European Journal of Human Genetics (2016) 24, 944–948

www.nature.com/ejhg

EJHG Open

SHORT REPORT

De novo variants in sporadic cases of childhood onset schizophrenia

Amirthagowri Ambalavanan[1], Simon L Girard[1], Kwangmi Ahn[2], Sirui Zhou[3], Alexandre Dionne-Laporte[4], Dan Spiegelman[4], Cynthia V Bourassa[4], Julie Gauthier[3], Fadi F Hamdan[3], Lan Xiong[5], Patrick A Dion[4], Ridha Joober[1,6], Judith Rapoport[2] and Guy A Rouleau*[,1,4,7]

Childhood-onset schizophrenia (COS), defined by the onset of illness before age 13 years, is a rare severe neurodevelopmental disorder of unknown etiology. Recently, sequencing studies have identified rare, potentially causative *de novo* variants in sporadic cases of adult-onset schizophrenia and autism. In this study, we performed exome sequencing of 17 COS trios in order to test whether *de novo* variants could contribute to this disease. We identified 20 *de novo* variants in 17 COS probands, which is consistent with the *de novo* mutation rate reported in the adult form of the disease. Interestingly, the missense *de novo* variants in COS have a high likelihood for pathogenicity and were enriched for genes that are less tolerant to variants. Among the genes found disrupted in our study, *SEZ6, RYR2, GPR153, GTF2IRD1*, *TTBK1* and *ITGA6* have been previously linked to neuronal function or to psychiatric disorders, and thus may be considered as COS candidate genes.
European Journal of Human Genetics (2016) **24,** 944–948; doi:10.1038/ejhg.2015.218; published online 28 October 2015

INTRODUCTION

Schizophrenia is a major mental disorder characterized by a wide spectrum of symptoms, including delusions, hallucinations, disturbance of thinking processes and deterioration of social behaviors; the prevalence of schizophrenia is 1.1% in the US adult population.[1,2] Schizophrenia is a neurodevelopmental disorder characterized by social and cognitive developmental abnormalities, often with mild motor signs.[3,4] The average age of onset of schizophrenia is 18 years in men and 25 years in women.[5] However, there is a rare, very early onset form of the disease referred to as Child onset schizophrenia (COS), defined by onset at 13 years or younger. This rare type of schizophrenia is clinically and neurobiologically continuous with the adult-onset disorder.[6] COS patients have a high rate of comorbidity of developmental disorders such as autism spectrum disorder (ASD), motor developmental disorders and learning disorders.[6] A longitudinal study conducted on COS patients at the child psychiatry branch of National Institute of Mental Health (NIMH) reported the prevalence of COS to be ~1/40 000, with a gradual onset and outcome resembling adult cases.[7] As the prevalence of the disease is very low, less is known about the genetic architecture in COS.

Studies looking at family-based association and copy-number variations (CNVs) have identified some genetic variations underlying COS.[6,8,9] One study of structural variations (ie, microduplications and microdeletions) found that the overall frequency of CNVs is higher in COS patients compared with adult-onset patients and to population controls.[6,10,11]

In schizophrenia, our group found potentially deleterious *de novo* single base pair variants and showed that the rate of *de novo* variants in schizophrenia patients is higher than expected.[12,13] The aim of the current study was to measure the rate of *de novo* variants and to uncover possibly exclusive COS candidate genes, by analyzing *de novo* rare variants in sporadic COS trios.

MATERIALS AND METHODS

Samples and clinical characteristics

Seventeen sporadic COS cases (11 males and 6 females) and their unaffected parents were recruited for this study. Clinical information is included in the Supplementary Information. The mean age of onset was 9.8 years (range, 6–12 years old). Six of the patients were also diagnosed with ASD. We had previously examined the cases for CNV and the results were published in Ahn *et al*[11] (Supplementary Table 1). Informed consent was obtained from all participants.

Whole-exome sequencing

The exome capture of all 51 individuals in the COS trios was performed using SureSelect^{XT} Human All Exon V4 kit (Agilent Technologies Inc. Mississauga, ON, Canada). We prepared samples in two different batches. The first batch consisting of 13 COS trios (39 samples) was captured and sequenced using Illumina HiSeq 2000 at the McGill University and Genome Quebec Innovation Centre (Montreal, QC, Canada). The last batch of 4 COS trios (12 samples) was sequenced using the Illumina HiSeq 2000 platform at the Université de Montréal's Beaulieu-Saucier Pharmacogenomics Centre at the Montreal Heart Institute (Montreal, Canada).

Exome data analysis

The sequenced reads of all the samples from Illumina HiSeq2000 were aligned to the reference genome (GRCh37/hg19) using Burrow

[1]Department of Human Genetics, McGill University, Montreal, QC, Canada; [2]Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD, USA; [3]Department of Medicine, Université de Montréal, Montreal, QC, Canada; [4]Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada; [5]Department of Psychiatry, Université de Montréal, Montreal, QC, Canada; [6]Department of Psychiatry, Douglas Mental Health University Institute, Montreal, QC, Canada; [7]Montreal Neurological Institute and Hospital, Montreal, QC, Canada
*Correspondence: Dr G Rouleau, Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, 3801 University Street, Montreal, QC, Canada H3A 2B4. Tel: +1 514 398 3690; Fax: +1 514 398 8248; E-mail: guy.rouleau@mcgill.ca
Received 24 February 2015; revised 7 August 2015; accepted 18 August 2015; published online 28 October 2015

Wheeler's Algorithm.[14] The aligned reads were converted to binary format for the convenience of further analysis using SAM tools.[15] The efficiency of capture was calculated based on the percentage of exome covered by at least 20 reads (20×) of each sample. Samples obtained an average coverage of over 90% target covered at a depth of 20× (Supplementary Figure 1). The quality of coverage was assessed by the total number of reads mapped to corresponding regions in the reference genome, over the total number of uniquely mapped reads. Next, the variant calling was performed using Genome Analysis Tool Kit (GATK).[16] The variants were called for the sequenced reads available within the coverage region for each of the samples. This process identified single-nucleotide variants and small insertions or deletions at different levels of stringency based on their quality scores. The variants identified were annotated with ANNOVAR tool[17] to state the position of genes and their chromosomes, including their minor allele frequencies from publicly available databases.

De novo variant identification

We used an in-house program to segregate all the variants in COS trios. Our selection included: (1) all truncating variants, (2) splice site disrupting variants, (3) missense variants, and (4) synonymous variants. *De novo* variants are those that occur in the gametes before fertilization or immediately after fertilization.[18] We prioritized the *de novo* variants by excluding all inherited variants from each proband's parent's exome data, as a part of the filter. We determined the allele frequency of variants based on the variants reported in the Exome Variant Server (EVS; Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA, USA (URL: http://evs.gs.washington.edu/EVS/)).

Validation of candidate variants

To validate the candidate variants, we used polymerase chain reactions (PCR) and Sanger sequencing. Primers were designed for each of the potential *de novo* variants identified from COS cases and controls using Primer3.[19] Sanger sequencing was performed for every member of each family. The Sanger sequencing results were analyzed using Mutation Surveyor v.4.0 (SoftGenetics, Pennsylvania, USA).

RESULTS

In our 17 COS trios, we identified 20 exonic *de novo* variants (14 novel missenses, 2 novel deletions and 4 synonymous), as shown in Table 1. Among the 14 missense variants, (*RYR2* p. (Glu746Tyr)) had two nucleotide substitutions in the same codon which was termed a 'delins' (Supplementary Figure 3). By opposition to what we observed in one of our previous adult schizophrenia report[13] there were no nonsense variants identified in this study. Variants are available in

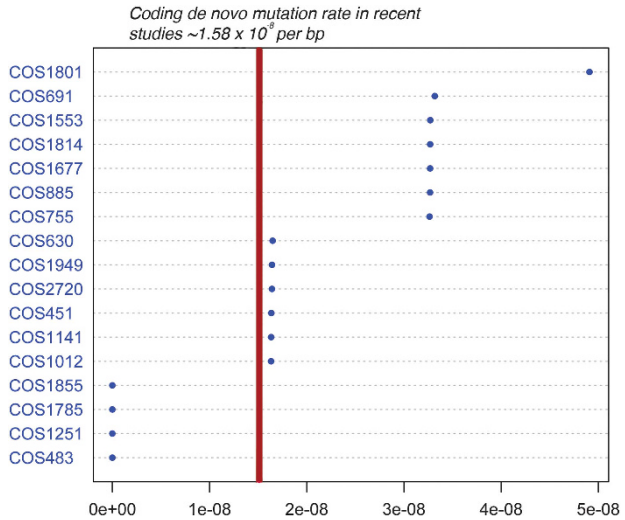

Figure 1 Dot plot shows the distribution of *de novo* variants in 17 COS probands. The *x* axis represents the *de novo* mutation rate per individual. The line in red indicates the fairly accurate coding *de novo* mutation rate reported so far.

Table 1 List of *de novo* variants in our study

Probands	*Position (hg19)*	*Gene*	*Nucleotide change*	*Amino acid change*	*PolyPhen-2*	*RVIS*
COS0755	chr13:g.39450228G>A	*FREM2*	NM_207361.4: c.8351G>A	p.(Arg2784Lys)	0.941	16.77
COS0755	chr17:g.27308427_27308435delGTGGTGGTA	*SEZ6*	NM_178860.2:c.684_692delTACCACCAC	p.(Thr229_Thr231del)	—	93.47
COS1012	chr19:g.54314055G>C	*NLRP12*	NM_001277126.1:c.858C>G	p.(=)	—	24.22
COS1141	chr12:g.89917133G>A	*GALNT4*	NM_001199781.1:c.1185C>T	p.(=)	—	76.41
COS1553	chr2:g.173368886_173368888delAAG	*ITGA6*	NM_001079818.1:c.3186_3188delAGA	p.(Glu1063del)	—	69.66
COS1553	chr7:g.73935594C>T	*GTF2IRD1*	NM_001199207.1:c.1069C>T	p.(Arg357Cys)	1	4.31
COS1677	chr1:g.22192291G>A	*HSPG2*	NM_005529.5:c.4233C>T	p.(=)	—	32.07
COS1677	chr17:g.37374260C>T	*STAC2*	NM_198993.2:c.257G>A	p.(Arg86Lys)	—	8.8
COS0885	chr1:g.6314749G>A	*GPR153*	NM_207370.1:c.217C>T	p.(Arg73Cys)	0.011	—
COS0885	chr17:g.79650824C>A	*ARL16*	NM_001040025.1:c.32G>T	p.(Arg11Leu)	0.023	53.19
COS1801	chr7:g.130139739C>T	*MEST*	NM_177524.1:c.532C>T	p.(Pro178Ser)	0.568	51.4
COS1801	chr3:g.73651603C>T	*PDZRN3*	NM_015009.1:c.820G>A	p.(Asp274Asn)	0.986	8.63
COS1801	chr1:g.237664043C>T	*RYR2*[a]	NM_001035.2:c.2236C>T	p.(Glu746Tyr)	1	0.05
COS1801	chr1:g.237664045A>C	*RYR2*[a]	NM_001035.2:c.2238A>C			
COS1814	chr6:g.43223506G>A	*TTBK1*	NM_032538.1:c.773G>A	p.(Arg258Gln)	1	23.69
COS1814	chr11:g.76751147C>G	*B3GNT6*	NM_138706.3:c.552C>G	p.(Asp184Glu)	0.998	—
COS0451	chr11:g.65429164A>G	*RELA*	NM_021975.3:c.329T>C	p.(Ile110Thr)	0.878	15.12
COS1870	chr20:g.48140626C>T	*PTGIS*	NM_000961.3:c.824G>A	p.(Arg275Gln)	1	29.49
COS0691	chr22:g.32021757C>T	*PISD*	NM_014338.3:c.45G>A	p.(=)	—	34.71
COS0630	chrX:g.153692763G>A	*PLXNA3*	NM_017514.3:c.1847G>A	p.(Arg616Gln)	0.001	0.94
COS2720	chr14:g.104251194C>T	*PPP1R13B*	NM_015316.2:c.215G>A	p.(Arg72Gln)	0.414	2.24

[a]There are two nucleotide base changes in the same codon that count for one missense variant.

ClinVar database at http://www.ncbi.nlm.nih.gov/clinvar/ (Accession: SCV000223956—SCV000223976). The total quality read bases for each individual was calculated based on the sequenced reads obtained from SureSelect Human All Exon V4 kit probes that were overlapping with regions from the Consensus CDS (CCDS). The *de novo* mutation rates were calculated based on the total number of *de novo* variants divided by the total number of callable bases. The observed *de novo* mutation rate in COS probands was 1.93×10^{-8} per base pair (Figure 1), based on the total number of callable bases for the 17 COS probands (518 Mb) and the 20 observed *de novo* variants (including small insertions and deletions). The *de novo* mutation rate in COS was slightly increased from the *de novo* mutation rate observed in recent studies; the binomial *P*-value was 0.25 (95% CI = 1.17×10^{-8}–2.98×10^{-8}; Table 2a). However, we found that the ratio of non-synonymous ($n = 15$) to synonymous ($n = 4$) *de novo* variants in the COS cases was increased, although not significantly when compared with the neutral expectation (ratio of 2.23:1) reported in previous findings[20] (Binomial *P*-value = 0.46; Table 2b).

We also investigated the possible effect of missense variants in all genes, based on the predictions of PolyPhen-2 (Polymorphism Phenotyping-2).[21] We compared the PolyPhen-2 scores of *de novo* variants found in COS cases with three different groups: *de novo* variants found in adult-onset schizophrenia cases previously reported by our group,[13] *de novo* variants found in published controls[22,23] and private inherited variants found in COS cases. Private inherited variants are the rare variants found exclusively in a given family. The comparison showed that the *de novo* variants were on average more severe than private inherited variants (Two tailed *T*-test *P*-value = 0.03; Figure 2a). We also found that published controls had a lower variant severity profile compared to diseased cohorts, although not significantly (Two-tailed *T*-test *P*-value = 0.29); severity being defined by impact on protein function (Figure 2a). The severity profile was similar in the distribution of schizophrenia and COS. Overall, our analysis showed increase in highly disruptive amino acid changes of *de novo* events, in comparison with inherited variants, and thus suggested a negative selection pressure on missense variants.

Furthermore, to explore the pathogenic potential of our *de novo* variants, we used Residual Variant Intolerant Score (RVIS).[24] RVIS is a gene based score and this genome-wide scoring system assesses the functional variation of human genes based on the single-nucleotide variants in EVS. The RVIS percentile gives an indication as to whether a gene is 'tolerant' or not to variants (ie, if it can be mutated without leading to a disease). This score is significantly correlated with genes known to cause Mendelian diseases.[24] We compared the RVIS percentile for genes harboring *de novo* variants in COS cases with the genes mutated in the three different groups mentioned in the comparison of PolyPhen-2 scores. Interestingly, we found that *de novo* variants in COS were more present in intolerant genes than the private inherited variants (Two tailed *T*-test *P*-value = 0.0375), as well as the *de novo* variants in published controls (Two tailed *T*-test *P*-value = 0.1366; Figure 2b). There was a very similar trend in the RVIS percentile distribution of schizophrenia and COS. Overall, we observed that *de novo* variants were more frequently in genes functionally intolerant to variations regardless of case–control status.

A likelihood analysis of PolyPhen-2 scores and RVIS percentile was also performed using EVS variants to assess the pathogenicity of random variants in a simulation (Supplementary Methods). We calculated the mean value for both PolyPhen-2 scores and RVIS percentile using permutations of a subset of randomly selected genes. We found that the mean PolyPhen-2 score of our identified *de novo* variants tend to be significantly higher than the random selection (*P*-value = 0.040). Furthermore, when applying the same method to RVIS percentile, we found that the *de novo* variant carrying genes had a higher global intolerance to variation (*P*-value = 0.010). Thus, this analysis further supported the notion that *de novo* variants in COS were pathogenic and in genes intolerant to variations not merely by chance (for more details see Supplementary Information and Supplementary Figure 2).

Table 2a *De novo* mutation rate comparison between our COS cohort and recent studies

Studies in Autism and Schizophrenia	*Study de novo mutation rate*	*Observed de novo mutation rate in COS*	*Binomial P-value*
Sanders *et al.*[27]	1.58×10^{-8}	1.93×10^{-8}	0.3842
Fromer *et al.*[25]	1.61×10^{-8}	1.93×10^{-8}	0.3901
Neale *et al.*[26]	1.50×10^{-8}	1.93×10^{-8}	0.2515

DISCUSSION

In this study, the coding *de novo* mutation rate in COS is 1.17 per exome, which is consistent with the results published in *de novo* mutation studies of other psychiatric diseases[13,25–29] (Supplementary Table 2). The missense variants identified in our COS subjects tend to be more severe than private inherited variants. This result supports the fact that *de novo* variants are more penetrant when compared with inherited variants, as they are not subjected to natural selection.[30] *De novo* variants are now thought to explain part of the heritability of complex neurodevelopmental disorders such as ASD, schizophrenia and intellectual disability.[13,22,28,31] Here we show that it might be the same for another related neurodevelopmental disorder, COS. As COS is an early-onset disease, we could hypothesize that *de novo* variants, which are enriched for more severe and deleterious effects (according to the results of PolyPhen-2 and RVIS), may have a greater role in disease than in the adult form of schizophrenia. However, the sample size of the current study does not allow us to draw such a conclusion and thus, a new study in a larger cohort is needed.

Among the 20 *de novo* variants identified, 6 were previously implicated in schizophrenia or other neuropsychiatric disorders. One such gene, *GPR153*, with a *de novo* missense variant in one of the subjects, has also been reported to have a *de novo* missense variant in a schizophrenia patient.[22] Interestingly, the same proband (COS885)

Table 2b Non-synonymous to synonymous ratio comparisons between *de novo* and inherited variants in our study

Cohort	*Variant types*	*Nonsynonymous*	*Synonymous*	*NS: S*
Childhood onset schizophrenia probands	Private inherited variants	3504	1664	2.11
	Common inherited variants (>5% MAF)	108 501	67 392	1.61
	De novo variants	16	4	4

Abbreviations: MAF, minor allele frequency; NS, nonsynonymous; S, Synonymous.

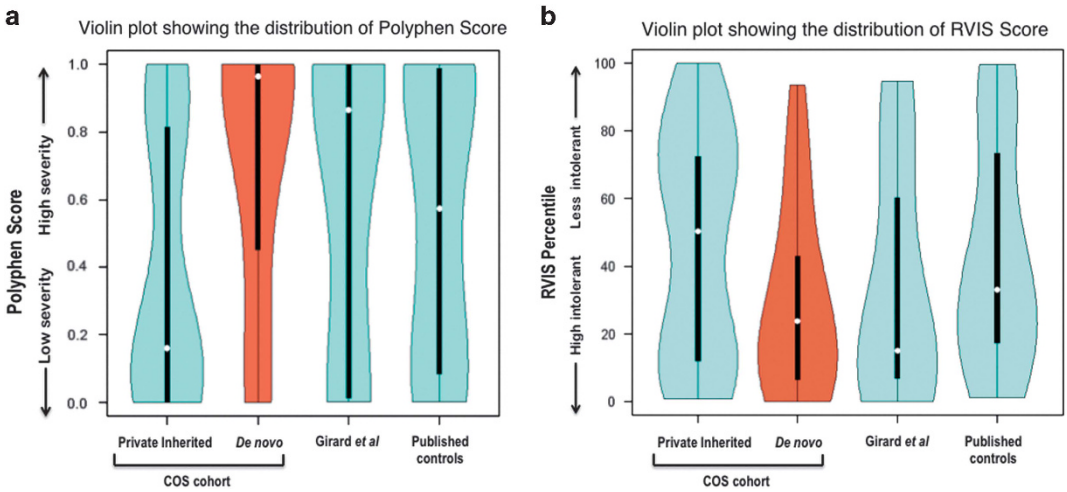

Figure 2 Violin plot showing the distribution of functional severity, predicted by bioinformatics scores, for the missense *de novo* variants and private inherited variants in COS probands. Also shown is the comparison of *de novo* variants in schizophrenia[13] reported by our group and other controls in recent studies.[22,23] The median is indicated by the white dot and the colored area shows the kernel distribution of the data. (**a**) PolyPhen-2 score—variant based and the higher score shows more severity. (**b**) RVIS, Residual Variant Intolerant Score is gene specific and lower score shows more intolerant for variation.

has a possibly disease-related CNV in 1q21.3 (Supplementary Information). If we exclude this patent from our study, then the *de novo* mutation rate is 1.84×10^{-8} per base pair, still similar to that seen in autism and schizophrenia. Similarly, a *de novo* missense variant in *STAC2*, which codes for cysteine-rich domain-containing protein 2, has previously been reported in a patient with schizophrenia.[25] In our study, we identified a synonymous variant in the same exon of this gene, suggesting a possible association with the disease. On the basis of the literature and severity score scale, some of the genes with *de novo* variants seem to be good candidates as COS predisposing genes. One of these is *SEZ6*, a closely related homolog of *SEZ6L2*, a candidate gene in autism.[32] Polymorphisms in *SEZ6L* have been associated with autism and bipolar disorder.[33] In addition, a *de novo* variant in *SEZ6* has been identified in a patient with intellectual disability.[31] As COS patients have comorbidity of autism and language disorders, *SEZ6* is a potential candidate gene for COS. Another gene harboring *de novo* variants, *RYR2*, codes for a calcium channel receptor and has been associated with autism.[34] Interestingly, one of the COS patients has two *de novo* nucleotide substitutions in the same codon of this gene (Supplementary Figure 4). These type of rare events may be termed as 'delins', and were explained in an autism study as *de novo* nucleotide substitutions that occur during allelic gene conversion events.[35] In addition, recurrent *de novo RYR2* and *RYR3* variants have been found in patients with infantile spasms and Lennox–Gastaut syndrome—classical epileptic encephalopathies.[23] Furthermore, RYR2 plays a major role in maintaining the calcium homeostasis and presynaptic function in hippocampal neurons.[36] All these evidences suggest that the variant identified in *RYR2* may affect synaptic organization and connectivity during the early development of the brain. Hence, *RYR2* is a good candidate gene for COS. We also identified a potentially damaging missense variant in *GTF2IRD1*, a gene previously associated with Williams–Beuren Syndrome, a neurodevelopmental disorder.[37] In addition, a deletion of the GTF2 transcription factor family was observed in patients with cognitive behavioral abnormalities, including language delay and non-social anxiety.[37] Similar clinical symptoms are observed in patients with COS (Supplementary Table 1). Since schizophrenia is believed to be a neurodevelopmental disorder, *GTF21RD1* is a potential candidate gene. *TTBK1*, one of the genes harboring a missense *de novo* variant in a COS patient, is brain specific and involved in neuronal and cognitive dysfunction.[38] Moreover, changes in expression of TTBK1 induce pathological effects in the hippocampal neurons of patients with Alzheimer's disease.[39] Therefore, *TTBK1* appears to be an attractive candidate gene for COS because of its expression in the hippocampus, a region whose dysfunction is thought to be involved in psychiatric diseases.[40] Several genes of the neuronal migration pathway have been previously implicated in schizophrenia. Among those genes, we identified a small deletion in *ITGA6* in one of the COS patients. Genes such as *ITGA3* and *LAMA2,* with *de novo* variants reported in schizophrenia, are known to have major roles in fetal development of brain.[29] *ITGA6*, of the same integrin gene family as *ITGA3*, could also be potentially involved in a pathway relating to neuronal migration.[41]

In conclusion, this study provides a list of interesting genes such as *SEZ6, RYR2, GPR153, GTF2ID1, TTBK1* and *ITGA6* that are candidates to be involved in the etiology of childhood onset schizophrenia. Although larger COS cohorts will be required to replicate these findings, it will be worthwhile to further investigate if these candidate genes are also involved in biological pathways associated with the disease, especially the early development of central nervous system.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The current study was supported by the Genome Canada and Genome Quebec (Grant No. RMF_92086). Bioinformatics analysis was supported by the Canadian Institutes of Health Research (CIHR). We thank the patients and the family members who agreed to participate in the study. We thank Daniel Rochefort and Sylvia Dobrezenicka for the technical support; also Edouard Henrion and Ousmane Diallo for their bioinformatics support. We also thank Claire Leblond, Pingxing Xie, Alanna Grant and Martine Therrien for their valuable suggestions and comments.

AUTHOR CONTRIBUTIONS

SLG, RJ, LX and GAR designed the study. JG selected the cases for exome capture and sequencing. JR recruited and diagnosed the cases to provide the

relevant clinical information. AA, AD-L and DS performed alignments and variant detection. AA performed the bioinformatics analysis and variant validation. AA, SLG, AK, CVB, SZ, PAD, JR and GAR wrote the paper.

1 Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK: The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry* 1993; **50**: 85–94.
2 Ramos CC, Rapoport PB, Brito Neto RV: Clinical and tympanometric findings in repeated recreational scuba diving. *Travel Med Infect Dis* 2005; **3**: 19–25.
3 Kirilina EA, Suvorov NI, Popova SS *et al*: Induction of differentiation in leukemic cell strains with myelopeptide-4. *Bull Exp Biol Med* 2005; **140**: 554–557.
4 Weinberger DR: Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry* 1987; **44**: 660–669.
5 Sham PC, MacLean CJ, Kendler KS: A typological model of schizophrenia based on age at onset, sex and familial morbidity. *Acta Psychiatr Scand* 1994; **89**: 135–141.
6 Addington AM, Rapoport JL: The genetics of childhood-onset schizophrenia: when madness strikes the prepubescent. *Curr Psychiatry Rep* 2009; **11**: 156–161.
7 Gochman P, Miller R, Rapoport JL: Childhood-onset schizophrenia: the challenge of diagnosis. *Curr Psychiatry Rep* 2011; **13**: 321–322.
8 Addington AM, Gornick M, Duckworth J *et al*: GAD1 (2q31.1), which encodes glutamic acid decarboxylase (GAD67), is associated with childhood-onset schizophrenia and cortical gray matter volume loss. *Mol Psychiatry* 2005; **10**: 581–588.
9 Gornick MC, Addington AM, Sporn A *et al*: Dysbindin (DTNBP1, 6p22.3) is associated with childhood-onset psychosis and endophenotypes measured by the Premorbid Adjustment Scale (PAS). *J Autism Dev Disord* 2005; **35**: 831–838.
10 Rapoport J, Ahn K: Rare variants and risk for schizophrenia: more support. *Biol Psychiatry* 2011; **70**: 1102–1103.
11 Ahn K, Gotay N, Andersen TM *et al*: High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol Psychiatry* 2014; **19**: 568–572.
12 Awadalla P, Gauthier J, Myers RA *et al*: Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet* 2010; **87**: 316–324.
13 Girard SL, Gauthier J, Noreau A *et al*: Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011; **43**: 860–863.
14 Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; **26**: 589–595.
15 Li H, Handsaker B, Wysoker A *et al*: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
16 McKenna A, Hanna M, Banks E *et al*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
17 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.
18 Ku CS, Vasiliou V, Cooper DN: A new era in the discovery of de novo mutations underlying human genetic disease. *Hum Genomics* 2012; **6**: 27.
19 Untergrasser ACI, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG: Primer3- new capabilities and interfaces. *Nucleic Acids Res* 2012; **40**: e115.
20 Kryukov GV, Pennacchio LA, Sunyaev SR: Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 2007; **80**: 727–739.
21 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
22 Xu B, Roos JL, Dexheimer P *et al*: Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011; **43**: 864–868.
23 Rauch A, Wieczorek D, Graf E *et al*: Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**: 1674–1682.
24 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**: e1003709.
25 Fromer M, Pocklington AJ, Kavanagh DH *et al*: De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
26 Neale BM, Kou Y, Liu L *et al*: Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
27 Sanders SJ, Murtha MT, Gupta AR *et al*: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
28 O'Roak BJ, Deriziotis P, Lee C *et al*: Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011; **43**: 585–589.
29 Gulsuner S, Walsh T, Watts AC *et al*: Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013; **154**: 518–529.
30 Bamshad M, Wooding SP: Signatures of natural selection in the human genome. *Nat Rev Genetics* 2003; **4**: 99–111.
31 Gilissen C, Hehir-Kwa JY, Thung DT *et al*: Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**: 344–347.
32 Konyukh M, Delorme R, Chaste P *et al*: Variations of the candidate SEZ6L2 gene on Chromosome 16p11.2 in patients with autism spectrum disorders and in human populations. *PLoS one* 2011; **6**: e17289.
33 Xu C, Mullersman JE, Wang L *et al*: Polymorphisms in seizure 6-like gene are associated with bipolar disorder I: evidence of gene x gender interaction. *J Affect Disord* 2013; **145**: 95–99.
34 Lu AT, Cantor RM: Allowing for sex differences increases power in a GWAS of multiplex Autism families. *Mol Psychiatry* 2012; **17**: 215–222.
35 Michaelson JJ, Shi Y, Gujral M *et al*: Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 2012; **151**: 1431–1442.
36 Wu B, Yamaguchi H, Lai FA, Shen J: Presenilins regulate calcium homeostasis and presynaptic function via ryanodine receptors in hippocampal neurons. *Proc Natl Acad Sci USA* 2013; **110**: 15091–15096.
37 Edelmann L, Prosnitz A, Pardo S *et al*: An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. *J Med Genet* 2007; **44**: 136–143.
38 Sato S, Xu J, Okuyama S *et al*: Spatial learning impairment, enhanced CDK5/p35 activity, and downregulation of NMDA receptor expression in transgenic mice expressing tau-tubulin kinase 1. *J Neurosci* 2008; **28**: 14511–14521.
39 Lund H, Cowburn RF, Gustafsson E *et al*: Tau-tubulin kinase 1 expression, phosphorylation and co-localization with phospho-Ser422 tau in the Alzheimer's disease brain. *Brain Pathol* 2013; **23**: 378–389.
40 Cabungcal JH, Counotte DS, Lewis EM *et al*: Juvenile antioxidant treatment prevents adult deficits in a developmental model of schizophrenia. *Neuron* 2014; **83**: 1073–1084.
41 Anton ES, Kreidberg JA, Rakic P: Distinct functions of alpha3 and alpha(v) integrin receptors in neuronal migration and laminar organization of the cerebral cortex. *Neuron* 1999; **22**: 277–289.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)