Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders

Till F. M. Andlauer (PhD)[1,2]*, Jose Guzman-Parra (PhD)[3]*, Fabian Streit (PhD)[4]*, Jana Strohmaier (PhD)[4], Maria José González (M.D.)[5], Susana Gil Flores (M.D.)[6], Francisco J. Cabaleiro Fabeiro (M.D.)[7], Francisco del Río Noriega (M.D.)[8], Fermin Perez Perez (M.D.)[5], Jesus Haro González (M.D.)[9], Guillermo Orozco Diaz (M.D.)[11], Yolanda de Diego-Otero (M.D.)[3], Berta Moreno-Kuestner (M.D.)[8], Georg Auburger (M.D.)[12], Franziska Degenhardt (M.D.)[13,14], Stefanie Heilmann-Heimbach (PhD)[13,14], Stefan Herms (MSc) [13,14,15], Per Hoffmann (PhD)[13,14,15,16], Josef Frank (MSc)[4], Jerome C. Foo (PhD)[4], Jens Treutlein (PhD)[4], Stephanie H. Witt (PhD)[4], Sven Cichon (PhD)[14,15,16], Manolis Kogevinas (M.D.)[17], Bipolar Disorder Working Group of the Psychiatric Genomics Consortium[§], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[§], Fabio Rivas (M.D.)[3], Fermín Mayoral (M.D.)[3], Bertram Müller-Myhsok (M.D.)[1,18,19], Andreas J. Forstner (M.D.)[13,14,15,20]*, Markus M. Nöthen (M.D.)[13,14]*, Marcella Rietschel (M.D.)[4]*[#]

[1] Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany
[2] Department of Neurology, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
[3] Department of Mental Health, University Regional Hospital of Málaga, Institute of Biomedicine of Málaga (IBIMA), Málaga, Spain
[4] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[5] Department of Mental Health, Hospital of Puerto Real, Cádiz, Spain
[6] Department of Mental Health, University Hospital of Reina Sofia, Cordoba, Spain
[7] Department of Mental Health, Hospital of Jaen, Jaen, Spain
[8] Department of Mental Health, Hospital of Jerez de la Frontera, Jerez de la Frontera, Cadiz, Spain
[9] Department of Mental Health, Hospital Punta de Europa, Algeciras, Spain
[10] Department of Psychiatry, Carlos Haya Regional University Hospital, Málaga, Spain
[11] Unidad de Gestión Clínica del Dispositivo de Cuidados Críticos y Urgencias del Distrito Sanitario Málaga - Coin-Guadalhorce, Málaga, Spain
[12] Department of Neurology, Goethe University Medical School, Frankfurt am Main, Germany
[13] Institute of Human Genetics, University Hospital Bonn, Bonn, Germany
[14] Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
[15] Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland
[16] Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany
[17] Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
[18] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.
[19] Institute of Translational Medicine, University of Liverpool, Liverpool, UK.
[20] Department of Psychiatry (UPK), University of Basel, Basel, Switzerland
*joint first and last authorship
[§]members of Bipolar and Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium are listed in the Supplementary Material.

#corresponding author:
Prof. Dr. Marcella Rietschel
Department of Genetic Epidemiology in Psychiatry
Central Institute of Mental Health
J5, D-68159 Mannheim, Germany
Tel.: +49-621-1703-6051
Fax: +49-621-1703-6055
Email: Marcella.Rietschel@zi-mannheim.de

Abstract

Multiplex families with a high prevalence of a psychiatric disorder are often examined to identify rare genetic variants with large effect sizes. In the present study, we analysed whether the risk for bipolar disorder (BD) in BD multiplex families is influenced by common genetic variants. Furthermore, we investigated whether this risk is conferred mainly by BD-specific risk variants or by variants also associated with the susceptibility to schizophrenia or major depression. In total, 395 individuals from 33 Andalusian BD multiplex families as well as 438 subjects from an independent, sporadic BD case-control cohort were analysed. Polygenic risk scores (PRS) for BD, schizophrenia, and major depression were calculated and compared between the cohorts. Both the familial BD cases and unaffected family members had significantly higher PRS for all three psychiatric disorders than the independent controls, suggesting a high baseline risk for several psychiatric disorders in the families. Moreover, familial BD cases showed significantly higher BD PRS than unaffected family members and sporadic BD cases. A plausible hypothesis is that, in multiplex families with a general increase in risk for psychiatric disease, BD development is attributable to a high burden of common variants that confer a specific risk for BD. The present analyses, therefore, demonstrated that common genetic risk variants for psychiatric disorders are likely to contribute to the high incidence of affective psychiatric disorders in the multiplex families. The PRS explained only part of the observed phenotypic variance and rare variants might have also contributed to disease development.

Introduction

Bipolar disorder (BD), characterized by alternating episodes of mania and depression, has a lifetime prevalence of approximately 1% and is a substantial contributor to disability throughout the world[1]. Nevertheless, reliable data concerning the aetiology of BD remain scarce. The heritability of BD is estimated to be above 70%[2-4], thus demonstrating an important genetic component in the development of the disorder. Genome-wide association studies (GWAS) in case/control samples have reported that common genetic risk factors with minor allele frequencies (MAF) of ≥1% explain a substantial proportion of the genetic risk for BD[5-11]: the heritability explained by such common variants is estimated to be 0.17-0.23 on a liability scale[12]. Common variants also make a substantial contribution to the development of schizophrenia (SCZ) and major depressive disorder (MDD)[13,14]. These three psychiatric disorders have a shared genetic component, whereby relatives of patients with BD have, in addition to BD, an increased risk for MDD and SCZ[15]. In fact, GWAS have shown that many genetic risk variants are associated with all three disorders[16-20].

Besides common variants with small individual effects, rare variants with large effect sizes may also contribute to BD development[21,22]. In theory, such highly penetrant variants should be enriched in families with a high prevalence of illness, termed multiplex families, in comparison to sporadic BD cases. However, it remains unclear whether and to what extent disease incidence in multiplex families is caused by rare variants, a high load of common variants, or a combination of both.

To elucidate the molecular genetic causes of BD, we established the Andalusian Bipolar Family (ABiF) study in 1997, which recruited BD multiplex families[23-25]. In the present analyses, we first investigated whether common genetic variants make a significant contribution to the occurrence of BD in ABiF families. Next, we examined whether BD development was attributable to (a) BD-specific risk variants, (b) variants conferring risk to all three disorders BD, MDD, and SCZ, or, (c), a combination of both. To this end, polygenic risk scores (PRS) based on GWAS of BD, MDD, and SCZ were calculated for and compared between ABiF family members and sporadic BD cases and unscreened controls from the same population.

Materials and Methods

Sample description

The ABiF study recruited BD multiplex families in Andalusia, Spain[23-25]. The present analyses included 395 members of 33 ABiF families. Diagnoses were assigned by two trained clinicians according to DSM IV criteria using the best estimate approach[23]. Diagnoses comprised (Table 1): BD, n=166 (FAM_{BD}); MDD, n=78 MDD (FAM_{MDD}); no history of an affective disorder n=151 ($FAM_{unaffected}$). Six unaffected individuals with a history of substance abuse were excluded from the analyses. Forty-four subjects have married into the families and had no parent in the ABiF cohort (36 unaffected; 8 MDD). Furthermore, an independent, previously reported Spanish BD case/control (CC) sample was analyzed[9].

Table 1: **Characteristics of 389 individuals from the 33 ABiF families.**

	BD (n=166)	**MDD** (n=78)	**Unaffected** (n=145)	**Significance level of differences between diagnostic groups**
Age at interview	40.5 (18.53)	44.5 (17.05)	44 (22.24)	χ^2=3.892; *p*=0.143
Age at onset	20 (7.41) *missing=3*	26 (11.86) *missing=1*		χ^2=17.482; *p*<0.001
Gender (% female)	103 (62.05)	54 (69.23)	51 (35.17)	χ^2=32.210; *p*<0.001
Married-in	0 (100.0)	8 (10.03)	36 (24.83)	χ^2=47.665; *p*<0.001
Educational level				χ^2=2.092; *p*=0.143
Primary school	118 (71.51)	53 (67.95)	102 (70.34)	
Secondary school	39 (23.63)	18 (23.08)	36 (24.83)	
University degree	8 (4.85) *missing=1*	7 (8.97)	7 (4.83)	
Severe impairment during the disorder	105 (65.62) *missing=6*	4 (5.55) *missing=6*		χ^2=71.931; *p*<0.001
History of psychosis	110 (66.26)	4 (5.13)		χ^2=101.218; *p*<0.001
History of suicide attempts	41 (24.70)	2 (2.56)	1 (0.69) *missing=1*	χ^2=51.670; *p*<0.001

Six unaffected individuals with a history of substance abuse were excluded from the analyses and are not shown in this table. *Missing*: number of individuals with missing data. Age was analysed using the Kruskal-Wallis test; median and median absolute deviation are shown. All other values were analysed using chi-squared (χ^2) tests with two degrees of freedom; number (n) and percentage (%) of subjects are shown. Note that of the 44 married-in individuals listed here, only 35 unaffected, married-in family members were part of the combined FAM+CC dataset. These subjects were excluded from analyses of the combined dataset unless specified otherwise.

After quality control (QC), the combined dataset of both cohorts comprised data from 384 FAM (163 FAM_{BD}, 73 FAM_{MDD}, 142 $FAM_{unaffected}$, and 6 $FAM_{unaffected}$ with a history of substance abuse) and 438 CC subjects (161 sporadic BD cases and 277 unscreened controls). This dataset contained 35 unaffected, married-in family members who were excluded from analyses using the combined sample (unless specified otherwise). A detailed description of QC procedures is provided in the Supplementary Methods.

The study was approved by the respective local ethics committees, and all participants provided written informed consent. For five adolescents (age 15-17 years), written informed consent was also obtained from parents.

Genotyping and Imputation

Genome-wide genotyping of the FAM sample was carried out using the Illumina Infinium PsychArray BeadChip (PsychChip). QC and population substructure analyses were performed in PLINK[26], as described in the Supplementary Methods. Genotyping and basic QC of the CC sample were conducted previously and are described elsewhere[9]. The genotypes of the CC dataset were merged with those of the FAM sample.

Genotype data were imputed to the 1000 Genomes phase 3 reference panel using SHAPEIT and IMPUTE2[27-29]. After imputation and post-imputation QC, the combined dataset of both cohorts contained 6862461 variants with an INFO metric of ≥0.8 and a MAF of ≥1%. The imputed FAM dataset without the CC subjects contained 8628089 variants.

Calculation of polygenic risk scores

PRS were calculated in *R* v3.3[30] using imputed genetic data. For each PRS, the effect sizes of variants below a selected *p*-value threshold, both obtained from large GWAS (training data), were multiplied by the imputed SNP dosage in the test data and then summed to produce a single PRS per threshold. The weighted PRS thus represent cumulative, additive risk. For each disorder, ten PRS based on different GWAS *p*-value thresholds were calculated, ranging from $p_{PRS}<5\times10^{-8}$ to $p_{PRS}<0.2$. The number of SNPs used for each PRS is shown in Supplementary Table S1. For additional details, see the Supplementary Methods.

For BD, MDD, and SCZ diagnoses, summary statistics of GWAS by the Psychiatric Genomics Consortium (PGC) were used as training data. For BD, the data freeze contained 20352 cases and 31358 controls, after exclusion of the Spanish samples[12]. For MDD and SCZ, published datasets were used. These contained 130664 cases and 330470 controls for MDD[14] and 33640 cases and 43456 controls for SCZ[13]. Variants with an INFO metric of <0.6 in the GWAS summary statistics were removed. For comparison, PRS for late-onset Alzheimer's disease (LOAD) were calculated, based on a GWAS by the International Genomics of Alzheimer's Project (IGAP) with 17008 cases and 37154 controls[31]. For further details, see the Supplementary Methods.

Shared psychiatric PRS were generated using all variants showing an association at $p<0.05$ in the GWAS of BD, SCZ, and MDD and for which effect sizes pointed in the same direction across studies. For this shared set of variants, *p*-values were calculated using random-effects meta-analysis.

Genome-wide inferred statistics (GWIS) were calculated as published elsewhere[32]. As recommended for this method, variants with an INFO metric of <0.9 or >1.1 were removed. Furthermore, 10000 random PRS for each of the ten *p*-value thresholds were calculated. To this end, random variants from across the genome were drawn, using the same number of variants as for the BD PRS at each threshold and random effect sizes from the pool of all available BD, SCZ, and MDD effects.

Statistical Analysis

PRS analyses were conducted in *R* using the function *polygenic* of the package *GenABEL*[33], which implements a linear mixed model that takes family structure into account. Test statistics, including 95% confidence intervals (CI), were calculated using bootstrapping (package *boot*[34,35]). Following the hypothesis that family members or subjects with a psychiatric diagnosis have increased PRS for psychiatric disorders, one-sided *p*-values were calculated for all PRS-based analyses. Since the residuals were not normally distributed, *p*-values were confirmed using permutation analysis (10000 permutations). Significance thresholds were corrected for multiple testing using the Bonferroni method. For further details see the Supplementary Methods.

In analyses of the combined FAM and CC dataset, sex was used as a covariate. In the analysis of FAM data alone, sex, the age at the interview, and whether an individual had married into the family were used as covariates. Age was not available for the CC controls.

Results

FAM_{BD} cases had higher psychiatric PRS than controls from the general population

We first examined whether FAM_{BD} cases, compared to independent $CC_{controls}$, had higher PRS for the three psychiatric disorders BD, SCZ, and MDD. Selected results for these comparisons are shown in Fig. 1 (*i.e.*, for each type of PRS, the test statistics for the PRS with the threshold p_{PRS} that showed the strongest association); full results

for all ten thresholds per PRS type are provided in Supplementary Figs. S1-S2 and Supplementary Table S2. On average, FAM_{BD} cases had higher BD PRS than $CC_{controls}$ at all ten thresholds. The most substantial support for an increased BD PRS was found with the threshold p_{PRS}=0.1 (one-sided $p=1.9\times10^{-20}$, significance threshold $\alpha=0.05/10=5\times10^{-3}$). FAM_{BD} cases also had significantly higher SCZ and MDD PRS than $CC_{controls}$. While the BD and SCZ PRS were increased across all thresholds, the association using MDD PRS was only significant for $p_{PRS}\geq0.1$.

BD and SCZ, and, to a lesser degree, BD and MDD, show a strong genetic correlation[14,16,18-20]. We, therefore, analysed whether variants associated with all three disorders contributed to the increased psychiatric PRS in FAM_{BD} cases. *Shared* PRS generated from the variants associated with BD, SCZ, and MDD were increased at $p_{PRS}\geq0.01$ in FAM_{BD} cases compared to $CC_{controls}$. The additional contribution of disorder-specific variants was examined using GWIS[32], which is a method previously applied to the analysis of genetic factors unique to either BD or SCZ. BD GWAS summary statistics were thus corrected for the MDD GWAS results to calculate a BD-MDD PRS less biased by the genetic correlation of BD with MDD. The BD-MDD PRS were significantly increased in FAM_{BD} cases compared to $CC_{controls}$ across all thresholds. When correcting BD summary statistics for SCZ, BD-SCZ PRS were only associated at the single $p_{PRS}=1\times10^{-4}$. Furthermore, the SCZ-BD GWIS PRS was significantly increased in FAM_{BD} cases at $p_{PRS}<1\times10^{-7}$, whereas the MDD-BD PRS was not. These results indicate that FAM_{BD} cases had a higher load of common risk variants for BD, SCZ, and MDD than the population-based controls. This difference was mainly attributable to variants specific to BD. However, the increased risk of FAM_{BD} cases was also influenced by variants that confer risk for all three psychiatric diagnoses.

Finally, to confirm that the FAM_{BD} cases had an increased PRS specifically for the tested psychiatric disorders but not for unrelated diseases, a PRS for late-onset Alzheimer's disease (LOAD) was calculated. No significant increase was found at any threshold. In addition, 10000 random PRS were generated for each of the ten thresholds. The threshold at which the lowest mean association *p*-value across all 10000 random PRS was observed ($p_{PRS}=1\times10^{-5}$; $p=0.496$) was selected for comparisons. Associations of the PRS for BD, MDD, and SCZ, and of the *Shared* PRS were significantly stronger than this random PRS in FAM_{BD} compared to $CC_{controls}$ (Table 2). These results demonstrate that the observed increase in the risk burden of FAM_{BD} cases compared to $CC_{controls}$ was not attributable to systematic, genome-wide genetic differences between the cohorts.

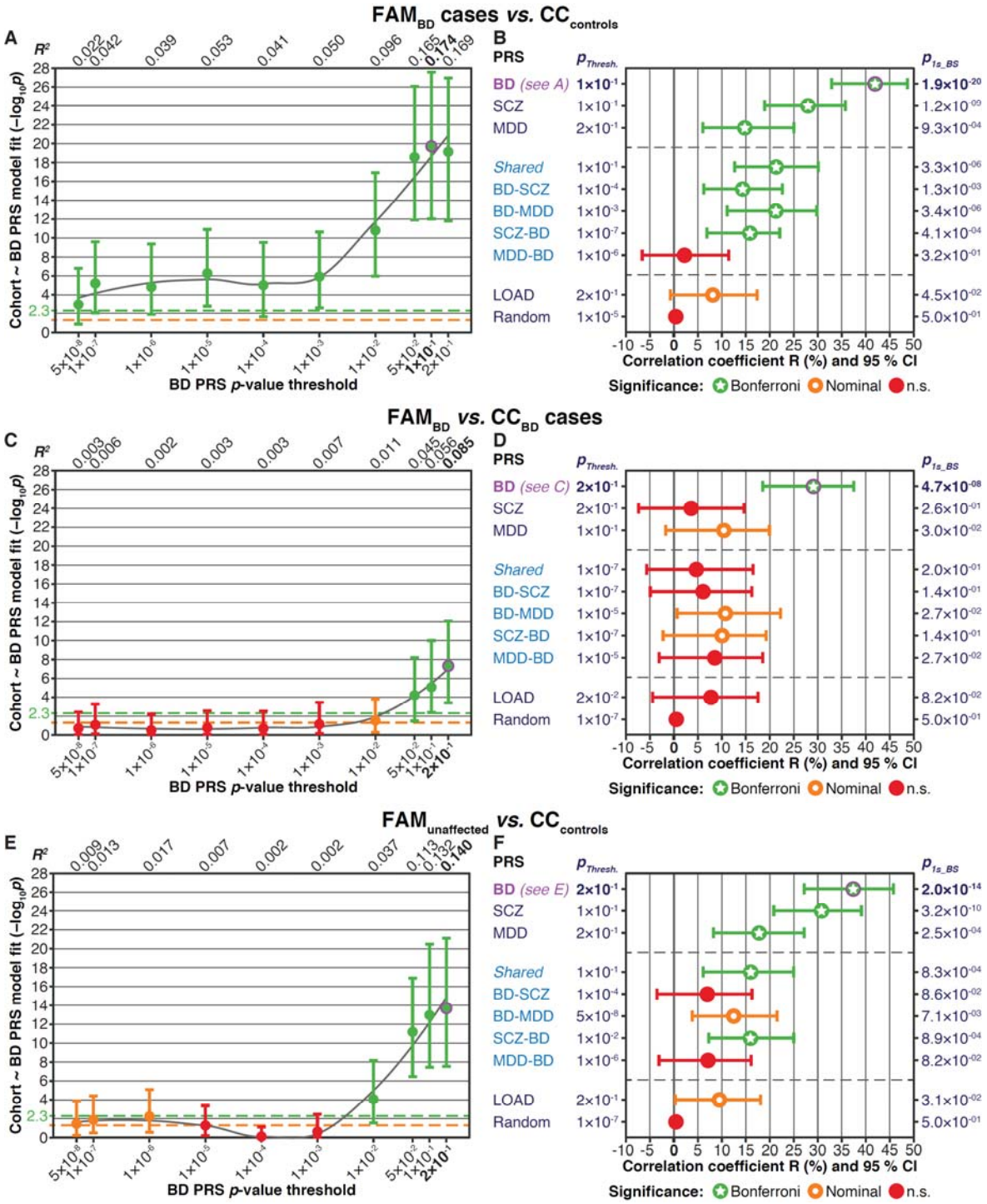

Figure 1: Association analyses comparing the various PRS between the FAM and CC samples. Married-in family members were excluded from these analyses. The plots show one-sided *p*-values, following the hypothesis that family members have higher PRS than individuals from the CC samples. **A-B:** Comparison of FAM_{BD} cases to $CC_{controls}$. **A:** FAM_{BD} cases had higher BD PRS across all ten p_{PRS} thresholds. The plot shows one-sided *p*-values (filled circles), and 95% CI calculated using bootstrapping. The various p_{PRS} thresholds are shown on the x-axis, and the association strength ($-\log_{10} p$) is shown on the y-axis. The coefficient of determination R^2 of the linear models is indicated at the top of the plot. Orange dashed line: nominal significance threshold (α=0.05), green dashed line: significance threshold after Bonferroni correction for multiple testing (α=0.05/10=0.005). Results for

each threshold are coloured in accordance with their degree of significance: red = not significant, orange = nominally significant, green = significant after correction for multiple testing. The top-associated PRS (p_{PRS}=0.1) is indicated in bold font and was marked by a magenta circle (also in B). **B:** For ten different PRS, this plot shows association statistics for the top-associated p_{PRS} thresholds. The x-axis shows the association strength as the coefficient of multiple correlation *R* and 95% CI calculated by bootstrapping. Note that *R* instead of R^2 has been used in this plot to allow for directionality. BD, SCZ, MDD: Standard PRS using the respective PGC GWAS summary statistics. *Shared*: Shared psychiatric PRS (SNPs with BD, MDD, SCZ p<0.05, random effects meta-analysis). BD-SCZ, BD-MDD: BD-specific GWIS PRS corrected for SCZ and MDD, respectively. SCZ-BD and MDD-BD: GWIS PRS for SCZ and MDD, each corrected for BD. LOAD: PRS for late-onset Alzheimer's disease. Random: Mean and CI of the 10000 random PRS at the p_{PRS} with the lowest mean association *p*-value of all random PRS across p_{PRS}. The column to the left of the plot: p_{PRS} with the strongest association. Supplementary Fig. S2 shows plots for all p_{PRS}. Column to the right: p_{1s_BS} = one-sided *p*-value calculated using bootstrapping. For full association test statistics, see Supplementary Table S2. Abbreviations: Bonferroni = significant after Bonferroni correction for multiple testing; nominal = nominally significant (p<0.05); n.s. = not significant. **C-D:** Comparison of FAM_{BD} cases and sporadic CC_{BD} cases. See Supplementary Fig. S3 and Table S3 for more detailed plots and full association test statistics. **E-F:** Comparison of $FAM_{unaffected}$ and $CC_{controls}$. See Supplementary Fig. S4 and Table S4 for more detailed plots and full association test statistics.

Table 2: Permutation test of 10000 random PRS at the *p*-value threshold showing the lowest mean association *p*-value (min. mean p_{PRS}).

Group	Comparison to disorder PRS	Min. mean p_{PRS}	Number of random PRS permutations	Association of random PRS ≥ disorder PRS (n)	*p*-value of permutation test	95% CI of the *p*-value
FAM_{BD} *vs.* $CC_{controls}$	BD	1×10^{-5}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	SCZ	1×10^{-5}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	MDD	1×10^{-5}	10000	109	1.1×10^{-2}	$9.0\times10^{-3}–1.3\times10^{-2}$
	Shared	1×10^{-5}	10000	3	3.0×10^{-4}	$6.2\times10^{-5}–8.8\times10^{-4}$
FAM_{BD} *vs.* CC_{BD}	BD	1×10^{-7}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	SCZ	1×10^{-7}	10000	3031	3.0×10^{-1}	$2.9\times10^{-1}–3.1\times10^{-1}$
	MDD	1×10^{-7}	10000	781	7.8×10^{-2}	$7.3\times10^{-2}–8.4\times10^{-2}$
	Shared	1×10^{-7}	10000	2428	2.4×10^{-1}	$2.3\times10^{-1}–2.5\times10^{-1}$
$FAM_{unaffected}$ *vs.* $CC_{controls}$	BD	1×10^{-7}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	SCZ	1×10^{-7}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	MDD	1×10^{-7}	10000	33	3.3×10^{-3}	$2.3\times10^{-3}–4.6\times10^{-3}$
	Shared	1×10^{-7}	10000	75	7.5×10^{-3}	$5.9\times10^{-3}–9.4\times10^{-3}$
FAM_{MDD} *vs.* $CC_{controls}$	BD	2×10^{-1}	10000	0	$<1\times10^{-4}$	$0–3.7\times10^{-4}$
	SCZ	2×10^{-1}	10000	426	4.3×10^{-2}	$3.9\times10^{-2}–4.7\times10^{-2}$
	MDD	2×10^{-1}	10000	2	2.0×10^{-4}	$2.4\times10^{-5}–7.2\times10^{-4}$
	Shared	2×10^{-1}	10000	118	1.2×10^{-2}	$9.8\times10^{-3}–1.4\times10^{-2}$

The number of random PRS that showed the same or a stronger association (*i.e.*, one-sided *p*-values) in the linear model comparing PRS between the FAM and CC samples was counted to generate a permutation *p*-value. CI = confidence interval, calculated using the *R* package *binom* (method: *exact*).

FAM_{BD} cases had higher BD PRS than sporadic BD cases

We next compared the FAM_{BD} cases to sporadic BD cases from Andalusia. The BD PRS was significantly higher in FAM_{BD} than in CC_{BD} cases at p_{PRS}≥0.05 (Fig. 1C-D, Supplementary Figs. S1 and S3, Supplementary Table S3). However, no other type of PRS was increased in FAM_{BD} compared to CC_{BD} cases. The association of BD PRS was significantly stronger than the association of any random PRS in this analysis (Table 2), which confirmed that the FAM_{BD} cases carried an exceptionally high load of common risk variants for BD.

Unaffected family members showed higher psychiatric PRS than CC controls

In the comparison of $FAM_{unaffected}$ to $CC_{controls}$, PRS for BD, SCZ, MDD, and the *Shared* PRS were significantly higher in unaffected family members (Fig. 1E-F, Supplementary Figs. S1 and S4, Supplementary Table S4). $FAM_{unaffected}$ individuals showed no increase in BD/MDD-specific GWIS, LOAD, or random PRS. However, their SCZ-BD GWIS PRS were higher. The associations of the psychiatric PRS were significantly stronger than the association of random PRS (Table 2). $FAM_{unaffected}$ individuals, therefore, had a substantial risk load for all three psychiatric disorders and did not show a specific increase in risk for BD.

FAM_{BD} cases had an increased PRS specifically for BD

The preceding analyses have established that all family members showed an increased burden of common risk variants for BD, MDD, and SCZ and that FAM_{BD} cases appeared to carry a particularly high load of BD-associated variants.

To investigate FAM_{BD} cases further, we compared them directly to $FAM_{unaffected}$. Here, only the BD PRS was significantly higher in FAM_{BD} (Fig. 2A-B, Supplementary Figs. S1 and S5, Supplementary Table S5).

Notably, while the BD-specific GWIS PRS were significantly higher, the *Shared* PRS were not. These results corroborate the observation that, while all family members had an increased load of risk variants for BD, SCZ, and MDD, the FAM_{BD} cases showed an especially high burden of variants that conferred a specific risk for BD.

Effects of assortative mating on BD PRS in family members

The high burden of psychiatric risk in the ABiF families may be attributable to assortative mating. To investigate this hypothesis, we analysed the married-in family members without a parent within the FAM sample. Eight of the 44 individuals who had married into the families had a diagnosis of MDD and none of BD (Table 1). While the unaffected married-in individuals had higher BD PRS than $CC_{controls}$, their BD PRS was not higher than the PRS of other $FAM_{unaffected}$ (Fig 2C, Supplementary Fig. S6, Supplementary Table S6). We also examined possible anticipation of BD in the families: neither did the BD PRS increase significantly over generations nor did the age at onset decrease over time (Fig. 2D, Supplementary Fig. S7, Supplementary Table S7).

FAM_{MDD} cases had higher psychiatric PRS than $CC_{controls}$

Comparisons of the FAM_{MDD} cases to $CC_{controls}$ demonstrated that FAM_{MDD} had significantly higher BD, MDD, and *Shared* PRS (Supplementary Figs. S8-S9, Supplementary Table S8). However, none of these PRS were significantly increased when comparing FAM_{MDD} to $FAM_{unaffected}$ (Supplementary Fig. S10, Supplementary Table S9). Notably, in both comparisons, FAM_{MDD} showed an increase in SCZ-MDD GWIS PRS but not in SCZ-BD GWIS PRS.

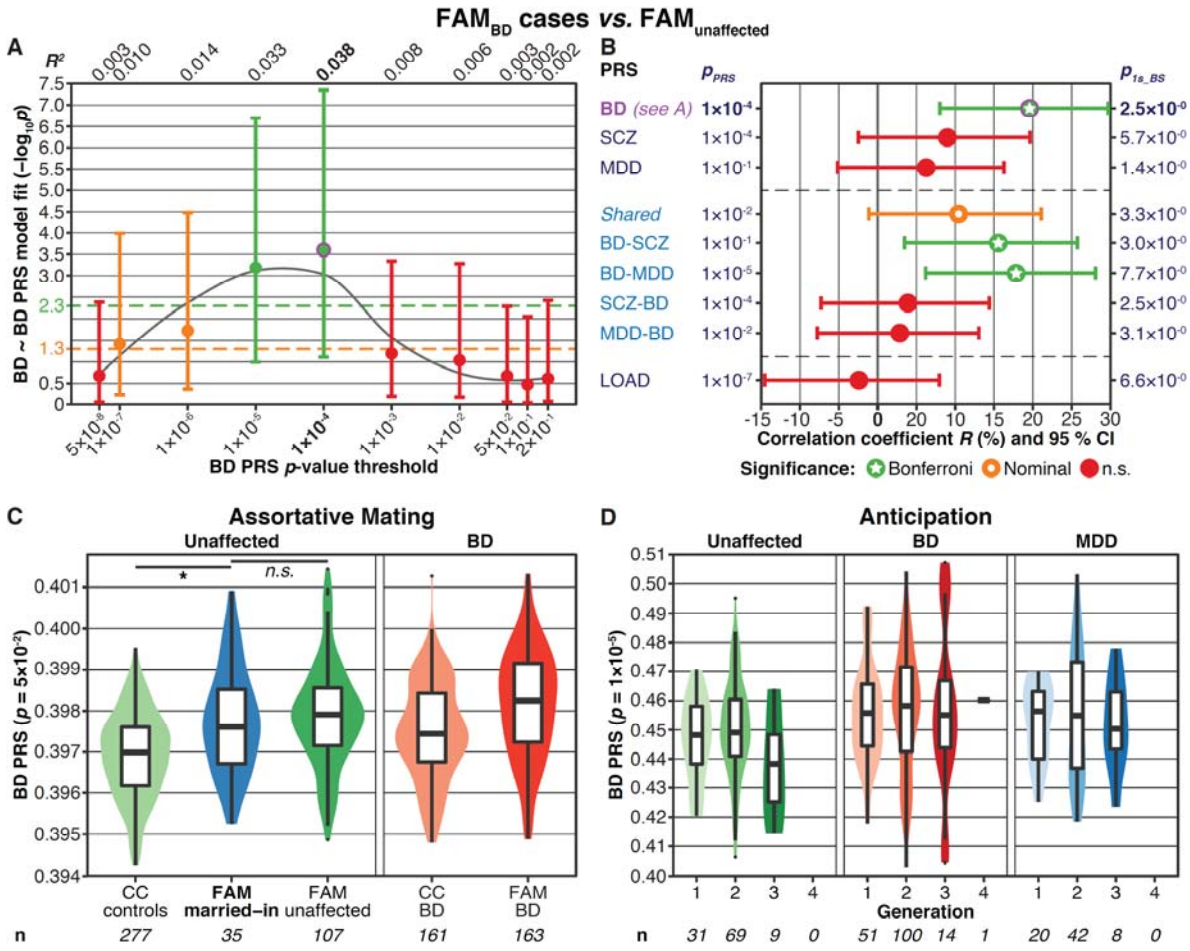

Figure 2: A-B: Association analyses comparing different PRS in FAM_{BD} cases to $FAM_{unaffected}$. The plots show one-sided *p*-values, following the hypothesis that BD cases have higher PRS than unaffected individuals. Further details of the plots are as described in the legend for Fig. 1. See Supplementary Fig. S5 and Table S5 for more detailed plots and full association test statistics. **C-D:** Analyses of assortative mating (**C**) and anticipation (**D**). These plots were not adjusted for covariates; n = sample size. The y-axis shows the PRS values. **C:** Assortative mating. The plot shows violin- and boxplots of the BD PRS (p_{PRS}=0.05), comparing unaffected, married-in individuals with no parent among the ABiF families to other FAM and CC subjects. At p_{PRS}=0.05, married-in family members showed the highest BD PRS compared to $CC_{controls}$ (p=1.1×10^{-5}, Supplementary Fig. S6A and Table S6). The BD PRS of married-in individuals was not significantly higher than the PRS of $FAM_{unaffected}$ at any p_{PRS} (p≥0.186, Supplementary Fig. S6B and Table S6). Covariate used: gender. One-sided *p*-values were calculated, following the hypothesis that married-in individuals have higher PRS than other unaffected subjects. **D:** Anticipation: the BD PRS did not increase across generations. The plot shows violin- and boxplots of the BD PRS (p_{PRS}=1×10^{-5}) across different generations of the FAM sample for the three diagnosis groups. At p_{PRS}=1×10^{-5}, the association of the BD PRS with generation was strongest (p=0.45, Supplementary Fig. S7A and Table S7). Married-in family members were excluded from this analysis. Covariates used: gender, age at the interview, diagnostic group. One-sided *p*-values were calculated, following the hypothesis that the PRS increase across generations.

Discussion

The main aim of the present study was to investigate whether common genetic variants contribute to the occurrence of BD in 33 BD multiplex families from Andalusia, Spain. For this purpose, PRS were calculated for the family members and compared to PRS of independent controls from the Spanish population and sporadic BD cases. In comparison to $CC_{controls}$, FAM_{BD} cases had significantly increased PRS for the three psychiatric disorders BD, SCZ, and MDD, higher BD+SCZ+MDD *Shared* PRS, and higher BD-specific PRS. Unaffected family members had, compared to $CC_{controls}$, significantly increased PRS for BD, SCZ, and MDD, and higher *Shared* PRS. When comparing FAM_{BD} to sporadic CC_{BD} cases, only the BD PRS were significantly higher in the FAM sample. Also within the ABiF families, only the BD PRS were higher in FAM_{BD} cases than in $FAM_{unaffected}$. These findings indicate that, compared to non-familial samples from Spain, the family members had a strongly elevated baseline risk for the genetically correlated psychiatric disorders BD, MDD, and SCZ. While this genetic burden confers a higher risk for mental illness in general, FAM_{BD} cases were characterized by a particularly high number of BD-specific risk variants.

Although both the FAM and CC samples were recruited in Spain[9], minor population differences may have influenced the present results. Even if such minor differences existed, it is unlikely that they caused the highly significant associations observed for the psychiatric PRS, given that the pairwise genetic relationship matrix was used as random effects in the association analyses. Nevertheless, three analyses were conducted to confirm that systematic differences between the genotype data of FAM, $CC_{controls}$, and CC_{BD} samples did not distort our findings: First, we did not find significant differences between the cohorts in a population substructure analysis (see Supplementary Fig. S11 and Supplementary Methods). Second, PRS for LOAD were not significantly increased in family members in any analysis. Since LOAD shows no genetic correlation with BD, MDD, or SCZ[14,36,37], this adds further support to the specificity of our analyses. Third, we generated random PRS and conducted permutation tests to determine how often random PRS showed the same or a more extreme association strength than PRS for psychiatric disorders. These analyses confirmed that when a psychiatric disorder PRS was significantly increased in family members, this association was stronger than for random PRS. We thus conclude that the high psychiatric PRS observed in family members compared to controls cannot be attributed to population or technical differences between the cohorts.

While the ABiF families are characterized by a high prevalence of BD and MDD, the family members showed increased SCZ PRS compared to controls. This increase could be an indirect consequence of the genetic correlation between BD and SCZ[14,16,18-20]. Interestingly, the majority of FAM_{BD} cases were diagnosed with BD-I (115 of 166 cases), which is the BD subtype with the highest genetic correlation to SCZ[12,38]. Furthermore, family members also had higher *Shared* PRS than $CC_{controls}$. SCZ shows a higher heritability and lower prevalence than MDD does, and the number of cases in the GWAS used for calculating the SCZ PRS was higher than the number of cases in the BD GWAS. The SCZ GWAS is therefore likely to have had higher statistical power than the GWAS of BD and MDD[39]. In consequence, the SCZ PRS may have included more cross-disorder signals with smaller effects than the PRS of BD and MDD. If family members had an increased *Shared* risk burden, this cross-disorder risk may have rendered them vulnerable to psychiatric disorders in general, with the high BD PRS then shaping the final BD diagnosis outcome. Notably, none of the family members included in the present analyses has been diagnosed with SCZ. Given that the recruitment strategy focused on BD multiplex families, this is likely to be attributable to ascertainment bias. Of note, the analyses of

FAM_{MDD} cases are discussed in the Supplementary Data.

The BD PRS of FAM_{BD} cases were significantly higher than those of $CC_{controls}$ across all p_{PRS} thresholds (Fig. 1A). Similarly, the BD PRS of CC_{BD} cases were higher than in $CC_{controls}$ across all thresholds (Supplementary Fig. S12, Supplementary Table S10). By comparison, the increase of the BD PRS in FAM_{BD} cases relative to $FAM_{unaffected}$ was less pronounced, albeit significant at two p_{PRS} thresholds (Fig. 2A). Depending on the genetic architecture of the disorder in question and the statistical power of GWAS and PRS analyses, PRS may be expected to differ less in a family-based cohort. In a previous study of BD multiplex families[40], in which the PRS was calculated using a different GWAS[10], the BD PRS was not significantly higher in cases compared to unaffected family members. In contrast, another study found that PRS for LOAD significantly distinguished between Alzheimer's disease cases and controls in familial cohorts[41].

Our analyses showed that ABiF family members carried high BD PRS compared to the average population. Although individuals who married into these families also had higher BD PRS than $CC_{controls}$, their BD risk load was similar to other $FAM_{unaffected}$. At the time of the interview, none of the married-in family members had a diagnosis of BD. Nevertheless, their increased BD PRS indicate that weak assortative mating may have occurred. Unaffected individuals with an above average BD PRS may display sub-threshold characteristics of BD, such as a broader range of emotions[42-44]. Consistent with the observation that married-in subjects did not have higher BD PRS than the other $FAM_{unaffected}$, no increase in BD PRS was found across generations. However, assortative mating may have contributed to the establishment and maintenance of a high genetic risk load for BD in these families. Furthermore, assortative mating may have already occurred in previous generations, for which no DNA was available. Of note, DNA was neither available for all ABiF family members of the current generations, limiting the scope of the analysis of assortative mating.

The present study generated substantial evidence that members of the ABiF families, including unaffected subjects, carried a higher risk burden of common genetic risk variants for the psychiatric disorders BD, SCZ, and MDD than the average population. A plausible hypothesis is that this polygenic load of common risk variants is a major contributor to the high incidence of BD and MDD in these families. Although a previous investigation of a single ABiF pedigree already described a high BD PRS in cases, no rare causal variants were identified[25]. Research suggests that for high prevalence complex diseases, even cases accumulated within families may typically be influenced by a polygenic risk burden[45]. However, given that the PRS explained only a fraction of the phenotypic variance (up to an R^2 of 17.4% in the present analyses), rare mutations likely also played an important role in each of the families. Sequencing studies carried out in multiplex families have suggested a function of rare variants in the aetiology of BD46-48. To date, however, it has proven difficult to identify replicable causal associations between rare variants and BD susceptibility. Of note, we analysed genotype data from the 33 ABiF families as an average across families and not separately per family. Thus, the degree to which common and rare variants shaped the emergence of psychiatric disorders may vary between families. To further enhance our understanding regarding the aetiology of BD, future studies should aim at identifying such rare variants in single ABiF families and determine the contribution of rare variants to disease development in comparison to the impact of common variants.

Supplementary information is available at bioRxiv.org.

Acknowledgments

The study was supported by the German Federal Ministry of Education and Research (BMBF), through the Integrated Network IntegraMent, under the auspices of the e:Med programme (grants 01ZX1314A to MMN and SC; 01ZX1314G to MR; 01ZX1614J to BMM), by the German Research Foundation (DFG grants FOR2107; RI908/11-1 to MR; NO246/10-1 to MMN; MU1315/8-2 to BMM), and by the Swiss National Science Foundation (SNSF grant 156791 to SC). MMN is a member of the DFG-funded cluster of excellence ImmunoSensation. The PGC has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320). We thank the research participants and employees of 23andMe, Inc. for their contribution to the MDD meta-analysis published in (14). We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for the present analyses. See the Supplementary Data for extended Acknowledgments.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

1 Merikangas KR, Jin R, He J-P, Kessler RC, Lee S, Sampson NA *et al.* Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry* 2011; **68**: 241–251.
2 Craddock N, Sklar P. Genetics of bipolar disorder. *Lancet* 2013; **381**: 1654–1662.
3 Edvardsen J, Torgersen S, Røysamb E, Lygren S, Skre I, Onstad S *et al.* Heritability of bipolar spectrum disorders. Unity or heterogeneity? *J Affect Disord* 2008; **106**: 229–240.
4 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.
5 Charney AW, Ruderfer DM, Stahl EA, Moran JL, Chambert K, Belliveau RA *et al.* Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. *Transl Psychiatry* 2017; **7**: e993.
6 Hou L, Bergen SE, Akula N, Song J, Hultman CM, Landén M *et al.* Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Human Molecular Genetics* 2016; **25**: 3383–3394.
7 Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol Psychiatry* 2017; **511**: 421.
8 Lee MTM, Chen C-H, Lee CS, Chen CC, Chong MY, Ouyang WC *et al.* Genome-wide association study of bipolar I disorder in the Han Chinese population. *Mol Psychiatry* 2011; **16**: 548–556.
9 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. *Nat Commun* 2014; **5**: 3339.
10 Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics* 2011; **43**: 977–983.
11 Cichon S, Mühleisen TW, Degenhardt FA, Mattheisen M, Miró X, Strohmaier J *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet* 2011; **88**: 372–381.
12 Stahl E, Breen G, Forstner A, McQuillin A, Ripke S, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium *et al.* Genomewide association study identifies 30 loci associated with bipolar disorder. 2018. doi:10.1101/173062.
13 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.

14 Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* 2018; **50**: 668–681.
15 Song J, Bergen SE, Kuja-Halkola R, Larsson H, Landén M, Lichtenstein P. Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. *Bipolar Disord* 2015; **17**: 184–193.
16 Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of Psychiatric Genomics Consortium *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* 2014; **19**: 1017–1024.
17 Forstner AJ, Hecker J, Hofmann A, Maaser A, Reinbold CS, Mühleisen TW *et al.* Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. *PLoS ONE* 2017; **12**: e0171595.
18 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.
19 Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 2013; **45**: 984–994.
20 Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu, Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* 2018; **173**: 1705–1715.e16.
21 McClellan J, King M-C. Genomic analysis of mental illness: a changing landscape. *JAMA* 2010; **303**: 2523–2524.
22 Kerner B. Toward a Deeper Understanding of the Genetics of Bipolar Disorder. *Front Psychiatry* 2015; **6**: 105.
23 Guzman-Parra J, Rivas F, Strohmaier J, Forstner A, Streit F, Auburger G *et al.* The Andalusian Bipolar Family (ABiF) Study: Protocol and sample description. *Rev Psiquiatr Salud Ment* 2017. doi:10.1016/j.rpsm.2017.03.004.
24 Schumacher J, Kaneva R, Jamra RA, Díaz GO, Ohlraun S, Milanova V *et al.* Genomewide scan and fine-mapping linkage studies in four European samples with bipolar affective disorder suggest a new susceptibility locus on chromosome 1p35-p36 and provides further evidence of loci on chromosome 4q31 and 6q24. *Am J Hum Genet* 2005; **77**: 1102–1111.
25 Collins AL, Kim Y, Szatkiewicz JP, Bloom RJ, Hilliard CE, Quackenbush CR *et al.* Identifying bipolar disorder susceptibility loci in a densely affected pedigree. *Mol Psychiatry* 2013; **18**: 1245–1246.
26 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
27 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
28 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* 2012; **44**: 955–959.
29 Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* 2013; **10**: 5–6.
30 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2017.
31 Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* 2013; **45**: 1452–1458.
32 Nieuwboer HA, Pool R, Dolan CV, Boomsma DI, Nivard MG. GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. *Am J Hum Genet* 2016; **99**: 917–927.
33 Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. *PLoS ONE* 2013; **8**: e65395.

34 Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions. Comprehensive R Archive Network (CRAN), 2017.
35 Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.
36 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015; **47**: 1236–1241.
37 Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J *et al.* Analysis of shared heritability in common disorders of the brain. *Science* 2018; **360**: eaap8757.
38 Song J, Kuja-Halkola R, Sjölander A, Bergen SE, Larsson H, Landén M *et al.* Specificity in Etiology of Subtypes of Bipolar Disorder: Evidence From a Swedish Population-Based Family Study. *Biol Psychiatry* 2017. doi:10.1016/j.biopsych.2017.11.014.
39 Levinson DF, Mostafavi S, Milaneschi Y, Rivera M, Ripke S, Wray NR *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biol Psychiatry* 2014; **76**: 510–512.
40 Fullerton JM, Koller DL, Edenberg HJ, Foroud T, Liu H, Glowinski AL *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 Neuropsychiatr Genet* 2015; **168**: 617–629.
41 Tosto G, Bird TD, Tsuang D, Bennett DA, Boeve BF, Cruchaga C *et al.* Polygenic risk scores in familial Alzheimer disease. *Neurology* 2017; **88**: 1180–1186.
42 Peyrot WJ, Robinson MR, Penninx BWJH, Wray NR. Exploring Boundaries for the Genetic Consequences of Assortative Mating for Psychiatric Traits. *JAMA Psychiatry* 2016; **73**: 1189–1195.
43 Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry* 2016; **73**: 354–361.
44 Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM *et al.* Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci* 2015; **18**: 953–955.
45 Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. *European Journal of Human Genetics* 2010; **18**: 1039–1043.
46 Goes FS, Pirooznia M, Parla JS, Kramer M, Ghiban E, Mavruk S *et al.* Exome Sequencing of Familial Bipolar Disorder. *JAMA Psychiatry* 2016; **73**: 590–597.
47 Cruceanu C, Schmouth J-F, Torres-Platas SG, Lopez JP, Ambalavanan A, Darcq E *et al.* Rare susceptibility variants for bipolar disorder suggest a role for G protein-coupled receptors. *Mol Psychiatry* 2017; **3**: 575, 577.
48 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.

Supplementary Information

Supplementary Material:

PDF document containing the Supplementary Methods, Discussion, References, Acknowledgments, and Figures S1-S12, legends for the Supplementary Tables and the full list of PGC authors. A detailed table of contents is provided at the beginning of the document.

Supplementary Tables:

Excel worksheet containing the Supplementary Tables S1-S10. Detailed legends for the Supplementary Tables are provided in the Supplementary Material PDF document. The Supplementary Tables can be accessed under goo.gl/3u1C66.