majority of the previously implicated CNVs in schizophrenia. About 2.5% of patients with schizophrenia and 0.9% of controls carry a large, detectable CNV at one of these loci. Routine CNV screening may be clinically appropriate given the high rate of known deleterious mutations in the disorder and the comorbidity associated with these heritable mutations.

Declaration of interest
None.

Copy number variants (CNVs) are chromosomal rearrangements involving large segments of DNA (from 1000 and up to several million base pairs in length) that can be deleted, duplicated, inverted or translocated. A number of pathogenic CNVs are known to cause clinically recognisable syndromes, such as Williams–Beuren syndrome (WBS), Angelman/Prader–Willi syndrome (AS/PWS) and velocardiofacial syndrome (VCFS). Some CNVs are associated with a highly variable phenotype, which often includes both developmental and neuropsychiatric disorders. For example, 22q11.2 deletions, the first CNV to be associated with schizophrenia, also causes VCFS, which is characterised by intellectual disability and a number of medical problems such as palatal and skeletal anomalies and cardiac defects. Several large and rare CNVs have now been implicated in the aetiology of schizophrenia, reviewed by Malhotra & Sebat. They have been shown to substantially increase risk for defects. Several large and rare CNVs have now been implicated in the aetiology of schizophrenia, reviewed by Malhotra & Sebat.

As some of the CNVs are very rare (found in less than 1 in 1000 patients), notwithstanding the relatively large data-sets examined so far (ranging between ~5000 and 14000 people with schizophrenia), it is not clear that all those that have been implicated are true risk factors for the disorder. Thus, for nine of the loci that have received the strongest support in the literature, fewer than 15 observations have been made in people with schizophrenia. These nine loci are: duplications at 1q21.1, AS/PWS region, the WBS region, the 17q12 region, deletions at 1q21.1,10 the WBS region,16 the AS/PWS region,8 at VIPR2,16,17 at 16p13.117 and deletions at 3q29,10,12 distal 16p11.2,18 17q1219 and 17p12.20 The evidence for six loci is based only on single, albeit very large, studies: deletions at 17q12,19 distal 16p11.218 and 17p1220 and duplications at 1q21.1,19 AS/PWS region and at 16p13.117. Moreover, the rate of two loci, deletions at 15q11.2 and duplications at 16p13.11, has recently been challenged by a study that indicated that the rates among controls might be higher than originally reported, and no excess in individuals with schizophrenia was seen for these two CNVs in one of the largest studies. Finally, each previous report focused on the identification of one, or only a small number of CNV loci, and several reports used partially overlapping data-sets. Consequently, the rate of all previously reported risk CNVs has not yet been evaluated in a single independent data-set that is not biased by the inclusion of the original discovery sample. We set out to evaluate these specific CNVs in a sample of 6882 people with schizophrenia that passed quality control. This represents the largest single CNV data-set yet reported in schizophrenia, and nearly doubles the sample size of patients from the total world literature for many loci. For control individuals, we used publicly available data from 6316 samples genotyped on arrays with a resolution similar to the case group. These data-sets are completely independent from samples used in previous studies or reviews that established the rates of these CNVs in schizophrenia.

Method
Samples
Case group
We collected data on patients with schizophrenia (case group) in two waves, which we call (a) CLOZUK and (b) CardiffCOGS
Copy number variations at 15 schizophrenia-associated loci

(25) All CNVs were subsequently excluded due to poor quality control or for being duplicates. The CLOZUK sample has been described elsewhere.21 The CardiffCOGS (n = 571) is a sample of patients with clinically diagnosed schizophrenia recruited from community, in-patient and voluntary sector mental health services in the UK. Interview with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) instrument22 and case-note review was used to arrive at a best-estimate lifetime diagnosis according to DSM-IV criteria.23

Control group
As controls we used publicly available data downloaded from dbGAP (www.ncbi.nlm.nih.gov/gap). To avoid CNV detection biases between arrays with different probe densities and platforms, we chose data-sets genotyped with Illumina arrays that had a high overlap with the probes used to genotypt the case group. The following data-sets were used: 1491 participants from the USA that took part in a study on smoking cessation, 3102 participants from the USA who took part in a study on melanoma and 1869 participants from Germany who took part in a study of refractive error (KORA study). Participants in the ‘smoking’ and ‘melanoma’ data-sets are cases or controls from these studies, whereas the KORA data-set is a population-based study where participants had refractive error measurements. The ethnicities of participants were derived from principal components analysis (PCA). A total of 91.4% of samples that passed quality control of participants were derived from European ancestry. Full details on these data-sets are presented in the online supplement (section 1).

Genotyping and quality control filtering
Raw intensity files from each data-set were independently processed to account for potential batch effects. PennCNV24 was used for CNV detection. To avoid cross-platform biases, we restricted CNV calling to the 520,766 probes present on all arrays used. Samples were excluded if any one of the following standard quality control statistics constituted an outlier within their source data-set: log R ratio (LRR) standard deviation, B-allele frequency (BAF) drift, wave factor (WF) and total number of CNVs (online supplement, section 2). Out of 15,591 samples with array data, 393 were excluded due to poor quality control or for being duplicates of the same individual after testing for identity-by-descent. The 8.6% non-European participants were retained in the analysis, to ensure that our data were comparable with those in recent reviews of CNV loci in schizophrenia.14 The numbers and ethnicities of these participants are listed in online Table DS2. The final numbers after exclusions for quality control and duplicates was 6882 in the case group and 6316 in the control group. The quality control process for individual CNVs is detailed in the online supplement (section 2). Briefly, CNVs were included if they were >10 kb in size and had a frequency <1%, (applying filters with PLINK version 1.07).25 All CNVs were subsequently required to pass a median z-score outlier method of validation26 that helps to remove false-positive CNVs, and to identify any missed CNVs.

Statistical analysis
For the analysis of CNV loci we used Fisher’s exact test (1-tailed as we were testing prior hypotheses). We performed a meta-analysis by adding our new data to those in the literature, applying a 2-tailed Fisher’s exact test to the combined data. As these CNVs have been shown to be subjected to strong negative selection pressure, their frequencies in the population essentially reflect the mutation rate and selection pressure operating against them.1,27 They are therefore less likely to be subjected to population stratification caused by genetic drift, in the way common variants are (although there are some examples of ethnic differences in the mutation rates). Therefore, in the interests of clarity, we do not stratify the sample by ethnicity but provide the full breakdown of the data in the different ethnic groups in online Table DS3. The results do not change appreciably if they are restricted to only the European populations that comprise 91.4% of the sample.

There is no accepted convention as to what constitutes genome-wide significance for a CNV locus. Girirajan and colleagues1 reported 72 recurrent CNVs that can cause a neuro-developmental disorder, although in total, 120 genomic regions are potentially prone to recurrent CNVs because they are flanked by segments of high homology, called segmental duplications.28 This suggests that a Bonferroni correction for multiple testing of recurrent CNVs (those flanked by segmental duplications) might require a P-value of <4.1 × 10^{-4} to be accepted as a significant association for this type of CNV (P = 0.05/120). Regarding associations with individual genes, a conservative Bonferroni correction would require correcting for the testing of ~20000 genes, or P < 2.5 × 10^{-5} (P = 0.05/20000).

Choice of CNVs for analysis
The list of previously implicated CNVs was taken from the largest meta-analysis to date.14 To this list we added three loci: exon-disrupting deletions at the NRXN1 gene, as there is consensus that they increase risk for developing schizophrenia;7 5 deletions at distal 16p11.2, the evidence for which was published after the above review;18 and duplications at the WBS region, a locus that just failed to reach significance in that review, but received support in a subsequent study.16 For duplications at the AS/PWS region, we tested their parental origin using a DNA methylation-sensitive high-resolution melting curve analysis,29 as previous research suggested that the maternal ones are specifically implicated6 (online supplement, section 5).

Results
The rates of CNVs among the case and control groups are presented in Table 1. For 13 out of the 15 CNVs, we found higher rates in the case than in the control group. For six of these, the difference was nominally significant in this new sample alone (Table 1).

Of the loci where previous evidence was modest, the most striking result was for AS/PWS duplications, where we found eight in patients and none in controls (P = 0.0055). When combined with previous data (Table 2, for a more detailed version that includes results from previous studies see online Table DS6), this CNV is now a strongly supported schizophrenia risk variant. Moreover, as this is an imprinted locus, we tested the parental origin of our eight duplications, and all were shown to be maternal in origin, similar to the original publication.6 Duplication at 16p13.11, another previously weakly associated CNV, now also shows strong evidence in the combined data (Table 2).
The only instances in our study where CNVs were more common in the control than in the case group concern duplications at VIPR2 and deletions at distal 16p11.2. In neither case is the excess of CNVs in the control group significant. In the case of VIPR2, meta-analysis is no longer supportive, whereas the evidence for association at distal 16p11.2 remains nominally significant (Table 2 and Table D56). Analysis of only individuals of European descent gave essentially the same results. The distribution of CNVs in the different data-sets and ethnic groups is presented in Table D53.

In the present sample, which is not subject to the potential bias of including the original studies that discovered the associations, about 2.5% of the case group and 0.9% of the control group carry one of the CNVs in Table 1, a highly significant excess ($P = 1.4 \times 10^{-12}$). Only four individuals in the case group carry two of these CNVs (online supplement, section 6).

In the analysis of all 15 loci in the combined data (Table 2), all but one of the CNVs showed significant evidence of association. For 11 of these, the significance surpasses the threshold for multiple testing correction that we suggest in the Method ($P < 4.1 \times 10^{-4}$). For many of them the statistical significance is greatly improved compared with the previous results, most strikingly for 15q11.2, AS/PWS, 16p11.31, 16p11.2 and 22q11.2, where the $P$-values were strengthened by several orders of magnitude.

### Discussion

In an analysis of the largest single schizophrenia sample to date, we establish more accurate estimates of risk from individual CNVs in an independent sample, and estimate the total burden of susceptibility conferred by this group of CNVs. The vast majority

---

**Table 1** Findings from our data-set for previously implicated copy number variation (CNV) loci in schizophrenia

<table>
<thead>
<tr>
<th>Locus</th>
<th>Position in Mb</th>
<th>Case group (n = 6882)</th>
<th>Control group (n = 6316)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1 del</td>
<td>chr1:146,57-147,39</td>
<td>12 0.17 1 0.016</td>
<td>11.03 (1.43-84.88)</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>1q21.1 dup</td>
<td>chr1:146,57-147,39</td>
<td>8 0.12 5 0.079</td>
<td>1.47 (0.48-4.49)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>NR007 del</td>
<td>chr2:50,15-51,26</td>
<td>11 0.16 0 0.00</td>
<td>NA (1.25-∞)</td>
<td>7.7 \times 10^{-4}</td>
<td></td>
</tr>
<tr>
<td>3q29 del</td>
<td>chr3:195,73-197,34</td>
<td>4 0.058 0 0.00</td>
<td>NA (0.44-∞)</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>WBS dup</td>
<td>chr7:72,74-74,14</td>
<td>3 0.044 1 0.016</td>
<td>2.75 (0.29-26.48)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>VIPR2 dup</td>
<td>chr7:158,82-158,94</td>
<td>1 0.015 6 0.095</td>
<td>0.15 (0.02-1.27)</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>15q11.2 del</td>
<td>chr15:22,80-23,09</td>
<td>44 0.64 26 0.41</td>
<td>1.56 (0.96-2.53)</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>AS/PWS dup</td>
<td>chr15:24,82-24,43</td>
<td>8 0.12 0 0.00</td>
<td>NA (0.90-∞)</td>
<td>0.0055</td>
<td></td>
</tr>
<tr>
<td>15q13.3 del</td>
<td>chr15:31,13-32,48</td>
<td>4 0.058 2 0.032</td>
<td>1.84 (0.34-10.03)</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>16p13.11 dup</td>
<td>chr16:15,31-16,30</td>
<td>24 0.35 12 0.19</td>
<td>1.84 (0.92-3.68)</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>16p11.2 distal del</td>
<td>chr16:28,82-29,05</td>
<td>0 0.00 2 0.032</td>
<td>NA (0-3.82)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>16p11.2 dup</td>
<td>chr16:29,64-30,20</td>
<td>27 0.39 0 0.00</td>
<td>NA (3.09-∞)</td>
<td>2.3 \times 10^{-4}</td>
<td></td>
</tr>
<tr>
<td>17p12 del</td>
<td>chr17:14,16-15,43</td>
<td>4 0.058 3 0.047</td>
<td>1.22 (0.27-5.47)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>17q12 del</td>
<td>chr17:34,81-36,20</td>
<td>1 0.015 0 0.00</td>
<td>NA (0.11-∞)</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>22q11.2 del</td>
<td>chr22:19,02-20,26</td>
<td>20 0.29 0 0.00</td>
<td>NA (2.8-∞)</td>
<td>2.2 \times 10^{-4}</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>171 2.48 58 0.92</td>
<td>1.4 \times 10^{-12}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2** Combined results of previous studies and the current data-set

<table>
<thead>
<tr>
<th>Locus</th>
<th>Case group (n)</th>
<th>Control group (n)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1 del</td>
<td>1.3 \times 10^{-9}</td>
<td>0.17 (33/19,056)</td>
<td>8.35 (4.65-14.99)</td>
<td>4.1 \times 10^{-13}</td>
</tr>
<tr>
<td>1q21.1 dup</td>
<td>2.0 \times 10^{-6}</td>
<td>0.13 (21/16,247)</td>
<td>3.45 (1.92-6.20)</td>
<td>9.9 \times 10^{-5}</td>
</tr>
<tr>
<td>NR007 del</td>
<td>7.9 \times 10^{-7}</td>
<td>0.18 (33/18,752)</td>
<td>9.01 (4.44-18.29)</td>
<td>1.3 \times 10^{-5}</td>
</tr>
<tr>
<td>3q29 del</td>
<td>2.3 \times 10^{-8}</td>
<td>0.082 (14/17,005)</td>
<td>57.65 (7.58-438.44)</td>
<td>1.5 \times 10^{-4}</td>
</tr>
<tr>
<td>WBS dup</td>
<td>5.5 \times 10^{-7}</td>
<td>0.066 (24/12,269)</td>
<td>67.63 (2.58-49.93)</td>
<td>6.9 \times 10^{-7}</td>
</tr>
<tr>
<td>VIPR2 dup</td>
<td>0.006</td>
<td>0.11 (15/14,218)</td>
<td>1.54 (0.77-3.09)</td>
<td>0.27</td>
</tr>
<tr>
<td>15q11.2 del</td>
<td>2.2 \times 10^{-7}</td>
<td>0.59 (116/19,547)</td>
<td>2.15 (1.71-2.68)</td>
<td>2.5 \times 10^{-6}</td>
</tr>
<tr>
<td>AS/PWS dup</td>
<td>0.014</td>
<td>0.083 (12/14,464)</td>
<td>13.20 (3.72-46.77)</td>
<td>5.6 \times 10^{-4}</td>
</tr>
<tr>
<td>15q13.3 del</td>
<td>2.1 \times 10^{-11}</td>
<td>0.14 (26/18,571)</td>
<td>7.52 (3.98-14.19)</td>
<td>4.0 \times 10^{-5}</td>
</tr>
<tr>
<td>16p13.11 dup</td>
<td>0.03</td>
<td>0.31 (37/12,029)</td>
<td>2.30 (1.57-3.36)</td>
<td>5.7 \times 10^{-5}</td>
</tr>
<tr>
<td>16p11.2 distal del</td>
<td>0.0014</td>
<td>0.063 (13/20,732)</td>
<td>3.39 (1.21-9.52)</td>
<td>0.017</td>
</tr>
<tr>
<td>16p11.2 dup</td>
<td>3.2 \times 10^{-14}</td>
<td>0.35 (58/16,772)</td>
<td>11.52 (6.86-19.34)</td>
<td>2.9 \times 10^{-5}</td>
</tr>
<tr>
<td>17p12 del</td>
<td>0.0004</td>
<td>0.094 (12/17,773)</td>
<td>3.62 (1.73-7.57)</td>
<td>0.0012</td>
</tr>
<tr>
<td>17q12 del</td>
<td>0.004</td>
<td>0.036 (5/14,024)</td>
<td>6.64 (1.78-24.72)</td>
<td>0.0072</td>
</tr>
<tr>
<td>22q11.2 del</td>
<td>1.0 \times 10^{-30}</td>
<td>0.29 (56/19,084)</td>
<td>NA (28.27-∞)</td>
<td>4.4 \times 10^{-80}</td>
</tr>
</tbody>
</table>

**Table 1 Notes:**
- del, deletion; dup, duplications; NA, not applicable; WBS, Williams–Beuren syndrome; AS/PWS, Angelman/Prader–Willi syndrome.
- a. Copy number variation positions are in UCSC Build 37. Significant results are in bold (using Fisher exact test, 1-tailed).

**Table 2 Notes:**
- del, deletion; dup, duplications; NA, not applicable; WBS, Williams–Beuren syndrome; AS/PWS, Angelman/Prader–Willi syndrome.
- a. For a more detailed version of this table that includes the CNV frequency, % (n/N) from previous studies see online Table D56. $P$-values are based on Fisher exact test, 2-tailed.
of patients in this study were recruited on the basis that they have a
diagnosis of treatment-resistant schizophrenia according to their
psychiatrist and were taking clozapine for that indication. The
availability of clinician diagnoses allowed us to exclude the limited
number of samples from individuals with diagnoses other than
schizophrenia. The convention for psychiatric samples has been
that patient inclusion is based on research diagnoses arrived at
following detailed interview and phenotyping procedures (as is
the case in the CardiffCOGS sample in this study). As genome
analysis has become more affordable than the establishment of a
formal research diagnosis, the latter has now become the limiting
step for exploiting the microarray technology. The CLOZUK
sampling method offers a pragmatic approach to recruit unusually
large numbers of patients with schizophrenia, and nearly doubles
the number of patients with schizophrenia analysed in the
previous literature. The use of such samples is supported by
evidence that with the use of operationalised criteria, clinician
diagnoses of schizophrenia have high specificity and positive
predictive values when validated against research-based
approaches. Additionally, we have reported findings that
support the validity of the individuals in CLOZUK as a
schizophrenia sample with genetic data, by demonstrating that
of the most strongly associated schizophrenia alleles in the
Psychiatric Genetics Consortium Stage 1, 85% (66/78) showed
the same direction of effect in the CLOZUK sample, sign test
P = 1.7 x 10^-14. In the present study, the findings of very similar
rates of susceptibility CNVs in the CLOZUK sample compared with
previous samples (Table D56), recruited using conventional
methods, support the comparability of the two types of samples.

In order to reduce the potential bias of using different arrays,
we used only Illumina platforms and only analysed those probes
common to all arrays. We used the z-score method to both
validate each CNV and check whether any CNV in the regions
in Table 1 had been missed.

The current study provides support for most previously
implicated CNVs, as we found higher rates in the case group than
in the control group for 13 of the 15 loci. The support
is particularly strong for duplications at 16p11.2 (P = 2.3 x 10^-4)
and at the AS/PWS critical region (P = 0.0055) and for deletions
at 22q11.2 (P = 2.2 x 10^-6), 1q21.1 (P = 0.0027), NRXN1
(P = 7.7 x 10^-4) and 15q11.2 (P = 0.046). All eight duplications
at the AS/PWS region were of maternal origin (online supplement,
section 5), thus supporting the original report. Two loci:
deletions at 15q11.2 and duplications at 16p13.11, that were not
supported in two recent papers, also receive support in the current
study and the statistical significance of their overall
association with schizophrenia is strengthened by several orders
of magnitude (Table 2).

Four of the loci in Table 2 do not surpass a significance
threshold that corrects for the multiple testing of large recurrent
CNVs (P = 4.2 x 10^-5), or for individual genes (P < 2.5 x 10^-7)
(see Method): duplications at VIPR2, and deletions at distal
16p11.2, 17p12 and 17q12.

**Burden of schizophrenia-associated CNVs**

Overall, 2.5% of the case group v. 0.9% of the control group carry
one or more of these CNVs. This is highly significant in this
completely independent data-set (P = 1.4 x 10^-14). They are
associated with a range of odds ratios and each locus clearly makes
a different contribution to the increase in risk (Table 2). They are
also known to increase risk for other disorders, such as epilepsy
(15q11.223 and 15q13.333), congenital heart disease (1q21.144 and
22q11.235), attention-deficit hyperactivity disorder (16p13.1126)
and obesity (distal 16p11.227), and all but two (at VIPR2 and
17p12) increase risk for developmental delay and autism spectrum
disorders. The overall contribution is modest but the effect size is
sufficiently large to suggest that if seen in a patient, it is very likely
to be relevant to the disorder, although not sufficient to account
for the disease.

**Summary of the findings for the individual loci**

1q21.1 deletions and duplications

Deletions at 1q21.1 were among the first implicated loci. Our
new data provide strong support for their role in schizophrenia
(P = 0.0027), with a frequency in the case group identical to that
in previous reports (0.17%). Confirming an approximately
eightfold excess of deletions among patients, with an extremely
strong statistical support of P = 4.1 x 10^-15 in the combined
literature (Table 2). Duplications at this locus have only been
implicated in one study, but with the addition of our data
(although not significant on its own), the evidence for
duplications at this locus in schizophrenia is now stronger
(P = 9.9 x 10^-5).

NRXN1 deletions

The gene NRXN1 encodes for a presynaptic cell adhesion protein,
which binds with postsynaptic proteins called neuroligins and
plays a vital role in the formation, maintenance and release of
neurotransmitters in synapses. Exonic deletions disrupting this
gene have been consistently implicated in schizophrenia, autism
spectrum disorder. The current study confirms their role, as we found 11 exonic deletions in the case group (0.16%) and
none in the control group. P = 7.7 x 10^-4. This brings the
overall significance in the combined literature to P = 1.3 x 10^-11
(easily surpassing our multiple testing correction threshold for
individual genes of P < 2.5 x 10^-4), with approximately a ninefold
excess in the case group (Table 2). The positions of exon-disrupting
CNVs at this locus in our new data-set are shown in the online
supplement, section 4.

3q29 deletions

The role of this CNV in schizophrenia was first reported by Mulle
et al. and confirmed by Levinson et al. The finding of four
individuals in the case group with deletions and none in the
control group just fails to reach significance in our independent
sample (P = 0.074) but can be regarded as supportive independent
confirmation. With an overall significance of P = 1.5 x 10^-3, it is
another locus where the evidence for a role in schizophrenia is
very strong. Only one such deletion has been found in nearly
70,000 controls, indicating it is highly penetrant (Table 2).

WBS duplications

The reciprocal duplication of the WBS region was first implicated
as increasing the risk for autism. The region was implicated in
schizophrenia after the finding of a de novo duplication and
reached statistical support in a large collaborative study. The
finding of three individuals in the case group and one in the
control group with this duplication in our study constitutes
modest support, but the overall strength of the evidence remains
strong, at P = 6.9 x 10^-9.

VIPR2 duplications

The evidence for duplications disrupting this gene comes from two
studies that used largely overlapping samples. The overall
evidence from the previous literature was modest: P = 0.086. As
we found duplications in six people in the control group and only
one in the case group, the evidence in favour of this locus is
nevertheless significant in the combined literature (P = 0.27). The
positions of CNVs at this locus in our new data-set are shown in the online supplement (section 4). Most CNVs at this gene are large and covered with a high number of probes on the arrays (medians of 381 kb and 66 probes, details in online Table DS4), therefore they should be called reliably on these arrays. We also examined the region with all available probes for CNV calling on the different arrays, and with the z-score method, and found no additional CNVs that had been missed. The rate in our control group is slightly higher than in previous studies (0.095% v. 0.059%, Table 1 and Table DS6), although this difference is not significant (P = 0.4).

15q11.2 deletions
This was one of the first CNVs implicated in schizophrenia. However, a recent report of a higher rate in controls than that reported in the original paper and the lack of support in the study by Levinson et al clearly indicated the need for replication. Here we found independent significant evidence for association (P = 0.046), that strengthens the evidence in the combined analysis to \( P = 2.5 \times 10^{-10} \) (Table 2).

AS/PWS duplications
The modest prior evidence for the role of this duplication of the AS/PWS critical region in schizophrenia comes from a single publication with just four observations in patients. We found another eight CNVs in our case group and none in the control group (P = 0.0055), thus substantially strengthening this finding. Even more notably, by using DNA methylation-sensitive high-resolution melting-curve analysis of the SNRPN locus, which does not require DNA from the parents, we found that all eight duplications are of maternal origin, as in the original study. These findings further underline the importance of imprinted genes (those genes subject to parent of origin specific epigenetic regulation) in the aetiology of psychosis and other neurodevelopmental disorders. These duplications are among the most common genetic susceptibility factors for autism spectrum disorder, where they are found in nearly 1:500 cases.

15q13.3 deletion
This is also among the first implicated CNVs in schizophrenia, and received further strong support in the study by Levinson et al. It was found at a ~tenfold higher rate in patients with schizophrenia, with a strong statistical support in the previous literature, \( P = 2.1 \times 10^{-11} \). Although we found only a twofold excess in our case group in our new sample, at 0.058% v. 0.032%, \( P = 0.38 \), its role as a susceptibility factor for schizophrenia remains very strong in the combined literature \( P = 4.0 \times 10^{-10} \). This locus also increases the risk for epilepsy.

16p13.11 duplication
The previous evidence for this CNV comes mostly from a single study and the overall number of analysed patients with schizophrenia so far, at 5147 individuals, is smaller than in our new data-set. Two recent studies weakened the evidence for this locus, and in the recent review by Mallotra & Sebat it had very weak statistical support, \( P = 0.03 \). Here we found 24 CNVs in the case group and 12 in the control group, an excess that just fails to reach significance, \( P = 0.056 \). Combined with the earlier data, our study strengthens the statistical support for this locus in schizophrenia to \( P = 5.7 \times 10^{-5} \). This CNV has also been implicated in attention-deficit hyperactivity disorder.

16p11.2 duplication
This is our strongest finding, with a \( P = 2.3 \times 10^{-8} \) in the discovery sample alone, and a combined evidence at \( P = 2.9 \times 10^{-4} \), with an odds ratio of over 11. This duplication is also one of the strongest autism spectrum disorder CNV risk factors.

16p11.2 distal deletion
This CNV was suggested to confer susceptibility to schizophrenia by Guha et al. We found no support for this locus in the present study, with two deletions in the control group and none in the case group. However, the lack of deletions among our case group could be as a result of ascertainment bias. Obesity is found in at least 50% of 16p11.2 distal deletion carriers. Clozapine produces the most severe weight gain among all antipsychotics and the most common cause for the reluctance of UK psychiatrists to prescribe clozapine is the potential weight gain. It is therefore possible that psychiatrists are less likely to give clozapine to carriers of 16p11.2 distal deletion (as such patients are more likely to be obese already), thus potentially reducing the frequency of this CNV in the CLOZUK sample. The combined result from the literature retains the statistical support for the role of this locus \( P = 0.017 \), but it clearly requires testing in further data-sets.

17p12 deletion
This deletion causes the neurological disorder hereditary neuropathy with liability to pressure palsies and was implicated in schizophrenia on the basis of only eight observations in the original case group. Although the overall statistical evidence is still in favour of it increasing the risk for schizophrenia \( P = 0.0012 \), the support is not compelling, raising the need for further replication. It does not increase the risk for either developmental delay or autism, unlike most other loci discussed here, and it is possible that the original report was a false-positive finding.

17q12 deletion
Originally this locus had been known to cause renal cysts and diabetes. It was identified as a susceptibility CNV for schizophrenia with only four observations in the case group in a study that also implicated its role in autism spectrum disorder. We found only one deletion in a patient, a rate a quarter of that in the original study. Although we found no deletions in the control group, the overall P-value of all data is not sufficiently robust to definitively conclude that this is a schizophrenia susceptibility locus \( P = 0.0072 \). Therefore, this locus also requires testing in further schizophrenia data-sets.

22q11.2 deletion
This was the first CNV implicated in schizophrenia and has received extensive replication over the years. It affects ~40 genes and leads to a variety of physical anomalies. The latest review found a rate of 0.30% in patients and 0% in controls. This rate is practically identical to our new sample, where we found 20 carriers in the case group (0.29%) and none in the control group. It remains the most significantly associated CNV in schizophrenia: \( P = 4.4 \times 10^{-4} \).

Implications
Out of 15 previously implicated CNV loci, 11 are now strongly associated with schizophrenia from the combined results of the previous literature and our new data. The evidence for the remaining four loci should be regarded as still equivocal and requiring further investigation. Our findings indicate that approximately 2.5% of individuals with schizophrenia carry at least one known pathogenic CNV. The odds ratios of these CNVs...
in relation to schizophrenia range between ~2 and >50 and nearly all of them are also associated with a range of other neurodevelopmental disorders, such as autism spectrum disorder and intellectual deficit.\(^1\) Moreover, a number of the individual pathogenic CNVs are associated with particular physical disease phenotypes, such as congenital heart disease (1q21.1 and 22q11.2), microcephaly (1q21.1, 3p29 and 16p11.2) and obesity (16p11.2 distal),\(^3,4,25,26\) Given their frequency, these findings suggest that routine screening for CNVs should be made available and that the results will have immediate implications for genetic counselling, and given their comorbidity with other medical disorders, for patient management as well. The robust identification of 11 relatively high penetrance risk alleles for schizophrenia also offers promise for biological research aimed at developing animal and cellular models for the identification of novel disease mechanisms and drug targets.

**Funding**

The work at Cardif University was funded by Medical Research Council (MRC) Centre (G0800509 and Program Grants (G0801410) and the European Community’s Seventh Framework Programme (HEALTH-F2-2010-241909 Project EU-GEI), and an MRC PhD Studentship to A.T.G.S., B.C.D. and C.G.C. have received funding from the MRC and the Wellcome Trust, UK. A.R.I. has received funding from the Biotechnology and Biological Sciences Research Council UK, the Wellcome Trust and the Leverhulme Trust. This work was supported by a clinical research fellowship to J.T.R.W. from the MRC/Welsh Assembly Government and the Margaret Temple Award from the British Medical Association. The schizophrenia samples were genotyped at the Broad Institute, USA, and funded by a philanthropic gift to the Stanley Center for Psychiatric Research.

**Acknowledgements**

We thank the participants and clinicians who took part in the CardifFROGGY study. We acknowledge Andrew Iles, David Parlow, Carissa Phillipat and Sophie Canton for their work in recruitment, interviewing and rating. For the CLOZUK sample we thank Novartis for their guidance and cooperation. We also thank staff at The Doctor’s Laboratory, in particular Lisa Levett and Andrew Levett, for help and advice regarding sample acquisition. We acknowledge Kiran Mamprapagada, Lesley Ilates, Catherine Brenser and Lucinda Hopkins for laboratory sample management.

The authors acknowledge the contribution of data from outside sources: (a) Genetic Architecture of Smoking and Smoking: Cessation accessed through dbGAP. Study Accession: phs000044-A.v1.t1. Funding support for genotyping, which was performed at the Center for Inherited Disease Research (CIDR), was provided by 1X01HG005274-01 (CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200700006C). Assistance with genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies (GENEVA) Coordinating Center (U01 HG004444). Funding support for collection of samples and samples was provided by the Genetic Study of Nicotine Dependence (COGEN; P01 CA089192) and the University of Wisconsin Transdisciplinary Tobacco Use Research Center (PS0011906; PS0014726). (b) High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation, dbGAP Study Accession: phs000187.v1.p1: research support to collect data and develop an application to support this project was provided by 1PS001693905, 1PS001693907, SR0152011746, and 1S01CA133926. (c) Genetic Epidemiology of Refractive Error in the General Population (KORA) (Cooperative Gesundheitsforschung in der Region Augsburg) Study, dbGAP Study Accession: phs000303.v1.p1. Principal investigators: Dr. Dietrich Stolerman, University of Pennsylvania, Philadelphia, Pennsylvania, USA, H. Erich Wichmann, Institut für Humangenetik, Helmholtz- Zentrum München, Germany; National Eye Institute, National Institutes of Health, Bethesda, Maryland, USA. Funded by H01 EY02428, National Institutes of Health, Bethesda, Maryland, USA.

**Correspondence:** George Kirov, MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. Kimberly C. Dyer, Study Centre for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; Michael C. O’Donovan, FRCPsych, PhD, Michael J. O’Donovan, FRCPsych, PhD, George Kirov, MRC Psych, PhD, Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK.

**References**


22. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990; 47: 589–93.


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–75.


Online supplement

1. Samples and Genotyping Information

Genotyping of 7,129 schizophrenic patients was performed at the Stanley Centre for Psychiatric Research at the Broad Institute, Cambridge, Massachusetts on two different arrays: HumanOmniExpress-12v1 (Omni Express chip), and HumanOmniExpressExome-8v1 (Combo chip). The Combo chip contains, in addition to the SNPs from the Omni Express chip, those from the Illumina HumanExome-12v1_A (Exome chip). The 7,129 schizophrenic samples were genotyped in three batches: batch one consisted of 2,469 samples genotyped on the Omni Express chip, while batch two (3,621 samples) and batch 3 (1,039 samples) were genotyped on the Combo chip (Table DS1).

To obtain a large control dataset we searched the Database of Genotypes and Phenotypes (dbGaP), for publicly available raw intensity data on controls genotyped on Illumina arrays that have a similar coverage to the arrays used to genotype the cases, and that have not been used in
previous studies on schizophrenia that implicated these loci. The datasets obtained for this study are summarised in Table DS1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source (accession ID)</th>
<th>Array (N probes on the array)</th>
<th>N Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia Batch 1</td>
<td>Broad Institute</td>
<td>HumanOmniExpress-12v1 (730,525)</td>
<td>2,469</td>
</tr>
<tr>
<td>Schizophrenia Batch 2</td>
<td>Broad Institute</td>
<td>HumanOmniExpressExome-8v1 (951,117)</td>
<td>3,621</td>
</tr>
<tr>
<td>Schizophrenia Batch 3</td>
<td>Broad Institute</td>
<td>HumanOmniExpressExome-8v1 (951,117)</td>
<td>1,039</td>
</tr>
<tr>
<td>The Genetic Architecture of Smoking and Smoking Cessation</td>
<td>dbGaP (phs000404.v1.p1)</td>
<td>Illumina HumanOmni2.5 (2,443,179)</td>
<td>1,491</td>
</tr>
<tr>
<td>High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation</td>
<td>dbGaP (phs000187.v1.p1)</td>
<td>Illumina HumanOmni1_Quad_v1-0-B (1,051,295)</td>
<td>3,102</td>
</tr>
<tr>
<td>Genetic Epidemiology of Refractive Error in the KORA Study</td>
<td>dbGaP (phs000303.v1.p1)</td>
<td>Illumina HumanOmni2.5 (2,443,179)</td>
<td>1,869</td>
</tr>
</tbody>
</table>

**Table DS1.** Summary of case and control datasets prior to quality control filtering.

The following is a brief description of the control samples:

**“Smoking” study:** Participants were of European and African American ancestry (see below) who took part in a nicotine dependence study in the USA. Participants were excluded based on evidence of psychosis history, clinically significant depression symptoms, other severe mental illness, and contraindications to smoking cessation medications. Smokers comprise ~87% and non-smokers controls ~13% of the sample.

**“Melanoma” study:** consists of cases and controls of European ancestry recruited in the USA. Approximately 2/3 of the sample are people affected with malignant melanoma, the rest are controls without a history of cancers. The goal of this study was to identify novel susceptibility and outcome-related genes for melanoma.
“KORA” study: This study is called "Kooperative Gesundheitsforschung in der Region Augsburg" which translates as “Cooperative Health Research in the Region of Augsburg”. This is a population based study of adults randomly selected from 430,000 inhabitants living in Augsburg and 16 surrounding counties in Germany, with refractive error measurements. They are of European ancestry.

The ethnicities of the cases were provided by the psychiatrists and the ethnicities for the controls were provided as supplementary notes along with the raw intensity data, usually based on self-reports. We also derived ethnicities from principle components analysis (PCA) that combined the case/control genotypes with hapmap genotypes. PCA estimates took precedence in the few instances where the results from PCA and self/psychiatrist-report differed. We split the data by ancestry into those with a European (cases = 6,530, controls =5,894), African (cases = 263, controls = 478) or ‘other’ (cases = 336, controls = 90) origin. As the non-European individuals constitute very small numbers, we group the ethnic groups they comprise of into “other” ethnicities in Table DS2, where we list the numbers of individuals in the different samples.
**Figure S1a.** PCA plots for PCA1 and PCA2 for the samples in the data-sets, together with HapMap individuals. Different samples and reported ethnic origins from different samples are shown in different colours. At the top are individuals from China and Japan. On the bottom right are individuals of African origin. Most individuals, who are white Europeans, are at the bottom left-hand corner (with PCA1 and PCA2 values of ~0.00). Individuals next to them are of mixed or Indian origin, and they were classed as “others” in our analysis. The cut-offs used to assign the “other” ethnic origins are shown on Figure S1b, which shows a zoomed-in part of the above figure.
Figure S1b. Cut-offs used to define other ethnic groups from the European ones.

2. Sample and CNV Quality Control (QC)

Each case and control dataset was analysed independently to avoid batch effects. From these samples the raw intensity data was processed using Illumina Genome Studio software (v2011.1). Each dataset had their own egt cluster file defined on their own data in order to generate accurate Log R ratios and B-allele frequencies for subsequent CNV detection. PennCNV \(^1\) was used to call CNVs from the data following the standard protocol and adjusting for GC content. A total of four different Illumina array platforms were used for genotyping the case and control datasets, each with different SNP probe sets. Comparing CNVs called across arrays with different probe sets can be highly problematic as a particular locus may be adequately covered in one array and
not in another. To make the CNV calling comparable across the different arrays used, we only analysed the 520,766 probes present on all arrays.

Sample level QC was performed using the QC metrics generated by PennCNV. These include: Log R ratio (LRR) standard deviation, B-allele frequency drift, wave factor and total number of CNVs called per person. Samples were excluded if for any one of these metrics they constituted an outlier in their source dataset. The number of samples excluded from each dataset is shown in Table DS2. The proportion of excluded samples in the control datasets should not be taken as evidence about different quality of the datasets, as some of these had already been filtered for quality before being downloaded from dbGaP.

Samples were checked for duplicates with an Identify by Decent (IBD) analysis using PLINK \(^2\) and the duplicate with the better QC was retained.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total Excluded</th>
<th>Total Retained</th>
<th>Ethnicity (excluded/retained)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>European</td>
</tr>
<tr>
<td>SCZ</td>
<td>247</td>
<td>6882</td>
<td>223/6307</td>
</tr>
<tr>
<td>Smoking</td>
<td>3</td>
<td>1488</td>
<td>1/938</td>
</tr>
<tr>
<td>Melanoma</td>
<td>131</td>
<td>2971</td>
<td>131/2955</td>
</tr>
<tr>
<td>KORA</td>
<td>12</td>
<td>1857</td>
<td>12/1857</td>
</tr>
</tbody>
</table>

**Table DS2.** Number of patients and control samples, before and after QC, divided into ethnicities.

After the poorly performing or duplicated samples had been identified and excluded, all CNVs went through a rigorous QC filtering. Firstly, raw CNVs in the same sample were joined
together if the distance separating them was less than 50% of their combined length. CNVs were then excluded if they were either less than 10kb, covered by less than 10 probes, overlapped with low copy repeats by 50% of their length (using PLINK, v 1.07 ²) or had a probe density (CNV size/number of probes) of less than 20k, as defined by the probe density distribution for all CNVs. The remaining high quality CNVs from each dataset were then merged together and CNV loci with a frequency greater than 1% were filtered out using PLINK ². Here, a CNV contributes to a locus if it overlaps by 50% of the locus length.

The remaining rare CNVs were required to pass a median z-score outlier method of validation. This method is detailed in Kirov et al, 2012 ³. Briefly, the z-score method standardises SNP probe intensities for each individual across all SNP probes and then standardises the intensity of each SNP probe across all individuals. These rounds of standardisation help reduce noise created by natural fluctuations in probe intensity. A median z-score value for all standardised probe intensities within a putative CNV region is used to assess copy number, with true deletions and duplications represented as outliers in the samples median z-score distribution (Figure S2). The z-score histograms of CNVs with marginal z-scores (between -6 to -4 and +2 to +4) were manually inspected, and those that had ambiguous z-scores, were visually inspected using the Illumina GenomeStudio v2011.1 software. This resulted in 2,569 CNVs being filtered out from the data. All CNVs from the 15 loci discussed in this paper, that had been called by PennCNV, but didn’t have confirmatory z-scores (< -6 or > +4), were visually inspected for their LRR and B-allele frequencies.
Figure S2. Typical z-score histograms. (A) Example of a z-score histogram of a true positive CNV deletion. (B) Example of a z-score histogram of a false positive CNV deletion. (C) Example of a z-score histogram of an ambiguous CNV deletion. The numbers on the x-axis are normalised z-score values. The red arrows indicate the z-scores for the CNVs in question, called by PennCNV in these regions.
3. **CNVs in previously implicated loci in the different sub-groups of cases and controls.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1 del</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1q21.1 dup</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>NRXN1 del</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3q29 del</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WBS dup</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VIPR2 dup</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15q11.2 del</td>
<td>44</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>PWS dup</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15q13.3 del</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16p11.11 dup</td>
<td>22</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>16p11.2 distal del</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16p11.2 dup</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17p12 del</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17q12 del</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DiGeorge/VCFS del</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>166</td>
<td>5</td>
<td>14</td>
<td>10</td>
<td>7</td>
<td>27</td>
</tr>
</tbody>
</table>

**Table DS3.** Numbers of CNVs in previously implicated loci found in the different sub-groups of cases and controls. All African, Indian, East Asian, mixed or other ethnicities are grouped together as “others”. There are 16 non-European people in the Melanoma study sample, they are not shown separately, as none of them had a CNV from the list in the table.

4. **Details on CNVs in the 15 loci**

The following images (Figure S3) show the positions of the CNVs in the 15 loci discussed in this paper. The positions of the recurrent CNVs are usually determined by the segmental duplications flanking these loci. For the two individual genes in the table, *NRXN1* and *VIPR2*, the CNVs are non-recurrent, and have different breakpoints. For clarity, we also present their exact positions in Table DS4. All positions in the table and figures are in hg19 (build37).
Figure S3. Positions of CNVs in the regions discussed in this paper. “Common SNPs” refer to the probes common on all Illumina arrays used for calling CNVs.

Figure S3. A) Deletions at 1q21.1.
Figure S3 B) Duplications at 1q21.1

Figure S3 C) Exonic deletions in NRXN1. Intronic deletions are not shown.
Figure S3 D) Deletions at 3q29.

Figure S3 E) Duplications at the WBS region.
Figure S3 F). Exonic duplications at VIPR2.

Figure S3 G). Deletions at 15q11.2. One small deletion in a control, that only intersects one of the four genes in the region, is retained, in order to be conservative in our analysis.
Figure S3 H). Duplications at the AS/PWS region.

Figure S3 I). Deletions at 15q13.3.
Figure S3 J). Duplications at 16p13.11.

Figure S3 K). Deletions at distal 16p11.2.
Figure S3 L). Duplications at 16p11.2.

Figure S3 M). Deletions at 17p12 involving the PMP22 gene.
Figure S3 N). Deletions at 17q12.

Figure S3 O). Deletions at 22q11.2.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>Start Position</th>
<th>End Position</th>
<th>Type</th>
<th>Case/Control</th>
<th>Size</th>
<th>Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50342832</td>
<td>51258887</td>
<td>Deletion</td>
<td>Case</td>
<td>916055</td>
<td>198</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50733094</td>
<td>50930181</td>
<td>Deletion</td>
<td>Case</td>
<td>197087</td>
<td>48</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50763645</td>
<td>51002606</td>
<td>Deletion</td>
<td>Case</td>
<td>238961</td>
<td>60</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50793781</td>
<td>50924358</td>
<td>Deletion</td>
<td>Case</td>
<td>130577</td>
<td>39</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50837494</td>
<td>51120335</td>
<td>Deletion</td>
<td>Case</td>
<td>282841</td>
<td>67</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>50937036</td>
<td>51331157</td>
<td>Deletion</td>
<td>Case</td>
<td>394121</td>
<td>69</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>51013647</td>
<td>51294096</td>
<td>Deletion</td>
<td>Case</td>
<td>280449</td>
<td>46</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>51058745</td>
<td>51483527</td>
<td>Deletion</td>
<td>Case</td>
<td>424782</td>
<td>61</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>51058745</td>
<td>51171457</td>
<td>Deletion</td>
<td>Case</td>
<td>112712</td>
<td>21</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>51191432</td>
<td>51345515</td>
<td>Deletion</td>
<td>Case</td>
<td>154083</td>
<td>21</td>
</tr>
<tr>
<td>NRXN1</td>
<td>2</td>
<td>51227254</td>
<td>51304238</td>
<td>Deletion</td>
<td>Case</td>
<td>76984</td>
<td>10</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158532706</td>
<td>158914170</td>
<td>Duplication</td>
<td>Case</td>
<td>381464</td>
<td>68</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158433732</td>
<td>158917998</td>
<td>Duplication</td>
<td>Control</td>
<td>484266</td>
<td>85</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158500805</td>
<td>158948951</td>
<td>Duplication</td>
<td>Control</td>
<td>448146</td>
<td>90</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158649004</td>
<td>158948951</td>
<td>Duplication</td>
<td>Control</td>
<td>299947</td>
<td>58</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158526184</td>
<td>158948951</td>
<td>Duplication</td>
<td>Control</td>
<td>422767</td>
<td>66</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158649004</td>
<td>158948951</td>
<td>Duplication</td>
<td>Control</td>
<td>299947</td>
<td>46</td>
</tr>
<tr>
<td>VIPR2</td>
<td>7</td>
<td>158803848</td>
<td>158878791</td>
<td>Duplication</td>
<td>Control</td>
<td>74943</td>
<td>19</td>
</tr>
</tbody>
</table>

Table DS4. Exonic CNVs found in NRXN1 and VIPR2. CNV coordinates are in UCSC build37.
5. Resolving the parental origin of duplications at the AS/PWS critical region

In order to determine the parent of origin of duplications at the AS/PWS (15q11-q13) we took advantage of the differential methylation of maternal and paternal copies of the DMR (differential methylated region) within the SNRPN gene. Briefly, DNA samples were subjected to bisulfite conversion using the EZ methylation kit Gold (Cambridge Bioscience, Cambridge, UK). These bisulfite treated samples were then amplified (Sensimix HRM kit, Bioline, UK) and the PCR product subjected to high-resolution melting curve analysis as previously described \(^4,5\). High resolution melt (HRM) data were analysed using dedicated HRM software (Rotor-Gene Q series software) (Figure S4A and B). Normalisation regions for the leading/trailing ranges were set at 73-79/86-93 °C. The genotyping function in the software was used to assign control samples for five individuals with two copies. Using automated calling with an 85% confidence percentage threshold all samples were correctly assigned to one of the two groups. All eight unknown duplication samples were designated as maternal in origin.
Figure S4. High-resolution melting curves (A) and difference plots (B) for eight AS/PWS duplications and five normal control samples. Lines represent averages of three replicates for each sample. Only maternal duplications and normal controls are seen.
6. Discovery samples that carry more than one previously implicated SCZ CNV

Following the work by Girirajan et al, (2012)\(^6\), we examined how many individuals had more than one susceptibility CNV. Samples carrying more than one previously implicated CNV are shown in Table DS5. We find only four cases to have double hits, labelled “Case 1-4” in Table DS5. No double hits were found in controls. This excess of double hits in cases is likely to be due to chance as there are ~3 times more CNVs in cases. This study is underpowered to test individual loci for the presence of additional CNVs, as was done by Girirajan et al, 2012\(^6\) where about 100 CNVs were observed for some individual loci, close to our combined total. Although double hits should be expected in CNVs with lower pathogenicity, we observe a case with a deletion at 1q21.1 and duplication of the AS/PWS region, and one with a deletion at 15q13.3 and duplication at 16p11.2, all of which are among the most pathogenic ones.

<table>
<thead>
<tr>
<th>locus</th>
<th>N CNVs in 6882 Cases</th>
<th>Case Double Hits</th>
<th>Freq Cases</th>
<th>N CNVs in 6316 Controls</th>
<th>Control Double Hit</th>
<th>Freq Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1del</td>
<td>12</td>
<td>Case 1 Case 2</td>
<td>0.17</td>
<td>1</td>
<td></td>
<td>0.016</td>
</tr>
<tr>
<td>1q21.1dup</td>
<td>8</td>
<td></td>
<td>0.12</td>
<td>5</td>
<td></td>
<td>0.079</td>
</tr>
<tr>
<td>NRXN1 del</td>
<td>11</td>
<td></td>
<td>0.16</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>3q29 del</td>
<td>4</td>
<td></td>
<td>0.058</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>WBS dup</td>
<td>3</td>
<td></td>
<td>0.044</td>
<td>1</td>
<td></td>
<td>0.016</td>
</tr>
<tr>
<td>VIPR2 dup</td>
<td>1</td>
<td></td>
<td>0.015</td>
<td>6</td>
<td></td>
<td>0.095</td>
</tr>
<tr>
<td>15q11.2 del</td>
<td>44</td>
<td>Case 1</td>
<td>0.64</td>
<td>26</td>
<td></td>
<td>0.41</td>
</tr>
<tr>
<td>PWS/AS dup</td>
<td>8</td>
<td>Case 2</td>
<td>0.12</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>15q13.3 del</td>
<td>4</td>
<td>Case 3</td>
<td>0.058</td>
<td>2</td>
<td></td>
<td>0.032</td>
</tr>
<tr>
<td>16p13.11 dup</td>
<td>24</td>
<td>Case 4</td>
<td>0.35</td>
<td>12</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>16p11.2 distal del</td>
<td>0</td>
<td></td>
<td>0.00</td>
<td>2</td>
<td></td>
<td>0.032</td>
</tr>
<tr>
<td>16p11.2 dup</td>
<td>27</td>
<td>Case 3</td>
<td>0.39</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>17p12 del</td>
<td>4</td>
<td></td>
<td>0.058</td>
<td>3</td>
<td></td>
<td>0.047</td>
</tr>
<tr>
<td>17q12 del</td>
<td>1</td>
<td>Case 4</td>
<td>0.015</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>DiGeorge/VCFS del</td>
<td>20</td>
<td></td>
<td>0.29</td>
<td>0</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Totals</td>
<td>171</td>
<td></td>
<td>2.48</td>
<td>58</td>
<td></td>
<td>0.92</td>
</tr>
</tbody>
</table>
Table DS5. Samples with more than one CNV at previously implicated loci. Four cases, labelled “Case 1-4”, are found to harbour CNVs at two different pathogenic loci. No control sample had a double hit.
7. **Table DS6.** Comparison of previous results on CNVs with the combined results after adding the data from the current study. For the 22q11.2 deletion, ORs (odds ratios) cannot be estimated (NA, not available), as no deletion is observed in controls, therefore we only provide the lower/upper bounds of the OR. The results in the previous studies are based on the review by Malhotra & Sebat 7, four loci are updated.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Position</th>
<th>Previous results</th>
<th>Combined results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Case</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P Value</td>
<td>Reference</td>
</tr>
<tr>
<td>1q21.1 del</td>
<td>chr1:146,57-147,39</td>
<td>0.17% (21/12,174)</td>
<td>0.021% (16/75,505)</td>
</tr>
<tr>
<td>1q21.1 dup</td>
<td>chr1:146,57-147,39</td>
<td>0.14% (13/9,365)</td>
<td>0.033% (19/57,730)</td>
</tr>
<tr>
<td>NRXN1 del</td>
<td>chr2:50,15-51,26</td>
<td>0.19% (22/11,880)</td>
<td>0.022% (10/44,845)</td>
</tr>
<tr>
<td>3q29 del</td>
<td>chr3:195,73-197,34</td>
<td>0.099% (10/10,123)</td>
<td>0.0016% (1/63,649)</td>
</tr>
<tr>
<td>WBS dup</td>
<td>chr7:72,74-74,14</td>
<td>0.076% (11/14,387)</td>
<td>0.0036% (1/28,139)</td>
</tr>
<tr>
<td>VIPR2 dup</td>
<td>chr7:158,83-158,94</td>
<td>0.19% (14/7,336)</td>
<td>0.059% (11/18,499)</td>
</tr>
<tr>
<td>15q11.2 del</td>
<td>chr15:22,80-23,09</td>
<td>0.57% (72/12,665)</td>
<td>0.27% (201/75,486)</td>
</tr>
<tr>
<td>Angelman /</td>
<td>chr15:24,82-28,43</td>
<td>0.053% (4/7,582)</td>
<td>0.0073% (3/41,370)</td>
</tr>
<tr>
<td>Prader-Willi dup</td>
<td>chr15:31,13-32,48</td>
<td>0.19% (22/11,689)</td>
<td>0.018% (13/74,106)</td>
</tr>
<tr>
<td>15q13.3 del</td>
<td>chr15:15,51-16,30</td>
<td>0.25% (13/5,147)</td>
<td>0.13% (81/62,973)</td>
</tr>
<tr>
<td>16p13.11 del</td>
<td>chr16:15,51-16,30</td>
<td>0.094% (13/13,850)</td>
<td>0.015% (3/19,954)</td>
</tr>
<tr>
<td>16p11.2 distal del</td>
<td>chr16:28,82-29,05</td>
<td>0.31% (31/9,890)</td>
<td>0.033% (19/56,752)</td>
</tr>
<tr>
<td>16p11.2 dup</td>
<td>chr16:29,64-30,20</td>
<td>0.14% (8/5,891)</td>
<td>0.024% (14/59,086)</td>
</tr>
<tr>
<td>17p12 del</td>
<td>chr17:14,16-15,43</td>
<td>0.056% (4/7,142)</td>
<td>0.0059% (4/68,131)</td>
</tr>
<tr>
<td>17q12 del</td>
<td>chr17:34,81-36,20</td>
<td>0.30% (36/12,202)</td>
<td>0.000% (0/70,739)</td>
</tr>
</tbody>
</table>
8. Supplementary References

Analysis of copy number variations at 15 schizophrenia-associated loci
Elliott Rees, James T. R. Walters, Lyudmila Georgieva, Anthony R. Isles, Kimberly D. Chambert, Alexander L. Richards, Gerwyn Mahoney-Davies, Sophie E. Legge, Jennifer L. Moran, Steven A. McCarroll, Michael C. O'Donovan, Michael J. Owen and George Kirov
Access the most recent version at DOI: 10.1192/bjp.bp.113.131052

Supplementary material can be found at: http://bjp.rcpsych.org/content/suppl/2013/11/25/bjp.bp.113.131052.DC1
This article cites 48 articles, 11 of which you can access for free at: http://bjp.rcpsych.org/content/204/2/108#BIBL
To obtain reprints or permission to reproduce material from this paper, please write to permissions@rcpsych.ac.uk
You can respond to this article at /letters/submit/bjprcpsych;204/2/108
Published by The Royal College of Psychiatrists

To subscribe to The British Journal of Psychiatry go to: http://bjp.rcpsych.org/site/subscriptions/