

Loss of function variants contribute to schizophrenia risk and point to *SETD1A* as a candidate susceptibility gene

Atsushi Takata¹, Bin Xu¹, Iuliana Ionita-Laza², J. Louw Roos³, Joseph A. Gogos^{4,5,*}, Maria Karayiorgou^{1,6,*}

¹Department of Psychiatry, Columbia University Medical Center, New York, NY 10032, USA

²Department of Biostatistics, Columbia University Medical Center, New York, NY 10032, USA

³Department of Psychiatry, Pretoria University, Weskoppies Hospital, Pretoria 0001, Republic of South Africa

⁴Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, USA

⁵Department of Physiology & Cellular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

⁶New York State Psychiatric Institute, New York, NY 10032, USA

*Corresponding authors: mk2758@columbia.edu, jag90@columbia.edu

SUMMARY

Loss-of-function (LOF) (i.e. nonsense, splice site and frameshift) variants that lead to disruption of gene function are likely to contribute to the etiology of neuropsychiatric disorders. Here, we perform a systematic investigation of the role of both *de novo* and inherited LOF variants in schizophrenia using exome sequencing data from 231 case and 34 control trios. We identify two *de novo* LOF variants in the *SETD1A* gene, which encodes a subunit of histone methyltransferase, a finding unlikely to have occurred by chance, and provide evidence for a more general role of chromatin regulators in schizophrenia risk. Transmission pattern analyses reveal that LOF variants are more

likely to be transmitted to affected individuals than controls. This is especially true for private LOF variants in genes intolerant to functional genetic variation. These findings highlight the contribution of LOF mutations to the genetic architecture of schizophrenia and provide important insights into disease pathogenesis.

INTRODUCTION

Schizophrenia (SCZ) is a severe, common psychiatric disorder with a strong genetic component (Rodriguez-Murillo et al., 2012). Elucidating the genomic architecture of SCZ and identifying specific risk genes and affected pathways can inform the underlying disease pathophysiology and lead to identification of novel treatment targets.

Recent findings make increasingly clear that rare *de novo* copy number variants (CNVs) and *de novo* protein-altering mutations (defined as point substitutions or single nucleotide variants [SNVs], and small insertions/deletions [indels]) contribute to the risk of SCZ and a number of other brain disorders, such as autism and intellectual disability (Karayiorgou et al., 1995; Kirov et al., 2012; Malhotra et al., 2011; Xu et al., 2008; de Ligt et al., 2012; Girard et al., 2011; Gulsuner et al., 2013; Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012b; Rauch et al., 2012; Sanders et al., 2012; Xu et al., 2012; Xu et al., 2011). *De novo* mutations are often transmitted by relatively asymptomatic carriers in afflicted families (Karayiorgou et al., 2012; Rodriguez-Murillo et al., 2012) thus contributing to the heritable component of these disorders. The relative contribution of *de novo* or inherited mutations to each disorder remains to be determined but it is expected to correlate with the impact of the disease on fitness and fecundity. Thus, investigation of both variant types should be particularly important for diseases with partial reduction in fecundity, such as SCZ. Among *de novo* and inherited variants, ones that lead to loss-of-function (LOF) by disrupting protein-coding genes have a high probability of being deleterious and are of

great interest in disease etiology (Veltman and Brunner, 2012).

This study was designed to evaluate the role that *de novo* and inherited LOF variants play in conferring SCZ risk. First, our investigation of the impact of *de novo* variation confirmed an excess of *de novo* LOFs in SCZ patients and led to the identification of a new candidate risk gene (*SETD1A*) harboring two *de novo* LOF indels in two individuals with SCZ. Second, our investigation of the impact of transmitted variation revealed a relative increase in the transmission of private and rare LOF variants to SCZ probands compared to controls, especially in genes that are more intolerant to functional genetic variation. Collectively, these findings highlight the contribution of LOF mutations to the genetic architecture of SCZ and provide important insights into disease pathogenesis.

RESULTS

Systematic search for *de novo* LOF indels identifies two mutations in *SETD1A*

Recently, we sequenced a total of 795 exomes from 231 parent-proband case trios enriched for sporadic SCZ cases, as well as 34 control trios (Xu et al., 2012) and using GATK (McKenna et al., 2010) and Dindel (Albers et al., 2011) we identified 14 indels (ten frameshift and four in-frame variants) with a validation rate of ~20%. The low accuracy of variant calls for *de novo* indels, suggested that we might have missed a considerable number of indels, including LOF frameshift indels. Here, we undertake a more extensive analysis of *de novo* LOF indels, using two recently developed software packages specifically designed to detect indels from high-throughput sequencing data by applying split-read based methods, PRISM (Jiang et al., 2012) and Pindel (Ye et al., 2009). Initial variant calling using PRISM and Pindel, produced 24 and 36 candidate *de novo* LOF indels, respectively. Of these, 16 candidates overlapped (for a total of 44 unique indels using the two programs), and all of the ten LOF indels validated in our previous study (nine in cases and one in controls) (Xu et al., 2012) were included

among them (Figure S1). The remaining 34 candidates for previously unidentified *de novo* LOF indels (29 in cases and five in controls) were manually inspected using the Integrative Genomics Viewer (Robinson et al., 2011) for supporting reads in the proband but not in either of the parents. 20 candidates in cases and all five candidates in controls were excluded by this step due to lack of reads supporting the indel in the proband or existence of multiple reads supporting the indel in either of the parents. We tested the remaining nine candidates by Sanger sequencing, and five novel *de novo* LOF indels (all in cases) were validated (Table 1). (A complete list of *de novo* LOF mutations in our cohort is shown in Table S1). The total number of individuals carrying *de novo* LOF variants (nonsense, splice site and frameshift variants) in our case and control datasets were, respectively, 25 and 1, resulting in a 3.7-fold enrichment of *de novo* LOF mutation carriers in SCZ subjects (25/231; 10.8% in cases and 1/34; 2.9% in controls), although the enrichment is not statistically significant. We also asked whether the load of *de novo* LOF variants correlates with any of four available clinical variables (severity of the disease [based on clinical course and functional outcome, see Supplemental Experimental Procedures for detailed information], age at disease onset, history of childhood learning disabilities, and comorbidity of mental retardation). We observed a significant correlation between history of childhood learning difficulties and the number of *de novo* LOF variants per-individual ($p = 0.0006$, Spearman's rank correlation $r = 0.26$), indicating a higher incidence of childhood learning difficulties in patients with *de novo* LOF variants.

Notably, two of the five validated LOF indels were found in the same gene, *SETD1A* (Figure 1), with two case probands each carrying a frameshift *de novo* indel and a *de novo* indel changing the canonical splice acceptor site sequence from AG to GG. The other three validated *de novo* LOF indels were located within the *CEP104*, *ITPR2* and *INHBC* genes (Table 1 and Figure S2). *SETD1A* encodes for a component of a histone methyltransferase complex that produces mono-, di-, and tri-methylated

Table 1. List of the Newly Identified De Novo Loss-of-Function Variants in 265 [Trios](#)

Position (hg19)	Reference / Variant Allele	Type	Effect	Property	Gene	Pop.	Diagnosis
chr1:3753245-3753259	CAGGGACTCTGCTG / -	D	Canonical splice site + Inframe	c.1120-3_1131 del, p.A374_L377 del	<i>CEP104</i>	U.S.	SCZAFF
chr12: 26816636-26816638	CTG / AGGTCAGTGTC	D + I	Frameshift	c.1693_1695 del3ins11, p.Q565fs	<i>ITPR2</i>	Afr.	SCZ
chr12: 57843366	A / -	D	Frameshift	c.619 del1, p.Q207fs	<i>INHBC</i>	Afr.	SCZAFF
chr16: 30976335	C / -	D	Frameshift	c.1272 del1, p.D424fs	<i>SETD1A</i>	Afr.	SCZ
chr16: 30992058-30992059	AG / -	D	Canonical splice site	c.4582-2_4582-1 del2	<i>SETD1A</i>	U.S.	SCZ

D, deletion; I, insertion, Afr., Afrikaner; SCZ, schizophrenia, SCZAFF, schizoaffective disorder

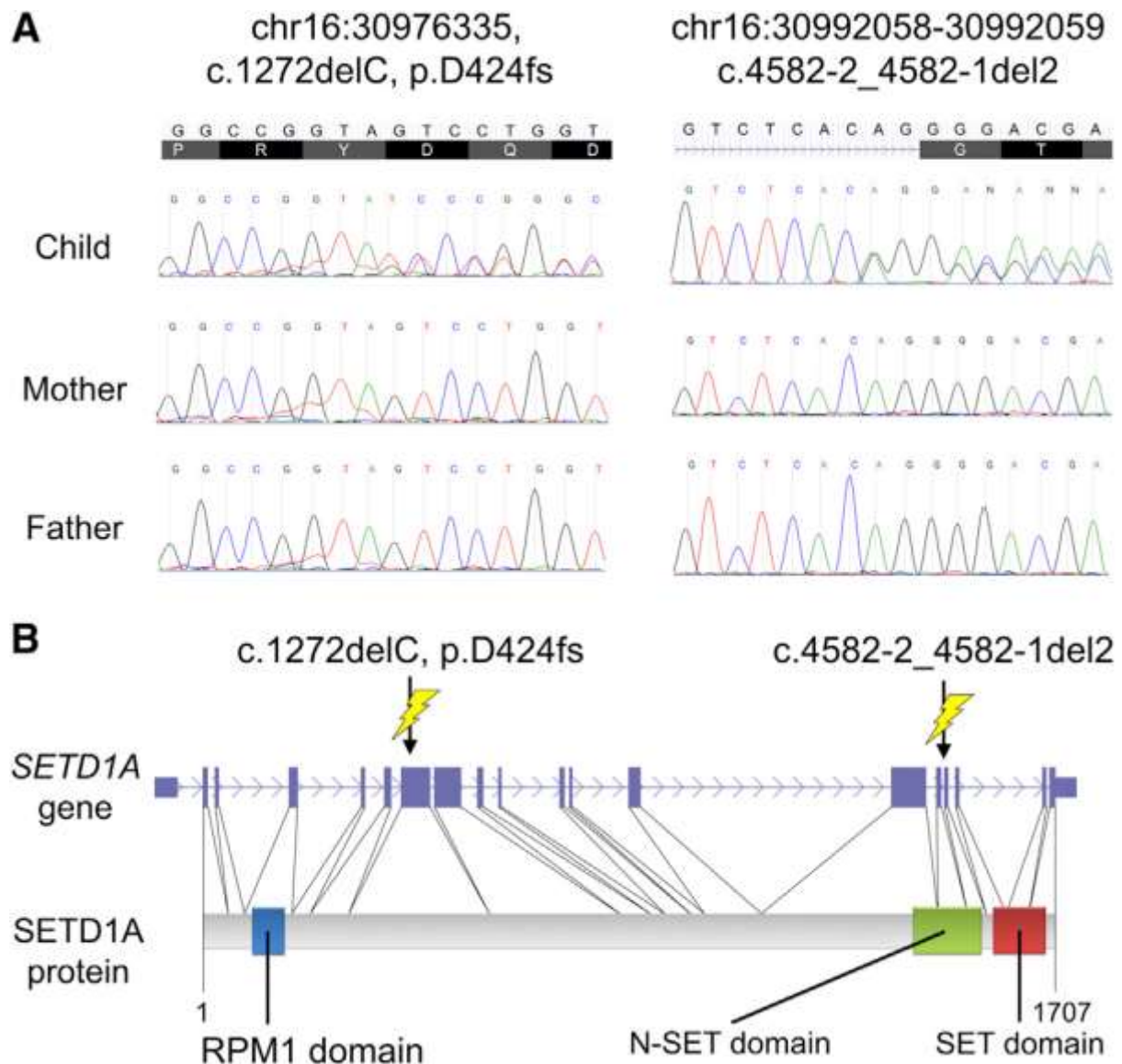


Figure 1. Two *de novo* LOF indels in *SETD1A*

(A) Two *de novo* LOF indels in the *SETD1A* gene confirmed by Sanger sequencing. (B) Structure of the *SETD1A* gene and the SETD1A protein along with the positions of the two LOF indels. The *de novo* frameshift indel variant (D424fs) (left) creates an early stop codon and leads to protein truncation. The *de novo* indel variant c.4582-2_4582-1del2 variant (right) changes the canonical splice acceptor site sequence adjacent to exon 16 from AG to GG.

histone H3 at Lysine 4 (H3K4) (Miller et al., 2001; Roguev et al., 2001). The *de novo* frameshift indel (D424fs) in the first proband creates an early stop codon and leads to a predicted protein truncation. The *de novo* indel changing the canonical splice acceptor site (c.4582-2_4582-1del2 variant) in the second proband is predicted to lead to loss of exon 16 and disruption of N-SET domain that plays an important role for H3K4 methyltransferase activity (Dehe et al., 2006; Schlichter and Cairns, 2005). We used a rather conservative approach to assess the statistical significance of observing at least two *de novo* LOF mutations in the same gene that takes into account the length of the gene and the gene specific GC content. Based on the number of *de novo* LOFs found in our cases (a total of 25: 18 coding LOFs [nonsense SNVs and frameshift indels] and 7 splice site LOFs [canonical splice site SNVs and indels affecting canonical splice site]) we determined that two or more *de novo* LOF mutations in a single gene were unlikely to occur by chance ($p = 0.035$). As mentioned above, because increased rates of *de novo* LOFs in patients with psychiatric disorders have now been reported in most of the published exome studies in psychiatric and neurodevelopmental disorders, this is a conservative analysis since it conditions on the number of observed *de novo* LOF mutations in cases (Girard et al., 2011; Iossifov et al., 2012; O’Roak et al., 2012b; Sanders et al., 2012; Xu et al., 2012; Xu et al., 2011). We used an additional method to evaluate the significance of observing at least two LOF *de novo* mutations in *SETD1A*, which results in a p-value that is specific to *SETD1A*. Specifically, we first estimated the probability of a LOF *de novo* mutation per chromosome for *SETD1A*, and then used a Poisson model for the probability of observing two or more LOF *de novo* events in this gene. The key element in this calculation is specifying the mutation rate for *SETD1A*. To achieve this, we made use of the recently developed TADA (Transmission And De novo Association) software (He et al., 2013), which estimates the mutation rate per gene based on its exonic length and its nucleotide content (Sanders et al., 2012). For *SETD1A*, we obtained a LOF mutation rate per chromosome of 4.75×10^{-6} . The resulting

probability of two or more LOF mutations is 2.4×10^{-6} . Assuming that there are 20,000 genes, the genome-wide p value is 0.048, very similar to our first analysis.

Considering the phenotypic presentation of the mutation carriers, it is notable that both SCZ cases carrying *de novo* LOF mutations in *SETD1A* also manifest symptoms consistent with a secondary diagnosis of obsessive-compulsive disorder (OCD). Detailed description of the clinical presentation of these two individuals is available in Supplemental Information. For both, SCZ is the primary lifetime diagnosis and the one associated with their long-term disability. Overall, clinical information for obsessive-compulsive symptoms is available for 205 of the 231 probands in our SCZ cohort, with 17 probands meeting full diagnostic criteria for secondary, comorbid OCD. The proportion of subjects with OCD in our SCZ cohort (17/205, 8.3%) is within the range reported in previous studies (7.8–26%, Bottas et al., 2005). Given the number of subjects with comorbid OCD in our SCZ cohort, occurrence of *de novo* LOF mutations in *SETD1A* was significantly associated with SCZ comorbid with OCD ($p = 0.002$, Barnard's test). Whether this highly similar phenotypic outcome suggests that patients with LOF mutations in *SETD1A* could define a new clinical subtype of SCZ comorbid with OCD symptoms (Bottas et al., 2005) remains to be determined by additional sequencing studies of larger cohorts of both SCZ and OCD patients well-phenotyped for psychotic and obsessive compulsive symptoms.

Damaging *de novo* variants highlight a role of chromatin regulators in SCZ

Further analysis aiming at prioritizing genes and variants for their likely pathogenicity provided additional evidence for a role of both *de novo* damaging mutations and chromatin regulators. Specifically, application of a recently established scoring system that makes use of human polymorphism data and reflects the intolerance of each gene against functional genetic variation in the general human population (residual variation intolerance score [RVIS]) (Petrovski et al., 2013), to all *de novo* mutations observed in

individuals with SCZ in our and other studies (Girard et al., 2011; Gulsuner et al., 2013; Xu et al., 2012), as well as in control individuals from family-based WES studies for SCZ, ASD and intellectual disability (ID) (Gulsuner et al., 2013; Iossifov et al., 2012; O’Roak et al., 2012b; Rauch et al., 2012; Sanders et al., 2012; Xu et al., 2012), revealed significant enrichment of genes intolerant to variation among the genes hit by *de novo* LOF and missense variants (but not by *de novo* silent variants) in cases (Figure S3A). No significant enrichment of any type of *de novo* variant was observed within the control samples. Two-dimensional (2D) plotting analyses that used the RVIS percentiles for genes (y-axis) and PolyPhen-2 quantitative scores (Adzhubei et al., 2010) for missense mutations (x-axis) revealed that *de novo* missense variants located within genes intolerant to variation and predicted to be damaging by PolyPhen-2 (quantitative score ≥ 0.95) are enriched in cases and therefore are more likely to contribute to disease risk (Figure S3B). A similar pattern of excess of genes intolerant to variation among genes hit by functional *de novo* mutations was also reported in cohorts ascertained for epilepsy, ID and ASD (Petrovski et al., 2013), suggesting that intolerant genes carrying functional variants, including LOF ones, are more likely to be associated with disease.

Based on these observations, in the combined dataset of family-based WES studies of SCZ, we selected a total of 62 genes that are intolerant to variation and harbor *de novo* LOF or damaging missense (PolyPhen-2 scores ≥ 0.95) variants: 38 genes from our studies (36 from the Xu et al., 2012 study plus *SETD1A* and *ITPR2* identified in the current analysis) as well as 25 genes from two other studies (Girard et al., 2011; Gulsuner et al., 2013) (*de novo* damaging variants in *KIAA1109* were found across datasets). Gene-set enrichment analysis using DAVID (The Database for Annotation, Visualization and Integrated Discovery) (Huang et al., 2009), revealed “chromatin regulator” as the most significantly enriched term among these 62 genes (Figure S4A and Table S2, also see Supplemental Information). Direct and indirect interactions among chromatin regulator genes harboring *de novo* damaging variants,

including *SETD1A*, *RBBP5*, *TRRAP2*, *UBR5*, *KDM2B* and *KDM5C*, are shown in Figure S4B.

The inherited component of SCZ risk is enriched in rare LOF variants

Recent WES studies in ASD provided evidence that inherited LOF mutations were enriched in affected individuals (Lim et al., 2013; Yu et al., 2013). Therefore, detailed analysis of rare inherited variants promises to provide important insights toward our understanding of the familial component of SCZ risk. To evaluate evidence for such etiologies in SCZ, we studied the transmission patterns of common and rare SNVs and indel variants identified from our WES data.

Because recent whole genome and exome studies suggested that many deleterious and disease-related variants, especially LOF variants, exist at low frequency in the human population (Lim et al., 2013; MacArthur et al., 2012; The 1000 Genomes Project Consortium, 2012; Zhu et al., 2011), we first partitioned the variants into disjoint classes according to their minor allele frequencies (MAF) in the parents as follows: private variants: observed only once in the parents; rare variants: observed more than once and $MAF \leq 0.05$; common variants: $MAF > 0.05$.

For the functional annotation and classification of variants, we used the SnpEff software (Cingolani et al., 2012). By following the definition of functionality of variants provided in this software, we identified three classes: nonsense variants, variants disrupting start and stop codons, canonical splice site variants and frameshift variants were classified as “LOF” variants; missense variants and in-frame indels were classified as “MODERATE” effect variants; and silent variants, variants generating start codons in untranslated regions and nonsynonymous variants generating alternative start codons were classified as “LOW” effect variants.

When we analyzed the transmission (number of transmissions from parents to offspring) to untransmission (number of untransmissions) ratio (T:U ratio) of LOF

Table 2. Numbers of Transmitted and Untransmitted Alleles in Cases and Controls

Cohort	Transmitted	Untransmitted	T:U ratio	RTR (Case/Control)	p Value
Loss-of-Function Variants (Nonsense, Canonical Splice Site, and Frameshift Variants)					
Private (hit once in parental population)					
Cases	1,682	1,945	0.86	1.49	0.0086
Controls	159	274	0.58	–	–
Rare (nonprivate and frequency ≤ 0.05)					
Cases	5,948	7,168	0.83	1.31	0.0051
Controls	902	1,425	0.63	–	–
Common (frequency > 0.05)					
Cases	26,350	27,724	0.95	1.08	0.037
Controls	3,732	4,231	0.88	–	–
MODERATE-Effect Variants					
Private					
Cases	26,558	25,250	1.05	0.99	0.607
Controls	2,286	2,155	1.06	–	–
Rare					
Cases	146,052	158,361	0.92	1.10	0.022
Controls	20,603	24,663	0.84	–	–
Common					
Cases	920,641	958,085	0.96	1.02	0.023
Controls	127,551	135,706	0.94	–	–
LOW-Effect Variants					
Private					
Cases	16,622	15,249	1.09	0.99	0.361
Controls	1,464	1,335	1.10	–	–
Rare					
Cases	129,196	136,780	0.94	1.05	0.176
Controls	17,610	19,562	0.90	–	–
Common					
Cases	1,148,993	1,177,446	0.98	1.01	0.058
Controls	159,769	165,749	0.96	–	–

T:U ratio, transmitted to untransmitted ratio; RTR, relative transmission ratio (T:U ratio in cases/T:U ratio in controls). Detailed definition for MODERATE- and LOW-effect variants are described in the Supplemental Experimental Procedures. p values were calculated by one-sided permutations.

variants, we observed a tendency towards undertransmission in both cases and controls, i.e. $T:U < 1$ (Table 2, private LOFs: 0.86, rare LOFs: 0.83 and common LOFs: 0.95 in cases; private LOFs: 0.58, rare LOFs: 0.63 and common LOFs: 0.88 in controls). A $T:U$ ratio of 1 is expected for neutral loci and it is what we observe for the common synonymous (low-effect) variants ($T:U = \sim 1$) (Table 2). There are several possible explanations for the observed distortion in the transmission ratios of the LOF variants. One possibility is that this is the result of selection acting against these LOF variants, a mutation load-reducing mechanism that facilitates elimination of deleterious mutations per generation and maintains higher mean fitness. For example, if some of these variants are associated with embryonic lethality or early onset of deleterious developmental defects (Edmonds et al., 1982), some of the carriers could be eliminated immediately in the next generation. Selection may also be acting prior to fertilization in the form of higher failure rate in gametogenesis, lower chance for fertilization (Otto and Hastings, 1998), or in the form of cellular processes related to selfish spermatogonial selection (Goriely et al., 2013). In addition, the combination of LOF variants from the two parents in the offspring may lead to more deleterious effects (Crow, 2000). Finally, as we generally collected subjects without severe comorbid diseases or developmental disorders for both controls and cases, LOF variants linked to such maladies should be represented less frequently among the transmitted alleles.

Although we observed global undertransmission of LOFs in both cases and controls, the deviation from the 1:1 ratio was more attenuated in cases than in controls. Therefore, we then compared the $T:U$ ratio in cases and controls to calculate the relative transmission ratio ($RTR = T:U \text{ ratio in cases} / T:U \text{ ratio in controls}$). Since sequencing and genotyping errors are expected to affect cases and controls equally, contrasting $T:U$ ratios in cases vs. controls should provide a robust result. We observed significantly higher RTRs than expected for all allele frequency classes of LOF variants (Table 2, private LOFs: $RTR = 1.49$, $p = 0.0086$; rare LOFs: $RTR = 1.31$, $p = 0.0051$;

common LOFs: RTR = 1.08, $p = 0.037$, one-sided permutation test). These results suggest that LOF variants identified in cases, especially the private and rare ones, are enriched in disease causing variants.

To further confirm that our observation was not due to errors in genotyping calls we analyzed “MODERATE” effect variants and “LOW” effect variants. Unlike LOF variants, the T:U ratios for MODERATE and LOW effect variants were close to 1:1 for all allele frequency classes (Table 2). Therefore, the global undertransmission of LOF variants we observed is unlikely due to sequencing errors but rather due to selection acting against the LOF variants. When we calculated RTRs, we observed significantly higher RTRs for the rare and common MODERATE effect variants, but the effect sizes observed were more modest than LOFs (Table 2, private MODERATE: RTR = 0.99, $p = 0.607$; rare MODERATE: RTR = 1.10, $p = 0.022$; common MODERATE: RTR = 1.02, $p = 0.023$, one-sided permutation test). For the variants with LOW effect, we did not obtain any significant results (Table 2). These findings indicate less contribution from variants classified as MODERATE or LOW effect to the disease etiology, than from LOF variants. We also performed an analysis excluding all European American case trios and confirmed that the observed results were not influenced by the combination of two population groups (Table S3, see also Supplemental Information).

Application of RVIS provided further evidence for a role of selection pressure in shaping the transmission patterns of inherited variants. We observed further increase in RTR for private LOF variants in genes intolerant to functional variation (Figure 2, RTR = 2.13, $p = 0.0193$, one-sided permutation test) compared to private LOFs in all genes (RTR = 1.49). RTR for private LOF variants in tolerant genes was 1.35 ($p = 0.025$). Difference of RTRs between private LOF variants in intolerant and tolerant genes showed marginal significance ($p = 0.054$). By contrast, a difference of RTRs between intolerant and tolerant genes was not apparent for other classes of variants (rare and common LOFs, all categories of missense and silent variants). These results

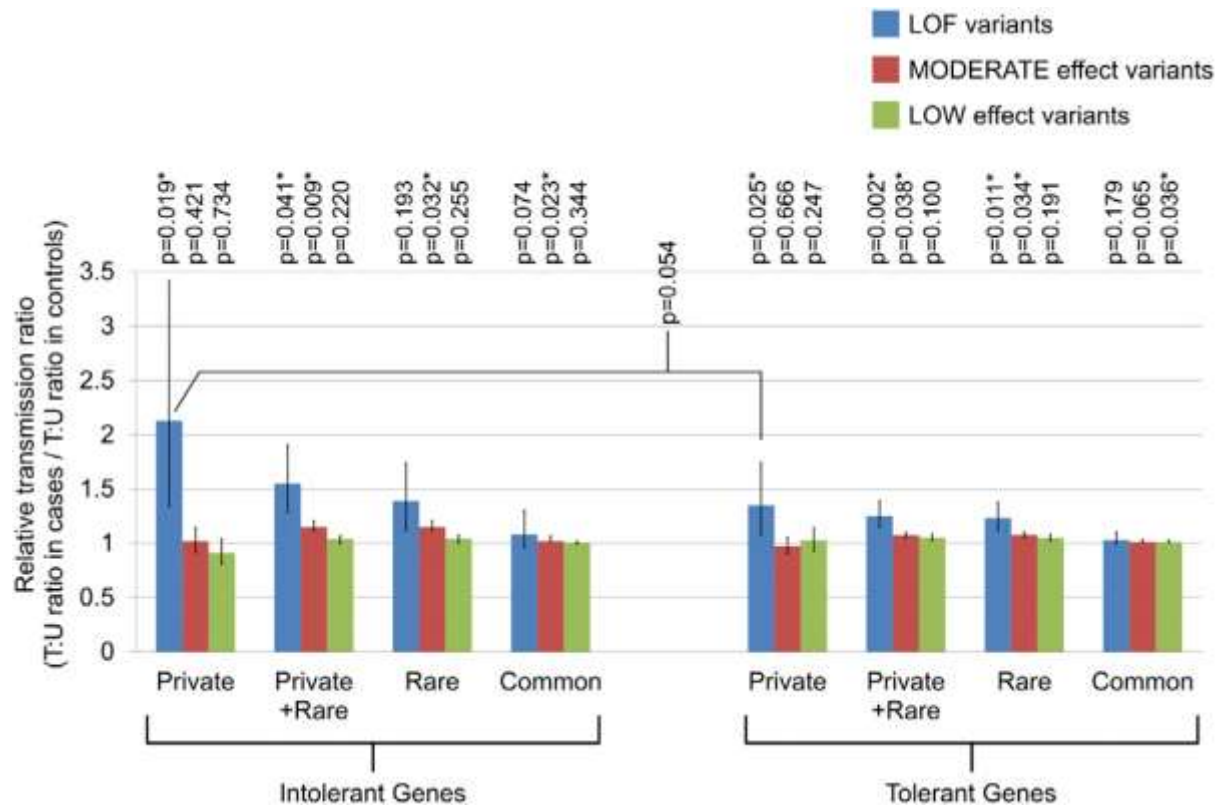


Figure 2. Prominent overtransmission of private LOF variants in intolerant genes

Relative transmission ratios (RTRs) are displayed for variants categorized by their functionality (blue: LOF [loss-of-function] variants; red: MODERATE effect variants; and green: LOW effect variants), minor allele frequencies (private, rare or common) and gene intolerance. Detailed definition for LOF, MODERATE and LOW effect variants is provided in the Experimental Procedures. $*p < 0.05$. Error bars indicate 95% confidential intervals. P values for RTRs were calculated by one-sided permutation analysis with random shuffling of the case-control status.

suggest that private LOFs in intolerant genes in cases are more likely to contribute to disease etiology than other variants. It should be also noted that global undertransmission of private and rare LOF variants was prominent for intolerant genes both in cases and controls (private LOF variants: T:U = 0.81 in cases and 0.38 in controls, rare LOF variants: T:U = 0.57 in cases and 0.41 in controls), further indicating that the observed global undertransmission of LOF variants is not due to sequencing errors.

We also asked whether the load of inherited LOF variants correlates with four clinical variables (see above, also described in Supplemental Experimental Procedures) and observed a significant correlation between severity scores of the disease and per-individual number of private LOF variants ($p = 2.77 \times 10^{-5}$, Spearman's rank correlation $r = 0.29$). This finding highlights the importance of private inherited LOFs for disease manifestation. Thus, both analysis of transmission patterns and phenotypic correlations suggest that the inherited component of SCZ risk is enriched in private LOF variants, especially ones in genes intolerant to functional variation. A gene-set enrichment analysis showed that despite high genetic heterogeneity, there is an enrichment of a few biological processes and/or functions among such genes, most notably enrichment of terms related to ATPase activity (see Supplemental information and Table S4).

Finally, to explore individual promising candidate genes further, we cross-compared the list of intolerant genes harboring at least one private LOF variant in SCZ with lists of candidate genes from previously published literature (see Supplemental information and Table S5). We found several genes in common that represent promising candidate genes worthy of follow up analysis. In particular, when we cross-compared the list of intolerant genes harboring private LOFs with previously published literature on CNVs associated with neuropsychiatric and neurodevelopmental diseases (Malhotra and Sebat, 2012), we found two genes (*LZTR1* and *RFC2*) in common. *LZTR1*, encoding a protein exclusively localized to the Golgi

network helping stabilize the Golgi complex (Nacak et al., 2006), is located within the 3-Mb 22q11.2 deletion locus, a finding consistent with recent evidence establishing a strong link between the Golgi apparatus and neuronal phenotypes associated with the 22q11.2 microdeletion (Mukai et al., 2008; Xu et al., 2013). *RFC2* is located on the chromosome 7q11.2 Williams–Beuren syndrome (WBS) critical locus, encodes for a protein playing a role in response to DNA damage, and may be involved in microcephaly and growth retardation observed in WBS patients (Merla et al., 2010). This finding is also consistent with recent findings implicating DNA repair mechanisms in the pathogenesis of SCZ and ASD (Ionita-Laza et al., 2014).

DISCUSSION

In this study, we performed a systematic investigation of the role of both *de novo* and inherited LOF variants in order to elucidate the role that gene-disruptive events play in the genetic architecture of SCZ.

We confirmed an excess of *de novo* LOFs in SCZ patients and identified a single gene (*SETD1A*) harboring two *de novo* LOF indels in two individuals with SCZ. *SETD1A* encodes a catalytic subunit of the histone methyltransferase protein complex, named Set/COMPASS (complex protein associated with Set1). This complex mediates mono-, di-, and tri-methylation of H3K4, a mark of active gene transcription (Miller et al., 2001; Roguev et al., 2001). In humans, Set/COMPASS is comprised of one of the six catalytic subunits (*SETD1A*, *SETD1B*, *KMT2A* [also known as *MLL*], *KMT2D* [*MLL2*], *KMT2C* [*MLL3*], and *KMT2B* [*MLL4*]), four common subunits (*WDR5*, *RBBP5*, *ASH2L*, and *DPY30*) and several subtype-specific proteins (Mohan et al., 2010; Shilatifard, 2012). Interestingly, *de novo* mutations in other subunits comprising Set/COMPASS have been reported in SCZ and ASD cases: a missense variant in *RBBP5* in SCZ (Gulsuner et al., 2013), as well as a nonsense variant in *KMT2C* and a missense variant in *KMT2A* in ASD (Neale et al., 2012; O’Roak et al., 2012b). It is also noteworthy that the ATP-dependent

chromatin helicase CHD8, in which recurrent *de novo* LOF mutations have been identified in ASD patients (O'Roak et al., 2012a; O'Roak et al., 2012b), directly interacts with common subunits of Set/COMPASS and plays a role in the regulation of H3K4 trimethylation (Yates et al., 2010).

Implication of *SETD1A* involvement, as well as the observed enrichment of chromatin regulators among genes hit by *de novo* damaging variants in SCZ, are in line with accumulating data indicating a role of chromatin related genes in various psychiatric and neurodevelopmental disorders (Ben-David and Shifman, 2013; Ronan et al., 2013). Involvement of chromatin modification pathways in SCZ was also suggested from analysis integrating data from *de novo* CNVs, *de novo* SNVs and genome-wide association studies (GWAS) (Gilman et al., 2012). Although there was no statistically significant enrichment of chromatin related pathways among genes at loci identified in the largest GWAS for SCZ so far, some genes involved in chromatin regulation, such as *PBRM1* and *SETD8*, were found in the proximity of single nucleotide polymorphisms (SNPs) with genome-wide significant association (Ripke et al., 2013).

H3K4 trimethylation, in particular, is known to be essential for neuronal differentiation (Wynder et al., 2005) and fear memory formation (Gupta et al., 2010). The H3K4 trimethylation status in prefrontal cortex neurons is dynamically regulated from the late prenatal period to early-adulthood (Shulha et al., 2013). Interestingly, there are reports of decreased H3K4 trimethylation at the *GAD1* promoter region in the prefrontal cortex of female SCZ patients (Huang et al., 2007), as well as of upregulation of *Gad1* H3K4 trimethylation induced by clozapine but not by haloperidol (Huang et al., 2007). It is also noteworthy that a *de novo* LOF variant in *KDM5C*, encoding JARID1C that mediates demethylation of H3K4, was found in another SCZ subject in our cohort (Xu et al., 2012). Mutations in *KDM5C* have also been linked to non syndromic X-linked mental retardation (Jensen et al., 2005).

Application of RVIS identified *SETD1A* within the first percentile of the most

intolerant genes against functional variants including LOF and missense variants (Petrovski et al., 2013). When we calculated in the same manner intolerance scores specifically against LOF variants using data from Exome Variant Server (<http://evs.gs.washington.edu/EVS>), *SETD1A* was included in the fourth percentile of the genes most intolerant against LOF variants. In addition, there is no LOF variant disrupting *SETD1A* in the 1000 Genomes Project database (The 1000 Genomes Project Consortium, 2012), another large-scale data source of human genetic variation. The high intolerance of *SETD1A* specifically against LOF variants further supports that the identification of two *de novo* LOF mutations in our SCZ cohort is unlikely to have occurred by chance.

Our current analysis using PRISM and Pindel revealed five previously unidentified *de novo* LOF indels. The fact that these LOFs were not identified in our previous analysis using Dindel and its default filtering parameters is primarily due to the low coverage of these variant sites in relation to their high GC content, or their complex structure. Specifically, for four of these LOFs (*CEP104*, *INHBC*, and the two *SETD1A* indels) the coverage of their genomic position was 5x-13x, with 56-70 GC% content in the surrounding 100bp regions (50bp upstream and 50bp downstream). For the fifth LOF (*ITPR2* indel) the genomic position was well covered (42x coverage, GC% content = 51), but the variant was probably missed in our previous analysis because of its complexity and relatively long length (a combination of an 11bp insertion and a 3bp deletion). Therefore, our current analysis using two different software and custom filtering criteria may be advantageous for identifying indels with complex structures or at low coverage regions. It is also notable that the *SETD1A* frameshift indel and the *CEP104* indel were solely identified by PRISM and Pindel, respectively. This indicates that the sensitivity of detecting *de novo* indels could be increased substantially by using multiple software.

Identification of relatively increased transmission of private and rare inherited

LOF variants in SCZ probands highlighted the significance of such events as risk factors for SCZ, while correlation of mutational load with disease severity implicated them in clinical outcome. Similar to the *de novo* CNVs and SNVs these mutations are distributed across many different genes and therefore would have been missed through previous association or linkage studies. Results from our WES family-based analysis are consistent with recent case-control studies performing targeted resequencing of candidate genes, which also reported excess of rare LOF variants in SCZ patients (Hu et al., 2013; Kenny et al., 2013).

Results from the current study and the findings in other studies for *de novo* SNVs and indels (Girard et al., 2011; Xu et al., 2012; Xu et al., 2011) collectively highlight the significance of LOF mutations in the overall genetic etiology of SCZ. In addition, LOF events provide further collective insight into the heritable component of SCZ, which has not yet been accounted for by *de novo* CNVs and SNVs. Finally, identification of *SETD1A* as one of the first genes recurrently hit by *de novo* LOF mutations in SCZ, strongly suggests that further studies using large sample sets enriched in truly sporadic cases as well as intensive analysis of currently available datasets focusing on LOF variants will lead to the discovery of additional recurrently affected genes, a key step towards elucidating the neural mechanisms underlying susceptibility to SCZ and designing novel therapies (Karayiorgou et al., 2012).

EXPERIMENTAL PROCEDURES

Analysis of *de novo* LOF indels

Short to middle-sized indels were called by PRISM (Jiang et al., 2012) and Pindel (Ye et al., 2009) with default parameters. Candidate *de novo* LOF variants were first selected as described in Supplemental Experimental Procedures, then manually inspected using the Integrative Genomics Viewer (Robinson et al., 2011) and finally subjected to Sanger sequencing for validation.

Transmission analysis of inherited variants

Transmission analysis was performed by using genotype data generated as described in Supplemental Experimental Procedures. To assess the statistical significance of the relative overtransmission of LOF variants in affected individuals as compared to controls, we performed permutation testing by randomly permuting the case/control labels of the trios in our datasets. Detailed information is described in Supplemental Experimental Procedures.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

We are grateful to all the families who participated in this research. We thank S.L. Lundy for valuable assistance with clinical database maintenance. We also thank S. Levy and the HudsonAlpha Genomics Services Laboratory for sequencing. This work was partially supported by National Institute of Mental Health (NIMH) grants MH061399 (to M.K.) and MH097879 (to J.A.G.) and the Lieber Center for Schizophrenia Research at Columbia University. A.T. was supported by the JSPS Postdoctoral Fellowship for Research Abroad. B.X. was partially supported by a National Alliance for Research in Schizophrenia and Depression (NARSAD) Young Investigator Award.

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nature methods* 7, 248-249.

Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., and Durbin, R.

(2011). Dindel: accurate indel calls from short-read data. *Genome research* 21, 961-973.

Ben-David, E., and Shifman, S. (2013). Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular psychiatry* 18, 1054-1056.

Bottas, A., Cooke, R.G., and Richter, M.A. (2005). Comorbidity and pathophysiology of obsessive-compulsive disorder in schizophrenia: is there evidence for a schizo-obsessive subtype of schizophrenia? *Journal of psychiatry & neuroscience : JPN* 30, 187-193.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80-92.

Crow, J.F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature reviews Genetics* 1, 40-47.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., *et al.* (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *The New England journal of medicine* 367, 1921-1929.

Edmonds, D.K., Lindsay, K.S., Miller, J.F., Williamson, E., and Wood, P.J. (1982). Early embryonic mortality in women. *Fertility and sterility* 38, 447-453.

Dehe, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodriguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A., *et al.* (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *The Journal of biological chemistry* 281, 35404-35412.

Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A., Karayiorgou, M., and Vitkup, D. (2012). Diverse types of genetic variation converge on functional gene networks

involved in schizophrenia. *Nat. Neurosci.* 15, 1723–1728.

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.

Goriely, A., McGrath, J.J., Hultman, C.M., Wilkie, A.O., and Malaspina, D. (2013). "Selfish spermatogonial selection": a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. *The American journal of psychiatry* 170, 599-608.

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.

Gupta, S., Kim, S.Y., Artis, S., Molfese, D.L., Schumacher, A., Sweatt, J.D., Paylor, R.E., and Lubin, F.D. (2010). Histone methylation regulates memory formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30, 3589-3599.

He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., *et al.* (2013). Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS genetics* 9, e1003671.

Hu, X., Zhang, B., Liu, W., Paciga, S., He, W., Lanz, T.A., Kleiman, R., Dougherty, B., Hall, S.K., McIntosh, A.M., *et al.* (2013). A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. *Molecular psychiatry*.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57.

Huang, H.S., Matevossian, A., Whittle, C., Kim, S.Y., Schumacher, A., Baker, S.P., and Akbarian, S. (2007). Prefrontal dysfunction in schizophrenia involves mixed-lineage

leukemia 1-regulated histone methylation at GABAergic gene promoters. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, 11254-11262.

Ionita-Laza, I., Xu, B., Makarov, V., Buxbaum, J.D., Roos, J.L., Gogos, J.A., and Karayiorgou, M. (2014). Scan statistic-based analysis of exome sequencing data identifies FAN1 at 15q13.3 as a susceptibility gene for schizophrenia and autism. *Proceedings of the National Academy of Sciences of the United States of America* 111, 343-348.

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.

Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P., *et al.* (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *American journal of human genetics* 76, 227-236.

Jiang, Y., Wang, Y., and Brudno, M. (2012). PRISM: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics* 28, 2576-2583.

Karayiorgou, M., Flint, J., Gogos, J.A., Malenka, R.C., Genetic, and Neural Complexity in Psychiatry Working, G. (2012). The best of times, the worst of times for psychiatric disease. *Nature neuroscience* 15, 811-812.

Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., Lasseter, V.K., *et al.* (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proceedings of the National Academy of Sciences of the United States of America* 92, 7612-7616.

Kenny, E.M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., Kelleher, E., Ennis, S., Tropea, D., Anney, R., *et al.* (2013). Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Molecular psychiatry*.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., *et al.* (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry* 17, 142-153.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., *et al.* (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77, 235-242.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., *et al.* (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823-828.

Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., *et al.* (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* 72, 951-963.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 148, 1223-1241.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20, 1297-1303.

Merla, G., Brunetti-Pierri, N., Micale, L., and Fusco, C. (2010). Copy number variants at Williams-Beuren syndrome 7q11.23 region. *Human genetics* 128, 3-26.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* 98, 12902-12907.

Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews Cancer* 10, 721-728.

Mukai, J., Dhillia, A., Drew, L.J., Stark, K.L., Cao, L., MacDermott, A.B., Karayiorgou, M., and Gogos, J.A. (2008). Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. *Nature neuroscience* 11, 1302-1310.

Nacak, T.G., Leptien, K., Fellner, D., Augustin, H.G., and Kroll, J. (2006). The BTB-kelch protein LZTR-1 is a novel Golgi protein that is degraded upon induction of apoptosis. *The Journal of biological chemistry* 281, 5065-5071.

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.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., *et al.* (2012a). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619-1622.

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.* (2012b). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246-250.

Otto, S.P., and Hastings, I.M. (1998). Mutation and selection within the individual. *Genetica* 102-103, 507-524.

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.

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Ende, 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.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics* 45, 1150-1159.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nature biotechnology* 29, 24-26.

Rodriguez-Murillo, L., Gogos, J.A., and Karayiorgou, M. (2012). The genetic architecture of schizophrenia: new mutations and emerging paradigms. *Annual review of medicine* 63, 63-80.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* 20, 7137-7148.

Ronan, J.L., Wu, W., and Crabtree, G.R. (2013). From neural development to cognition: unexpected roles for chromatin. *Nature reviews Genetics* 14, 347-359.

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.

Schlichter, A., and Cairns, B.R. (2005). Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *The EMBO journal* 24, 1222-1231.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry* 81, 65-95.

Shulha, H.P., Cheung, I., Guo, Y., Akbarian, S., and Weng, Z. (2013). Coordinated cell type-specific epigenetic remodeling in prefrontal cortex begins before birth and continues into early adulthood. *PLoS genetics* 9, e1003433.

The 1000 Genomes Project Consortium (2012). An integrated map of genetic variation

from 1,092 human genomes. *Nature* 491, 56-65.

Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. *Nature reviews Genetics* 13, 565-575.

Wynder, C., Hakimi, M.A., Epstein, J.A., Shilatifard, A., and Shiekhattar, R. (2005). Recruitment of MLL by HMG-domain protein iBRAF promotes neural differentiation. *Nature cell biology* 7, 1113-1117.

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. (2012). 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.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature genetics* 40, 880-885.

Yates, J.A., Menon, T., Thompson, B.A., and Bochar, D.A. (2010). Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8. *FEBS letters* 584, 689-693.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865-2871.

Yu, T.W., Chahrour, M.H., Coulter, M.E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D.A., Adli, M., Malik, A.N., *et al.* (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77, 259-273.

Zhu, Q., Ge, D., Maia, J.M., Zhu, M., Petrovski, S., Dickson, S.P., Heinzen, E.L., Shianna, K.V., and Goldstein, D.B. (2011). A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *American journal of human genetics* 88, 458-468.