OPEN

Molecular Psychiatry (2016) 21, 885–893

www.nature.com/mp

IMMEDIATE COMMUNICATION

Exome sequencing for bipolar disorder points to roles of *de novo* loss-of-function and protein-altering mutations

M Kataoka[1,2,6], N Matoba[1,3,6], T Sawada[1], A-A Kazuno[1], M Ishiwata[1], K Fujii[1,4], K Matsuo[5], A Takata[1] and T Kato[1]

Although numerous genetic studies have been conducted for bipolar disorder (BD), its genetic architecture remains elusive. Here we perform, to the best of our knowledge, the first trio-based exome sequencing study for BD to investigate potential roles of *de novo* mutations in the disease etiology. We identified 71 *de novo* point mutations and one *de novo* copy-number mutation in 79 BD probands. Among the genes hit by *de novo* loss-of-function (LOF; nonsense, splice site or frameshift) or protein-altering (LOF, missense and inframe indel) mutations, we found significant enrichment of genes highly intolerant (first percentile of intolerant genes assessed by Residual Variation Intolerance Score) to protein-altering variants in general population, an observation that is also reported in autism and schizophrenia. When we performed a joint analysis using the data of schizoaffective disorder in published studies, we found global enrichment of *de novo* LOF and protein-altering mutations in the combined group of bipolar I and schizoaffective disorders. Considering relationship between *de novo* mutations and clinical phenotypes, we observed significantly earlier disease onset among the BD probands with *de novo* protein-altering mutations when compared with non-carriers. Gene ontology enrichment analysis of genes hit by *de novo* protein-altering mutations in bipolar I and schizoaffective disorders did not identify any significant enrichment. These results of exploratory analyses collectively point to the roles of *de novo* LOF and protein-altering mutations in the etiology of bipolar disorder and warrant further large-scale studies.

Molecular Psychiatry (2016) **21,** 885–893; doi:10.1038/mp.2016.69; published online 24 May 2016

INTRODUCTION

Bipolar disorder (BD) is a severe mental disorder characterized by recurrent manic and depressive episodes with a lifetime prevalence ~1%.[1,2] Epidemiological studies have consistently indicated significant contribution of genetic factors to the etiology of BD, and its heritability (broad-sense heritability that considers both the additive and non-additive genetic factors) calculated from the concordance rates in monozygotic and dizygotic twins is ~85%.[3] Because of this high heritability a number of genetic studies for BD have been conducted. However, linkage studies using family samples could not robustly identify causative genes.[4] More recently genome-wide association studies (GWAS) have identified single-nucleotide polymorphisms certainly associated with BD in several genes such as *CACNA1C* encoding a plasma membrane L-type Ca^{2+} channel,[5,6] whereas the effect size of each robustly associated single-nucleotide polymorphism is tiny (typically only increases the risk by 1.1–1.2 fold) and a large part of the disease heritability cannot be explained even if hundreds of thousands of weakly associated common single-nucleotide polymorphisms are considered.[7] This phenomenon is termed as the 'missing heritability', and potential roles of rare variants that have not been investigated in GWAS are drawing increasing attention.

As one of the proofs for the hypothesis that rare variants could have a significant role in the genetic architecture of BD, recent whole-genome sequencing of 200 individuals from 41 bipolar families showed the role of rare transmitted variants in neuronal excitability related genes including those encoding calcium channels.[8] In addition to rare transmitted variants, *de novo* (newly arising) mutations could be another class of genetic variation that explains a part of the missing heritability (note that *de novo* mutations explain a part of 'broad-sense' heritability while these are not inherited[9]). Indeed, recent studies have implicated contribution of *de novo* copy-number variations (CNVs) to the risk for BD,[10,11] particularly patients with an early onset.[10] In addition to CNVs, *de novo* point mutations, whose frequency increases as paternal age advances,[12] could have an important role in BD, because it has been well known that an older age of the father increases risk for BD in the offspring.[13,14] However, to our knowledge there has been no study reporting results of comprehensive analysis of *de novo* point mutations in BD, whereas their significant contribution to the genetic architectures of various neuropsychiatric disorders including autism spectrum disorder (ASD)[15–21] and schizophrenia[22–28] has been consistently reported in recent whole-exome sequencing (WES) studies that analyze all coding exons in the human genome.

To address the question whether *de novo* point mutations contribute to the genetic architecture of BD, here we performed, to the best of our knowledge, the first trio-based WES study for BD by analyzing 237 exomes.

[1]Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Saitama, Japan; [2]Department of Child Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; [3]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; [4]Department of Psychiatry, Dokkyo Medical University School of Medicine, Tochigi, Japan and [5]Division of Neuropsychiatry, Department of Neuroscience, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan. Correspondence: Dr A Takata or Dr T Kato, Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, 2-1 Hirosawa Wako, Saitama 351-0198, Japan.
E-mail: atakata@brain.riken.jp or kato@brain.riken.jp
[6]These authors contributed equally to this work and are listed in alphabetical order.
Received 9 September 2015; revised 16 February 2016; accepted 14 March 2016; published online 24 May 2016

MATERIALS AND METHODS

Studied subjects

We used DNA samples from 79 trios with a BD proband (56 with bipolar I disorder (BDI) and 23 with bipolar II disorder (BDII)) and unaffected parents. All the probands were diagnosed with BDI or BDII based on the DSM (Diagnostic and Statistical Manual of Mental Disorders) IV criteria by trained psychiatrists. All the parents were screened for mental disorders by structured interview using M.I.N.I. (Mini International Neuropsychiatric Interview).[29] See Supplementary Information for more detailed information. The study was approved by the First Committee of Research Ethics of RIKEN Wako Institute and the Institutional Review Board of Yamaguchi University Hospital.

Exome sequencing, data processing, variant calling and identification of *de novo* mutations

Genomic DNA from either peripheral blood or saliva was subjected to target capturing using the SureSelect^{XT} Human All Exon kits V4, V5 or V5 + mitochondria (Agilent Technologies, Santa Clara, CA; detailed information is available in Supplementary Table 1). WES was performed by using either the HiSeq2000 or HiSeq2500 (Illumina, San Diego, CA, USA) with paired-end 101 bp reads. Generated sequence data (fastq files) were processed by using the pipeline with BWA-MEM[30] (version 0.7.5a), SAMtools,[31] Picard (version 1.92, http://picard.sourceforge.net/) and GATK[32] (version 2.6–4). Variant calls were made by using the GATK best practices recommendations.[33] Identified candidates for *de novo* mutations were subjected to validation experiments by amplification using PCR followed by standard Sanger sequencing. *De novo* CNVs were identified by using exome hidden Markov model (XHMM)[34–36] and copy-number inference from exome reads (CoNIFER)[37] and validated by comparative genome hybridization arrays. See Supplementary Information for more detailed information.

Analyses of genes hit by *de novo* mutations using RVIS

We assigned Residual Variation Intolerance Score (RVIS) representing 'gene intolerance' to protein-altering genetic variation in general population to each gene hit by a *de novo* mutation using Dataset 2 of Petrovski *et al.*[38] Among the 70 genes with *de novo* mutations in our BD cohort, 66 genes were assigned for RVIS and used for the subsequent analyses. In Petrovski *et al.*,[38] the hypothesized probability that a *de novo* mutation hit the first quartile of the most intolerant genes was calculated as 38% considering gene sizes. This was because,

$$\frac{\text{Total coding length of the first quartile of intolerant genes}}{\text{Total coding length of all genes}} = 0.38.$$

We obtained the data of coding length for each gene assigned for RVIS using UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables) and calculated that

$$\frac{\text{Total coding length of the first percentile of intolerant genes}}{\text{Total coding length of all genes}} = 0.04.$$

Therefore if a mutation is randomly generated, the probabilities to hit the first percentile and the first quartile of the intolerant genes are 4% and 38%, respectively. On the basis of these hypothesized probabilities (4% for the first percentile and 38% for the first quartile), we evaluated whether the first percentile and the quartile of 'intolerant genes' are enriched among the genes hit by three types of mutations (that is, loss-of-function (LOF), protein-altering and synonymous) in our cohort using one-tailed binominal exact test. For example, there are 53 genes with a *de novo* protein-altering mutation in BD for which RVIS is available. If 53 mutations are randomly generated, the expected number of first percentile of intolerant genes hit by a protein-altering mutations is $53 \times 0.04 = 2.1$, whereas we observed that six genes were among the first percentile of the intolerant genes in the real data set for BD. On the basis of these observations, we performed one-tailed binominal exact test with the following numbers; the number of success (x) = 6, the number of trials (n) = 53 and the hypothesized probability of success (p) = 0.04. The corresponding commend for R software we used for the analysis was *binom.test (x = 6, n = 53, p = 0.04, alternative = "greater")*.

Enrichment analysis of LOF and protein-altering *de novo* mutations in case subjects

Global enrichment of *de novo* LOF and protein-altering mutations in case subjects were examined by one-tailed Fisher's exact test using the reported data of 1911 unaffected siblings.[39] For an extended analysis, published data on schizoaffective disorder (SAD)[24,27] was combined with the current data set of BD. Details are described in Supplementary Information.

Procedures for gene ontology (GO) enrichment analysis of *de novo* LOF and protein-altering mutations and other analyses are detailed in Supplementary Information.

RESULTS

Identification of *de novo* point mutations in bipolar disorder

We performed WES of 237 DNA samples from a cohort consisting of 79 probands affected with BD and their parents without major psychoses (that is, BD, schizophrenia and SAD, Supplementary Table 1). On average 92.5% of the targeted exome regions were covered by 20 or more reads at the individual level. At the trio level, on average 88.9% of the targets were covered by 20 or more reads in all three members (Supplementary Figure 1). Among the 79 trios, we identified 71 *de novo* point mutations (single-nucleotide variations (SNVs) and short insertion/deletions (indels)) in 70 genes, which were validated by Sanger sequencing (Table 1 and Supplementary Table 2). As two missense mutations in *DOCK10* were identified in the same individual on the same sequencing reads with an interval of four bases, we considered these mutations as a single missense mutation event (thus there are 70 *de novo* events). These 70 *de novo* events comprise of 64 SNVs (including the composite mutation noted above) and six indels, including four nonsense mutations, one canonical splice site mutation, 45 missense mutations, 14 synonymous mutations, four frameshift indels (of which two mutations directly introduce a stop codon) and two inframe indels. The number of *de novo* mutations in each proband ranged from zero to four. Forty-two probands carried one or more *de novo* mutations. Per-individual number of *de novo* mutations was 0.89, which is similar to those reported in the largest family-based WES studies for ASD[20] (0.94 for ASD and 0.84 for unaffected siblings, based on re-annotated data using our analytical pipelines) and schizophrenia[26] (0.90 for schizophrenia) to date.

Identification of *de novo* CNV from exome sequencing data

We next analyzed CNVs using our WES data. For this purpose we used two software, XHMM[34–36] and CoNIFER,[37] both of which were specifically developed to detect CNVs from WES data sets. We identified a *de novo* deletion of approximately 0.2 Mbp at 3q29 including *ATP13A3*, *TMEM44*, *LSG1* and *FAM43A* (approximated position in hg19 = chr3: 194.2 - 194.4M), which was confirmed by comparative genome hybridization arrays (Supplementary Figure 2). This *de novo* deletion is located at ~1.3 Mbp upstream of the known 3q29 deletion syndrome locus, whose nominal association with BD was recently reported.[40] In addition to direct disruption of the genes included in the CNV region, this deletion may have some regulatory impact on genes in the 3q29 deletion syndrome locus.

Table 1. List of 71 *de novo* point mutations in 79 trios with BD probands

ID	*BD type*	*Chromosome*	*Start position*	*Reference*	*Alternative*	*Mutation type*	*Gene symbol*	*Amino-acid change*[a]
103	I	17	43192727	G	A	Missense	*PLCD3*	p.A515V
105	I	4	3251006	C	T	Synonymous	*MSANTD1*	p.G19G
106	I	2	128176049	T	G	Missense	*PROC*	p.S9R
107	I	1	171673653	C	T	Missense	*VAMP4*	p.R140H
107	I	X	21995301	A	G	Missense	*SMS*	p.K151R
110	I	1	39735167	—	C	Frameshift insertion	*MACF1*	p.V266fs
111	I	11	6653775	C	T	Missense	*DCHS1*	p.D990N
115	I	3	197497100	G	A	Missense	*FYTTD1*	p.S161N
116	I	11	64621996	C	—	Frameshift deletion[b]	*EHD1*	p.V472X
116	I	16	18820961	T	C	Missense	*SMG1*	p.Y3640C
116	I	18	65179594	C	T	Missense	*DSEL*	p.G761D
117	I	12	53565177	C	T	Missense	*CSAD*	p.R194H
118	I	9	104341502	C	T	Synonymous	*GRIN3A*	p.L969L
118	I	14	23862939	T	A	Missense	*MYH6*	p.D955V
119	I	11	124793786	T	C	Missense	*HEPACAM*	p.D183G
120	I	11	397415	—	CCA	Inframe insertion	*PKP3*	p.S305delinsSH
120	I	12	71978307	C	T	Synonymous	*LGR5*	p.Y839Y
120	I	X	73811786	G	A	Missense	*RLIM*	p.S455F
121	I	10	27459678	C	T	Missense	*MASTL*	p.S597F
122	I	8	139164319	T	A	Missense	*FAM135B*	p.D800V
122	I	9	35403597	G	A	Splice site	*UNC13B*	—
122	I	11	6291989	A	G	Missense	*CCKBR*	p.D256G
122	I	22	21989506	G	C	Missense	*CCDC116*	p.G385A
123	II	18	58039428	A	G	Missense	*MC4R*	p.V52A
127	II	4	187177191	C	T	Missense	*KLKB1*	p.T512I
130	I	15	51828860	A	G	Missense	*DMXL2*	p.L606S
132	I	1	63284902	C	A	Synonymous	*ATG4C*	p.S207S
134	I	2	197707468	A	C	Synonymous	*PGAP1*	p.T869T
134	I	16	847034	C	T	Missense	*CHTF18*	p.P1101L
135	I	4	68688179	C	T	Missense	*TMPRSS11D*	p.R378Q
135	I	12	121691252	G	A	Missense	*CAMKK2*	p.R311C
205	I	1	110280787	G	A	Nonsense	*GSTM3*	p.R100X
205	I	7	42017240	C	T	Missense	*GLI3*	p.V577I
209	II	2	225662602	A	G	Missense	*DOCK10*	p.S1531P
209	II	2	225662607	T	C	Missense	*DOCK10*	p.N1529S
212	II	14	105957989	G	A	Missense	*C14orf80*	p.V21M
213	II	6	7182273	C	G	Synonymous	*RREB1*	p.P43P
213	II	11	75591045	G	A	Nonsense	*UVRAG*	p.W131X
214	I	17	56387403	—	TCC	Inframe insertion	*BZRAP1*	p.E1272delinsEE
215	I	2	55544865	C	T	Missense	*CCDC88A*	p.R1146Q
215	I	7	151859861	T	A	Nonsense	*KMT2C*	p.K3601X
215	I	10	73494079	C	G	Missense	*CDH23*	p.P1401R
215	I	14	74567890	G	C	Missense	*LIN52*	p.E82Q
217	II	13	21436873	A	T	Nonsense	*XPO4*	p.Y100X
219	I	1	33558984	C	T	Missense	*AZIN2*	p.A185V
219	I	9	110249695	C	—	Frameshift deletion	*KLF4*	p.S327fs
220	I	2	223536557	C	A	Missense	*MOGAT1*	p.T18K
221	I	17	61889463	A	T	Missense	*DDX42*	p.N524Y
223	II	7	150935311	C	T	Synonymous	*CHPF2*	p.V621V
301	II	5	16704730	G	A	Missense	*MYO10*	p.A756V
303	II	9	100444622	A	C	Missense	*XPA*	p.F255C
311	II	1	11896118	C	T	Missense	*CLCN6*	p.H630Y
311	II	11	47354513	G	A	Synonymous	*MYBPC3*	p.T1114T
311	II	11	57970920	T	G	Missense	*OR1S2*	p.Q245P
311	II	22	21336724	T	C	Missense	*LZTR1*	p.S22P
312	II	6	143085951	C	T	Missense	*HIVEP2*	p.A1835T
312	II	19	54966705	C	T	Synonymous	*LENG8*	p.G328G
314	I	6	90410529	T	C	Missense	*MDN1*	p.K2825R
316	II	15	44092967	T	C	Synonymous	*HYPK*	p.S103S
316	II	17	11738078	A	G	Missense	*DNAH9*	p.M3124V
317	I	4	56726636	G	C	Missense	*EXOC1*	p.D62H
317	I	10	5924978	A	G	Synonymous	*ANKRD16*	p.Y280Y
319	I	15	67528983	G	A	Synonymous	*AAGAB*	p.Y83Y
319	I	22	39627761	G	A	Missense	*PDGFB*	p.R108W
321	II	1	165324728	C	G	Synonymous	*LMX1A*	p.S23S
321	II	1	209799134	C	T	Missense	*LAMB3*	p.G612E
321	II	6	151669922	C	T	Synonymous	*AKAP12*	p.H132H
321	II	19	16625364	C	G	Missense	*C19orf44*	p.L598V
324	I	6	64401764	A	C	Missense	*PHF3*	p.K776T
325	I	14	53513547	A	—	Frameshift deletion[b]	*DDHD1*	p.L881X
325	I	20	39794889	C	T	Missense	*PLCG1*	p.P619S

Abbreviation: BD, bipolar disorder. [a]Annotations for the longest transcript. [b]Frameshift deletions directly introducing a stop codon.

De novo LOF and protein-altering mutations in BD preferentially hit intolerant genes

We next analyzed properties of genes hit by *de novo* mutations in BD in the context of 'gene intolerance'. Recently Petrovski *et al.*[38] developed RVIS, a scoring system to assess intolerance of individual genes to protein-altering variants based on large-scale WES data of general population. By using RVIS it has been demonstrated that genes hit by *de novo* LOF (nonsense, splice site and frameshift; all of these are expected to totally disrupt the protein sequence) and protein-altering (that is, LOF, missense and inframe indel) mutations in ASD and schizophrenia are enriched for the first quartile (25 percentile) of intolerant genes,[28,38] and rare transmitted LOF variants in the first percentile of highly intolerant genes are enriched in ASD probands when compared with their healthy siblings.[41]

Among 70 genes hit by *de novo* mutations in BD, 66 genes were assigned for RVIS. When we analyzed gene intolerance by classifying mutations according to their predicted functionality (that is, LOF, protein-altering or synonymous), we found that genes with *de novo* LOF mutations or protein-altering mutations are significantly enriched for the first percentile of intolerant genes (Figures 1a, $P = 4.5 \times 10^{-3}$ for LOF and $P = 0.019$ for any protein-altering mutations, one-tailed binominal exact test, see Materials and Methods for details), whereas there was no enrichment in genes with synonymous mutations ($P = 1$). A similar trend was observed when we analyzed for the first quartile of intolerant genes ($P = 0.079$ for LOF, $P = 0.067$ for protein-altering and $P = 0.79$ for synonymous mutations). These findings equivalent to those for ASD and schizophrenia, for which roles of *de novo* mutations are established, suggest contribution of *de novo* mutations, particularly *de novo* LOF and protein-altering mutations in intolerant genes, to the genetic etiology of BD.

According to these results and previous WES studies for ASD and schizophrenia mostly reporting global enrichment of *de novo* LOF and protein-altering mutations in case subjects,[20–22,24,25,27,28,42] we also tested whether there is global excess of these mutations in BD. For this purpose, we compared the numbers of *de novo* LOF or protein-altering mutations and *de novo* synonymous mutations between our BD probands and a large cohort of control subjects (1911 unaffected siblings in Iossifov *et al.*[20]), because this method using synonymous mutations as an internal control should be resistant to potential artifacts caused by comparison of data from different studies. There was no statistically significant enrichment of *de novo* LOF and protein-altering mutations in BD (odds ratio (OR) = 1.48, $P = 0.244$, for LOF mutations, OR = 1.30, $P = 0.233$ for protein-altering mutations, one-tailed Fisher's exact test, Figure 1b).

Considering relationship between *de novo* mutations and clinical phenotypes in our BD cohort, we observed that there is significant difference in age of onset between the probands carrying one or more *de novo* protein-altering mutations and the probands with no protein-altering mutation (Figure 1c, two-tailed Student's *t*-test, $P = 0.013$, average age of onset ± s.d. = 21.6 ± 6.1 in mutation carriers and 25.9 ± 8.8 in non-carriers), while there was no significant difference in age of ascertainment between these two groups.

Global enrichment of *de novo* LOF and protein-altering mutations in BDI and SAD

As a moderate sample size in our cohort limit the statistical power, we next performed a joint analysis by combining our data set and the published data for patients with SAD[43] characterized by both the mood and psychotic symptoms and shares genetic background with BD[44] (no. of trios = 143; 79 BD from our cohort, 63 SAD from Xu *et al.*,[24] and one SAD from MaCarthy *et al.*[27]). In the analysis comparing this combined group of BD and SAD to controls (1911 unaffected siblings in Iossifov *et al.*[20] as described above), we observed a trend toward enrichment of *de novo* LOF and protein-altering mutations ($P = 0.085$, OR = 1.73 for LOF mutations, $P = 0.073$, OR = 1.47 for protein-altering mutations, Figure 1b). In addition, when we focused on the severer group, patients with BDI or SAD, there was statistically significant enrichment of *de novo* LOF and protein-altering mutations in the case group ($P = 0.030$, OR = 2.30 for LOF mutations, $P = 0.021$, OR = 1.87 for protein-altering mutations, Figure 1b), whereas further large-scale studies should be required to conclude enrichment of these mutations.

GO enrichment analysis of the genes hit by *de novo* protein-altering mutations

On the basis of the results of our analyses showing global enrichment of *de novo* protein-altering mutations in BDI and SAD, we next exploratory investigated whether there are specific GO terms overrepresented among the genes hit by these mutations in the combined group of BDI and SAD.

We first performed a GO enrichment analysis with the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.7)[45,46] using the list of genes hit by *de novo* protein-altering mutations in BDI and SAD as an input (no. of genes = 75, Supplementary Table 3). There was no GO term with significant enrichment after performing correction for multiple testing. Non-significant trend of enrichment was found for nine GO terms including 'calcium ion binding (GO:0005509)', 'serine-type peptidase activity (GO:0008236)' and 'tissue morphogenesis (GO:0048729)' (Figure 2a). To test whether the suggestive enrichment of these terms are noteworthy, we performed a simulation analysis by randomly selecting 75 *de novo* protein-altering mutations, equal to the number of mutations in BDI and SAD, from the list of *de novo* protein-altering mutations reported in control subjects[20] 10 000 times (Figure 2b, see Supplementary Information for details). We counted how many times nominally significant enrichment of a given GO term was observed, and the probability to see significant enrichment among 10 000 trials was considered as the *P*-value. If the enrichment observed in our DAVID analysis is explained by artifacts due to use of genes harboring *de novo* mutations as an input (for example, the input genes should be biased toward large genes because such genes have higher chance to be hit by *de novo* mutations), significant enrichment should not be observed in this simulation analysis. Three GO terms showed non-significant results in the simulation analysis. We observed nominally significant enrichment (uncorrected *P*-value < 0.05 in the simulation analysis) of six GO terms; 'calcium ion binding (GO:0005509)', 'tissue morphogenesis (GO:0048729)', 'serine-type peptidase activity (GO:0008236)', 'serine hydrolase activity (GO:0017171)', 'embryonic morphogenesis (GO:0048598)' and 'protein homodimerization activity (GO:0042803)'. However, this enrichment was no more significant after the correction for multiple testing with the number of terms subjected to the simulation analysis, except for 'serine-type peptidase activity (GO:0008236)' and 'serine hydrolase activity (GO:0017171)'. Individual genes with a *de novo* protein-altering mutation included in each term are detailed in Figure 2a.

New candidate genes for bipolar disorder

The list of genes harboring *de novo* mutations in BD could help identification of promising new candidate genes for BD.

According to the 'ascertainment differentials', the differences in the frequencies of each class of mutation in two populations[20,47] calculated from per-individual rates of *de novo* LOF or protein-altering mutations in our BD cohort and control subjects (1911 unaffected siblings in Iossifov *et al.*,[20] see Supplementary Information for details), roughly 22% and 9% of *de novo* LOF and protein-altering mutations in BD could contribute to the diagnosis of BD, respectively. This indicates that genes with

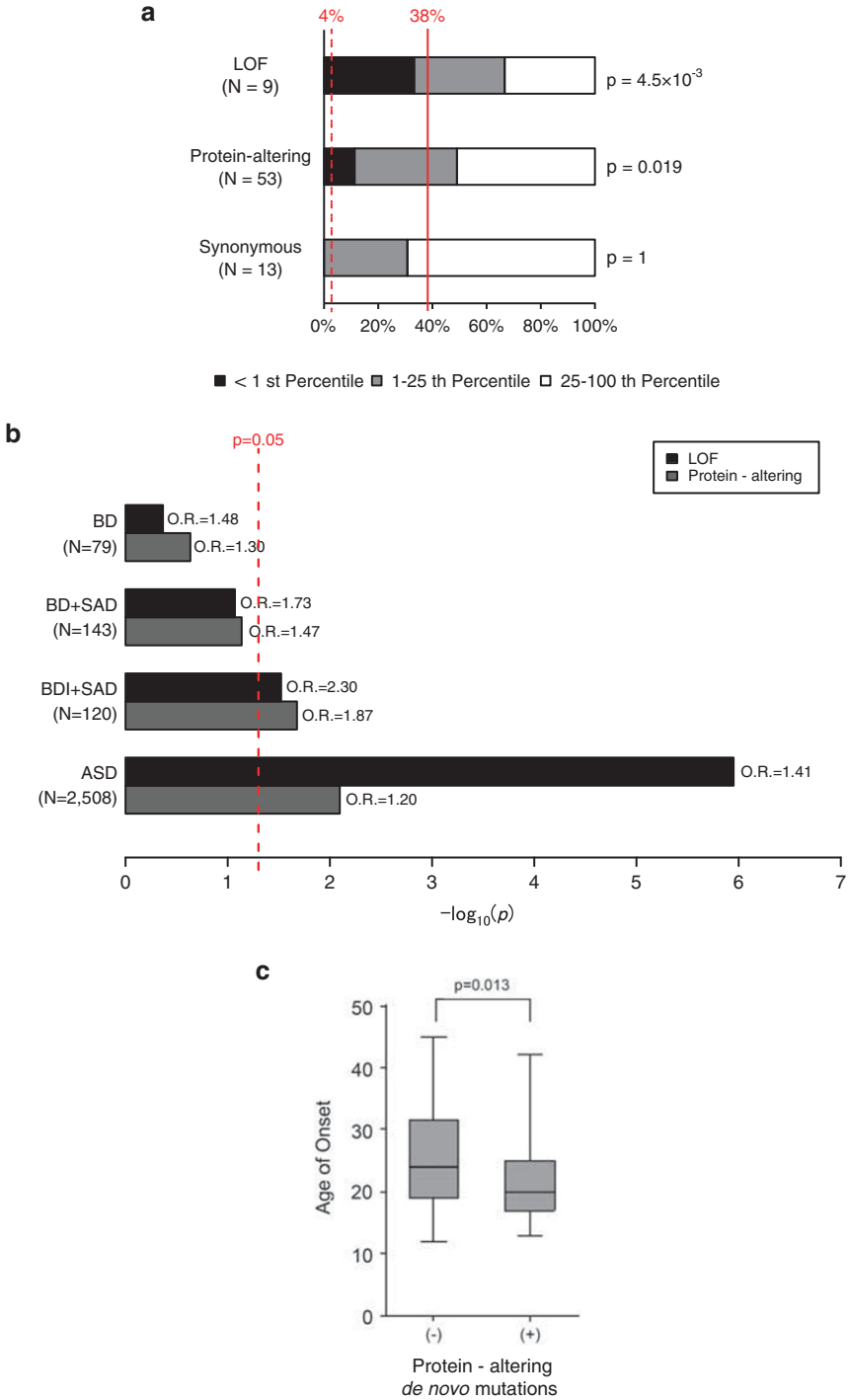

Figure 1. Roles of *de novo* loss-of-function (LOF) and protein-altering mutations in bipolar disorder. (**a**) Proportion of the genes hit by different types of *de novo* mutations according to their gene intolerance. Gene intolerance to protein-altering variants in general population was assessed by using Residual Variation Intolerance Score (RVIS).[38] Black, < 1st percentile of intolerant genes; gray, 1–25th percentiles of intolerant genes; white, rest of the genes (25–100th percentiles). Dashed and solid red lines indicate expected proportion for the first percentile and the first quartile of intolerant genes considering gene sizes (4 and 38%, see Materials and Methods for detailed procedures). Enrichment *P*-values for the first percentile of intolerant genes calculated by one-tailed binominal tests are shown on the right side of the bars. (**b**) Enrichment analyses of *de novo* LOF and protein-altering mutations in case groups. Bars indicate statistical significance ($\log_{10}$ *P*-values) for enrichment of *de novo* LOF (black) and protein-altering (gray) mutations in each disease group. Enrichment was evaluated by comparing the numbers of *de novo* LOF or protein-altering mutations and synonymous mutations between each disease group, and controls (1911 unaffected siblings in Iossifov *et al.*[20]) with one-tailed Fisher's exact test. Data for autism spectrum disorder (ASD) were shown as a reference.[20] Red dashed line indicates $P = 0.05$. (**c**) Box plots of age of onset for bipolar disorder (BD) probands with or without protein-altering *de novo* mutations. Average ages of onset between the two groups were compared by two-tailed Student's *t*-test. Box plots of median values with hinges at the 25th and 75th percentiles and whiskers extending to the highest and lowest values are shown. BDI, bipolar I disorder; OR, odds ratio; SAD, schizoaffective disorder.

a

GO Terms		Genes	Count	P_DAVID	P_Simulation (raw/corrected)
GO:0005509	calcium ion binding	*CAMKK2, CDH23, DCHS1,* ***EHD1****,* ***MACF1****, PLCD3, PLCG1, PROC, (CANX, CAMK4, MBTPS1)*	11	0.0033	0.0082/0.0738
GO:0008236	serine-type peptidase activity	*PROC, TMPRSS11D, (KLK13, MBTPS1, RBP3)*	5	0.0054	0.0013/0.0117
GO:0017171	serine hydrolase activity	*PROC, TMPRSS11D, (KLK13, MBTPS1, RBP3)*	5	0.0056	0.0013/0.0117
GO:0048729	tissue morphogenesis	*GLI3,* ***KLF4****,* ***MACF1****, MYH6, (HECTD1)*	5	0.0057	0.0101/0.0909
GO:0048598	embryonic morphogenesis	*CDH23, GLI3,* ***KLF4****,* ***MACF1****, (HECTD1)*	5	0.033	0.006/0.054
GO:0042803	protein homodimerization activity	*CCDC88A, MYH6, PDGFB, (DPYD, PML)*	5	0.049	0.0077/0.0693

b

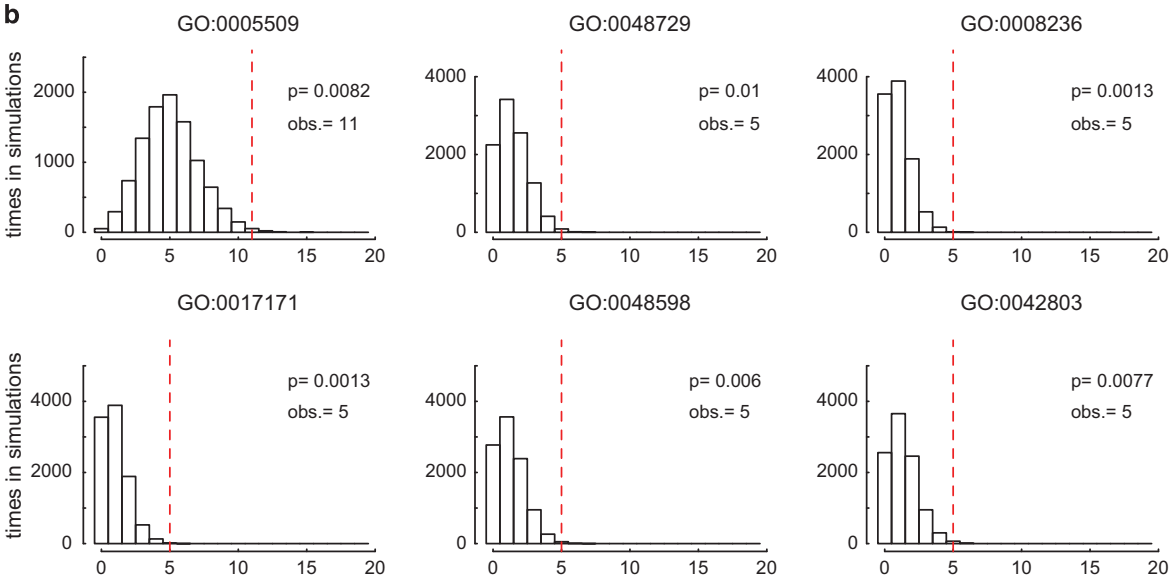

Figure 2. Gene ontology enrichment analysis of the genes hit by *de novo* protein-altering mutations. (**a**) Six gene ontology (GO) terms nominally enriched among the genes hit by *de novo* protein-altering mutations in the combined group of bipolar I disorder (BDI) and schizoaffective disorder (SAD). P_DAVID indicates *P*-values calculated by DAVID (The Database for Annotation, Visualization and Integrated Discovery)[45,46] (uncorrected raw P-values). P_Simulation indicates *P*-values calculated by a simulation analysis using the data of *de novo* protein-altering mutations in control subjects (1911 unaffected siblings in Iossifov *et al.*[20]). For P_Simulation, both the raw *P*-values and *P*-values corrected for the number of terms subjected to the simulation analysis (# = 9, Bonferroni procedure) were noted. Boldface indicates genes with a *de novo* loss-of-function (LOF) mutation in bipolar disorder (BD). Genes with *de novo* mutations identified in SAD[24,27] are shown in parentheses. (**b**) Histograms represent the distribution of hit counts in the simulation analyses (10 000 iterations) for seven GO terms. Dotted lines indicate the observed hit counts (obs.) and the corresponding *P*-values.

de novo LOF mutations should be particularly enriched for genuine disease susceptibility genes. Among nine genes with a *de novo* LOF mutation, we found enrichment of genes highly intolerant to functional variation as described above. These genes hit by a LOF mutation despite their high intolerance, *EHD1*, *KLF4*, *KMT2C*, *MACF1*, *UNC13B* and *XPO4* (Table 2), should be good candidates for disease susceptibility genes.

Previous studies have pointed out genes hit by multiple *de novo* protein-altering mutations, particularly LOF mutations, are highly likely to be genuine disease genes.[48–50] Although there was no gene with two or more *de novo* protein-altering mutations in our BD cohort or the combined group of BD and SAD, when we compared the list of genes hit by *de novo* protein-altering mutations in BD with the list for schizophrenia (excluding known cases of SAD) we found six genes hit by *de novo* protein-altering mutations in BD and also in schizophrenia (*BZRAP1*, *DNAH9*, *GLI3*, *KMT2C*, *LCT* and *MACF1*, Table 2). Although the P-values for observed numbers of *de novo* protein-altering mutations in these genes (calculated by the procedures described in Samocha *et al.*,[49] see Supplementary Information for details) do not reach to the exome-wide significance threshold ($P = 2.5 \times 10^{-6}$, considering the number of coding genes, Table 2), some of them could be good candidates for genes associated with BD and schizophrenia that share genetic risk factors.[51,52]

DISCUSSION

In this study reporting results of the first trio-based WES for BD, we identified 71 *de novo* point mutations and one *de novo* CNV in 79 probands. By exploring the properties of *de novo* mutations and genes hit by these mutations in BD, we observed significant enrichment of *de novo* LOF mutations hitting genes highly intolerant to functional variants as well as a trend toward global excess of *de novo* LOF and protein-altering mutations. In the joint analysis combining our data of BD and the data of SAD in published studies, we observed global enrichment of LOF and protein-altering mutations in the severer group of patients consisting of BDI and SAD, implicating contribution of these mutations to the genetic etiology. Our analysis of relationship between clinical phenotypes and *de novo* protein-altering mutations revealed significantly earlier age at onset in probands with one or more such mutations than non-carriers. This observation is in line with a previous study reporting stronger association of *de novo* CNVs with early-onset BD.[10] Although further large-scale studies are required to prove the pathogenic role of *de novo* mutations in BD, given the moderate sample size and statistical significance in this study, our observations would be credible considering the fact that similar results have been reported for ASD and schizophrenia. *De novo* SNVs reportedly increase with advanced paternal age.[12]

Table 2. New candidate genes for BD from whole-exome sequencing

Intolerant genes hit by de novo *LOF mutations*

Gene symbol	*BD type*	*Mutation type*	*RVIS*	*RVIS percentile*	
MACF1	I	Frameshift	−3.92	0.21	Most intolerant
UNC13B	I	Splice site	−2.89	0.59	
KMT2C	I	Nonsense	−2.52	0.91	
EHD1	I	Nonsense	−1.38	4.39	
XPO4	II	Nonsense	−0.78	12.97	
KLF4	I	Frameshift	−0.45	24.19	Less intolerant

Genes hit by de novo *protein-altering mutations in BD and schizophrenia*

Gene symbol	*BD type*	*Mutation type in BD*	*Mutation type in schizophrenia*	*Number of observed mutations*	*Number of expected mutations*[a]	P-*value*[b]
BZRAP1	I	Inframe	Missense	2	0.130	0.0074
GLI3	I	Missense	Missense	2	0.130	0.0075
DNAH9	II	Missense	Frameshift	2	0.295	0.034
KMT2C	I	Nonsense	Missense	2	0.299	0.035
MACF1	I	Frameshift	Missense	2	0.355	0.048

Abbreviations: BD, bipolar disorder, LOF, loss-of-function; RVIS, Residual Variation Intolerance Score. [a]Number of expected mutations were calculated by using per-gene mutation rates provided in Samocha *et al.*[49] [b]*P*-values calculated by invoking Poisson distribution probabilities. Lower RVIS and RVIS percentile indicate that the gene is less tolerant to protein-altering variants.

In a disease with constant prevalence rate despite reduced reproduction fitness, genetic risk factors are assumed to be constantly supplied as *de novo* mutations. Because the risk for BD is associated with advanced paternal age[13,14] and BD patients show reduced reproduction fitness,[53,54] albeit less so than in schizophrenia and ASD,[53] it is plausible that de novo SNVs have a role in etiology of BD.

In our GO enrichment analyses of genes hit by *de novo* protein-altering mutations in BDI and SAD, we identified nominal enrichment of six GO terms. Among them, enrichment of 'serine-type peptidase activity (GO:0008236)' and 'serine hydrolase activity (GO:0017171)' remained significant when we considered the number of terms subjected to our simulation analysis, suggesting potential involvement of this pathway in the pathophysiology of BD. In addition, identification of 'calcium ion binding (GO:0005509)' as one of nominally enriched terms should be of interest, because this result is in line with findings in previous genetic, biochemical and pharmacological studies for BD. For instance, association of several voltage-dependent calcium channel genes such as *CACNA1C and CACNA1B* was reported in large-scale GWAS for BD.[5,55] A pathway analysis of GWAS data suggests enrichment of calcium channel-related pathways among the genes empirically associated with BD.[56] Studies of peripheral blood cells have consistently demonstrated altered intracellular calcium signaling in BD.[57,58] Lithium, the first-line therapeutic drug for BD, modulates inositol-mediated pathways[59] and thereby regulate calcium ion release from the endoplasmic reticulum.[60] When we performed a GO enrichment analysis using DAVID[45,46] by integrating the data of *de novo* protein-altering mutations in our study, common single-nucleotide polymorphisms associated with BD in a large-scale GWAS[5] and rare CNVs implicated in BD[40] (total no. of unique input genes = 229, see Supplementary Information for details), 'calcium signaling pathway (hsa04020)' was the only term significantly enriched after performing correction for multiple testing ($P_{corrected} = 6.4 \times 10^{-3}$, Bonferroni correction; note that significant enrichment of 'calcium signaling pathway (hsa04020)' after correction was not observed when we submitted candidate genes from each study). This result of an integrative analysis indicates that various types of genetic evidence for BD could converge on the calcium signaling pathway.

When we looked at individual genes carrying a *de novo* mutation to search for promising candidate genes, we found that *KMT2C*, *MACF1* and *UNC13B* are hit by a *de novo* LOF mutation despite their extremely high intolerance to protein-altering variants[38] (Table 1). *KMT2C* (also known as *MLL3*) encodes a catalytic subunit of histone methyltransferase protein complex specifically mediating mono-, di- and tri-methylation of histone H3 at lysine 4 (H3K4). Identification of a *de novo* LOF mutation in this gene that is also hit by *de novo* protein-altering mutations in ASD and schizophrenia[17,21,26] could be in line with accumulating evidence pointing out important roles of chromatin regulator genes in neuropsychiatric disorders.[21,27,28,61,62] *MACF1* (also known as *ACF7*) encodes the microtubule actin cross-linking factor 1 protein that has an essential role in integration of microtubule dynamics.[63] This gene categorized as a 'calcium ion binding (GO:0005509)' gene is involved in calcium-induced reorganization of the cytoskeleton.[64] Interestingly, binding of MACF1 to microtubules is regulated by GSK3β,[65] a key enzyme implicated in the mechanism of action of lithium.[66] *UNC13B* (also known as *MUNC13*) encodes a presynaptic protein with an essential role in synaptic vesicle priming. Although this gene is not classified as a 'calcium ion binding (GO:0005509)' gene, UNC13B forms a complex with calmodulin, a calcium ion binding protein, and this complex regulates synaptic vesicle priming and synaptic efficacy in response to residual calcium ion signals,[67] further suggesting involvement of the calcium signaling pathway in the BD pathophysiology. Besides these three genes, *EHD1* could be another promising candidate gene in the context of possible relationship between BD and the calcium signaling pathway. This gene carrying a *de novo* LOF mutation in BD encodes a calcium binding protein involved in regulation of synaptic endocytosis and exocytosis.[68,69] In addition, our preliminary experiments indicate that the *de novo* frameshift mutation in *EHD1* at the last exon of this gene cause expression of the protein lacking EF-hand calcium binding domain by escaping nonsense-mediated mRNA decay (Supplementary Figure 3). This truncated form of EHD1 protein may have a dominant negative effect. These four genes could be

particularly good candidates for disease susceptibility genes for BD, and it would be worthwhile to subject these genes to target resequencing.[70]

In summary, we performed the first trio-based WES study for BD and demonstrated potential roles of *de novo* protein-altering mutations and calcium-related genes in the disease etiology. These findings are in accordance with the results of previous WES studies for other neuropsychiatric disorders with reduced fecundity such as ASD and schizophrenia, and with the evidence from various types of studies for BD. Our results could provide important insights into the genetic architecture and biology of BD, and warrant further large-scale studies in order to understand the roles of *de novo* and rare mutations in BD more precisely, and to identify robust disease-associated genes/mutations with a large effect size.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the patients and their families for their participation. We thank Dr Jared C Roach at the Institute for Systems Biology and Dr Takaoki Kasahara at RIKEN BSI for valuable discussion and insightful comments. We also thank Ms Atsuko Komori at RIKEN BSI for technical assistance. We are grateful to Mr Kenichiro Harada and Ms Masako Hirotsu at Yamaguchi University Hospital for assistance of collecting data. We thank the Research Resources Center at the RIKEN BSI, especially Mr Keisuke Fukumoto and Ms Fujiko Sakai for exome library preparation, sequencing and array CGH analysis. We are grateful to RIKEN GeNAS for sequencing. This work was supported by the Prime Minister's special fund for 'conquering depression and dementia through advanced circuit research' for RIKEN BSI and in part by the RIKEN Junior Research Associate Program. TK received a grant from Takeda Pharmaceutical Co., Ltd outside of this work. KM received funding for an unrelated study from Otsuka Pharmaceutical Co., Ltd..

AUTHOR CONTRIBUTIONS

MK collected human samples, assessed clinical phenotypes of participants, assisted exome library preparation, performed Sanger validation, analyzed relationship between clinical phenotypes and *de novo* mutations, and contributed to the analyses for Figure 1 and Table 1. NM developed the analytical pipelines for sequencing data processing, identified candidates for *de novo* mutations, analyzed properties of *de novo* mutations and genes hit by these mutations, and contributed to the analyses for Figures 1 and 2, Tables 1 and 2. KF, KM and TK performed clinical assessment and collected human samples. AK assisted Sanger validation and exome library preparation. MI assisted DNA extraction. TS performed analysis of EHD1 protein. AT supervised bioinformatics analyses. TK conceived and, AT and TK co-organized the project. NM and AT prepared the initial draft and NM, MK, AT and TK wrote the final manuscript. All authors read and approved the manuscript.

REFERENCES

1 Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RMA, Petukhova M *et al.* Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 2007; **64**: 543–552.
2 Kawakami N. Large scale epidemiology study of the prevalence of mental disorders: World Mental Health Japan Survey Second. *A report of the Health Labour Sciences Research Grant from The Ministry of Health Labour and Welfare* (H25-Seishin-Ippan-006), Tokyo, 2014.
3 McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry* 2003; **60**: 497–502.
4 Kato T. Molecular genetics of bipolar disorder and depression. *Psychiatry Clin Neurosci* 2007; **61**: 3–19.
5 Sklar P, Smoller JW, Fan J, Ferreira MAR, Perlis RH, Chambert K *et al.* Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008; **13**: 558–569.
6 Mühleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J *et al.* Genome-wide association study reveals two new risk loci for bipolar disorder. *Nature Commun* 2014; **5**: 3339.
7 Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**: 294–305.
8 Ament SA, Szelinger S, Glusman G, Ashworth J, Hou L, Akula N *et al.* Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci USA* 2015; **112**: 3576–3581.
9 Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 2014; **46**: 881–885.
10 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**: 951–963.
11 Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N *et al.* De novo CNVs in bipolar affective disorder and schizophrenia. *Hum Mol Genet* 2014; **23**: 6677–6683.
12 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**: 471–475.
13 Chudal R, Gissler M, Sucksdorff D, Lehti V, Suominen A, Hinkka-Yli-Salomäki S *et al.* Parental age and the risk of bipolar disorders. *Bipolar Disord* 2014; **16**: 624–632.
14 Frans EM, Sandin S, Reichenberg A, Lichtenstein P, Långström N, Hultman CM. Advancing paternal age and bipolar disorder. *Arch Gen Psychiatry* 2008; **65**: 1034–1040.
15 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**: 242–245.
16 O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011; **43**: 585–589.
17 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**: 246–250.
18 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**: 237–241.
19 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**: 285–299.
20 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
21 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
22 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**: 860–863.
23 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**: 864–868.
24 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; **44**: 1365–1369.
25 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**: 518–529.
26 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
27 McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol Psychiatry* 2014; **19**: 652–658.
28 Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron* 2014; **82**: 773–780.
29 Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; **59**(Suppl 2): 22–33.
30 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.

31 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
32 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
33 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**: 491–498.
34 Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM *et al.* Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* 2012; **91**: 597–607.
35 Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwara Y *et al.* Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am J Hum Genet* 2013; **93**: 607–619.
36 Fromer M, Purcell SM. Using XHMM software to detect copy number variation in whole-exome sequencing data. *Curr Protoc Hum Genet* 2014; **81**: 23.1–23.21.
37 Krumm N, Sudmant P, Ko A. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012; **22**: 1525–1532.
38 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.
39 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
40 Green EK, Rees E, Walters JTR, Smith K-G, Forty L, Grozeva D *et al.* Copy number variation in bipolar disorder. *Mol Psychiatry* 2016; **21**: 89–93.
41 Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K *et al.* Excess of rare, inherited truncating mutations in autism. *Nat Genet* 2015; **47**: 582–588.
42 Howrigan D, Neale B, Samocha K, Moran J, Chambert K, Rose SA *et al.* The emerging landscape of schizophrenia risk conferred by *de novo* coding mutations. *World Congress of Psychiatric Genetics*, Toronto, 16-20 Oct 2015.
43 Nardi AE, Nascimento I, Freire RC, de-Melo-Neto VL, Valença AM, Dib M *et al.* Demographic and clinical features of schizoaffective (schizobipolar) disorder--a 5-year retrospective study. Support for a bipolar spectrum disorder. *J Affect Disord* 2005; **89**: 201–206.
44 Cardno AG, Owen MJ. Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. *Schizophr Bull* 2014; **40**: 504–515.
45 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
46 Huang DW, Sherman BT. Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1–13.
47 Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH *et al.* Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci USA* 2015; **112**: E5600–E5607.
48 Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 2013; **155**: 997–1007.
49 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014; **46**: 944–950.
50 Hoischen A, Krumm N, Eichler EE. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* 2014; **17**: 764–772.
51 Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 2009; **373**: 234–239.
52 International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
53 Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* 2013; **70**: 22–30.
54 Mansour H, Kandil K, Wood J, Fathi W, Elassy M, Ibrahim I *et al.* Reduced fertility and fecundity among patients with bipolar I disorder and schizophrenia in Egypt. *Psychiatry Investig* 2011; **8**: 214–220.
55 Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E *et al.* Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry* 2009; **14**: 252–260.
56 Nurnberger JI, Koller DL, Jung J, Edenberg HJ, Foroud T, Guella I *et al.* Identification of pathways for bipolar disorder: a meta-analysis. *JAMA Psychiatry* 2014; **71**: 657–664.
57 Warsh JJ, Andreopoulos S, Li PP. Role of intracellular calcium signaling in the pathophysiology and pharmacotherapy of bipolar disorder: current status. *Clin Neurosci Res* 2004; **4**: 201–213.
58 Kato T. Role of mitochondrial DNA in calcium signaling abnormality in bipolar disorder. *Cell Calcium* 2008; **44**: 92–102.
59 Berridge MJ, Downes CP, Hanley MR. Neural and developmental actions of lithium: a unifying hypothesis. *Cell* 1989; **59**: 411–419.
60 Mikoshiba K. Inositol 1,4,5-trisphosphate receptor. *Trends Pharmacol Sci* 1993; **14**: 86–89.
61 Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Mol Psychiatry* 2013; **18**: 1054–1056.
62 Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet* 2013; **14**: 347–359.
63 Kodama A, Karakesisoglou I, Wong E, Vaezi A, Fuchs E. ACF7: an essential integrator of microtubule dynamics. *Cell* 2003; **115**: 343–354.
64 Karakesisoglou I, Yang Y, Fuchs E. An epidermal plakin that integrates actin and microtubule networks at cellular junctions. *J Cell Biol* 2000; **149**: 195–208.
65 Wu X, Shen Q-T, Oristian DS, Lu CP, Zheng Q, Wang H-W *et al.* Skin stem cells orchestrate directional migration by regulating microtubule-ACF7 connections through GSK3β. *Cell* 2011; **144**: 341–352.
66 Jope RS. Lithium and GSK-3: one inhibitor, two inhibitory actions, multiple outcomes. *Trends Pharmacol Sci* 2003; **24**: 441–443.
67 Junge HJ, Rhee J-S, Jahn O, Varoqueaux F, Spiess J, Waxham MN *et al.* Calmodulin and Munc13 form a Ca^{2+} sensor/effector complex that controls short-term synaptic plasticity. *Cell* 2004; **118**: 389–401.
68 Wei S, Xu Y, Shi H, Wong S-H, Han W, Talbot K *et al.* EHD1 is a synaptic protein that modulates exocytosis through binding to snapin. *Mol Cell Neurosci* 2010; **45**: 418–429.
69 Yap CCC, Lasiecka ZM, Caplan S, Winckler B. Alterations of EHD1/EHD4 protein levels interfere with L1/NgCAM endocytosis in neurons and disrupt axonal targeting. *J Neurosci* 2010; **30**: 6646–6657.
70 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**: 1619–1622.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)