Supplementary material

1) Samples **p 2**

2) Genotyping and CNV calling in the Bulgarian samples **p 5**

3) Filtering putative *de novo* CNVs at higher stringency using a median z-score outlier method. **p 5**

4) Validation of CNVs with Agilent arrays and overall evaluation of QC for *de novo* and non-*de novo* calls. **p 7**

5) Parental origin of the *de novo* CNVs **p 9**

6) Detailed results for each *de novo* **p 10**

7) Case control analysis **p 27**

8) Size of the *de novo* CNVs **p 28**

9) Family history **p 29**

10) Gene set curation **p 30**

11) Comparison of gene sets hit by *de novo* CNVs in cases with those hit by CNVs in Bulgarian controls or hit by *de novo* CNVs in Icelandic controls. **p 32**

12) Testing for enrichment in number of *de novo* CNVs hitting gene sets by random placement **p 34**

13) Partitioning the PSD CNV signal **p 36**

14) Genes in the NMDAR and ARC gene sets (Table S9) **p 39**

15) Overlap between the loci affected by de novos in schziophrenia probands from the current study with other datasets **p41**

16) De novo CNVs identified in the Icelandic control sample **p44**

17) Supplementary references **p 45**

1) Samples:

***De novo* study**: Families from Bulgaria were recruited if an offspring had schizophrenia or schizoaffective disorder, and both parents were available and agreed to give blood samples. Recruitment took place between 1999 and 2004 in several psychiatric hospitals in Bulgaria. Ethical Committee approval was obtained from each of these hospitals. All probands and all parents received an Information Sheet and signed Informed Consent Forms. All participants had attended mainstream schools, which at the time in Bulgaria, excluded people with mental retardation. Probands were either in- or out-patients at the time of the study but each had a history of hospitalisation. A team of psychiatrists was trained in using the rating scales and methods of the study. (The names of psychiatrists who recruited patients are included in the Acknowledgements). We used the SCAN instrument to perform an interview for psychotic and mood symptoms[1]. This instrument has been translated into Bulgarian and validated by one of its authors (A. Jablensky). Consensus diagnoses were made according to DSMIV criteria on the basis of an interview and inspection of hospital notes by two clinicians (GK and IN). If consensus was not attained, the patient was re-interviewed by IN or GK, and was excluded if consensus could still not be reached. In addition, approximately 23% of the sample was selected at random and re-interviewed by GK or IN. Hospital notes were also collected for affected relatives in order to confirm their diagnoses. The final sample consisted of 698 families, with 720 offspring, of which 107 had schizoaffective disorder and the rest had schizophrenia.

30ml of peripheral venous blood was collected and DNA was extracted within 48 hours (without freezing), using a phenol-chlorophorm method.

Following genotyping with Affymetrix 6.0 arrays and rigorous quality control (QC) (see below) we had good quality CNV data on 662 offspring, in 638 families (23 families with 2 offspring and one with 3 offspring).

We also obtained Affymetrix 6.0 genotyping data from 605 unrelated controls with no history of severe psychiatric disorder recruited by us in Bulgaria[2]. For the present study, CNVs were called using the same QC criteria and calling algorithm as used in the trios and the CNVs served as controls in the primary pathway/gene set analyses reported here.

Icelandic control *de novos*: In total 5,928 samples, constituting 2,623 offspring with both parents available, genotyped using Illumina (San Diego, CA, USA) bead arrays (HumanHap317 (59.2%), HumanHap370 (32.9%) and HumanHap 1M (7.9%)) were analysed. Subsets of these subjects are from the population sample screened for *de novos* in a previous publication[3]. In the present analysis, we have excluded incomplete trios (parent-offspring duos) and probands known to be affected with neurodevelopmental/psychiatric disorders (schizophrenia, autism, Attention Deficit Hyperactivity Disorder (ADHD), mental retardation and bipolar affective disorder). BeadStudio (version 2.0) was used to call genotypes, normalise the signal intensity data, and establish the log R ratio and B allele frequency at every SNP according to the standard Illumina protocols. Samples passing standard SNP-based quality control procedure and with a SNP call rate higher than 0.97 were used in the analysis. PennCNV (10.1101/gr.6861907), a free, open-source tool, was used for copy number variation detection. The input data for PennCNV are log R ratio (LRR), a normalized measure of the total signal intensity for the two alleles of the SNP and B allele frequency (BAF), a normalized measure of the allelic intensity ratio of the two alleles. These

values are derived with the help of control genotype clusters (HapMap samples), using the Illumina BeadStudio software. PennCNV employs a hidden Markov model (HMM) to analyse the LRR and BAF values across the genome. CNV calls are made based on the probability of a given copy state at the current marker as well as on the probability of observing a copy state change from the previous marker to the current one. PennCNV uses a built-in correction model for GC content (doi:10.1093/nar/gkn556). A subset of markers, those present on all genotyping chips listed above, were used for calling the CNVs in the Icelandic sample (the "317 content"). CNVs spanning ten consecutive markers were included in the analysis. All *de novo* events were visually inspected using DosageMiner (a software developed by DeCODE) and all trios were tested for inheritance errors genome wide.

As with the Bulgarian sample, we only included autosomal CNVs, those with frequency ≤1% in the sample, and those not spanned by ≥50% by LCRs. This resulted in the identification of 59 *de novo* CNVs, an autosomal *de novo* rate of 2.2%, less than half the rate (5.1%) we observed in cases (see main text). The set of Icelandic *de novo* CNVs had a similar size distribution (see Section 8) to those we observed in schizophrenia cases, suggesting that with respect to subsequent gene enrichment analyses, the sample of control CNVs are a suitable comparator group for the case *de novos*.

In order to explore whether the lower resolution of the Illumina 317k array resulted in a lower rate of *de novos* in controls relative to cases, we undertook multiple sensitivity analyses. 1) We excluded those Bulgarian *de novo* CNVs that were covered with <58 probes on the Affymetrix 6.0 array (to allow for a 5.8 fold increase in probe density in that array; Affymetrix 6.0 contains ~1,850,000 million probes, while Illumina 317k contains ~317,000 probes, or 5.8 times less). This still resulted in a two-fold higher rate of *de novo* CNVs (4.2%) rate in schizophrenia probands, p=0.005. 2) We restricted analysis to those *de novos* ≥100kb size in both datasets, (and covered by ≥58 probes in the Bulgarian sample). The rate remained two-fold higher: 3.6% vs 1.8%, p=0.004. 3) In the size range ≥500kb, the difference was 2.8 fold higher: (2.1% in cases, vs. 0.76% in controls, p=0.004). Thus, the enrichment for *de novos* was greater for larger CNVs than for smaller CNVs, an outcome that is incompatible with the hypothesis that enrichment is due to greater sensitivity of *de novo* detection in cases. Under that scenario, enrichment in cases would be biased towards small CNVs. However, this is clearly not what is observed (see also Figure 1 of the main text). Details on the size distribution of the two *de novo* sets are presented in Section 8. Further information is presented in **Table S2**, size of *de novos*. 4) We checked how many of our schizophrenia *de novos* should have been missed by the Illumina HumanHap 300 array, as they were covered by <10 probes of that array (see figures in Supplementary section 6). Six would have been missed, which still results in ~2-fold higher rate in cases than in the Icelandic controls (4.2% vs. 2.2%, p=0.005). Of these 6, only 1 CNV is greater than 50kb (chr4:79944612-80081979; 137kb). Since our sensitivity analyses reveal a significant excess of *de novos* across the size range, and more specifically, in the size range well above the largest poorly covered CNV, and since the case CNVs also show an excess compared with the autism control *de novos* (which have good coverage across all our de novos), the relatively low coverage is unlikely to be an important confounder . We also note that a) none of these CNVs hit members of the NMDAR/ARC complexes and their removal does not influence the results of our pathway analyses and b) pathway analyses of the *de novos* restricted to very large CNVs also implicate ARC and NMDAR complexes (page 18, last paragraph).

Further Analyses: In order to gain additional information on the pathogenic relevance of the *de novo* CNVs we observed in Bulgarian cases, we examined further available samples. We chose to analyse large studies that had published their genome-wide data on CNV tracks or provided the raw intensity data. For multigenic loci, we considered CNVs that spanned the *de novo* locus by 50% or more[4], while if it hit a single gene, we considered CNVs that intersected any exon of that gene (i.e. we did not count intronic CNVs in the additional samples, but list their numbers in Section 6). The significance of the excess occurrence of such CNVs in the combined samples was tested with Fisher Exact test.

Bulgarian trios:
In the 662 Bulgarian trios we examined the regions containing *de novos* for other CNVs that were transmitted or non-transmitted from carrier parents, and counted this sample as consisting of 662 cases and 662 controls.

International Schizophrenia Consortium (ISC):
We used the publicly available data published as part of the ISC study[2]. The study included 3391 cases and 3181 controls genotyped with Affymetrix 6.0 or 5.0 arrays. 328 Bulgarian cases from the ISC study are probands in the current study, although their parents were not genotyped as part of that study. For analysis of CNV loci in additional datasets, those 328 Bulgarian samples were excluded, as their CNVs are counted in the *Bulgarian trios* sample. However, just for the record, in Section 6 we point out if any *de novo* CNV was also observed in a proband in the ISC**.** We restricted all analyses to CNVs ≥100kb (this was the cut-off used in this study[2, 5]).

*UK/WTCCC**:***
The samples and analysis methods are described in Kirov et al[5]. Briefly, we analysed 471 schizophrenia cases using the Affymetrix GeneChip 500K Mapping Array as part of the same pipeline of the WTCCC study[6], and 2792 controls that were used in that study. We had used a conservative method to call CNVs, namely they had to be called independently on both *Sty* and *Nsp* arrays that make up the 500K arrays, with each being ≥100kb and covered with ≥10 probes. No additional QC measures were applied in the current analysis. These CNVs are available with the Supplementary material of that study[5].

Molecular Genetics of Schizophrenia (MGS):
We used the publicly available raw intensity data published as part of the MGS study[7] that consisted of cases and controls from both European American and African American ancestry. All samples from this study were genotyped on Affymetrix 6.0 arrays at the BROAD Institute. We performed similar in house CNV filtering and Z-Score validation to that described for the Bulgarian Trios (MeZOD, see below). Following QC, 2215 cases and 2556 controls of European American ancestry and 977 cases and 881 controls of African American ancestry remained for analyses. We note that the number of samples available to us at the time of the download is smaller than the ones presented in the Levinson et al[7] paper.

Japanese study:
We used the publicly available CNV data from a Japanese population published as the Ikeda et al[8] study. This study includes 519 cases and 513 controls. CNVs from this data are restricted to those >= 100kb, as published.

These five sample sets amount to 7,907 independent cases and 10,585 controls (after excluding the 328 overlapping Bulgarian probands).

2) Genotyping and CNV calling in the Bulgarian Samples:

Initial CNV calls: All offspring and parents were genotyped with Affymetrix 6.0 arrays at the Broad Institute, USA. Analysis was performed using Genotyping Console 4.0 software, one batch of arrays at a time. Each batch contained 70-90 arrays. We used the default parameters of this software with the exception that we set the minimum CNV size to ≥ 10kb and a minimum coverage of ≥10 probes. Individuals with >50 CNVs of that size and type were excluded as these reflect poor quality CNV genotyping (e.g.Kirov et al[5]) and were clear outliers from the distribution. PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) was then used to exclude CNV loci with a frequency >1% in the sample, using the "--cnv-overlap 0.67" command (i.e. two thirds overlap) to define overlapping regions.

Preliminary de novo analysis:

For our first pass screening for *de novos*, we targeted autosomal CNVs that were ≥15 kb and were covered by ≥15 probes. We then excluded 67 potential *de novo* CNVs because more than 50% of their length overlapped segmental duplications (Low Copy Repeats, LCRs) based upon the 'Segmental Dups' track in UCSC (http://genome.ucsc.edu/cgi-bin/hgTables) as such CNVs are prone to false calls. We also excluded the families of 8 probands who had ≥10 apparent *de novo* CNVs each, as this is indicative of technical artefacts in these samples. Calls compatible with a *de novo* CNV were made if the CNV in the child was spanned ≤50% of its length by a CNV in either parent. A script for calling such overlaps is available upon request.

3) Filtering putative *de novo* CNVs at higher stringency using a median z-score outlier method.

The signal for each probe was measured by the Log2 ratios that were derived from PennCNV[9]. We then used a slight modification of the algorithm (MeZOD) reported by McCarthy et al[10] to evaluate putative *de novo* CNVs. This is a three-stage process: (i) the signal from each probe on an individual array is assigned a z-score based upon the distribution of all probe signals on that array (individual-wise standardization) (ii) each resulting z-score for each probe from (i) is assigned a new z-score based upon the distribution of all individual z-scores for that probe (probe-wise standardization) (iii) the median of the z-scores for all probes within a region of interest from (ii) is calculated and displayed as a histogram. Outlier detection is performed by visual inspection of the histogram.

Software to perform this analysis and visualize the results can be obtained from http://x004.psycm.uwcm.ac.uk/~dobril/z_scores_cnvs

The results of median z-score outlier analysis of 271 potential *de novo* CNVs meeting our initial relaxed criteria were as follows.

1. Patterns highly suggestive of a true *de novo* event (Figure 1a in the main text) were observed for 40 putative *de novos.*
2. Ambiguous patterns were observed for 33 putative *de novos*. (Figure 1b in the main text).
3. A CNV that had not been called by the Genotyping Console in a parent was detected in a parent at the same locus for 117 of the events. Those were rejected (Figure S1a).
4. Two or more common copy number state 'bins' were present for 36 of the putative events. As this indicates the presence of a common CNV, these were also rejected. (Figure S1b).
5. No outlier signal indicative of a CNV was found in 45 of the putative events. These were also rejected. (Figure S1c).

Post hoc, of 122 potential *de novo* events that had already been rejected because >50% of the CNV included low copy repeat (LCRs) sequence, 117 (96%) were rejected using the z-score method, supporting our decision to exclude such CNVs at the first pass stage.

Our decision to analyse only CNVs ≥15 kb and defined by ≥15 probes was also supported *post hoc* by the z-score method. Out of 90 potential *de novo* CNVs called by the Genotyping Console in the size range 10-15kb and covered with 10-15 probes, only five were accepted at a relaxed criterion as a possible *de novo*, three of them being ambiguous.

Figure S1. Examples of median z-score histogram analysis. (note further examples in Figure 1 in the main text)

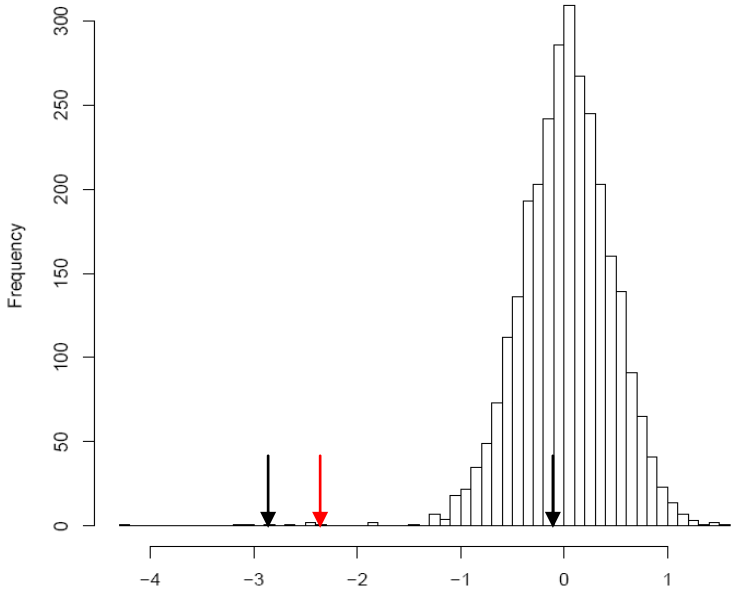

Figure S1a. Example of a true CNV deletion falsely called as a *de novo*. The CNV is correctly called by the Genotyping Console in the affected child (red arrow). The CNV is also clearly present in a parent (black arrow), but had not been called by the Genotyping Console.

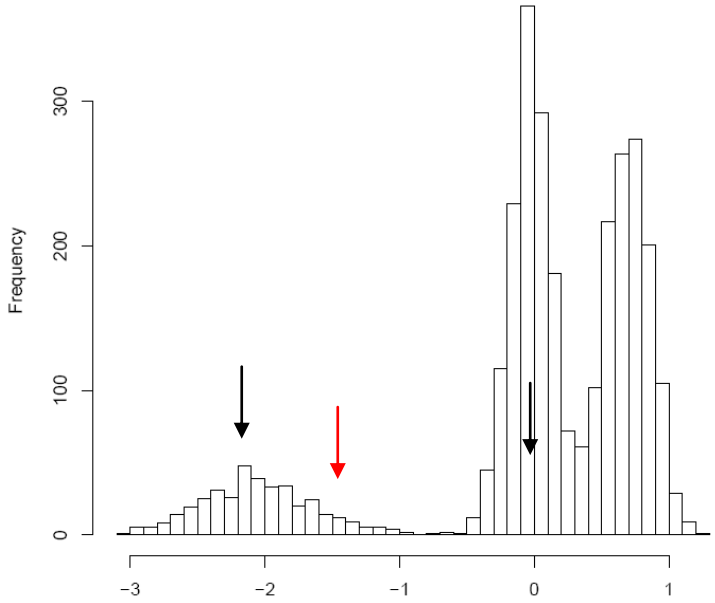

Figure S1b. **A region with several common copy number states.** The *de novo* call in this particular occasion is false, as the deletion has probably come from the parent, but CNVs in all such regions were excluded. The CNV was not called in either parent by the Genotyping Console.

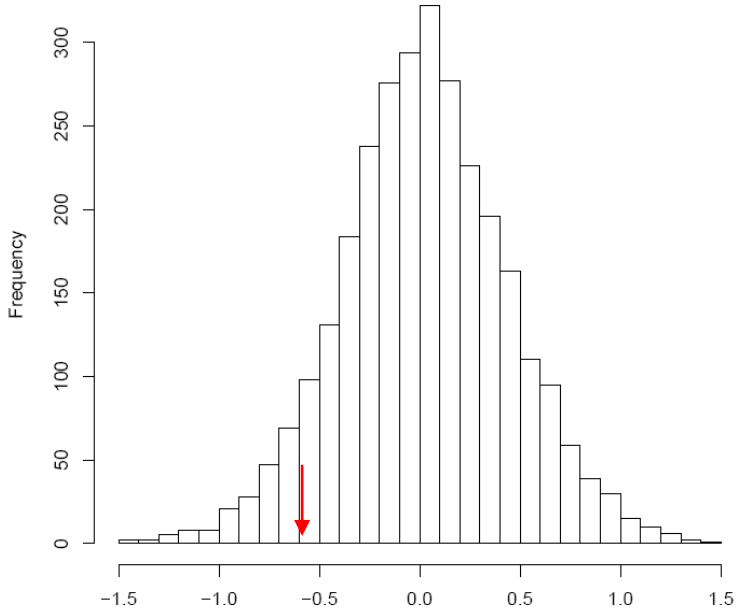

Figure S1c. Example of a false positive *de novo* deletion call (in red). There are no outliers in this region.

4) Validation of CNVs with Agilent arrays and overall evaluation of QC for *de novo* and non-*de novo* calls.

In order to maximise capture of *de novo* CNVs, we intentionally used relaxed criteria to define CNVs for follow up based upon the z-score method. Thus, we identified 73 potential

de novo CNVs, of which 40 were classified as highly suggestive and 33 as ambiguous. These were taken forward for analysis using custom Agilent SurePrint G3 Human CGH Microarrays, 8x60K. We placed 26,477 probes on 361 regions. In addition to the 73 putative *de novo* CNVs, these included all other rare CNVs (sample frequency <1%) in the proband that had been transmitted from parents. These were included to facilitate sample tracking since each composite set of rare CNVs identifies a unique trio. Between 50 and 200 probes were used to cover each CNV (depending on CNV size) including 20kb-200kb of the putative flanking regions (proportional to CNV size). The rest of the array was populated with the standard probes from the 8x60 arrays. In each experiment, the parents of a proband were hybridised against each other (CNVs at all potential *de novo* loci are rare, and it is therefore extremely unlikely that both parents would carry the same CNV). Probands were hybridised against another proband who did not share a CNV at the same locus. An example of the Agilent array output is provided below (Figure S2).

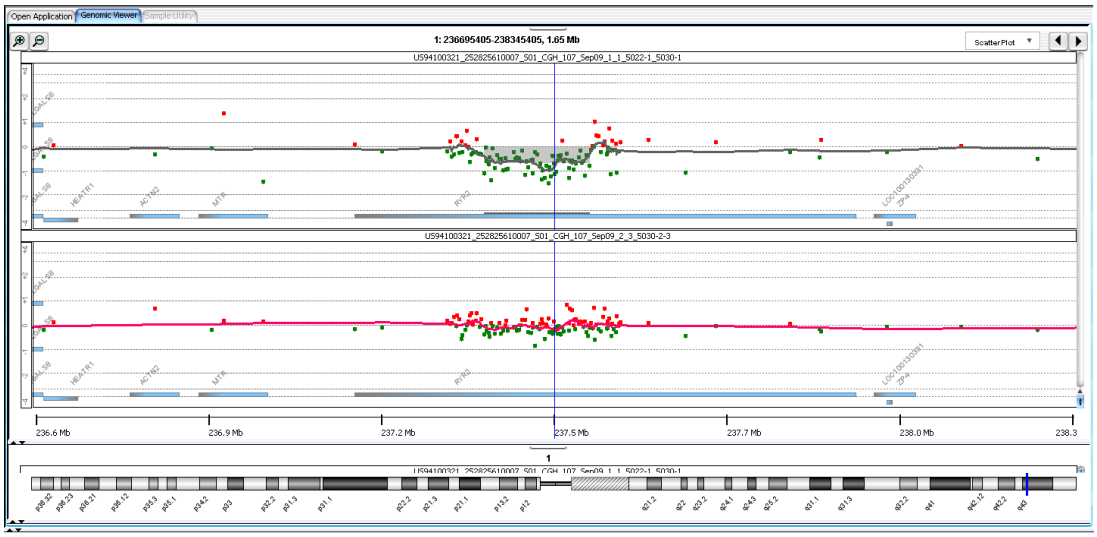

Figure S2. Agilent validation of a *de novo* deletion on chr1:235475280-235639644. The coordinates in the title of the figure are given according to hg18 to allow the reader to cross reference with Table 1 of the main text. However, the software used by Agilent is annotated according to hg19 (NCBI build 37), so the coordinates differ from those in Table 1. In hg19 the corresponding region is chr1:237408657-237573021 as shown above in the figure. Top: proband hybridised against another proband (showing a deletion), bottom: parents hybridised against each other (no evidence for deletion). The region of suspected *de novo* CNV, including a flanking sequence, is densely saturated with custom-designed probes, while the remaining sequence is covered at reduced density with the standard probes from the 8x60K array.

Of the 40 suggestive and 33 ambiguous calls, 31 and 3 respectively were confirmed on the Agilent array as *de novos*.

Evaluation of the false-positive rate of (non *de novo*) CNV detection in controls and in cases.

We can estimate the likely false positive error rate arising from our criteria for CNV calling in the Bulgarian *controls* (very few of which are expected to be *de novo*) by examining the full set of CNVs in *cases and their parents*. In cases, after z-score filtering using the highly suggestive (the strictest) threshold as applied to *de novos* that we subsequently applied to the controls, we identified 1688 CNVs (inherited and *de novo*). Excluding the 40 potential de novos (called with the z-score method as highly suggestive) the remaining 1648 were independently validated as they were clearly present in a parent. Thus, 1648 CNV calls can

be considered true positives. However, of the 40 putative *de novos* at the highly suggestive threshold, 31 CNVs were confirmed *de novos*, and were therefore not false CNV calls. This leaves 9 that had been erroneously called, giving a false positive call rate of 0.5% (9/1688).

5) Parental origin of the *de novo* CNVs

For 21 *de novos*, we were able to determine the likely parent of origin of the mutation, and in a smaller number of cases, the mechanism of the *de novo* formation. This was done by analysis of the informative SNPs, as implemented in PennCNV (May 2010 version), with modifications that we implemented for the analysis of duplications (to allow for the different patterns in intra- and inter-chromosome Non- Allelic Homologous Recombination, NAHR). PennCNV also produces a significance score for whether the CNV has arisen on the paternal or maternal genome based on the number of informative SNPs. Of the 21 *de novos* in which we could determine the likely parent of origin, we had evidence from at least one SNP that 14 had occurred in the paternal line and 7 in the maternal line but this was not statistically significant (p=0.13) (Table 1 in the main text). Itsara et al[11] observed a slight maternal excess in the origin of *de novo* CNVs in probands with autism or asthma (26:21). For 18 of the 21 cases in our study the information from SNPs reached nominal levels of significance for parental origin. (Although the information on the remaining three does not reach statistical significance, there were no SNPs suggesting occurrence on the genome of the other parent (Table 1 in the main text), so we consider it most likely that the parental origin of these events is also identified correctly, as there was only a single occurrence of a SNP contradicting the trends in the remaining CNVs). The trend for a peternal excess was stronger for CNVs that were not flanked by LCRs, 8:2 for paternal vs. maternal origin. While this difference was not significant either (p=0.06), it suggests that other mutational mechanisms resulting in *de novo* CNVs could be more likely to occur on the paternal genome. A similar strong parent-of origin bias for the formation of CNVs not mediated by NAHR was observed by Hehir-Kwa et al (2011)[12] in a larger study of *de novo* CNVs in mental retardation.

In four cases, we were also able to determine whether the duplication had arisen because of inter- or intra-chromosome NAHR. The following informative SNP calls in a family give an example of a duplication that has occurred by maternal intra-chromosomal NAHR:

Father	Mother	Offspring
AA	BB	ABB (mutation of maternal genome, intra- or inter-chromosomal)
AA	AB	ABB (mutation of maternal genome, DNA from one maternal chromosome duplicated, intra-chromosomal)

Previous research on the mechanisms of CNV *de novo* formation[13] showed that inter-chromosomal and intra-chromatidal NAHR are much more common than inter-chromatidal ones. This results in a more than 2-fold greater rate of deletions than duplications at such loci, because intra-chomatidal NAHR generates a deletion and a circular DNA molecule, which lacks a centromere and cannot segregate at cell division, i.e. does not result in a duplication that is transmitted. Where we could determine the mechanism, we found two cases of intra-chromosomal NAHR (in the Williams-Beuren syndrome reciprocal duplication and the 16p11.2 duplication) and two for inter-chromosomal NAHR (duplications at 15q11.2, and 15q13.1).

6) Detailed results for each *de novo*

All coordinates in the paper refer to the UCSC human genome assembly NCBI build 36 (hg18). The ISC data are available in NCBI build 35 (hg17). The CNV positions were converted into hg18 for the figures. Phenotypic details of CNV carriers are provided where the CNV is of particular interest.

chr1:144101459-144503409.
The *de novo* deletion in our proband covers the region of the **Thrombocytopenia Absent Radius (TAR Syndrome)** (Figure S3). TAR syndrome (MIM ID #274000) is caused by a 200-kb minimally deleted region at 1q21.1 that encompasses at least 12 known genes including *HFE2*, *TXNIP*, *POLR3GL*, *ANKRD34A*, *LIX1L*, *RBM8A*, *GNRHR2*, *PEX11B*, *ITGA10*, *ANKRD35*, *PIAS3*, and *NUDT1*[14]. Individuals with TAR syndrome almost always have bilateral absence of the radius, while the thumbs are always present. Other features include malformations of the skeleton (limbs, ribs, and vertebrae), heart, and genitourinary system. Thrombocytopenic episodes decrease with age, most children with TAR Syndrome having normal platelet counts by school age (reviewed in Toriello[15]). Cognitive development is usually normal, and features of psychosis or autism have not yet been documented as part of the syndrome. The prevalence of the TAR syndrome is estimated at 0.5:100,000-1:100,000[15]. Haploinsufficiency of the deleted region is not sufficient to cause TAR syndrome, leading to the suggestion that at least two unlinked alleles, rare deletion and another common variant, are required for the syndrome to occur[14].

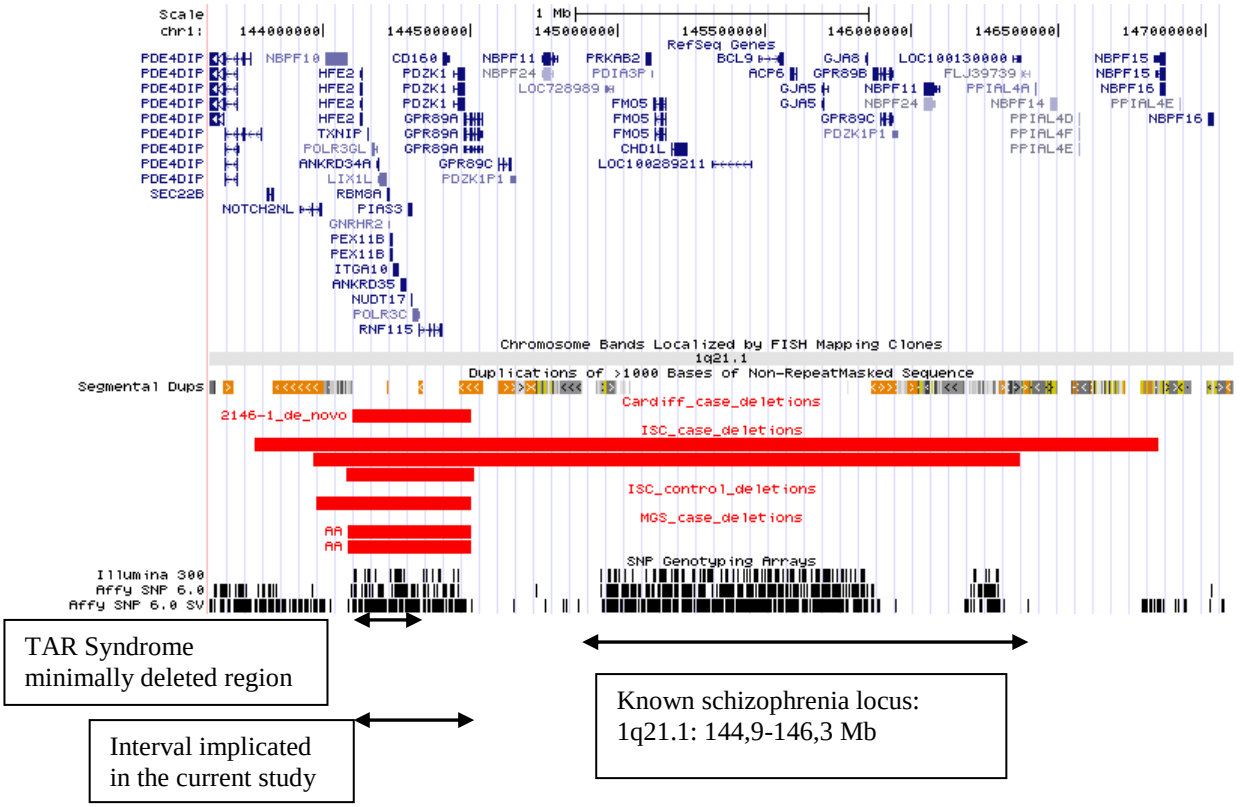

Figure S3. Deletions at chr1:144101459-144503409. In this and all subsequent figures, subjects in the MGS sample are indicated with AA (African American) or EA (European American). All figures include the positions of probes on the Affymetrix 6.0 array

(separately for SNPs: Affy SNP 6.0; and copy number probes: Affy SNP 6.0 SV), and those on the Illumina HumanHap300 array.

The minimally deleted region in patients with the TAR syndrome overlaps the region deleted in our proband. In the MGS study[7], we find 2 African American case deletions in the TAR region. In the ISC[2] data, CNVs in one case and one control cover the interval, while in two more cases, the CNVs are much larger and correspond to the so called "larger form of the 1q21.1 deletion"[3] also known as class II deletions[16]. The larger types of deletions were reported in four individuals with schizophrenia in the study of Stefansson et al[3]. We find no additional deletions in the Kirov et al[5] or Ikeda et al[8] studies. In summary, for the full dataset we have 4 cases and 1 control CNV deletions in the TAR region (excluding the 2 larger class II deletions and including the *de novo*), 1-sided Fishers Exact p = 0.11. Involvement in schizophrenia for deletions in the TAR region has not been proposed before and our result is not significant, even without correction for multiple testing. However a recent study strengthens this finding, as it implicated this locus in developmental delay, with deletions found in 47 cases and 2 controls, in a sample of 15,767 children affected with developmental delay and various congenital anomalies, and 8,329 controls, $p = 3.3 \times 10^{-7}$.[17]

The proband with a *de novo* is a 49-year old university graduate, who held a responsible job and fathered two children. No physical anomalies are documented in his notes or interview record. He suffers with a schizoaffective disorder (bipolar type) with a late age at onset of 40 years. Family history is negative. 10 informative SNPs suggest that the deletion had occurred on the paternal genome, and none suggest a deletion on the maternal genome (10:0, p=0.002).

chr1:235475280-235639644. We found a single *de novo* 164kb deletion, which disrupts several exons of the **RYR2** gene (Figure S4, *RYR2* in the trios sample). *RYR2* encodes a ryanodine receptor found in cardiac muscle sarcoplasmic reticulum. Mutations in this gene are associated with stress-induced polymorphic ventricular tachycardia, sudden cardiac death and arrhythmogenic right ventricular dysplasia (MIM ID *180902). RYR2 is part of the PSD complex (see Table 2 and Figure 2 in the main text). No additional deletions disrupting the *RYR2* gene were found in the remaining studies. The proband with the *de novo* was included in the ISC[2] study.

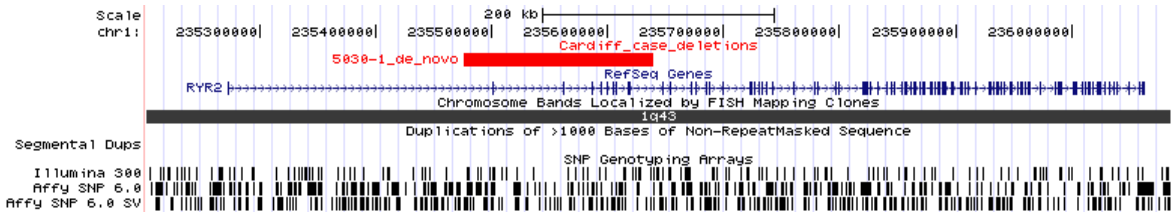

Figure S4. *RYR2* in the trios sample.

CNVs in *NAP5*, chr2:133504420-133879778. One *de novo* deletion of 375kb was identified in a proband. Three transmitted CNVs (T), two non-transmitted CNVs (NT), and the *de novo* CNV disrupted exons (Figure S5). The ISC[2] study identified the *de novo* in the same individual from the current study. Three more controls in that study had deletions in this

gene, all intersecting exons. No deletions disrupting this gene were found in the Kirov et al[5] or Ikeda et al[8] studies. In the MGS study[7], we find 2 case deletions that intersect exons. In addition, one more transmission in a trio was intronic, as were 3 MGS controls (not shown in the figure). NAP5 is one of the genes implicated in a recent genome-wide study on schizophrenia and bipolar disorder [18]. Overall we do not find support for this gene despite it being hit by a *de novo*.

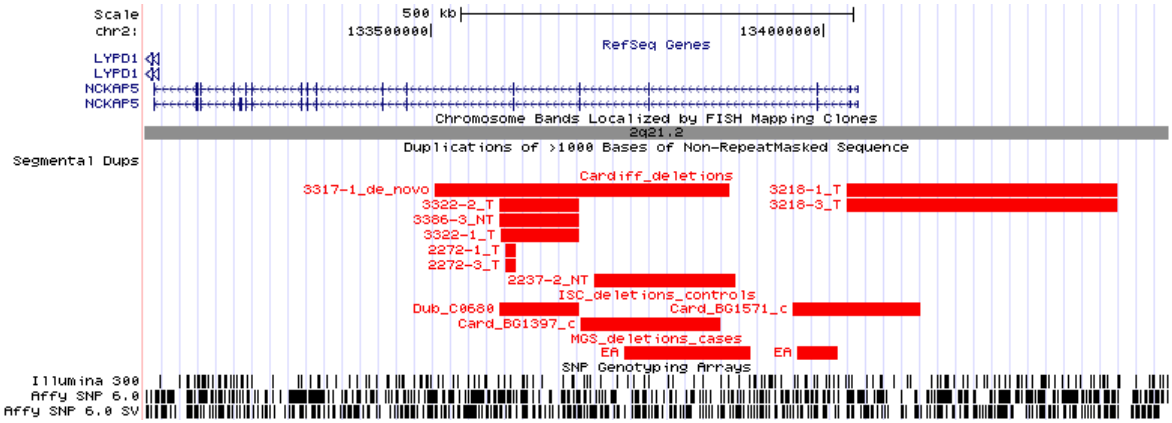

Figure S5. NAP5 in the trios, ISC and MGS data. In this and all other figures in the Supplementary material NNNN-1 is proband, NNNN-2 is father, NNNN-3 is mother and NNNN-4 is affected sibling. T stands for transmitted and NT for non-transmitted.

chr3:109330592-110198715. A single *de novo* duplication of 868kb was found. There are 9 known genes intersected by the CNV: *DZIP3, GUCA1C, HHLA2, IFT57, KIAA1524, MORC1, MYH15, RETNLB, TRAT1*. In the ISC [2] study no case, but one control carried a duplication spanning >50% of this region (Figure S6). No such duplication was found in the remaining studies. The duplication occurred on the paternal genome (11:0 informative SNPs, p=0.001).

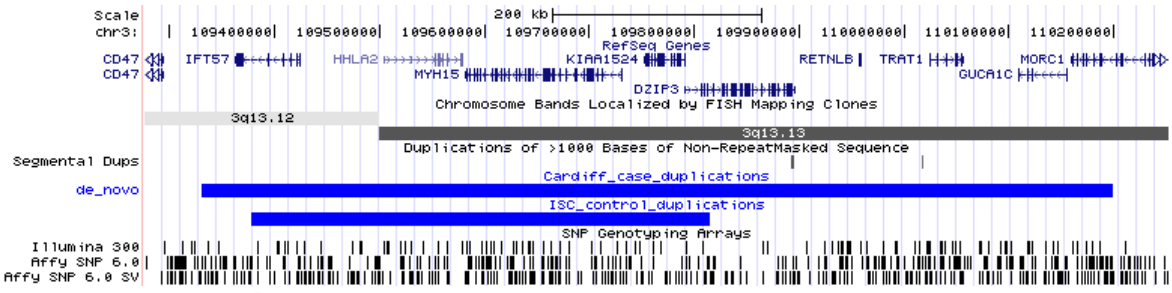

Figure S6. CNVs at chr3:109330592-110198715.

3q29: **deletion at chr3:197185548-198825231** (Figure S7). One case had a *de novo* deletion in this region (the same case deletion is reported in the ISC[2] study). While this paper was being prepared, another team identified a *de novo* deletion in this interval in a patient with

schizophrenia[19]. Another such case is reported by Walsh et al[20], among 233 patients with schizophrenia, and one by Magri et al[21], among 172 such patients. Using an extended set of cases and controls, including the ISC[2] data which includes our proband, Mulle et al[19], found a significantly increased rate of this deletion in cases: overall there were six 3q29 deletions in a total of 7,545 cases, compared to one deletion in 39,748 controls. The combined meta-analysis performed by that team refined the odds ratio estimate to 16.98 (p = 0.0097, 95% CI: 1.36–1198.4). This locus was confirmed as a schizophrenia-associated locus by Levinson et al [22] who found another five deletions in cases and none in controls, among 3,945 cases and 3,611 controls.

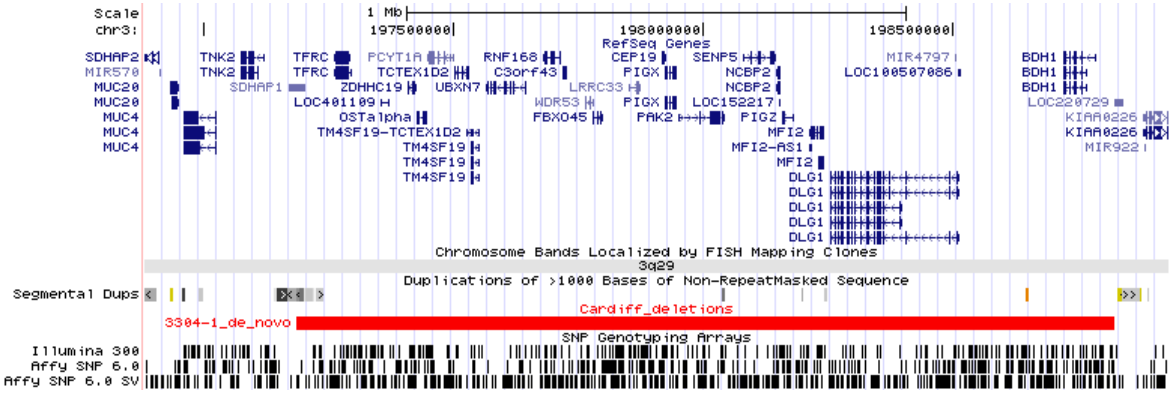

Figure S7. Deletion at 3q29. No results from the other studies are shown, as this is a known locus for schizophrenia.

The region is involved in the known genomic disorder called **3q29 microdeletion syndrome**[23, 24]. It affects 21 known genes, among which a good candidate for CNS disorders is ***DLG1***, which is part the PSD-95, ARC and NMDAR complexes (see main text). *DLG1* is a homolog of *DLG3*, mutation of which causes non-syndromic X-linked mental retardation[25]. Another homolog, *DLGAP2*, was implicated in the study on autism by Pinto et al[26]. We also observed *de novos* in the related genes *DLG2* and *DLGAP1* that are also part of the PSD-95, ARC and NMDAR complexes as discussed in the main text and below.

The 3q29 microdeletion syndrome presents with a varying phenotype that includes mental retardation, microcephaly, language delay, autistic features and mild facial dysmorphisms. The deletion is present in up to 1:1,000 people with idiopathic mental retardation[23]. Our proband is a 45-year old single female, who had below-average school marks, but attended a mainstream school. Age at onset was 19, with quite typical schizophrenia presentation of persecutory delusions, auditory hallucinations, and a relapsing course of illness. There is no family history of psychiatric illness. SNP analysis revealed that the deletion had occurred on the paternal genome (56:0 informative SNPs, $p=2.78\times10^{-17}$).

chr4:70935504-70969553. We detected a single 34 kb *de novo* deletion. It affects two genes: *HTN1* and *HTN3*. (*CSN1S2A* shown in the Figure S8 is a pseudogene). No CNVs hit this gene in the MGS study[7] and as the CNV and the genes are <100kb, we did not examine this further in the remaining studies.

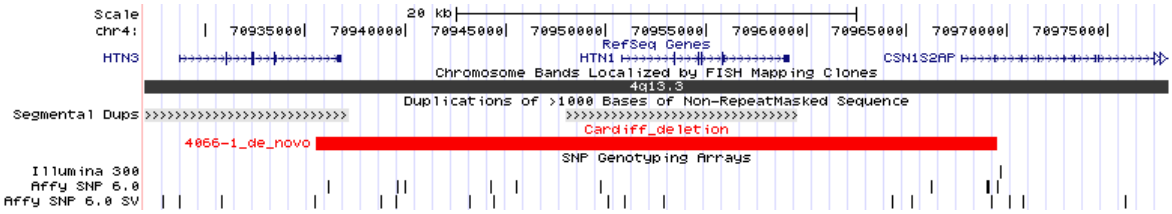

Figure S8. chr4:70935504-70969553.

chr4:79944612-80081979. We detected a single *de novo* 137 kb deletion in the trios sample (Figure S9). Two genes are within the interval: *BMP2K* and *PAQR3*. There is one deletion in a case in the ISC[2] study that intersects the same two genes, and none in the remaining studies.

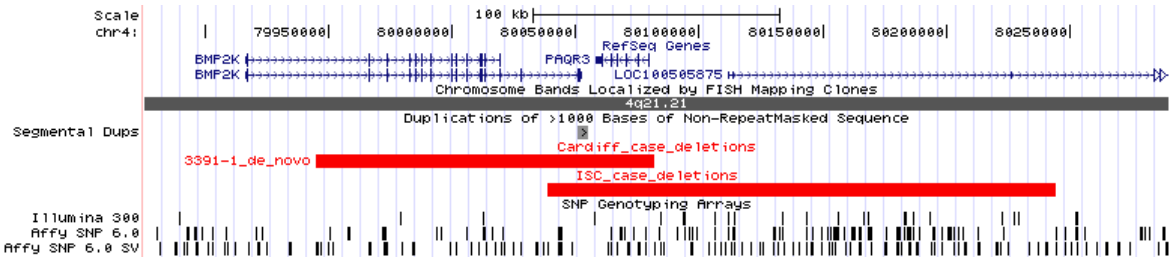

Figure S9. chr4:79944612-80081979.

chr6:68675955-68761101. We detected an 85 kb duplication that does not span any known gene and was not investigated in the additional datasets.

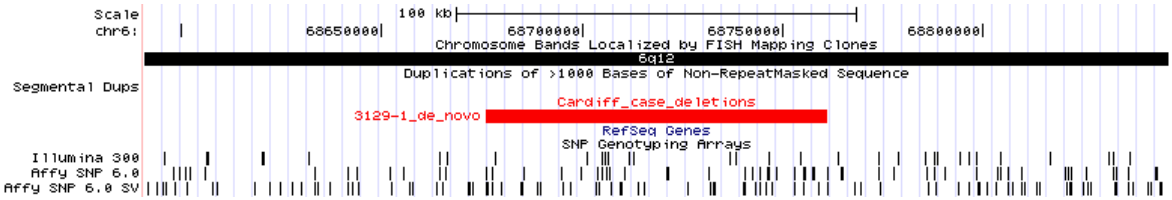

Figure S10 chr6:68675955-68761101.

chr7:38260614-38307187. We detected a single 74 kb *de novo* deletion spanning the full length of *TARP*. (Figure S11). In the MGS study[7] we find 3 case deletions and 6 control deletions intersecting *TARP* exons, thus giving no support for the involvement of this gene in the pathogenesis of schizophrenia. However, as the deletion and the gene are <100kb, we did not examine this locus in the remaining datasets.

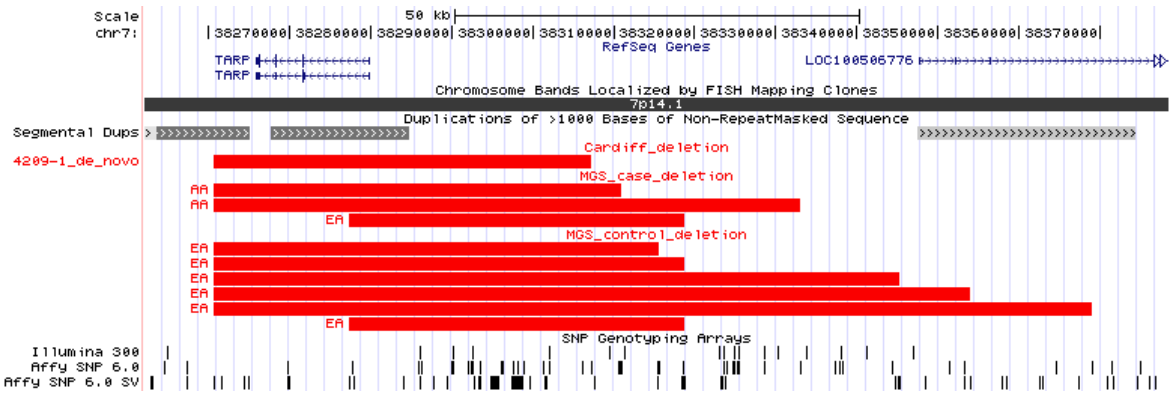

Figure S11. chr7:38260614-38307187, TARP gene.

chr7:72390286-76445231. We found a *de novo* duplication of 4Mb that spans the full 1.55Mb region that is deleted in **Williams Beuren Syndrome** (WBS, OMIM #194050). (Figure S12).

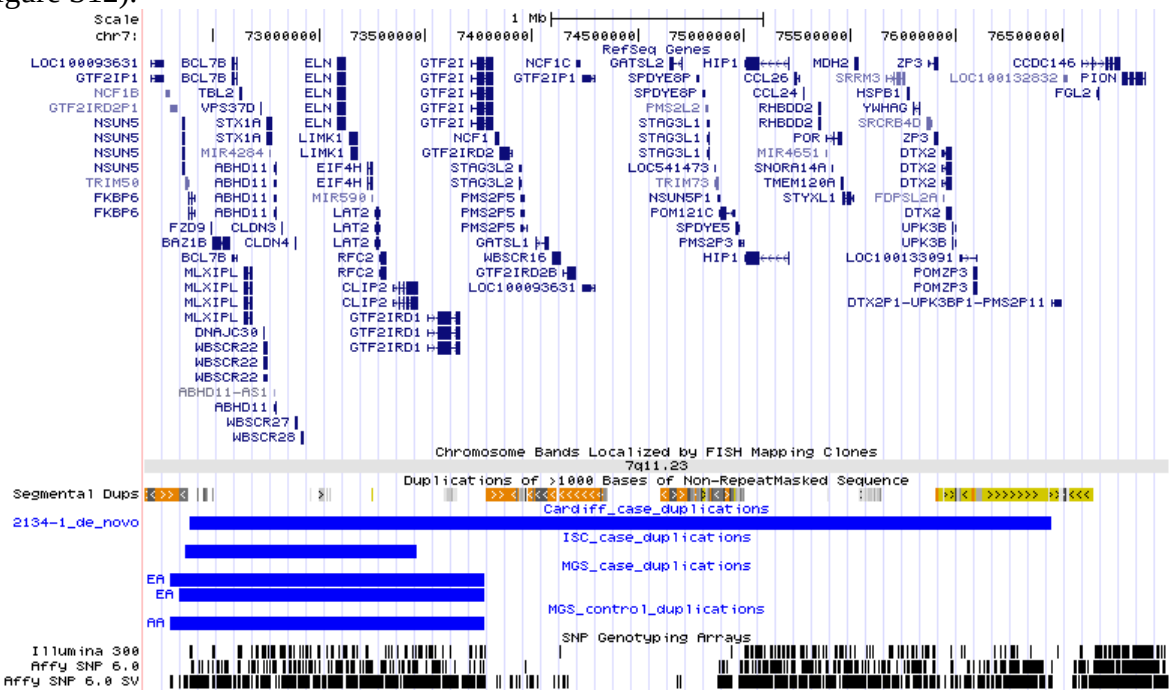

Figure S12. Duplications in the Williams Beuren Syndrome region. The smaller duplications in the ISC[2] and the MGS studies[7] cover the WBS region flanked by segmental duplications (bottom of the figure). The four duplications in ASD cases reported in the Sanders et al[27] study cover the same smaller interval. Our *de novo* duplication extends over a larger interval.

WBS is a well-characterised microdeletion syndrome caused by the reciprocal deletion at 7q11.23[28]. However the duplication of this region is also associated with a recognised phenotypic syndrome[29]. The estimated population frequency of the duplication is 1:13,000-1:20,000, and carriers show variable speech delay, normal cognitive ability to moderate mental retardation, facial dysmorphisms, autism, ADHD and problems with social

interaction. All carriers described by Van der Aa et al[29] were pre-pubertal, so it is not known whether the duplication is associated with schizophrenia or psychosis in adulthood. Our proband with a *de novo* duplication had an early age at onset of psychotic symptoms at 17 years. He suffers with a disorganised type of schizophrenia, with thought disorder, thought withdrawal and delusions, but no prominent hallucinations. The hospital discharge summary describes speech delay, but he finished a mainstream school with good marks. He had been treated by child psychiatrists for ADHD. SNP analysis revealed the duplication had occurred on the maternal genome, 45:1 informative SNPs, ($p=7.91x10^{-16}$). The most likely mechanism of the mutation is intrachromosomal non-allelic homologous recombination (NAHR) mediated by the flanking LCRs (36:0 informative SNPs suggested that one chromosome was duplicated).

Recently Sanders et al[27] implicated duplications in the WBS interval as risk factors for autism. Given the other neurodevelopmental features of 7q11.23 duplication syndrome and their role in ASD, we consider it very likely that this CNV is also of relevance to the presence of psychosis in schizophrenia. In the current datasets we find a total of 4 case and 1 control duplications (including the *de novo*) among 7907 cases and 10585 controls, (1-sided Fisher Exact test p = 0.11, without correction for multiple testing). Clearly, larger studies will be required to provide statistical support for this locus, due to its rarity.

One of the genes within the WBS region, ***STX1A***, is part of the NMDAR complex (see main text and Table S9). One more gene (***YWHAG***) within the larger *de novo* duplication found in this study is also part of this complex. Distal deletions at 7q11.23, including *HIP1* and *YWHAG* were recently identified in patients with intellectual disabilities, epilepsy and neurobehavioural problems[30]

chr7:127275795-127447967. We detected a 172 kb *de novo* deletion that affects two genes: *C7orf54* and *SND1* (Figure S13). No other CNV in the tested samples intersects the SND1 gene.

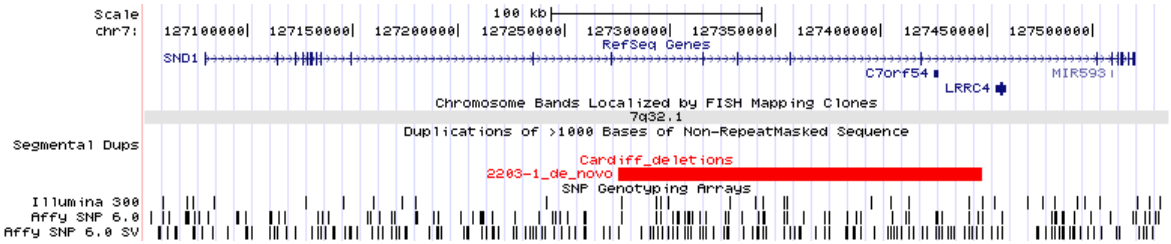

Figure S13. chr7:127275795-127447967.

chr8:4121968-4299810. We detected a 178kb *de novo* deletion intersecting an exon of the CSMD1 gene (Figure S14). Two non-transmissions were also found to disrupt exons in this gene. In addition, we find one case deletion in the ISC[2] study (in addition to our proband), one case and two controls in the MGS study[7] and none in the remaining studies that intersect *CSMD1* exons. However, we also found five transmissions, 6 non-transmissions, two case deletions from the ISC study[2] and 20 cases and 23 controls from the MGS study[7] that hit *CSMD1* introns (intronic CNVs not shown in figure). Common variation in this gene has

been implicated at genome-wide significance in a large GWAS study of schizophrenia including approximately 20,000 cases and 40,000 controls[31]. The evidence for any role of CNVs in this gene in schizophrenia is unresolved.

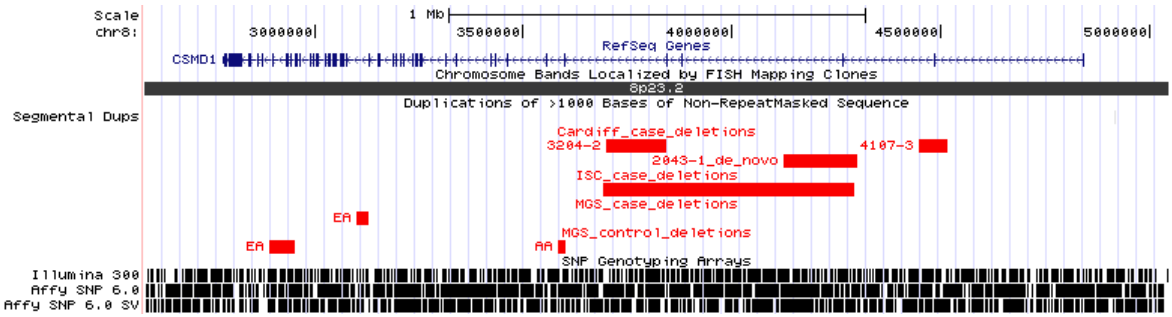

Figure S14. Deletions in *CSMD1*.

chr8:10066862-10155414. We detected an 89kb *de novo* deletion disrupting exons of the MSRA gene (Figure S15). The CNV is <100kb, however the gene is larger, so we examined all other datasets for CNVs intersecting the *MSRA* gene. One larger deletion, which intersected several more genes, was found in a case in the ISC[2] study. Two smaller deletions in MGS[7] cases also intersect exons, while one is intronic (Figure S15). No CNVs were found to intersect MSRA in the remaining studies. Methionine sulfoxide reductase A knockout (MsrA-/-) mice have been found to maintain a larger dopamine reserve pool than wild-type control mice, and that this pool is readily mobilized[32]. A widely accepted theory postulates an imbalance between cortical and subcortical dopamine systems in the brains of patients affected with schizophrenia[33]. As we find 4 deletions in cases (including our *de novo*) and none in controls that intersect this gene among the 7907 cases and 10585 controls, our study provides support for its involvement in schizophrenia, 1-sided Fisher Exact test p = 0.03, but we note that this p-value is not corrected for multiple testing.

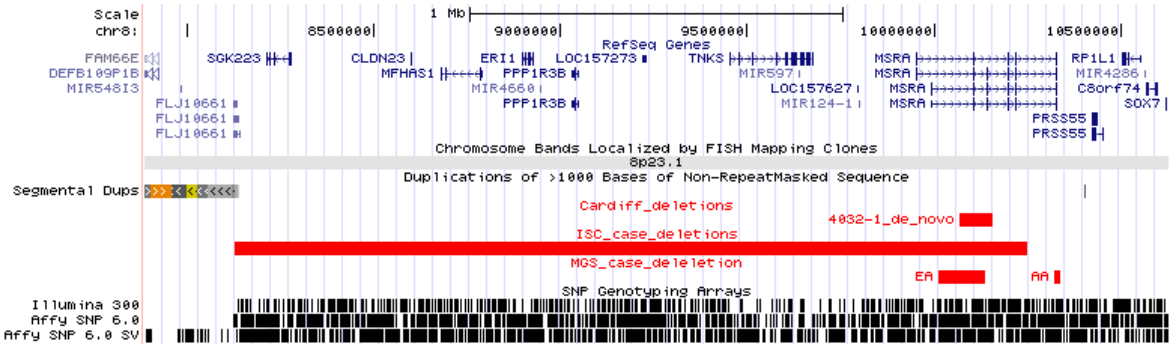

Figure S15. *MSRA* deletions.

chr9:16310745-16327782. This 17kb deletion does not hit any genes and was not examined further (Figure S16).

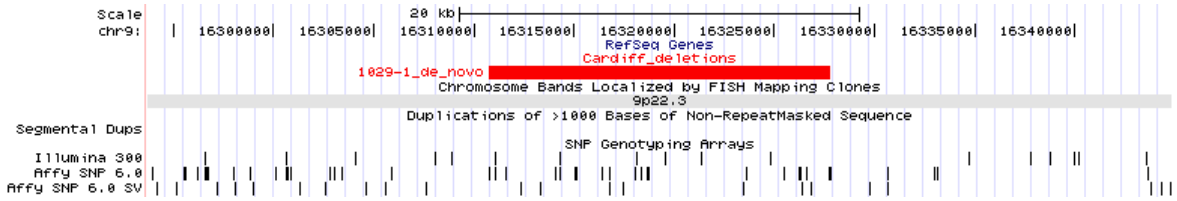

Figure S16 chr9:16310745-16327782

chr9:110859131-111433199. We observed a 574kb *de novo* duplication that was not found in any other cases or controls (Figure S17). It spans over 4 genes: *C9orf4*, *C9orf5*, *EPB41L4B*, *PTPN3*.

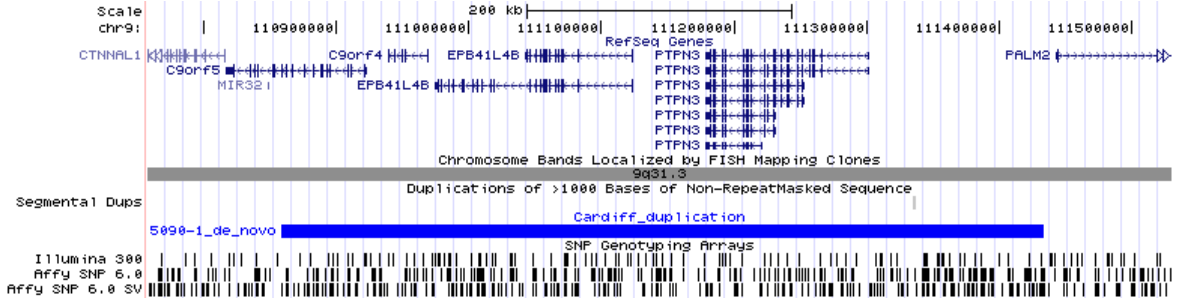

Figure S17. chr9:110859131-111433199.

EHMT1: Two *de novo* CNVs affected this gene: duplication at chr9:139762152-139797423 and deletion at chr9:139769564-139792102, both affecting exons (Figure S18). This is a very good candidate gene since it is involved in mental retardation and the **chromosome 9q subtelomere deletion syndrome (9qSTDS)**[34]. The syndrome is caused by haplo-insufficiency of *EHMT1*, a gene whose protein product (Eu-HMTase1) is a histone H3 Lys 9 (H3-K9) methyltransferase. The role of this gene in the syndrome was established by the identification of three patients with features of the syndrome and either mutations or a balanced translocation in *EHMT1*. Point mutations in *EHMT1* cause a similar phenotype to those with submicroscopic deletions[34]. Affected individuals have severe hypotonia with speech and gross motor delay, micro- or brachycephaly, other facial dysmorphisms, obesity and heart defects. A significant minority have epilepsy and/or behavioural and sleep disturbances[34]. A recent study found 60 deletions disrupting *EHMT1* among 15,767 children affected with developmental delay and various congenital anomalies, and none among 8,329 controls ($p=8.5 \times 10^{-12}$)[17].
The *de novo* CNVs are 22kb and 35 kb however, the gene itself spans over 200kb of genomic sequence, so we examined for CNVs in the other studies. Three more cases (two duplications and one deletion, all in the ISC[2] study) and one control (duplication, MGS study[7]) carried CNVs that intersected exons in this gene. One more case deletion in the MGS study[7] was intronic (not shown in the figure). No CNVs were found in the remaining studies. The presence of five CNVs in cases and only one in a control, together with the known involvement of this gene in cognitive phenotypes, makes it a strong candidate for a schizophrenia locus. This trend is not significant in the extended sample of 7,907 cases and 10,585 controls: uncorrected 1-sided Fisher Exact test p = 0.055.

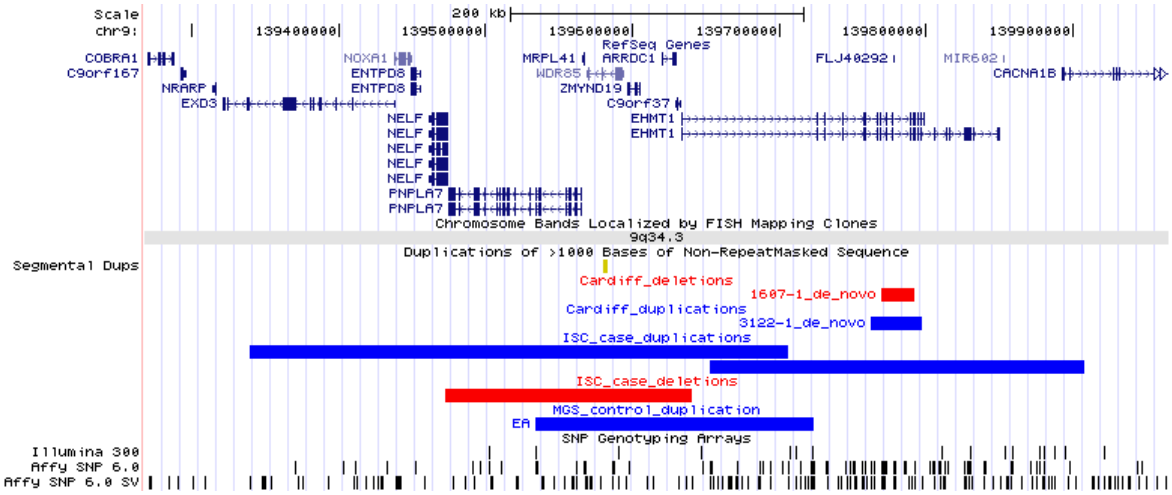

Figure S18. **CNVs affecting *EHMT1***.

Descriptions of adults with 9qSTDS are scarce, but one study[35] reported that five out of six patients who had reached adulthood had developed severe psychiatric symptoms. All five had been under psychiatric care and prescribed psychotropic medication. Abnormalities included apathy, aggressive episodes, psychosis or (autistic) catatonia, bipolar mood disorder, and regression in daily function and cognitive abilities. Our cases share some of these features: The case with a duplication was a 37 year old female, with an age at onset of 20, and a chronic course of schizophrenia, with delusions of reference, bizarre delusions, and a prominent apathy. Her father and grandfather suffered with depression. Our proband with a deletion is a 36 year old female with poor school results and hebephrenic type of schizophrenia. Her illness started with gradual onset during puberty. She lives a vagrant life-style, does not sleep at night, does not wash for months, has thought disorder, occasional auditory hallucinations, and elated and inappropriate behaviour. She has a monozygotic twin who is also schizophrenic (age at onset 24). The proband also has a child with mental retardation, and her maternal grandmother is also schizophrenic. No DNA was available to examine if the CNV was present in the monozygotic twin or her own child, and no dysmorphic features were recorded in the discharge summaries of the proband or her twin. Interestingly, both our cases with *de novo* CNVs were reported as being overweight, but exact BMI data is not available (these were the only two cases among the 33 probands with *de novos* that were reported as overweight). Increased body weight is reported in 50% of cases with 9q Subtelomeric Deletion Syndrome[35].

DLG2. We found two *de novo* deletions involving *DLG2*, one of which disrupted an exon (Figure S19). In addition, in the trios there are two transmissions (one disrupting exons) and three non-transmissions (one disrupting an exon). Both individuals with *de novo* CNVs in the current study were included in the ISC study[2]. In addition, one case and one control have CNVs that intersect exons in the ISC study[2], while one more case has an intronic CNV. In the MGS study[7] there were four deletions in cases and none in controls that intersect exons. There were however another 9 case and 6 control intronic deletions in that study (the figure shows only the exonic deletions, and the intronic *de novo* deletion). We also find one exonic case CNV and two intronic control CNVs in the Ikeda et al[8] study. In summary, we find 8 exon-disrupting CNVs in 7907 cases (or transmitted CNV) and two in 10585 controls (including the one non-transmitted CNV), giving an uncorrected 1-sided Fisher Exact test

p=0.02. **DLG2** is a member of the PSD-95, ARC and NMDAR complexes (see main text and discussions on other DLG genes).

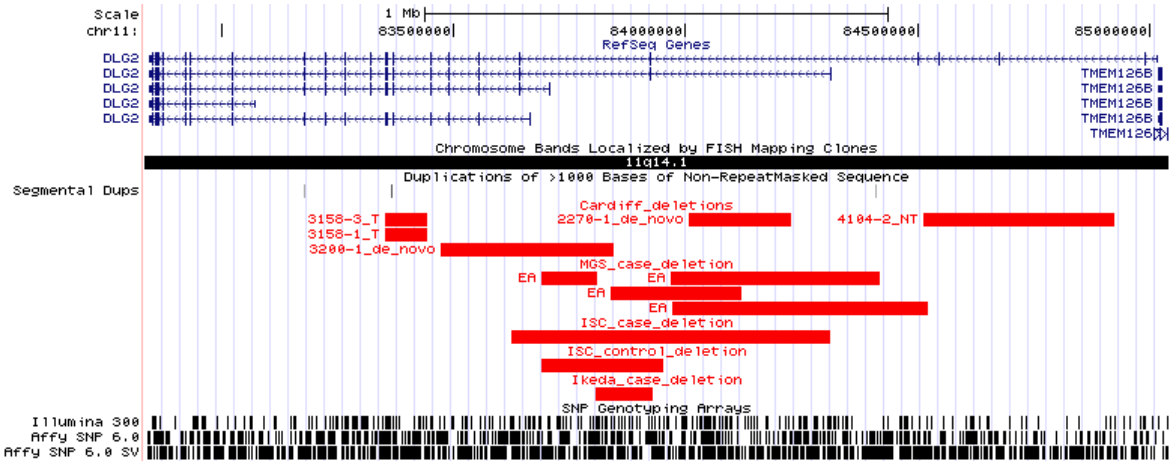

Figure S19. ***DLG2*** **CNVs**. Nomenclature is as described in Figure S5. Note that one of the *de novos* is intronic.

Phenotypic summaries of cases with *DLG2* CNVs

3158-1, transmitted from mother, exon-disrupting, negative family history. Male, very poor school marks, age at onset 31. Persecutory and grandiose delusions, auditory and olfactory hallucinations, disorganised behaviour.

3200-1 *de novo*, exon-disrupting, negative family history. Female with good school results, persecutory delusions, thought disorder, ticks, choreiform movements, stereotypies, apathy.

3223-1 transmitted from mother, not exon-disrupting, negative family history. Female, below average school marks, Auditory and somatic hallucinations, age at onset 24, persecutory delusions.

2270-1 *de novo*, but not disrupting exons. Negative family history. Female, poor school results, onset after childbirth aged 20, auditory hallucinations, persecutory delusions, thought disorder, odd behaviour.

4104-2 exon-disrupting, father carrier, not ill, not transmitted.

1601-3 not disrupting an exon, mother carrier, not ill, not transmitted.

chr12:111723795-111776045. This deletion disrupts 3 exons of the ***RPH3A*** gene (Figure S20). No CNVs disrupting this gene were found in the other datasets. Interestingly, RPH3A has been found to interact with neurexins[36]. Deletions at the neurexin 1 gene (*NRXN1*) are an established risk factor for schizophrenia[22, 37]

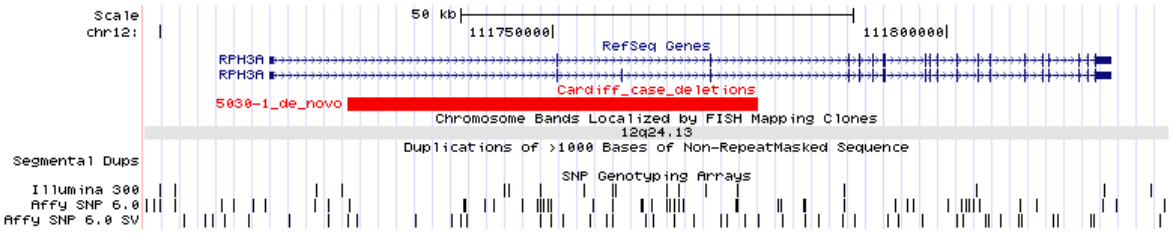

Figure S20. **chr12:111723795-111776045,** ***RPH3A***

chr12:130388037-130659530. The 271kb *de novo* does not intersect any gene (figure S21), so it was not examined further.

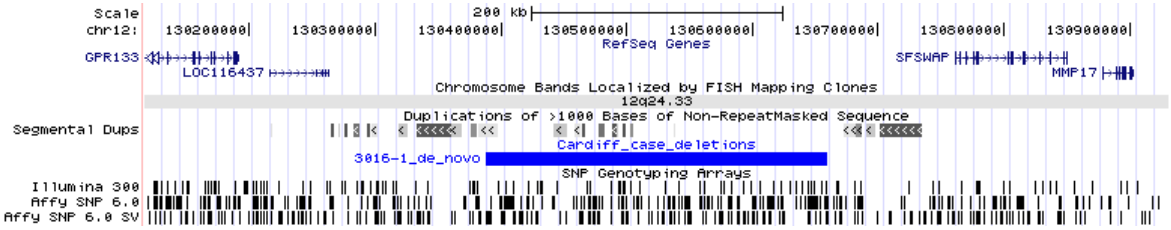

Figure S21 chr12:130388037-130659530

chr13:40319620-41182276. This 862kb deletion affects 8 known genes: *C13orf15, ELF1, KBTBD6, KBTBD7, KIAA0564, MTRF1, NARG1L, WBP4*. The same person is reported in the ISC[2] dataset. No additional deletions spanning >50% of this region were found in the other datasets (Figure S22).

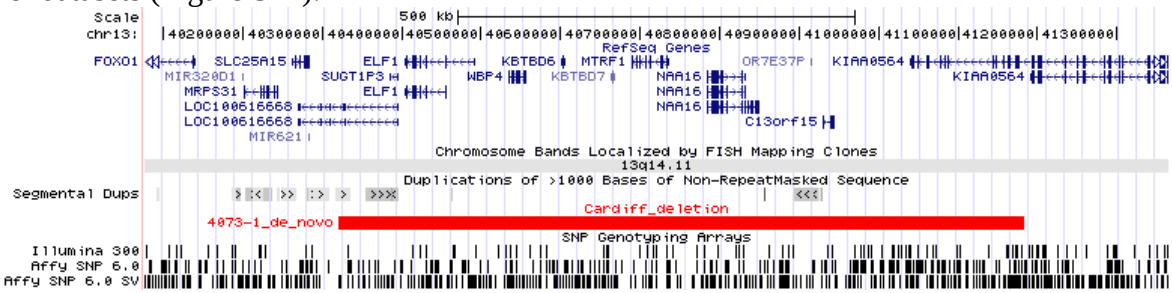

Figure S22. chr13:40319620-41182276

chr14:34464771-34627720. This 162kb deletion affects three known genes: *FAM177A1, PPP2R3C, SRP54* (Figure S23). No additional deletions spanning >50% of this region were found in the other datasets.

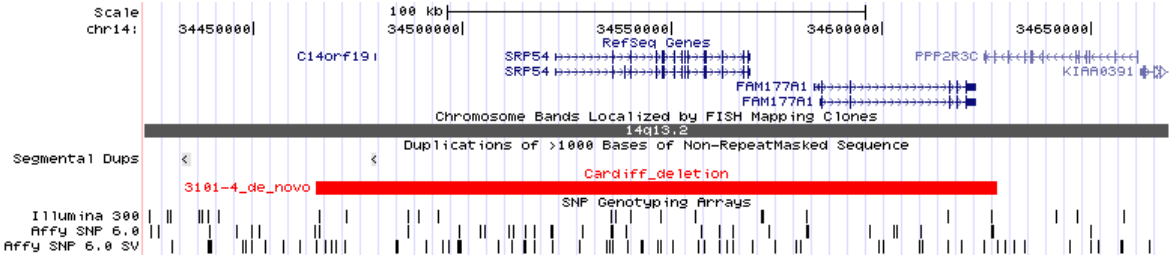

Figure S23. chr14:34464771-34627720.

15q11.2. Deletions at this locus have been proposed as risk factors for schizophrenia[3, 5]. We find two *de novo* deletions and two *de novo* duplications covering the interval. One of the cases with a *de novo* deletion has a brother treated for paranoid disorder.

In addition to the *de novos*, there are 3 transmissions and 3 non-transmissions of deletions and the respective ratio is 3:4 for duplications (Figure S24, 15q11.2 in the trios sample). The rate of the reciprocal duplications has not been reported as increased in schizophrenia cases before. The finding of so many *de novos* at this locus indicates that selection pressure must be operating on both deletions and duplications at this locus as these exist at low frequency in the general population[38]. The highest rate of this deletion is in patients with idiopathic generalized epilepsy: 1% [39]
The gene **CYFIP1** is within this interval and is part of the ARC complex (see main text).

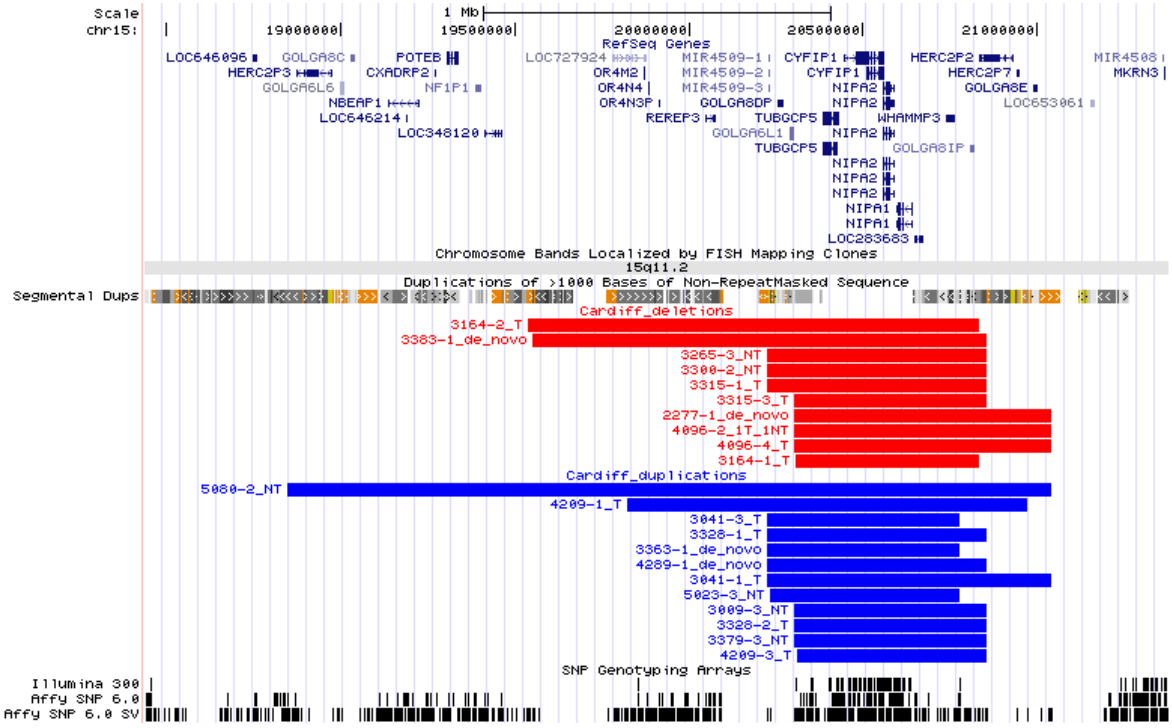

Figure S24. 15q11.2 CNVs in the trios sample. Nomenclature is as described in Figure S5. 4096-2 has transmitted the deletion to 4096-4, and not transmitted it to his other affected child, 4096-1, who is not shown in the figure.

15q13.1. We found a *de novo* duplication of 1.5Mb (Figure S25). This was already presented in detail in our previous paper[40], as the only *de novo* CNV that we found in 93 trios, while working with a much lower-resolution array. It contains four genes: *APBA2*, *KIAA0574*, *NDNL2*, *TJP1*. ***TJP1*** is part of the NMDAR complex (see main text and Table S9. The proband is not in the ISC[2] study. We found one additional duplication CNV within this region in a case from the MGS study[7] and no CNVs in the remaining studies.

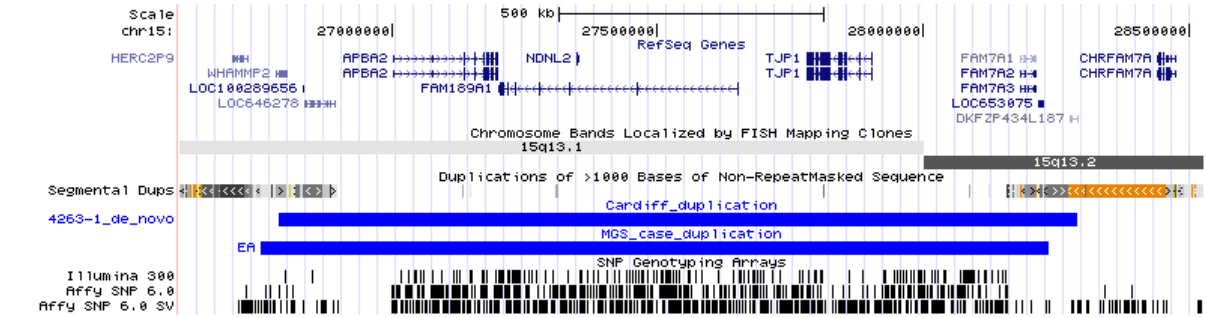

Figure S25. 15q13.1 duplication

15q13.3. This is a well established schizophrenia associated deletion[2, 3]. We find two *de novo* deletions that cover the full length of this region (Figure S26). Both probands with *de novo* deletions are in the ISC[2] study, but until now it had not been established that they are *de novo*.

Phenotypic data for carriers of 15q13.3 *de novo* deletions:

3024-1, *de novo* deletion of chr15:28707904-30326817, male, age at onset 31, paranoid schizophrenia, negative family history. Obtained very low marks in a mainstream school, but managed to work in as a tradesman. The main symptoms are persecutory delusions, aggression, auditory hallucinations and apathy. The mutation occurred on the paternal genome (25:0 informative SNPs, $p=5.96e^{-08}$).

2144-1 *de novo* deletion of chr15:28707904-30299500, female, age at onset 32, paranoid schizophrenia, negative family history. Achieved very good school results, had persecutory delusions, aggressive behaviour, olfactory hallucinations. The mutation occurred on the paternal genome (93:0 informative SNPs, $p=2.02e^{-28}$)

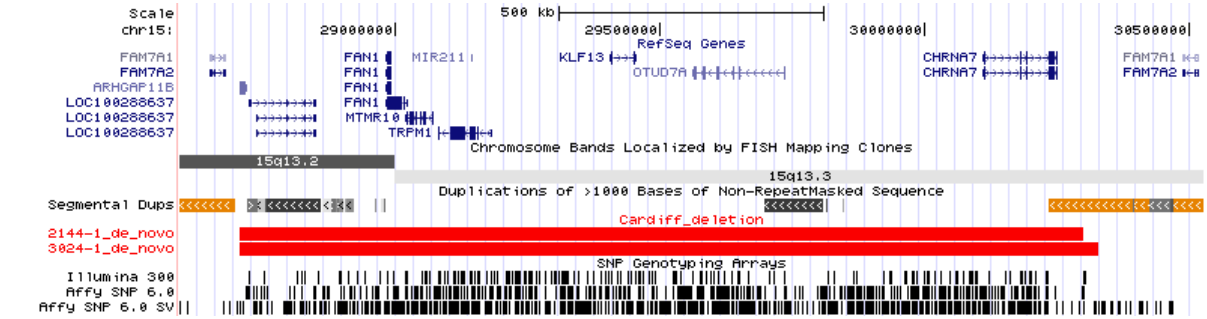

Figure S26. 15q13.3 CNVs in the trios sample. CNVs from other datasets are not shown, as this is an established schizophrenia locus.

16p11.2, duplication. This is a well established schizophrenia CNV locus[10]. We find one *de novo* duplication and one more transmission in this region, (Figure S27, 16p11.2 duplications in the trios sample). Both probands were reported in the ISC[2] and in the McCarthy meta-study[10] but it had not been established that one of them is a *de novo*. One gene from the NMDAR complex is within this region: ***MAPK3.***

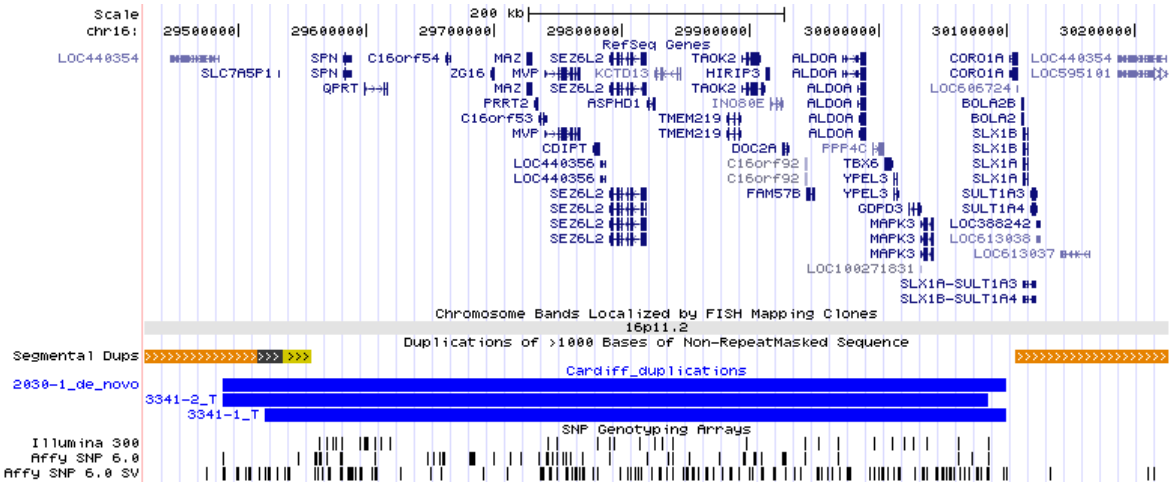

Figure S27. 16p11.2 duplications in the trios sample. CNVs from other datasets are not shown, as this is an established schizophrenia locus.

Phenotypic data on carriers of 16p11.2 duplications.
Our proband 2030-1 with a *de novo* duplication is female, with excellent school marks, negative family history, relatively late age at onset of 35 of paranoid schizophrenia. The presentation is unremarkable for schizophrenia, with prominent auditory hallucinations, thought disorder, delusions of reference and hypochondriacal delusions. SNP analysis revealed the duplication had occurred on the paternal genome, with 21:0 informative SNP, $p=9.54e^{-07}$, most likely by intrachromosomal NAHR, (18:0 informative SNPs).
3341-1 carrier of duplication, transmitted from the father. The proband is a female with a poor school achievement, age at onset 26 of paranoid schizophrenia. She had auditory hallucinations, aggressive behaviour, thought disorder, posturing, and a fairly good recovery between episodes. There is no family history (including the carrier father).

chr18:3515935-4332609, deletion of 817kb affecting the *DLGAP1* and *LOC284215* genes (Figure S28). We found additional deletions disrupting these genes in one case and one control in the MGS study[7], and none in the remaining studies. ***DLGAP1*** is a good candidate gene: it is part of the PSD-95, ARC and NMDAR complexes (see main text and discussions on the other DLG genes).

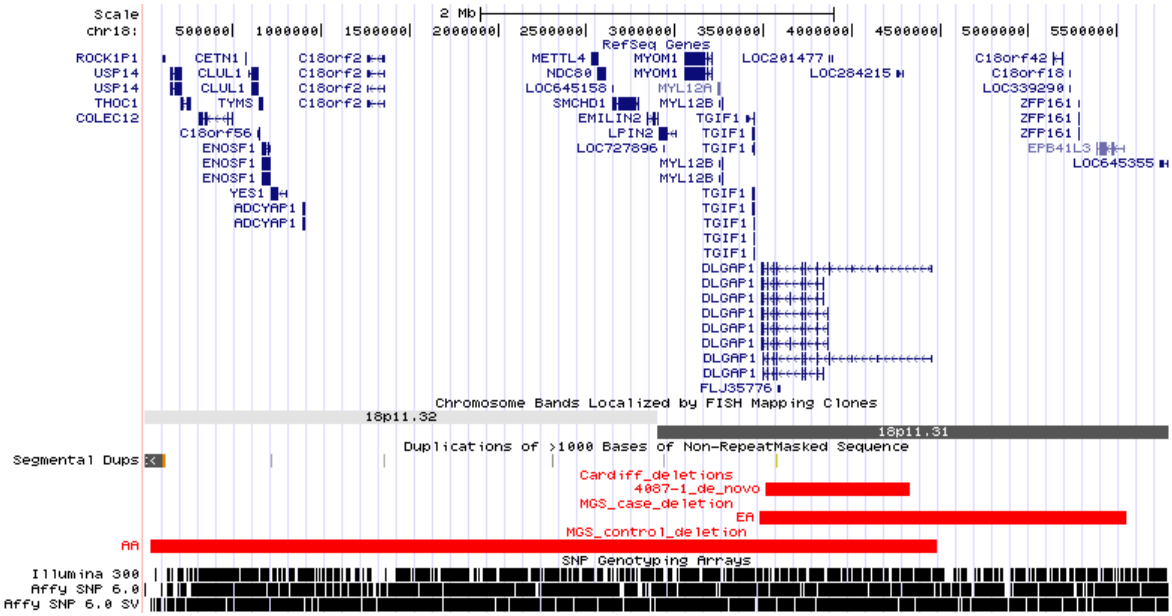

Figure S28. *DLGAP1* deletions.

Deletion at chr20:14694326-14863051, affecting *MACROD2*. This 169kb *de novo* deletion does not affect exons (Figure S29). Single cases of *de novo* deletions in *MACROD2* that intersect exons are reported in the autism studies by Pinto et al[26] and Sanders et al[27]. There were 64 other CNVs in the Cardiff sample, most of them <100kb, that were predominantly non-transmitted and did not intersect exons. Similar complex pictures are observed in the additional studies, with more controls having deletions within the *MACROD2* gene (data not presented). This suggests this particular *de novo* CNV is an incidental finding rather than a CNV of direct relevance to schizophrenia.

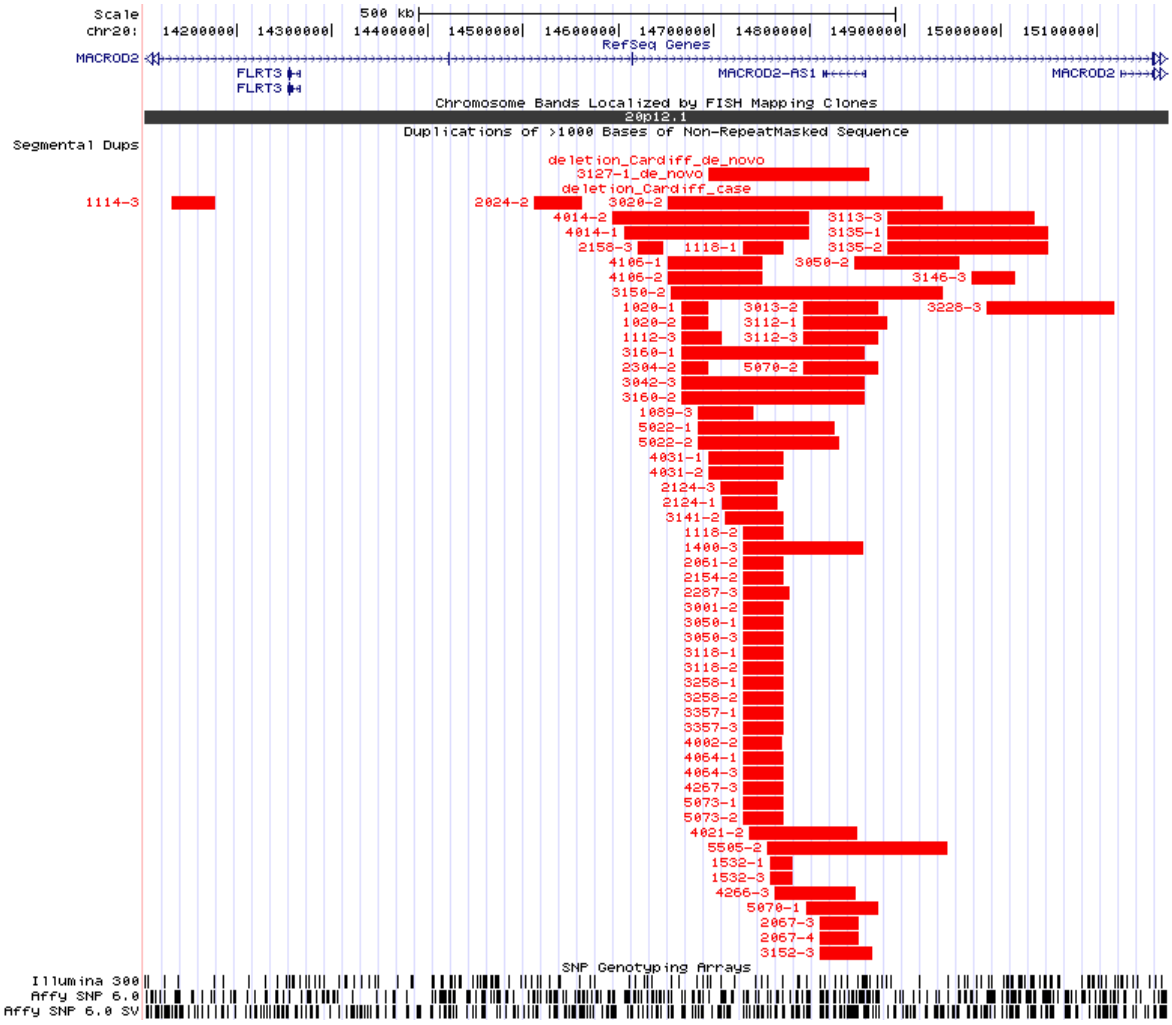

Figure S29. *MACROD2* in the trios data.

chr21:22698250-22778244. We detected a single *de novo* deletion of 80kb that does not affect any gene (figure S30). It was not examined in the other datasets.

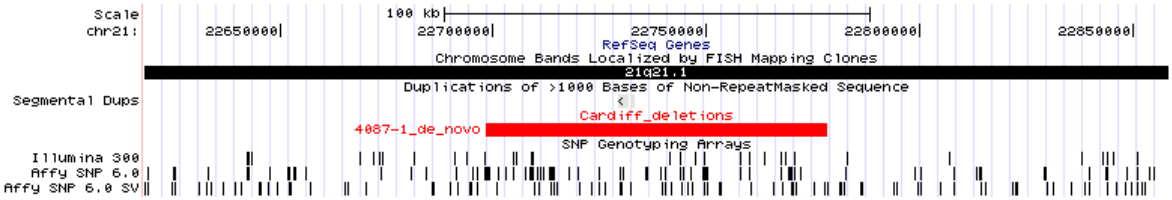

Figure S30. chr21:22698250-22778244

7) Case-Control Analysis.

We undertook an analysis of the frequency of the CNVs at loci where we had observed a *de novo* CNV in the case control sample (7,907 independent cases and 10,585 controls). To be conservative, we excluded loci that were already known to be over-represented in schizophrenia prior to this study and were found as *de novos* in the current study (3q29, 15q11.2, 15q13.3, 16p11.2). We also exclude CNVs at *MACROD2* for the reasons listed above. We found 32 CNVs in cases and 22 in controls (details listed in the above section and summarised Table S1 below). This corresponds to a rate of 0.4% in cases and 0.21% in controls, a nearly 2-fold enrichment (1-sided Fisher Exact test p = 0.012).

Cytoband	Location	Type	size (bp)	Genes	MGS (Case/ Con)[7]	ISC (Case/ Con)[2]	Kirov (Case/ Con)[5]	Trios (T/NT)	Ikeda (Case/ Con)[8]	Total (case+ de novo /con)	Fishers Exact test (1-tail)
1q21.1	chr1:144101459-144503409	Del	401,950	16 genes, TAR region	2/0	1/1	0/0	0/0	0/0	4/1	0.11
1q43	chr1:235475280-235639644	Del	164,364	RYR2	0/0	0/0	0/0	0/0	0/0	1/0	0.43
2q21.2	chr2:133504420-133879778	Del	375,358	NAP5	2/0	0/3	0/0	3/2	0/0	6/5	0.31
3q13.12	chr3:109330592-110198715	Dupl	868,123	9 genes	0/0	0/1	0/0	0/0	0/0	1/1	0.67
4q13.3	chr4:70935504-70969553	Del	34,049	3 genes	0/0			0/0		1/0	0.43
4q21.21	chr4:79944612-80081979	Del	137,367	BMP2K, PAQR3	0/0	1/0	0/0	0/0	0/0	2/0	0.18
7p14.1	chr7:38260614-38307187	Del	46,573	TARP	3 /6			0/0		4/6	1
7q11.23	chr7:72390286-76445231	Dupl	4,054,945	38 genes, WBS region	2/1	1/0	0/0	0/0	0/0	4/1	0.11
7q32.1	chr7:127275795-127447967	Del	172,172	C7orf54, SND1	0/0	0/0	0/0	0/0	0/0	1/0	0.43
8p23.2	chr8:4121968-4299810	Del	177,842	CSMD1	1 /2	1/0	0/0	0/2	0/0	3/4	1
8p23.1	chr8:10066862-10155414	Del	88,552	MSRA	2 /0	1/0	0/0	0/0	0/0	4/0	**0.03**
9q31.3	chr9:110859131-111433199	Dupl	574,068	4 genes	0/0	0/0	0/0	0/0	0/0	1/0	0.43
9q34.3	chr9:139769564-139792102 del chr9:139762152-139797423 dupl	Dupl+ Del		EHMT1	0/1	3/0	0/0	0/0	0/0	5/1	0.055
11q14.1	chr11:83472750-83842973	Del	370,223	DLG2	4/0	1/1	0/0	1/1	1/0	8/2	**0.02**
12q24.13	chr12:111723795-111776045	Del	52,250	RPH3A	0/0	0/0	0/0	0/0	0/0	1/0	0.43
13q14.11	chr13:40319620-41182276	Del	862,656	8 genes	0/0	0/0	0/0	0/0	0/0	1/0	0.43
14q13.2	chr14:34464771-34627720	Del	162,949	3 genes	0/0	0/0	0/0	0/0	0/0	1/0	0.43
15q13.1	chr15:26785056-28289366	Dupl	1,504,310	4 genes	1/0	0/0	0/0	0/0	0/0	2/0	0.18
18p11.31	chr18:3515935-4332609	Del	816,674	DLGAP1, FLJ35776	1 /1	0/0	0/0	0/0	0/0	2/1	0.39

Table S1. CNVs in other datasets that intersect *de novo* CNV loci found in this study. Known schizophrenia loci, those without genes and those at *MACROD2* are removed, leaving 19 loci. *De novo* events in the current study are included the statistical analysis. Two individual results (MSRA and DLG2 regions) are significant when tested with Fisher Exact test (shown in bold), but none is significant if correction for multiple testing of 19 independent loci is applied: a Bonferroni corrected threshold of significance is P=0.0025.

8) Size of the *de novo* CNVs.

D*e novo* CNVs in Bulgarian cases were similar in size (median 320.8kb) to the Icelandic control *de novos* (median 259.4kb, Mann-Whitney U test n.s.) and the Autism control *de novos* (median = 227.0kb, n.s.), Table S2. Both schizophrenia and Icelandic sets of *de novos* are significantly larger than the 1367 CNVs in the Bulgarian controls which were typed on the same array and filtered according to the same criteria (including the z-score method) as the Bulgarian *de novos* (median = 67.6kb, Mann-Whitney U test $P<10^{-6}$ for both comparisons). The size distributions of CNVs of the four groups are presented in a survival graph in Figure 1 of the main text, and in Table S3, while their median size is shown in Table S2.

De novo source	N	Median, bp
Schizophrenia de novos	34	320858
Icelandic control de novos	59	259355
Autism Sanders control denovos	14	226968
Bulgarian control CNVs	1367	67615

Table S2. **Median size of *de novo* and control CNVs in the different datasets**.

We also present the proportion of the different types of CNVs in a range of size bins in Figure S31. Please also see the sensitivity analysis in Section 1 in which we compare the size distributions of *de novo* CNVs in cases and Icelandic controls.

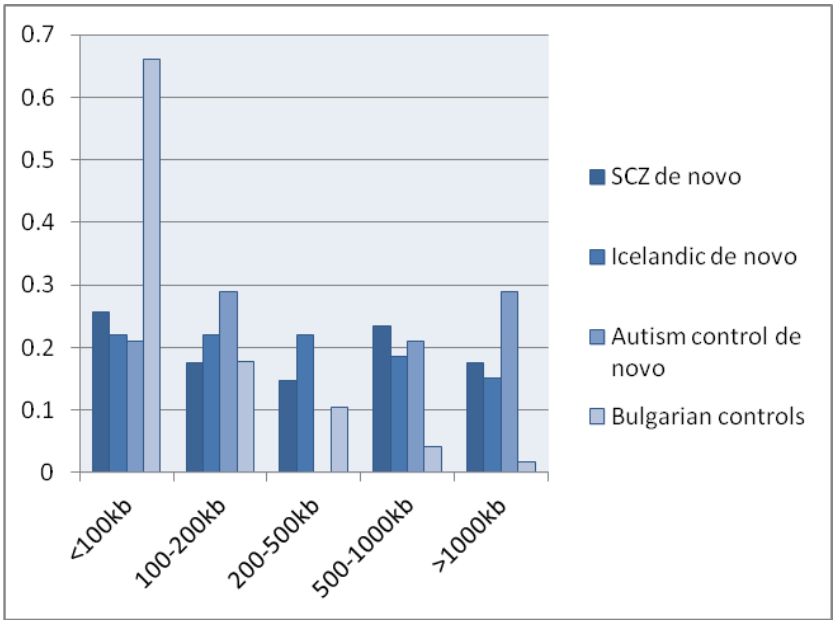

Figure S31. Size of CNVs. Proportion of different *de novo* and control (presumably mostly inherited) CNVs according to size ranges. *De novo* CNVs (in schizophrenia, Icelandic controls and autism controls) are over-represented in the size ranges >500kb, and under-represented <100kb size.

study	Array	N	All CNVs	>100kb	>200kb	>500kb	>1Mb
This study	Affymetrix 6.0	662 schizophrenia probands	N=34 5.1%	N=25 3.8%	N=19 2.9%	N=14 2.1%	N=6 0.9%
Sanders et al 2011[27]	Illumina 1M	872 unaffected siblings	N=14 1.6%	N=11 1.3%	N=7 0.8%	N=7 0.8%	N=4 0.5%
DeCODE Controls:	Illumina317 (59.2%) Illumina 370 (32.9%) Illumina 1M (7.9%)	2623 controls	N=59 2.2%	N=46 1.8%	N=33 1.3%	N=20 0.8%	N=9 0.3%
Controls Sanders + DeCODE		3495	N=73 2.0%	N=57 1.6%	N=40 1.1%	N=27 0.8%	N=13 0.4%

Table S3. Size of *de novo* CNVs in schizophrenia cases and two control samples.

9) Family history

There were 61 probands with a parent with schizophrenia, schizoaffective or other psychotic disorder. For the purpose of this study of *de novo* CNVs, we did not count as family history positive those cases with an affected sibling, affected children of probands, and one case with an affected monozygotic twin, as *de novo* events are still compatible with the inheritance in these families. One *de novo* was found in the 61 probands with a positive family history, a rate (1.6%) that is similar to that in the control subjects from Iceland (2.2%), those in the Sanders study[27] (1.6%) or in other previous studies (summarised by Sanders, 2011[27]). In contrast, 33 *de novos* were observed among the remaining 601 probands with a negative family history (as defined by the absence of the above criteria), a rate of 5.5%. Thus, the increased rate of *de novo* CNVs (in comparison to controls) was restricted to those without a positive family history of schizophrenia or schizoaffective disorder, a finding that is qualitatively in agreement with an earlier study[41], although in the present study, the difference between the two groups defined by family history is not statistically significant (p=0.35, two-tail Fisher Exact Test).

10) Gene set curation

The sources of proteomic data used to define gene sets in Table 2 of the main text were as follows.

Human PSD

In brief, PSDs were isolated from 9 human neocortex samples removed at neurosurgical biopsy and grouped into 3 pools of 3 samples for proteomic analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS). The consensus PSD consisted of proteins identified in all triplicates. The isolation and characterisation is reported elsewhere[42].

PSD complexes

PSD-95
Fernandez and colleagues[43] used mice carrying a genetically targeted tandem affinity purification tag in the PSD-95 locus to isolate PSD-95 complexes which were characterised using LC-MS/MS. We used their 'core' set of proteins, being those identified in at least three independent tandem purifications.

NMDAR
We took genes found in the NMDAR interaction network constructed by Pocklington and colleagues[44] (Table S9. ARC-NMDAR genes). This used high confidence protein-protein interactions curated from the literature to link the NMDAR to proteins previously identified by profiling of NMDAR complexes isolated from mouse forebrain using immunoprecipitation and peptide affinity methods[32, 33].

ARC
Using the same methods described for isolating PSD-95 complexes[43], protein complexes bound to the mouse Arc protein were identified (EF, MOC, JCC, SGNG, manuscript in preparation) (Table S9. ARC-NMDAR genes).

mGluR5
Farr and colleagues[45] used immunoprecipitation to isolate mGluR5 complexes from rat brain lysates.

In order to analyse these complexes as strict sub-components of the PSD (and to ensure their relevance to human synapse function), any genes not found in the human PSD were removed.

Presynapse

Presynaptic vesicles
This was taken unmodified from the publication of Takamori and colleagues[46], who characterised synaptic vesicle proteins purified from rat brain using LC-MS/MS.

Active zone
The presynaptic active zone contains elements of both the release site machinery (governing vesicle fusion and retrieval to/from the membrane) and docked synaptic vesicles. This gene

set was taken from (Morciano et al. 2009)[47], who combined nanoLC ESI MS/MS and MALDI-TOF-MS identification of proteins in samples purified from rat brain.

Others

All remaining gene sets were curated from Foster et al[48], who used protein correlation profiling (comparing mass spectrometry intensity profiles to those of known marker proteins for individual organelles) to generate high quality protein localisation data for generic subcellular components, using tissue from mouse liver.

Rat and mouse genes were mapped to human using orthology files HMD_Human5.rpt and HMD_RatHuman5.rpt downloaded from the MGI database (18/3/2010).

Mapping genes to chromosomal locations

Chromosomal locations for Build 36.3 were downloaded from NCBI, and all features for which the evidence code was not 'identical' (i.e. localisation was ambiguous) were removed. Feature ids were then checked against current gene annotation (file gene_info.gz downloaded from NCBI on 17/02/2010), and any ids not annotated as 'protein-coding' (e.g. pseudo-genes, unknown features, RNAs…) were removed. Ids were further checked against the HUGO Gene Nomenclature Committee (HGNC) and any whose chromosome (as annotated by NCBI) did not match that given in HGNC were also removed. The resulting set of genes and their chromosomal locations were used in all following analyses.

11) Comparison of gene sets hit by *de novo* CNVs in cases with those hit by CNVs in Bulgarian controls or hit by *de novo* CNVs in Icelandic controls.

This analysis was presented in the main text, here we provide more detail. The numbers of genes in the various gene sets hit by the 34 *de novo* CNVs were compared to those hit by 1367 CNVs from 605 Bulgarian unaffected controls. Here, and in all subsequent analyses, a gene was assumed to be "hit" by a CNV if a CNV overlapped any part of its length, according to Build 36.3 (see previous paragraph). The controls were genotyped at the same laboratory on the same arrays (Affymetrix 6.0). The methods for control CNV analysis were as for the *de novos*, including evaluation with the z-score method, accepting only "suggestive calls", (see Section 3). Although not subsequently validated with Agilent arrays, as discussed in Section 4 of the Supplementary material, z-score validated CNVs have a low false positive call rate (see Section 4). This can be expected to be even lower for CNVs that are not ascertained to be putative *de novos*, the rarity of *de novos* conferring relatively unfavourable signal to noise characteristics for this class of CNV in comparison with CNVs in general.

The impact of biases relating to gene and CNV size has been discussed by Raychaudhuri et al[49]. To overcome those biases, an excess of genes hit by the *de novo* CNVs was tested by fitting the following logistic regression model to the combined set of CNVs:

logit (pr(case)) = CNV size + Total number of genes hit outside the gene set + number of genes hit in the gene set.

and comparing the change in deviance between it and the model

logit (pr(case)) = CNV size + Total number of genes hit outside the gene set

P-values for a one-sided test of an *excess* of genes hit in the gene set by *de novo* CNVs are given in Table 2 of the main text.

By comparing case to control CNVs, this analysis allows for the possibility of non-random CNV location unrelated to disease (i.e. CNVs tend to occur in specific locations of the genome and this is unrelated to case status, both in cases and controls). The inclusion of CNV size in the regression allows for the fact that *de novo CNVs* are larger than typical CNVs (and thus likely to hit more genes, regardless of function). Inclusion of the total number of genes hit outside the gene set in the regression corrects for case CNVs hitting more genes overall (regardless of function) than control CNVs.

We use the same method of analysis to test if the *de novos* found in cases with schizophrenia hit more genes from the gene sets than do the 59 *de novos* found in the Icelandic controls. Those data are presented in Table 2 in the main text.

Testing enrichment of genes hit in Gene Ontology gene sets

GO sets were taken from the gene2go file available at the NCBI ftp site on 28/7/10. Gene sets were analyzed as above. Analysis was restricted to gene sets with at least 10 gene hits (case+control) to remove the possibility of small gene sets being counted as significant from a small number of CNV hits. The 10 sets most significantly enriched for genes hit by *de novo*

CNVs among all GO categories are shown below in Table S4. Category GO:45202, corresponding to the synapse, was by two orders of magnitude the most significantly enriched for *de novo* CNV hits ($p=9.60x10^{-9}$).

Pathway number	No. Genes	No. gene hits (denovo)	No. gene hits (control)	P-value (filtered)	
GO:45202	367	14	35	9.60E-09	synapse
GO:15674	200	14	24	1.95E-06	di-, tri-valent inorganic cation transport
GO:19717	92	8	14	4.23E-06	synaptosome
GO:70838	153	13	22	4.62E-06	divalent metal ion transport
GO:31267	96	6	7	1.55E-05	small GTPase binding
GO:17016	86	6	7	1.55E-05	Ras GTPase binding
GO:51020	107	6	10	2.97E-05	GTPase binding
GO:44456	271	9	23	4.23E-05	synapse part
GO:51015	50	5	9	7.41E-05	actin filament binding
GO:45211	150	6	15	9.27E-05	postsynaptic membrane

Table S4. Enrichment of GO gene sets for *de novo* CNV hits.

Enrichment of genes hit in Gene Ontology synapse gene sets

Analysis of GO categories also allowed us to undertake a more systematic analysis of synapse related genes. The enrichment analysis described above was applied to the gene sets defined by the subcategories of GO:45202 ("synapse"). This was performed in three different ways:
a) using all genes in the gene sets, *b)* removing genes that are part of the ARC or NMDAR gene sets defined above, *c)* removing genes that are part of the PSD gene set defined above. These analyses were performed using all 1367 filtered control CNVs and the results are shown in Table S5.

Pathway number	No. Genes	No.gene hits (denovo)	No. gene hits (control)	P-value (all genes)	P-value (noARC/ NMDAR)	P-value (no PSD)	Description
GO:45202	367	14	35	9.60E-09	0.017	0.049	synapse
GO:44456	271	9	23	4.23E-05	0.044	0.143	synapse part
GO:45211	150	6	15	9.27E-05	0.133	0.110	postsynaptic membrane
GO:30672	44	3	0	5.35E-04	4.22E-04	0.036	synaptic vesicle membrane
GO:14069	81	4	11	3.05E-03	0.929	0.926	postsynaptic density
GO: 8021	84	3	4	0.105	0.077	0.470	synaptic vesicle
GO:31594	24	0	1	0.576	0.573	0.573	neuromuscular junction
GO:48786	6	0	2	0.914	0.901	0.904	presynaptic active zone
GO:32279	10	0	2	0.914	0.901	0.904	asymmetric synapse
GO:42734	35	0	3	0.917	0.905	0.909	presynaptic membrane

Table S5. Enrichment analysis of subcategories of GO:45202 ("synapse") with and without ARC, NMDAR and PSD genes.

It can be seen from Table S5 that GO:45202 "synapse" is at least two orders of magnitude more significant than any of its children when all genes in the gene sets are used (P value all genes). When ARC and NMDAR genes are removed, GO:30672 "synaptic vesicle membrane" ($p=4.22\times10\text{-}4$), GO:44456 "synapse part" ($p=0.044$) and GO:45202 "synapse" ($p=0.017$) remain significant, although the significance of the latter two is greatly reduced.

When PSD genes are removed, only GO: 45202 "synapse" (p=0.049) and GO:30672 "synaptic vesicle membrane" (p=0.036) are nominally significantly enriched, and their significance is greatly reduced Thus it appears that the PSD gene set can account for nearly all of the signal observed in the GO gene sets.

12) Testing for enrichment in number of *de novo* CNVs hitting gene sets by random placement

To investigate the robustness of our results, and in particular the impact of using 'control' CNVs, we undertook a different analysis based upon comparing the number of *de novo* CNVs hitting each gene set with that found when CNV locations were randomized. Taking the 30 *de novo* CNVs that hit genes, we randomized (x100,000) their locations – ensuring that each random assignment hits at least one gene, and that the probability of a gene being hit was proportional to its length. In order to account for the possibly uneven distribution of *de novo* CNVs and/or gene sets between chromosomes, the randomized locations of each *de novo* were restricted to the chromosome on which it was actually found.

For each gene set we then calculated P(N) – the probability of N or more *de novos* hitting the set, where N is the actual number of *de novo* hits. While this can be estimated directly (by counting the proportion of randomised trials in which the total number of hits is $\geq$ N), such estimates become inaccurate when P(N) is small relative to the number of randomisation trials (e.g. for the ARC complex). To generate more accurate estimates of P(N), we made use of the fact that CNVs were not allowed to change chromosomes when their positions were randomised (so the number of CNVs on each chromosome is the same in each trial and much smaller than N). To do this, we first note that P(N) can be written as

$$P(N) = \sum P_1(n_1)\, P_2(n_2) \ldots P_{22}(n_{22})$$

where the sum is over all $(n_1, n_2, \ldots n_k)$ such that $n_1 + n_2 + \ldots + n_k \geq N$ and $P_m(n)$ is the probability of the gene set being hit by n CNVs on chromosome m (where n lies between 0 and the total number of *de novos* found on chromosome m). As there are only a few possible values of n for any one chromosome, $P_m(n)$ can be much more reliably estimated than P(N). As expected, estimates of P(N) calculated using this method agreed with those obtained by the direct approach where P(N) was large. Bonferroni corrected P(N) are given in Table S6.

Gene Set	N genes	N CNVs	P-value
PSD	664	14	0.0024
- ARC complex	25	8	2.21×10^{-8}
- NMDAR complex	59	7	2.95×10^{-4}
- PSD-95 complex	58	4	0.11
Presynapse	415	7	1
- Synaptic vesicle	333	7	0.72
- Active zone	165	2	1
Nucleus	160	5	0.075
Mitochondrion	189	3	1
Cytoplasm	263	3	1
Endoplasmic Reticulum (ER)	94	1	1

ER/Golgi-derived vesicles	94	0	1
Recycling Endosomes	65	0	1
Early Endosomes	17	0	1
Golgi	31	0	1
Plasma Membrane	50	0	1

Table S6. Enrichment in number of de novo CNVs hitting each gene set, compared to randomised placement of gene-hitting *de novos* (restricted to the same chromosomes).

This analysis confirmed the main findings of the previous analysis, namely, that the *de novo* CNVs we observed in schizophrenia preferentially hit genes encoding ARC and NMDAR complex proteins.

13) Partitioning the PSD CNV signal

Is the enrichment of *de novo* hits in PSD attributable to genes in the synaptic vesicle and/or nucleus complexes?

In the unconditional analysis (Table 2, main text) there is some weak evidence for enrichment of the 'synaptic vesicle' and 'nucleus' gene sets in *de novo* CNVs. To determine whether these are independent of the strong signal in the PSD, or indeed whether they can account for the PSD signal, we undertook conditional analyses. Since the synaptic vesicle and nucleus complexes overlap with, but are not subsets of, the PSD, the test for conditional enrichment of *de novo* hits in PSD was carried out as follows:
Denote by G1 the set of genes being conditioned on (this can be synaptic vesicle, nucleus or the two complexes combined), and by G2, the set of genes in PSD that are not in G1, and by G3, the set of genes not in PSD or G1.

The following two logistic regression models are fitted, and the deviances compared:

M1: logit (pr(case)) = size + #gene hits in G3 + #gene hits in set G1
M2: logit (pr(case)) = size + #gene hits in G3 + #gene hits in set G1+#gene hits in set G2.

Analyses were carried out using synaptic vesicle and nucleus separately as the conditioning set, and also conditioning on both synaptic vesicle and nucleus combined. The results are shown in Table S7.

Conditioning set	**Conditional p-value for enrichment of PSD *de novo* hits**
Synaptic vesicle	5.20×10^{-3}
Nucleus	5.24×10^{-4}
Combined	0.016

Table S7. **Enrichment p-values for the PSD gene set conditional on the synaptic vesicle and nucleus gene complexes.**

It can be seen that neither synaptic vesicle nor nucleus, nor their combination, account for the PSD signal.

Analyses were also performed to see whether there was enrichment for *de novo* CNV hits in the synaptic vesicle or nucleus complexes after conditioning on the PSD signal. The method described above was used, with PSD being the conditioning set G1, and genes in the synaptic vesicle (or nucleus) complex, but not in PSD, being the test set G2.

One-sided p-values for enrichment conditional on PSD are given below in Table S8.

Neither synaptic vesicle nor nucleus show significant enrichment for *de novo* hits conditioning on the *de novo* hits in PSD. This is unsurprising since all the genes hit by *de novos* in the synaptic vesicle or nucleus complexes are also members of PSD (see Table 2 of the main text)

Test set	p-value for enrichment of *de novo* hits conditional on PSD
Synaptic vesicle	0.955
Nucleus	0.663

Table S8. **Enrichment p-values for the synaptic vesicle and nucleus gene complexes conditional on PSD.**

Are the associations to ARC, NMDAR and PSD-95 independent?

This was tested as follows: For each pair of gene sets, we denote the set of genes in gene set 1 by G1. We denote by G2, the set of genes in gene set 2 that are not in gene set 1. The enrichment of genes hit by *de novo* CNVs in gene set 2 conditioning on the enrichment of *de novo* CNV hits in gene set 1 can be tested by fitting the following two logistic regression models and comparing their deviances:

M1: logit (pr(case)) = size + #non-PSD gene hits + #gene hits in set G1
M2: logit (pr(case)) = size + #non-PSD gene hits + #gene hits in set G1 + #gene hits in set G2

A significant p-value indicates that the enrichment of *de novo* CNV hits in gene set 2 is not attributable solely to an enrichment of *de novo* CNV hits in gene set 1. This analysis yielded the following results:

P-value for enrichment of CNV hits in NMDAR allowing for ARC = 0.019
P-value for enrichment of CNV hits in ARC allowing for NMDAR = 2.17×10^{-4}
P-value for enrichment of CNV hits in PSD-95 allowing for ARC = 0.603
P-value for enrichment of CNV hits in ARC allowing for PSD-95 = 4.22×10^{-4}

Thus, we conclude that the ARC gene set can explain the PSD-95 effect, but that the NMDAR is partially independently associated.

Do associations in ARC and NMDAR explain association with PSD?

Given that ARC and NMDAR are subsets of the PSD, and that each is independently associated with schizophrenia, we investigated whether the association observed to the wider PSD category could be accounted for by these two subcategories.

Denote the set of genes in ARC and/or NMDAR by G1. Denote the set of genes in PSD, but in neither ARC nor NMDAR, by G2. The enrichment of *de novo* CNV hits in PSD genes in G2 (that is, outside the ARC and NMDAR complexes) was tested by fitting the following two logistic regression models:

M1: logit (pr(case)) = size + #non-PSD gene hits + #gene hits in set G1
M2: logit (pr(case)) = size + #non-PSD gene hits + #gene hits in set G1 + #gene hits in set G2

The enrichment of *de novo* CNV hits in the PSD genes outside ARC and NMDAR conditional on the enrichment of hits in the ARC+NMDAR genes can be tested by comparing the deviance of model M1 to model M2. A significant p-value indicates that there is significant enrichment of *de novo* CNV hits in the PSD genes that are not contained within the ARC and NMDAR gene sets.

Comparison of deviances gave a p-value of 0.231. Thus, there is no evidence for enrichment of *de novo* CNV hits in PSD genes lying outside the ARC and NMDAR complexes. This in turn suggests that the observed enrichment of *de novo* CNV hits in the PSD gene set is driven by genes in ARC and NMDAR.

14) Genes in the NMDAR and ARC gene sets

The following table lists the genes in the ARC and NMDAR gene sets and indicates those hit by *de novo* CNVs.

gene	chr	start	end	NMDAR	ARC	*De novo*
GNB1	1	1706589	1812355	*		
CAPZB	1	19537860	19684579	*		
ATP1A1	1	116717359	116748919	*	*	
ATP1B1	1	167342571	167368584		*	
GLUL	1	180618292	180627573		*	
ATP2B4	1	201862551	201979832	*		
ACTN2	1	234916422	234993863	*		
SPTBN1	2	54601684	54749366	*		
MAP2	2	209997016	210307079	*		
IQSEC1	3	12914374	12983960		*	
CTNNB1	3	41216004	41256938	*		
GAP43	3	116825142	116922842	*		
GNB4	3	180599696	180652065	*		
DLG1	3	198255819	198509844	*	*	1 del
CRMP1	4	5873392	5945686		*	
PPP3CA	4	102163610	102487376	*		
SLC25A4	4	186301392	186305419		*	
HOMER1	5	78705542	78845456	*		
CAMK2A	5	149579248	149649529	*	*	
SYNGAP1	6	33495825	33529444	*		
RAC1	7	6380651	6410123	*		
RALA	7	39629687	39714244	*		
CAMK2B	7	44225416	44331749	*		
STX1A	7	72751476	72771898	*		1 dup
YWHAG	7	75794052	75826252	*		1 dup
GNB2	7	100109311	100114728	*		
PRKAR2B	7	106472414	106589492	*		
CAPZA2	7	116289799	116346549	*		
ARF5	7	127015753	127018988		*	
DLGAP2	8	1436976	1644049		*	
NEFM	8	24827212	24832511	*		
DPYSL2	8	26491338	26571610		*	
PTK2B	8	27224916	27372820	*		
YWHAZ	8	102000090	102034745	*		
GSN	9	123070201	123134941	*		
STXBP1	9	129414390	129494816	*	*	
DNM1	9	130005484	130057348	*		
GRIN1	9	139152663	139183028	*	*	
GLUD1	10	88800223	88844603		*	
INA	10	105026910	105040093	*		
SHANK2	11	69991609	70185520	*		
RAB6A	11	73064331	73149849	*		
DLG2	11	82843701	84312113	*	*	2 del
HSPA8	11	122433410	122438054		*	
GAPDH	12	6513918	6517797	*		

TUBA1B	12	47807832	47811571	*		
SYT1	12	78135414	78368974	*		
SLC25A3	12	97511534	97519909		*	
PTPN11	12	111340919	111432100	*		
CIT	12	118607981	118799475	*		
CYFIP1	15	20444125	20555044		*	2 del, 2 dup
TJP1	15	27779648	27901998	*		1 dup
MYO5A	15	50392602	50608540	*		
MAP2K1	15	64466679	64570936	*		
PKM2	15	70278424	70310738		*	
MAPK3	16	30032927	30042131	*		1 dup
YWHAE	17	1194590	1250267	*		
DLG4	17	7033933	7063754	*	*	
ALDOC	17	23924260	23928078	*		
NSF	17	42023398	42189997	*		
BAIAP2	17	76623557	76705827		*	
DLGAP1	18	3488837	3870135	*	*	1 del
CDH2	18	23784933	24011189	*		
MBP	18	72819777	72973762	*	*	
RAB3A	19	18168611	18175839	*		
ACTN4	19	43830167	43913010		*	
ATP1A3	19	47162574	47190222		*	
SHANK1	19	55856896	55912007	*		
PPP2R1A	19	57385046	57421483	*		
PRKCG	19	59077279	59102713	*		
PLCB1	20	8061296	8813547	*		
SNAP25	20	10147477	10236065	*		
SRC	20	35406502	35467235	*		
GNAS	20	56848190	56919642	*		
YWHAH	22	30670479	30683590	*		
IQSEC2	X	53278783	53327521		*	
DLG3	X	69581544	69639258	*		
PLP1	X	102918410	102934204		*	
SLC25A5	X	118486437	118489307		*	
FLNA	X	153230159	153256123	*		

Table S9. ARC-NMDAR genes. Start and End refer to the positions of the genes.
* indicates whether a gene encodes a protein that belongs to the ARC and/or NMDAR complex.

15) Overlap between the loci affected by de novos in schziophrenia probands from the current study and other datasets

Cytoband	Location: *de novos* in 662 schizophrenia probands from the current study	Type	Genes	Xu et al[41]	Icelandic control *de novos*:	Sanders et al[27]	Cooper et al[17]
1q21.1	chr1:144101459-144503409	Del	16 genes, TAR region				13/2 del, p=0.066
1q43	chr1:235475280-235639644	Del	RYR2				
2q21.2	chr2:133504420-133879778	Del	NAP5				
3q13.12	chr3:109330592-110198715	Dupl	9 genes				
3q29	chr3:197185548-198825231	Del	21 genes, 3q29 syndrome			1del	6/0 del, p=0.079
4q13.3	chr4:70935504-70969553	Del	HTN1,HTN3				
4q21.21	chr4:79944612-80081979	Del	BMP2K,PAQR3				
6q12	chr6:68675955-68761101	Dupl	-				
7p14.1	chr7:38260614-38307187	Del	TARP				
7q11.23	chr7:72390286-76445231	Dupl	38 genes, WBS region		(1 del)	4 dup	16/0 dupl, p=0.0011
7q32.1	chr7:127275795-127447967	Del	C7orf54,SND1				
8p23.2	chr8:4121968-4299810	Del	CSMD1		1 del		
8p23.1	chr8:10066862-10155414	Del	MSRA				
9p22.3	chr9:16310745-16327782	Del	-				
9q31.3	chr9:110859131-111433199	Dupl	4 genes				
9q34.3	chr9:139762152-139797423	Dupl	EHMT1			1 del	4/0 dup, p=0.18
9q34.3	chr9:139769564-139792102	Del					60/0 del, $p=8.5\times10^{-12}$
11q14.1	chr11:83472750-83842973	Del	DLG2				
11q14.1	chr11:84006106-84226064	Del					
12q24.13	chr12:111723795-111776045	Del	RPH3A				
12q24.33	chr12:130388037-130659530	Dupl	-				
13q14.11	chr13:40319620-41182276	Del	8 genes			1 del	
14q13.2	chr14:34464771-34627720	Del	3 genes				
15q11.2	chr15:19548923-20852202	Del	8 genes				94/19, $p=2.1\times10^{-5}$
15q11.2	chr15:20302446-21038975	Del	6 genes				
15q11.2	chr15:20224751-20777909	Dupl	5 genes		1 dup		64/36, p=0.66
15q11.2	chr15:20224751-20852202	Dupl					
15q13.1	chr15:26785056-28289366	Dupl	4 genes				
15q13.3	chr15:28707904-30326817	Del	7 genes			1 del, (1dup)	42/0, $p=1.8\times10^{-8}$
15q13.3	chr15:28707904-30299500	Del					
16p11.2	chr16:29488112-30099396	Dupl	31 genes		(2 del)	4dup (7del)	28/2, p=0.0004
18p11.31	chr18:3515935-4332609	Del	DLGAP1,FLJ35776				
20p12.1	chr20:14694326-14863051	Del	MACROD2		1 del	1del	
21q21.1	chr21:22698250-22778244	Del	-				

Table S10. Overlaps between the *de novos* in the current studies with other studies. We show overlaps with the following studies: Xu et al[41]: a study on *de novo* CNVs among 152 probands affected with schizophrenia; the overlap with the Icelandic *de novos* among 2623 controls from that population, part of the current study; Sanders et al[27]: a large study on *de novo* CNVs in 1124 autism probands; and Cooper et al[17]: a case-control study on loci implicated in developmental delay among 15767 cases and 8329 controls. The case control study by Cooper et al[17] is based on testing 45 established known genomic disorders regions, therefore not every one of our regions affected by *de novos* was tested. This study does not report the *de novo* rates in these loci (although de novo events have been reported for each of these loci in other studies), but we still present their data in order to demonstrate the very high overlap between our *de novo* loci and those implicated in developmental delay. We show overlaps with CNVs from the same type (del/dupl), and for completeness, for

the instances where both types are known to cause disease (shown in brackets in the table). Only three Icelandic control CNVs of the same type were found to overlap our loci, while seven were found in the much smaller sample of autism probands[27], and eight are in loci associated with developmental delay[17]. For the study by Cooper et al[17] on developmental delay we only present the results on the same type of CNV, and include the significance level reported in that study. There were no overlaps with the previous smaller study on *de novo* CNVs in schizophrenia by Xu et al[41].

16) De novo CNVs identified in the Icelandic control sample.

sample	chr	start	end	size_bp	N_markers	del/dup	N_genes	genes
Sample31	1	8263306	8814839	551533	36	dupl	2	RERE,SLC45A1
Sample26	1	234236524	234410280	173756	38	del	2	GPR137B,NID1
Sample27	1	241629387	241677072	47685	10	del	1	SDCCAG8
Sample61	2	238555681	238623290	67609	11	del	1	UBE2F
Sample52	3	2585327	3437953	852626	200	dupl	4	CNTN4,CRBN,IL5RA,TRNT1
Sample9	3	4078926	4188033	109107	13	del	0	
Sample56	3	67062966	67251220	188254	25	dupl	1	KBTBD8
Sample11	3	71237486	71819797	582311	90	dupl	2	EIF4E3,FOXP1
Sample14	3	95019980	99373057	4353077	241	del	14	ARL13B,ARL6,DHFRL1,DKFZp667G2110,EPHA6,GABRR3,MINA, NSUN3,OR5AC2,OR5H14,OR5H15,OR5H1,PROS1,STX19
Sample54	3	174806412	176937361	2130949	287	del	2	NAALADL2,NLGN1
Sample25	4	10007639	10110140	102501	13	dupl	2	MIST,ZNF518B
Sample23	4	189677818	190015659	337841	49	dupl	0	
Sample69	5	409381	491564	82183	14	dupl	1	AHRR
Sample39	5	20914020	23316296	2402276	180	dupl	1	CDH12
Sample59	5	104300738	104510644	209906	21	del	0	
Sample66	6	31468368	31559455	91087	37	dupl	2	HCP5,MICA
Sample6	6	118779768	119061916	282148	19	dupl	2	C6orf204,PLN
Sample43	6	162593816	162797372	203556	31	dupl	1	PARK2
Sample33	7	188894	295765	106871	16	del	1	FAM20C
Sample53	7	13422396	13681751	259355	39	del	0	
Sample24	7	15803157	16310427	507270	71	dupl	1	LOC729920
Sample3	7	72818575	73971272	1152697	68	del	16	CLDN3,CLDN4,CLIP2,EIF4H,ELN,GTF2I,GTF2IRD1,GTF2IRD2, LAT2,LIMK1,NCF1,PMS2L5,RFC2,STAG3L2,,WBSCR27WBSCR28
Sample20	7	98348058	98436427	88369	10	dupl	1	TRRAP
Sample33	7	157746991	158777552	1030561	109	del	5	FAM62B,NCAPG2,PTPRN2,VIPR2,WDR60
Sample59	8	4307750	4433352	125602	33	del	1	CSMD1
Sample21	8	11006485	11691234	684749	119	dupl	7	AMAC1L2,BLK,FAM167A,GATA4,MTMR9,NEIL2,XKR6
Sample44	8	25278082	25520009	241927	24	del	4	CDCA2,DOCK5,GNRH1,KCTD9
Sample67	8	43545847	43837769	291922	10	del	0	
Sample68	8	43545847	43837769	291922	10	del	0	
Sample1	9	468153	605310	137157	40	del	1	KANK1
Sample37	9	4605020	6110336	1505316	200	dupl	18	AK3,C9orf46,C9orf68,CD274,CDC37L1,ERMP1,INSL4,INSL6, JAK2,KIAA1432,KIAA2026,MLANA,PDCD1LG2,PPAPDC2, RANBP6RCL1,RLN1,RLN2

Sample63	9	27547919	27591332	43413	11	del	1	C9orf72
Sample64	9	108611914	108635956	24042	10	del	0	
Sample19	10	7917790	8021528	103738	11	del	1	TAF3
Sample12	10	7928373	8021528	93155	10	del	1	TAF3
Sample29	10	67880428	68013385	132957	19	del	1	CTNNA3
Sample46	11	6739977	6849442	109465	20	del	5	OR10A2,OR10A5,OR2AG1,OR2AG2,OR6A2
Sample2	11	84757067	85458970	701903	54	dupl	7	CCDC83,CCDC89,CREBZF,PICALM,SYTL2,TMEM126A,TMEM126B,DLG2*
Sample40	12	50505409	51036005	530596	83	del	11	ACVR1B,ACVRL1,ANKRD33,C12orf44,GRASP,KRT7,KRT80,KRT81, KRT83,KRT86,NR4A1
Sample8	12	98533988	98721262	187274	20	del	2	ANKS1B,FAM71C
Sample15	12	125155361	128399092	3243731	486	dupl	3	GLT1D1,SLC15A4,TMEM132D
Sample4	15	20455753	20635884	180131	43	dupl	3	CYFIP1,NIPA1,NIPA2
Sample7	15	32729249	32769236	39987	12	del	0	
Sample35	15	65176215	65747577	571362	44	del	4	FLJ11506,IQCH,MAP2K5,SMAD3
Sample10	16	15032942	16197033	1164091	112	del	11	ABCC1,ABCC6,C16orf45,C16orf63,KIAA0430,MPV17L,MYH11,NDE1, NTAN1,PDXDC1,RRN3
Sample62	16	15387380	16188508	801128	102	del	8	ABCC1,ABCC6,C16orf45,C16orf63,KIAA0430,MPV17L,MYH11,NDE1
Sample5	16	29563365	30062241	498876	31	del	24	ALDOA,ASPHD1,C16orf53,C16orf54,CCDC95,CDIPT,DOC2A,FAM57B, FLJ25404,GDPD3,HIRIP3,KCTD13,MAPK3,MAZ,,MVPPPP4C,PRRT2, QPRT,SEZ6L2,SPN,TAOK2,TBX6,TMEM219,YPEL3
Sample49	16	29563365	30085308	521943	32	del	24	ALDOA,ASPHD1,C16orf53,C16orf54,CCDC95,CDIPT,DOC2A,FAM57B, FLJ25404,GDPD3,HIRIP3,KCTD13,MAPK3,MAZ,,MVPPPP4C,PRRT2, QPRT,SEZ6L2,SPN,TAOK2,TBX6,TMEM219,YPEL3
Sample16	16	81445132	81491808	46676	14	del	1	CDH13
Sample42	16	86921984	87097635	175651	15	del	2	ZFPM1,ZNF469
Sample18	17	14041754	15390352	1348598	239	del	8	CDRT15,CDRT4,COX10,FAM18B2,FLJ45831,HS3ST3B1,PMP22,TEKT3
Sample34	17	31889664	33323543	1433879	118	dupl	16	AATF,ACACA,AP1GBP1,C17orf78,DDX52,DUSP14,GGNBP2,HNF1B, LHX1,LOC284100,MGC4172,MRM1,MYO19,PIGW,TADA2LZNHIT3
Sample70	17	53737476	53759348	21872	11	dupl	1	BZRAP1
Sample28	18	640968	923680	282712	41	dupl	5	ADCYAP1,C18orf56,ENOSF1,TYMS,YES1
Sample51	18	866096	900635	34539	10	dupl	1	ADCYAP1
Sample47	18	7084612	7387011	302399	48	del	2	LAMA1,LRRC30
Sample30	18	74985521	75321604	336083	43	dupl	2	ATP9B,NFATC1
Sample60	20	14874333	15177947	303614	60	del	1	MACROD2
Sample13	21	21763517	21833902	70385	23	del	1	NCAM2

Table S11: De novo CNVs identified in the Icelandic control sample.

* According to Figure S19, this CNV intersects one isoforms of the DLG2 gene on chr11:82843701-85015962. However according to the list of RefSeq genes downloaded from NCBI (according to build 36.3, see Section 10), which we used for the main statistical analysis presented in Table 2 of the main text, this CNV does not intersect the gene, as it has the coordinates chr11:82843701-84312113. Although a pre-requisite of unbiased genome-wide gene-set analyses is that they must be based upon systematic *a priori* definitions of gene boundaries in which the annotation of each gene is made independent of examining the CNV data, in order to find out how the inclusion of this gene to this CNV would affect our analysis, we have re-run the statistical analysis after including the extra DLG2 hit in a *de novo* CNV among the Icelandic control individuals. This makes only a slight change to our p-values for the comparison with the Icelandic control *de novos* in Table 2, but the ARC and NMDAR remain significant: ARC ($p=1.44x10^{-3}$) and NMDAR ($p=0.02$), while PSD-95 is no longer significant ($p=0.0575$).

17) Supplementary references:

1. 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**: 589-593.
2. International Schizophrenia Consortium (ISC). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008; **455**: 237-241.
3. Stefansson H, Rujescu D, Cichon S, Pietilainen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008; **455**: 232-236.
4. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet*. 2010; **376**: 1401-1408.
5. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, et al. Support for the involvement of large cnvs in the pathogenesis of schizophrenia. *Hum Mol Genet.* 2009; **18**: 1497-1503.
6. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; **447**: 661-678.
7. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, et al. Copy Number Variants in Schizophrenia: Confirmation of Five Previous Findings and New Evidence for 3q29 Microdeletions and VIPR2 Duplications. *Am J Psychiatry.* 2011; **168**: 302-316.
8. Ikeda M, Aleksic B, Kirov G, Kinoshita Y, Yamanouchi Y, Kitajima T, et al. Copy number variation in schizophrenia in the Japanese population. *Biol Psychiatry*. 2010; **67**: 283-286.
9. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007; **17**: 1665-1674.
10. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009; **41**: 1223-1227.
11. Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, et al. De novo rates and selection of large copy number variation. *Genome Res*. 2010; **20**: 1469-1481.
12. Hehir-Kwa JY, Rodriuez-Santiago B, Vissers LE, de Leeuq N, Pfund R, Perez-Jurado LA, et al. De novo copy number variations are mostly paternal in origin. 2011: Wellcome Trust Scientific Conferences. Genomic Disorders 2011: The Genomics of Rare Diseases. 23-26 March
13. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, et al. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet*. 2008; **40**: 90-95.
14. Klopocki E, Schulze H, Strauß G, Ott C-E, Hall J, Trotier F, et al. Complex Inheritance Pattern Resembling Autosomal Recessive Inheritance Involving a Microdeletion in Thrombocytopenia-Absent Radius Syndrome. *Am J Hum Genet*. 2007; **80**: 232-240.
15. Toriello HV. Thrombocytopenia Absent Radius Syndrome. In: Pagon RA, Bird TC, Dolan CR, Stephens K, editors. *GeneReviews* [Internet]. 2009: Seattle (WA): University of Washington, Seattle; 1993-2009 Dec 8.
16. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*. 2008; **40**: 1466-1471.
17. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011; **43**: 838-846.
18. Wang K-S, Liu X-F, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr Res. 2010; **124**: 192-199.

19. Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, et al. Microdeletions of 3q29 Confer High Risk for Schizophrenia. *Am J Hum Genet*. 2010; **87**: 229-236.
20. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008; **320**: 539-543.
21. Magri C, Sacchetti E, Traversa M, Valsecchi P, Gardella R, Bonvicini C, et al. New Copy Number Variations in Schizophrenia. *PLoS ONE*. 2010; **5**: e13422.
22. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, et al. Copy Number Variants in Schizophrenia: Confirmation of Five Previous Findings and New Evidence for 3q29 Microdeletions and VIPR2 Duplications. *Am J Psychiatry*. 2011: **168**:302-316.
23. Ballif B, Theisen A, Coppinger J, Gowans G, Hersh J, Madan-Khetarpal S, et al. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogenet*. 2008; **1**: 8.
24. Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, et al. 3q29 Microdeletion Syndrome: Clinical and Molecular Characterization of a New Syndrome. *Am J Hum Genet*. 2005; **77**: 154-160.
25. Raymond FL, Tarpey P. The genetics of mental retardation. *Hum Mol Genet*. 2006; **15**(suppl 2): R110-R6.
26. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010; **466**: 368-372.
27. Sanders Stephan J, Ercan-Sencicek AG, Hus V, Luo R, Murtha Michael T, Moreno-De-Luca D, et al. Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron*. 2011; **70**: 863-885.
28. Tassabehji M. Williams-Beuren syndrome: a challenge for genotype phenotype correlations. *Hum Mol Genet*. 2003; **12**(suppl 2): R229-R37.
29. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, et al. Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur J Med Genet*. 2009; **52**: 94-100.
30. Ramocki MB, Bartnik M, Szafranski P, Kolodziejska KE, Xia Z, Bravo J, et al. Recurrent Distal 7q11.23 Deletion Including HIP1 and YWHAG Identified in Patients with Intellectual Disabilities, Epilepsy, and Neurobehavioral Problems. *Am J Hum Genet*. 2010; **87**: 857-865.
31. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011 Sep 18. doi: 10.1038/ng.940. [Epub ahead of print].
32. Ortiz AN, Oien DB, Moskovitz J, Johnson MA. Quantification of reserve pool dopamine in methionine sulfoxide reductase A null mice. *Neuroscience*. 2011; **177**: 223-229.
33. Abi-Dargham A, Grace AG. Dopamine and schizophrenia. Chapter 20. In *Schizophrenia*, Eds: Weinberger, DR and Harrison, PJ. Blackwell Publishing Ltd. 2011. (Blackwell).
34. Stewart DR, Kleefstra T. The chromosome 9q subtelomere deletion syndrome. *Am J Hum Genet C Semin Med Genet*. 2007; **145C**(4): 383-92.
35. Kleefstra T, van Zelst-Stams WA, Nillesen WM, Cormier-Daire V, Houge G, Foulds N, et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. *J Med Genet*. 2009; **46**: 598-606.

36. Perin MS. Mirror Image Motifs Mediate the Interaction of the COOH Terminus of Multiple Synaptotagmins with the Neurexins and Calmodulin. *Biochemistry*. 1996; **35**: 13808-13816.
37. Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) Deletions in Schizophrenia. *Schizophr Bull*. 2009; **35**: 851-854.
38. Rees E, Moskvina V, Owen MJ, O'Donovan MC, Kirov G. De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiat. Biol Psychiatry*. 2011 Aug 18. [Epub ahead of print]
39. de Kovel CGF, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies *Brain*. 2010; **133**: 23-32.
40. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*. 2008; **17**: 458-465.
41. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008; **40**: 880-885.
42. Bayés A, van de Lagemaat LN, Collins MO, Croning MDR, Whittle IR, Choudhary JS, et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci*. 2011; **14**: 19-21.
43. Fernandez E, Collins MO, Uren RT, Kopanitsa MV, Komiyama NH, Croning MDR, et al. Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol Syst Biol*. 2009; **5**:269. Epub 2009 May 19.
44. Pocklington AJ, Cumiskey M, Armstrong JD, Grant SGN. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol Syst Biol*. 2006; **2:**2006.0023. Epub 2006 Jan 17.
45. Farr CD, Gafken PR, Norbeck AD, Doneanu CE, Stapels MD, Barofsky DF, et al. Proteomic analysis of native metabotropic glutamate receptor 5 protein complexes reveals novel molecular constituents. *J Neurochem*. 2004; **91**: 438-450.
46. Takamori S, Holt M, Stenius K, Lemke EA, Grønborg M, Riedel D, et al. Molecular Anatomy of a Trafficking Organelle. *Cell*. 2006; **127**: 831-846.
47. Morciano M, Beckhaus T, Karas M, Zimmermann H, Volknandt W. The proteome of the presynaptic active zone: from docked synaptic vesicles to adhesion molecules and maxi-channels. *J Neurochem*. 2009; **108**: 662-675.
48. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, et al. A Mammalian Organelle Map by Protein Correlation Profiling. *Cell*. 2006; **125**: 187-199.
49. Raychaudhuri S, Korn JM, McCarroll SA, Consortium TIS, Altshuler D, Sklar P, et al. Accurately Assessing the Risk of Schizophrenia Conferred by Rare Copy-Number Variation Affecting Genes with Brain Function. *PLoS Genet*. 2010; **6**(9): pii: e1001097.