


ARTICLE

Open Access

Contribution of common and rare damaging variants in familial forms of bipolar disorder and phenotypic outcome

Elisa Courtois^{1,2,3}, Mark Schmid^{1,2,3}, Orly Wajsbrot^{3,4}, Caroline Barau⁵, Philippe Le Corvoisier⁶, Bruno Aouizerate^{3,7}, Frank Bellivier^{3,8,9,10}, Raoul Belzeaux^{3,11}, Caroline Dubertret^{3,12}, Jean-Pierre Kahn^{3,4}, Marion Leboyer^{1,2,3,13}, Emilie Olie^{3,14}, Christine Passerieux^{3,15}, Mircea Polosan^{3,16}, Bruno Etain^{3,8,9,10} and Stéphane Jamain^{1,2,3} , and the FondaMental Advanced Centers of Expertise in Bipolar Disorders (FACE-BD)

Abstract

Genome-wide association studies on bipolar disorders (BD) have revealed an additive polygenic contribution of common single-nucleotide polymorphisms (SNPs). However, these SNPs explain only 25% of the overall genetic variance and suggest a role of rare variants in BD vulnerability. Here, we combined high-throughput genotyping data and whole-exome sequencing in cohorts of individuals with BD as well as in multiplex families with a high density of affected individuals in order to determine the contribution of both common and rare variants to BD genetic vulnerability. Using polygenic risk scores (PRS), we showed a strong contribution of common polymorphisms previously associated with BD and schizophrenia (SZ) and noticed that those specifically associated with SZ contributed more in familial forms of BD than in non-familial ones. The analysis of rare damaging variants shared by affected individuals in multiplex families with BD revealed a single interaction network enriched in neuronal and developmental biological pathways, as well as in the regulation of gene expression. We identified four genes with a higher mutation rate in individuals with BD than in the general population and showed that mutations in two of them were associated with specific clinical manifestations. In addition, we showed a significant negative correlation between PRS and the number of rare damaging variants specifically in unaffected individuals of multiplex families. Altogether, our results suggest that common and rare genetic variants both contribute to the familial aggregation of BD and this genetic architecture may explain the heterogeneity of clinical manifestations in multiplex families.

Introduction

Bipolar disorders (BD) are chronic, heterogeneous, and complex mental disorders with a worldwide prevalence of ~1%¹. Their etiopathology is poorly understood but genetic contribution arises as an important factor, with an estimated heritability ranging between 60% and 80%^{2,3}. Recent genome-wide association studies (GWAS) have demonstrated the contribution of common variants in the

increased vulnerability of developing BD⁴. Associated variants had a low effect size, but when combined in an additive model, they were able to significantly distinguish individuals with BD from controls⁵. However, these overall results explain only 25% of the genetic variance⁶, revealing a “missing heritability” and increasing the potential interest for rare penetrant variants.

The rapid development of next-generation sequencing led to the possibility to unravel the potential contribution of rare variants in BD. In other psychiatric disorders, e.g. autism spectrum disorder (ASD) and schizophrenia (SZ), whole-exome sequencing (WES) has already shown a higher frequency of de novo mutations in affected

Correspondence: Stéphane Jamain (stephane.jamain@inserm.fr)

¹INSERM U955, Psychiatrie Translationnelle, Créteil 94000, France

²Université Paris Est, Faculté de Médecine, Créteil 94000, France

Full list of author information is available at the end of the article

A list of members and their affiliations are listed at the end of the paper.

© The Author(s) 2020



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

individuals compared to control populations^{7,8}. However, few studies have been conducted yet in subjects with BD with a limited consensus on rare penetrant mutation burden and their relationship with low penetrant common variants^{9–15}.

Here, we combined high-throughput genotyping data and WES in cohorts of individuals with BD as well as in multiplex families with a high density of affected individuals in order to determine the relative contribution of common and rare variants in clinical manifestations of BD. We first calculated polygenic risk scores (PRS) for BD and SZ in a cohort of 445 patients with BD and 1636 individuals from the general population, and determine their impact on the family history of BD. We then used these scores in multiplex families and compared affected and unaffected individuals. Using WES, we identified rare damaging variants in constrained genes that were shared by affected subjects and highlighted impacted biological pathways. Association with BD was confirmed by comparing mutation rates of associated genes in an independent cohort of 241 subjects with BD and 21,071 non-Finnish European individuals from the non-psychiatric population of the Exome Aggregation Consortium (ExAC)¹⁶. Finally, we combined PRS and rare variants identified in families and showed different genetic architectures between affected and unaffected individuals in multiplex families.

Materials and methods

Multiplex families

A written informed consent was obtained from all subjects prior to participation in the study and research ethic board of the Pitié-Salpêtrière Hospital (CPP Ile de France VI) approved protocols and procedures.

We recruited eight multiplex families (Supplementary Fig. S1) of French ancestry with at least two individuals meeting the BD criteria of the fourth edition of the diagnostic and statistical manual for psychiatric disorders (DSM-IV)¹⁷. In total, we collected DNA for 38 individuals: 21 with BD, 1 with schizo-affective disorder (SAD), 1 with ASD, 1 with unipolar depression, 1 with major depressive episode, and 13 healthy individuals. Taking into account the high shared heritability between BD and SZ and to avoid too much heterogeneity, we considered, as “affected” subjects in multiplex families, those with a diagnosis of either a BD or a SAD. All other phenotypes and unaffected subjects were pooled under the name “unaffected”.

Cohorts

Two-hundred-and-eighty patients with BD were recruited in nine FondaMental Academic Centers of Expertise for Bipolar Disorder (FACE-BD) located in France¹⁸. This network is coordinated by the French

scientific cooperation foundation, Fondation FondaMental (www.fondation-fondamental.org). All patients met the DSM-IV diagnostic criteria for BD. In addition, a previously published cohort^{19,20} of 445 subjects with BD and 1636 control individuals of French origin has been used for genotyping analyses (Supplementary Fig. S2).

DNA extraction and genotyping

Genomic DNA was isolated from venous blood sample or from saliva. DNA extraction was performed as previously described²¹ for blood samples or using prepIT® L2P kit (DNA genotek, Kanata, ON, Canada), following the manufacturer's instructions. In multiplex families, 27 DNAs were genotyped for ~960,000 single-nucleotide polymorphisms (SNPs) using the OmniExpressExome-8v1-4_A1 beadchips (Illumina Inc., San Diego, CA, USA) and 11 DNAs were genotyped for ~660,000 SNPs, using the Infinium Global Screening Arrays-24 v2.0 (Illumina Inc.), according to the manufacturer's protocol (Supplementary Fig. S2). Quality control of genotypic data, family relationships, and ancestry were performed using the PLINK (v.1.9)²², the KING (v.2.1.3)²³, and the Eigensoft (v.7.2.1)²⁴ softwares, respectively. Briefly, samples with sex discordance, a heterozygosity rate different of the mean_{net} ± 0.2, a genotyping rate lower than 0.98 or without a European ancestry were excluded of analyses. SNPs with a genotyping rate lower than 0.95 were removed. Remaining SNPs were subsequently used for imputation in 37 samples using the Sanger imputation server. Prephasing was conducted using SHAPIT2 (v2.r790)²⁵ and imputation using PBWT²⁶ with 1000 Genomes phase 3 as reference panels²⁷. After imputation, only biallelic SNPs with a minor allele frequency (MAF) higher than 0.01 in the reference panel and an imputation score higher than 0.9 were used for further studies.

The previously published cohort has been genotyped using HumanHap550 bead arrays (Illumina Inc.) for individuals with BD and HumanHap300 bead arrays (Illumina Inc.) for controls and imputed by the Psychiatric Genomics Consortium (PGC) based on HapMap 2 CEU sample¹⁹.

Polygenic risk score analysis

PRS were calculated for each individual using the PRSice software (v.2.0.15)²⁸ and were based on the PGC summary statistics for BD⁴ or SZ²⁹ with only European individuals. The previously published cohort was included in the PGC BD summary statistics. Therefore, summary statistics excluding the French cohort have been generously provided by Dr. Eli Stahl, who conducted the last PGC BD analyses. Only genotyped or imputed independent SNPs ($r^2 < 0.2$) both in multiplex families and in the previously published cohort with a MAF higher than 0.01 and an imputation score higher than 0.9 were used for

PRS calculation. We estimated the best PRS threshold to discriminate patients from controls in our cohort by testing all p values threshold between 0.00001 and 0.5 with an incremental factor of 0.00001. We then selected the one showing the lowest p value and the highest R^2 . We obtained a p value threshold of 0.3319 for the BD-associated SNPs-based PRS (BD-PRS) and of 0.2025 for SZ-associated SNPs-based PRS (SZ-PRS) (Supplementary Figs. S3 and S4, respectively). In addition, at the threshold previously determined, we isolated SNPs that only contributed to BD-PRS (BD_{only}-PRS), i.e. only present in BD summary statistics with INFO > 0.9 and p value < 0.3319, from those that only contributed to SZ-PRS (SZ_{only}-PRS), i.e. only present in SZ summary statistic with INFO > 0.9 and p value < 0.2025, and from those that contributed both to BD and SZ, i.e. present in BD summary statistics with INFO > 0.9 and p value < 0.3319 and SZ summary statistics with INFO > 0.9 and p value < 0.2025. For the latter ones, we calculated a new PRS (BD–SZ-PRS) based on the average effect, i.e. the mean of the odds ratio (OR), observed in the SZ and the BD summary statistics. The correlation between PRS and the number of rare damaging variants in multiplex families was calculated using the sum of the scores (PRS_{sum}) of BD_{only}-PRS, SZ_{only}-PRS, and the BD–SZ-PRS for each individual.

Exome sequencing

Exons were captured using the SureSelect^{XT} Human All Exon V5, the SureSelect^{QXT} Human All Exon V6 or the SureSelect^{QXT} Human All Exon V6+ UTR libraries (Agilent Technologies, Santa Clara, CA, U.S.A.). WES was conducted on NextSeq500 with paired-end 75 bp reads, HiSeq2000 with paired-end 75 bp reads, or HiSeq2500 with paired-end 125 bp reads. Data processing consisted in using an home-made workflow based on Trimmomatic (v.0.36)³⁰, BWA-MEM (v.0.7.12)³¹, Picard (v.2.8.1), and GATK (v.3.7)³² softwares. A bed intersect file between the three different captures generated with BedOps (v.2.4.28)³³ and the human reference genome (hg19: GRCh37) were used. Variant calling was performed following the GATK best practices recommendations and annotation was made with SnpEff (v.4.3)³⁴ and bcftools (v1.6). Only variants passing quality filter of VariantRecalibration tool of GATK, with a depth of read higher than 10, a genotype quality score higher than 20, and no missingness in family were kept. Same quality control of individuals as for beadchips was applied using the PLINK, KING, and Eigensoft softwares except a genotyping rate lower than 0.94. After quality control, 34 samples were used in subsequent analysis. In order to identify relevant variants, we selected variants absent from the ExAC¹⁶ or the 1000 Genomes²⁷ databases or with a MAF lower than 0.01 in these two databases. We then selected damaging variants, i.e. missense variants, indels, frameshift, loss or

gain of stop or start codon, and those altering splice sites, with a phred CADD score higher than 15 (v1.4)³⁵, corresponding to the 5% most deleterious variants in the genome. We also selected variants in constrained genes, i.e. with a z -score higher than 3 for missense variants and inframe indels, or with a pLI > 0.9 for loss-of-function variants¹⁶. All the mutations identified have been checked by Sanger sequencing.

For the cohort of 280 subjects with BD, the same quality control was performed. In addition, related individuals and variants with a missingness higher than 0.03 were excluded. After quality control, 241 were used in subsequent analyses (Supplementary Fig. S2).

For general population, we downloaded vcf files (ftp://ftp.broadinstitute.org/pub/ExAC_release/release1/subsets/ExAC_nonpsych.r1.sites.vep.vcf.gz) corresponding to non-psychiatric individuals of the ExAC population¹⁶ and selected only non-Finnish European individuals. We then extracted damaging variants (missense or loss-of-function) with an MAF < 1% and which passed filter of VariantRecalibration tool of GATK in mutated genes identified in our multiplex families. Only variants covered at 95% of the maximum count of allele were considered and only genes with more than 70% of their variants meeting this criterion were considered.

Network and pathway analyses

Network and pathway analyses were performed using the R version of the STRING (v.10) database³⁶. Network was defined using a minimum interaction score of 0.4 and with active interaction sources as text-mining, co-expression, experiments, and database. Pathway analyses were based on the Gene Ontology (GO) resource^{37,38} and a custom background gene list of 18,016 genes. Only GO term including between 5 and 1800 genes and pathways with at least 10% of mutated genes were considered for enrichment analyses.

Gene length bias was controlled by randomly selecting 10,000 times the number of genes identified in multiplex families ($N = 34$) with a cumulative exon length equal to the mean size of mutated genes ($\pm 20\%$). GO terms were removed when they were found more than 500 times after 10,000 permutations.

Statistical analyses

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Statistical analyses were conducted with R (v.3.4.1). All variables were expressed as median and interquartile range and between-group comparisons were analyzed by Wilcoxon rank-sum tests. Group comparison proportions were analyzed using the Fisher's exact test or Chi-squared test according to sample size. Benjamini–Hochberg

correction was applied for multiple testing when comparing mutation rates. Differences were considered as significant when the false discovery rate (FDR) was lower than 0.05. Bonferroni correction was applied for PRS comparisons, multivariate PRS analyses, clinical data analyses, and linear regression between PRS_{sum} and the number of rare mutations. In all cases, p values were adjusted (p_{adj}) according to the number of tests, i.e. 12, 3, 5, and 2, respectively, and difference were considered as significant when the p_{adj} was lower than 0.05.

Results

SZ-specifically associated SNPs contribute only to familial forms of BD

Common polymorphisms have been demonstrated to explain up to 25% of the genetic variance in BD, more than half of which could be shared with SZ⁶. In order to determine whether the polygenic contribution was similar in familial and non-familial forms of BD, we used imputed genotyping data of 1,709,567 SNPs for 445 patients with BD, 1636 controls, and 37 individuals in multiplex families to calculate two PRS based on PGC BD⁴ and SZ²⁹ summary statistics. Both BD-PRS and SZ-PRS contributed to BD vulnerability for subjects with ($N=144$) and without ($N=218$) a family history of BD when compared with controls (Fig. 1). Although there was no difference in BD-PRS between individuals with and without a family history of BD in our cohort (Wilcoxon rank-sum test, $W=16,212$; $p=0.60$; $p_{\text{adj}}=1.00$; Fig. 1a), a higher SZ-PRS was observed for familial forms when compared to non-familial ones (Wilcoxon rank-sum test, $W=13,225$; $p=0.01$; $p_{\text{adj}}=0.12$; Fig. 1b). However, this difference did not resist to correction for multiple testing. We then split PRS with the aim of distinguishing SNPs that were specifically associated with BD (BD_{only}-PRS) or SZ (SZ_{only}-PRS), and those that contributed to both disorders (BD-SZ-PRS) and performed multivariate analyses (Table 1). While the SZ_{only}-PRS and the BD-SZ-PRS both contributed to the risk of developing BD in familial forms of the disease, only the BD-associated polymorphisms (BD_{only}-PRS and BD-SZ-PRS) contributed to the disease vulnerability in non-familial forms of BD. Consistent with this observation, we showed that SZ specific associated polymorphisms (SZ_{only}-PRS) were able to distinguish familial and non-familial forms of BD (Table 1).

In multiplex families, we observed that affected subjects had a higher BD-PRS than controls (Wilcoxon rank-sum test, $W=10,675$; $p=0.003$, $p_{\text{adj}}=0.04$; Fig. 1a). Although, BD-PRS was slightly higher in unaffected subjects in multiplex families, this difference did not resist to correction for multiple testing (Wilcoxon rank-sum test, $W=9199$; $p=0.04$, $p_{\text{adj}}=0.48$) and no difference was observed between affected and unaffected subjects (Wilcoxon rank-sum test, $W=182$; $p=0.68$, $p_{\text{adj}}=1.00$). In

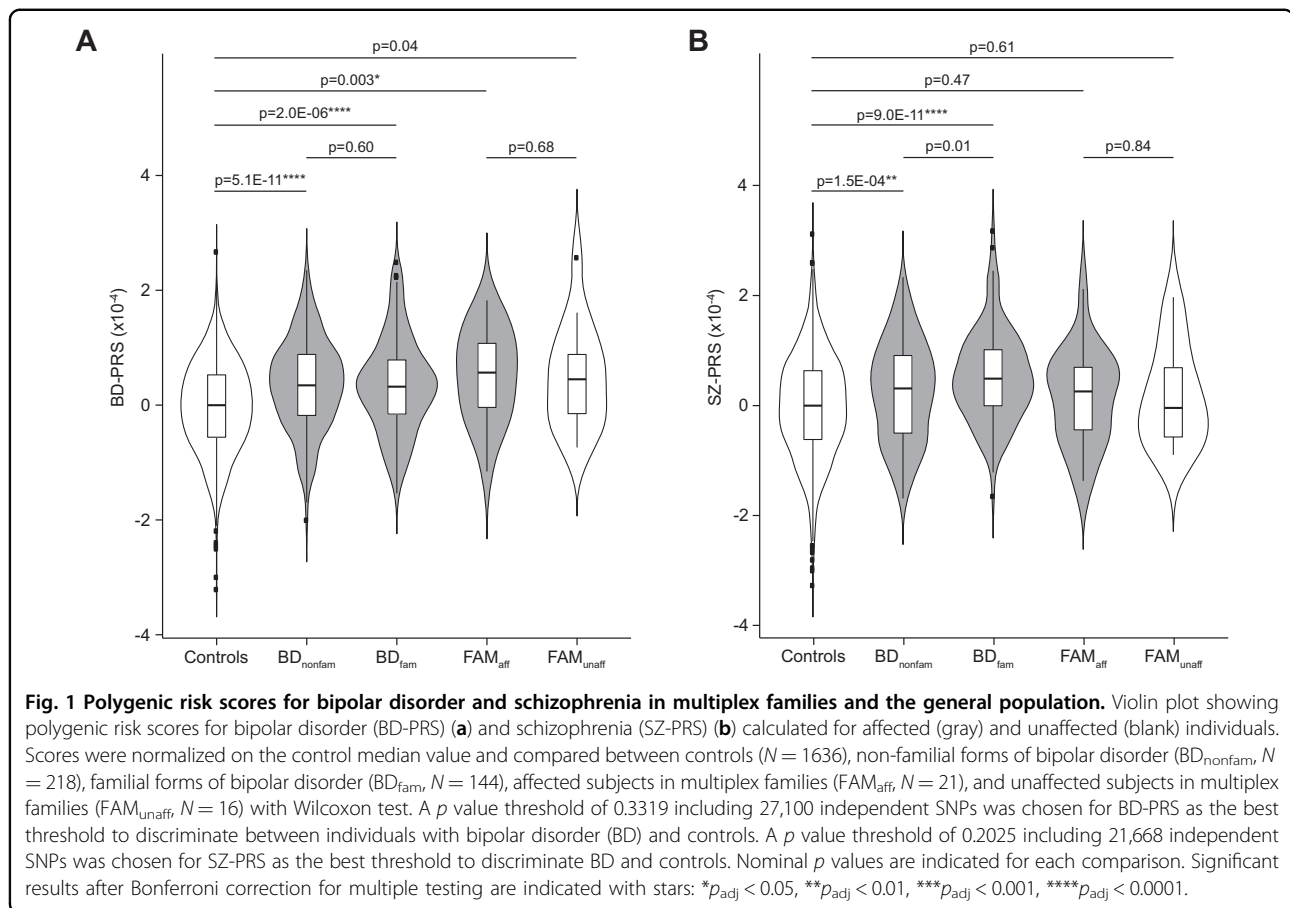
contrast, no higher SZ-PRS was observed neither for affected (Wilcoxon rank-sum test, $W=15,589$, $p=0.47$, $p_{\text{adj}}=1.00$) nor for unaffected (Wilcoxon rank-sum test, $W=12,124$, $p=0.61$, $p_{\text{adj}}=1.00$; Fig. 1b) subjects in multiplex families. These results suggested that common polymorphisms were unable to explain alone the clinical status differences observed in multiplex families of BD.

Rare damaging variants shared by affected subjects in multiplex families affect gene regulation or neuron development and morphogenesis

In order to determine whether rare variants may also contribute to the risk of developing BD in multiplex families, we carried out a WES on 34 subjects from the 8 multiplex families. The mean coverage was 91.5X with 94.7% of bases covered by 20 reads on average. We identified on average 38,347 variants per individual. Focusing on constrained genes has been shown in the past to be a relevant strategy to identify genes associated with multiple psychiatric disorders^{10,39}. Thus, after variants filtration, we had on average four rare deleterious variants in constrained genes shared by affected individuals per family (Supplementary Table S1) for a total of 34 mutations located in 34 different genes. We then conducted a network and pathway analysis with STRING software in order to determine whether the proteins encoded by these genes worked in interaction or belonged to the same biological pathways. We showed that genes with mutations have a protein–protein interaction enrichment ($p=0.01$) (Fig. S5), highlighting their biological connections. The GO terms analysis showed an overrepresentation of 63 biological pathways (FDR < 0.05), and all but three resisted to 10,000 random permutations when we checked for putative gene size bias (Supplementary Table S2). Combining GO data (Supplementary Tables S2–S4) with a review of the literature, we were able to gather functionally convergent biological pathways and showed that the 34 genes were involved either in the development and morphogenesis ($N=31$), or the regulation of gene expression ($N=25$), or both ($N=22$). In addition, we showed that 29 of them were expressed in neurons (Supplementary Fig. S5 and Supplementary Table S3).

Genes with shared damaging mutation in multiplex families are more frequently mutated in patients with BD than in the general population

In order to determine whether the genes with shared damaging mutations in multiplex families were associated with BD, we sequenced 241 additional independent individuals with BD and compared the mutation rate in patients with the 21,071 non-Finnish European individuals of the ExAC non-psychiatric population. We selected rare (MAF < 1%) missense or loss-of-function variants according to the type of mutations identified in



multiplex families in 21 genes. Four genes (*SMARCC2*, *WDR37*, *UPF2*, and *HELLS*) showed a higher mutation rate in individuals with BD than in the ExAC population after correction for multiple testing (Table 2).

Patients with mutations in BD-associated genes display a specific clinical profile

For the four genes with a higher mutation rate in patients with BD than in general population, we analyzed the clinical profile of mutated subjects. As the severity of illness may depend on its duration, we first checked that there was no difference in the duration of illness between subjects with and without mutations in associated genes (Wilcoxon rank-sum test, $W = 1635$, $p = 0.60$, $p_{adj} = 1.00$). In order to determine whether carrying a rare damaging mutation increased the severity of the disease, we calculated a severity score that combined the age at onset, the age at first hospitalization, as well as the number of total episodes and the number of hospitalizations per year of disease. Mutated individuals showed no difference in severity of the disease when compared with non-mutated ones (Wilcoxon rank-sum test, $W = 1497$, $p = 0.59$, $p_{adj} = 1.00$; Supplementary Fig. S6). Nevertheless, we observed particular clinical manifestations for

individuals with mutation in some of the BD-associated genes (Fig. 2). Thus, individuals with *SMARCC2* mutations had a predominant manic polarity, displaying more manic than depressive episodes in total ($\chi^2 = 15.3$, $df = 1$, $p = 9.1 \times 10^{-5}$, $p_{adj} = 4.6 \times 10^{-4}$; Fig. 2a). Furthermore, individuals with missense variants in *HELLS* displayed more mixed episodes per year of disease than non-mutated subjects (Wilcoxon rank-sum test, $W = 884$, $p = 0.006$, $p_{adj} = 0.03$; Fig. 2b). Although individuals with missense variants in *UPF2* seemed to have more depressive episodes than non-mutated patients, this difference was not significant (Wilcoxon rank-sum test, $W = 865$, $p = 0.49$, $p_{adj} = 1.00$; Fig. 2c). As only two patients had mutation in *WDR37* and only one had available clinical data, we were unable to perform statistical analysis for this gene.

Affected and unaffected subjects in multiplex families display different genetic pattern of vulnerability

We showed that BD- and SZ-associated polymorphisms both contributed to familial forms of BD. In addition, we identified rare variants in constrained genes that were shared by affected subjects in multiplex families and which may also contribute to the BD vulnerability. In

order to explore the genetic architecture explaining the clinical difference between affected and unaffected individuals in multiplex families, we computed a new PRS (PRS_{sum}) combining the information from BD- and SZ-

Table 1 Multivariate analyses in family and non-familial forms of bipolar disorder using polygenic risk scores based on bipolar disorder-specific-, schizophrenia-specific-, and bipolar disorder and schizophrenia shared-associated variants.

| | β | s.e. | z^a | p | p_{adj}^b |
|--|---------------|---------------|-------------|----------------|----------------|
| BD _{fam} vs. controls | | | | | |
| BD _{only} -PRS | 2477.5 | 1097.9 | 1.64 | 0.02 | 0.07 |
| SZ_{only}-PRS | 2892.4 | 931.6 | 2.26 | 0.002 | 0.006 |
| BD-SZ-PRS | 3624.7 | 739.5 | 4.90 | 9.5E-07 | 2.9E-06 |
| BD _{nonfam} vs. controls | | | | | |
| BD_{only}-PRS | 3022.6 | 934.6 | 3.23 | 0.001 | 0.004 |
| SZ _{only} -PRS | 44.7 | 767.0 | 0.06 | 0.95 | 1 |
| BD-SZ-PRS | 3162.6 | 598.6 | 5.28 | 1.3E-07 | 3.8E-07 |
| BD _{fam} vs. BD _{nonfam} | | | | | |
| BD _{only} -PRS | -270.2 | 1362.8 | -0.20 | 0.84 | 1 |
| SZ_{only}-PRS | 3229.7 | 1204.4 | 2.68 | 0.007 | 0.02 |
| BD-SZ-PRS | 384.6 | 847.0 | 0.45 | 0.65 | 1 |

BD bipolar disorder, BD_{fam} familial forms of BD, BD_{nonfam} non-familial forms of BD, PRS polygenic risk score, p nominal p value, s.e. standard error, SZ schizophrenia.

^aWald test.

^bAdjusted p value after Bonferroni correction for multiple testing. Significant

PRS and analyzed this score in regard to the number of rare mutations identified in each individual. Whereas no correlation was observed for affected individuals ($R^2 = -0.014$, $F = 0.76$, $p = 0.40$, $p_{adj} = 0.80$; Fig. 3a), a negative linear regression was observed between the PRS_{sum} and the number of rare variants in unaffected individuals ($R^2 = 0.32$, $F = 7.19$, $p = 0.02$, $p_{adj} = 0.04$; Fig. 3b). The observation that only unaffected individuals in multiplex families had either a high PRS_{sum} and low rate of rare damaging mutations or the opposite, suggested the need to reach a threshold of BD-associated variants to declare the disease and showed the important role of both common and rare variants in BD vulnerability.

Discussion

In this study, we explored the genetic architecture of familial forms of BD combining both common and rare variant analyses. PRS has been demonstrated to be a powerful tool to show the role of common polymorphisms in many psychiatric disorders as well as their shared heritability⁶. However, few studies looked at PRS in families of individuals with BD. Here, we report a higher BD-PRS in families of individuals with BD than in the general population, but no difference was observed between affected and unaffected subjects in multiplex families. Although the higher PRS in BD families is well supported by previous studies^{15,40,41}, those calculating a PRS for unaffected subjects reported inconsistent results depending on the number of subjects under study^{13,15,41}. Here, the small sample size in multiplex families prevent to exclude any differences between affected and

Table 2 Mutation rate in BD cohort and ExAC database for genes mutated in multiplex families.

| Gene symbol | Mutation type | Mutation rate in BD ^a (N = 249) | Mutation rate in ExAC cohort ^b (N = 21,071) | p value ^c | OR [95% CI] | FDR ^d |
|-------------|---------------|--|--|------------------------|-------------------|------------------|
| WDR37 | LoF | 0.004 | 0.0000 | 0.0001 | +∞ [24.42; +∞] | 0.002 |
| SMARCC2 | Missense | 0.016 | 0.004 | 0.0006 | 4.51 [2.20; +∞] | 0.006 |
| UPF2 | Missense | 0.020 | 0.007 | 0.003 | 3.01 [1.61; +∞] | 0.02 |
| HELLS | Missense | 0.018 | 0.007 | 0.008 | 2.72 [1.40; +∞] | 0.04 |
| HSPH1 | LoF | 0.002 | 0.0001 | 0.06 | 21.13 [0.87; +∞] | 0.21 |
| SETD2 | Missense | 0.028 | 0.018 | 0.06 | 1.6117 [0.96; +∞] | 0.21 |
| ATP2B2 | Missense | 0.016 | 0.010 | 0.15 | 1.56 [0.77; +∞] | 0.37 |
| VCL | Missense | 0.014 | 0.009 | 0.15 | 1.64 [0.76; +∞] | 0.37 |
| RBM14 | Missense | 0.012 | 0.007 | 0.16 | 1.6 [0.72; +∞] | 0.37 |
| RPTOR | Missense | 0.008 | 0.005 | 0.24 | 1.61 [0.54; +∞] | 0.50 |
| EIF3L | Missense | 0.004 | 0.002 | 0.3 | 1.86 [0.3254; +∞] | 0.57 |

OR [95% CI] odds ratio with 95% confidence interval, LoF loss-of-function variant.

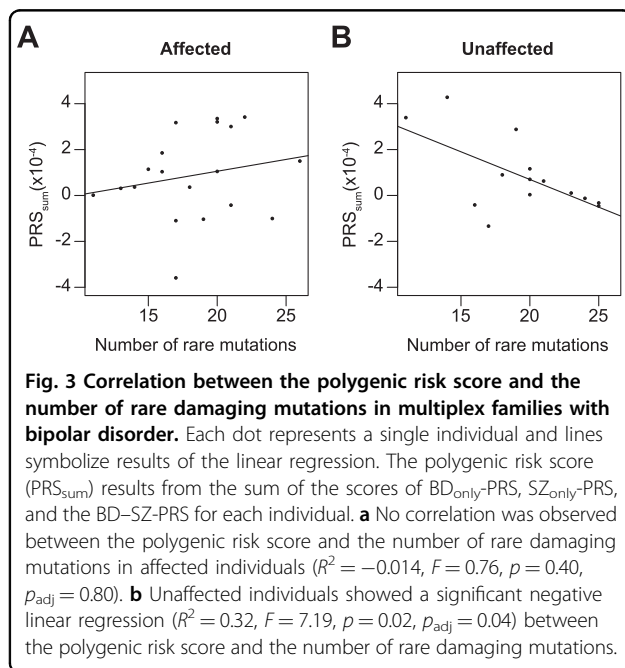
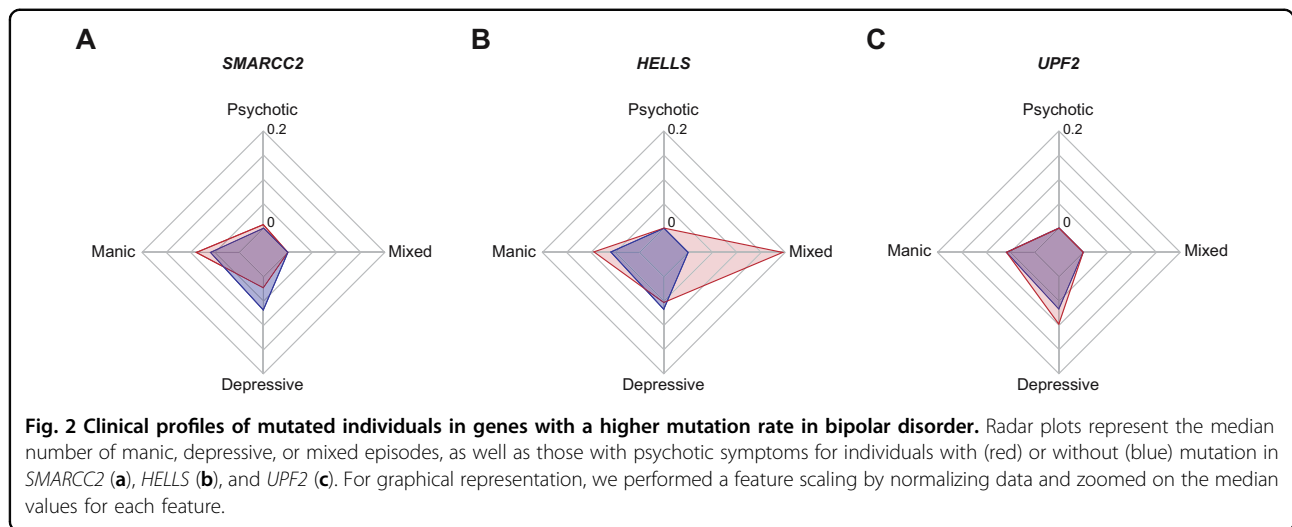
^aBD corresponding to 241 individuals of WES cohort +1 proband by family.

^bNon-Finnish and non-psychiatric European population of ExAC.

^cOne-sided Fisher exact test.

^dFalse discovery rate after Benjamini–Hochberg correction for multiple testing.

results ($p_{adj} < 0.05$) are indicated in bold.



unaffected individuals. Indeed, we estimated that we had a statistical power of 80% to show a large difference (Cohen’s $d = 0.7$) between affected and unaffected individuals, but only 12.5% chance to detect a small effect size ($d = 0.2$). These observations pinpoint the importance of having cohorts as large as possible to conduct such studies and replication on larger samples is required to exclude a difference in PRS between affected and unaffected individuals of multiplex families. Nevertheless, in the whole population of individuals with BD ($N = 362$), we have been able to show that the PRS based on SNPs that have been specifically associated with SZ were higher only in familial forms of BD. The difference in sample size between SZ and BD cohorts to calculate the summary

statistics might make the SZ-PRS more robust than the BD-PRS. However, according to the sample size of this cohort, we estimated that we had a statistical power higher than 80% to show a difference in BD-PRS between familial and non-familial forms of BD if this difference was similar to the one observed using SZ-PRS. SZ-PRS has already been shown to explain a shared genetic variance between SZ and BD^{5,6}. But difference between familial and non-familial forms of BD had never been reported yet. Thus, our results suggest that the shared genetic vulnerability between SZ and BD might concern mainly familial forms of BD. This is consistent with the observation of a high frequency of individuals with SAD in relatives of individuals with BD⁴². In the multiplex families that we studied, no significant difference between affected and unaffected subjects was detected. Although the small number of individuals in multiplex families limits our possibility to draw robust conclusion, this suggests this score alone was unable to explain the difference in clinical status. We thus investigated the impact of the rare genetic variation in constrained genes that were shared by affected individuals in multiplex families. Interestingly, the negative correlation that we observed between PRS and rare damaging variants specifically in unaffected individuals suggests that the balance between rare and common variants might explain the difference in clinical manifestation that we observed in multiplex families of BD. The small size of our sample forces us to consider these results with caution and replication on independent larger samples will be needed to confirm such hypothesis.

Genes with rare damaging mutations which were shared by affected subjects in multiplex families revealed a network of proteins involved in neuronal development or in the regulation of neuronal gene expression. These pathways have already been reported in numerous WES in

ASD or SZ^{7,8,39}, indicating a clear overlap in the etiology of BD with these disorders as already suggested^{9–11,14}. These findings are also consistent with results from GWAS and transcriptome analyses of SZ and, to a lesser extent of ASD^{6,43,44}, suggest that the regulation of gene expression during brain development can be a common pathogenic pathway impaired in BD, SZ, and ASD.

Here, we identified a set of four novel genes, for which the mutation rate in individuals with BD was significantly higher with respect to the general population. Among these genes, *SMARCC2* (MIM: 601734) encodes a scaffolding core subunit of the chromatin remodeling complex mSWI/SNF that controls the accessibility of DNA sequences to transcription factors. This complex plays an essential role during neuron development and the absence of *Smrcc2* in mice resulted in impaired embryonic and adult neurogenesis inducing cognitive dysfunction⁴⁵. Mutations in this gene have already been found in patients with ASD⁴⁶ or neurodevelopmental disorder⁴⁷, and mutations in another subunit of this complex, *SMARCA2*, have been reported in individuals with SZ^{48,49}.

HELLS (MIM: 603946), the second gene identified to be frequently mutated in patients with BD, belongs to the SNF2 subfamily of chromatin remodeling ATPase as *SMARCA2*. This protein is required for genome-wide methylation⁵⁰ and it is responsible of immunodeficiency centromeric region instability and facial anomalies (ICF) syndrome (MIM: 616910). The ICF syndrome is characterized by centromeric instability, as the cytogenetic hallmark, facial dysmorphism, and severe immunodeficiency, as well as developmental delay and intellectual deficit. In mice, *HELLS* was demonstrated as regulator of neural stem cell fate, affecting self-renewal and proliferation of neural progenitor cells⁵¹, suggesting a direct role in nervous system development.

The third gene frequently mutated in our cohort of individuals with BD is *UPF2* (MIM: 605529), a core component of the nonsense-mediated mRNA decay (NMD) pathway, a surveillance pathway that eliminates mRNA with premature translation termination codon. In *Drosophila*, *Upf2* plays a role for proper development of synapse architecture and synaptic vesicle efficacy⁵². Moreover, copy number variants in this gene were found in intellectual disabilities syndrome with DiGeorge syndrome associated⁵³ and de novo point mutations have been identified in a patient with SZ⁵⁴. Altogether, these observations suggest a role for *UPF2* in nervous system development.

Few things are known about the last gene, *WDR37*. Its protein belongs to the WD40 repeat domain family that is characterized by diverse cellular function, as chromatin assembly, RNA processing, immunity, or development⁵⁵. This interacting domain scaffolds protein–protein or

protein–DNA interactions. Although *WDR37* function is unknown, mice knockout for this gene show a larger brain than wild-type littermates⁵⁶. Interestingly, 57 copy number variants including this gene have been reported in the DECIPHER v9.30 released (<https://decipher.sanger.ac.uk/>). Among associated phenotypes, we reported developmental delay, seizures, intellectual disabilities, and ASD.

In summary, we demonstrated that both common and rare variants contribute to phenotypic outcome in multiplex families of BD, and our data suggest that the difference between the affected and unaffected status of individuals within multiplex families might come from the balance between the two types of variants. In addition, we confirmed the shared genetic vulnerability between BD, SZ, and ASD and suggest the familial forms of BD might share a higher heritability with SZ than non-familial ones. Finally, we highlighted a neurodevelopmental impairment in familial forms of BD. Altogether, our results shed new lights on the genetic architecture and biology of BD and highlight the importance of using large pedigrees with a high density of affected individuals to investigate the etiology of the disease.

Acknowledgements

This work was supported by the *Investissements d'Avenir* program managed by the *Agence Nationale pour la Recherche* (ANR) under references ANR-11-IDEX-0004-02 (Labex BioPsy) and ANR-10-COHO-10-01 (Cohorte PSY-COH). This work also received financial support from the ANR program on mental health and addiction under reference ANR-13-SAMENTA-0003-01 (ANTIDDOCT), from the International Foundation (IF), the Institut National pour la Santé et la Recherche Médicale (Inserm) and the Réseau Thématique de Recherche et de Soins en Santé Mentale (Fondation FondaMental®, Prix Henry Rousset). The Jamain's team is affiliated to the Paris School of Neuroscience (ENP) and the Bio-Psy Laboratory of Excellence. We are very grateful to patients with bipolar disorder, their family, and control subjects for their participation. The authors thank Dr. Eli Stahl and the Bipolar Disorder Group of the PGC for having provided PGC BD summary statistics excluding the French cohort. In addition, we thank the Clinical Investigation Centre and the Biological Resources Platform of Mondor Hospital for technical assistance FondaMental Advanced Centers of Expertise in Bipolar Disorders (FACE-BD).

FondaMental Advanced Center of Expertise in Bipolar Disorders (FACE-BD) Group

FACE-BD Clinical Coordinating Center (Fondation FondaMental): B. Etain, C. Henry, E. Olié, M. Leboyer, E. Haffen, and P. M. Llorca. FACE-BD Data Coordinating Center (Fondation FondaMental): V. Barteau, S. Bensalem, O. Godin, H. Laouamri, and K. Souryis. FACE-BD Clinical Sites and Principal Collaborators in France: AP-HP, DHU PePSY, Pôle de Psychiatrie et d'Addictologie des Hôpitaux Universitaires H Mondor, Créteil: S. Hotier, A. Pelletier, N. Drancourt, J. P. Sanchez, E. Saliou, C. Hebbache, J. Petrucci, L. Willaume, and E. Bourdin; AP-HP, GH Saint-Louis–Lariboisière–Fernand Widal, Pôle Neurosciences, Paris: F. Bellivier, M. Carminati, B. Etain, E. Marlinge, and J. Meheust; Hôpital C. Perrens, Centre Expert Trouble Bipolaire, Service de Psychiatrie Adulte, Pôle 3-4-7, Bordeaux: B. Antonioli, A. Desage, S. Gard, A. Jutant, K. Mbailara, I. Minois, and L. Zanouy; Département d'Urgence et Post Urgence Psychiatrique, CHRU Montpellier, Montpellier: C. Abettan, L. Bardin, A. Cazals, P. Courtet, B. Deffinis, D. Ducasse, M. Gachet, A. Henrion, E. Martinerie, F. Molière, B. Noisette, E. Olié, and G. Tarquini; Pôle de Psychiatrie, addictologie et pédopsychiatrie, Hôpital Sainte Marguerite, Assistance Publique Hôpitaux de Marseille: J. M. Azorin, R. Belzeaux, N. Correard, I. Muraccioli, F. Groppi, L. Lescalier, M. Rebattu, and N. Viglianese; Service de Psychiatrie et Psychologie Clinique, CHU de Nancy, Hôpitaux de Brabois, Vandoeuvre Les Nancy: R. Cohen, J.P. Kahn, M. Milazzo, and O. Wajsbrot-Elgrabli; Clinique Universitaire de Psychiatrie, CHU de Grenoble et des Alpes, Grenoble: T. Bougerol, B.

Fredembach, S. Garçon, P. Grignon, A. Perrin, and M. Polosan; Centre Hospitalier de Versailles, Service Universitaire de Psychiatrie d'adultes, Le Chesnay; A. M. Galliot, I. Grévin, A. S. Cannavo, N. Kayser, G. Liaskovsky, C. Passerieux, and P. Roux; Service de Psychiatrie, Centre Hospitalier Princesse Grace, Monaco; V. Aubin, I. Cussac, MA Dupont, J. Loftus, and I. Medecin; AHPH, Département de Psychiatrie, Hôpital Louis Mourier, Colombes, France; A. Bing, C. Dubertret and N. Mazer; CHU de Clermont-Ferrand, Centre Expert Troubles Bipolaires FondaMental, Clermont-Ferrand, France; P. M. Llorca, L. Samalin, L. Fournes, D. Lacelle, S. Pires, C. Doriât, and O. Blanc.

Author details

¹INSERM U955, Psychiatrie Translationnelle, Créteil 94000, France. ²Université Paris Est, Faculté de Médecine, Créteil 94000, France. ³Fondation FondaMental, Créteil 94000, France. ⁴Université de Lorraine, CHRU de Nancy et Pôle de Psychiatrie et Psychologie Clinique, Centre Psychothérapique de Nancy, Laxou 54520, France. ⁵AP-HP, Hôpital H. Mondor–A. Chenevier, Plateforme de Ressources Biologiques, Créteil 94000, France. ⁶Inserm, Centre d'Investigation Clinique 1430 and AHPH, Henri Mondor Hospital, Créteil 94000, France. ⁷Centre Expert Troubles Bipolaires, Service de Psychiatrie Adulte, Hôpital Charles-Perrens, Bordeaux 33000, France. ⁸AP-HP, GH Saint-Louis-Lariboisière-F. Widal, Département de Psychiatrie et de Médecine Addictologique, Paris 75010, France. ⁹Université Paris Diderot, Sorbonne Paris Cité, Paris 75010, France. ¹⁰Inserm, UMR-S1144, Paris 75010, France. ¹¹Pôle de Psychiatrie, Assistance Publique Hôpitaux de Marseille, INT-UMR7289, CNRS Aix-Marseille Université, Marseille 13009, France. ¹²AP-HP, Département de Psychiatrie, Hôpital Louis Mourier, INSERM U894, Université de Paris, Colombes 92700, France. ¹³AP-HP, DHU PePSY, Pôle de Psychiatrie et d'Addictologie des Hôpitaux Universitaires Henri Mondor, Créteil 94000, France. ¹⁴Département urgence et Post-urgence psychiatrique, CHU Montpellier, INSERM U1061, Université de Montpellier, Montpellier 34000, France. ¹⁵Service Universitaire de Psychiatrie d'Adultes, Centre Hospitalier de Versailles, Laboratoire HandiRESP, EA4047, UFR des Sciences de la Santé Simone Veil, Université de Versailles Saint-Quentin-En-Yvelines, Le Chesnay 78150, France. ¹⁶Université Grenoble Alpes, CHU de Grenoble et des Alpes, Grenoble Institut des Neurosciences (GIN) Inserm U 1216, La Tronche 38700, France

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-020-0783-0>).

Received: 20 September 2019 Revised: 14 January 2020 Accepted: 28 February 2020

Published online: 28 April 2020

References

- Ferrari, A. J. et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord.* **18**, 440–450 (2016).
- Kieseppä, T., Partonen, T., Haukka, J., Kaprio, J. & Lonnqvist, J. High concordance of bipolar I disorder in a nationwide sample of twins. *Am. J. Psychiatry* **161**, 1814–1821 (2004).
- Lichtenstein, P. et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
- Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
- International Schizophrenia Consortium. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Cross-Disorder Group of the Psychiatric Genomics Consortium. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
- De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
- Fromer, M. et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
- Goes, F. S. et al. Exome sequencing of familial bipolar disorder. *JAMA Psychiatry* **73**, 590–597 (2016).
- Kataoka, M. et al. Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol. Psychiatry* **21**, 885–893 (2016).
- Lescai, F. et al. Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder. *Transl. Psychiatry* **7**, e1034 (2017).
- Husson, T. et al. Identification of potential genetic risk factors for bipolar disorder by whole-exome sequencing. *Transl. Psychiatry* **8**, 268 (2018).
- Maaser, A. et al. Exome sequencing in large, multiplex bipolar disorder families from Cuba. *PLoS ONE* **13**, e0205895 (2018).
- Toma, C. et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. *Transl. Psychiatry* **8**, 65 (2018).
- Szatkiewicz, J. et al. The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. *Transl. Psychiatry* **9**, 60 (2019).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- American Psychiatric Association (eds). *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn (American Psychiatric Press, Washington, DC, 1994).
- Henry, C. et al. A French network of bipolar expert centres: a model to close the gap between evidence-based medicine and routine practice. *J. Affect. Disord.* **131**, 358–363 (2011).
- Psychiatric Gwas Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).
- Jamain, S. et al. Common and rare variant analysis in early-onset bipolar disorder vulnerability. *PLoS ONE* **9**, e104326 (2014).
- Etain, B. et al. A SNAP25 promoter variant is associated with early-onset bipolar disorder and a high expression level in brain. *Mol. Psychiatry* **15**, 748–755 (2010).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
- Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
- Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **30**, 1266–1272 (2014).
- The Genomes Project C. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
- Cingolani, P. 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)* **6**, 80–92 (2012).
- Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

37. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
38. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
39. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
40. Collins, A. L. et al. Identifying bipolar disorder susceptibility loci in a densely affected pedigree. *Mol. Psychiatry* **18**, 1245–1246 (2013).
41. Fullerton, J. M. et al. Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young At-Risk Individuals. *Am. J. Med. Genet. B* **168**, 617–629 (2015).
42. Laursen, T. M. et al. Family history of psychiatric illness as a risk factor for 624 schizoaffective disorder: a Danish register-based cohort study. *Arch. Gen. Psychiatry* **62**, 841–848 (2005).
43. Brainstorm Consortium et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**(6395), eaap8757 (2018).
44. Gandal, M. J. et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
45. Tuoc, T. et al. Ablation of BAF170 in developing and postnatal dentate gyrus affects neural stem cell proliferation, differentiation, and learning. *Mol. Neurobiol.* **54**, 4618–4635 (2017).
46. Neale, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
47. Machol, K. et al. Expanding the spectrum of BAF-related disorders: de novo variants in SMARCC2 cause a syndrome with intellectual disability and developmental delay. *Am. J. Hum. Genet.* **104**, 164–178 (2019).
48. Walsh, T. et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
49. Koga, M. et al. Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum. Mol. Genet.* **18**, 2483–2494 (2009).
50. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* **15**, 2940–2944 (2001).
51. Han, Y. et al. Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells. *Sci. Rep.* **7**, 1136 (2017).
52. Long, A. A. et al. The nonsense-mediated decay pathway maintains synapse architecture and synaptic vesicle cycle efficacy. *J. Cell Sci.* **123**, 3303–3315 (2010).
53. Nguyen, L. S. et al. Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. *Hum. Mol. Genet.* **22**, 1816–1825 (2013).
54. Gulsuner, S. et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
55. Schapira, M., Tyers, M., Torrent, M. & Arrowsmith, C. H. WD40 repeat domain proteins: a novel target class? *Nat. Rev. Drug Discov.* **16**, 773–786 (2017).
56. Kannan, M. et al. WD40-repeat 47, a microtubule-associated protein, is essential for brain development and autophagy. *Proc. Natl. Acad. Sci. USA* **114**, E9308–E9317 (2017).