Nature. Author manuscript; available in PMC 2014 November 19.

Published in final edited form as:
Nature. 2014 February 13; 506(7487): 179–184. doi:10.1038/nature12929.

De novo mutations in schizophrenia implicate synaptic networks

Menachem Fromer[1,4], **Andrew J. Pocklington**[2], **David H. Kavanagh**[2], **Hywel J. Williams**[2], **Sarah Dwyer**[2], **Padhraig Gormley**[3,6], **Lyudmila Georgieva**[2], **Elliott Rees**[2], **Priit Palta**[3,5,9], **Douglas M. Ruderfer**[1,2], **Noa Carrera**[2], **Isla Humphreys**[2], **Jessica S. Johnson**[1], **Panos Roussos**[1], **Douglas D. Barker**[4], **Eric Banks**[6], **Vihra Milanova**[7], **Seth G. Grant**[8], **Eilis Hannon**[2], **Samuel A. Rose**[4], **Kimberly Chambert**[4], **Milind Mahajan**[1], **Edward M. Scolnick**[4], **Jennifer L. Moran**[4], **George Kirov**[2], **Aarno Palotie**[3,6,9], **Steven A. McCarroll**[4,6,10], **Peter Holmans**[2], **Pamela Sklar**[1,11], **Michael J. Owen**[2,*], **Shaun M. Purcell**[1,4,12], and **Michael C. O'Donovan**[2]

[1]Division of Psychiatric Genomics in the Department of Psychiatry, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA [2]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK [4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA [5]Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia [6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA [7]Department of Psychiatry, Medical University, Sofia 1431, Bulgaria [8]Centre for Neuroregeneration, University of Edinburgh, EH16 4SB [9]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland [10]Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA [11]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA [12]Analytic and Translational Genetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114, USA

Summary

Inherited alleles account for most of the genetic risk for schizophrenia. However, new (*de novo*) mutations, in the form of large chromosomal copy number changes, occur in a small fraction of cases and disproportionally disrupt genes encoding postsynaptic proteins. Here, we show that small *de novo* mutations, affecting one or a few nucleotides, are overrepresented among glutamatergic postsynaptic proteins comprising activity-regulated cytoskeleton-associated protein

[*]Corresponding author: Michael J Owen, owenmj@cardiff.ac.uk.

Author Contributions The project was led in Cardiff by MCOD & MJO, in Mount Sinai by SMP & PS, at the Broad by SAMC & JLM, and at the Sanger by AP. HJW, JLM, KC, JSJ, DDB, MM & SAR were responsible for sample processing and data management. MF, HJW, PG, DMR, DHK, GK, ER & SD processed NGS data, annotated and validated mutations. LG, NC, IH, SD, HJW & SAR undertook validation of mutations and additional lab work. AJP, MF, DHK, SMP & PH co-ordinated/undertook the main bioinformatics/statistical analyses. ER, DMR, EB, PP, EH & PR performed additional analyses. SGG contributed additional insights into synaptic biology. Sample recruitment was led by GK and VM. The main findings were interpreted by MOD, MF, MJO, PH, GK, EMS, SAMC, DHK, AJP, AP, SMP, & PS who drafted the manuscript.

Author Information Data included in this manuscript have been deposited at dbGaP under accession number phs000687.v1.p1 and is available for download at http://www.ncbi.nlm.nih.gov/projects/gap/cgibin/study.cgi?study_id=phs000687.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

(ARC) and N-methyl-D-aspartate receptor (NMDAR) complexes. Mutations are additionally enriched in proteins that interact with these complexes to modulate synaptic strength, namely proteins regulating actin filament dynamics and those whose mRNAs are targets of fragile X mental retardation protein (FMRP). Genes affected by mutations in schizophrenia overlap those mutated in autism and intellectual disability, as do mutation-enriched synaptic pathways. Aligning our findings with a parallel case-control study, we demonstrate reproducible insights into aetiological mechanisms for schizophrenia and reveal pathophysiology shared with other neurodevelopmental disorders.

Schizophrenia is a disorder whose pathophysiology is largely unknown. It has a lifetime risk of about 1%, is frequently chronic and socially disabling, and is associated with an average reduction in lifespan of about 25 years. High heritability points to a major role for transmitted genetic variants[1]. However, it is also associated with a marked reduction in fecundity[2], leading to the hypothesis that alleles with large effects on risk might often occur *de novo* (mutations present in affected individual but not in either parent) to balance their elimination from the population by selection[3].

Of the known risk alleles for schizophrenia, the only ones definitively shown to confer considerable increments in risk are rare chromosomal copy number variants (CNVs)[1,4], which involve deletion or duplication of thousands of bases of DNA. As predicted by schizophrenia's association with decreased fecundity, these CNVs often occur *de novo* in the small proportion of cases in which they are found[5]. Exome sequencing technology now allows systematic scans of genes for *de novo* mutations at single-base rather than kilobase resolution. This approach has already implicated *de novo* loss-of-function (LoF) mutations in disorders in which, as in schizophrenia, *de novo* CNVs play a role, including autism spectrum disorder (ASD)[6–9] and intellectual disability (ID)[10,11]. In schizophrenia, the results from exome sequencing[12–14] do not yet support definitive conclusions, likely due to limited sample sizes.

We report the largest exome sequencing study of *de novo* mutations in schizophrenia to date, based upon genomic (blood) DNA from 623 schizophrenia trios. The primary aims were fourfold (Table 1 a-d). The first two aims were to establish a general case for the relevance of *de novo* mutations in schizophrenia by determining if *de novo* mutations affecting protein sequences either occur in schizophrenia at higher than expected rates (Table 1a) or are enriched among sets of genes implicated in the disorder through other approaches (Table 1b). The remaining two aims, the main motivation for the study, were to determine whether *de novo* mutations implicate specific pathogenic biological processes in schizophrenia (Table 1c) and to investigate the relationship between schizophrenia and other neurodevelopmental disorders (Table 1d). To test for reproducibility, and ensure robustness of the findings to study design, we shared our findings with an independent case-control exome sequencing study[15].

De novo mutation rates

We generated sequence data for a median of 93% of targeted exome bases at a depth of >10 reads, from which we generated putative *de novo* calls (ED Figures 1 and 2; Supplementary

Text). Using Sanger sequencing, we validated 637 *de novo* coding or canonical splice site variants (Table S1) in 617 probands (6 trios were excluded after QC), a rate of 1.032 mutations per trio. These comprised 482 nonsynonymous mutations, of which 64 were LoF (nonsense, splice, and frameshift). The remaining 155 mutations were silent and were therefore excluded from enrichment analyses.

The exome point mutation rate in schizophrenia was, adjusting for target coverage, 1.61×10^{-8} per base per generation, compatible with the population expectation of 1.64×10^{-8} (Supplementary Text). The mutation rate (corrected for experimental confounders, Supplementary Text) was associated with increasing paternal (p=0.005) and maternal (p=0.0003) age at proband birth. Given the high correlation between the two, we could not confidently distinguish independent parental age effects (Supplementary Text). As expected[16], most *de novo* mutations (79%) we could phase occurred on paternal chromosomes (Supplementary Text). The number of *de novos* per individual followed a Poisson distribution (ED Figure 3a) in line with previous studies of autism[6] and schizophrenia[13]. Nevertheless, LoF *de novo* mutations were more common in patients with relatively poor school performance (p=0.018; ED Figure 3b), but none of the other variables tested – family history, age at onset, gender, or having a *de novo* CNV – were significantly associated with mutation rates.

Compared with 731 controls from published datasets (Table S2), probands did not have a significant elevation in the relative rates of nonsynonymous to silent mutations, or LoF to missense mutations (Table 1a, Table 2). No differences were observed between schizophrenia cases with or without *de novo* CNVs or between those stratified by common allele risk scores (ED Table 1a). Consistent with their higher LoF mutation rate, those with school grades below the median had significantly elevated LoF:missense ratios compared to both controls (p=0.02) and cases with higher school grades (p=0.0095) (ED Table 1b, ED Figure 3b). In the absence of an effect of age at onset (that might affect school performance), this suggests LoF mutations occur preferentially in (the large proportion of) schizophrenia cases that have premorbid cognitive impairment[17]. All probands attended and graduated from mainstream schools, which excluded people with significant degrees of ID; moreover, recruiting psychiatrists were explicitly instructed to exclude people with known ID. Thus, the enrichment of LoF mutations in those with the poorest scholastic attainment cannot be attributed to the inclusion of individuals with severe ID, although this does not preclude the presence of individuals with mild ID among cases with low educational achievement.

Mutations in schizophrenia gene sets

Gene sets selected for independent evidence for relevance to schizophrenia showed enrichment ($p_{corrected}$=0.0007) of nonsynonymous *de novos* (Table 1b), indicating that a proportion of mutations are pathogenic for schizophrenia. Specifically, genes were recurrent for *de novos* more than expected (Table 1b, ED Table 2). Genes affected by nonsynonymous *de novo* mutations were also enriched for inherited rare risk alleles (Table 1b), with excess transmission of rare nonsynonymous alleles from parents to the affected probands, as well as enrichment in cases of rare (MAF < 0.001) gene-disruptive mutations in an independent

case-control exome sequencing study[15]. One gene, *TAF13*, encoding a subunit of the TFIID transcription initiation complex, contains two rare LoF mutations. This recurrence is significant even after genome-wide correction ($p=1\times10^{-6}$; $p_{corrected}=0.024$) (ED Table 2). Replication is necessary to firmly establish this as a susceptibility gene.

Mutations enriched in synaptic genes

Previous studies have suggested that CNVs in people with schizophrenia preferentially affect broadly-defined sets of synaptic genes[18,19]. Moreover, a detailed analysis of *de novo* CNVs based on gene sets constructed from experimental proteomics led us to propose that this synaptic enrichment could be explained by mutations affecting proteins closely associated with the N-methyl-D-aspartate (NMDA) receptor, which we refer to as the NMDAR complex, and proteins that interact with ARC (activity-regulated cytoskeleton-associated protein), referred to as the ARC complex[20]. Our primary functional hypothesis in the present study was that genes encoding proteins in the ARC and NMDAR complexes would be disproportionately impacted by *de novo* SNV and indel mutations. We additionally postulated that brain-expressed genes that are repressed by fragile X mental retardation protein (FMRP)[21] would also be enriched for *de novo* mutations because these have been shown to be enriched for *de novo* mutations in ASD[9]. Moreover, FMRP targets include multiple members of the NMDAR and ARC complexes.

We observed experiment-wide significant enrichment for nonsynonymous mutations among the synaptic gene sets (Table 1c), as well as specifically for NMDAR and ARC complexes (Tables 1c and 3, ED Figure 4). NMDAR and ARC complexes are closely associated elements central to regulating synaptic strength at glutamatergic synapses and have been implicated in cognition. NMDA signaling triggers multiple processes required for inducing synaptic plasticity[22], while ARC is involved in almost all known forms of synaptic plasticity including synaptic remodeling, the consolidation of changes in synaptic strength linked to memory and response to stress[23–25], and regulating synapse elimination during development[26], a process believed to be aberrant in schizophrenia[27].

FMRP targets were also enriched for nonsynonymous *de novo* mutations (Table 1c), even after NMDAR, ARC, and the broader group of postsynaptic density (PSD) genes were removed (ED Table 3). Given that loss of FMRP results in widespread deficits in synaptic plasticity[28], these findings again implicate pathogenic disruption of plasticity mechanisms in schizophrenia. Secondary analyses to dissect the FMRP enrichment by subdividing genes by gene ontology (GO)[29] membership did not identify significant categories.

Support for the candidate hypotheses were replicable and robust to study design. In the schizophrenia case-control study[15], rare (MAF < 0.001) LoF mutations were enriched in NMDAR ($p=0.02$), ARC ($p=1\times10^{-3}$), and FMRP target ($p=0.003$) sets. Across studies, LoF enrichments in the ARC complex were particularly striking -- 17 fold here (Table 3, ED Figure 4f) and 19 fold in the case-control study -- suggesting that disruption of ARC function has particularly strong effects on disease risk.

Aiming to identify hitherto unsuspected disease mechanisms, we undertook an hypothesis-free analysis based on the comprehensive GO annotations[29]. A single category (GO:

0051017) was significantly enriched for nonsynonymous *de novo* mutations ($p=6.6\times10^{-6}$) after correction for all GO categories ($p_{corrected}=0.032$). Genes in GO:0051017, assembly of actin filament bundles, are intimately involved in synaptic plasticity, and are functionally interconnected with ARC and NMDAR signalling (see Supplementary Text). Even after removal of genes overlapping with ARC/NMDAR sets, GO:0051017 remained 8 fold enriched for mutations (p=0.0011). Although not significant in the case-control dataset[15], this category was significantly enriched for *de novo* CNVs in a study of ASD[30]. It also includes *KCTD13*, the gene responsible for some of the phenotypes associated with CNVs at 16p11.2[31], duplication of which is a risk factor for schizophrenia[4]. *KCTD13* also maps to a schizophrenia genome-wide significant SNP locus[32].

Connectivity of mutated synaptic genes

Seeking further insights into synaptic pathology, we identified interactions involving proteins with *de novo* mutations using a synaptic interactome database[33] (Supplementary Text). Proteins with nonsynonymous *de novos* had more connectivity than expected among each other (Figure 1a) and with synaptic proteins in general, suggesting a greater than average importance to the synapse. Directly interacting proteins with *de novos* are involved in multiple processes regulating synaptic plasticity, particularly NMDA, AMPA, and kainate receptor trafficking, and the regulation of actin dynamics. These interactions involve genes not present in our pre-assigned NMDAR/ARC and actin filament complexes (Supplementary Text). Though our analyses highlighted postsynaptic processes, some of the interacting synaptic proteins with de novo mutations are presynaptic (Figure 1a, Supplementary Text, and ED Figure 4a). Pre- and postsynaptic proteins are, however, closely functionally related; indeed, trans-synaptic effects of presynaptic proteins on the regulation of AMPA receptor trafficking and NMDAR-dependent plasticity have recently been described[34].

We were unable to replicate a previous report of prenatal bias in brain expression for genes with schizophrenia *de novos*[13] using microarray or RNA-seq data (Supplementary Text, ED Table 4).

Overlaps between disorders

CNV loci associated with schizophrenia overlap with those seen in ASD, ID, and ADHD[1,4,35]. However, since pathogenic CNVs typically span multiple genes and are concentrated in a relatively small fraction of the genome[36], it is possible that this may not indicate cross-disorder effects at the level of specific genes. Therefore, we sought evidence for shared genetic aetiology between schizophrenia and both ID and ASD[37] by testing for overlap of genes affected by *de novo* mutations in schizophrenia, ASD, and ID.

Genes with *de novo* mutations in the current study overlapped those affected by *de novo* mutations in ASD[6–9] and ID[10,11] (Figure 1b, Table 1d, Table 4) but not controls (ED Table 5). Moreover, LoF mutations in schizophrenia were enriched even in the small subset of genes (N=7) with *recurrent* LoF *de novos* in ASD (p=0.0018) or ID (p=0.019), the mutations occurring in *SCN2A* (encoding an alpha subunit of voltage-gated sodium channels, a major mediator of neuronal firing and action potential propagation) and *POGZ*

(whose involvement in mitosis suggests a possible role in regulating neuronal proliferation[38]). *SCN2A* and *POGZ* are both now established ASD genes[39]. Other notable genes affected by LoF mutations in the present study for which there is prior support for LoF mutations in other neurodevelopmental disorders include *DLG2* and *SHANK1* (Supplementary Text). Thus, we now show overlap between schizophrenia, ASD, and ID at the resolution not just of loci or even individual genes, but even of mutations with similar functional (LoF) impacts.

Further pointing to shared disease mechanisms, ARC/NMDAR complexes (Table 3) and FMRP targets (ED Table 3) were enriched for *de novo* mutations in ID, and NMDAR and FMRP targets were also enriched in ASD. However, we also find differences between the disorders. In general, enrichment statistics were stronger for ASD and ID than schizophrenia, particularly for LoF mutations (Table 2), despite the relatively small number of ID trios. Genes and mutation sites were most highly conserved in ID, then ASD, with schizophrenia least conserved (Supplementary Text, ED Table 6). These findings suggest highly disruptive mutations play a relatively lesser role in schizophrenia, and also that the disorders differ by severity of functional impairment, consistent with the hypothesis of an underlying dimension of neurodevelopmental pathology[40] indexed by cognitive impairment, with ID at one extreme.

That the most damaging mutations reflect a gradient of neurodevelopmental impairment is further supported by the observation that, within schizophrenia, the highest rate of LoF mutations (ED Figure 3b) occurred in individuals likely to have the greatest cognitive impairment (lowest scholastic attainment), as does the observation that the LoF genic overlap between schizophrenia and both autism and ID is dependent on the *de novo* mutations (including *SCN2A* and *POGZ*) in those individuals (ED Table 1c). However, as noted above, the enrichment of LoF mutations in those with the poorest scholastic attainment cannot be attributed to the inclusion of individuals with severe ID. Moreover, when we exclude cases with low scholastic attainment, we still see significant enrichment of the synaptic pathways that are enriched in the full sample (Supplementary Text, ED Table 1c). Thus, our implication of synaptic protein complexes is not dependent on mutations present in a subset of cases with severely impaired cognitive function.

Discussion

In the largest exome-sequencing-based study of *de novo* mutations in schizophrenia, we demonstrate a convergence of *de novo* mutations on multiply defined sets of functionally related proteins, pointing to the regulation of plasticity at glutamatergic synapses as a pathogenic mechanism in schizophrenia. How disruption of these synaptic mechanisms impacts brain function to produce psychopathology cannot be answered by genetic studies alone, but our identification of *de novo* mutations in these gene sets provides the basis to address this. Our findings of overlaps between the pathogenic mechanisms underlying schizophrenia and those in autism and ID lend support to recent, controversial, suggestions that our understanding of these disorders might better be advanced by research that integrates findings across multiple disorders and places more emphasis on domains of

psychopathology, e.g., cognition, and their neurobiological substrates rather than current diagnostic categories[40,41].

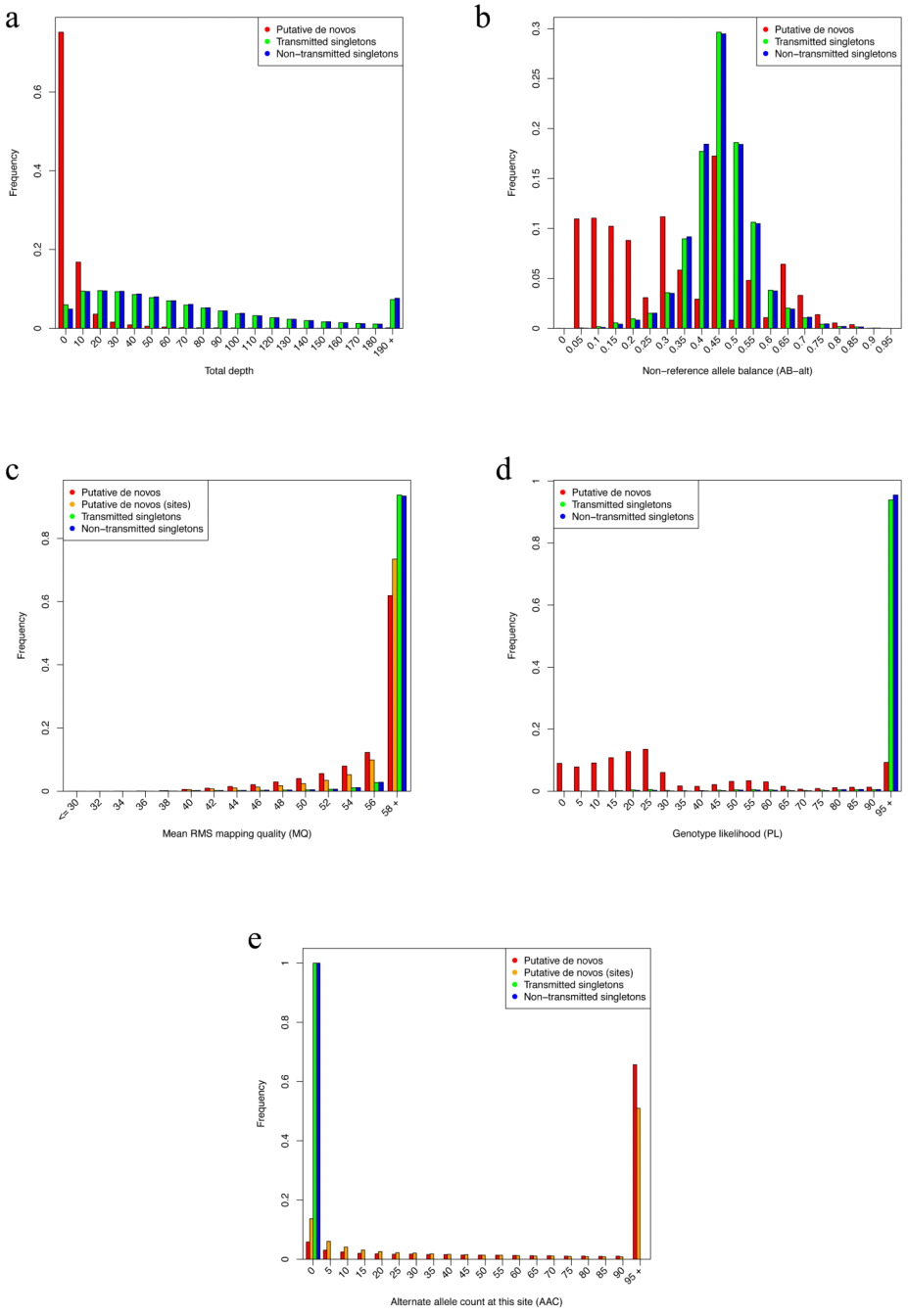
METHODS SUMMARY

Parent proband trios (N=623), where the proband had a history of hospitalization for schizophrenia or schizoaffective disorder, were recruited from psychiatric hospitals in Bulgaria. Probands attended mainstream schools which excluded people with ID (intellectual disability); all graduated with a pass. Exome DNA was captured from genomic DNA (whole blood), using either Agilent or Nimblegen array-based capture, and subjected to paired-end sequencing on Illumina HiSeq sequencers. The BWA/Picard/GATK pipeline was used for sequence alignment and variant calling. Putative *de novo* mutations were identified using Plink/Seq (http://atgu.mgh.harvard.edu/plinkseq) and were validated using Sanger sequencing. We used Plink/Seq to annotate mutations according to RefSeq gene transcripts (UCSC Genome Browser, http://genome.ucsc.edu). Mutation rate was tested for association with clinical and other covariates using a generalized linear model. Rates of functional classes of mutations in probands were compared with those in published controls using Fisher's exact test. Mutations were tested for recurrence, enrichment in candidate gene sets, and enrichment in genes affected by *de novo* mutations in previous studies using *dnenrich* (Supplementary Text). *Dnenrich* calculates one-sided p-values under a binomial model of greater than expected hits using randomly placed mutations accounting for gene size, sequencing coverage, tri-nucleotide contexts, and functional effects of the observed mutations. Candidate gene sets and studies of neuropsychiatric disease are described in the main and Supplementary Text. Primary hypotheses (Table 1) were Bonferroni corrected for multiple testing.

Full Methods and associated references are available in the Supplementary Text.

Europe PMC Funders Author Manuscripts

Extended Data

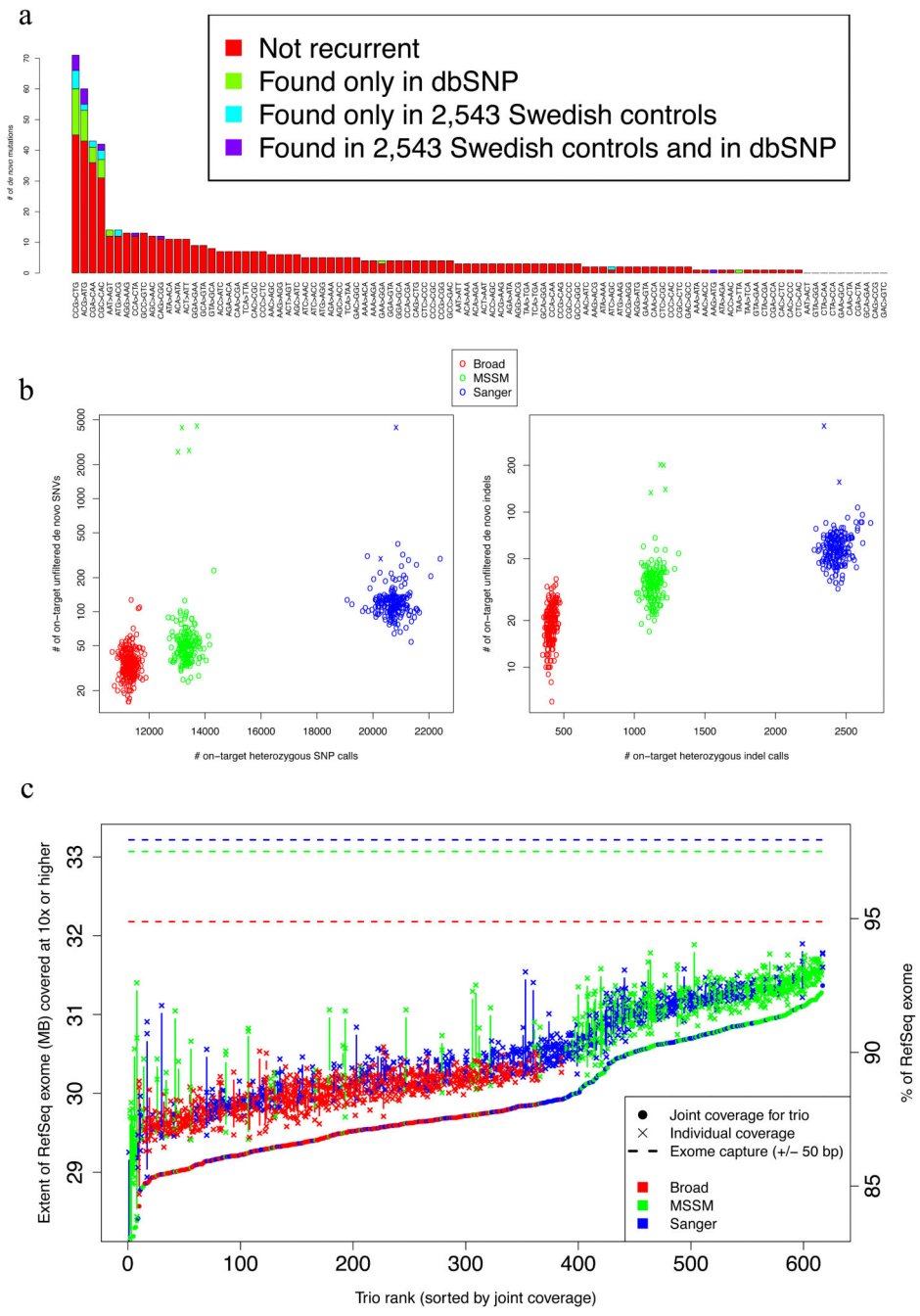


Extended Data Figure 1. Comparison of sequencing metrics for putative *de novo* calls and parental singletons

Putative *de novo* calls (child heterozygous, both parents homozygous reference) were compared with variants observed in only a single parent ("singletons"), in terms of (a) depth of all reads at the variant site [DP = depth], (b) fraction of reads with the alternate allele [AB = allele balance], (c) mapping quality of the reads at the site [MQ], (d) the likelihood of the heterozygous genotype [PL = Phred-scaled likelihood], and (e) the number of other samples

Europe PMC Funders Author Manuscripts

in the present study with a non-reference allele at that site [AAC = alternate allele count]. Distributions were calculated for putative *de novo* variants (red), or grouped by sites of putatively recurrent *de novos* (orange) when relevant, transmitted singletons (green), and non-transmitted singletons (blue).



Extended Data Figure 2. Metrics for *de novo* variants across cohorts and trios

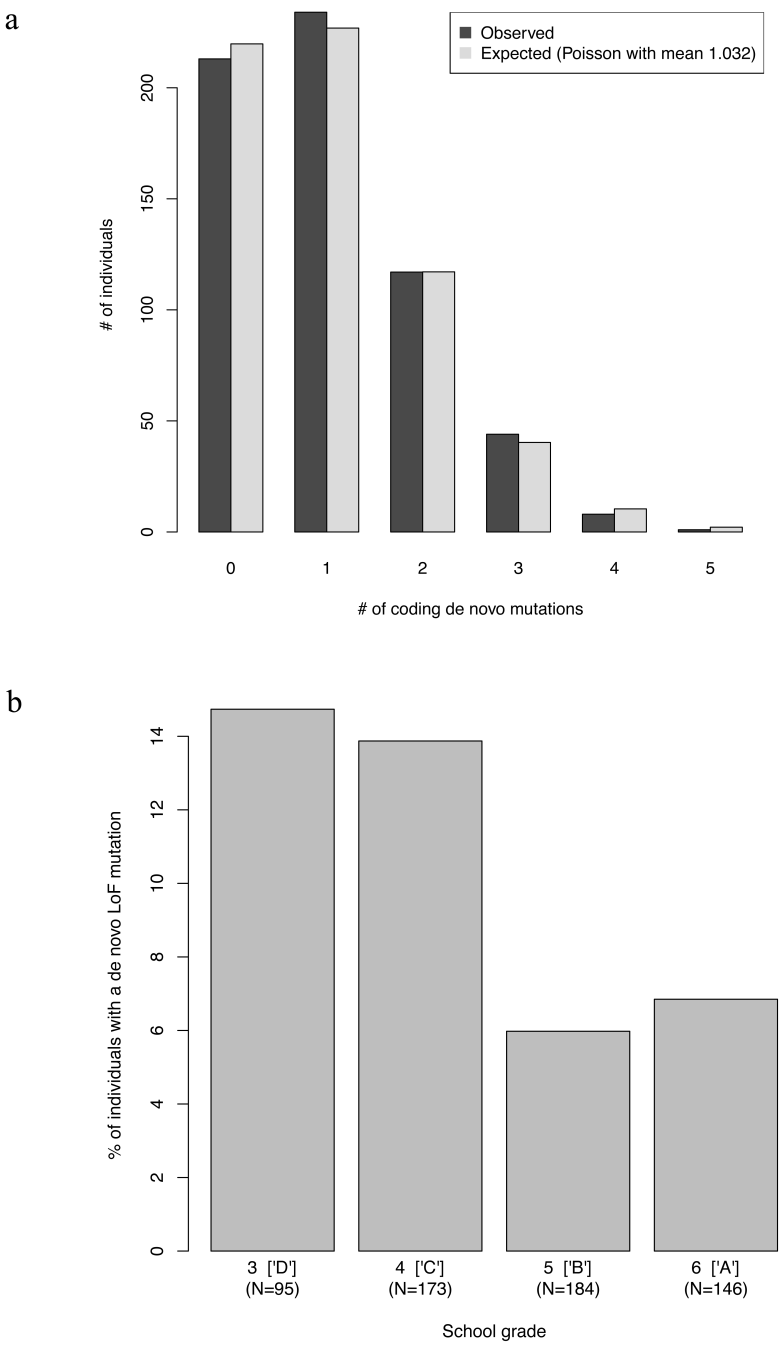
a. Rates of recurrence of *de novo* mutations for tri-nucleotide sequences. For each of 96 possible tri-nucleotide base contexts of single-base mutations (accounting for strand

symmetry by reverse complementarity), the number of observed *de novo* SNV is plotted (sorted by this count). Mutation counts are sub-divided into those not found in external data (red), those found in dbSNP (build 137, green), those found in controls in the parallel exome sequencing study[15] (cyan), and those found both in dbSNP and that study (purple)

b. Comparison of on-target heterozygous SNV and indel call rate with putative *de novo* mutation calls. For each proband, the number of heterozygous SNV and indel calls is compared with the number of putative *de novo* mutations (child heterozygous, both parents homozygous reference). Probands are colored by sequencing center (see Supplementary Text for differences in exome capture), and 6 trios are noticeable outliers from all others in terms of number of putative *de novos*.

c. Variation in sequencing coverage between and across trios and sequencing centers. For each trio, the number of bases covered by 10 reads or more for each member (marked by 'x') and the joint coverage[9] in all 3 members (marked by points) are plotted at corresponding horizontal points; trios are sorted in increasing order of joint coverage and colored by sequencing center (see Supplementary Text). The intersection of each exome capture with the RefSeq coding sequence is marked by respective dotted lines.

Europe PMC Funders Author Manuscripts



Extended Data Figure 3. *De novo* mutation counts and rates

a. The observed distribution of number of validated RefSeq-coding (see Supplementary Text) *de novo* mutations found for each trio (N=617) is compared with that expected from a Poisson distribution with a rate equal to the observed mean number of *de novos* (λ=1.032).

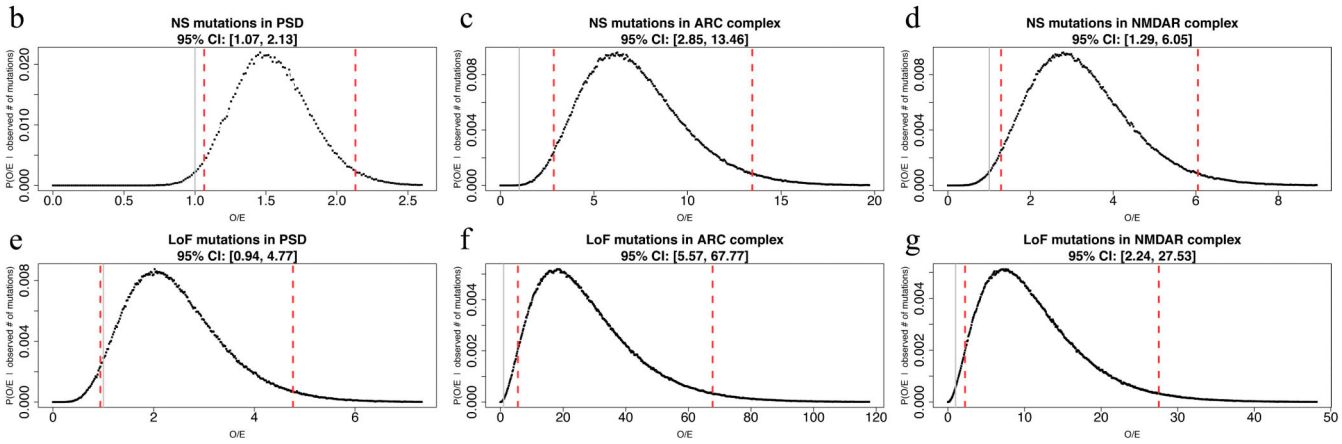
b. Deleterious mutation rate inversely correlates with academic performance. Individuals were grouped according to their final school grade (3-6, corresponding to D, C, B, A in the US system, http://www.fulbright.bg/en/p-Educational-System-of-Bulgaria-18/), and the proportion of individuals with one or more *de novo* loss-of-function (LoF) mutations is

plotted. See Supplementary Text for details on linear regression performed to evaluate association; note that 19 samples were removed from this analysis for missing parental age or school grade information, leaving a total of 598 trios.

a

Gene set	# genes	Mut type	# mut hitting set	p-value	Genes hit (counts)	*de novo* *CNV p-value (Kirov et al., 2012)	*de novo* CNV genes	Case-control CNVs p-value (Kirov et al., 2012)
PSD	681	NS	34	**0.019**	ACTN1, ANK1, BAIAP2(x2), BRSK1, CAPN5, **DLG1**, **DLG2**, EPB41, EPB41L1, FARSA, GIT1, GNB2, HSP90AA1, HSPA8(x2), ITSN1, KIF1A, MYH11, MYH9, MYO18A, NCKIPSD, NFASC, NRXN1, PLXNA1, PTK2B, RIMS1, SHANK1, SLC25A12, **SND1**, SORBS2, SRCIN1, UNC13A, YWHAZ	**4.50E-02**	ALDOA, CYFIP1, **DLG1**, **DLG2**, DLGAP1, HSPB1, MAPK3, MDH2, RPH3A, RYR2, **SND1**, STX1A, TAOK2, TJP1, YWHAG	-
		LoF	6	0.091	**DLG2**, HSP90AA1, HSPA8, ITSN1, NCKIPSD, SHANK1			
ARC complex	28	NS	6	**0.0005**	BAIAP2(x2), **DLG1**, **DLG2**, HSPA8(x2)	**2.51E-04**	CYFIP1, **DLG1**, **DLG2**, DLGAP1	0.14
		LoF	2	**0.005**	**DLG2**, HSPA8			
NMDAR complex	60	NS	6	**0.025**	**DLG1**, **DLG2**, GNB2, PTK2B, SHANK1, YWHAZ	**6.30E-03**	**DLG1**, **DLG2**, DLGAP1, MAPK3, STX1A, TJP1, YWHAG	**1.50E-03**
		LoF	2	**0.035**	**DLG2**, SHANK1			

* compared to control *de novo* CNVs



Extended Data Figure 4. Enrichment of *de novo* SNVs, indels, and CNVs in genes encoding postsynpatic complexes at glutamatergic synapses

a. Number of de novo mutations in postsynaptic complexes in current study (and genes affected) are shown alongside the most conservative estimate of *de novo* CNV enrichment from Kirov, et al.[20]. NS = nonsynonymous, LoF = loss-of-function. The NMDAR complex gene set was derived *a priori* from a published proteomics dataset[42]. To avoid investigator bias, we did not add additional members *post hoc*, thus omitting genes with *de novo* mutations and important NMDAR functions; these include *GRIN2A*, which encodes a subunit of the NMDA receptor itself, and *AKAP9* which directly anchors protein complexes involved in signalling at NMDA receptors[43]. $p<0.05$ are marked in bold.

b. - g. 95% credible intervals (CI) for fold-enrichment statistics of *de novo* mutations in postsynaptic gene sets (corresponding to enrichments in a. above, and as marked) were calculated from the posterior distributions of fold-enrichment (observed-to-expected = O/E) statistic values for individuals in this study. Point estimates of O/E are given in Table 3, and correspond to the distribution modes here. The 95% CI is marked by red vertical lines, and a null effect size (value of 1) is marked by a gray line. Note that LoF mutations in the large

Europe PMC Funders Author Manuscripts

PSD set are not significantly enriched, and thus the corresponding CI includes an effect size of 1. All posterior distributions were calculated using *dnenrich*, as described in the Supplementary Text.

Extended Data Table 1a

Stratification of *de novo* mutations based on polygenic burden, presence of a 'pathogenic' CNV, or poor scholastic achievement

Ratios of nonsynonymous to synonymous (NS:S) and loss-of-function to missense (LoF:missense) *de novo* mutations were compared (Fisher's exact test) between those found in individuals with a high polygenic score (top 50%) and those in the individuals in the bottom 50% of the polygenic score distribution (scores previously generated for this sample[44]). Individuals were additionally split based on the presence of a 'pathogenic' CNV. A 'pathogenic' CNV was defined as either a *de novo* CNV identified for these samples in Kirov, et al.[20], or a CNV associated with psychiatric disease[1]. P-values were computed using Fisher's exact test as in Table 2.

	Probands with top 50% of polygenic scores	Probands with bottom 50% of polygenic scores	Probands with top 50% of polygenic scores or a 'pathogenic' CNV	Probands with bottom 50% of polygenic scores and no 'pathogenic' CNV
NS	210	229	228	214
S	71	66	74	63
Ratio	2.96	3.47	3.08	3.4
p	0.43		0.63	
LoF	24	29	26	27
missense	182	196	198	183
Ratio	0.13	0.15	0.13	0.15
p	0.77		0.77	

Extended Data Table 1b

Ratios of NS:S and LoF:missense *de novo* mutations were compared (Fisher's exact test) between schizophrenia probands with poor scholastic performance and 1) controls; 2) probands with high scholastic performance. Nominally significant results ($p<0.05$) are marked in bold.

	Controls[7-10,13-14]	Probands with poor scholastic performance (school grades 3 or 4)	Probands with high scholastic performance (school grades 5 or 6)
NS	434	222	242
S	155	67	84
Ratio	2.8	3.3	2.9
p vs. poor scholastic performance	0.32	-	0.51

	Controls[7-10,13-14]	Probands with poor scholastic performance (school grades 3 or 4)	Probands with high scholastic performance (school grades 5 or 6)
LoF	49	40	23
missense	376	177	214
Ratio	0.13	0.23	0.11
p vs. poor scholastic performance	**0.021**	-	**0.0095**

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Extended Data Table 1c

Enrichment of *de novo* mutations (as calculated by *dnenrich*, see Supplementary Text) including all individuals, after excluding individuals with the lowest scholastic achievement (a school grade of 3), or excluding those with a 'pathogenic' CNV (see above) or a polygenic score in the top 50% of the distribution (see above). These secondary exclusion analyses were performed on those gene sets identified as significant in the analyses of the full set. p<0.05 are marked in bold.

		All probands				Exclude probands with lowest school grade (3)				Exclude probands with a 'pathogenic' CNV or with a polygenic score in the top 5% among probands			
		NS (482)		LoF (64)		NS (398)		LoF (49)		NS (423)		LoF (54)	
Gene set	**# of genes**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**
PSD	681	**0.019**	34	0.091	6	**0.0033**	32	**0.031**	6	**0.039**	29	**0.047**	6
ARC complex	28	**0.00048**	6	**0.005**	2	**0.0002**	6	**0.0036**	2	**0.0019**	5	**0.0048**	2
NMDAR complex	60	**0.025**	6	**0.035**	2	**0.036**	5	**0.02**	2	**0.045**	5	**0.026**	2
FMRP targets	784	**0.0094**	64	0.37	7	**0.0041**	56	0.28	6	**0.031**	54	0.55	5
actin filament bundle assembly	34	**6.57E-06**	8	1	0	**0.0023**	5	1	0	**0.0005**	6	1	0
autism LoF genes	128	**0.015**	11	**0.00072**	4	0.41	6	0.52	1	**0.025**	9	**0.0013**	3
ID LoF genes	30	0.27	1	**0.019**	1	1	0	1	0	0.22	1	**0.016**	1

Europe PMC Funders Author Manuscripts

Extended Data Table 2

Genes overlapped by two nonsynonymous *de novo* mutations in schizophrenia probands

Genes hit by nonsynonymous (NS) mutations in two different probands with schizophrenia (N=18) are listed, with the expected functional impact of those mutations and the nominal p-value for genic recurrence (calculated by *dnenrich*); for the single instance of two loss-of-function (LoF) alleles in a single gene (*TAF13*), the p-value for LoF recurrence is given in parentheses; this is bolded since it is significant after Bonferroni correction for multiple testing of all genes (see Supplementary Text). Also shown are the case/control counts from the parallel exome sequencing study[15] and the corresponding nominal p-value for association with schizophrenia.

18 genes with recurrent nonsynonymous de novos mutations (p = 0.0314)				
Gene	**De novo mutations**	**Nominal p-value for recurrence of NS (LoF) de novos**	**Case/control counts of rare (MAF < 0.001) LoF mutations in Purcell, et al.[15]**	**Nominal case/control p-value**
AKD1	frameshift, missense	0.0024	2/8	1
BAIAP2	codon-deletion, missense	0.00042	1/0	0.53
C7orf60	missense (x2)	0.00013	0/0	1
CD14	missense (x2)	0.00021	0/0	1
HSPA8	frameshift, missense	0.00035	0/0	1
HUWE1	missense (x2)	0.014	0/0	1
KIAA1244	missense (x2)	0.0041	0/0	1
KIF18A	missense (x2)	0.00063	1/0	0.52
LPHN2	missense, nonsense	0.0014	0/0	1
MUC6	missense (x2)	0.0059	3/5	1
NIPAL3	missense, nonsense	0.00017	0/0	1
NLRC5	missense (x2)	0.0025	3/4	1
PHC2	missense (x2)	0.00072	0/0	1
PHF7	missense, nonsense	9.80E-05	0/0	1
PIK3C2B	frameshift, missense	0.0024	3/0	0.11
PSPC1	missense, nonsense	0.00034	0/0	1
RYR3	missense (x2)	0.018	4/1	0.22
TAF13	frameshift, nonsense	1.5e-05 **(1.2e-06)**	1/0	0.53

Europe PMC Funders Author Manuscripts

Extended Data Table 3

Enrichment of *de novo* mutations in genes targeted by FMRP and conditional analysis of enrichment in postsynaptic density complexes. Enrichment was tested using *dnenrich* (Supplementary Text). Columns are as in Table 2, and p<0.05 are marked in bold.

	Mutations			Current study		SZ (Gulsuner)[14]				SZ (Xu)[13]				ASD[6-9]				ID[10,11]			
		NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)	
Genes tested	**# of genes**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**
FMRP targets (ALL)	784	**0.0094**	64	0.37	7	0.065	11	1	0	**0.027**	21	0.55	2	**0.003**	102	**0.0003**	26	**2.00E-05**	40	**0.00068**	10
FMRP targets not ARC complex	768	**0.011**	63	0.52	6	0.061	11	1	0	**0.023**	21	0.54	2	**0.0046**	100	**0.00052**	25	**2.00E-05**	35	**0.0094**	8
FMRP targets not NMDAR complex	753	**0.016**	61	0.67	5	0.055	11	1	0	0.062	19	0.84	1	**0.0096**	96	**0.0004**	25	**2.00E-05**	32	0.17	5
FMRP targets not ARC or NMDAR	745	**0.014**	61	0.67	5	0.053	11	1	0	0.059	19	0.84	1	**0.012**	95	**0.00088**	24	**2.00E-05**	31	0.35	4
FMRP targets excluding all PSD genes	615	**0.02**	51	0.68	4	**0.037**	10	1	0	0.12	15	0.77	1	**0.013**	80	**0.0094**	18	**2.00E-05**	29	0.22	4
ARC complex (ALL)	28	**0.00048**	6	**0.005**	2	1	0	1	0	1	0	1	0	0.22	3	0.22	1	**2.00E-05**	5	**0.0015**	2
ARC complex and FMRP target	16	0.46	1	0.068	1	1	0	1	0	1	0	1	0	0.26	2	0.14	1	**2.00E-05**	5	**0.00084**	2
ARC complex not FMRP targets	12	**6.00E-05**	5	**0.045**	1	1	0	1	0	1	0	1	0	0.47	1	1	0	1	0	1	0
NMDAR complex (ALL)	60	**0.025**	6	**0.035**	2	1	0	1	0	0.13	2	0.086	1	0.031	8	0.46	1	**2.00E-05**	8	**2.00E-05**	5
NMDAR complex and FMRP target	31	0.17	3	**0.016**	2	1	0	1	0	0.061	2	0.055	1	**0.031**	6	0.33	1	**2.00E-05**	8	**2.00E-05**	5
NMDAR complex not FMRP targets	29	**0.04**	3	1	0	1	0	1	0	1	0	1	0	0.36	2	1	0	1	0	1	0

Europe PMC Funders Author Manuscripts

Extended Data Table 4a

Brain expression biases of genes impacted by *de novo* mutations

Enrichment of *de novo* mutations (as calculated by *dnenrich*, see Supplementary Text) falling in genes with pre- or postnatal brain expression bias. Number and significance of overlap of mutations in schizophrenia, autism, and intellectual disability in genes with no brain expression bias, or with a preor postnatal expression bias in the brain, based on HBT data as used in Xu, et al.[13] (see Supplementary Text), for two brain regions; HPC = hippocampus, PFC = prefrontal cortex. Columns are as in Table 2, and p<0.05 are marked in bold.

		Mutations			Current study		SZ (Gulsuner)[14]				SZ (Xu)[13]				ASD[6-9]				ID[10,11]	
			NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)	
Brain region	Expression bias?	# of genes	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut
HPC	none	5373	0.72	106	0.1	19	0.72	14	0.52	3	0.41	33	0.48	5	0.36	186	0.54	30	0.95	25	0.8	6
	pre-natal	6444	0.45	175	0.81	22	0.12	30	0.91	3	0.32	53	0.52	8	**0.00028**	332	0.021	63	0.14	57	0.21	16
	post-natal	7299	0.13	196	0.63	22	0.78	23	0.21	6	0.82	47	0.44	8	1	258	0.99	37	0.33	57	0.57	12
PFC	none	4997	0.36	104	0.41	14	0.99	7	0.72	2	0.9	23	0.94	2	0.92	149	0.89	22	0.89	24	0.71	6
	pre-natal	6266	0.44	174	0.52	25	0.071	31	0.54	5	0.18	55	0.34	9	6.00E-05	333	0.00084	69	0.052	60	0.2	16
	post-natal	7853	0.34	200	0.59	24	0.37	29	0.5	5	0.59	54	0.35	9	0.97	294	0.99	39	0.68	55	0.69	12

Europe PMC Funders Author Manuscripts

Extended Data Table 4b

Enrichment of *de novo* mutations (as calculated by *dnenrich*, see Supplementary Text) based on RNA-seq-based brain expression and developmental trajectory. Number and significance of overlap of mutations in schizophrenia, autism, and intellectual disability in genes not expressed in the brain, highly expressed in the brain, or with a pre- or postnatal expression bias in the brain (rows), based on BrainSpan RNA-seq data (see Supplementary Text). Columns are as in Table 2, and p<0.05 are marked in bold.

	Mutations			**Current study**		**SZ (Gulsuner)[14]**				**SZ (Xu)[13]**				**ASD[6-9]**				**ID[10,11]**	
		NS (482)		**LoF (64)**		**NS (68)**		**LoF (12)**		**NS (137)**		**LoF (20)**		**NS (789)**		**LoF (134)**		**NS (141)**		**LoF (34)**	
Brain expression?	**# of genes**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**	**p**	**# mut**
Low	5851	0.89	118	0.97	11	0.48	19	0.17	5	0.31	40	0.44	6	0.99	185	1	22	1	23	1	2
High	9279	**0.0058**	264	**0.016**	40	0.45	34	0.57	6	0.12	74	0.34	11	**2.00E-05**	442	**0.00018**	86	**6.00E-05**	93	**0.0007**	26
Pre-natal	7962	0.33	225	0.39	32	0.74	29	0.97	3	0.2	68	0.65	9	**0.00054**	405	**0.00024**	83	0.35	67	0.21	19
Post-natal	2393	0.17	64	0.54	7	0.68	7	0.74	1	0.96	10	0.65	2	0.95	79	0.88	11	0.71	15	0.89	2

Europe PMC Funders Author Manuscripts

Extended Data Table 5

Comparison of genes hit by *de novo* mutations between this study and other disease studies and control individuals

Each set of columns gives the number of mutations (either nonsynonymous (NS) or loss-of-function (LoF)) and enrichment p-value (as calculated by *dnenrich*, see Supplementary Text) in the set of genes hit by *de novos* in the study listed in the corresponding row. For example, the first two rows detail the significance of the overlap of the mutations from other studies of disease hitting the genes hit by mutations in this study. Nominally significant p-values (<0.05) are marked in bold. Disease sets and functional classes are as listed in Table 2.

	Mutations (N)	Current study				SZ (Gulsuner)[14]				SZ (Xu)[13]				ASD[6-9]				ID[10,11]				Controls[7-10,13-14]			
		NS (482)		LoF (64)		NS (68)		LoF (12)		NS (137)		LoF (20)		NS (789)		LoF (134)		NS (141)		LoF (34)		NS (434)		LoF (49)	
Gene set		p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut	p	# mut
Current study	NS (464)			-		0.16	6	**0.03**	3	0.85	4	0.29	1	**0.016**	55	**0.0066**	15	**0.044**	13	0.26	3	0.61	21	0.72	2
	LoF (63)					**0.014**	3	0.088	1	1	0	1	0	**0.0023**	14	**0.00012**	7	**0.019**	4	**0.002**	3	1	0	1	0
SZ (Gulsuner)[14]	NS (67)	0.22	6	**0.021**	3			-		0.31	2	0.16	1	0.47	7	0.11	3	0.67	1	1	0	0.48	4	1	0
	LoF (12)	0.051	3	0.11	1					0.21	1	1	0	0.15	3	0.21	1	0.21	1	1	0	1	0	1	0
SZ (Xu)[13]	NS (136)	0.79	5	1	0	0.25	2	0.15	1			-		0.083	16	1	0	0.13	4	**0.012**	3	0.081	10	0.49	1
	LoF (20)	0.24	2	1	0	0.13	1	1	0					1	0	1	0	**0.0026**	3	**4.00E-05**	3	0.21	2	1	0
ASD[6-9]	NS (743)	0.14	45	0.023	9	0.49	6	0.13	3	0.32	14	1	0			-		**6.00E-05**	24	**2.00E-05**	12	0.26	38	0.64	4
	LoF (128)	**0.015**	11	**0.00072**	4	0.11	2	0.17	1	1	0	1	0					**2.00E-05**	8	**2.00E-05**	5	0.36	8	0.52	1
ID[10,11]	NS (132)	**0.032**	9	**0.031**	1	0.56	1	0.14	1	0.14	2	**0.01**	1	**2.00E-05**	24	**2.00E-05**	7			-		0.37	7	0.46	1
	LoF (30)	0.27	1	**0.019**	1	1	0	1	0	**0.046**	1	**0.0062**	1	**2.00E-05**	15	**2.00E-05**	5					1	0	1	0
Controls[7-10,13-14]	NS (424)	0.59	21	1	0	0.41	4	1	0	0.15	9	0.26	2	0.062	41	0.31	8	0.48	7	1	0			-	
	LoF (49)	0.6	2	1	0	1	0	1	0	0.44	1	1	0	0.42	4	0.45	1	0.45	1	1	0				

Extended Data Table 6a

Mammalian conservation at *de novo* mutation sites and of genes hit by *de novo* SNVs

Mann-Whitney rank test of the Genomic Evolutionary Rate Profiling (GERP) score (see Supplementary Text) distributions of nonsynonymous (NS) *de novo* mutations between pairs of phenotypes, with significant pairwise comparisons (p<0.05) in bold.

		ID[10,11]	ASD[6-9]	Current study
	median (N)	4.89 (141)	4.72 (780)	4.48 (481)
ID[10,11]	4.89 (141)	-	**0.028**	**0.00053**
ASD[6-9]	4.72 (780)	0.972	-	**0.013**
Current study	4.48 (481)	0.999	0.987	-

Extended Data Table 6b

Mann-Whitney rank test of the median GERP scores of genes (see Supplementary Text) hit by nonsynonymous *de novo* mutations between pairs of phenotypes, with significant comparisons (p<0.05) in bold.

		ID[10,11]	ASD[6-9]	Current study
	median (N)	4.75 (141)	4.27 (780)	4.2 (481)
ID[10,11]	4.75 (141)	-	**0.00015**	**0.000009**
ASD[6-9]	4.27 (780)	0.9999	-	0.166
Current study	4.2 (481)	1.0000	0.834	-

Extended Data Table 6c

Linear modeling of variant GERP and per-gene GERP was employed to test whether the differences observed in (a) were driven by those observed in (b). The coefficients and p-value of the variant GERP score (from the joint linear models) are shown, where, for example, “ID > Current study” indicates a test of whether the conservation at sites of *de novo* mutations in ID is greater than that of mutations in SZ (from the current study), after correcting for the fact that the mutations in ID hit genes with greater overall conservation (b). p<0.05 are marked in bold.

Comparison	Coefficient	P-value
ID > ASD	0.052	0.270
ID > Current study	**0.102**	**0.044**
ASD > Current study	0.039	0.079

Logistic regression model for (X > Y): type ~ gene_gerp + variant_gerp

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Work in Cardiff was supported by Medical Research Council (MRC) Centre (G0800509) and Program Grants (G0801418) and the European Community's Seventh Framework Programme (HEALTH-F2-2010-241909 (Project EU-GEI)).

Work at the Icahn School of Medicine at Mount Sinai was supported by the Friedman Brain Institute, the Institute for Genomics and Multiscale Biology (including computational resources and staff expertise provided by the Department of Scientific Computing), and National Institutes of Health grants R01HG005827 (SMP), R01MH099126 (SMP), and R01MH071681 (PS).

Work at the Broad Institute was funded by Fidelity Foundations, the Sylvan Herman Foundation, philanthropic gifts from Kent and Liz Dauten, Ted and Vada Stanley, and an anonymous donor to the Stanley Center for Psychiatric Research.

Work at the Wellcome Trust Sanger Institute was supported by The Wellcome Trust (grant numbers WT089062 and WT098051) and also by the European Commission FP7 project gEUVADIS no. 261123 (PP).

We would like to thank Mark Daly, Ben Neale, and Kaitlin Samocha for their helpful discussions and providing unpublished autism data. We would also like to acknowledge Mark de Pristo, Stacey Gabriel, Timothy J. Fennel, Khalid Shakir, Charlotte Tolonen, and Hardik Shah for their help in generating and processing the various data sets.

References

1. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. 2012; 13:537–51. [PubMed: 22777127]
2. Bundy H, Stahl D, MacCabe JH. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. Acta Psychiatr. Scand. 2011; 123:98–106. [PubMed: 20958271]
3. McClellan JM, Susser E, King M-C. Schizophrenia: a common disease caused by multiple rare alleles. Br. J. Psychiatry. 2007; 190:194–9. [PubMed: 17329737]
4. Malhotra D, Sebat J. Genetics: Fish heads and human disease. Nature. 2012; 485:318–9. [PubMed: 22596152]
5. Rees E, Moskvina V, Owen MJ, O'Donovan MC, Kirov G. De novo rates and selection of schizophrenia-associated copy number variants. Biol. Psychiatry. 2011; 70:1109–14. [PubMed: 21855053]
6. Neale BM, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012; 485:242–5. [PubMed: 22495311]
7. O'Roak BJ, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012; 485:246–50. [PubMed: 22495309]
8. Sanders SJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485:237–41. [PubMed: 22495306]
9. Iossifov I, et al. De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron. 2012; 74:285–299. [PubMed: 22542183]
10. Rauch A, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012; 380:1674–82. [PubMed: 23020937]
11. De Ligt J, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 2012; 367:1921–9. [PubMed: 23033978]
12. Girard SL, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nat. Genet. 2011; 43:860–3. [PubMed: 21743468]
13. Xu B, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat. Genet. 2012; 44:1365–9. [PubMed: 23042115]
14. Gulsuner S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013; 154:518–29. [PubMed: 23911319]
15. Purcell S. A polygenic burden of rare disruptive mutations in schizophrenia. Nature.

16. Kong A, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012; 488:471–5. [PubMed: 22914163]
17. Mesholam-Gately RI, Giuliano AJ, Goff KP, Faraone SV, Seidman LJ. Neurocognition in first-episode schizophrenia: a meta-analytic review. Neuropsychology. 2009; 23:315–36. [PubMed: 19413446]
18. Glessner JT, et al. Strong synaptic transmission impact by copy number variations in schizophrenia. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:10584–9. [PubMed: 20489179]
19. Malhotra D, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron. 2011; 72:951–63. [PubMed: 22196331]
20. Kirov G, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry. 2012; 17:142–53. [PubMed: 22083728]
21. Darnell JC, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell. 2011; 146:247–61. [PubMed: 21784246]
22. Malenka RC, Nicoll RA. NMDA-receptor-dependent synaptic plasticity: multiple forms and mechanisms. Trends Neurosci. 1993; 16:521–7. [PubMed: 7509523]
23. Bramham CR, et al. The Arc of synaptic memory. Exp. Brain Res. 2010; 200:125–40. [PubMed: 19690847]
24. Korb E, Finkbeiner S. Arc in synaptic plasticity: from gene to behavior. Trends Neurosci. 2011; 34:591–8. [PubMed: 21963089]
25. Shepherd JD, Bear MF. New views of Arc, a master regulator of synaptic plasticity. Nat. Neurosci. 2011; 14:279–84. [PubMed: 21278731]
26. Mikuni T, et al. Arc/Arg3.1 Is a Postsynaptic Mediator of Activity-Dependent Synapse Elimination in the Developing Cerebellum. Neuron. 2013; 78:1024–1035. [PubMed: 23791196]
27. Feinberg I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? J. Psychiatr. Res. 17:319–34. [PubMed: 7187776]
28. Darnell JC, Klann E. The translation of translational control by FMRP: therapeutic targets for FXS. Nat. Neurosci. 2013 at <http://www.ncbi.nlm.nih.gov/pubmed/23584741>.
29. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 2000; 25:25–9. [PubMed: 10802651]
30. Gilman SR, et al. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron. 2011; 70:898–907. [PubMed: 21658583]
31. Golzio C, et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature. 2012; 485:363–7. [PubMed: 22596160]
32. Steinberg S, et al. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Hum. Mol. Genet. 2011; 20:4076–81. [PubMed: 21791550]
33. Von Eichborn J, et al. SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations. Nucleic Acids Res. 2013; 41:D834–40. [PubMed: 23143269]
34. Aoto J, Martinelli DC, Malenka RC, Tabuchi K, Südhof TC. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. Cell. 2013; 154:75–88. [PubMed: 23827676]
35. Williams NM, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet. 2010; 376:1401–8. [PubMed: 20888040]
36. Girirajan S, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N. Engl. J. Med. 2012; 367:1321–31. [PubMed: 22970919]
37. Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry. 2011; 198:173–5. [PubMed: 21357874]
38. Nozawa R-S, et al. Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation. Nat. Cell Biol. 2010; 12:719–27. [PubMed: 20562864]
39. Buxbaum JD, et al. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. Neuron. 2012; 76:1052–6. [PubMed: 23259942]

40. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going… but still not gone. Br. J. Psychiatry. 2010; 196:92–5. [PubMed: 20118450]
41. Insel TR. NIMH · Transforming Diagnosis. NIMH Dir. Blog. 2013 at <http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml>.

Extended Data References

42. Pocklington AJ, Armstrong JD, Grant SGN. Organization of brain complexity--synapse proteome form and function. Brief. Funct. Genomic. Proteomic. 2006; 5:66–73. [PubMed: 16769682]
43. Coba MP, et al. TNiK is required for postsynaptic and nuclear signaling pathways and cognitive function. J. Neurosci. 2012; 32:13987–99. [PubMed: 23035106]
44. Ruderfer DM, et al. A family-based study of common polygenic variation and risk of schizophrenia. Mol. Psychiatry. 2011; 16:887–8. [PubMed: 21483432]

Europe PMC Funders Author Manuscripts

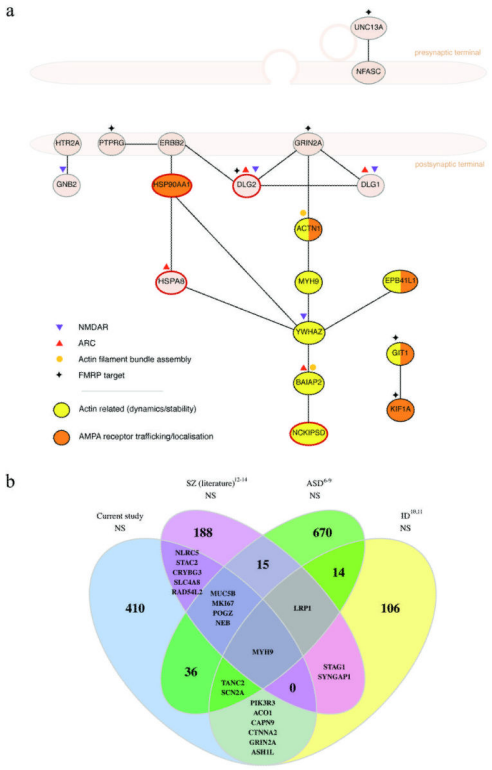


Figure 1. *De novo* mutations from schizophrenia affect genes in the synapse and genes impacted in other neuropsychiatric diseases

a. Synaptic protein-protein interactions between proteins affected by nonsynonymous *de novo* mutations in schizophrenia. Interactions were retrieved from the expert-curated lists in the SynSysNet database (http://bioinformatics.charite.de/synsysnet/) and plotted to show their general pre/postsynaptic localization. Genes belonging to various functional sets are as marked, and the 4 genes with LoF mutations are noted with a red outline. Proteins with nonsynonymous *de novos* had more than expected direct interconnections (p=0.008), which was consistent with more overall connectivity to synaptic proteins as a whole (p=0.005).

b. Overlap of genes bearing nonsynonymous (NS) *de novo* mutations in schizophrenia, autism, and intellectual disability. Overlaps of 6 or fewer genes are listed by name. See Extended Data Table 5 for statistical significance of these overlaps; see Table 2 and text for disease sets.

Europe PMC Funders Author Manuscripts

Table 1

Summary results for primary hypotheses

Hypotheses are grouped into four broad categories (a-d). Each is comprised of sub-tests from which we derive global evidence for the broad category (see Supplementary Text). For category a) the broad p-value was generated using Fisher's exact method on the missense, silent, and loss-of-function (LoF) mutation counts. P-values for category b) were generated using Fisher's combined probability test to combine the sub-tests. For c) and d) we combined all genes from each sub-test into a single geneset and evaluated enrichment using *dnenrich* (see main and Supplementary Text). For categories b - d, we separately evaluated two classes of mutation, nonsynonymous (NS) and LoF, making 7 tests in total. Corrected p-values for the broad categories are adjusted by Bonferroni correction for 7 tests. P-values <0.05 (corrected for broad category tests, uncorrected for sub-tests) in bold.

	Hypothesis category	P-value (corrected)		Sub-tests of primary hypotheses	Subtest details	P-value (uncorrected)	
(a)	Increased rates of *de novo* mutations	1.00		NS:S ratio compared to controls[7-10,13-14]	Table 2	0.43	
				LoF:missense ratio compared to controls[7-10,13-14]		0.37	
		NS	LoF			NS	LoF
(b)	Genic recurrence in SCZ	**0.0007**	0.25	Genic recurrence of *de novo* mutations (current study)	ED Table 2	**0.03**	0.20
				Enrichment in SZ (literature[12-14]) NS *de novo* genes	Table 4, ED Table 5	0.59	0.21
				Increased case/control[15] ratio of rare (MAF < 0.1%) LoF variants in *de novo* genes	Purcell, et al.[15]	**0.0003**	**0.0075**
				Excess transmission of NS singletons (current study) in *de novo* genes	-	**0.01**	0.29
				Enrichment in SZ CNV (literature[1,20]) genes	-	0.29	0.66
(c)	Enrichment in candidate genes	**0.0098**	1.00	Enrichment in ARC/NMDAR genes[20]	Table 3, ED Figure 4	**0.0008**	**0.006**
				Enrichment in PSD genes, excluding ARC/NMDAR genes[20]	-	0.24	0.53
				Enrichment in FMRP target genes[9]	ED Table 3	**0.009**	0.37
(d)	Enrichment in autism/ID *de novo* genes	0.17	**0.0055**	Enrichment in autism LoF *de novo* genes[6-9]	Table 4, ED Table 5	**0.02**	**0.0007**
				Enrichment in ID LoF *de novo* genes[10,11]		0.27	**0.02**

Europe PMC Funders Author Manuscripts

Table 2

Ratios of functional classes of *de novo* mutations across various samples

Classes of *de novo* mutation in the present study, previous studies of schizophrenia (Gulsuner[14] and Xu[13]), and in all studies of schizophrenia combined (SZ (ALL)), which includes this study and an additional small study[12]. ASD = Autism Spectrum Disorder, ID = Intellectual Disability. Controls are unaffected individuals or unaffected siblings of probands with ASD or SZ. To control for factors that influence estimates of absolute rates (sequencing depth, calling, parental age, etc.), we tested for differences between the ratios of classes of *de novo* mutations (nonsynonymous to silent, loss-of-function to missense) in the disorder groups and the controls, using Fisher's exact test. Nominally significant p-values (<0.05) are bold. NS = nonsynonymous, S = synonymous (silent), LoF = loss-of-function.

	Controls[7-10,13-14]	**Current study**	**Schizophrenia (ref.14)**	**Schizophrenia (ref.13)**	**Schizophrenia all (refs 12-14)**	**Autism spectrum disorder[6-9]**	**Intellectual disability[10,11]**
Nonsynonymous	434	482	68	137	702	789	141
Synonymous	155	155	29	27	211	255	25
Ratio	2.8	3.1	2.3	5.1	3.3	3.1	5.6
P vs. Controls	-	0.43	0.46	**0.0097**	0.18	0.41	**0.0027**
Loss-of-function	49	64	12	20	100	134	34
Missense	376	408	56	113	588	638	104
Ratio	0.13	0.16	0.21	0.18	0.17	0.21	0.33
P vs. Controls	-	0.37	0.17	0.29	0.17	**0.0072**	**0.0003**

Europe PMC Funders Author Manuscripts

Table 3

Enrichment of *de novo* mutations in postsynaptic protein complexes

Statistical significance for enrichment of *de novo* mutations in glutamatergic postsynaptic gene sets[20]. Nominally significant p-values (<0.05), as calculated by *dnenrich* (see Supplementary Text), are marked in bold. # mut = mutation counts in each set. O/E = observed-to-expected ratio of mutational hits (fold-enrichment statistic) calculated by *dnenrich*. Samples and classes of mutations are as Table 2. Total numbers of mutations for each class in each sample are given in parentheses. Additional details for the current study, including genes and 95% credible intervals (CI) for the O/E statistics, are given in Extended Data Figure 4.

		Current study						Schizophrenia (ref. 14)		Schizophrenia (ref. 13)		Schizophrenia all (refs 12–14)		Autism spectrum disorder[6-9]		Intellectual disability[10,11]	
		Nonsynonymous (482)			Loss-of-function (64)			Nonsynonymous (68)	Loss-of-function (12)	Nonsynonymous (137)	LoF (20)	Nonsynonymous (702)	Loss-of-function (100)	Nonsynonymous (789)	Loss-of-function (134)	Nonsynonymous (141)	Loss-of-function (34)
Gene set	**genes *(N)***	***P***	**No. mut.**	**O/E**	***P***	**# mut**	**O/E**	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***
Postsynaptic density	681	**0.019**	34	1.46	0.091	6	1.92	0.84	0.45	0.65	0.64	0.091	0.12	0.47	0.064	**0.0015**	**4.00E-05**
ARC complex	28	**0.00048**	6	6.06	**0.005**	2	17.42	1	1	1	1	**0.0035**	**0.015**	0.22	0.22	**2.00E-05**	**0.0015**
NMDAR complex	60	**0.025**	6	2.74	**0.035**	2	6.99	1	1	0.13	0.086	**0.016**	**0.011**	**0.031**	0.46	**2.00E-05**	**2.00E-05**

Table 4

Overlap between genes hit by *de novo* mutations in this study and other phenotypes

Number of mutations (# mut) in present study in gene sets derived from previous studies. P-values are calculated by *dnenrich* for enrichment of mutations in the gene sets from previous studies (see Supplementary Text). Nominally significant p-values (<0.05) are in bold. Disease sets and mutation classes are as Table 2. Additional comparisons are given in Extended Data Table 5.

		Current study (mutations)			
		Nonsynonymous (482)		Loss-of-function (64)	
Gene set	**Mutation class (*N* genes)**	**P**	**# mut**	***P***	**# mut**
Schizophrenia (ref. 14)	Nonsynonymous (67)	0.22	6	**0.021**	3
	Loss-of-function (12)	0.051	3	0.11	1
Schizophrenia (ref. 13)	Nonsynonymous (136)	0.79	5	1	0
	Loss-of-function (20)	0.24	2	1	0
Autism spectrum disorder[6-9]	Nonsynonymous (743)	0.14	45	0.023	9
	Loss-of-function (128)	**0.015**	11	**0.00072**	4
Intellectual disability[10,11]	Nonsynonymous (132)	**0.032**	9	**0.031**	1
	Loss-of-function (30)	0.27	1	**0.019**	1
Controls[7-10,13-14]	Nonsynonymous (424)	0.59	21	1	0
	Loss-of-function (49)	0.6	2	1	0