



Published in final edited form as:

*Mol Psychiatry*. 2014 June ; 19(6): 652–658. doi:10.1038/mp.2014.29.

## **De novo Mutations in Schizophrenia Implicate Chromatin Remodeling and Support a Genetic Overlap with Autism and Intellectual Disability**

Shane E. McCarthy<sup>1,¶</sup>, Jesse Gillis<sup>1</sup>, Melissa Kramer<sup>1</sup>, Jayon Lihm<sup>1</sup>, Seungtai Yoon<sup>1</sup>, Yael Berstein<sup>1</sup>, Meeta Mistry<sup>2</sup>, Paul Pavlidis<sup>2</sup>, Rebecca Solomon<sup>1</sup>, Elena Ghiban<sup>1</sup>, Eric Antoniou<sup>1</sup>, Eric Kelleher<sup>3</sup>, Carol O'Brien<sup>3</sup>, Gary Donohoe<sup>3</sup>, Michael Gill<sup>3</sup>, Derek W. Morris<sup>3</sup>, W. Richard. McCombie<sup>1,4,¶</sup>, and Aiden Corvin<sup>3,¶</sup>

<sup>1</sup>The Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724, USA <sup>2</sup>Department of Psychiatry and Centre for High-throughput Biology, The University of British Columbia, Vancouver, Canada <sup>3</sup>Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin 2, Ireland <sup>4</sup>The Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724, USA

### **Abstract**

Schizophrenia is a serious psychiatric disorder with a broadly undiscovered genetic etiology. Recent studies of *de novo* mutations (DNM) in schizophrenia and autism have reinforced the hypothesis that rare genetic variation contributes to risk. We carried out exome sequencing on 57 trios with sporadic or familial schizophrenia. In sporadic trios, we observed a ~3.5-fold increase in the proportion of nonsense *de novo* mutations (DNMs) (0.101 vs. 0.031, empirical  $P=0.01$ , BH-corrected  $P=0.044$ ). These mutations were significantly more likely to occur in genes with highly ranked probabilities of haploinsufficiency ( $P=0.0029$ , corrected  $P=0.006$ ). DNMs of potential functional consequence were also found to occur in genes predicted to be less tolerant to rare variation ( $P=2.01\times 10^{-5}$ , corrected  $P=2.1\times 10^{-3}$ ). Genes with DNMs overlapped with genes

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

¶Correspondence: Dr. S. McCarthy, Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724, USA. Telephone: +516-422-4165, Fax: +516-422-4109, mccarthy@cshl.edu. Dr. W.R. McCombie, Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724, USA. Telephone: +516-422-4083, Fax: +516-422-4109, mcombie@cshl.edu. Prof. Aiden Corvin, Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine and Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. Telephone: + 353 1 896 2468, acorvin@tcd.ie.

### **AUTHOR CONTRIBUTIONS**

S.E.M., W.R.M., and A.C., designed the study. D.W.M., E.K., C.O.B., G.D., M.G., A.C collected all genomic DNA samples for the study. A.C made the final diagnoses. M.K., R.S., E.G., E.A., organized and performed exome capture, sequencing, data handling and validation genotyping. M.M and P.P provided data from Neurocarta and co-expression meta-analysis data. S.E.M, J.G., J.L., S.Y. performed the analysis. S.E.M., J.G., S.Y. R.S., W.R.M. and A.C. wrote the paper with comments and critique from all co-authors.

### **CONFLICT OF INTEREST**

W.R.M. has participated in Illumina sponsored meetings over the past four years and received travel reimbursement and an honorarium for presenting at these events. Illumina had no role in decisions relating to the study/work to be published, data collection and analysis of data and the decision to publish.

W.R.M. has participated in Pacific Biosciences sponsored meetings over the past three years and received travel reimbursement for presenting at these events.

W.R.M. is a founder and shared holder of Orion Genomics, which focuses on plant genomics and cancer genetics.

implicated in autism (e.g. *AUTS2*, *CDH8*, *MECP2*) and intellectual disability (ID) (e.g. *HUWE1* and *TRAPPC9*), supporting a shared genetic etiology between these disorders. Functionally *CHD8*, *MECP2* and *HUWE1* converge on epigenetic regulation of transcription suggesting that this may be an important risk mechanism. Our results were consistent in an analysis of additional exome based sequencing studies of other neurodevelopmental disorders. These findings suggest that perturbations in genes which function in the epigenetic regulation of brain development and cognition could have a central role in the susceptibility to, pathogenesis, and treatment of mental disorders.

## Keywords

*De novo* Mutations; Schizophrenia; Autism; CHD8; MECP2

---

## Introduction

Schizophrenia is a complex brain disorder affecting perception, thinking, behavior, cognition and social functioning. The disorder affects about 1% of the adult population and is a huge burden for those diagnosed, their families, and society. As is the case with most psychiatric disorders, schizophrenia is a syndromal diagnosis based on observed behavior, duration of symptoms and impaired function rather than on a biological understanding of disease etiology. This has significantly hindered progress in developing more precise diagnosis and better therapeutics to improve patient outcomes.

Schizophrenia is substantially heritable making it a target for human genetics research. As genomic technologies have improved a wide spectrum of genetic risk factors has emerged, encompassing common and rare risk variants, but also suggesting significant genetic heterogeneity within the patient population. Published genome-wide association studies (GWAS) have confirmed at least 20 common loci of small effect<sup>1, 2</sup>, with many more likely to be detected as sample sizes increase<sup>3</sup>. From GWAS data, it has also been possible to estimate that common risk variants account for at least a quarter of the genetic contribution to schizophrenia risk<sup>4</sup> and that genetic risk overlaps with other psychiatric disorders, in particular bipolar disorder<sup>5, 6</sup>. Studies of rare variation identified recurrent copy number variants (CNV) which have a moderate or large effect on schizophrenia risk<sup>7-9</sup> but also implicate *de novo* mutation (DNM) mechanisms as a critical source of private, large effect risk variants in schizophrenia<sup>10, 11</sup>. Significantly, almost all of the confirmed CNVs are also risk factors for other neurodevelopmental disorders including autism, intellectual disability (ID) and seizure disorder. In many instances, for example the 1q21.1 deletion originally identified as a risk factor for schizophrenia, the CNV actually has a substantially greater effect on risk for developmental delay, intellectual disability, and autistic spectrum disorder<sup>12-14</sup>.

Exome sequencing studies of parent offspring trios<sup>15, 16</sup> and the accrued risk associated with greater paternal age<sup>17</sup> suggest that an increased rate of DNMs disrupting gene function (e.g. missense and nonsense mutations), could play a significant role in schizophrenia susceptibility. Similar findings have been reported for other severe neurodevelopmental

disorders, including autism<sup>18–21</sup> and intellectual disability<sup>22–24</sup>. Although the rate of functional DNMs may be increased in neurodevelopmental disorders, the genetic heterogeneity, the abundance of loss of function mutations in the genome of healthy individuals, and the abridgement in our understanding of immediate mutational effects on gene function and downstream biological processes makes pinpointing or prioritizing specific mutations difficult. In addition to more exome sequencing studies of trios with neurodevelopmental disorders, analytical approaches that overcome inherent analytical biases (e.g. the limited curation of biological resources) are necessary to elucidate disease pathogenesis.

In this study we have sequenced the exome of 171 individuals representing 42 sporadic and 15 familial trios with schizophrenia or a related psychotic condition to identify additional risk mutations. In our primary analysis we test the hypothesis that the rate of functional DNM is increased in the sporadic and familial trios group compared to the expected rate in unaffected individuals. We perform a hypothesis-free over-representation analysis using the ontology and annotations from Neurocarta to determine whether the genes with DNM were enriched in other neurodevelopmental disorders<sup>25</sup>. We assess if genes with DNMs in our dataset are over represented among highly specific chromatin remodeling genes based on protein domain data<sup>26, 27</sup> and in chromatin remodeling genes previously implicated in mental disorders<sup>28</sup>. Finally we evaluate the robustness of our findings in recently published exome data sets of schizophrenia, autism and intellectual disability trios.

## MATERIALS/SUBJECTS AND METHODS

### Sample Collection

All participants gave written informed consent according with local research ethics committee approval. Participants were screened for psychiatric disorder by a trained clinician and cases were interviewed using a structured clinical interview (Structured Clinical Interview for DSM-IV (SCID-P) (ISBN:0880489324)<sup>29</sup>. Diagnosis of a major psychotic disorder was made by the consensus lifetime best estimate method using DSM-IV criteria with all available information including interview, family history (or staff) report and chart review. All cases were over 18 years of age, of Irish origin (having all four grandparents born in Ireland) and had been screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition. Family History was defined by the absence or presence of psychosis in 1° or 2° relatives. Further details on ascertainment methodology are provided in (ref<sup>30</sup>) and information on the family history, age at onset, illness course and other clinical indices of included trios are provided in Supplementary Information.

### Exome Capture and Sequencing

Exome capture DNA library was performed using the Solution Phase Exome Capture method<sup>31</sup>, which is a compilation and optimization of the Bioo Scientific NEXTflex™ (Illumina Compatible) Sequencing Kit and NimbleGen, Inc SeqCap EZ Library and Technology Note: Targeted Sequencing with NimbleGen SeqCap EZ Libraries and Illumina TruSeq ® DNA Sample Preparation kit. Briefly, sonicated genomic DNA ranging

from 1–5µg was used to create TruSeq Barcoded libraries. Approximately 1µg of the pre-capture library was hybridized with the NimbleGen's SeqCap EZ Human Exome Library v2.0 probes (NimbleGen SeqCap EZ Exome User's Guide and TechNote for paired-end libraries.) for 72 hours at 47°C. Following dual PCR enrichment and QC evaluation, samples with Bioanalyzer traces resulting in broad peaks ranging from 250bp–850bp and producing the highest peak around 400bp (DNA insert plus adaptors) were pooled. Libraries were sequenced on 3 or 4 lanes on a HiSeq 2000 with a Paired End 101 run including a 7 reads indexing run for the barcode detection. Additional methodological details are provided in the Supplementary Information.

### Data Processing and Variant Calling

Sequence reads from the Illumina HiSeq 2000 runs were demultiplexed using the Illumina Casava v1.8 pipeline, aligned to hg19 using the BWA aligner<sup>32</sup>, allowing 2 mismatches in the 30-base seed. Alignments were then paired, imported to binary (bam) format, sorted and indexed using SAMtools<sup>32</sup>. Picard was then used to fix any mate pair information altered by the sorting. Bamtools<sup>33</sup> was used to filter alignments to retain only properly paired reads (reads aligned with appropriate insert size and orientation). PCR duplicates were removed using Picard. Bamtools<sup>33</sup> was then used to select alignments with a minimum mapping quality score of 20. Target coverage for each NimbleGen exome capture was assessed using Picard's HSMetrics utility, and both depth and breadth of coverage were reviewed for each sample. The Genome Analysis Toolkit<sup>34</sup> GATK was used for local read realignment around indels, and for base quality score recalibration using corrections for base position within the Illumina read, for sequence context, and for platform-reported quality. Filtering criteria are provided in the Supplementary Materials. Finally, the variant calls were processed with snpEff v 2\_0\_5b<sup>35</sup> to provide annotation.

### De Novo SNV and INDEL Discovery

Any proband SNV that was not present in either parent was considered for further analysis. All proband variants required >10x in all members of the trio. Variants were also filtered for segmental duplications and presence in the Exome Variant Server (EVS6500). Indels were filtered similar to Iossifov et al<sup>18</sup>, any Indel called in a proband was removed from further analysis if at least one read from either parent backed the Indel call. In addition 15 reads supporting the reference allele were required in all three members of the trio. Finally at least 5% of the reads in the proband were required to support the indel. Further Details are provided in the Supplementary Materials.

### Variant Annotation

Putative de novo variants affecting protein function were selected based on the SNPEFF annotations of these data. In line with the annotation of existing data sets ANNOVAR to annotated the selected SNVs and Indels<sup>36</sup>. Scores from SIFT<sup>37</sup>, PolyPhen2 (REF<sup>38</sup>), LRT<sup>39</sup>, MutationTaster<sup>40</sup>, were used to predict the amino acid effects on protein function. PhyloP<sup>41</sup>, GERP ++ (Ref<sup>42</sup>) and phastCons 46 way scores were determined to measure conservation of variant sites. The variants were then tested for allele frequency in the 1000 Genomes and again by the Exome Variant Server (6500).

### Confirming Relatedness

Trio relatedness was confirmed using a combination of two measures the Glabitz score<sup>43</sup> and the Square Sum of the difference in the number of alternative alleles between two individuals. We used a set of 17,855 common SNPs with alternative allele frequency between 0.45 and 0.55 with thresholds for relatedness scores at 0.79 and 9,000 for the Glabitz Score and Square Sum, respectively.

### SNV Mutation Rate and Differences in Functional Annotations

Mutation rates were determined within targets with a mean coverage >10x in each proband and respective parents jointly. Confidence Intervals for the mutation rates were determined using a two sided binomial exact test. These rates were compared to prior estimates described and tabulated in the supplementary note using binomial exact tests. We compared the distribution of missense, nonsense and silent mutations to data ascertained in recent studies further described in the Supplementary Materials (Table S8). Binomial exact test were used to determine the significance of the missense to silent and nonsense to missense ratios.

### Analysis of Haploinsufficiency and Residual Variation Intolerance Score (RVIS)

Functional analyses were limited to DNMs defined by the following categories (1) Broadly Damaging: de novo missense variants considered by one or more prediction algorithm to possibly alter or damage gene function; (2) Nonsense mutations; (3) Likely Gene Disruptive (LGD) mutations such as nonsense, frameshifts and canonical splice-sites; (4) LGD+ Broadly Damaging. For comparison we also assessed silent variants. Gene based gene-based probability of exhibiting haploinsufficiency and RVIS scores based on ESP6500 (All\_0.1%) were obtained from Huang et al<sup>44</sup> from Petrovski et al<sup>45</sup> respectively. The distribution of haploinsufficiency probabilities and ranked RVIS scores for genes in the defined mutational categories were compared to the remainder of the genome using a two-sided Wilcoxon rank sum test. To rule out a potential bias in on our observation due to gene size, 10,000 permutations were performed selecting genes randomly controlling for gene length and GC content.

### Disease Gene Ontology

We used 26762 human-phenotype associations across 6788 human genes in 2178 phenotype categories from Neurocarta<sup>46</sup> to assess the enrichment of disease ontologies in our list of genes. We treated the functional annotation of genes as being drawn from a pool of 19099 genes in UCSC GoldenPath “known genes” table. We used the hypergeometric distribution to calculate the significance of overlaps between Neurocarta phenotypes and DNM gene sets. The ontology data was filtered to disorders with 10 or more reliable gene-phenotype relationships (>0.4) in order to flatten representation biases, leaving 72 of the original 2178 phenotypes. Bootstrap permutations (10,000) were performed sampling genes with probabilities determined by coding length and GC content. Reported p-values were multiple-test corrected using the Benjamini-Hochberg correction.

## Enrichment of De Novo Mutations in Chromatin Modifier Genes

Pfam domain information for all genes in the genome was obtained from the UCSC genome browser and matched to logs odds ratio (LOR) for 3,469 Pfam protein domains obtained from Pu et al 2010<sup>26</sup>. Pfam domains with no LOR value as  $\min(\text{LOR})-1$ . A list of chromatin modifiers implicated in mental disorders was obtained from Ronan et al (Ref<sup>28</sup>). We used the hypergeometric test to calculate the significance of overlaps between chromatin modifiers (LOR>5 and the disease list) and DNM sets. Bootstrap permutations (10,000) using the same parameters as in the Gene Ontology analysis.

## Assessing Additional Neurodevelopmental and Healthy Trio Exome Data sets

Our cross disorder analysis included data from published trio-based exome studies representing two schizophrenia cohorts (Xu et al<sup>47</sup> and Gulsuner et al<sup>48</sup>); four Autism data sets (O’Roak et al<sup>20</sup>, Neale et al<sup>19</sup>, Iossifov et al<sup>18</sup>, Sanders et al<sup>21</sup>) and two Intellectual Disability cohorts (Rauch et al<sup>23</sup> and de Ligt et al<sup>22</sup>). For the purpose of this analysis, Afrikaner and US cohorts reported by Xu et al<sup>47</sup> in their schizophrenia study were analyzed separately. De novo SNV calls in these data were annotated and filtered using the same pipeline applied to our data. Additional analytical details are provided in the supplementary materials.

## URLS

PICARD: <http://picard.sourceforge.net/>

UCSC Hg19: <http://genome.ucsc.edu/>

snpEff: [http://snpeff.sourceforge.net/faq.html#What\\_effects\\_are\\_predicted?](http://snpeff.sourceforge.net/faq.html#What_effects_are_predicted?)

Exome Variant Server <http://evs.gs.washington.edu/EVS/>

1K Genomes Project: <http://www.1000genomes.org/>

RVIS: <http://chgv.org/GenicIntolerance/>

## Results

### De Novo Variant Discovery in Schizophrenia Trios

We performed whole exome sequencing on 57 complete parent-parent-offspring trios with schizophrenia or a related psychiatric condition, composing 42 “sporadic” trios and 15 “familial trios” defined by the absence or presence of psychosis in 1<sup>o</sup> or 2<sup>o</sup> relatives, respectively (Supplementary Information, Table S1). On average, 94.2M properly paired reads mapped to the human exome reference (target size~36MB) for sporadic trios providing a mean coverage of 67X with over 90% of the exome covered at 10X or greater (Supplementary Information, Table S2). The number of mapped reads and mean coverage was higher for familial trios, however with little gain in the breadth of coverage at 10x (Supplementary Information, Table S2).

Proband calls were filtered for coverage (>10X), for parental variants, and for presence in the Exome Variant Server 6500 and 1000 Genomes (URL). Fifty-nine exonic *de novo* variants validated by Sanger Sequencing including 58 *de novo* SNVs (dnSNVs), one *de novo* dinucleotide variant (dnDNV) (Table S3). The combined effect of both adjacent dnDNV nucleotide substitutions introduced a stop codon in *SEC31A* and therefore the dnDNV was considered a nonsense variant in down stream analysis.

Of the 59 exonic dnSNVs, 47 and 12 were present in sporadic and familial trios, respectively. In sporadic trios, 28/47, 5/47 and 14/47 DNMs were classified as missense nonsense and silent mutations respectively. In familial trios the, 10/12 and 2/10 DNMs were classified as missense and silent, respectively. The number of dnSNVs per sporadic trio was higher (1.12) than for familial trios (0.8) however, the difference was not consistently significant (Supplementary Information). In both cohorts the distribution of DNMs was consistent with an expected Poisson distribution (Supplementary Information, Table S4, S5). The overall mutation rates observed in sporadic ( $1.62 \times 10^{-8}$ ) and familial trios ( $1.16 \times 10^{-8}$ ) were within range of rates observed in previous studies (Supplementary Information, Table S6). We did not see a correlation between paternal age and the number of DNMs per trio in this relatively small dataset. Likely *de novo* INDELS were filtered similarly to Iossifov *et al.* (Supplementary Information). Six *de novo* INDELS (dnINDs) were detected and validated by Sanger sequencing five of which were predicted to generate amino acid frameshifts and present in the sporadic trios

### Distribution of Exonic DNMs in Schizophrenia Trios

In contrast to the distribution of exonic DNMs in healthy trios previously used in exome sequencing studies of autism (Supplementary Information, Table S7), there was no significant difference in the proportion of missense DNM in the sporadic trios however the proportion of nonsense DNMs was increased approximately 3.5-fold (0.101 vs. 0.031, empirical  $P=0.01$ , BH-corrected  $P=0.044$ , Supplementary Information, Table S8). Furthermore, the ratio of nonsense to missense DNMs in sporadic trios was also significantly greater than expected ( $P=0.01$ ) (Supplementary Table 9). In familial trios, although there was proportionally more missense than silent variants (ratio 5:1), the difference was not significant. Nonsense mutations were not identified in the familial schizophrenia trios.

### Haploinsufficiency and Intolerance Analysis of Genes with Exonic DNMs

To prioritize likely candidate mutations based on functional impact, the distributions of haploinsufficiency and Residual Variation Intolerance Score (RVIS)<sup>45</sup> were analyzed in five mutation groups defined by (1) Broadly damaging missense: DNMs that are potentially damaging by one or more prediction algorithms; (2) Nonsense; (3) Likely gene disruptive (LGD): nonsense, frameshifts and splice sites; (4) LGD and Broadly Damaging Missense and (5) silent mutations.

Relative to genomewide predications, genes with nonsense DNM in sporadic trios had significantly higher probabilities of haploinsufficiency (Supplementary Figure 2  $P=0.0029$ , BH-corrected  $P=0.015$ ). This remained significant after simulations controlling for gene size

and GC content ( $P=0.0012$ , BH-corrected  $P=0.006$ ). All but one gene ranked in the top 15% of probable haploinsufficient genes (Table 1). The genes in other mutational group did not show significantly higher probabilities of haploinsufficiency

Similarly, RVIS scores of genes with nonsense DNMs ranked significantly higher relative to genome-wide predictions ( $P=0.0013$ , BH-Corrected  $P=0.0022$ ). This effect was even more evident for genes with broadly damaging missense DNMs ( $P=2\times 10^{-4}$ , BH-Corrected  $P=5\times 10^{-4}$ ). These results remained significant in simulations controlling for gene size and GC content (Nonsense:  $P=0.0027$ , BH-corrected  $P=0.0067$ ; Broadly Damaging:  $P=0.0081$ , BH-corrected  $P=0.014$ ). Collectively, the RVIS scores for genes with LGD and broadly damaging mutations ranked significantly higher relative to the remainder of the genome before and after controlling for gene size and GC content ( $P=2.01\times 10^{-5}$ , GC-Size-BH-Corrected= $2.1\times 10^{-3}$ ). Genes with silent DNMs did not show any difference relative to the genome. Genes ranked in the top 15% RVIS intolerant scores are shown in Table 1.

Notably, among genes ranked in the top 15% of RVIS scores, nonsense DNMs in sporadic trios were identified in Chromodomain Helicase DNA Binding Protein 8 (*CHD8*), Autism Susceptibility Locus 2 (*AUTS2*), Histone Lysine Methyltransferase 2 gene (*MLL2*). Prior genetic evidence suggests that *CHD8*, *AUTS2* and *MLL2* may have an important role in the risk and pathogenesis of neurodevelopmental disorders. Broadly damaging DNMs discovered in other genes implicated in neurodevelopmental disorders such methyl-DNA binding protein, *MECP2*, E3 ubiquitin-protein ligase *HUWE1*, and Trafficking Protein Particle Complex 9 (*TRAPPC9*) may also be of etiological relevance in this cohort. Recurrent broadly damaging missense DNMs in *PITPNM1* observed in this study and another schizophrenia cohort suggests this gene may also be of importance.

### Enrichment Analysis of DNMs in ASD/ID Implicated Genes

In a hypothesis-free over-representation analysis using high quality disease ontology annotations ( $\geq 10$  genes,  $>0.4$  quality score, yielding 72 phenotypes) from Neurocarta<sup>25</sup>, the ontologies, “autism spectrum disorders”, “autistic disorders” and “intellectual disability” were the most significantly over-represented disorders in all mutational categories assessed except for silent mutations. After correcting for assessment across this broad set of phenotypes, these disorders remained the highest ranked disorders in all mutation categories, with significant evidence for enrichment of LGD+Broadly damaging de novo’s in “autistic disorders” (BH-corrected  $P=0.02$ ). The enrichment of autism and intellectual disability among genes with de novo mutations was significant in repeated simulations controlling for gene size and GC content. This provides support for a specific overlap between schizophrenia and autism at the gene level. Genes factoring in this enrichment were, *CHD8*, *MECP2*, *AUTS2*, *HUWE1*, and *TRAPPC9*.

### DNMs in Chromatin Modifiers

The convergent molecular functions of *CHD8*, *MECP2* and *HUWE1* support growing hypotheses that epigenetic regulation of transcription could represent a shared molecular “risk” mechanism in neurodevelopmental disorders, including autism and ID<sup>2849</sup>. Indeed, across all mutational categories except silent mutations, there was a significant over-



representation of genes associated with mental disorders involved in chromatin organization, the most significant of which was observed for genes with LGD+Broadly damaging de novos (BH-corrected  $P=7\times 10^{-6}$ ). This was significant in repeated simulations that also controlled for gene size and GC content (BH-corrected  $>0.0001$ ). Overall, nonsense DNMs were significantly enriched among a set of 419 genes characterized by domains highly specific ( $LOR>5$ ) to chromatin modification ( $P=0.0046$ , BH-corrected  $P=0.023$ ) (Supplementary Information) This association was largely contributed by *CHD8* and *MLL2*.

### Consistency Across Exome Sequencing Studies of other Neurodevelopmental Disorders

To validate this finding we assessed the robustness of our results using data from nine larger exome sequencing studies of trios with neurodevelopmental disorders including schizophrenia ( $n=3$ ), autism ( $n=4$ ), intellectual disability ( $n=2$ ) as well as six “healthy” siblings/controls (Supplementary Information).

Following consistent annotation and filtering, a significant increase in the proportion of de novo nonsense mutations was observed in 3 of 9 neurodevelopmental datasets but only 1 of 6 unaffected datasets. (Supplementary Information, Table S10). Including the current study, this increase in the proportion of nonsense variation was observed in 40% (4/10) of the neurodevelopmental cohorts. Six of nine additional disease data sets (66%) had significantly more haploinsufficient genes in one of the functional classes with potentially damaging or disrupting mutations, compared to only one control data set (Supplementary Table S11). Similarly, 7/9 additional disease data sets had significantly more RVIS-based intolerant genes with broadly damaging or disrupting mutations compared to 2/6 control data sets. Combined with the current study this increase was observed in 80% of exome-based neurodevelopment cohorts (Supplementary Table S12).

The only disease ontologies that remained significant in these additional data sets after correction for multiple tests and gene size were “intellectual disability”, “autistic disorder”, “autism spectrum disorder” and “infantile epilepsy”. An over-representation of functional mutations was specifically observed in 3/9 neurodevelopmental data sets (Supplementary Table S15). Including our data, these ontologies were enriched in 4/10 (40%) disease data sets analyzed, but not in any control data set.

Finally, 3 out of 9 additional neurodevelopmental data sets had an over-representation of LGD mutations in chromatin remodeling genes implicated in mental disorders (Supplementary Table S14). Including our study, this enrichment was observed in 40% (4/10) of the neurodevelopmental data sets analyzed but not in any control data set. Although, the broader but highly specific chromatin modifier gene set was not over-represented among nonsense mutations in any of the additional neurodevelopmental data sets analyzed, they were enriched in other functional mutation classes in 5/9 neurodevelopmental data sets (6/10 including our data) compared to 1 control data set (Supplementary Table S12).

## Discussion

Motivated by the growing interest in identifying ultra rare, potentially highly penetrant, genetic variants underlying the pathogenesis of psychiatric and neurodevelopmental disorders, we describe the exome sequencing 57 parent-offspring trios with schizophrenia or a related psychotic disorder. In our analysis of sporadic trios, we observed a higher than expected proportion of nonsense DNMs. We also found that genes with potentially functional mutations ranked significantly less intolerant to rare variation, complementing recently proposed hypotheses that DNMs may be significant risk factors for sporadic schizophrenia<sup>16, 47</sup>. We also provide supporting evidence that schizophrenia shares a genetic etiology with autism and ID and highlight specific genes with roles in chromatin modification proposing a potential molecular disease mechanism shared by these diseases. The analysis of additional exome studies of neurodevelopmental disorders supports these findings.

Although categorical family history information represents a relatively crude measure of genetic liability, in our dataset, this distinction was sufficient to identify significant increases in the rate of nonsense mutations in sporadic trios relative to familial and healthy controls reinforcing recently emerging hypotheses that DNMs with a greater likelihood of disrupting gene function could play a significant etiological role in neurodevelopmental disorders<sup>16, 18–24</sup>. While DNM in sporadic cases is insufficient to confirm causality, 4 of 5 nonsense DNMs occurred in genes with high probabilities of haploinsufficiency and 3 of 5 genes have been previously implicated in other neurodevelopmental disorders (*CHD8*, *MLL2*, *AUTS2*) increasing the possibility they are highly sensitive to inactivating mutations and significant risk factors for schizophrenia. Larger studies will be important in guiding future gene discovery and widening our perspective on the genetic architecture and allelic diversity of the disorder<sup>50</sup>.

Consistent with the growing epidemiological<sup>51, 52</sup> and genetic evidence for a shared etiology between neurodevelopmental disorders, we identified an overlap between genes in several trio-based exome studies sets and autism as well as intellectual disability. Interestingly over-representation of genes with potentially functional mutations was restricted to affected trios and not unaffected trios. Although the enrichment of these diseases was not evident across all of the neurodevelopmental disorder data sets analyzed, autism, ID and schizophrenia are unlikely to represent single disease entities and a substantial genetic and etiological heterogeneity is captured by current neurodevelopmental disorder classification. The analysis of much larger cohorts will be required to identify key points of similarity and difference within these patient groups and as well as support for continuing revisions of gene-phenotype ontologies and quality assignments in Neurocarta.

In contrast to GWAS or CNV studies, the granularity of exome data allows the ability to pinpoint, at a higher resolution, potential molecular risk genes and mechanisms tractable to further investigation. We found a significant over representation of potentially functional DNMs in genes containing domains necessary for editing, reading, writing of histone posttranslational marks and DNA methylation complementing neurobiological findings that epigenetic and retrotransposition regulation play an important role in

neurodevelopment<sup>53–55</sup>. Mutations were especially enriched in chromatin modifying genes already implicated in mental disorders such as *CHD8*, *MECP2*, and *HUWE1* suggesting the importance of genes that have evolutionarily impacted the epigenetic regulation of brain development and cognitive function in humans<sup>26, 56</sup> as having a shared central role in the susceptibility to, pathogenesis, and treatment of neurodevelopmental diseases.

Few of the de novo variants discovered in sporadic trios are expected to be highly penetrant mutations<sup>21</sup> (Supplementary Material). Consequently, gene prioritization based on disease and function ontologies (e.g. Neurocarta, GO), and network inference (e.g. protein-protein interaction), is important. However, these approaches are not without limitations. Despite systematic ascertainment, the population of gene-phenotype databases with insufficient literature evidence or weak experimental validity can mislead gene-disease ontologies. Furthermore the ambiguity between mutational effects on gene function and biological processes can bias variant prioritization based on protein-protein interactions and molecular pathways towards genes that are well-studied, are of high node degree or are multifunctional. Probabilistic approaches based on recurrence may circumvent biological information but if the likelihood of recurrence is low, this approach could have limited application in diseases with broad genetic heterogeneity<sup>57, 58</sup>. Alternatively developing a network level understanding of genetic mechanisms by combining complementary genomic datasets that are incognizant of the underlying biology of disease to may be critical to prioritize genes and providing novel insights into disease pathogenesis (Supplementary Information).

We have preliminarily used gene expression data from the prefrontal cortex, brain and non-brain to build novel co-expression networks and identified a brain-region specific role for other DNMs, suggesting a new approach to prioritizing future candidate disease-causing variants (Supplementary Information). We found that genes with putatively highly damaging variants were preferentially not co-expressed (low connectivity) with one another in the prefrontal cortex in both the control and schizophrenia conditions. This low connectivity did not extrapolate to the full brain or non-brain network, suggesting that genes such as *AUTS2* and *NIP7* may have particular, independent functions in networks specifically in the prefrontal cortex. The low connectivity of likely disrupting variants that we observed is consistent with the low node degree of “nonessential” disease genes<sup>59</sup> which are likely under less but still purifying selection than essential higher node degree hub genes<sup>60, 61</sup>. Furthermore the specificity of this low connectivity in the prefrontal cortex, a region of the brain highly relevant to schizophrenia, is concordant with the confined expression of nonessential disease genes to specialized tissues. This is in contrast to hub genes of high node degree, which show widespread expression in multiple tissues. We suggest tailoring of co-expression networks to control for both brain region and disease state represents a potentially interesting approach to determine molecular mechanisms for further research and supports the utility of complementing standard methods with agnostic approaches to prioritize novel disease candidates on genome-wide scales.

In conclusion, our results indicate that potentially functional and deleterious DNMs may contribute to the risk of schizophrenia and are consistent with prior exome studies. Genetic and phenotypic diversity represent a challenge for population-based association approaches

and may require a broader inclusion of neurodevelopmental phenotypes in assessment of identified risk genes. Our results provide a defined set of genes that support the genetic overlap between schizophrenia and autism, some of which may have a role in chromatin modeling and epigenetic regulation. The identification of these biological functions as potential contributors to the etiology of schizophrenia is, until very recently very unexpected but is consistent with recent observations in autism genetics (ref CHD8)<sup>58</sup> A caveat of our study is sample size and it will be necessary to assess these findings in larger cohorts. However, given the genetic and biological heterogeneity of neurodevelopmental disorders, novel findings require increased granularity which may be shadowed by the reliability of larger cohort studies on formal network analysis that innately rely on limited and biased annotations. As the number of exome studies increase, a more refined set of genes spanning the broad heterogeneity of autism, schizophrenia and other psychiatric disorders will emerge and the incorporation of complementary genomic data may elucidate pathways and mechanisms such as epigenetic regulation that are critical to the development, and ultimately the treatment, of neuropsychiatric conditions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Stephanie Muller, Gloria Cheang, Senem Mavruk, Manasa Kolli, Nabil Azamy, Anthony DeSantis, Patricia Mocombe for contributing their expertise and help to perform exome capture, sequencing, validations and genotyping. We are grateful to Elodie Portales-Casamar for assistance with Neurocarta. We thank Jianchao Yao for helpful discussion and feedback.

Funding for this study was provided by grants from T. and V. Stanley to S.E.M., J.G., M.K., J.L., S.Y., Y.B., R.S., E.G., E.A., and W.R.M and from Science Foundation Ireland (08/IN.1/B1916) to A.C., M.G. and D.M.. P.P. was supported by NIH Grant GM076990 and salary awards from the Michael Smith Foundation for Health Research and the Canadian Institutes for Health Research

## References

1. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 2011; 43(10):969–976.
2. Ripke S, O’Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* 2013; 45(10):1150–1159.
3. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron.* 2010; 68(2):182–186.
4. Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet.* 2012; 44(3): 247–250.
5. Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009; 460(7256):748–752.
6. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013; 381(9875):1371–1379. [PubMed: 23453885]

7. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008; 455(7210):232–236. [PubMed: 18668039]
8. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11. 2 and autism. *N Engl J Med*. 2008; 358(7):667–675. [PubMed: 18184952]
9. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11. 2 are associated with schizophrenia. *Nat Genet*. 2009; 41(11):1223–1227.
10. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet*. 2000; 9(9):1415–1423. [PubMed: 10814723]
11. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008; 40(7):880–885. [PubMed: 18511947]
12. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*. 2011; 72(6):951–963.
13. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol*. 2012
14. Haldeman-Englert, C.; Jewett, T. 1q21.1 Microdeletion. In: Pagon, RA.; Adam, MP.; Bird, TD.; Dolan, CR.; Fong, CT.; Stephens, K., editors. *GeneReviews*. Seattle (WA): 1993.
15. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet*. 2011; 43(9):864–868.
16. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet*. 2011; 43(9):860–863. [PubMed: 21743468]
17. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 2012; 488(7412):471–475. [PubMed: 22914163]
18. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012; 74(2):285–299. [PubMed: 22542183]
19. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012; 485(7397):242–245. [PubMed: 22495311]
20. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012; 485(7397):246–250. [PubMed: 22495309]
21. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012; 485(7397):237–241. [PubMed: 22495306]
22. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*. 2012; 367(20):1921–1929. [PubMed: 23033978]
23. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*. 2012; 380(9854):1674–1682. [PubMed: 23020937]
24. Vissers LE, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nat Genet*. 2010; 42(12):1109–1112. [PubMed: 21076407]
25. Portales-Casamar E, Ch'ng C, Lui F, St-Georges N, Zoubarev A, Lai AY, et al. Neurocarta: aggregating and sharing disease-gene relations for the neurosciences. *BMC Genomics*. 2013; 14:129. [PubMed: 23442263]
26. Pu S, Turinsky AL, Vlasblom J, On T, Xiong X, Emili A, et al. Expanding the landscape of chromatin modification (CM)-related functional domains and genes in human. *PLoS One*. 2010; 5(11):e14122. [PubMed: 21124763]

27. Turinsky AL, Turner B, Borja RC, Gleeson JA, Heath M, Pu S, et al. DAnCER: disease-annotated chromatin epigenetics resource. *Nucleic Acids Res.* 2011; 39(Database issue):D889–894. [PubMed: 20876685]
28. Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet.* 2013; 14(5):347–359. [PubMed: 23568486]
29. American Psychiatric Association., American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV.* 4. American Psychiatric Association; Washington, DC: 1994. Task Force on DSM-IV; p. xxviip. 886
30. Strange A, Riley BP, Spenser CA, Morris DW, Pirinen M, O’Dushlaine CT, et al. Genome-wide association study implicates HLA-C\*01:02 as a risk factor at the MHC locus in schizophrenia. *Biol Psychiatry.* 2012 In Press.
31. Green, MR.; Russell, DW. *Molecular cloning: a laboratory manual.* 4. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y: 2012.
32. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–1760. [PubMed: 19451168]
33. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics.* 2011; 27(12):1691–1692. [PubMed: 21493652]
34. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43(5):491–498. [PubMed: 21478889]
35. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. 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 (Austin).* 2012; 6(2):80–92. [PubMed: 22728672]
36. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38(16):e164. [PubMed: 20601685]
37. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009; 4(7):1073–1081. [PubMed: 19561590]
38. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4):248–249. [PubMed: 20354512]
39. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res.* 2009; 19(9):1553–1561. [PubMed: 19602639]
40. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.* 2010; 7(8):575–576. [PubMed: 20676075]
41. Siepel, A.; Pollard, KSDH. New methods for detecting lineage-specific selection. *Proceedings of the 10th International Conference on Research in Computational Molecular Biology (RECOMB 2006); 2006.* p. 190-205.
42. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 2010; 6(12):e1001025. [PubMed: 21152010]
43. Glaubitz JC, Rhodes OE, Dewoody JA. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol Ecol.* 2003; 12(4):1039–1047. [PubMed: 12753222]
44. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* 2010; 6(10):e1001154. [PubMed: 20976243]
45. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 2013; 9(8):e1003709. [PubMed: 23990802]
46. Zoubarev A, Hamer KM, Keshav KD, McCarthy EL, Santos JR, Van Rossum T, et al. Gemma: A resource for the re-use, sharing and meta-analysis of expression profiling data. *Bioinformatics.* 2012
47. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet.* 2012

48. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*. 2013; 154(3):518–529. [PubMed: 23911319]
49. Houston I, Peter CJ, Mitchell A, Straubhaar J, Rogaev E, Akbarian S. Epigenetics in the human brain. *Neuropsychopharmacology*. 2013; 38(1):183–197. [PubMed: 22643929]
50. Walsh CA, Engle EC. Allelic diversity in human developmental neurogenetics: insights into biology and disease. *Neuron*. 2010; 68(2):245–253. [PubMed: 20955932]
51. Morgan VA, Croft ML, Valuri GM, Zubrick SR, Bower C, McNeil TF, et al. Intellectual disability and other neuropsychiatric outcomes in high-risk children of mothers with schizophrenia, bipolar disorder and unipolar major depression. *Br J Psychiatry*. 2012; 200(4):282–289. [PubMed: 22241931]
52. Sullivan PF, Magnusson C, Reichenberg A, Boman M, Dalman C, Davidson M, et al. Family History of Schizophrenia and Bipolar Disorder as Risk Factors for Autism Family History of Psychosis as Risk Factor for ASD. *Arch Gen Psychiatry*. 2012:1–5.
53. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, et al. L1 retrotransposition in neurons is modulated by MeCP2. *Nature*. 2010; 468(7322):443–446. [PubMed: 21085180]
54. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009; 460(7259):1127–1131. [PubMed: 19657334]
55. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011; 479(7374):534–537. [PubMed: 22037309]
56. Vogel C, Chothia C. Protein family expansions and biological complexity. *PLoS Comput Biol*. 2006; 2(5):e48. [PubMed: 16733546]
57. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*. 2012; 151(7):1431–1442. [PubMed: 23260136]
58. O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*. 2012; 338(6114):1619–1622. [PubMed: 23160955]
59. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007; 104(21):8685–8690. [PubMed: 17502601]
60. Khurana E, Fu Y, Chen J, Gerstein M. Interpretation of genomic variants using a unified biological network approach. *PLoS Comput Biol*. 2013; 9(3):e1002886. [PubMed: 23505346]
61. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012; 335(6070):823–828. [PubMed: 22344438]

**Table 1**

De novo mutations in the Top 15% Ranked RVIS Genes

Chr <sup>d</sup>	Pos	Function <sup>b</sup>	Gene	AA Change	Base Change	Trio	FH <sup>c</sup>	Diagnosis <sup>d</sup>	AAO <sup>e</sup>	Sex <sup>f</sup>	Disease <sup>g</sup>	HI <sup>h</sup>	LOR>5 <sup>i</sup>	RVIS Rank <sup>j</sup>
12	49420670	nonSNV	MLL2	p.R5027X	C>T	51	F	SCZ	21	M	KS	0.647 9.9%	+	0.06
1	12316498	misSNV	VPS13D	p.R260W	C>T	19	F	SCZ	18	F	EV	0.231 38.1%	-	0.14
11	68204419	misSNV	LRP5	p.C1355R	T>C	53	F	SCZ	17	F	EV	0.066 88.0%	-	0.24
X	53654776	misSNV	HUWE1	p.V433I	G>A	4	F	SCZ	31	F	MR	0.816 5.4%	-	0.4
14	21860919	nonSNV	CHD8	p.S2173X	C>A	37	F	SCZ	18	M	ASD	0.637 10.3%	+	1.18
7	69364416	nonSNV	AUTS2	p.R152X	C>T	58	F	SCZ	15	F	ASD, EP	0.741 7.3%	-	1.82
8	140922446	misSNV	TRAPPC9	p.R970Q	G>A	37	F	SCZ	18	M	ID, MR	0.143 56.4%	-	2.77
16	14340403	misSNV	MKL2	p.R429H	G>A	32	F	SCZ	19	M	ASD	0.316 27.9%	-	3.27
11	3697753	misSNV	NUP98	p.R1650P	G>C	16	F	SCZ	14	M	AML	0.932 2.6%	-	5.01
11	67267698	misSNV	PITPNM1	p.R279W	C>T	53	F	SCZ	17	F	SCZ	0.106 70.4%	-	8.2
4	83770051,2	nonSNV	SEC31A	p.P764X	CC>AT	15	F	SCZ	24	M		0.328 26.7%	-	8.63
11	62414716	misSNV	INTS5	p.V946F	G>T	24	F	SCZ	15	M		0.274 32.4%	-	8.95
10	48390369	misSNV	RBP3	p.R170P	G>C	12	F	SA	19	M		0.340 25.8%	-	9.02
X	153296711	misSNV	MECP2	p.R202C	C>T	17	F	SCZ	24	F	ASD, RS	0.359 24.3%	+	10.46
2	43958706	misSNV	PLEKHH2	p.E970K	G>A	39	F	SCZ	18	F		0.104 71.0%	-	11.68
10	111642192	misSNV	XPNPEP1	p.K347Q	A>C	49	F	SCZ	19	M		0.155 52.8%	-	14.02

<sup>a</sup>Chromosome

<sup>b</sup>Exonic functions. misSNV, missense SNV; nonSNV, nonsense SNV. The de novo dinucleotide in SEC31A creates one stopcodon, however ANNOVAR annotations for both individual positions are provided in Supplementary Table S3.

<sup>c</sup>FH: Family History. F: Sporadic; T: Familial

<sup>d</sup>Diagnosis. SCZ: Schizophrenia; SA: Schizoaffective

<sup>e</sup>AAO. Age at Onset

<sup>f</sup>Sex M: Male; F: Female

<sup>g</sup>Associated Diseases. KS: Kabuki Syndrome; ASD: Autism Spectrum Disorder; SCZ: Schizophrenia; EV: Exudative Vitreoretinopathy; MR: Mental Retardation; EP: Epilepsy; ID: Intellectual Disability; AML: Acute Myeloid Leukemia; RS: Rett Syndrome



Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

<sup>i</sup>Hi: Probability of Haploinsufficiency and percentile rank from Huang et al 44  
<sup>j</sup>Genes with Pfam domains with LOR>1 that they are present or absent in chromatin modifying 26  
<sup>j</sup>Residual Variation Intolerance Ranks based on ESP6500 (ALL\_0.1%) 45