SUPPLEMENTAL INFORMATION

De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia

Atsushi Takata[1,6,8], Iuliana Ionita-Laza[2], Joseph A. Gogos[3,4], Bin Xu[1,6,7,*] and Maria Karayiorgou[1,5,7,*]

[1]Department of Psychiatry, Columbia University Medical Center, New York, NY 10032, USA
[2]Department of Biostatistics, Columbia University Medical Center, New York, NY 10032, USA
[3]Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, USA
[4]Department of Physiology & Cellular Biophysics, Columbia University Medical Center, New York, NY 10032, USA
[5]New York State Psychiatric Institute, New York, NY 10032, USA
[6]These authors contributed equally to this work
[7]Co-senior authors
[8]Present address: Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Saitama, 351-0198, Japan

*Correspondence to Maria Karayiorgou (mk2758@cumc.columbia.edu) or Bin Xu (bx2105@cumc.columbia.edu)

SUPPLEMENTAL INVENTORY

SUPPLEMENTAL FIGURES

SUPPLEMENTAL TABLES

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

SUPPLEMENTAL TEXT

Different Experimental and Analytical Pipelines

2. Enrichment Analysis of *De Novo* Synonymous Mutations within the Fetal Brain-derived DNase I Footprints in SCZ and ASD
3. Enrichment Analyses of *De Novo* Synonymous Mutations Affecting miRNA Binding Sites in the Coding Regions, Codon Optimality and RNA Secondary Structure in SCZ and ASD
4. Confirmatory Analyses for Enrichment of Near-Splice Site Synonymous Mutations Changing ESR in ASD and Mutations within the Frontal Cortex-Derived DHS in SCZ
5. Enrichment Analyses of Various Types of *De Novo* Synonymous Mutations in Epileptic Encephalopathies (EE)

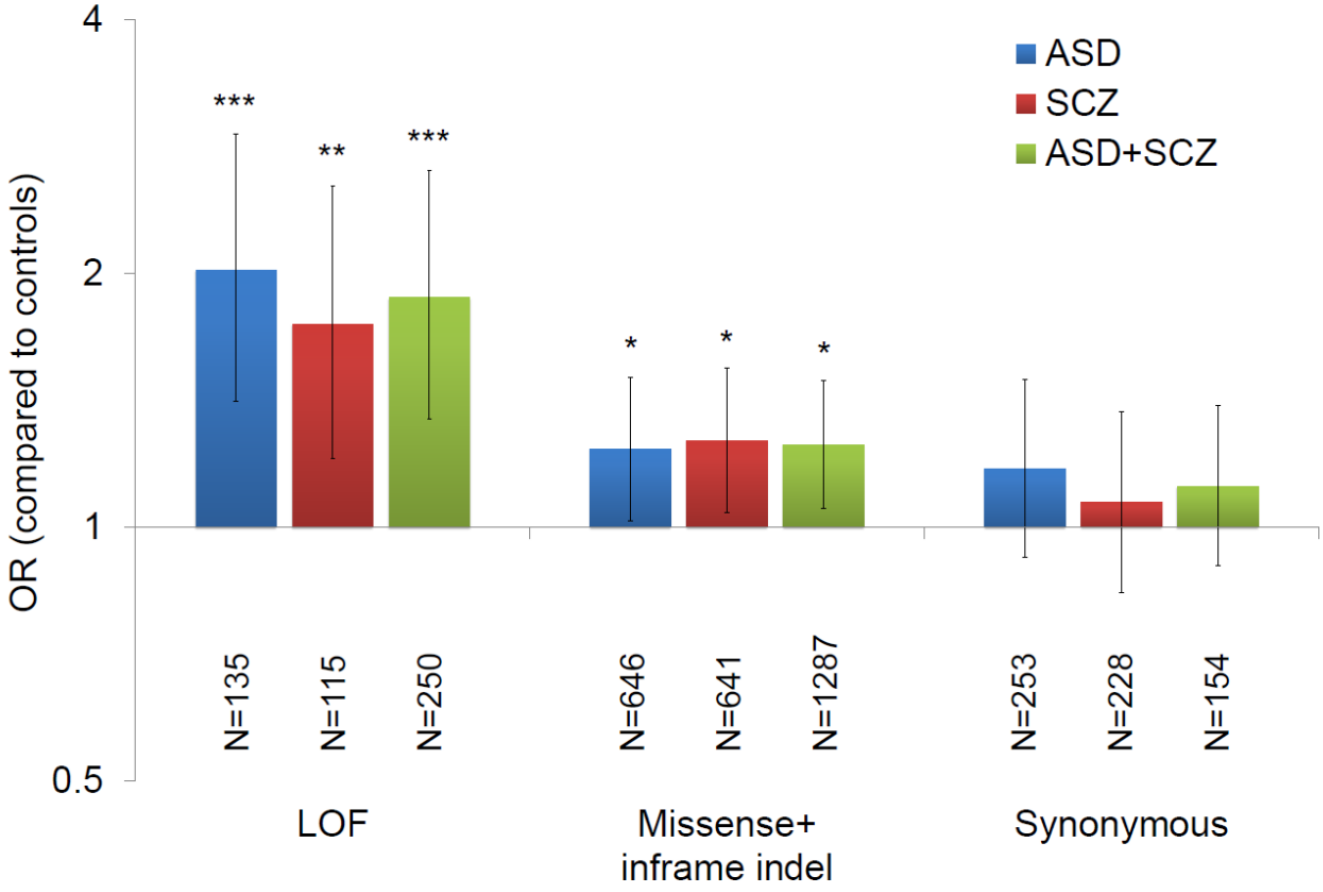

Figure S1. Global Enrichment Analysis of *De Novo* Mutations in ASD and SCZ, Related to Figures 2 and 3

Global enrichment of different types of *de novo* mutations (i.e. loss-of-function [LOF] mutations [nonsense, frameshift and canonical splice site mutations], missense and inframe indel mutations, as well as synonymous mutations) in ASD and SCZ was analyzed by using data from the following published family-based whole-exome sequencing studies: ASD: (Fromer et al., 2014; Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012); SCZ: (Fromer et al., 2014; Girard et al., 2011; Gulsuner et al., 2013; McCarthy et al., 2014; Xu et al., 2012b; Xu et al., 2011); and controls (including both unrelated healthy subjects and unaffected siblings): (Gulsuner et al., 2013; Iossifov et al., 2012; O'Roak et al., 2012; Rauch et al., 2012; Sanders et al., 2012; Xu et al., 2012b). The numbers below the bars indicate the number of each type of mutations. P values were calculated by 1-tailed Fisher's exact tests. Error bars indicate 95% CI. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$.

A

Study	# individuals	# synonymous mutations	Per individual # of mutations
Fromer et al	695	182	0.26
Iossifov et al	686	154	0.22
Sanders et al	438	66	0.15
Xu et al	265	32	0.12
ORoak et al	259	83	0.32
Gulsuner et al	189	49	0.26
Neale et al	175	50	0.29
McCarthy et al	54	15	0.28
Rauch et al	20	7	0.35
Girard et al	14	0	0.00

B

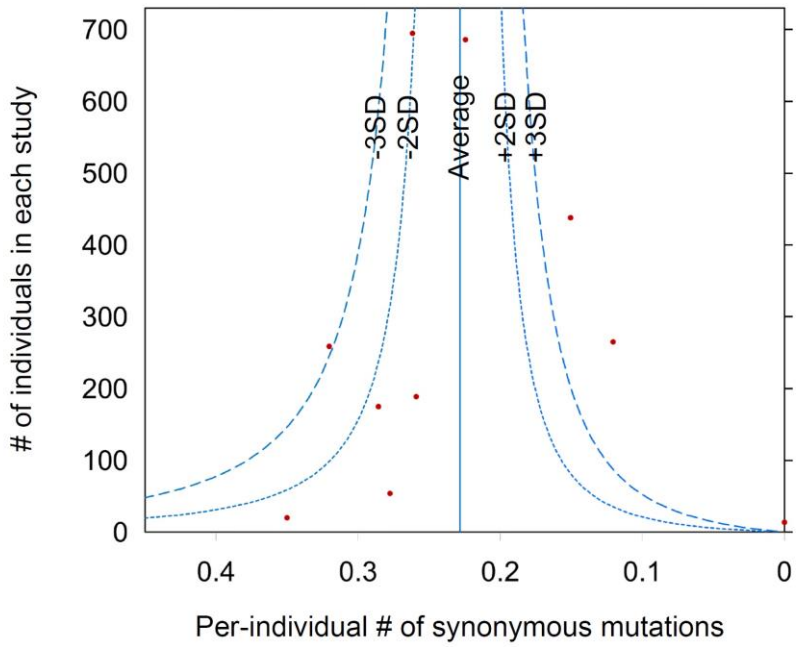

Figure S2. Per-individual Number of Synonymous Mutations in Each Study and a Funnel Plotting Analysis, Related to Figures 2 and 3

(**A**) Number of individuals sequenced, number of *de novo* synonymous mutations detected and per-individual number of *de novo* synonymous mutations in each study. (**B**) Funnel plotting of per-individual number of synonymous mutations in each study (red dots). We evaluated asymmetry of the plotting by a regression test and found no significant asymmetry (p = 0.65).

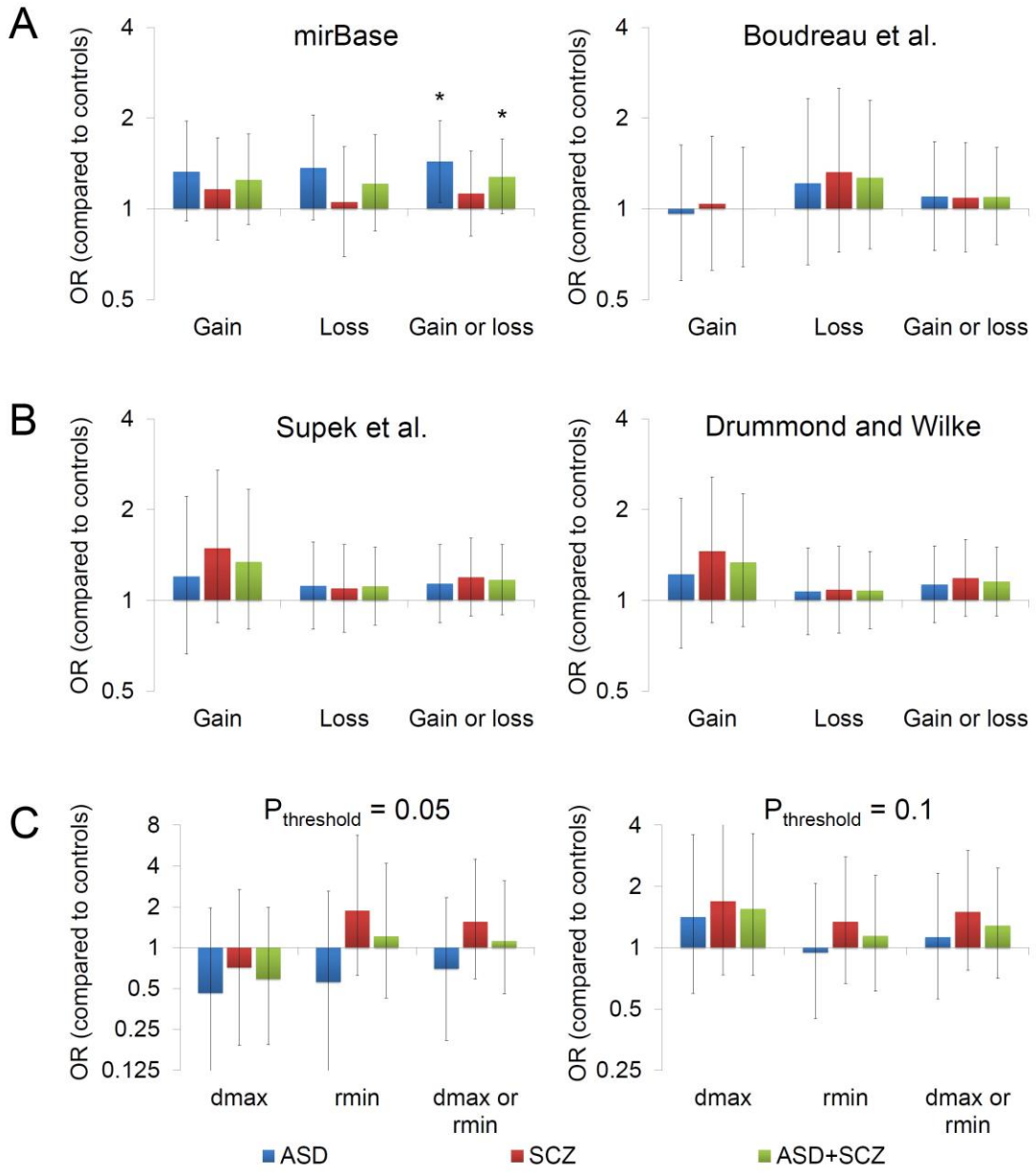

Figure S3. Enrichment Analyses of *De Novo* Synonymous Mutations Affecting miRNA Binding, Codon Optimality or RNA Secondary Structure, Related to Figures 2 and 3

Results of enrichment analyses of *de novo* synonymous mutations affecting miRNA binding sites in the coding regions (**A**), codon optimality (**B**) and RNA secondary structure (**C**) are shown. (**A**) MiRNA targets provided by mirBase (Kozomara and Griffiths-Jones, 2014) (left) and (Boudreau et al., 2014) (right) were used. (**B**) Optimal codons defined in (Supek et al., 2014) (left) and Drummond and Wilke(Drummond and Wilke, 2008) (right) were used. (**C**) Two different threshold p values (0.05 [left] and 0.1 [right]) were used. See also **Experimental Procedures** for details. P values were calculated by 1-tailed Fisher's exact tests. Error bars indicate 95% CI. * $p < 0.05$.

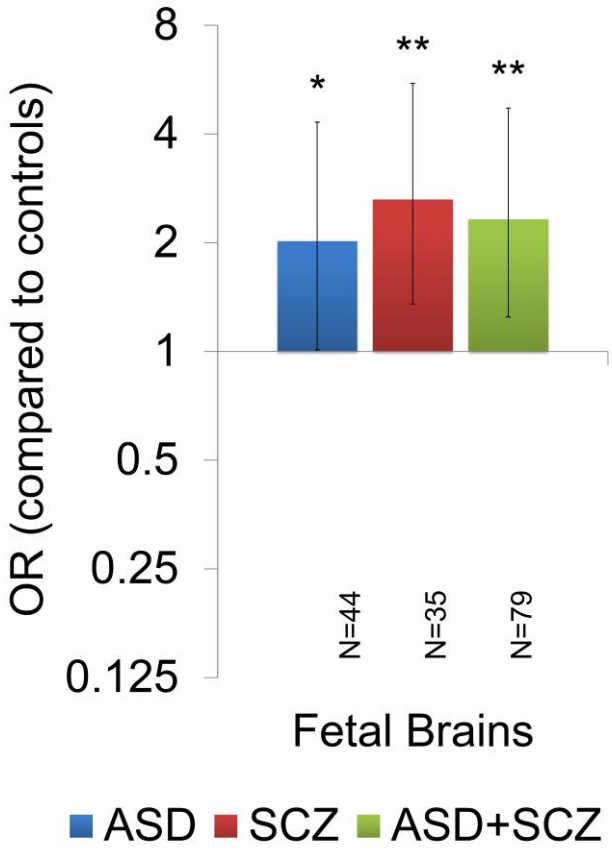

Figure S4. Enrichment Analysis of *De Novo* Synonymous Mutations within DNase I hypersensitive sites using the Roadmap Epigenomics datasets, Related to Figure 3

Enrichment analysis of *de novo* synonymous mutations within brain-derived DHS using an independent dataset from the Roadmap Epigenomics Project (Roadmap Epigenomics et al., 2015) (IDs: E081 and E082, both derived from fetal brains). The numbers below the bars indicate the number of each type of synonymous mutations. Error bars indicate 95% CI. P values were calculated by 1-tailed 2×2 Fisher's exact tests. *$p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$. ASD; autism spectrum disorders, OR; odds ratio, SCZ; schizophrenia, SS; splice site.

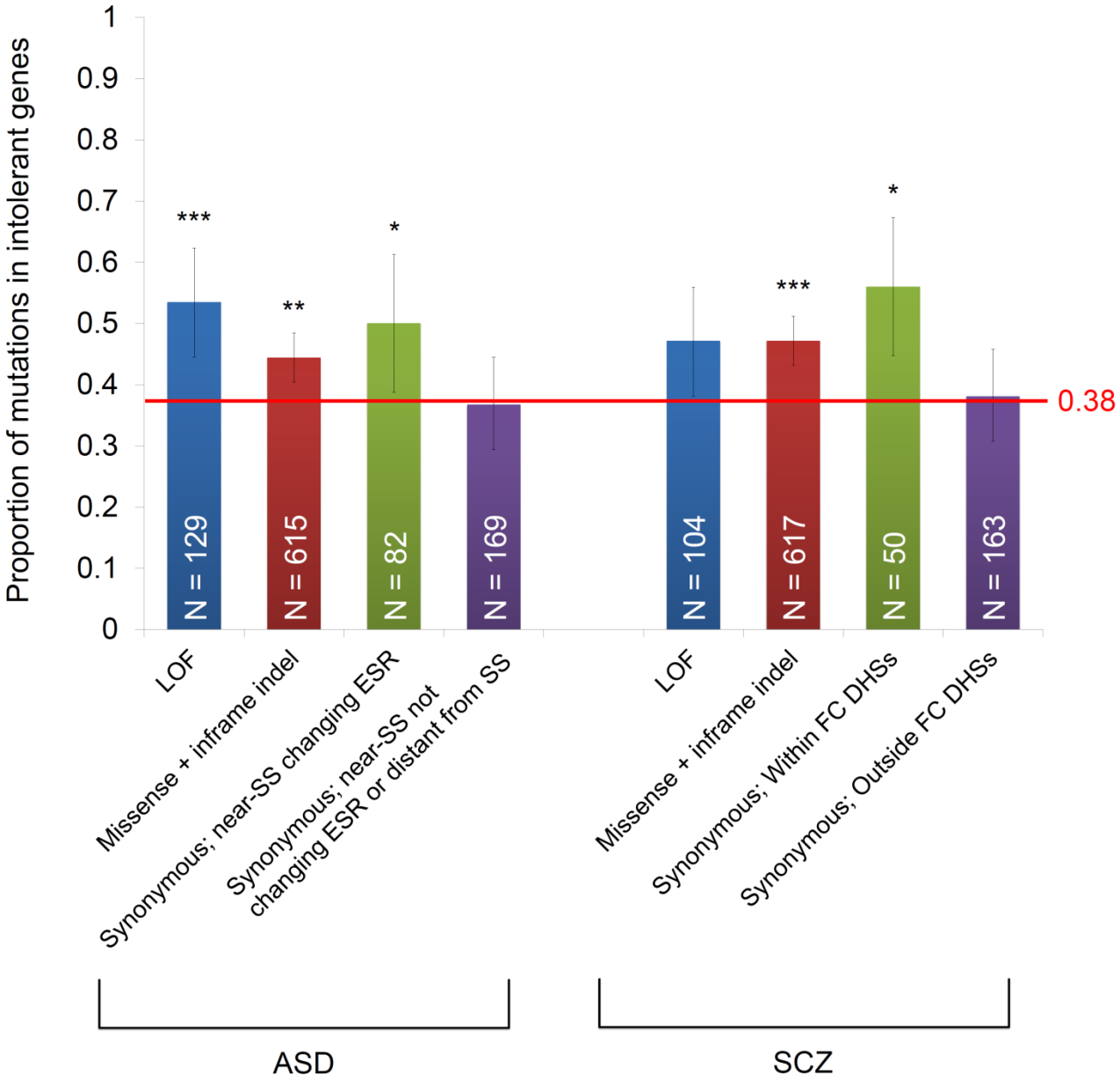

Figure S5. Proportion of *De Novo* Mutations in Genes Intolerant to Functional Genetic Variation, Related to Figures 2 and 3

Bars indicate the proportion of *de novo* mutations found in genes intolerant to functional variation (within the most intolerant quartile), which were defined by using Residual Variation Intolerance Score (RVIS) (Petrovski et al., 2013). P values were calculated by binominal exact tests with the hypothesized probability of success = 0.38, which is the theoretical likelihood of finding *de novo* mutations in the first 25th percentile of the most intolerant genes, considering the gene size (red line) (Petrovski et al., 2013). Error bars indicate 95% CI. For each category of mutations, the number of mutations for which RVIS is available is shown at the bottom of the bar. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$. DHS: DNase I hypersensitive site; ESR: exonic splicing regulator; FC: frontal cortex; LOF: loss-of-function; SS: splice site.

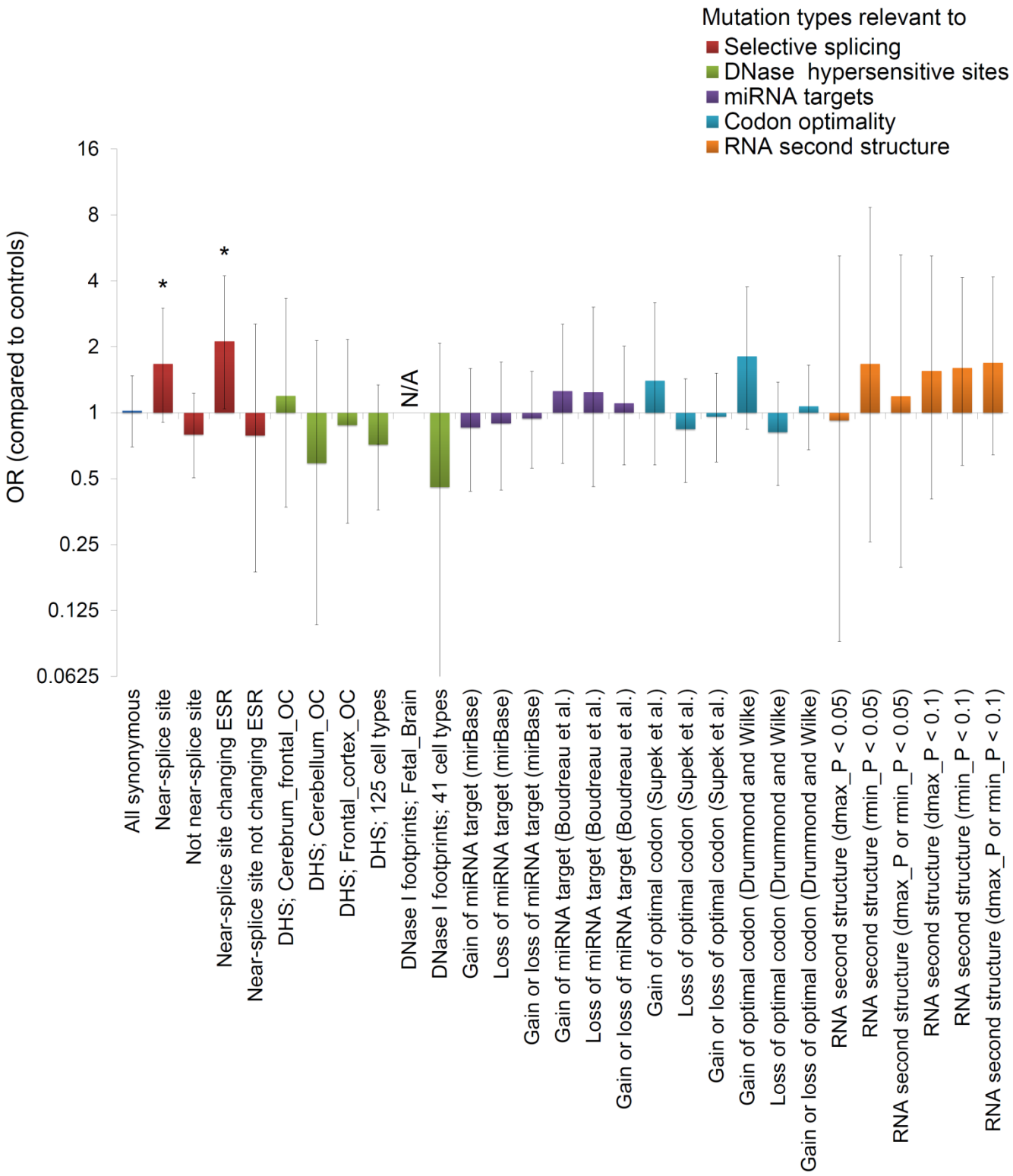

Figure S6. Enrichment Analyses of *De Novo* Synonymous Mutations in Epileptic Encephalopathies (EE), Related to Figures 2 and 3

Results of enrichment analyses of various types of *de novo* synonymous mutations found in EE (Epi4K_Consortium et al., 2013) are shown. P values were calculated by 1-tailed Fisher's exact tests. Error bars indicate 95% CI. An odds ratio (OR) for mutations in fetal-brain derived DNase I footprints for which no mutation was found in the case group is not displayed (shown as N/A). * $p < 0.05$.

Table S1A. List of the Whole Exome Sequencing Studies from Which *De Novo* Mutation Data Were Derived for This Study, Related to Figure 1

Table S1B. List of the *De Novo* Synonymous Mutations Used in This Study with Corresponding Annotations, Related to Figure 1

These tables are provided as a separate spreadsheet file

Table S2. Summary of Enrichment Analyses for Various Types of *De Novo* Synonymous Mutations, Related to Figures 2 and 3

This table is provided as a separate spreadsheet file

Table S3. Results of Functional Gene-set Enrichment Analyses, Related to Figures 2 and 3

Category of the term	ID of the term	Description	Source	BH-corrected P value	Hit Count in Query List	Hit Count in Genome	Hit in Query List
Genes with near-splice site *de novo* synonymous mutations changing ESR in ASD (*N* = 82)							
GO: Molecular Function	GO:0005245	voltage-gated calcium channel activity		6.20E-03	4	35	*CACNA2D2,CACNA1A,CACNA1G,CACNA1E*
GO: Molecular Function	GO:0015085	calcium ion transmembrane transporter activity		4.62E-02	5	127	*CACNA2D2,CACNA1A,SLC8A3,CACNA1G,CACNA1E*
GO: Cellular Component	GO:0005891	voltage-gated calcium channel complex		4.58E-03	4	36	*CACNA2D2,CACNA1A,CACNA1G,CACNA1E*
GO: Cellular Component	GO:0034704	calcium channel complex		9.99E-03	4	52	*CACNA2D2,CACNA1A,CACNA1G,CACNA1E*
GO: Cellular Component	GO:0097458	neuron part		4.78E-02	13	1053	*ITGA2,MYO5B,SLC17A6,RAB2A,CACNA1A,SLC8A3,CACNA1G, CACNA1E,DBNL,NFASC,PTPRN2,LPHN1,GRIK5*
Mouse Phenotype	MP:0008840	abnormal spike wave discharge		4.46E-03	4	19	*CACNA2D2,CACNA1A,CACNA1G,KCNH3*
Mouse Phenotype	MP:0003216	absence seizures		1.46E-02	4	32	*CACNA2D2,CACNA1A,CACNA1G,KCNH3*
Mouse Phenotype	MP:0000753	paralysis		1.46E-02	6	109	*SCYL1,GALC,CACNA2D2,CACNA1A,DBNL,NFASC*
Mouse Phenotype	MP:0000948	nonconvulsive seizures		1.76E-02	4	37	*CACNA2D2,CACNA1A,CACNA1G,KCNH3*
Pathway	106517	Depolarization of the Presynaptic Terminal Triggers the Opening of Calcium Channels	BioSystems: REACTOME	7.71E-03	3	12	*CACNA2D2,CACNA1A,CACNA1E*
Pathway	P00037	Ionotropic glutamate receptor pathway	PantherDB	1.01E-02	4	46	*SLC17A6,CACNA1A,CACNA1E,GRIK5*
Pathway	106096	Integration of energy metabolism	BioSystems: REACTOME	1.45E-02	5	108	*ACLY,CACNA2D2,CACNA1A,CACNA1E,PPP2R1B*
Pathway	169352	Regulation of Wnt-mediated beta catenin signaling and target gene transcription	BioSystems: Pathway Interaction Database	4.21E-02	4	79	*TCF4,TNIK,CACNA1G,HDAC2*
Gene Family	CACN	Calcium channels	genenames.org	2.14E-07	4	17	*CACNA2D2,CACNA1A,CACNA1G,CACNA1E*
Genes with *de novo* synonymous mutations distant from the nearest splice site (> 30 bp) or near-splice site mutations not changing ESR in ASD (*N* = 171)							
Gene Family	MUC	Mucins	genenames.org	2.89E-03	3	20	*MUC5B,MUC17,MUC4*
Genes with near-splice site *de novo* synonymous mutations changing ESR in controls (*N* = 23)							
		No significantly enriched term (BH-corrected p value < 0.05)					
Genes with *de novo* synonymous mutations within the frontal cortex-derived DHS in SCZ (*N* = 60)							
		No significantly enriched term (BH-corrected p value < 0.05)					

ASD: autism spectrum disorders; BH: Benjamini–Hochberg; DHS: DNase I hypersensitive sites; ESR: exonic splicing regulator; SCZ: schizophrenia

Table S4. Disease Liability Explained by the Different Types of *De Novo* Mutations, Related to Figures 2 and 3

Disease	Mutation Type	Liability explained	Per-individual mutation rate	Relative risk compared to controls
ASD	Loss-of-function	1.30%	0.129	1.93
	Missense and inframe indels	0.10%	0.619	1.18
	Near-splice site synonymous (all)	1.10%	0.097	1.91
	Near-splice site synonymous changing ESR	1.50%	0.079	2.50
SCZ	Loss-of-function	0.74%	0.113	1.68
	Missense	0.10%	0.628	1.20
	Synonymous within frontal cortex-derived DHS	0.60%	0.059	1.87

ASD; autism spectrum disorders, DHS; DNase I Hypersensitive site, ESR; exonic splicing regulator, SCZ; schizophrenia

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

List of *De Novo* Synonymous Mutations

Information for *de novo* synonymous mutations in ASD, SCZ and control subjects was obtained from the following whole-exome sequencing studies; ASD: (Fromer et al., 2014; Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012); SCZ: (Fromer et al., 2014; Girard et al., 2011; Gulsuner et al., 2013; McCarthy et al., 2014; Xu et al., 2012b; Xu et al., 2011); and controls (including both unrelated healthy subjects and unaffected siblings): (Gulsuner et al., 2013; Iossifov et al., 2012; O'Roak et al., 2012; Rauch et al., 2012; Sanders et al., 2012; Xu et al., 2012b). The list of these *de novo* synonymous mutations along with annotations for their potential impact on regulatory elements is provided in **Table S1**.

Evaluation of Enrichment for Near-Splice Site Synonymous Mutations

Information for the distance to the nearest splice site for each mutation was annotated by using SeattleSeq Annotation 138 (http://snp.gs.washington.edu/SeattleSeqAnnotation138/). To evaluate enrichment of synonymous mutations in the case groups, we first randomly distributed observed number of mutations among each group for 1,000 times and estimated the number of individuals carrying one or more mutation. After that, enrichment was evaluated by using 1-tailed 2×2 Fisher's exact tests (rows: cases and controls, columns: number of individuals carrying and not carrying mutations). Kolmogorov-Smirnov tests were performed to compare the distribution of the distance to the nearest splicing site for synonymous mutations between cases and controls.

Analysis of Near-Splice Site Synonymous Mutations Predicted to Change Exonic Splicing Regulators

To identify near-splice site synonymous mutations likely to change exonic splicing

enhancer and silencer (ESE and ESS) sequences, we used the following lists of predicted ESEs and ESSs; 1) the 200 hexamers with the highest "ESEseq" score and the 200 hexamers with the lowest "ESSseq" score in (Ke et al., 2011), in which ESE and ESS activity of all possible hexamers (4,096 patterns) were experimentally tested using human cells, 2) 238 ESEs in the RESCUE-ESE server (http://genes.mit.edu/burgelab/rescue-ese/), which provides a list of ESE candidates predicted by a computational analysis of exon-intron and splice site composition (Fairbrother et al., 2002) and 3) 176 ESSs (FAS-hex2 dataset) in the FAS-ESS server (http://genes.mit.edu/fas-ess/), which provides a list of ESS candidates identified by an *in vivo* screening using a splicing reporter system (Wang et al., 2004). We used the list of ESE and ESS sequences found in either one of these three datasets as the reference (in total 401 ESEs and 315 ESSs). We compared the sequences of 6 bp windows including the mutation site in the transcribed product (six hexamers can be generated for each mutation; hexamers at the positions -5~0, -4~+1, -3~+2, -2~+3, -1~+4, and 0~+5 when the mutation is at position 0) to the reference lists of ESE and ESS sequences, and assessed whether each mutation causes at least one gain or loss of ESE/ESS events. We considered all synonymous mutations that cause any type of exonic splicing regulator (ESR) changes (i.e. ESE gain, ESE loss, ESS gain or ESS loss) as the mutations changing ESR. Enrichment of synonymous mutations changing or not changing ESR in the case groups was evaluated by using 1-tailed 2×2 Fisher's exact tests after randomly distributing observed number of mutations among the corresponding population (same procedure as described in **Evaluation of Enrichment for Near-Splice Site Synonymous Mutations**).

Machine-Learning Based Prediction of Impact of Synonymous Mutations on Splicing Regulation

Prediction of impact of synonymous mutations on splicing regulation was performed by

using SPANR (Splicing-based Analysis of Variants; available at http://tools.genes.toronto.edu/) (Xiong et al., 2015). Predicted impact of synonymous mutations on splicing events was evaluated by absolute values of difference in percent of transcripts with the exon spliced-in (PSI: percent-spliced-in) between reference and mutation alleles (|dPSI|). The averages of |dPSI| for synonymous mutations in cases and controls were compared by 1-tailed t-test to test whether synonymous mutations in cases show significantly higher |dPSI| than those in controls. Analyses were performed with all the synonymous mutations (not restricted to the mutations within 30 bp of the nearest splice site because SPANR can analyze mutations that are within 300 bp of the nearest splice site (Xiong et al., 2015)).

Analysis of Synonymous Mutations within DNase I Hypersensitive Sites (DHS) and DNase I Footprints

To evaluate enrichment of synonymous mutations in DHS, the following data were downloaded from the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables); the wgEncodeRegDnaseClusteredV2 table (integrated data derived from 125 cell types) from the DNase Clusters track, and the Cerbrm frnt Pk table (derived from Cerebrum_frontal_OC: primary frozen frontal cerebrum tissue from NICHD donor IDs 1104 [Rep B1], 602 [Rep B2], 1442 [Rep B3], all Caucasian), the Cerbllm Pk table (derived from Cerebellum_OC: primary frozen cerebellum tissue from NICHD donor IDs 1104 [Rep B1], 602 [Rep B2], 1442 [Rep B3], all Caucasian) and the Front crtx Pk table (derived from Frontal_cortex_OC: primary ventromedial prefrontal cortex, from KPBBB donor IDs 673 [Rep B1] and 913 [Rep B2]) from the Duke DNaseI HS track. To search for data derived from normal brain tissues for which DNase-seq data are available, information in the ENCODE Experiment Matrix (https://genome.ucsc.edu/ENCODE/dataMatrix/encodeDataMatrixHuman.html) and the ENCODE Common Cell Types table

(https://genome.ucsc.edu/encode/cellTypes.html) was utilized. We did not use data from neuronal cancer cells, astrocytes and brain vascular cells. Data of DNase I footprints derived from 41 cell types (Neph et al., 2012) were downloaded from the EMBL European Bioinformatics Institute ftp server (ftp://ftp.ebi.ac.uk/pub/databases/ensembl/encode/supplementary/integration_data_jan2011/byDataType/footprints/jan2011/). This dataset includes one brain derived data (fBrain: data derived from fetal brain tissues). Synonymous mutations intersecting DHS or DNase I footprints were identified by using BEDtools (Quinlan and Hall, 2010). Enrichment of synonymous mutations in DHS or DNase I footprints in the case groups was evaluated by using 1-tailed 2×2 Fisher's exact tests after randomly distributing observed number of mutations among the corresponding population. For synonymous mutations found within the fetal brain-derived DNase I footprints, we analyzed whether these mutations preferentially hit known TF binding motifs. For this purpose, we used TF binding motifs listed in the Factorbook, an online registry of TF binding data led by the ENCODE project (Wang et al., 2012). Enrichment of mutations hitting TF binding sites was tested by 1-tailed 2×2 Fisher's exact test (rows: numbers of mutations within or outside the footprints, columns: numbers of mutations hitting or not hitting known TF binding sites; in this analysis we did not consider case-control status for mutations, while all mutations within DNase I footprints were identified in case subjects). An additional enrichment analysis of synonymous mutations within brain-derived DHS was performed by using independent datasets available in the Roadmap Epigenomics Project (Roadmap Epigenomics et al., 2015) (http://egg.wustl.edu/roadmap/web_portal/index.html). In this project there are two datasets derived from brain tissues for which original DNase-seq data are available (IDs: E081 and E082, both derived from fetal brains). Enrichment of synonymous mutations within brain-derived DHS in SCZ and ASD was tested by using the imputed gapped peak call datasets for these tissues (E081-DNase.imputed.gappedPeak.bed and

E082-DNase.imputed.gappedPeak.bed). We considered mutations within either of the E081 or E082 peak calls as mutations within DHS.

Analysis of Synonymous Mutations Changing miRNA Binding Site Sequences

To evaluate enrichment of synonymous mutations changing miRNA binding sites in the coding regions, we used lists of miRNA seed sequences defined by 1) high confidence miRNAs in mirBase (Kozomara and Griffiths-Jones, 2014) (# of unique target sequences = 815) and 2) miRNA binding sites in human postmortem brains that were identified by using high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) (Boudreau et al., 2014) (# of unique target sequences = 382; derived from Table S5 of the corresponding publication). We compared the sequences of 7 bp windows including the mutation site in the transcribed product (seven heptamers can be generated for each mutation; heptamers at the positions -6~0, -5~+1, -4~+2, -3~+3, -2~+4, -1~+5 and 0~+6 when the mutation is at the position 0) and miRNA seed heptamers (7m8: positions 2-8 of processed miRNAs and 7A1: positions 2–7 followed by an adenine) in the reference datasets described above, and evaluated whether a synonymous mutation can cause at least one loss or gain of miRNA binding site events. Enrichment of synonymous mutations changing miRNA binding sites (loss, gain or loss/gain) in the case groups was evaluated by using 1-tailed 2×2 Fisher's exact tests after randomly distributing observed number of mutations among the corresponding population.

Analysis of Synonymous Mutations Altering Codon Optimality

To determine whether synonymous mutations cause loss or gain of the optimal codons, we used definition of the optimal codons described in (Supek et al., 2014) (Table S6 of the corresponding publication) and (Drummond and Wilke, 2008) (Table S2 of the corresponding publication). For each mutation, triplet codon sequences for reference

and variant alleles in the transcribed product were annotated by using Variant Effect Predictor (http://www.ensembl.org/Homo_sapiens/Tools/VEP) (McLaren et al., 2010). As expected, optimal codons were much more frequently found in the reference alleles (335/701 in the reference and 115/701 in the variant alleles with the definition in Supek et al., $p = 4.0 \times 10^{-37}$, and 369/701 in the reference and 169/701 in the variant alleles with the definition in Drummond and Wilke, $p = 2.8 \times 10^{-28}$, 2-tailed 2×2 Fisher's exact test). Enrichment of synonymous mutations causing loss, gain or loss/gain of the optimal codons in the case groups was evaluated by using 1-tailed 2×2 Fisher's exact tests after randomly distributing observed number of mutations among the corresponding population.

Evaluation of Impact of Synonymous Mutations on RNA Secondary Structures

To predict impact of synonymous mutations on RNA secondary structures, we used RNAsnp (Sabarinathan et al., 2013) with default parameters and the Mode 1, which compute the effect of variants by using global folding. Sequences and 5' UTR lengths of mRNAs encoded by genes with synonymous mutations were obtained from the UCSC Genome Browser (https://genome.ucsc.edu/). The position of each mutation in mRNA was calculated with the obtained 5' UTR length and the coding position annotated by SeattleSeq Annotation 138 (http://snp.gs.washington.edu/SeattleSeqAnnotation138/). We used information for the longest transcripts when multiple isoforms existed. We used the p value calculated with maximum Euclidean distance (d_{max}) and the p value calculated with minimum Pearson correlation coefficient (r_{min}) with two different thresholds ($p < 0.05$ and 0.1) to assess the significance. Enrichment of synonymous mutations likely to change RNA secondary structures in the case groups was evaluated by using 1-tailed 2×2 Fisher's exact tests after randomly distributing observed number of mutations among the corresponding population.

Permutation Analyses Testing Enrichment of Near-Splice Site Synonymous Mutations Changing ESR in ASD and Mutations within the Frontal Cortex-Derived DHS in SCZ

We performed permutation analyses by randomly shuffling case-control status one million times. As a part of the data for ASD subjects in (Fromer et al., 2014) (ARRA_Broad_wave3 dataset in Supplementary Table 3 of the corresponding paper) has no information at the individual sample level (i.e. which individual has which mutations), we excluded these mutations (26 mutations including seven near-splice site synonymous mutations changing ESR) from the permutation analyses.

Enrichment Analysis of Mutations in Genes Intolerant to Functional Variation

The Residual Variation Intolerance Score (RVIS), which reflects gene intolerance to functional mutations, was obtained from (Petrovski et al., 2013). We considered the first 25^{th} percentile of the genes with the smallest RVIS (in this scoring system, genes with smaller scores are more intolerant to functional genetic variation) as "intolerant" genes, because this threshold was shown to be useful in enriching for genes associated with neurodevelopmental diseases (Petrovski et al., 2013). Mutations for which RVIS was not available were excluded from the analysis. Enrichment of mutations in intolerant genes were evaluated by binominal exact tests with the hypothesized probability of success = 0.38, which is the theoretical likelihood of finding *de novo* variants in intolerant genes, considering the gene size (Petrovski et al., 2013).

Functional Gene-Set Enrichment Analysis of Genes with Potentially Functional Synonymous Mutations

Functional gene-set enrichment analysis was performed by using the ToppGene Suite (Chen et al., 2009) with default parameters for the following categories of functional annotations; GO: Molecular Function; GO: Biological Process; GO: Cellular Component, Mouse Phenotype, Human phenotype, Pathway and Gene Family. Terms with two or

less hit counts in the query list were excluded from the results. Additional enrichment analysis of synaptic genes was performed by using lists of synaptic genes in Synaptome Database (Pirooznia et al., 2012) (including 1875 RefSeq genes in all synaptic regions, 1744 genes in postsynapse, 336 genes in presynapse and 209 genes in the active zone) and Genes to Cognition (G2C) Synapse Proteomics Datasets (Bayes et al., 2011) (including 1420 RefSeq genes in postsynapse). Significance of the enrichment was assessed by performing 1-tailed 2×2 Fisher's exact tests (rows: numbers of synaptic and non-synaptic genes among genes hit by mutations, columns: numbers of synaptic and non-synaptic genes among all RefSeq genes within the target regions of SeqCap EZ Human Exome Library v2.0 [18,901 genes in total]).

Estimating the Relative Contribution of the Different Types of *De Novo* Mutations to Variability on the Liability Scale

Using a threshold liability model, we estimated the contribution to variability on the liability scale (Gaugler et al., 2014; Witte et al., 2014; Wray et al., 2010) of various classes of functional *de novo* mutations. Results are reported in **Table S4** along with per-individual mutation rates and relative risks compared to controls. Note that our estimates for *de novo* LOF and missense mutations in ASD are very similar to the ones in (Gaugler et al., 2014).

Interpretation of Genes with Multiple Functional *De Novo* Mutations Including Synonymous Mutations Affecting Regulatory Elements

Significance of the observation of multiple functional *de novo* mutations including synonymous mutations potentially affecting regulatory elements in the same gene in the current dataset was evaluated according to a recently developed framework for the interpretation of *de novo* mutations (Samocha et al., 2014). Probability of observing each type of mutation (i.e. nonsense, splice site, frameshift, missense and synonymous

mutation) in each gene was obtained from Supplementary Table 2 of (Samocha et al., 2014). Probability of observing potentially functional synonymous mutations was calculated by multiplying the mutation rate for all synonymous mutations described in Supplementary Table 2 of (Samocha et al., 2014) and the proportion of potentially functional synonymous mutations among all synonymous mutations in the control subjects of the current dataset (23/154 for the near-splice site synonymous mutations changing ESR and 23/154 for the synonymous mutations within the frontal cortex-derived DHS, those are the same by coincidence). Expected numbers of functional *de novo* mutations were then calculated from the gene-based probability of mutations and the number of subjects in the current dataset (1,043 for ASD and 1,021 for SCZ). The observed and expected numbers of functional *de novo* mutations were compared and Poisson distribution probabilities were invoked to determine the statistical significance. For genes with potentially functional synonymous mutations and LOF mutations, we used the probability of observing LOF or potentially functional synonymous mutations to calculate significance (we considered that potentially functional synonymous mutations can be as damaging as LOF mutations according to the observed ORs for potentially functional synonymous mutations). For genes with potentially functional synonymous mutations and missense mutations, we used the probability of observing LOF, missense or potentially functional synonymous mutations. Genome-wide correction for multiple testing was performed using the number of the genes in the table described above (Supplementary Table 2 of (Samocha et al., 2014), *N* = 18,271).

Joint Analysis Using New Large-Scale Datasets for ASD

Newly published large-scale datasets of *de novo* mutations in ASD for the joint analysis were obtained from Supplementary Table 2 of (Iossifov et al., 2014) and Supplementary Table 3 of (De Rubeis et al., 2014). Because sample overlap between

these new studies (De Rubeis et al., 2014; Iossifov et al., 2014) and previous studies (Fromer et al., 2014; Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012) could not always be verified at the individual sample level, we compared the numbers of near-splice site synonymous mutations changing ESR and the other synonymous mutations in cases and controls with 1-tailed Fisher's exact test to evaluate enrichment of these mutations in ASD. Distance to the nearest splice site and impact on ESR for each mutation were determined as described above. We combined the list of *de novo* synonymous mutations in these two new datasets with the mutations in **Table S1** and removed any redundant ones. Redundancy was determined based on position of the mutation (since the probability of having >1 *de novo* mutations in the same position is very low in current sample sizes) and, additionally, on family ID when available. The combined dataset comprises 1,562 synonymous mutations (1,046 in ASD subjects and 516 in controls), which is three to four times larger than the initial dataset (253 in ASD subjects and 151 in controls).

SUPPLEMENTAL TEXT

Consideration of Potential Biases Due to Combining Data from Studies Using Different Experimental and Analytical Pipelines

In this study we combined the data of synonymous mutations in different whole-exome sequencing (WES) studyies to analyze enrichment of potentially functional synonymous mutations in cases. Considering that different studies may use different experimental and analytical pipelines, possible biases due to combining the data from multiple studies should be taken into consideration. Indeed, there was some variation in per-individual mutation numbers among these studies (**Figure S2A**). However, when we evaluated funnel plot asymmetry by plotting the per-individual number of synonymous mutations in each study, there was no significant asymmetry ($p = 0.65$, calculated by a regression test with R software, **Figure S2B**), suggesting that per-individual number of synonymous mutations in each study is evenly distributed according to the sample sizes and there is no particular study confounding the result.

In addition, if enrichment of mutations in case groups is explained by potential biases due to combining data from multiple studies (e.g. in a case that the studies sequenced more case subjects used more relaxed variant call criteria and thereby the number of mutations in cases is greater than controls in the combined data set), there should be global enrichment of all types of mutations when the number of mutations (hence the statistical power) is sufficient. However, this was not the case. We actually observed specific enrichment of near-splice site synonymous mutations in ASD and SCZ, while there was no significant enrichment of synonymous mutations distant from the nearest splice site or when synonymous mutations were considered all together (**Figures 2** and **S1**). In addition, the number of synonymous mutations distant from the nearest splice site is greater than the number of near-splice site mutations in all groups (**Table S2**). Therefore, the analysis for synonymous mutations distant from the nearest splice site should be more statistically powerful, but indeed we did not detect

significant enrichment of these mutations. Furthermore, we observed tissue- or brain region-specific enrichment of synonymous mutations within DNase I hypersensitive sites (DHS) and DNase I footprints (**Figure 2**). The observed results (frontal cortex- and fetal brain-specific enrichment) are in accordance with our current understanding of the pathophysiology of ASD and SCZ likely involving dysfunction of neural circuits in the frontal cortex in early developmental stages (Insel, 2010; State and Sestan, 2012), further suggesting that the observed enrichment of particular subsets of synonymous mutations is unlikely to be explained by technical biases. We, therefore, conclude that our main findings cannot be explained by any potential biases that might exist due to combining data from studies using different experimental and analytical pipelines.

Enrichment Analysis of *De Novo* Synonymous Mutations within the Fetal Brain-derived DNase I Footprints in SCZ and ASD

We evaluated enrichment of *de novo* synonymous mutations within DNase I footprints, i.e. depressions of DNase signals within DHS that indicate the precise genomic location bound by regulatory factors (Kavanagh et al., 2013). We did not see enrichment of synonymous mutations within the footprints in any of the case groups when the analysis was based on the integrated data of DNase I footprints derived from 41 cell types (Neph et al., 2012); however, similar to the analysis of mutations in DHS, when we focused on the data derived from fetal brain tissues (one of the 41 cell types used above), we observed significant enrichment of synonymous mutations within DNase I footprints in SCZ ($p = 0.012$, 8 in cases and none in controls) and ASD+SCZ ($p = 0.039$, 11 in cases and none in controls). While the small number of mutations found in these regions limits the statistical power to robustly detect enrichment, it is notable that the *de novo* synonymous mutations within the fetal brain-derived DNase I footprints were highly enriched for mutations that also hit known TF binding motifs ($p = 2.3 \times 10^{-5}$, 1-tailed 2×2 Fisher's exact test, 4/11 mutations within the DNase I footprints and

10/624 mutations outside the DNase I footprints hit the TF binding motifs, see **Supplemental Experimental Procedures**), suggesting potential impact of these mutations on the TF binding affinity in brain tissues.

Enrichment Analyses of *De Novo* Synonymous Mutations Affecting miRNA Binding Sites in the Coding Regions, Codon Optimality and RNA Secondary Structure in SCZ and ASD

Various lines of evidence have supported contribution of miRNA dysregulation to the pathogenesis of neuropsychiatric diseases (Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Stark et al., 2008; Xu et al., 2012a; Xu et al., 2013). Although miRNAs typically bound with 3' UTR sequences cannot be analyzed in the currently available exome sequencing data, it was also recently reported that a substantial portion of miRNA binding sites can be detected in the coding regions (Boudreau et al., 2014), while their functional properties have not been well characterized yet. We therefore tested whether there is enrichment of synonymous mutations that alter sequences of miRNA binding sites in the coding regions. In the analysis using seed sequences (7m8: positions 2–8 of processed miRNAs and 7A1: positions 2–7 followed by an adenine) for high confidence miRNAs provided by the miRBase (Kozomara and Griffiths-Jones, 2014) as reference, we observed enrichment of mutations causing any type of miRNA binding site changes (i.e. loss or gain of the target site) in ASD ($p = 0.011$, OR = 1.42, 1-tailed 2×2 Fisher's exact test) and ASD+SCZ ($p = 0.041$, OR = 1.28) (**Figure S3A**, left). However, no significant enrichment in the case groups was observed when a list of miRNA binding sites experimentally derived from human postmortem brain tissues (Boudreau et al., 2014) was used (**Figure S3A**, right).

While most amino acids are encoded by several different codons, some codons are used more frequently than others. The major cause for this codon usage bias is assumed to be that the preferred optimal codons are translated more accurately

and/or more efficiently (Hershberg and Petrov, 2008). The synonymous mutations could potentially alter the codon optimality and in turn alter the translation accuracy and/or efficiency. We then examined whether there is enrichment of such *de novo* synonymous mutations by using the definitions of optimal codons described in (Supek et al., 2014) and (Drummond and Wilke, 2008). Our analysis indicated that there was no significant enrichment of synonymous mutations causing gain, loss or gain/loss of optimal codons in the case groups (**Figure S3B**).

Another way the synonymous variants can alter the gene expression is to potentially change the secondary structure and thereby the function of RNAs. A recent study demonstrated that a substantial proportion of SNPs can alter local RNA structure (Wan et al., 2014). We also tested this possibility by using RNAsnp (Sabarinathan et al., 2013). We did not observe any significant enrichment of mutations that are likely to affect mRNA secondary structures in the affected subjects as compared to the controls (**Figure S3C**).

Taken together, our results suggest that *de novo* synonymous mutations might contribute only minimally to the disease risk by globally affecting miRNA binding to the coding regions, codon optimality or RNA secondary structure, though there might be individual mutations strongly associated with the diseases through these mechanisms.

Confirmatory Analyses for Enrichment of Near-Splice Site Synonymous Mutations Changing ESR in ASD and Mutations within the Frontal Cortex-Derived DHS in SCZ

To further confirm enrichment of near-splice site synonymous mutations changing ESR in ASD and synonymous mutations within the frontal cortex-derived DHS in SCZ, we performed permutation analyses (note that we had to exclude 26 mutations in ASD that could not be verified at the individual sample level). We observed very similar p values when compared to those obtained by the Fisher's exact test; $p = 5.0 \times 10^{-5}$ for near-splice site synonymous mutations changing ESR in ASD (the p value calculated by

Fisher's exact test = 3.4×10^{-5}) and p = 0.00078 for synonymous mutations within the frontal cortex-derived DHS in SCZ (the p value calculated by Fisher's exact test = 0.00079).

In addition to these technical considerations, we tested enrichment of *de novo* synonymous mutations likely to affect splicing regulation in ASD and SCZ by applying a different computational tool, SPANR (Splicing-based Analysis of Variants) (Xiong et al., 2015). By analyzing *de novo* synonymous mutations in cases and controls with SPANR (see **Supplemental Experimental Procedures**), we observed that the average of predicted impact of synonymous mutations on splicing was significantly higher in ASD and SCZ than in controls (|dPSI (difference in percent-spliced-in)|; 0.61 ± 0.08 in ASD, 0.56 ± 0.09 in SCZ and 0.33 ± 0.06 in controls, p = 0.006 for ASD and p = 0.027 for SCZ, 1-tailed t-test). With respect to enrichment of *de novo* synonymous mutations within brain-derived DHS in SCZ, we confirmed enrichment of these synonymous mutations particularly in SCZ using additional independent fetal brain-derived DHS datasets recently published by the Roadmap Epigenomics Project (Roadmap Epigenomics et al., 2015) (p = 0.0010 and OR = 2.63 for SCZ, p = 0.024 and OR = 2.02 for ASD, p = 0.0023 and OR = 2.32 for ASD+SCZ, **Figure S4**). These results indicate robust enrichment of these types of potentially functional *de novo* synonymous mutations in ASD and SCZ.

Enrichment Analyses of Various Types of *De Novo* Synonymous Mutations in Epileptic Encephalopathies (EE)

By using WES data for epileptic encephalopathies (EE) (Epi4K_Consortium et al., 2013) (56 *de novo* synonymous mutations in 264 EE probands), we evaluated whether particular types of potentially "functional" *de novo* synonymous mutations are enriched in EE with the same methods applied to ASD and SCZ. We observed significant enrichment of near-splice site synonymous mutations in EE (**Figure S6**, p = 0.044, OR = 1.67, 2×2 Fisher's exact test). In addition, as we observed for ASD and SCZ, near-splice

site synonymous mutations changing ESR showed further enrichment (**Figure S6** p = 0.014, OR = 2.12). On the other hand, no significant enrichment was observed for other types of potentially functional synonymous mutations (mutations within DHS and mutations affecting miRNA targets, codon optimality and RNA structure), though the smaller number of mutations limits the statistical power. Nevertheless, the observed results suggest a common role of near-splice site synonymous mutations across different neuropsychiatric diseases.

SUPPLEMENTAL REFERENCES

Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nature neuroscience *14*, 19-21.

Boudreau, R.L., Jiang, P., Gilmore, B.L., Spengler, R.M., Tirabassi, R., Nelson, J.A., Ross, C.A., Xing, Y., and Davidson, B.L. (2014). Transcriptome-wide discovery of microRNA binding sites in human brain. Neuron *81*, 294-305.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research *37*, W305-311.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S.*, et al.* (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341-352.

Epi4K_Consortium, Epilepsy Phenome/Genome, P., Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T.*, et al.* (2013). De novo mutations in epileptic encephalopathies. Nature *501*, 217-221.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007-1013.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M.*, et al.* (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179-184.

Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J.*, et al.* (2014). Most genetic risk for autism resides with common variation. Nature genetics *46*, 881-885.

Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O.*, et al.* (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. Nature genetics *43*, 860-863.

Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of, S., Group, P.S.*, et al.* (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518-529.

Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. Annual review of genetics *42*, 287-299.

Insel, T.R. (2010). Rethinking schizophrenia. Nature *468*, 187-193.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E.*, et al.* (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A.*, et al.* (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285-299.

Kavanagh, D.H., Dwyer, S., O'Donovan, M.C., and Owen, M.J. (2013). The ENCODE project: implications for psychiatric genetics. Molecular psychiatry *18*, 540-542.

Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J., and Chasin, L.A. (2011). Quantitative evaluation of all hexamers as exonic splicing elements. Genome research *21*, 1360-1374.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research *42*, D68-73.

McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E.*, et al.* (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Molecular psychiatry.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics *26*, 2069-2070.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V.*, et al.* (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242-245.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K.*, et al.* (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83-90.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D.*, et al.* (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246-250.

Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS genetics *9*, e1003709.

Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Huganir, R.L., Goes, F.S., Potash, J.B., and Zandi, P.P. (2012). SynaptomeDB: an ontology-based knowledgebase for synaptic genes. Bioinformatics *28*, 897-899.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing

genomic features. Bioinformatics *26*, 841-842.
Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N.*, et al.* (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674-1682.
Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.
Sabarinathan, R., Tafer, H., Seemann, S.E., Hofacker, I.L., Stadler, P.F., and Gorodkin, J. (2013). RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. Human mutation *34*, 546-556.
Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A.*, et al.* (2014). A framework for the interpretation of de novo mutation in human disease. Nature genetics.
Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L.*, et al.* (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237-241.
Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011). Genome-wide association study identifies five new schizophrenia loci. Nature genetics *43*, 969-976.
Stark, K.L., Xu, B., Bagchi, A., Lai, W.S., Liu, H., Hsu, R., Wan, X., Pavlidis, P., Mills, A.A., Karayiorgou, M.*, et al.* (2008). Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nature genetics *40*, 751-760.
State, M.W., and Sestan, N. (2012). Neuroscience. The emerging biology of autism spectrum disorders. Science *337*, 1301-1303.
Supek, F., Minana, B., Valcarcel, J., Gabaldon, T., and Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. Cell *156*, 1324-1335.
Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E.*, et al.* (2014). Landscape and variation of RNA secondary structure across the human transcriptome. Nature *505*, 706-709.
Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y.*, et al.* (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome research *22*, 1798-1812.
Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831-845.

Witte, J.S., Visscher, P.M., and Wray, N.R. (2014). The contribution of genetic variants to disease depends on the ruler. Nature reviews Genetics.

Wray, N.R., Yang, J., Goddard, M.E., and Visscher, P.M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. PLoS genetics *6*, e1000864.

Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R.*, et al.* (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science *347*, 1254806.

Xu, B., Hsu, P.K., Karayiorgou, M., and Gogos, J.A. (2012a). MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. Neurobiology of disease *46*, 291-301.

Xu, B., Hsu, P.K., Stark, K.L., Karayiorgou, M., and Gogos, J.A. (2013). Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. Cell *152*, 262-275.

Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012b). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics *44*, 1365-1369.

Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature genetics *43*, 864-868.