ARTICLE

doi:10.1038/nature12975

A polygenic burden of rare disruptive mutations in schizophrenia

Shaun M. Purcell[1,2,3,4,5], Jennifer L. Moran[1]*, Menachem Fromer[1,2,3,4]*, Douglas Ruderfer[2,3]*, Nadia Solovieff[4], Panos Roussos[2,3], Colm O'Dushlaine[1], Kimberly Chambert[1], Sarah E. Bergen[1,6], Anna Kähler[6], Laramie Duncan[1,4,5], Eli Stahl[2,3], Giulio Genovese[1], Esperanza Fernández[7,8], Mark O. Collins[9], Noboru H. Komiyama[9], Jyoti S. Choudhary[9], Patrik K. E. Magnusson[6], Eric Banks[5], Khalid Shakir[5], Kiran Garimella[5], Tim Fennell[5], Mark DePristo[5], Seth G. N. Grant[10], Stephen J. Haggarty[1,4,11], Stacey Gabriel[5], Edward M. Scolnick[1], Eric S. Lander[5], Christina M. Hultman[6], Patrick F. Sullivan[12], Steven A. McCarroll[1,5,13] & Pamela Sklar[2,3,14]

Schizophrenia is a common disease with a complex aetiology, probably involving multiple and heterogeneous genetic factors. Here, by analysing the exome sequences of 2,536 schizophrenia cases and 2,543 controls, we demonstrate a polygenic burden primarily arising from rare (less than 1 in 10,000), disruptive mutations distributed across many genes. Particularly enriched gene sets include the voltage-gated calcium ion channel and the signalling complex formed by the activity-regulated cytoskeleton-associated scaffold protein (ARC) of the postsynaptic density, sets previously implicated by genome-wide association and copy-number variation studies. Similar to reports in autism, targets of the fragile X mental retardation protein (FMRP, product of *FMR1*) are enriched for case mutations. No individual gene-based test achieves significance after correction for multiple testing and we do not detect any alleles of moderately low frequency (approximately 0.5 to 1 per cent) and moderately large effect. Taken together, these data suggest that population-based exome sequencing can discover risk alleles and complements established gene-mapping paradigms in neuropsychiatric disease.

Genetic studies of schizophrenia (MIM 181500) have demonstrated a substantial heritability[1,2] that reflects common and rare alleles at many loci. Genome-wide association studies (GWAS) continue to uncover common single nucleotide polymorphisms (SNPs) at novel loci[3]. Rare or *de novo* genic deletions and duplications (copy-number variants (CNVs)) have been firmly established, including risk variants at 22q11.2, 15q13.3 and 1q21.1 (refs 4, 5). One notable outcome of these large-scale, genome-wide investigations is the degree of polygenicity, consistent with thousands of genes and non-coding loci harbouring risk alleles[3,6–9].

Nonetheless, progress has been made in implicating biological systems and quantifying shared genetics among related psychiatric disorders (for example, refs 10, 11), such as identifying common variants in calcium ion channel genes affecting schizophrenia and bipolar disorder[12] and *de novo* CNVs affecting genes encoding members of the postsynaptic density (PSD) proteome[13], in particular members of the neuronal ARC protein and *N*-methyl-D-aspartate receptor (NMDAR) postsynaptic signalling complexes.

Here we apply massively parallel short-read sequencing to assay a substantial portion of variation that previously was essentially invisible: rare coding point mutations (single nucleotide variants (SNVs)) and small insertions and deletions (indels). Although previous schizophrenia studies have applied sequencing, the results have been inconclusive, reflecting limited sample sizes or a focus on small numbers of candidate genes[14–17]. Exome-sequencing studies of *de novo* mutations published to date have neither demonstrated an increased rate in schizophrenia, nor conclusively implicated individual genes[18,19], although some data suggest a link with particular classes of gene, such as those with higher brain expression in early fetal life[19]. *De novo* studies in intellectual disability[20,21] and autism[22–25] have, however, made considerable progress in identifying large-effect alleles and the underlying gene networks.

In this study, we sought to identify the alleles, genes or gene networks that harbour rare coding variants of moderate or large effect on risk for schizophrenia by exome sequencing 5,079 individuals, selected from a Swedish sample of more than 11,000 individuals. Previous analyses of the full sample (Supplementary Information section 1) have demonstrated an enriched burden of rare CNVs and a polygenic common variant component[3]. We generated high-coverage exome sequence to ensure sufficient sensitivity to detect and genotype alleles observed in only one heterozygous individual (singletons, implying an allele frequency of ~1 in 10,000, although the true population frequency will typically be rarer).

The high baseline rate of rare, neutral mutations makes it difficult to detect rare alleles that increase risk for common diseases[26]. Although power can be increased by jointly testing groups of variants in a gene[27], association testing across all genes is likely to be under-powered at current sample sizes. Indeed, a recent application of population-based exome sequencing in autism did not identify genes[28], despite moderately large sample size and the success of the *de novo* paradigm. Furthermore, many confirmed results from candidate-gene sequencing studies of nonpsychiatric disease still fall short of exome-wide significance[29].

We therefore adopted a top-down strategy in which we studied a large set of genes with a higher likelihood of having a role in schizophrenia, on the basis of existing genetic evidence (Supplementary Information section 7). We focused on ~2,500 genes implicated by unbiased,

[1]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [3]Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [4]Analytic and Translational Genetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [5]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden. [7]Center for Human Genetics, KU Leuven, 3000 Leuven, Belgium. [8]VIB Center for Biology of Disease, 3000 Leuven, Belgium. [9]Proteomic Mass Spectrometry, The Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. [10]Genes to Cognition Programme, Centre for Clinical Brain Sciences and Centre for Neuroregeneration, The University of Edinburgh, Edinburgh EH16 4SB, UK. [11]Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [12]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA. [13]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [14]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.
*These authors contributed equally to this work.

large-scale genome-wide screens, including GWAS, CNV and *de novo* SNV studies, testing for enrichment of rare alleles in cases. To prioritize individual genes, we characterized emerging signals with respect to the genes and frequency and type of mutations. We coordinated analysis with an independent trio exome-sequencing study (Fromer *et al.*[30], this issue) and note key points of convergence below.

After alignment and variant calling of all samples jointly, we removed 11 subjects with low-quality data along with likely spurious sites and genotypes (Supplementary Information sections 2 and 3). Per individual, 93% of targeted bases were covered at ≥10-fold (81% at ≥30-fold). The final data set comprised 2,536 cases and 2,543 controls (Extended Data Table 1a and Extended Data Fig. 1a). Cases and controls had similar technical sequencing metrics, including total coverage, proportion of deeply covered targets, and overall proportion of non-reference alleles (Extended Data Table 1b). We observed 635,944 coding and splice-site passing variants of which 56% were singletons. Using Sanger sequencing and Exome Chip data on these samples, we determined high specificity and sensitivity for singletons (Supplementary Information section 3).

We annotated variants with respect to RefSeq and combined five *in silico* algorithms to predict missense deleteriousness (Extended Data Table 1c and Supplementary Information section 4). As expected, allelic types more likely to affect protein function showed greater constraint: 69% of nonsense variants were singletons, compared to 58% of missense and 51% of silent variants. Primary analyses tested (1) disruptive variants (nonsense, essential splice site and frameshifts, $n = 15{,}972$ alleles with minor allele frequency (MAF) $< 0.1\%$); (2) disruptive plus missense variants predicted to be damaging by all five algorithms ($n = 50{,}369$); and (3) disruptive plus missense variants predicted to be damaging by at least one algorithm ($n = 233{,}575$). These groups are labelled disruptive, NS_{strict} and NS_{broad}, in which NS indicates nonsynonymous. We also stratified most analyses by allele frequency: (1) singletons; (2) up to 0.1% (ten or fewer minor alleles); and (3) up to 0.5% (50 or fewer minor alleles). In the main gene set analyses, we empirically corrected for multiple testing over the nine combinations of these factors (Supplementary Information section 7).

The most significant SNV or indel association ($P = 5 \times 10^{-8}$) was for a common missense allele in *CCHCR1*, in the major histocompatibility complex (MHC), a known risk locus; this top SNP was in linkage disequilibrium with many other schizophrenia-associated SNPs in the MHC. All $P < 10^{-5}$ variants were for either common alleles or a few instances of likely aberrant variants that had escaped earlier filtering (Supplementary Information section 5). We performed two series of gene-based tests: a one-sided burden test of an increased rare allele rate in cases, and the SNP-set (sequence) kernel association test (SKAT[27]), which allows for risk and protective effects. For both tests, the distribution of gene-based statistics broadly followed a global null (Extended Data Fig. 1b).

Considering only disruptive variants, the genic test yielding the lowest nominal *P* value was for *KYNU* (kynureninase), showing ten variants in cases and zero in controls (Extended Data Table 2 and Supplementary Table 1); one novel nonsense mutation at chr2:143713804 (g.468T>A; p.Y156*) was observed in seven cases and not present in either the Exome Variant Server (http://evs.gs.washington.edu/EVS/) or 1000 Genomes Project (http://www.1000genomes.org/). Although previous studies have suggested links between the kynurenine pathway and schizophrenia[31], our *P* value of 1.7×10^{-3} does not withstand correction for multiple testing, even if considering only the 246 genes with ≥10 rare disruptive mutations capable of achieving a nominally significant result.

A polygenic burden of rare coding variants

We evaluated a polygenic burden of rare coding variants in cases, first selecting 2,546 genes (~10% of the exome) on the basis of previous genetic studies that we proposed to be enriched for schizophrenia-associated mutations (Supplementary Information section 6). Sources included genome-wide CNV studies[5,13], GWAS[3,12,32] and exome sequencing of *de novo* mutations[18,19,30]. In our sample, these genes had a significantly higher rate of rare (MAF $< 0.1\%$), disruptive mutations in cases compared to controls ($P = 10^{-4}$ for 1,547 versus 1,383 mutations). The enrichment was unlikely to represent technical or ancestry-related artefact because the *P* values controlled for potential differences in exome-wide burden in cases and controls, and because we observed no differences exome wide ($P = 0.24$). Furthermore, enrichment *P* values were empirically derived by permuting phenotypes within subgroups of cases and controls, matched on exome-wide identity-by-state, experimental batch and sex; the above result withstood correction for multiple testing (Table 1). We observed similar results for rarer (singletons, $P = 8 \times 10^{-4}$) and more frequent (MAF $< 0.5\%$, $P = 2 \times 10^{-4}$) alleles. We also observed case enrichment for the strictly defined set of damaging mutations (NS_{strict}, $P = 1.5 \times 10^{-3}$) but not the broader set (NS_{broad}, $P = 0.13$).

Table 1 | Gene set analysis of primary schizophrenia candidate gene sets

Variant type	Gene set/subset	*n* genes	Singletons	MAF < 0.1%	MAF < 0.5%
Disruptive	Primary	2,546	**0.0008**	**0.0001**	**0.0002**
NS_{strict}			0.0059	**0.0015**	0.0110
NS_{broad}			0.0986	0.1295	0.1126
Disruptive	SCZ *de novo* genes				
	Exome sequencing (disruptive)[18,19,30]	87	0.0319	**0.0007**	**0.0003**
	Exome sequencing (nonsyn.)[18,19,30]	611	0.0053	**0.0011**	0.0055
	Copy number variants				
	de novo CNV genes[13]	234	0.0234	0.0039	0.0124
	SCZ-associated CNV genes[5]	345	0.3308	0.4596	0.4376
	GWAS				
	Voltage-gated calcium channel genes[12]	26	**0.0019**	0.0214	0.0212
	Common SNPs ($P < 10^{-4}$ intervals)[3]	479	0.1794	0.0368	0.0037
	miR-137 targets[32]	446	0.6573	0.5609	0.4747
	Synaptic genes				
	PSD (human core)[13]	685	0.0808	0.1154	0.1256
	ARC[13]	28	**0.0016**	**0.0014**	**0.0014**
	NMDAR network[13]	61	0.0158	0.0251	0.0252
	PSD-95[13]	65	**0.0017**	**0.0009**	**0.0010**
	mGluR5[13]	39	0.1327	0.0900	0.0902

Enrichment test empirical *P* values for rare (singleton; MAF < 0.1%; MAF < 0.5%) variants from disruptive, NS_{strict} and NS_{broad} sets. *P* values represent the relative case enrichment compared to average exome-wide case/control difference. Bolded values are significant at $P^{corrected} < 0.05$. Initial comparison corrects (based on the empirical distribution of minimum *P* values) for the nine correlated tests (top panel). The bottom panel focuses on the 12 subsets of the primary gene set, for disruptive variants only as they showed the greatest enrichment for the entire primary set. Again, bold values are significant after correcting for the 36 tests performed.
SCZ, schizophrenia.

This enrichment suggests a polygenic burden of rare variants. Although not so marked as to be detectable at the exome-wide level given the sample size, it is relatively concentrated in genes that were found to be associated with schizophrenia by other methods. The mean allelic effect was not large: in the primary comparison, the odds ratio was 1.12 (1.04–1.20, 95% confidence interval) for each MAF < 0.1% disruptive mutation; 46% of cases carried one or more allele in this primary set (0.62 per case) compared to 41% of controls (0.55 per control). At two extremes, the modest mean effect could represent either that a subset of mutations are fully penetrant or that every allele is associated but increases risk by only 12%, similar to common alleles from GWAS. To extract subsets of potentially stronger-effect alleles, we individually tested the constituent gene sources (Table 1 and Extended Data Fig. 1c), focusing on disruptive variants as they showed the strongest omnibus enrichment. For disruptive mutations, eight out of 12 sets were nominally significant ($P < 0.05$), indicating that the initial observation was not driven by a single category.

ARC, PSD-95 and calcium ion channel genes

Three of the smaller significantly enriched sets (the ARC and PSD-95 (encoded by *DLG4*) complexes and calcium ion channel genes) had odds ratios >5. We observed enrichment ($P = 1.6 \times 10^{-3}$) of disruptive mutations among the 28 ARC complex genes: nine mutations in nine genes (all singletons) in cases and zero in controls, yielding an odds ratio of 19.2 (2.4–2,471, 95% confidence interval; Extended Data Table 2). Along with the NMDAR gene set (also significantly enriched), ARC genes largely accounted for the overall PSD enrichment ($P = 4 \times 10^{-8}$) in ref. 13, in which four ARC genes had one or more *de novo* CNVs. Of note, in an independent exome-sequencing study in trios, Fromer *et al.*[30] found that the ARC gene set was enriched ($P = 5 \times 10^{-4}$) for nonsynonymous *de novo* SNVs and indels, with four genes harbouring six mutations (Extended Data Table 7). The other PSD gene set with strong enrichment ($P = 9 \times 10^{-4}$; odds ratio = 5.1, 1.8–19.2, 95% confidence interval) was the PSD-95 complex, which contains 65 genes and overlaps with ARC. PSD genes are very highly conserved and have critical roles in excitatory neural signalling components, as well as dendrite and spine plasticity. Further categorization of neuronal genes on the basis of subcellular localization[13] (Extended Data Table 3a) or associated mouse and human phenotypes[33] did not yield further enrichment.

The other subset yielding a large odds ratio of 8.4 (2.03–77, 95% confidence interval) was the 26 voltage-gated calcium ion channel genes (12 cases, one control; disruptive singletons, $P = 2 \times 10^{-3}$, although the effect is attenuated when including recurrent alleles: 15/8 cases/controls, $P = 0.021$, see Extended Data Table 2). The singleton enrichment was predominantly driven by the pore-forming α_1 and auxiliary $\alpha_2\delta$ subunits; of the α_1 subunits, the Ca_V1/L-type genes carried the most case mutations, including two in *CACNA1C*, a gene implicated by GWAS of bipolar disorder and schizophrenia[3,10]. Calcium signalling is involved in many cell functions including the regulation of gene expression[34] and is critical for modulating synaptic plasticity[35]. In a secondary analysis of proteins found in the nano-environment of the calcium channel[36], we observed independent enrichment for other ion channel transporters (Supplementary Table 1), odds ratio 9.1 (2.2–83) for singletons ($P = 1 \times 10^{-3}$; 13/1 disruptive alleles).

Convergence with *de novo* studies

A line of convergence across studies was that genes carrying nonsynonymous *de novo* mutations[18,19,30] were enriched for rare disruptive mutations in cases ($P = 1 \times 10^{-3}$; Table 1 and Extended Data Table 6a, b). We observed a similar result for the smaller class of genes carrying disruptive *de novo* mutations ($P = 7 \times 10^{-4}$, from 47 genes in our study); these genes included *UFL1* (5/0 disruptive mutations, $P = 0.03$; 7/0 NS_{strict}, $P = 0.008$), *SYNGAP1* (4/0 NS_{strict}, $P = 0.04$) and *SZT2* (18/9 NS_{strict}, $P = 0.049$). *SYNGAP1* (synaptic Ras GTPase activating protein 1) is a component of the NMDAR PSD complex[37] and mutations in this gene are known to cause intellectual disability and autism[38].

Genes under previously associated CNV regions did not show significant enrichment of rare disruptive mutations, although there was an enrichment of NS_{strict} mutations ($P = 0.0044$; Extended Data Table 4). Of the 11 CNV regions, only the 3q29 locus, which contains multiple genes including *DLG1* (ref. 4), was significant ($P = 0.0006$) and withstood correction for multiple testing.

Autism/intellectual disability genes and FMRP targets

We next tested, as a single set, the 2,507 genes representing autism and intellectual disability candidates (Supplementary Information section 6), which yielded only nominal significance ($P < 0.05$) for disruptive and NS_{strict} variants and no test survived correction for multiple testing (Table 2). Considering the 12 constituent sets, genes from autism *de novo*

Table 2 | Gene set analysis of secondary autism/intellectual disability candidate gene sets

Variant type	Gene set/subset	*n* genes	Singletons	MAF < 0.1%	MAF < 0.5%
Disruptive	Autism/ID	2,507	0.029	0.043	0.049
NS_{strict}			0.052	0.008	0.013
NS_{broad}			0.532	0.619	0.287
		n genes	Min. $P^{corrected}$ (for 9 × 12 = 108 tests)		
Disruptive, NS_{strict} and NS_{broad}	*De novo* genes (exome sequencing)				
	Autism (disruptive) [22–25]	128	1.000		
	Autism (nonsyn.) [22–25]	743	1.000		
	ID (disruptive) [20,21]	30	0.070		
	ID (nonsyn.) [20,21]	132	0.995		
	Neurodevelopmental candidates				
	ASD candidates[39]	112	1.000		
	ID candidates[39]	196	1.000		
	Autism PPI networks				
	CHD8 network[24]	6	1.000		
	49-gene network[24]	49	1.000		
	74-gene network[24]	74	1.000		
	Fragile X mental retardation protein targets				
	Darnell targets[40]	788	**0.010**		
	Ascano targets[42]	939	0.997		
	Ascano FMRP/autism overlap[42]	93	0.993		

Enrichment test empirical *P* values for the secondary (autism/intellectual disability) gene set. As in Table 1, the top panel shows uncorrected *P* values; tests significant after multiple test correction are in bold (that is, all $P^{corrected} > 0.05$). Because no class of variant is significant after multiple test correction for the omnibus test (top panel), we applied and corrected for all 108 tests (nine conditions by 12 subsets) in the bottom panel. The single-category FMRP targets (Darnell *et al.*[40]) mainly reflect disruptive and NS_{strict} singleton enrichment.
ASD, autism spectrum disorder; ID, intellectual disability; PPI, protein–protein interaction.

studies showed no enrichment (Extended Data Fig. 1c), despite greater sample size and number of disruptive *de novo* mutations. There was no evidence for autism or intellectual disability genes curated from the literature[39] or for genes in the protein–protein-interaction-derived subnetworks built around autism *de novo* mutations[24].

The nominal omnibus signals arose largely from the Darnell *et al.* list of FMRP targets[40]. FMRP is encoded by the gene *FMR1* (the locus of the Mendelian fragile X syndrome repeat mutation) and is an RNA-binding protein that regulates translation and is needed at synapses for normal glutamate receptor signalling and neurogenesis[41]. Targets of FMRP are enriched for *de novo* mutations in autism[22,40,42]; here we find significant enrichment of disruptive singletons ($P = 1.4 \times 10^{-3}$; 289/223 case/control count; odds ratio = 1.3). These FMRP targets overlap with PSD genes (Extended Data Table 3b), although were still enriched independently (Supplementary Information section 6). In addition, these genes were enriched in GWAS of this sample ($P < 10^{-3}$, Supplementary Information section 9). Whereas the Darnell list is derived from mouse brain, a second recently reported FMRP target list[42] was generated from cultured human embryonic kidney cells, using a different experimental approach (Supplementary Information section 6). This list has relatively little overlap with Darnell targets and, in contrast to the Darnell list, does not show any enrichment for rare case mutations, for GWAS loci, or comparable overlap with PSD genes (Extended Data Table 3b).

Our results are perhaps surprising: unlike Fromer *et al.*[30], we did not observe direct evidence for overlap at the individual gene level with autism and intellectual disability, despite CNV studies showing pleiotropic effects of individual loci. Nonetheless, at the broader level of gene sets, all three disorders showed enrichment for FMRP targets; autism and intellectual disability *de novo* mutations also showed strong enrichment in several PSD complexes enriched in our study, including NMDAR, PSD-95 and (for intellectual disability) ARC. At the least, our results suggest that any overlap is far from complete, although more refined analyses in larger samples will be needed before a clearer picture can emerge of which genes and pathways are shared and which are specific to one disease.

Characterizing enrichment by variant type

To further characterize the observed enrichment with respect to mutational function and frequency, we created a single 'composite' set of 1,796 genes comprising all members of the most prominently enriched sets (Supplementary Table 2). Rare disruptive mutations in this set were present in 990 cases and 877 controls (for singletons, 645–530). Cases carrying rare disruptive mutations did not appear to be phenotypically or clinically unusual in terms of sex, ancestry, history of drug abuse, general medical conditions plausibly aetiologically related to psychosis, or epilepsy, although they did have a higher rate of admissions noting comorbid intellectual disability compared to other cases ($P = 0.009$; Extended Data Table 2b).

Figure 1 shows composite set enrichment across a range of conditions. As this set merges other sets showing enrichment, it necessarily shows enrichment; it was not, however, due to confounding effects of ancestry, sex or experimental wave (Supplementary Information section 8). It was primarily driven by singleton nonsense mutations across a large number of genes, as it was removed or greatly attenuated when either singleton or nonsense mutations were excluded. Considered alone, neither splice-site, frameshift, missense, silent or noncoding mutations showed enrichment at $P < 0.01$. Different ways of defining damaging missense mutations did not substantively affect results. Considering only nonsynonymous coding variants present on Exome Chip, we did not observe enrichment. Rather, enrichment mainly reflected novel variants (Extended Data Table 5b), which is expected as most rare variants in our study are novel. We also took an alternative approach, whereby instead of filtering variants on frequency, we excluded genes with any control disruptive variants before calculating the burden of case alleles; the composite set was still highly enriched ('case-unique' in

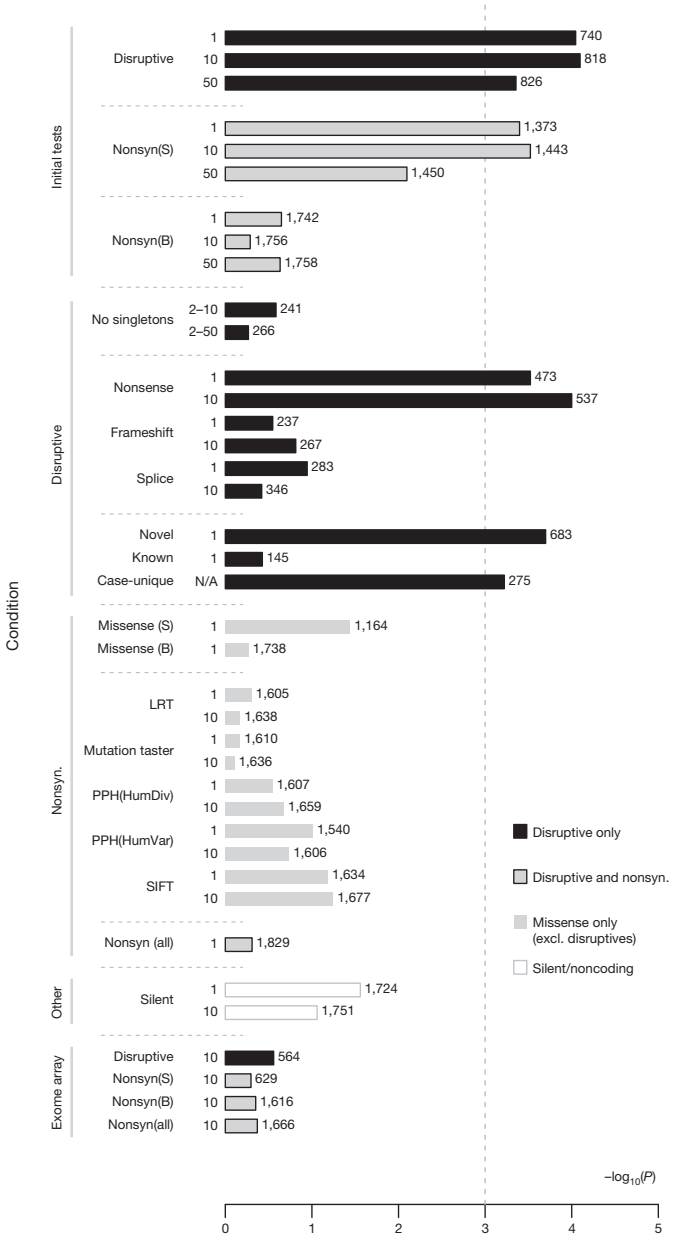


Figure 1 | Composite set gene set analysis, stratified by mutation type. Statistical significance (*x* axis) for the composite gene set stratified by type and frequency of mutation and other variables. Numbers to the right of each bar represent the number of genes with at least one mutation in that category for the composite set. (S) represents strictly defined damaging missenses; (B) represents the broadly defined group. Nonsyn (all) represents all nonsynonymous mutations. Numbers to the left of the bars (1, 10, 50) represent the minor allele count threshold (i.e. 1 indicates a singleton-only analysis); here the ranges 2–10 and 2–50 represent analyses that excluded singletons; N/A indicates that no allele-wise threshold was used. The source of deleteriousness prediction algorithms (LRT, MutationTaster, PPH and SIFT) is described in the Supplementary Information. For the exome array contrasts, Exome Chip sites were tested using the exome sequence calls.

Fig. 1; see Extended Data Table 5b and Supplementary Information section 7). Finally, the enrichment could not be attributed to only a small number of variants or genes (Extended Data Fig. 2a).

These findings do not preclude potentially important effects from other classes of rare variation in specific genes or other gene sets, although exploratory analyses of generic gene sets (for example, based on Gene Ontology terms) did not unambiguously identify novel signals after correction for multiple testing (Supplementary Information section 7).

We found preferential enrichment in genes with high brain expression, but not for genes with a prenatally biased developmental trajectory (Extended Data Fig. 3). In fact, greater enrichment came from postnatally biased genes. Finally, although greatly attenuated compared to disruptive mutations, other categories displayed nominal ($0.01 < P < 0.05$) enrichment in Fig. 1 and strictly defined damaging missense mutations alone showed enrichment for ARC and NMDAR gene sets (32/15 for ARC, $P = 0.007$; Extended Data Tables 5a and 7). Although rare coding alleles other than ultra-rare nonsense mutations will undoubtedly contribute to risk, it will probably prove harder still to elucidate such effects.

Rare variants, CNVs and common GWAS variants

We quantified the relative impact of common SNPs (indexed by a genome-wide polygene score from independent GWAS samples[32]), rare CNVs (the burden of genic deletions) and disruptive mutations in the composite set. Considering the same 5,079 individuals, all three classes of variation were uncorrelated and significantly, independently and additively enriched in cases compared to controls. From logistic regression, the relative effect sizes (reduction in model R^2) were 5.7%, 0.2% and 0.4% for GWAS, rare CNV and rare coding variants, respectively (Supplementary Information section 8). Although not a complete assessment, it indicates that for the current sets of identifiably enriched alleles, common GWAS variants account for an order-of-magnitude more heritability than this set of rare variants does. However, these estimates will be diluted to varying degrees, owing to associated variants being included. As a consequence of this, and also the fact that true risk variants outside of composite set genes were not considered here, this estimate represents a conservative lower bound on the contribution of rare coding variation.

Discussion

We have demonstrated a polygenic burden that increases risk for schizophrenia, primarily comprising many ultra-rare nonsense mutations distributed across many genes. Implicating individual genes remains challenging, as genes that contributed to the highest-ranked sets typically had unremarkable P values, often around 0.5 with the gene containing only one or two rare mutations. Nonetheless, we were able to detect several small and highly enriched sets, notably of genes related to calcium channels and the postsynaptic ARC complex. Across these ~50 genes, ~1% of cases carried a rare disruptive mutation likely to have a considerable impact on risk. However, reported effect sizes will have a tendency to over-estimate true population values (Supplementary Information section 5).

We add to previous work that has implicated disruption of synaptic processes in schizophrenia[13]. The PSD is comprised of supramolecular multiprotein complexes that detect and discriminate patterns of neuronal activity and regulate plasticity processes responsible for learning[43]. Members of the membrane-associated guanylate kinase (MAGUK) family of scaffold proteins, such as PSD-95, have a key role in assembling ~2 MDa complexes comprising calcium channels, including the glutamate-gated NMDAR, voltage-gated calcium channels and ARC[36,44]. The genetic disruption of MAGUKs and their associated components result in specific cognitive impairments in mice and humans[45]. One possibility is that the genetic risk identified here reflects altered tuning in calcium-dependent signalling cascades, triggered by NMDAR[46] and L-type calcium channels[47], mediated by postsynaptic MAGUK signalling complexes driving ARC synthesis.

Although we cannot yet use rare mutations to partition patients into more homogeneous clinical subgroups, this will remain a central goal for future sequencing studies. The few population-based common-disease exome-sequencing studies published to date, in psychiatric (for example, ref. 28) and non-psychiatric (for example, ref. 48) diseases, have not been successful in finding individual genes showing significant enrichment. Our study yields similar findings for individual genes, but yields positive results when considering gene sets. These current findings are likely to foreshadow the definitive identification of individual genes in larger cohorts, following the trajectory of GWAS and other genetic studies of complex disease.

METHODS SUMMARY

Sample ascertainment. Cases with schizophrenia were identified through the Swedish Hospital Discharge Register[3]. Case inclusion criteria: ≥2 hospitalizations with a discharge diagnosis of schizophrenia, both parents born in Scandinavia, age ≥18 years. Case exclusion criteria: hospital register diagnosis of any disorder mitigating a confident diagnosis of schizophrenia. Controls were randomly selected from Swedish population registers. Control inclusion criteria: never hospitalized for schizophrenia or bipolar disorder, both parents born in Scandinavia, age ≥18 years. All subjects provided informed consent; institutional human subject committees approved the research.

Sequencing. The samples (2,536 cases, 2,543 controls) were sequenced using either the Agilent SureSelect Human All Exon Kit (29 Mb, $n = 132$) or the Agilent SureSelect Human All Exon v.2 Kit (33 Mb). Sequencing was performed by IlluminaGAII or Illumina HiSeq2000. Sequence data were aligned and variants called by the Picard (http://picard.sourceforge.net)/BWA[49]/GATK[50] pipeline. Validation of selected variants used Sanger sequencing. On the basis of validation and Exome Chip data, we estimated high sensitivity and specificity of singleton calls. BAM and VCF files are available in the dbGaP study phs000473.v1 (http://research.mssm.edu/statgen/sweden).

Analysis. We used PLINK/SEQ (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000473.v1.p1) to annotate variants according to RefSeq gene transcripts (UCSC Genome Browser, http://genome.ucsc.edu). Single-site association used Fisher's exact test; primary gene-based association used a burden test and the sequence kernel association test[27]. Analyses controlled for ancestry and quality control metrics. Gene sets were evaluated on the empirical distribution of the sum of individual gene burden statistics, and incorporated an empirical correction for multiple testing. Odds ratios with 95% confidence intervals used penalized maximum likelihood (Firth's method) for low cell counts. See Supplementary Information for further details. Summary results are posted at http://research.mssm.edu/statgen/sweden/.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 24 December 2013.
Published online 22 January 2014.

1. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60,** 1187–1192 (2003).
2. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373,** 234–239 (2009).
3. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45,** 1150–1159 (2013).
4. Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* **168,** 302–316 (2011).
5. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev. Genet.* **13,** 537–551 (2012).
6. Stone, J. L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).
7. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).
8. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genet.* **44,** 247–250 (2012).
9. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148,** 1223–1241 (2012).
10. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nature Genet.* **43,** 977–983 (2011).
11. Moreno-De-Luca, D. *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.* **87,** 618–630 (2010).
12. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).
13. Kirov, G. *et al.* *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–153 (2012).
14. Need, A. C. *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* **91,** 303–312 (2012).

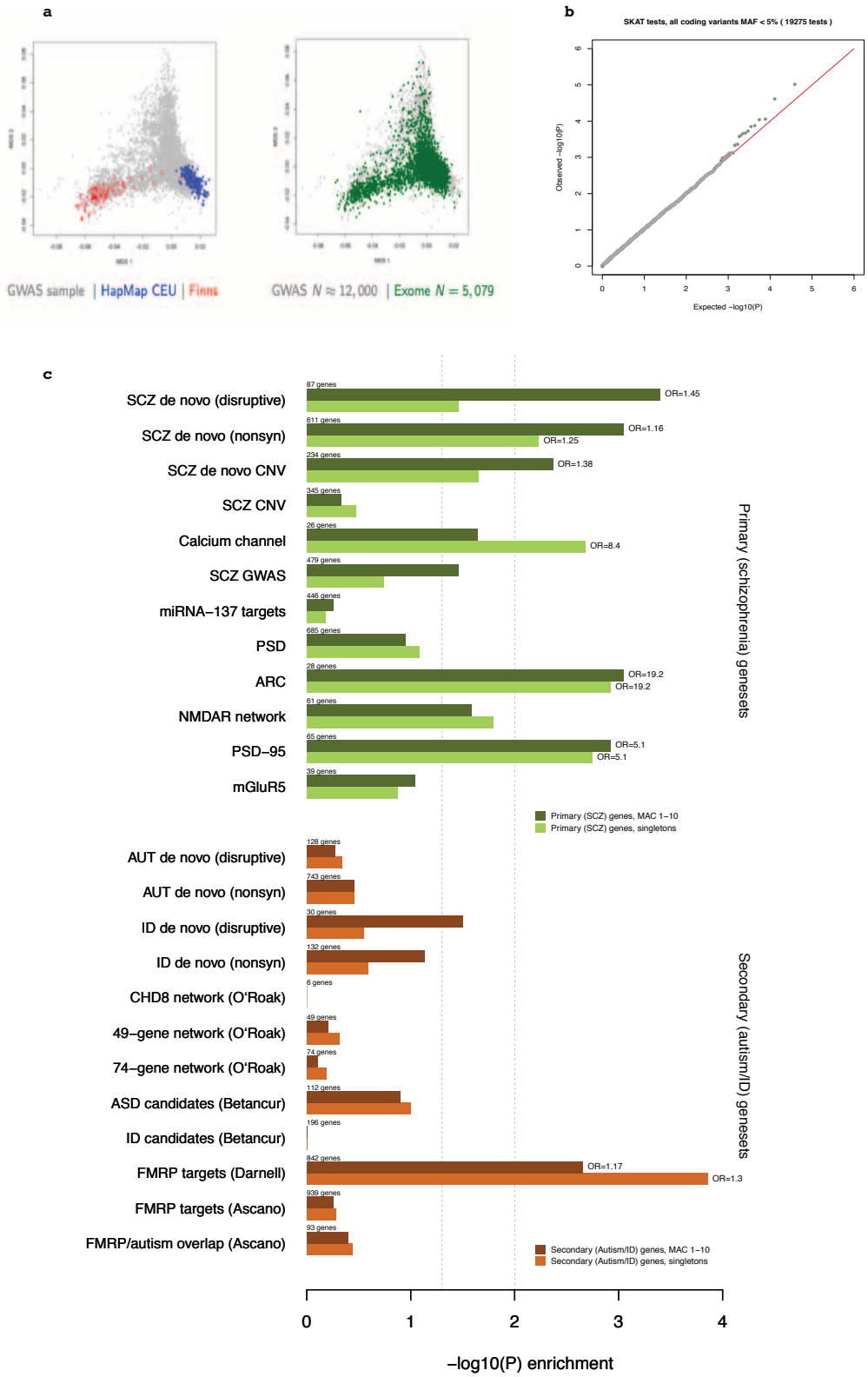
15. Crowley, J. J. *et al.* Deep resequencing and association analysis of schizophrenia candidate genes. *Mol. Psychiatry* **18,** 138–140 (2013).
16. Takata, A. *et al.* A population-specific uncommon variant in *GRIN3A* associated with schizophrenia. *Biol. Psychiatry* **73,** 532–539 (2013).
17. Timms, A. E. *et al.* Support for the *N*-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA Psychiatry* **3,** 1–9 (2013).
18. Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genet.* **43,** 860–863 (2011).
19. Xu, B. *et al. De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature Genet.* **44,** 1365–1369 (2012).
20. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367,** 1921–1929 (2012).
21. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380,** 1674–1682 (2012).
22. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74,** 285–299 (2012).
23. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485,** 242–245 (2012).
24. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485,** 246–250 (2012).
25. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237–241 (2012).
26. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336,** 740–743 (2012).
27. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89,** 82–93 (2011).
28. Liu, L. *et al.* Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet.* **9,** e1003443 (2013).
29. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genet.* **44,** 623–630 (2012).
30. Fromer, M. *et al. De novo* mutations in schizophrenia implicate synaptic networks. *Nature* http://dx.doi.org/10.1038/nature12929 (this issue).
31. Linderholm, K. R. *et al.* Increased levels of kynurenine and kynurenic acid in the CSF of patients with schizophrenia. *Schizophr. Bull.* **38,** 426–432 (2012).
32. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genet.* **43,** 969–976 (2011).
33. Bayés, À. *et al.* Characterisation of the proteome, diseases and evolution of the human postsynaptic density. *Nature Neurosci.* **14,** 19–21 (2011).
34. Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392,** 933–936 (1998).
35. Yasuda, R., Sabatini, B. L. & Svoboda, K. Plasticity of calcium channels in dendritic spines. *Nature Neurosci.* **6,** 948–955 (2003).
36. Müller, C. S. *et al.* Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. *Proc. Natl Acad. Sci. USA* **107,** 14950–14957 (2010).
37. Komiyama, N. H. *et al.* SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J. Neurosci.* **22,** 9721–9732 (2002).
38. Berryer, M. H. *et al.* Mutations in *SYNGAP1* cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. *Hum. Mutat.* **34,** 385–394 (2013).
39. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380,** 42–77 (2011).
40. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).
41. Callan, M. A. & Zarnescu, D. C. Heads-up: new roles for the fragile X mental retardation protein in neural stem and progenitor cells. *Genesis* **49,** 424–440 (2011).
42. Ascano, M., Jr *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386 (2012).
43. Migaud, M. *et al.* Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. *Nature* **396,** 433–439 (1998).
44. Fernández, E. *et al.* Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol. Syst. Biol.* **5,** 269 (2009).
45. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neurosci.* **16,** 16–24 (2013).
46. Steward, O. & Worley, P. F. Selective targeting of newly synthesized *Arc* mRNA to active synapses requires NMDA receptor activation. *Neuron* **30,** 227–240 (2001).
47. Waltereit, R. *et al.* Arg3.1/Arc mRNA induction by Ca^{2+} and cAMP requires protein kinase A and mitogen-activated protein kinase/extracellular regulated kinase activation. *J. Neurosci.* **21,** 5484–5493 (2001).
48. Albrechtsen A. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56,** 298–310 (2013).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler Transform. *Bioinformatics* **25,** 1754–1760 (2009).
50. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful for the participation of all subjects contributing to this research, and to the collection team that worked to recruit them: E. Flordal-Thelander, A.-B. Holmgren, M. Hallin, M. Lundin, A.-K. Sundberg, C. Pettersson, R. Satgunanthan-Dawoud, S. Hassellund, M. Rådstrom, B. Ohlander, L. Nyrén and I. Kizling. We acknowledge funding support from National Institutes of Health (NIH)/National Institute of Mental Health (NIMH) ARRA Grand Opportunity grant NIMH RC2 MH089905 (S.M.P., P.S.), the Sylvan Herman Foundation, the Stanley Center for Psychiatric Research, the Stanley Medical Research Institute, NIH/National Human Genome Research Institute (NHGRI) grant U54HG003067 (E.S.L.), NIH/NIMH grant R01 MH095088 (S.J.H.), NIH/NIMH grant R01 MH091115 (S.J.H.), the Tau Consortium (S.J.H.), NIH/NIMH grant R01 MH099126 (S.M.P.), NIH/NHGRI grant R01 HG005827 (S.M.P.), NIH/NIMH grant R01 MH077139 (P.F.S.), NIH/NIMH grant R01 MH095034 (P.S.), NIH/NIMH grant T32 MH017119 (L.D.), the Friedman Brain Institute at Mount Sinai School of Medicine, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, an ALF grant from Swedish County Council, the Söderström Königska Foundation, the Netherlands Scientific Organization (NWO 645-000-003), the Wellcome Trust, Genes to Cognition Program, The Medical Research Council and European Union projects GENCODYS no. 241995, EUROSPIN no. 242498 and SYNSYS no. 242167 (E.F., M.O.C., N.H.K., J.S.C., S.G.N.G.). Work at the Icahn School of Medicine at Mount Sinai was also supported by the Institute for Genomics and Multiscale Biology (including computational resources and staff expertise provided by the Department of Scientific Computing). The funders had no role in study design, execution, analysis or manuscript preparation.

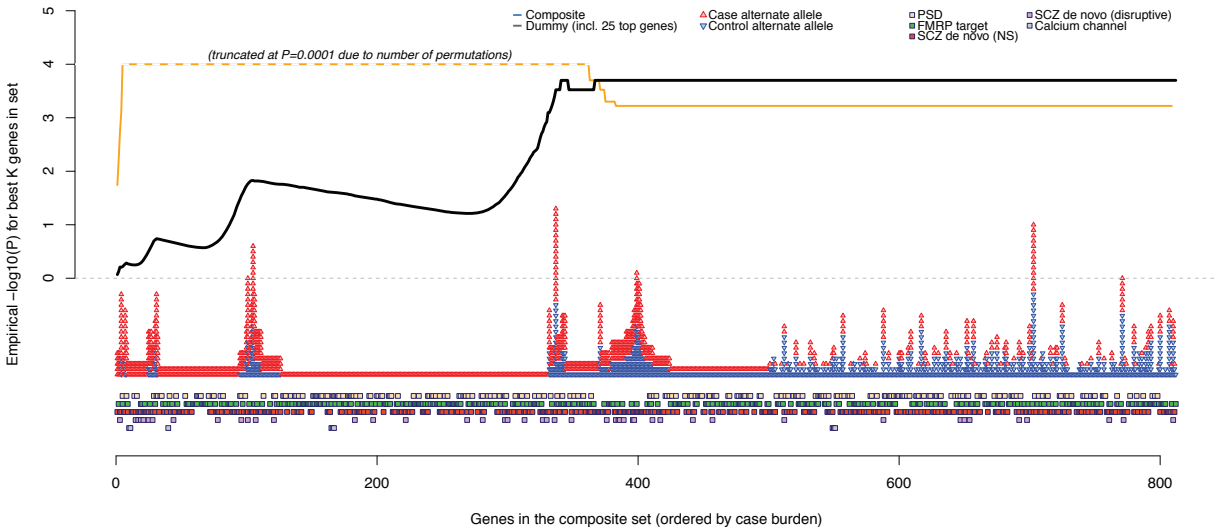
Author Contributions Project leadership: S.M.P., J.L.M., P.F.S., S.A.M., C.M.H., P.S. Sample collection and phenotyping: A.K., P.K.E.M., P.F.S., K.C., J.L.M., C.M.H. Sample processing and data management: K.C., D.R., M.F., J.L.M. Sequencing and variant calling: M.d.P., E.B., K.S., K.G., T.F., S.G. Primary statistical analysis: D.R., M.F., S.M.P. Additional analyses: L.D., E.S., G.G., S.J.H., N.S., P.R., C.O'D., S.E.B. Determination of synaptic gene sets: S.J.H., E.F., M.O.C., N.H.K., J.S.C., S.G.N.G. Interpretation of main findings: S.M.P., E.M.S., E.S.L., S.J.H., M.F., P.F.S., S.A.M., P.S. Primary drafting of manuscript: S.M.P., P.F.S., S.A.M., E.S.L., P.S. Production and approval of the final manuscript: all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.P. (shaun@broadinstitute.org)



Extended Data Figure 1 | Ancestry and association summaries. **a**, Multidimensional scaling plot of ancestry in the full Swedish GWAS sample, in which each point represents one individual; the left panel superimposes HapMap CEU and Finnish samples and the right panel highlights (in green) the subset of the full Swedish sample for whom we have exome sequence data. **b**, Quantile–quantile plot for gene-based SKAT results (MAF $< 5\%$ coding variants). Similar, or more conservative, profiles were obtained for other subsets of variants. **c**, Case enrichment of rare (MAF $< 0.1\%$) and singleton disruptive mutations for the constituent sets of the primary/schizophrenia gene set (top panel in green) and the secondary (autism/intellectual disability) gene set (bottom panel in orange). The primary set is enriched in cases (MAF $< 0.1\%$; disruptive mutations, $P = 10^{-4}$; singletons, $P = 8 \times 10^{-4}$; significant after correction for multiple testing) whereas the autism/intellectual disability set shows only a modest trend ($P = 0.04$ for MAF $< 0.1\%$ and $P = 0.03$ for singletons) and is not significant after correction. x axis represents $-\log_{10}(P)$; OR, odds ratio. Number of genes is for total in the set (whether or not they had a rare variant).

a



b

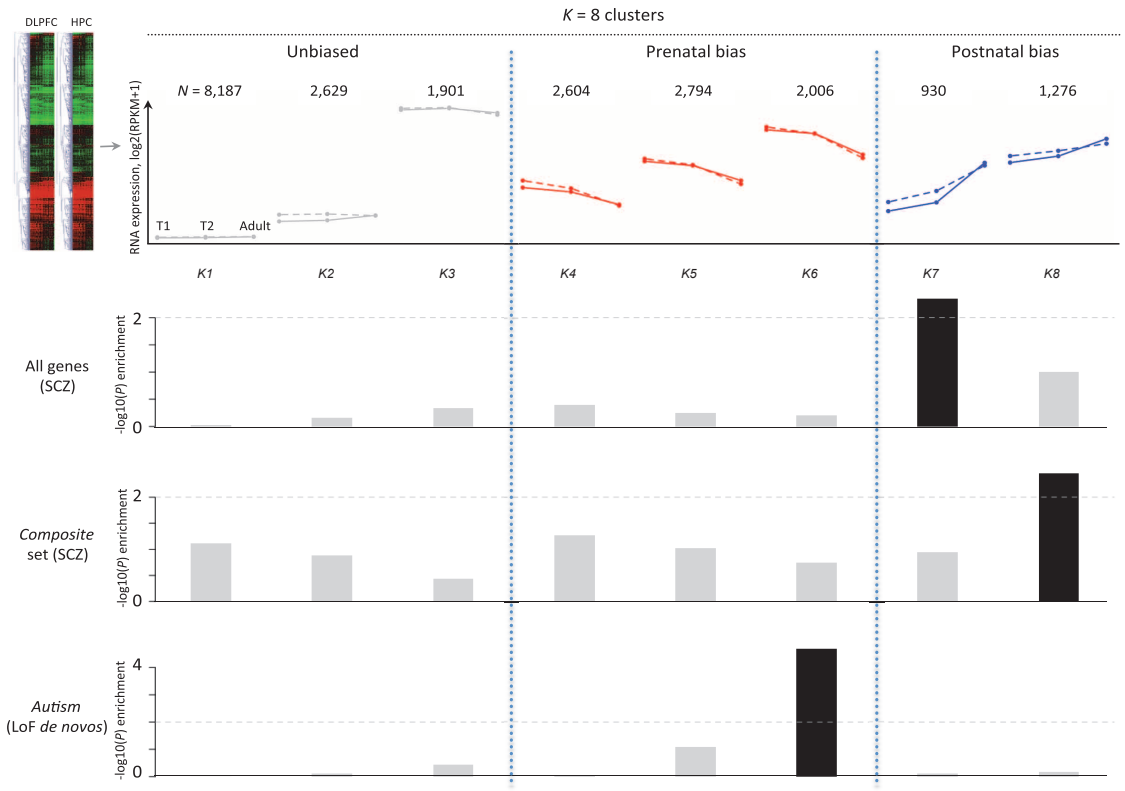
Characteristic of composite set carriers (cases only)	# admissions OR	# admissions P	>0 admissions OR	>0 admissions P	>1 admissions OR	>1 admissions P	>5 admissions OR	>5 admissions P
Hospital admissions								
Duration of hospitalization for SCZ	1.02	0.610	1.02	0.512	1.02	0.625	1.02	0.663
Total number of admissions	1.00	0.957	1.00	0.968	1.02	0.789	1.02	0.690
Year of first admission	1.00	0.861	1.00	0.884	1.00	0.895	1.00	0.924
Year of most recent admission	1.01	0.230	1.01	0.246	1.01	0.235	1.01	0.254
Drug abuse	1.00	0.401	0.96	0.675	0.90	0.400	0.77	0.122
General medical condition plausibly etiologically related to psychosis	0.99	0.596	1.12	0.421	0.96	0.831	0.85	0.664
Epilepsy	1.03	0.494	0.76	0.348	0.81	0.582	1.01	0.990
Intellectual disability	**1.05**	**0.009**	**1.41**	**0.044**	**1.59**	**0.019**	**2.03**	**0.018**
Demographics								
Male	0.89	0.187	0.90	0.199	0.90	0.222	0.90	0.230
In homogeneous subset	1.11	0.287	1.10	0.323	1.10	0.316	1.11	0.280
Finnish ancestry	0.95	0.736	0.94	0.684	0.95	0.719	0.96	0.783

Extended Data Figure 2 | Genic and phenotypic subset analyses for the composite set. **a**, Individual gene-ranking of composite set genes. Genes are ranked by their case burden of rare disruptive mutations, from left to right, for the composite set. The squares along the bottom indicate to which sets each gene belongs. The red and blue triangles represent case and control counts for each gene. The lines above represent the statistical significance of the best test for this set: that is, the significance of the top *K* genes, evaluated by permutation. The black line represents results for the real data (disruptive MAF $< 0.1\%$ composite set analysis). The orange line represents the dummy condition, in which we artificially constructed a set in which the number of genes, statistical enrichment, odds ratio and case/control counts were similar to the real composite set. However, this set included the 25 top-ranked genes from individual gene-based tests (disruptive MAF $< 0.1\%$ variants), with the remainder selected at random. The profile of the best test line is markedly different between the real and dummy gene sets (note: truncated at $P = 0.0001$ reflecting the number of permutations performed). Whereas the dummy *P* value climbs quickly and then drops to the final aggregate result, the true composite set line continues to climb after 200 genes, indicating that many genes with a single disruptive mutation contribute to the observed set enrichment (rather than a relatively small proportion of the 1,796 genes accounting for the majority of the signal, as in the dummy set). **b**, Phenotypic characteristics of cases carrying mutations. Relationship between clinical and demographic measures in schizophrenia cases in relation to carrying one or more composite set disruptive risk alleles (MAF $< 0.1\%$). Hospital Discharge Registry (HDR) data (ICD9 codes) were available on 979 of the 990 case carriers. All *P* values (uncorrected) are two-sided from a case-only joint logistic regression of carrier status (one or more risk alleles) on all admission and demographic variables including year of first and last admissions. The four pairs of columns represent analyses in which we varied the way in which the HDR admission data were represented (for drug abuse, general medication condition, epilepsy and intellectual disability). # admissions, independent variables are the untransformed number of admissions; $>X$ admissions, independent variable is binary 0/1 variable representing whether individuals had more than *X* admissions. Of all clinical/demographic measures considered, we observed a nominally significant increased likelihood that cases carrying a disruptive allele in the composite set have increased rates of secondary diagnoses of intellectual disability compared to other cases (based on HDR ICD9 codes).

a

		Bias: (higher in)			Prenatal			Postnatal	
	Average level:	None	Low/average	High	Low	Higher	Highest	Low	High
BrainSpan : 8-class	**Geneset**	**K1**	**K2**	**K3**	**K4**	**K5**	**K6**	**K7**	**K8**
	n genes	8,187	2,629	1,901	2,604	2,794	2,006	930	1,276
	All genes	0.8653	0.7511	0.8719	0.4672	0.7913	0.3518	0.1639	**0.0005**
	Composite	0.1295	0.0447	0.3896	**0.0057**	0.1395	0.4188	0.1862	**0.0014**
	PSD genes	0.9027	0.2103	0.1043	0.7012	0.4265	0.8340	0.0244	0.0210
	FMRP targets	0.1245	0.2581	0.4230	0.0351	0.5382	0.1616	0.3797	**0.0008**
	SCZ *de novos (NS)*	0.1138	0.1593	0.5044	**0.0089**	0.1855	0.2325	0.2541	0.1367
BrainSpan : 4-class		**None/low**		**High**	**Prenatal**			**Postnatal**	
	n genes	2,748		3,688	3,540			1,058	
	All genes	0.8504		**0.0020**	0.6946			0.0255	
	Composite set	0.0322		**0.0017**	0.1535			**0.0006**	
	PSD	0.3982		0.1043	0.9692			**0.0008**	
	FMRP targets	0.0257		0.0100	0.1918			0.0131	
	SCZ *de novo* (NS)	0.0500		0.0145	0.0359			0.0504	
Human Brain Transcriptome : Xu et al. (ref. 19) 3-way classifications, HPC			**Unbiased**		**Prenatal**			**Postnatal**	
	n genes		6,578		6,841			8289	
	All genes		0.6467		0.7281			0.0517	
	Composite set		0.1009		0.2876			**0.0001**	
	PSD		0.3693		0.9691			0.1043	
	FMRP targets		0.5629		0.1264			**0.0022**	
	SCZ *de novo* (NS)		0.0397		0.1395			**0.0108**	
Human Brain Transcriptome : Xu et al. (ref. 19) 3-way classifications, DLPFC			**Unbiased**		**Prenatal**			**Postnatal**	
	n genes		6,255		6,669			8,784	
	All genes		0.5076		0.7538			0.0894	
	Composite set		0.2143		0.2054			**0.0001**	
	PSD		0.5949		0.8329			0.1043	
	FMRP targets		0.3205		0.4257			**0.0005**	
	SCZ *de novo* (NS)		0.1383		0.1189			**0.0043**	

b



Extended Data Figure 3 | Stratified enrichment analysis *P* values by developmental trajectory of expression in brain (BrainSpan and Human Brain Transcriptome (HBT) data sets). **a**, Uncorrected *P* values for a set of exploratory analyses in which we stratified genes in the enrichment analyses by their developmental profile of brain expression. We used four schemes to classify genes as 'brain expressed' and/or 'biased' with respect to prenatal or postnatal expression (see Supplementary Information section 6 for details). We merged data on the hippocampus and dorsolateral prefrontal cortex for the BrainSpan classifications; to mirror the classification of Xu *et al.*[19] we kept separate these two groupings for the HBT data set. Results presented for MAF $< 0.1\%$ disruptive variants; similar results were obtained for singletons with the exception that the 'K4' prenatal enrichment signals were no longer significant. In general, the most consistent enrichment across variant classes, classification schemes and brain regions emerges for postnatally biased genes with high brain expression. **b**, Analysis of exome variants by developmental expression trajectory in human brain. Genes are grouped by cluster analysis of human postmortem brain expression into eight developmental trajectories, using RNA-sequencing data from the BrainSpan project. The top row gives the number of genes per cluster and the cluster centres in $\log_2$-scaled RPKM (reads per kilobase per million) values; solid and dotted solid lines indicate dorsolateral prefrontal cortex (DLPFC) and hippocampus (HPC), respectively. The bottom two rows show enrichment in the current study, relative to the exome-wide average, for singleton disruptive mutations in cases compared to controls, either subsetting all genes by expression profile (first row), or considering only genes in the composite set (second row). In both cases, we only observed nominally ($P < 0.01$) significant enrichment for genes that are postnatally biased. By contrast, a list of genes with loss-of-function (LoF) *de novo* mutations (compiled and reported in Fromer *et al.*[30]) shows strong enrichment for prenatal bias (see Fromer *et al.*[30] for details on how *de novo* enrichment was calculated). Alternative approaches to classifying genes as prenatally or postnatally biased led to similar conclusions (Supplementary Information section 6).

Extended Data Table 1 | Sample and detected variant properties

a

Case/control status by sex (*P* = 5e-11)	Status	Male	Female	Female (%)
	Control	1291	1252	49%
	Case	1520	1016	40%
Case/control status by ancestry (*P* = 2e-12)	**Ancestry**	**Control**	**Case**	**Case (%)**
	Swedish	2356	2197	48%
	Finnish	139	274	66%

b

Sample and sequencing metrics	Cases	Controls	*P* (case vs. control)
n	2536	2543	.
n (pre-QC)	2546	2545	.
Total number of reads	100,532,755	100,079,333	0.62
Filtered, unique reads aligned	68,940,753	68,339,964	0.26
Filtered, unique bases aligned	5,106,614,996	5,070,497,844	0.34
Mean target coverage	89.98	89.55	0.53
Percentage of target bases covered > 10x	92.83	92.85	0.55
Percentage of target bases covered > 20x	87.30	87.30	0.93
Percentage of target bases covered > 30x	81.13	81.07	0.63
Percentage of targets w/out any bases covered at 2x	1.72	1.72	0.60
Mean number of non-reference genotypes per individual (unfiltered)	18772.9	18786.6	0.13
Mean number of on-target singletons per individual (unfiltered)	49.6	49.0	0.38
Mean dbSNP % per individual	98.3970%	98.3969%	1.00

c

Property	Variant type	*n*	Mean MAC	% singleton
All alternate alleles		635,944	103.37	56%
Functional class				
Noncoding		61,416	142.03	53%
Silent		185,336	152.85	51%
Missense		342,561	69.52	58%
Non-essential splice site		25,450	127.04	54%
Nonsense		9,022	20.68	69%
Essential splice-site		4,394	16.18	70%
Frameshifting indel		3,461	9.46	79%
In silico annotation of missenses				
LRT		168,437	34.55	62%
MutationTaster		167,316	19.90	63%
PolyPhen2 (HumDiv)		130,719	28.84	62%
PolyPhen2 (HumVar)		91,156	24.74	64%
SIFT		140,345	43.85	61%
Primary variant groupings for analysis				
Singletons	Gene disruptive	12,047	1.00	100%
	NS_{strict}	36,542	1.00	100%
	NS_{broad}	160,229	1.00	100%
<0.1% MAF (1-10 alleles)	Gene disruptive	15,972	1.56	75.4%
	NS_{strict}	50,369	1.65	72.5%
	NS_{broad}	233,575	1.78	68.6%
<0.5% MAF (1-50 alleles)	Gene disruptive	16,523	2.24	72.9%
	NS_{strict}	52,545	2.51	69.5%
	NS_{broad}	248,217	3.04	64.6%

a, Numbers of individuals in the final data set, after individual-level QC. Finnish ancestry was inferred by multidimensional scaling. *P* values from Fisher's exact test. **b**, Technical metrics for the cases and controls (after individual-level QC); *P* values for two-sided test of case/control differences (*t*-test). **c**, Properties of variants detected by exome sequencing. Counts (N) and minor allele counts (MAC) for various classes of variant in the main exome data set, following all QC. Missense deleteriousness prediction algorithms and how they were combined are described in Supplementary Information section 4.

Extended Data Table 2 | Genes prioritized as more likely to harbour large-effect alleles

Class	Gene	Singletons	MAF < 0.1%	Notes
ARC/PSD complex				
	CYFIP1	1/0	1/0	SCZ *de novo* (CNV)
	BAIAP2	1/0	1/0	SCZ *de novo* (NS)
	DLG1	1/0	1/0	SCZ *de novo* (NS), SCZ de novo (CNV)
	SLC25A3	1/0	1/0	
	GLUD1	1/0	1/0	
	CAMK2A	1/0	1/0	FMRP target
	ATP1B1	1/0	1/0	AUT *de novo* (disruptive); FMRP target
	IQSEC2	1/0	1/0	ID *de novo* (disruptive); FMRP target
	MBP	1/0	1/0	FMRP target
	Total	9/0	9/0	
		P = 0.0016	*P* = 0.0014	
		OR = 19.2 (2.4 - 2471)	*OR* = 19.2 (2.4 - 2471)	
PSD-95 geneset				
	ABLIM1	1/0	1/0	
	ACO2	1/0	1/0	FMRP target
	ANKS1B	3/1	3/1	
	ATP1B1	1/0	1/0	AUT *de novo* (disruptive); FMRP target
	ATP5A1	1/0	1/0	FMRP target
	BAIAP2	1/0	1/0	SCZ de novo (NS)
	CAMK2A	1/0	1/0	FMRP target
	CAMK2B	2/0	2/0	FMRP target
	DLG1	1/0	1/0	SCZ *de novo* (NS), SCZ de novo (CNV)
	GAPDH	1/0	1/0	
	IQSEC2	1/0	1/0	ID *de novo* (disruptive); FMRP target
	NRXN1	1/0	1/0	SCZ de novo (NS); AUT de novo (disruptive); FMRP target
	PRDX1	0/1	0/1	
	PRDX2	0/1	0/1	
	SUCLA2	1/0	1/0	AUT *de novo* (disruptive)
	SYNGAP1	1/0	1/0	SCZ de novo (disruptive); ID de novo (disruptive); FMRP target
	Total	17/3	17/3	
		P = 0.0017	*P* = 0.0009	
		OR = 5.1 (1.8 - 19.2)	*OR* = 5.1 (1.8 - 19.2)	
Voltage-gated calcium ion channel genes				
	CACNA1B	1/0	1/0	FMRP target
	CACNA1C	2/0	2/0	SCZ & BP GWAS hit
	CACNA1H	1/0	3/0	
	CACNA1S	2/0	2/3	SCZ & AUT *de novos* (NS)
	CACNA2D1	1/0	1/0	PSD
	CACNA2D2	3/0	3/0	
	CACNA2D3	0/0	3/0	AUT *de novo* (disruptive)
	CACNA2D4	1/0	2/4	
	CACNB2	0/1	0/1	
	CACNB4	1/0	1/0	PSD
	Total	12/1	15/8	
		P = 0.0021	*P* = 0.021	
		OR = 8.4 (2.03 - 77)	*OR* = 2.1 (0.97 - 4.9)	
Top disruptive gene-based test				
	KYNU	3/0	10/0	
		P = 0.13	*P* = 0.0017	
			OR = 21.2 (2.7 - 2725)	

Individual gene case/control counts, odds ratios and *P* values for genes from primary gene sets with odds ratios >5, and *KYNU* (top-ranked individual gene). Odds ratios are calculated using Firth's method (penalized maximum likelihood logistic regression) and shown with 95% confidence intervals. *P* values are empirical, uncorrected one-sided burden tests. FMRP target annotations are based on the Darnell *et al.*[40] list only. Supplementary Table 1 lists singleton variant and genotype information for the genes listed here.

Extended Data Table 3 | Extended results for all PSD gene sets

a

Set	*n* genes	Disruptive			NS_{strict}			NS_{broad}		
		Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%	Singletons	MAF < 0.1%	MAF < 0.5%
PSD (human core)	685	0.0729	0.1019	0.1083	0.0058	0.1045	0.1285	0.0866	0.5827	0.3743
ARC	28	0.0016	0.0013	0.0014	**0.0004**	0.0018	0.0047	0.2830	0.4607	0.3542
NMDAR network	61	0.0154	0.0229	0.0225	**0.0001**	0.0007	0.0005	0.0012	0.1426	0.0420
mGluR5	39	0.1299	0.0861	0.0862	0.0628	0.0837	0.0900	0.0302	0.2192	0.1329
PSD-95	65	0.0015	0.0008	0.0008	0.0027	0.0204	0.0393	0.2992	0.3722	0.1147
Pre-synapse	431	0.0187	0.0983	0.1458	0.2327	0.1811	0.3600	0.1306	0.7597	0.6515
Pre-synaptic active zone	173	0.0518	0.0487	0.0482	0.6162	0.6641	0.7082	0.8439	0.9918	0.9554
Synaptic vesicle	344	0.1030	0.3133	0.4151	0.1439	0.1093	0.2466	0.0375	0.3423	0.2718
Cytoplasm	271	0.5851	0.1793	0.1034	0.8983	0.5351	0.6007	0.1861	0.1084	0.1192
Early Endosomes	17	0.8917	0.7860	0.7826	0.2891	0.2420	0.2139	0.1472	0.4019	0.4123
Endoplasmic Reticulum	97	0.3005	0.1882	0.2612	0.6615	0.2805	0.5036	0.6194	0.5044	0.7878
ER/Golgi-derived vesicles	94	0.4258	0.2678	0.3644	0.3001	0.4977	0.6239	0.3063	0.6944	0.7649
Golgi	31	0.5130	0.5493	0.5481	0.1998	0.0921	0.1338	0.0074	0.1628	0.4178
Mitochondrion	197	0.0141	0.0259	0.0178	0.4351	0.0860	0.0671	0.6999	0.2112	0.6079
Nucleus	167	0.1790	0.3029	0.2900	0.0626	0.1512	0.2728	0.2006	0.3340	0.3196
Plasma membrane	50	0.7940	0.5659	0.5635	0.9416	0.8059	0.8091	0.8944	0.5531	0.3028
Recycling Endosomes/ trans-Golgi network	68	0.1502	0.0944	0.1556	0.5349	0.4514	0.5359	0.0902	0.0862	0.1862

b

		PSD (human core) (ref. 13)		FMRP targets (Darnell, ref. 40)		FMRP targets (Ascano, ref. 42)	
	n	*n*	%	*n*	%	*n*	%
PSD (human core)	685	.	.	170	25%	80	12%
ARC	28	.	.	16	57%	2	7%
NMDAR network	61	.	.	32	52%	5	8%
mGluR5	39	.	.	25	64%	7	18%
PSD-95	65	.	.	30	46%	4	6%
Pre-synapse	431	213	49%	87	20%	31	7%
Pre-synaptic active zone	173	121	70%	50	29%	10	6%
Synaptic vesicle	344	162	47%	72	21%	27	8%
Cytoplasm	271	77	28%	16	6%	23	8%
Early Endosomes	17	6	35%	2	12%	1	6%
Endoplasmic Reticulum	97	13	13%	5	5%	4	4%
ER/Golgi-derived vesicles	94	24	26%	7	7%	4	4%
Golgi	31	2	6%	2	6%	4	13%
Mitochondrion	197	57	29%	6	3%	12	6%
Nucleus	167	19	11%	7	4%	19	11%
Plasma membrane	50	16	32%	6	12%	5	10%
Recycling Endosomes/trans-Golgi network	68	19	28%	3	4%	7	10%
Total	1509	685	45%	170	11%	80	5%

a, Full PSD gene set association results. For all nine (three annotation levels by three frequency levels), *P* values for enrichment of all gene sets described and tested in Kirov *et al.*[13]. In addition to the PSD genes (top five rows), enrichment statistics for presynaptic genes, and neuronal genes clustered on the basis of subcellular location are given. Although the *P* values presented are uncorrected, we performed this analysis correcting for all $9 \times 17 = 153$ tests (by considering the distribution of the minimum empirical *P* value across tests and sets, as described in the Supplementary Information). The values in bold are significant ($P^{corrected} < 0.05$) after correction for multiple testing. Both ARC and NMDAR network are significant after multiple test correction, for the singleton NS_{strict} category. (Note: for ARC the disruptive singleton category is, as reported in the primary test, highly significant and withstands correction for multiple testing in that context; in this broader, less focused analysis it yields $P^{corrected} = 0.17$; the majority of $P^{corrected}$ values (not shown) are 1.00.) **b**, PSD and FMRP-target gene sets: descriptive statistics and overlap. Overlap between Darnell *et al.*[40] and Ascano *et al.*[42] FMRP targets and PSD genes: for example, 57% (16 out of 28) of ARC genes are in the Darnell FMRP list. By contrast, only 7% (2 out of 28) are in the Ascano list. There is a similar trend across the three other major PSD subsets considered here: NMDAR network, PSD-95 and *mGluR5* genes. Conversely, 22% of Darnell targets are in the PSD (human core) compared to only 9% of Ascano targets.

Extended Data Table 4 | Association results for individual CNV regions

		Disruptive		NS_{strict}	
Group	**Genes**	**Singletons**	**MAF < 0.1%**	**Singletons**	**MAF < 0.1%**
CNV loci	All	0.3279	0.4557	0.0843	**0.0044**
	1q21.1	0.4533	0.6966	0.3205	0.1832
	2p16.3	0.4775	0.3580	0.4703	0.2750
	3q29	0.1054	0.0068	0.0123	**0.0006**
	7q36.3	0.8642	0.5750	0.6411	0.2688
	7q11.23	0.7199	0.6800	0.3329	0.2207
	15q11.2	0.3208	0.1616	0.5138	0.1362
	15q13.3	0.0883	0.3976	0.3672	0.2746
	16p13.11	0.4194	0.3775	0.8346	0.4124
	16p11.2	0.1613	0.1240	0.0655	0.0974
	17q12	0.7377	0.5205	0.4313	0.1385
	22q11.21	0.9386	0.9977	0.5456	0.8754

	Gene	**A/U**	**Gene name**
3q29 genes	*DLG1*	5/0	discs, large homolog 1 (Drosophila)
	RNF168	5/1	ring finger protein 168, E3 ubiquitin protein ligase
	CEP19	2/0	centrosomal protein 19kDa
	LRRC33	2/0	leucine rich repeat containing 33
	PAK2	2/0	p21 protein (Cdc42/Rac)-activated kinase 2
	PCYT1A	5/2	phosphate cytidylyltransferase 1, choline, alpha
	PIGX	6/3	phosphatidylinositol glycan anchor biosynthesis, class X
	FBXO45	1/0	F-box protein 45
	NCBP2	1/0	nuclear cap binding protein subunit 2, 20kDa
	PIGZ	1/0	phosphatidylinositol glycan anchor biosynthesis, class Z
	TFRC	1/0	transferrin receptor (p90, CD71)
	ZDHHC19	1/0	zinc finger, DHHC-type containing 19
	C3orf43	4/2	chromosome 3 open reading frame 43
	MFI2	15/12	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
	SLC51A	1/1	solute carrier family 51, alpha subunit
	BDH1	0/1	3-hydroxybutyrate dehydrogenase, type 1
	TCTEX1D2	0/1	Tctex1 domain containing 2
	WDR53	0/1	WD repeat domain 53

Focused enrichment analysis of genes under schizophrenia-associated CNV regions. The top panel presents omnibus *P* values testing all genes/regions (bold indicates significance after correction for the four tests, $P^{corrected} = 0.016$ for NS_{strict} MAF $< 0.1\%$ variants). This enrichment arises solely from the 3q29 locus (middle panel; bold indicates significance after correction of the 44 tests performed, $P^{corrected} = 0.024$ for 3q29). Genes and NS_{strict} case/control counts for the 3q29 region (bottom panel).

Extended Data Table 5 | Further stratification of enrichment analyses by class of variant

a

Primary geneset	Disruptive singletons P	n	A/U	OR	Damaging missense (strict) singleton P	n	A/U	OR
All primary genes	0.0008	905	852/716	1.20	0.0393	1357	2080/2001	1.04
SCZ *de novo* genes (refs. 18,19,30)								
Exome sequencing (disruptive)	0.0349	40	56/38	1.48	0.5613	53	121/120	1.01
Exome sequencing (nonsyn)	0.0059	332	384/309	1.25	0.2776	393	750/736	1.02
Copy number variants (refs. 5,13)								
de novo CNV genes	0.0224	64	61/40	1.53	0.0593	90	125/112	1.12
SCZ-associated CNV genes	0.3378	72	65/55	1.19	0.0310	111	148/119	1.25
GWAS (refs. 3,5,12)								
Voltage-gated calcium channel genes	0.0021	9	12/1	8.40	0.4629	18	37/35	1.06
Common SNPs (P < 1e-4 intervals)	0.1832	185	165/146	1.14	0.9246	268	359/395	0.91
miR-137 targets	0.6643	140	98/100	0.99	0.1415	263	376/361	1.05
Synaptic genes (ref. 13)								
PSD (human core)	0.0824	219	172/145	1.19	0.0070	394	646/581	1.12
ARC	0.0012	9	9/0	19.20	0.0069	19	32/15	2.14
NMDAR network	0.0162	17	17/5	3.42	0.0003	34	76/45	1.70
PSD-95	0.0018	16	17/3	5.10	0.0218	34	44/30	1.47
mGluR5	0.1335	10	9/3	3.02	0.1715	22	52/36	1.45

b

Geneset	Case-unique burden analysis P	n genes(A/U)	A(U)	Known variants Singletons P	n	MAF < 0.1% P	n	Novel variants Singletons P	n	MAF < 0.1% P	n
Composite	0.0006	829(275/214)	378(297)	0.3733	145	0.2202	226	0.0002	683	0.0005	744
Primary	0.0022	1026(325/265)	440(367)	0.1003	191	0.0417	299	0.0074	831	0.0058	910
SCZ *de novo* genes (refs. 18,19,30)											
Exome sequencing (disruptive)	0.0018	47(16/6)	29(12)	0.5514	13	0.2196	24	0.0362	35	0.0010	40
Exome sequencing (nonsyn)	0.0037	371(108/80)	159(116)	0.3647	94	0.2064	144	0.0142	302	0.0071	326
Copy number variants (refs. 5,13)											
de novo CNV genes	0.1267	79(25/17)	32(24)	0.0156	13	0.0116	24	0.0819	59	0.0679	67
SCZ-associated CNV genes	0.7971	90(20/23)	24(32)	0.0355	23	0.0069	34	0.6081	63	0.9187	76
GWAS (refs 3,5,12)											
Voltage-gated calcium channel	0.0129	9(5/1)	10(1)	0.4922	2	0.7077	3	0.0006	7	0.0022	8
P < 1e-4 intervals	0.1079	211(65/55)	91(78)	0.0394	44	0.0409	69	0.4425	164	0.1882	180
miR-137 targets	0.4498	156(52/50)	67(60)	0.9939	14	0.9972	22	0.3846	133	0.2757	147
Synaptic genes (ref. 13)											
PSD (human core)	0.2234	244(92/79)	113(109)	0.5348	25	0.3072	40	0.1091	205	0.1629	226
ARC	0.0008	9(9/0)	9(0)	.	0	.	0	0.0016	9	0.0013	9
NMDAR network	0.0105	21(13/4)	18(4)	1.0000	1	0.6905	2	0.0075	16	0.0085	19
PSD-95	0.0137	16(13/2)	14(2)	0.1218	1	0.1559	1	0.0034	15	0.0022	15
mGluR5	0.1363	11(7/1)	8(1)	.	0	0.1458	1	0.1427	10	0.1826	11
Secondary (autism/ID)	0.0916	1249(348/314)	479(471)	0.1679	226	0.3543	352	0.1834	1041	0.0807	1143
De novo genes (exome sequencing)											
Autism (disruptive) (refs. 22-25)	0.662	65(17/17)	20(26)	0.4463	18	0.0781	23	0.6161	50	0.7009	56
Autism (nonsyn) (refs. 22-25)	0.220	407(101/96)	143(154)	0.3198	89	0.4487	133	0.6656	336	0.4960	369
ID (disruptive) (refs. 20, 21)	0.262	8(4/1)	4(2)	1.0000	1	0.2747	3	0.3558	8	0.0578	8
ID (nonsyn) (refs. 20,21)	0.052	69(22/18)	35(28)	0.1303	14	0.0934	26	0.5331	62	0.3368	66
Neurodevelopmental candidates											
ASD candidates (ref. 39)	0.110	37(12/6)	16(7)	0.5824	9	0.7543	14	0.0484	24	0.0429	29
ID candidates (ref. 39)	0.994	88(14/28)	16(38)	0.6553	16	0.7056	24	0.9488	74	0.9556	82
Autism PPI networks											
CHD8 network (ref. 24)	1.000	1(0/1)	0(1)	.	0	.	0	1.0000	1	1.0000	1
49-gene network (ref. 24)	0.796	19(3/7)	4(16)	0.4755	5	0.7326	7	0.6081	30	0.7231	33
74-gene network (ref. 24)	0.654	33(6/13)	10(28)	0.6438	4	0.6667	4	0.7285	17	0.8411	19
Fragile X mental retardation protein targets											
Darnell (ref. 40) targets	0.022	341(131/95)	169(133)	0.3048	39	0.3889	61	0.0007	288	0.0022	309
Ascano (ref. 42) targets	0.449	517(134/131)	187(200)	0.5571	83	0.7281	128	0.5261	439	0.4089	482
Ascano (ref. 42) FMRP/autism	0.423	33(10/6)	12(12)	0.0384	5	0.4624	11	0.6088	23	0.3954	28

a, Gene set analyses for damaging missense mutations only. For the primary gene set and the 12 constituent subsets, a comparison of disruptive versus (strictly defined) damaging missenses, that is, an independent set of variants. The omnibus result for the primary test is modest ($P = 0.04$) and did not withstand correction for multiple testing: as illustrated in Fig. 1 and the main text, the bulk of the enrichment signal we observe comes from (singleton) disruptive mutations. Nonetheless, specific gene sets such as ARC and the NMDAR network are highly and independently enriched for missense variants. *N* represents the number of genes with at least one mutation of this class observed in the sample. A/U represent case/control counts of non-reference genotypes. OR represents the odds ratio (not corrected for exome-wide rates) estimated by Firth's method for sets with small cell counts. All tests are empirical and one-sided (higher values expected in cases) as described in the main text and Methods. **b**, Enrichment analyses of novel and case-unique disruptive mutations. For primary and secondary gene sets (and constituent subsets) as well as the composite set: results of alternative burden analyses. First, focusing only on genes without any control disruptive variants; no further frequency filter is imposed. Here 'N genes(A/U)' indicates the number of genes with at least one disruptive variant, followed by the number of genes with case-only disruptive mutations and (for comparison) the number with control-only disruptive mutations. The 'A(U)' column gives the number of case variants in the case-only genes: the test statistic is based on the empirical distribution of this count. The U in this field represents the similar quantity for controls (not explicitly used in the statistic). The second set of analyses represent standard burden/enrichment tests (that is, as Tables 1 and 2) but stratified for novel versus known disruptive variants, according to dbSNP and the Exome Sequencing Project/Exome Variant Server (ESP/EVS) database. Novel variants show greater enrichment, although most rare variants observed in our study (both in cases and in controls) are novel, so tests of novel variants will have greater power.

Extended Data Table 6 | Gene set analysis of *de novo* genes from schizophrenia exome-sequencing studies

a

		Disruptive			NS_{strict}			NS_{broad}		
Set	***n* genes**	**Singletons**	**MAF < 0.1%**	**MAF < 0.5%**	**Singletons**	**MAF < 0.1%**	**MAF < 0.5%**	**Singletons**	**MAF < 0.1%**	**MAF < 0.5%**
Fromer et al. (ref. 30) disruptive	63	0.1484	0.0075	0.0034	0.7401	0.7324	0.6264	0.3347	0.1660	0.2536
Fromer et al. (ref. 30) nonsyn	464	**0.0004**	**0.0003**	0.0016	0.0341	0.0057	0.0892	0.4688	0.4842	0.6547
Girard & Xu (refs. 18,19) disruptive	24	0.0342	0.0082	0.0082	0.0423	0.0412	0.0602	0.0774	0.1106	0.0891
Girard & Xu (refs. 18,19) nonsyn	151	0.6916	0.4124	0.4186	0.1510	0.1326	0.1285	0.2258	0.3420	0.1790
Combined SCZ disruptive	87	0.0319	**0.0007**	**0.0003**	0.3355	0.3162	0.2692	0.1325	0.0757	0.1016
Combined SCZ nonsyn	611	0.0053	0.0011	0.0055	0.0192	0.0024	0.0379	0.3408	0.4064	0.4472

b

Gene	***De novo* study (type)**	**Test**	***n***	**A/U**	***P***	**Gene name**
ALDH1L2	Fromer (ref. 30) nonsyn	disruptive	6	10/3	0.028	aldehyde dehydrogenase 1 family, member L2
CACNA1S	Fromer (ref. 30) nonsyn	NS_{strict}	23	28/15	0.031	calcium channel, voltage-dependent, L type, alpha 1S subunit
DLG1	Fromer (ref. 30) nonsyn	NS_{strict}	4	5/0	0.021	discs, large homolog 1 (Drosophila)
IGSF22	Fromer (ref. 30) nonsyn	disruptive	5	5/0	0.043	immunoglobulin superfamily, member 22
JARID2	Fromer (ref. 30) nonsyn	NS_{strict}	5	5/0	0.041	jumonji, AT rich interactive domain 2
LAMA4	Fromer (ref. 30) nonsyn	NS_{strict}	8	12/4	0.041	laminin, alpha 4
NBEA	Fromer (ref. 30) nonsyn	NS_{strict}	5	5/0	0.025	neurobeachin
POLL	Fromer (ref. 30) nonsyn	disruptive	4	4/0	0.042	polymerase (DNA directed), lambda
PTK2B	Fromer (ref. 30) nonsyn	NS_{strict}	4	4/0	0.044	PTK2B protein tyrosine kinase 2 beta
SHKBP1	Fromer (ref. 30) nonsyn	NS_{strict}	9	15/4	0.018	SH3KBP1 binding protein 1
SULF2	Fromer (ref. 30) nonsyn	NS_{strict}	5	7/0	0.007	sulfatase 2
SYNGAP1	Xu (ref. 19) LoF	NS_{strict}	4	4/0	0.043	synaptic Ras GTPase activating protein 1
SZT2	Xu (ref. 19) LoF	NS_{strict}	22	18/9	0.049	seizure threshold 2 homolog (mouse)
TANC1	Fromer (ref. 30) nonsyn	NS_{strict}	14	17/4	0.002	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
TEP1	Xu (ref. 19) nonsyn	NS_{strict}	26	25/14	0.048	telomerase-associated protein 1
UFL1	Fromer (ref. 30) LoF	NS_{strict}	4	7/0	0.008	UFM1-specific ligase 1
UFL1	Fromer (ref. 30) LoF	disruptive	2	5/0	0.029	UFM1-specific ligase 1

a, Test of case enrichment of rare variants in cases compared to controls, for genes with one or more *de novos* in Fromer *et al.*[30], Xu *et al.*[19] and/or Girard *et al.*[18]. The *P* values in bold are significant at $P^{corrected} < 0.05$, correcting for all $3 \times 3 \times 6 = 54$ tests reported. **b**, Genes nominally significant (no correction) that had an observed *de novo* in one of the schizophrenia studies.

Extended Data Table 7 | Summary of observed likely N.A, deleterious variants in ARC genes across studies

	Current study				
ARC gene (*n*=28)	**Disruptive**	**Damaging missense (strict)**	***de novo* CNV (Kirov et al., ref. 13)**	***de novo* SNV (Fromer et al., ref. 30)**	***de novo* SNV in ID (refs. 20,21)**
ACTN4		3/1			
ARF5					
ATP1A1		3/0			
ATP1A3		2/1			
ATP1B1	1/0	1/0			
BAIAP2	1/0			NS(x2)	
CAMK2A	1/0	1/0			
CRMP1		1/3			
CYFIP1	1/0	4/1	2 del; 2 dup		
DLG1	1/0	2/0	1 del	NS	
DLG2		2/3	2 del	LoF	
DLG4		1/2			NS
DLGAP1		2/0	1 del		
DLGAP2		1/0			
DPYSL2		0/1			
GLUD1	1/0	1/0			
GLUL		2/0			
GRIN1		2/0			
HSPA8				LoF & NS	
IQSEC1		4/1			
IQSEC2	1/0	0/1			LoF
MBP	1/0	0/1			
PKM2					
PLP1					
SLC25A3	1/0				
SLC25A4					
SLC25A5					
STXBP1					LoF, NS(x2)
Counts:	9 / 0	32 / 15	8 CNVs	6 SNVs	5 SNVs
***P*-value:**	0.0016	0.0069	0.00025	0.0005	0.00002

For the 28 ARC genes, a summary of which genes had singleton disruptive, or damaging missense, variants in the current study, compiled alongside the genes with *de novo* CNVs or SNVs observed in Kirov *et al.*[13] or Fromer *et al.*[30] as well as the intellectual disability (ID) *de novo* genes (compiled in Fromer *et al.*[30]). The *P* values at the bottom indicate that in each comparison the ARC gene set was significantly enriched.