Environmental factors contribute to risk of bipolar disorder (BD), but how environmental factors impact the development of psychopathology within the context of elevated genetic risk is unknown. We herein sought to identify epigenetic signatures operating in the context of polygenic risk for BD in young people at high familial risk (HR) of BD. Peripheral blood-derived DNA was assayed using Illumina PsychArray, and Methylation-450K or -EPIC BeadChips. Polygenic risk scores (PRS) were calculated using summary statistics from recent genome-wide association studies for BD, major depressive disorder (MDD) and cross-disorder (meta-analysis of eight psychiatric disorders). Unrelated HR participants of European ancestry (n = 103) were stratified based on their BD-PRS score within the HR-population distribution, and the top two quintiles (High-BD-PRS; n = 41) compared against the bottom two quintiles (Low-BD-PRS; n = 41). The High-BD-PRS stratum also had higher mean cross-disorder-PRS and MDD-PRS (ANCOVA p = 0.035 and p = 0.024, respectively). We evaluated DNA methylation differences between High-BD-PRS and Low-BD-PRS strata using linear models. One differentially methylated probe (DMP) (cg00933603; p = 3.54 × 10^{-7}) in VAR52, a mitochondrial aminoaacyl-tRNA synthetase, remained significantly hypomethylated after multiple-testing correction. Overall, BD-PRS appeared to broadly impact epigenetic processes, with 1,183 genes mapped to nominal DMPs (p < 0.05); these displayed convergence with genes previously associated with BD, schizophrenia, chronotype, and risk taking. We tested poly-methylomic epigenetic profiles derived from nominal DMPs in two independent samples (n = 54 and n = 82, respectively), and conducted an exploratory evaluation of the effects of family environment, indexing cohesion and flexibility. This study highlights an important interplay between heritable risk and epigenetic factors, which warrant further exploration.

INTRODUCTION

Bipolar disorder (BD) is a highly heritable mental illness, for which genetic factors explain ~60–85% of risk variance [1]; the remaining variance is explained by non-genetic factors, including environmental contributors. First-degree relatives of probands with BD have 5–10-fold increased risk of developing BD themselves [1–3], and are at increased risk of broader psychopathology [4, 5], including major depression, anxiety, behavioural and substance use disorders; therefore, young first-degree relatives of those with BD are considered high risk (HR) for later mental illness. The elucidation of clinical antecedents of the BD prodrome is an active research area [6–10], but specific precursors are heterogeneous, and biomarkers of risk trajectories are a research priority.

Genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) associated with increased disease risk, each with small individual effect [11–13]. Many disease-associated SNPs are shared amongst psychiatric disorders, with substantial genetic correlation between BD and schizophrenia, and BD and major depressive disorder (MDD) [13–17]. GWAS have demonstrated the polygenic nature of BD, where common SNPs identified to date collectively account for ~25% of the estimated heritability [12–15]. Indeed, individuals with polygenic risk scores (PRS) at the extremes of a population distribution have substantially altered risk of developing a psychiatric disorder—those in the top 10% of the BD-PRS distribution have an odds ratio of 9.3 of developing BD compared to the lowest decile [13]—yet the predictive capacity of PRS is currently limited [18], with inadequate sensitivity and specificity as PRS will only capture part of the genetic contribution [19]. SNP-based heritability falls short of heritability estimates from family, twin and epidemiologic studies; [20] the latter approaches typically employ models that include additive genetic and common/unique environmental factors (and their interaction), whereas SNP-based heritability estimates typically only model additive genetic variance [20].
Moreover, some of the “missing heritability” may also reside in rare variants not captured by GWAS [21–23], and/or relate to other familial factors that are not encoded in nucleotide-level DNA sequence variation.

Family history is the strongest current predictor of future BD [7] and is sometimes posited as a proxy for genetic transmission, but likely reflects dynamic gene-environment interplay from preconception through the life span. Twin studies also indicate environmental contribution to BD [24]. Several environmental risk factors have been posited—including prenatal infections, childhood maltreatment, and psychological stress [25]—but the attributable impact of such factors is small and not disease-specific. Families that include a parent with BD have lower parent-reported cohesion compared to families with no parental psychiatric disorders [26], and while offspring-centred reports are less common, child reports of lower cohesion and adaptability, and higher conflict environments are also noted [27]; these factors potentially confound family history with environmental elements. Furthermore, children who experienced early maternal loss have a four-fold increased risk of BD [28], suggesting that early trauma and altered home-environment dynamics may influence disease trajectory. Indeed, children are influenced by both genes and environment provided by parents, which cannot be easily disambiguated to apportion causation [29]. Moreover, environmental effects are likely conditional on genetic factors, which are only partially appraised by existing gene-environment studies [30]. Also, genetic nurture may include indirect genetic effects from parental genes that are not transmitted [31]. Thus, there is increasing interest in the role of specific gene-environment interactions, as well as the mechanisms by which risk factors interact in the development of BD [32].

Environmental factors can potentially impact gene expression through epigenetic modulations [33]. One of the most studied epigenetic processes is DNA methylation; the addition of a methyl group to the \( \text{S} \) cytosine of a cytosine-guanine sequence (CpG) [33]. To date, there are no large-scale epigenome-wide studies in BD [34], although several candidate-gene epigenetic studies [35, 36] and pharmaco-epigenomic studies of antipsychotic medicines [37] have been performed. Furthermore, GWAS signals for schizophrenia—a condition that shares genetic overlap with BD [17, 38]—are enriched in human-specific methylated regions [39], implying mechanistic overlap between genetic and epigenetic risk. Interestingly, a recent epigenetic element-based transcriptome-wide association study identified genes that contribute to BD heritability beyond those explained by GWAS-associated SNPs [40], suggesting that epigenetic regulation may further contribute to heritability. Epigenetic factors may even contribute to transgenerational genomic regulation [41], although this mechanism of genomic transmission in humans remains controversial [42].

To identify biomarkers for illness-onset, a prospective longitudinal approach is required. One small prospective study that compared HR participants who developed BD or MDD \((n = 22)\) to those who remained well \((n = 23)\) identified 22,543 nominally differentially methylated CpGs \(p < 0.05\) [43], although no probe passed epigenome-wide correction for multiple testing. To date, the characterisation of methylation differences relating to polygenic burden in individuals with familial risk for BD has not been performed. Thus, we undertook the identification of differentially methylated positions (DMPs) in young HR participants, stratified by genetic burden of common BD-associated risk alleles, to characterise epigenetic modifications operating beyond polygenic additive risk. Methylation signatures derived from the cumulative effect of multiple CpGs were then validated in independent samples of BD cases, HR, and controls. Finally, the impact of family environment measures on methylation signatures were explored, to further understand relationships between familial and environmental risk factors associated with the development of psychopathology.

### MATERIALS AND METHODS

An overview of the approach is summarised in Fig. 1.

#### Study participants

Australian participants were aged 12–30 years and recruited as previously described [44, 45]. Briefly, HR and BD participants were recruited from families who had previously participated in BD family studies, specialised BD research clinics, mental health consumer organisations, or response to public notices. BD cases met DSM-IV criteria for BD-type I (BD-I) or type-II (BD-II). HR participants were the children or siblings of individuals with DSM-IV diagnoses of BD-I, BD-II, or schizoaffective disorder–bipolar-type (SABP), and did not personally have threshold diagnoses of these conditions at baseline (participants with sub-threshold BD “not otherwise specified” (BD-NOS) were not excluded). Control participants (CON) were recruited via print/electronic media, and noticeboards in universities and local communities. CON had no personal or familial (first-degree) history of BD-I, BD-II, recurrent unipolar disorder, SABP, schizophrenia, recurrent substance abuse or psychiatric hospitalisation, and no second-degree relative with a past mood-disorder hospitalisation or history of psychosis. Written informed consent was obtained from all participants, with additional parental consent for participants aged <16 years. This study was approved by University of New South Wales Human Research Ethics Committee (HREC Protocol 09/097).

High-risk participants from the United States were aged 12–21 years, and recruited from Indiana University, University of Michigan, Johns Hopkins University, and Washington University in St. Louis (with site-specific IRB approval), as previously described [46]. The US sites used recruitment criteria and clinical assessments identical to those used at the Australian site.

#### Clinical assessments

Structured interviews were administered by staff with extensive clinical background and after specific training in each instrument, which comprised a battery of structured clinical interviews, self-report questionnaires and clinician rated assessments. The Family Interview for Genetic Studies (FIGS) [47] was administered to each participant or a member of their family to determine family history of mood or psychotic disorder. To determine psychiatric diagnoses—diagnoses for participants aged 12–21, both parents and their parent/s completed the Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Aged Children–Present and Lifetime Version (K-SADS-BP; v2, July 2009) [46, 48]. To determine psychiatric diagnoses for participants aged 22–30 (including BD-probands), the Diagnostic Interview for Genetic Studies (DIGS; v4.0/BD, July 2005) [49] was administered. Consensus DSM-IV diagnoses were determined by two clinicians (i.e., psychiatrists with child specialty training, clinical psychologists, or clinical social workers) using best-estimate methodology [50], using the K-SADS or DIGS, FIGS, and medical records (where available). Functional capacity was determined via Global Assessment of Functioning (GAF) [51] or Clinical Global Impression (CGI) [52] rating scales.

One to four follow-up clinical interviews were performed on HR and control participants to identify emergent psychopathology. Diagnoses were assigned a confidence rating on a 4-point scale; only those that met full DSM-IV criteria received a confidence rating ≥3. As previously described by Frankland et al. [8], participants with a best-estimate diagnosis of BD were categorised as either threshold converters (i.e. with BD-I or BD-II; confidence level 3–4) or sub-threshold converters (i.e. with BD-NOS; confidence level of 1–2); the latter diagnosis was made when participants did not meet the minimum 4 day duration criterion, but otherwise met all full symptom criteria for hypomania. Participants (HR or control) reaching threshold or sub-threshold criteria at follow-up assessment for diagnosis of BD-I, BD-II, BD-NOS, with a confidence of ≥2 were considered “BD-syndromic”, and those with no clinical diagnoses after follow-up were considered “disorder-free”.

#### Genotyping and polygenic risk scores

Peripheral blood samples were collected, DNA extracted, and genotyping performed on PsychArray-24 BeadChip as previously described [54] (details in Supplementary methods). An independent sample of BD cases \(n = 264\); described in [13] as “neuc1”) and controls \(n = 1115\) [55] were employed.
to infer optimal $p$ value threshold for generation of PRS ($p_T$; Supplementary Methods, Fig. S1).

Polygenic risk scores (PRS) were generated using summary statistics from the PGC GWAS for BD ($n = 31,358$ controls and $n = 20,352$ cases; excluding 6,201 participants from GAIN, BMAU, FAT2, MICH datasets which contained parents of some of the study participants) [12]. Data for MDD ($n = 561,190$ controls and $n = 246,362$ cases) [56] and cross-disorder ($n = 494,162$ controls and $n = 232,964$ cases with anorexia nervosa, attention-deficit/hyperactivity disorder, autism spectrum disorder, BD, MDD, obsessive-compulsive disorder, schizophrenia, and Tourette syndrome) [17] were also used. PRS were calculated in a single analysis for all participants simultaneously using PRSice v2.2.13 [57], employing unrelated individuals (identity-by-descent $\pi_{\text{hat}} < 0.1$; Supplementary Methods) identified as European-ancestry by multidimensional scaling analysis (Supplementary Methods), and comparing BD cases to controls to define optimal $p_T$ to explain the largest variance (Nagelkerke pseudo-$R^2$); phenotypes for HR were coded unknown/missing (-9). Polygenic scoring employed linkage disequilibrium clumping, removal of strand-ambiguous SNPs, and SNPs with MAF < 0.05 or low imputation quality (INFO < 0.8), prior to PRS computation and use in subsequent analyses.

Demographic comparisons
Group differences were examined using univariate general linear ANCOVA models, using SPSS Statistics for Windows v26 (IBM Corp., Armonk, NY). For PRS group level comparisons in the extended cohort, generalised estimation equations (GEE) were employed to account for relationships within families which contained multiple relatives, and included genotype-derived MDS components (C1 and 2) as covariates.

Methylation analysis
Epigenome-wide methylation profiling. Methylation quantification was performed on DNA derived from peripheral blood in four batches: batches 1 and 2 (both $n = 96$) employed the Illumina HumanMethylation450K array, and batches 3 and 4 ($n = 127$ and 96, respectively) employed the Illumina MethylEPIC BeadChip. The R package methyl [58] was used to extract signal intensities and initial quality control (QC) within each batch (Supplementary Material). The R package Shinymethyl [59] was used to conduct principal component analysis and visualise outliers within each batch. After removal of 11 technical replicates and 20 samples failing QC procedures, 384 participants remained (round 1 $n = 88$, round 2 $n = 96$, round 3 $n = 118$, round 4 $n = 82$) – 425,453 CpG probes that passed QC and were represented on both 450k and EPIC chips then underwent further filtering for EWAS as outlined below. To avoid potential bias in PRS due to ethnicity [60], the cohort was restricted to unrelated European-ancestry participants (Fig. S2), which were divided into Discovery and Validation sets (described below and in Fig. 1). Normalisation of methylation $\beta$ values employed meffil, including technical covariates (sentrix position and sentrix group; the latter unique across batches) [58]. Normalisation of batches 1–3 was performed simultaneously across all three batches, and formed the Discovery Set and Validation Set 1, whereas batch 4 was normalised separately as Validation Set 2.

Tissue source, smoking status and cell count estimates. DNA was derived from three peripheral blood sources: whole blood, ficoll/buffy-coat, and isolated lymphoblasts (details in Supplementary Material). To account for inter-individual differences in cellular proportions, estimation of six cell types (B cells, CD4T, CD8T, granulocytes, monocytes and natural killer) was performed using meffil’s gse35069 profile [58]. Tissue source and cell count estimates (Fig. S3) were included as covariates, as described below.

Lifetime tobacco use collected via KSADS was available for ~60% of participants in this study, and current smoking status was available for ~16% ($n = 36$ out of 218 individuals; $n = 5$ current users, all in Validation Set 2). To predict smoking status in all participants, normalised $\beta$ values at cg00475921 were used (Fig. S4), where $\beta < 0.75$ were classified as probable smokers [61].

Discovery set: HR youth stratified by BD-PRS. To stratify HR participants on the basis of their personal burden of BD-associated SNP variants, the distribution of BD-PRS within the HR sample was divided into quintiles; this stratification was undertaken on the basis that individuals with PRS in the extremes of a population distribution have substantially altered risk of...
developing a psychiatric disorder [13]. Quintiles were chosen to balance maximisation of the available sample, with exclusion of individuals with intermediate scores reflecting ‘average’ SNP burden. Thus two comparator groups were defined: the top two quintiles (40% of participants) formed the High-BD-PRS group (n = 41) and the bottom two quintiles (40% of participants) formed the Low-BD-PRS group (n = 41) (Table S1, Fig. 2).

**Discovery EWAS: high vs. low-BD-PRS within HR youth**

CpG selection: Probes were restricted to those which were: (1) variable in our data (n = 249,436)—defined by >5% variability in the difference between the 10th to 90th percentile range of β values across batches 1–3 [62], and (2) reported to be variable in blood (n = 244,724), as defined by Edgar et al. [63], and 3) correlated between blood and brain (n = 42,364; with mean r ≥ ±0.3 across prefrontal cortex, superior temporal gyrus and entorhinal cortex) as defined by Hannon et al. [62]. Restricting the discovery EWAS to variable blood-brain correlated probes had benefits in: (1) reducing the EWAS search space, and (2) focusing on CpGs that are likely to be functionally relevant in the primary disease-affected tissue. Importantly, ‘probe SNPs’ as annotated in the Illumina manifest file were removed prior to EWAS, to exclude direct effects of sequence variation out to 50 bp (n = 6,455). The final probe set included 35,907 CpGs, and β values for each probe were transformed into M-values using the ‘beta2m’ function in R package lumi [64].

Surrogate variable analyses: To identify and account for residual unmeasured variation that was not corrected by normalisation of technical covariates (sentrix position/chip/batch), the ‘be’ method in R package SVA [65] was used to identify surrogate variables considering BD-PRS stratum as the variable of interest (Low-BD-PRS = 0, High-BD-PRS = 1). Four significant surrogate variables (P < 0.05) were identified and included in downstream analysis.

**EWAS: DMPs** were identified using the R package limma [66], employing a linear regression model, with normalised M-values as outcome and BD-PRS stratum as predictor (Low-BD-PRS = 0, High-BD-PRS = 1). Covariates included sex, age, two ethnicity components (MDS C1 and C2; Supplementary Material), tissue source, blood-cell count estimates and four surrogate variables.

**Bias correction:** The R package bacon [67] was applied to correct test-statistic inflation, to minimise shift in distribution of effect sizes and risk of false-positive findings. Empirical null estimates were generated using a Gibbs Sampling algorithm in a Bayesian framework, based on 5000 iterations with burn-in period of 2000. DMPs whose inflation- and bias-corrected p values exceeded p < 1.39 × 10−6 (Bonferroni adjustment for 35,907 tests at α = 0.05) were considered epigenome-wide significant.

**Pathway analyses**

DMPs at two p value thresholds (the Q–Q plot inflection point of p < 0.002, and p < 0.05; uncorrected) were mapped to genes using two complementary methods: (1) physical mapping of probes based on UCSC hg19 base-pair coordinates (as per Illumina manifest file), and (2) functional mapping of cis-regulatory regions using the Genomic Regions Enrichment of Annotations Tool (GREAT; v4.0.4) [68]. GREAT enables functional mapping of DMPs in non-coding regulatory regions (i.e. outside coding-sequence boundaries)—we employed ‘basal plus extension’ mode that includes genes 5 kb upstream, 1 kb downstream, plus 5 kb distal binding. The combined output of both methods created a gene list for pathway interrogation.

Finally, Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) [69] was employed with default parameters (i.e. not excluding MHC region) to examine enrichment of genes harbouring DMPs in biological function and pathway categories, and enrichment of gene expression across tissues [70]. Enrichments were considered significant if false discovery rate (FDR) q ≤ 0.05.

**Supplementary EWAS: post-hoc sensitivity analysis for interpretation of DMP functional enrichments**

As the discovery EWAS employed a restricted set of 35,907 CpG probes that were blood-brain correlated [62], we reasoned that probe pre-selection might bias functional enrichments in genes mapped to DMPs. Therefore, a supplementary EWAS was conducted using 214,352 probes that passed QC and were variable in blood [63] and variable in our data (but without applying the blood-brain correlation filter), employing identical procedures as the primary EWAS. The multiple-testing correction threshold was adjusted to $p < 2.33 \times 10^{-7}$ for 214,352 tests at α = 0.05 [71].

**Validation using poly-methylomic profile scores**

A ‘poly-methylomic profile score’ (PMPS) is a quantitative metric reflecting the degree of methylation at multiple sites across the epigenome (similar to calculating a PRS from GWAS summary statistics for genotypic data), as previously described [72]. A PMPS was calculated for each participant using effect sizes and p-values for DMPs from the discovery PRS-stratified EWAS, and applied to two independent validation sets (Fig. 1), as described below. Normality was assessed via Shapiro-Wilk tests in SPSS, and effect sizes reported as partial eta squared ($\eta^2_p$).

**Validation Set 1**

To determine whether PMPS provided a replicable index of BD-PRS, we employed 67 control participants (European-ancestry, unrelated ($\pi_hat < 1699$ controls, 354 HR, 355 BD; $M = 0.0077695, SD = 7.543 \times 10^{-5}$)). The boxes and whiskers inside the violins indicate the 25–75th percentiles, and 1.5 times the interquartile range, respectively. The dashed rectangles indicate the stratification of HR participants into High-BD-PRS vs. Low-BD-PRS strata, as defined by the top and bottom two quintiles (i.e. 40% of the HR distribution), for the epigenome-wide association study. HR high risk, BD bipolar disorder, CON control, PRS polygenic risk score.
Validation Set 2
To determine whether the PMPs indexed effects related to the development of psychopathology, 82 independent unrelated individuals of European ancestry were selected from batch 4. BD cases were recruited in Australia, and 68% of HR and controls were from US sites. A ‘BD-syndromic’ group (n = 44) comprised BD cases (n = 20) plus HR individuals who developed BD-related psychopathology (n = 24), and a ‘disorder-free’ group (n = 38) comprised controls (n = 26) plus HR who remained well (n = 12).

PMPs were generated using ρ<0.05 (n = 1,891 CpGs, after QC exclusions), and extended ‘BD-syndromic’ vs. ‘disorder-free’ groups, using univariate GLM, as described above. BD-PRS and a BD-PRS×Group interaction term were also added to the model.

Sixty-six individuals in Validation Set 2 (80%) completed the FACES-II scale [53]. Exploratory regression analysis evaluating the impact of family covariates. BD-PRS and a BD-PRS×Group interaction term were evaluated using univariate GLM, including BD-PRS group (n = 264) and cross-disorder-PRS (ANCOVA T < 0.01), and the cross-disorder-PRS was optimal at p = 0.026 (Fig. 3). The quantile-quantile inflection point was defined at p < 0.002. Post-hoc adjustment for epigenetically-inflected smoking exposure (using p0.00575921 as a covariate) did not substantially alter the findings (cg00933603, p = 3.10 × 10−7; Table S2, Fig. S5).

Putative biological impact of differentially methylated genes
The DMPs above the inflection point (p < 0.002; n = 82; Table S2) and nominally significant DMPs (p < 0.05; n = 1957) were physically and functionally mapped to 66 and 1260 genes, respectively. Genes with DMPs above the inflection included MLC1, ESRT1, KCKNS, L1CAM, CPEB1 and GABBR2, previously associated with BD [73] (Table S2; Fig. S7). EWAS-significant probe cg00933603 lies in exon 2 of VARS2, and has a cis-mQTL SNP rs2532928 (formerly known as rs116537083) [74], that correlates with VARS2 cortical expression (GTEx v8; [75] p = 9.10 × 10−29, Fig. S8).

In FUMA analyses, the 1183 DMPs were enriched in genes previously associated with several relevant neurobehavioral phenotypes in the GWAS catalog, including schizophrenia, chronotype and general risk tolerance, which were represented in the top 15 enrichments (adjusted p < 0.01; Table 1 and Table S3). The genes that mapped to DMPs were mostly upregulated in brain regions, including the hypothalamus, anterior cingulate cortex and amygdala (adjusted p = 4.40 × 10−16, 7.08 × 10−16 and 1.69 × 10−15, respectively; Fig. S9). Top gene ontology enrichment categories from FUMA

Group comparisons: High vs. Low-BD-PRS strata within HR participants
Within the HR group, the High-BD-PRS stratum had significantly higher mean MDD-PRS (ANCOVA F = 5.32, p = 0.024, η²p = 0.064) and cross-disorder-PRS (ANCOVA F = 4.57, p = 0.035, η²p = 0.055) than the low-BD-PRS stratum. Epigenetically inferred smoking status analysis identified only two probable smokers, and these were equally represented in PRS group strata (Fig. S4). There was no difference in global functioning (GAF) between High-BD-PRS vs. Low-BD-PRS groups (M ± SD = 86.0 ± 9.3 vs. 84.8 ± 8.4, p = 0.54) (Fig. S6).

Differentially methylated positions: high vs. low-BD-PRS strata within HR
EWAS was performed using a linear regression framework, using the final probe set of 35,907 variable CpGs that were blood-brain correlated. Following minor correction for bias and inflation (−0.0022 and 0.97, respectively; corrected to −0.0049 and 1, respectively), one probe (cg00933603, p = 3.54 × 10−7) exceeded the epigenome-wide association threshold (pT < 1.39 × 10−6; FDR = 0.026) (Fig. 3). The quantile-quantile inflection point was defined at p < 0.002. Post-hoc adjustment for epigenetically-inflected smoking exposure (using p0.00575921 as a covariate) did not substantially alter the findings (cg00933603, p = 3.10 × 10−7; Table S2, Fig. S5).

RESULTS
Polygenic risk scores
BD-PRS at ρ = 0.105 optimally distinguished the independent BD group (n = 264) from controls (n = 1115), with p = 5.36 × 10−12 and R² = 0.057 (n = 214,928 SNPs; Fig. S1). The optimal MDD-PRS (for BD vs. controls) was at ρ = 0.017 (n = 187,029 SNPs, p = 1.89 × 10−8; R² = 0.037), and the cross-disorder-PRS was optimal at ρ = 0.229 (n = 269,339 SNPs, p = 1.76 × 10−10; R² = 0.049). At a group level, BD-PRS distinguished HR (n = 355) from both BD and control groups, where HR had higher mean BD-PRS than controls (n = 1699; GEE p = 0.0003) and lower BD-PRS than BD (n = 319; GEE p = 0.0006). The BD-PRS exhibited an overlapping population distribution across clinical groups (Fig. 2).

Fig. 3 PRS-stratified EWAS in high-risk participants. A Quantile-Quantile plot, indicating observed vs. expected p values from 35,907 probes. Bias and inflation corrected to −0.00049 and 1, respectively. The dashed line indicates the inflection point at p = 0.002. B Manhattan plot indicating the genomic location of differentially methylated probes. The vertical line indicates the multiple-testing correction threshold at α = 0.05 for epigenome-wide association, based on 35,907 probes (p < 1.39 × 10−6). The location of the top DMP, cg00933603, is indicated. Covariates included age, sex, MDS C1 and C2, tissue source, six blood-cell components, and four surrogate variables. log logarithm; P p value.

Translational Psychiatry (2022) 12:310
including synaptic and neuronal cellular components; canonical pathways involving synaptic function/interactions and signal transmission; and biological processes relating to cell-cell adhesion and signalling functions (Table 2, Table S4).

Supplementary EWAS: post-hoc sensitivity analysis for interpretation of DMP functional enrichments

No DMP survived multiple-testing correction in the supplementary EWAS using 214,352 probes regardless of blood-brain correlation (Fig. S10). The top DMP from the primary PRS-EWAS (cg00933603; \( p = 3.54 \times 10^{-7} \)) was the third top DMP (\( p = 4.38 \times 10^{-9} \)) in this secondary EWAS (Table S5).

Pathway analysis in FUMA showed that brain tissues continue to exhibit the strongest enrichment with an equivalently-sized gene list derived from DMPs regardless of blood–brain correlation. The top brain regions were cortex, anterior cingulate cortex and amygdala (\( p = 1.58 \times 10^{-14}, p = 1.35 \times 10^{-11} \) and \( p = 2.39 \times 10^{-11} \), respectively; Fig. S11). Gene ontology enrichments implicate neuronal-related functions and GWAS catalog overlaps, with chronotype represented in the top 15 enrichments (Tables 1 and 2, Tables S6 and S7).

Test of generalisability of epigenetic signature using polymethylome profile score

Validation Set 1. PMPS calculated using three \( p \) value thresholds (\( p_T \)) were normally distributed (Shapiro–Wilk \( p > 0.05 \)). At the most stringent threshold (\( p_T = 0.002 \)), no association between PMPS and BD-PRS strata in controls was observed (\( F = 1.00, p = 0.323, \eta^2 p = 0.023 \)). However, the PMPS at \( p_T = 0.05 \) (\( n = 1,891 \) CpGs) showed stronger associations with BD-PRS strata (\( F = 11.57, p < 0.002, \eta^2 p = 0.212 \)) than the PMPS at \( p_T = 0.01 \) (\( F = 7.11, p = 0.011, \eta^2 p = 0.142 \)); PMPS was higher in controls with High-BD-PRS (\( p_T = 0.05; M \pm SE = 11.84 \pm 1.27 \)) compared to low BD-PRS (\( M \pm SE = 8.17 \pm 1.34 \)). Thus we considered \( p_T = 0.05 \) to be optimal for use in Validation Set 2 (Table S9).

Validation Set 2

Associations between PMPS and PRS or clinical status in BD and CON groups: PMPS at \( p_T = 0.05 \) was normally distributed within each clinical group (Shapiro–Wilk \( p > 0.05 \)). PMPS did not differ between BD and control groups in a basic model that did not covary for BD-PRS (\( F = 2.39, p = 0.131, \eta^2 p = 0.066 \); Table S9). Following the inclusion of BD-PRS and BD-PRSxGroup term in the model, a significant association between PMPS and BD group was observed (\( F = 7.20, p = 0.011, \eta^2 p = 0.184 \); Table S9), in conjunction with a significant BD-PRSxGroup interaction (\( F = 7.07, p = 0.012, \eta^2 p = 0.181 \)) and a suggestive effect of BD-PRS (\( F = 3.99, p = 0.054, \eta^2 p = 0.111 \)).

Associations between PMPS and PRS or clinical status in ‘BD-syndromic’ and ‘disorder-free’ groups: Extended-clinical group definitions were utilised based on BD-psychopathology in HR (Table S8). HR meeting threshold or sub-threshold diagnoses (HR\text{unwell}) demonstrated lower functional capacity (CGI scores).
Table 2. FUMA top 10 gene ontology (GO) and canonical pathway enrichment categories of ~1,183 genes that map to differentially methylated probes in primary EWAS (left) and comparable gene list from supplementary EWAS that included all CpGs regardless of blood-brain correlation (right).

| RANK | Gene set | N overlap (N genes) | p | adjP | Gene Set | N overlap (N genes) | p | adjP |
|------|----------|---------------------|---|-----|----------|---------------------|---|-----|
| **GO biological processes** | | | | | | | | |
| 1 | Biological adhesion | 128 (1404) | 7.88E | −26 | 5.79E | −22 | Neurogenesis | 152 (1594) | 1.70E | 1.25E |
| 2 | Homophilic cell adhesion via plasma membrane adhesion molecules | 41 (165) | 1.37E | −24 | 5.04E | −21 | Neuron differentiation | 128 (1343) | 2.47E | 9.08E |
| 3 | Cell–cell signalling | 137 (1638) | 4.77E | −24 | 1.17E | −20 | Positive regulation of biosynthetic process | 160 (1966) | 1.78E | 4.37E |
| 4 | Cell–cell adhesion via plasma membrane adhesion molecules | 50 (271) | 2.20E | −23 | 4.05E | −20 | Positive regulation of gene expression | 154 (1955) | 4.99E | 9.18E |
| 5 | Cell–cell adhesion | 87 (819) | 4.78E | −22 | 7.03E | −19 | Central nervous system development | 99 (972) | 2.25E | 3.31E |
| 6 | Synaptic signalling | 73 (712) | 8.10E | −18 | 9.92E | −15 | Sensory organ development | 70 (535) | 7.72E | 9.46E |
| 7 | Ion transport | 121 (1663) | 1.53E | −16 | 1.61E | −13 | Positive regulation of RNA biosynthetic process | 132 (1592) | 1.43E | 1.51E |
| 8 | Positive regulation of molecular function | 121 (1740) | 4.19E | −14 | 1.33E | −11 | Biological adhesion | 100 (1030) | 4.63E | 4.25E |
| 9 | Synapse assembly | 31 (175) | 1.63E | −14 | 1.33E | −11 | Biological organ morphogenesis | 120 (1404) | 1.24E | 1.01E |
| 10 | Synapse organization | 47 (404) | 6.41E | −14 | 4.71E | −11 | Growth | 96 (979) | 1.62E | 1.19E |
| **GO cellular components** | | | | | | | | |
| 1 | Intrinsic component of plasma membrane | 145 (1697) | 2.26E | −26 | 2.26E | −23 | Neuron part | 150 (1709) | 3.48E | 3.48E |
| 2 | Synapse | 107 (1169) | 8.30E | −22 | 4.16E | −19 | Cell junction | 113 (1275) | 1.23E | 6.17E |
| 3 | Neuron part | 130 (1709) | 2.66E | −22 | 8.88E | −19 | Plasma membrane region | 104 (1185) | 1.16E | 3.14E |
| 4 | Cell junction | 106 (1275) | 1.53E | −16 | 3.83E | −13 | Synapse | 103 (1169) | 1.33E | 3.14E |
| 5 | Synapse part | 85 (932) | 2.05E | −17 | 4.11E | −15 | Neuron projection | 110 (1301) | 1.57E | 3.14E |
| 6 | Neuron projection | 102 (1301) | 3.68E | −16 | 6.15E | −14 | Cell projection part | 115 (1438) | 1.56E | 2.60E |
| 7 | Plasma membrane region | 94 (1185) | 2.61E | −22 | 4.01E | −19 | Intrinsic component of plasma membrane | 124 (1697) | 7.63E | 1.09E |
| 8 | Postsynapse | 61 (610) | 1.35E | −14 | 1.68E | −12 | Somatodendritic compartment | 77 (818) | 1.82E | 2.27E |
| 9 | Cell projection part | 103 (1438) | 8.84E | −14 | 9.84E | −12 | Cytoskeletal part | 119 (1639) | 5.43E | 6.04E |
| 10 | Whole membrane | 111 (1647) | 4.82E | −13 | 4.82E | −11 | Synapse part | 81 (932) | 2.69E | 2.70E |
| **GO molecular function** | | | | | | | | |
| 1 | Calcium ion binding | 79 (693) | 5.41E | −22 | 8.91E | −19 | Sequence-specific DNA binding | 106 (1114) | 1.15E | 1.90E |
| 2 | Ribonucleotide binding | 123 (1885) | 1.40E | −22 | 9.17E | −19 | Regulatory region nucleic acid binding | 91 (934) | 1.09E | 1.90E |
| RANK | Gene set                                                                 | N overlap (N genes) | p     | adjP | Gene Set                                                                 | N overlap (N genes) | p     | adjP |
|------|--------------------------------------------------------------------------|---------------------|-------|------|--------------------------------------------------------------------------|---------------------|-------|------|
|      | **Primary EWAS (blood-brain correlated CpGs)**a                          |                     |       |      | **Supplementary EWAS (all CpGs)**a                                       |                     |       |      |
| 2.23E−13 | 1.83E−10                                                                 | DNA binding transcription factor activity | 128 (1691) | 1.37E−18 | 7.49E−16                                                                | 2.96E−20 | 2.43E−17 |
| 3    | Drug bindingb                                                            | 112 (1718)          | 3.09E−12 | 1.70E−09 | 9.91E−09 Sequence-specific double-stranded DNA bindingb                  | 83 (860)           | 2.59E−18 | 1.06E−15 |
| 4    | Adenyl nucleotide binding                                                | 101 (1536)          | 2.41E−11 | 9.28E−08 | 8.10E−08 Double-stranded DNA bindingb                                     | 88 (952)           | 3.60E−18 | 1.19E−15 |
| 5    | Sequence-specific DNA bindingb                                           | 78 (1114)           | 3.01E−10 | 9.28E−08 | 1.70E−07 Drug bindingb                                                   | 125 (1718)         | 7.84E−17 | 2.01E−14 |
| 6    | Regulatory region nucleic acid bindingb                                  | 69 (934)            | 3.54E−10 | 9.28E−08 | 7.49E−12 Ribonucleotide bindingb                                          | 133 (1885)         | 8.55E−17 | 2.01E−14 |
| 7    | Transmembrane transporter activity                                        | 74 (1038)           | 3.95E−10 | 9.28E−08 | 7.49E−12 Sequences-specific DNA bindingb                                  | 118 (1706)         | 2.24E−12 | 3.69E−12 |
| 8    | Double-stranded DNA bindingb                                             | 69 (952)            | 7.90E−10 | 1.62E−07 | 1.06E−04 Adenyl nucleotide bindingb                                       | 114 (1536)         | 5.06E−16 | 1.04E−13 |
| 9    | Sequence-specific double-stranded DNA bindingb                           | 64 (860)            | 1.17E−09 | 2.13E−07 | 7.90E−16 Cytoskeletal protein bindingb                                     | 83 (948)           | 7.90E−16 | 1.44E−13 |
| 10   | Ion transmembrane transporter activity                                    | 64 (870)            | 1.84E−09 | 2.76E−07 | 1.06E−04 Identical protein bindingb                                       | 118 (1706)         | 2.24E−12 | 3.69E−12 |
|      | **Canonical pathways (Reactome)**                                        |                     |       |      |                                                                          |                     |       |      |
| 1    | Neuronal systemb                                                          | 52 (410)            | 7.87E−17 | 1.73E−13 | Neuronal systemb                                                          | 44 (410)           | 8.06E−12 | 1.77E−12 |
| 2    | Protein proteinb interactions at synapses                                 | 21 (87)             | 5.20E−13 | 5.71E−10 | Extracellular matrix organization                                         | 35 (299)           | 1.03E−07 | 1.13E−07 |
| 3    | Transmission across chemical synapsesb                                   | 33 (269)            | 8.46E−11 | 6.20E−08 | Transport of small molecules                                              | 55 (728)           | 1.31E−08 | 9.61E−06 |
| 4    | Neurotransmitter receptors and postsynaptic signal transmissionb         | 28 (204)            | 1.66E−09 | 9.10E−08 | Axon guidance                                                             | 45 (551)           | 2.90E−08 | 1.38E−05 |
| 5    | Neurexins and neuropeptides                                              | 15 (56)             | 2.21E−10 | 9.71E−08 | Developmental biology                                                     | 72 (1100)          | 3.14E−08 | 1.38E−05 |
| 6    | Activation of NMDA receptors and postsynaptic events                     | 17 (92)             | 6.95E−09 | 2.55E−06 | NABA matriosomeb                                                          | 68 (1024)          | 4.42E−08 | 1.62E−05 |
| 7    | Disease                                                                  | 68 (1072)           | 1.88E−07 | 5.92E−05 | Signalling by receptor tyrosine kinases                                   | 39 (468)           | 1.43E−07 | 4.17E−05 |
| 8    | NABA core matriosomeb                                                    | 26 (274)            | 1.39E−06 | 0.00382 | Protein protein interactions at synapsesb                                 | 15 (87)            | 1.52E−07 | 4.17E−05 |
| 9    | CREB1 phosphorylation through NMDAR-mediated activation of RAS signalling| 8 (28)              | 2.21E−06 | 0.00054 | Neurotransmitter receptors and postsynaptic signal transmissionb         | 23 (204)           | 3.14E−07 | 7.29E−05 |
| 10   | GPCR ligand binding                                                      | 35 (453)            | 2.94E−06 | 0.000598 Transmission across chemical synapsesb                          | 27 (269)           | 3.32E−07 | 7.29E−05 |

*RANK rank of enrichment category on the basis of adjusted p-value, N genes number of genes in category, N overlap number of DMP genes in category. DMP differentially methylated position, P p value, adjP adjusted p value.

aData are presented for an equivalent number of genes in both primary and supplementary EWAS that map to differentially methylated probes. Full outputs of GO and canonical pathway enrichment are provided in Tables S4 and S7.

Categories that are represented in top 10 for both primary EWAS using blood-brain selected probes, and unselected supplementary EWAS.
than HR who remained well (HR الوالد) (Fig. S12). Overall, the ‘BD-syndromic’ group had poorer global functioning (GAF) than the ‘disorder-free’ group (M ± SD = 74.0 ± 14.7 vs. 92.7 ± 6.0, p = 2.75 × 10−6) (Fig. S6).

PMPS was not associated with psychopathology-group without BD-PRS in the model (F = 0.39, p = 0.533, ηp2 = 0.006). Following the inclusion of BD-PRS and BD-PRS × Group term in the model, a significant association between PMPS and BD group was observed (F = 5.55, p = 0.021, ηp2 = 0.076; Table S9), in conjunction with a significant BD-PRS × Group interaction (F = 5.52, p = 0.022, ηp2 = 0.075) and a non-significant effect of BD-PRS (F = 2.33, p = 0.132, ηp2 = 0.033).

A post-hoc sensitivity analysis of the impact of psychiatric medication use on PMPS was conducted within the BD-syndromic group, which showed no evidence of medication effects on PMPS (Supplementary Results). Exploratory regression analysis in the subset of participants with FACES data (n = 66; omnibus model p = 0.013), showed that PMPS was not associated with FACES (Wald−χ2 = 0.671, p = 0.413), but BD-PRS, psychopathology-group, and BD-PRS × Group were significant predictors of PMPS (Wald−χ2 = 9.55, p = 0.002; Wald−χ2 = 9.29, p = 0.002; Wald−χ2 = 10.17, p = 0.001; respectively), while the FACES × Group interaction was non-significant (Wald−χ2 = 3.24, p = 0.072).

**Sensitivity analysis for population stratification**

To ensure our results were robust to residual population stratification within our selected study samples, the most divergent four ‘nearest neighbour’ individuals were identified by pairwise IBS distances, and excluded if Z > 4 (n = 5 individuals affected by total; n = 1 discovery EWAS, n = 3 validation set 1, n = 1 validation set 2). Removal of one participant from discovery EWAS with outlier values on C4 did not substantially alter the findings (effect size r = 0.9999; absolute β difference = 1.43 ± 1.49%). For validation set 1, three ‘nearest neighbour’ exclusions resulted in a slight attenuation of association between PMPS at p1 = 0.05 and BD-PRS strata (F = 8.946, p = 0.005, ηp2 = 0.183). For validation set 2, ‘nearest neighbour’ exclusions (n = 1) revealed almost identical results and effect sizes (ηp2 ± 0.001).

**DISCUSSION**

After stratifying HR individuals based on their personal burden of BD-associated common genetic variants indexed by BD-PRS and restricting the epigenome search space to variable probes with blood-brain correlation, we found a single epigenome-wide significantly differentially methylated probe (cg00933603), located in an active regulatory element in exon 2 of the VARS2 gene, which lies in the major histocompatibility complex region (hg19/chr6:25–34 Mb). VARS2 encodes a mitochondrial aminocyl-tRNA synthetase involved in mitochondrial protein synthesis. Mitochondrial abnormalities are evident in BD [76] and other psychiatric illnesses [77], and loss-of-function mutations in VARS2 have been previously associated with mitochondrial encephalopathies [78], epilepsy [79], and schizophrenia [80]. Furthermore, an EWAS of depressive symptoms in 724 monozygotic Danish twins identified a differentially methylated region in a putative active enhancer of VARS2, which spanned 9-probes including cg00933603 [81]. In our study, VARS2 was hypomethylated in individuals with a high polygenic burden for BD. Hypomethylation in the 5′ region of a gene can promote gene expression [82], which in the case of VARS2, might play a role in phasic dysregulation of mitochondrial bioenergetics associated with BD [83]. Further characterisation of this ubiquitously expressed functionally relevant gene in the pathophysiology of BD is required.

Beyond the top DMP, disease-associated genetic variants may broadly impact epigenetic processes, supporting the supposition that DNA methylation may mediate genetic risk [84], potentially via long-range epigenetic networks [85]. Of 1,183 genes that mapped to nominally significant DMPs, some were previously associated with BD or schizophrenia, including MLC1, ESR1, KCKNS, L1CAM, CPEB1 and GABBR2 [73]. In addition, 693 of these genes contained CpGs independently nominally associated with later development of BD and MDD in individuals at high familial risk of mood disorder [43]. Although overlap between the specific DMPs implicated in that prior study and our present analysis was low, gene-level convergence provides supporting evidence for specific differentially methylated genes in developing affective psychopathology.

While the subgrouping of HR participants was primarily focused on genetic predisposition to BD, the BD-PRS strata did not exclusively index BD-associated risk—a higher burden of cross-disorder and MDD-associated variants was observed—which supports a relative lack of specificity of PRS. It is therefore possible that the methylation signatures identified herein may show pleiotropic effects across these related psychiatric disorders.

Functional analyses indicated that genes differentially expressed in the brain were over-represented amongst the differentially methylated genes, particularly in the hypothalamus and anterior cingulate cortex (ACC). The hypothalamus is part of the hypothalamic-pituitary-adrenal (HPA) axis, the key stress-response system, which has been shown to be dysregulated in mood disorders [86]. In addition, the ACC plays an important role in cognitive functions and emotional regulation. Structural imaging meta-analyses have shown grey matter reduction in ACC [87], and smaller hippocampal volumes associated with BD [88, 89], thus methylation changes impacting these tissues are potentially of relevance. Overlap between genes mapped to DMPs and genes implicated in relevant phenotypes via GWAS— including BD and schizophrenia, chronotype, and risk tolerance —suggest convergence of epigenetic and genetic signals. Sleep disruption and chronotype have long been posited as hallmark features of BD [90], as well as targets of psychiatric medication [91], and may influence medication response [92]. Risky behaviour is a common symptom in manic phases and is genetically correlated with BD [93], highlighting relevant pathways within our defined epigenetic signals.

Restricting the discovery EWAS to putatively functionally relevant CpGs that reflect methylation status in the primary disease-affected tissue [62, 63] enabled reduction of the EWAS search space and application of a more permissive epigenome-wide significance threshold than previously modelled [71], but may consequently bias enrichment analyses. Our secondary EWAS employed all variable CpGs, regardless of blood-brain correlation, and revealed similar over-representation in brain tissue, pathway enrichments, and overlap of top DMPs with an equivalent-sized gene list; supporting the primary methodological approach.

Utilising nominally differentially methylated CpGs from PRS-stratified EWAS in generating a poly-methylomic profile score (PMPS) [72]—akin to calculating PRS from GWAS summary statistics for genotypic data [14]—we validated the impact of BD-PRS on methylation signature. We found significant associations between BD-PRS and PMPS in controls, demonstrating the generalisability of the PMPS as an index of PRS, regardless of family history of BD. Consistent with the dependency of the PMPS on PRS background, when BD-PRS was included in the statistical models testing differences between groups in PMPS, a higher mean poly-methylomic profile score was observed in those with BD-symptomatology compared to disorder-free individuals. This highlights the important interplay between environmental, epigenetic and genetic risk factors, which may lead to the development of psychopathology, and is consistent with a multifactorial liability model [94].

While there was no significant effect of FACES on PMPS in our exploratory analysis, this study cannot rule out the potential influence of family environment on methylation signature, given
the small sample size with available FACES that was used. The PMPS appears to at least partially reflect environmental differences related to the burden of BD-associated variants, however, increasing the sample size for discovery EWAS may further elucidate the mechanisms underlying the development of psychopathology. Longitudinal investigation of epigenetic markers in HR individuals based on their clinical status over time may also reveal useful epigenetic signatures of the BD prodrome.

**Limitations**

This study was restricted to participants of European-ancestry, due to substantial ethnicity-specific effects on PRS [60]. We note that residual population substructure not indexed by the covariates we employed (i.e. the first two MDS components or four surrogate variables of methylation data) may influence the results reported herein. However, expanding the existing 15 essential covariates to include additional MDS components would further reduce the ‘subjects per variable’ ratio, and compromise model stability and accuracy of estimation of regression coefficients [95]. This trade-off in design of our statistical model for analysis of this cohort may lead to under-correction of fine-scale population structure that is evident in European population [96, 97], although sensitivity analysis showed little effect of excluding potential outliers. To permit the study design to include both discovery and validation samples, sub-setting of the cohort was required, which limited sample size. While we employed a Bayesian method to reduce potential for EWAS bias [67], the methylation effects reported herein require validation. We mapped DMPs both physically and functionally [68], to comprehensively define genes associated with DMPs in both coding and non-coding genomic regions. However, genes were not prioritised for inclusion in enrichment analysis based on the number nor location of DMPs within each mapped gene, nor was the list restricted to one gene per DMP, thus enrichment signals may have been impacted by inclusion of genes whose expression was not affected by DMPs. Furthermore, in silico mapping of potentially cross-reactive CpG probes to homologous genome locations [98] may impact gene annotations and enrichments. Moreover, FUMA does not account for the number of CpGs per gene, therefore enrichment results may be biased towards categories with larger genes that contain more probes (i.e., with greater stochastic chance to be differentially methylated) [99].

One universal limitation of epigenome-wide studies in psychiatric disorders is the accessibility to brain tissue, as there is considerable variation in DNA methylation across tissues and cell types [100, 101]. We addressed this design limitation by covarying for inter-individual variability in blood-cell components, and focusing on CpGs which are blood-brain correlated [62]. However, use of blood-derived DNA may exclude relevant brain-specific CpGs that are not variable in this surrogate tissue. We acknowledge that: (1) participants were at different follow-up points, with potential limitations on capacity for defining ‘syndromic’ status that is dependent on length of clinical follow-up, and (2) PMPS reflects baseline methylation signatures, and timing of emergence of symptoms in relation to proximity to blood draw is not accounted for. Furthermore, while only two participants in the discovery EWAS were epigenetically inferred to be probable smokers (and were equally represented in PRS strata), smoking exposure may be a relevant confounder for future replication. Finally, psychiatric medication use was not controlled for due to the small numbers of individuals exposed, and although sensitivity analysis indicated non-significant effect of medication on PMPS, this remains an important caveat.

Although discovery EWAS in HR participants stratified by BD-PRS can reveal epigenetic markers associated with genetic risk for developing BD, many genetic variants that may impact BD risk are not indexed by PRS [18]. Indeed, familial risk encompasses all classes of genomic variation—including structural and rare variants—which may track more closely than PRS with disease-status in families with strong family history [21]. While we excluded annotated ‘probe SNPs’ from our analysis [102], the existence of polymorphic genotypes (e.g., rare or recently discovered DNA variants) that are present in our cohort and underlie CpGs could potentially lead to bias [103]. Conversely, we note that standard clumping procedures employed herein (based on r2) for generation of PRS included multiple MHC variants, thus PRS group differences may be inflated by multiple genetic effects from this region, which may amplify apparent effect sizes of methylation differences in the MHC region. The MHC region was also not excluded in FUMA enrichment analysis. Finally, limited availability of family environment data impacted power to detect effects, which require larger samples to definitively characterise. Thus extension to larger cohorts, and potential examination of the impact of other environmental factors, including stressful life-events, should be considered in future.

**CONCLUSIONS**

To our knowledge, this is the first report to investigate DNA methylation differences amongst individuals at high risk of BD, in the context of their personal polygenic background for BD. While only a single site in the VAR2 gene exceeded epigenome-wide significance, many CpG sites were nominally differentially methylated, which related to neurological pathways and functions associated with risk of psychopathology. Methylation profiling in independent validation sets confirmed the relationships between methylation signatures and genetic background, finding that methylation profiles may also partially reflect differences in family environment. Further larger-scale studies are needed to examine the impact of environmental factors in the relationship between familial risk and development of psychopathology.

**DATA AVAILABILITY**

The data that support the findings of this study are available in the UNSW data-archive “ResData” at www.dataarchive.unsw.edu.au, under Research Data Management Plan reference number DQ237303, or are available on request from the authors.

**REFERENCES**

1. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. Am J Med Genet C Semin Med Genet. 2003;123C:48–58.
2. Craddock N, Jones I. Genetics of bipolar disorder. J Med Genet. 1999;36:585–94.
3. Craddock N, Sklar P. Genetics of bipolar disorder. Lancet. 2013;381:1654–62.
4. Mesman E, Nolen WA, Reichart CG, Wals M, Hillegers MH. The Dutch bipolar offspring study: 12-year follow-up. Am J Psychiatry. 2013;170:542–9.
5. Perich T, Lau P, Hadzis-Pavlovic D, Roberts G, Frankland A, Wright A, et al. What clinical features precede the onset of bipolar disorder? J Psychiatr Res. 2015;62:71–7.
6. Van Meter AR, Burke C, Youngstrom EA, Faedda GL, Correll CU. The bipolar prodrome: meta-analysis of symptom prevalence prior to initial or recurrent mood episodes. J Am Acad Child Adolesc Psychiatry. 2016;55:543–55.
7. Duffy A, Vandevelde C, Heffer N, Preisig M. The clinical trajectory of emerging bipolar disorder among the high-risk offspring of bipolar parents: current understanding and future considerations. Int J Bipolar Disord. 2017;5:37.
8. Frankland A, Roberts G, Holmes-Preston E, Perich T, Levy F, Lenroot R, et al. Clinical predictors of conversion to bipolar disorder in a prospective longitudinal familial high-risk sample: focus on depressive features. Psychol Med. 2018;48:1713–21.
9. Duffy A, Goodday S, Keown-Stoneman C, Grof P. The emergent course of bipolar disorder: observations over two decades from the canadian high-risk offspring cohort. Am J Psychiatry. 2019;176:720–9.
10. Sandstrom A, MacKenzie L, Pizzo A, Fine A, Rempel S, Howard C, et al. Observed psychopathology in offspring of parents with major depressive disorder, bipolar disorder and schizophrenia. Psychol Med. 2020;50:1050–6.
11. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 2011;43:977–83.
64. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–8.
65. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.
66. Smyth GK. Limma: linear models for microarray data, in Bioinformatics and computational biology solutions using R and bioconductor. Springer: 2005; p. 397–420.
67. van Iterson M, van Zwet EW, Heijmans BT, BIOS Consortium. Controlling bias and inflation in epigenome-and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18:19.
68. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501.
69. Watanabe K, Taskesen E, Van Bochoven A, Posthumus D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1–11.
70. Watanabe K, Mirkov MU, de Leeuw CA, van den Heuvel MP, Posthumus D. Genetic mapping of cell type specificity for complex traits. Nat Commun. 2019;10:1–13.
71. Saffati A, Silver MJ, Zavattari P, Moi L, Cumbalano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol. 2018;42:20–33.
72. Watkeys OJ, Cohen-Woods S, Quédé Y, Cairns MJ, Owers B, Fullerton JM, et al. Derivation of poly-methylic profile scores for schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry. 2020;101:109925.
73. Pfeifer J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:DS85–DS85.D1.
74. McRae AF, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, et al. Identification of 55,000 Replicated DNA Methylation QTL. Sci Rep. 2018;8:17605.
75. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:550–5.
76. Scaini G, Andrews T, Lima CN, Benevenuto D, Streck EL, Quevedo JJM. Mitochondrial basis of bipolar disorder. Neurosci Biobehav Rev. 2017;74:1–11.
77. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.
78. Diodato D, Melchionda L, Haack TB, Dallabona C, Barufi S, Ligante D, et al. The interaction of family environment with polygenic risk. 2022. Submitted.
79. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. 2015;68:627–36.
80. Cook JP, Mahajan A, Morris AP. Fine-scale population structure in the UK Biobank: implications for genome-wide association studies. Hum Mol Genet. 2020;29:2803–11.
81. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–9.
82. Chen Y-A, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8:203–9.
83. Makosimovic J, Oshlack A, Pipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. 2021;22:173.
84. Reinus LE, Acedovo N, Joerink M, Pershagen G, Dahlen S-E, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLOS one. 2012;7:e41361.
85. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Bull A, Yurovsky A, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLOS Genet. 2015;11:e1004958.
86. Price ME, Cotton AM, Lam LL, Farre P, Emberley E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6:4.
87. LaBarre BA, Goncearenco A, Petykowska HM, Jaratlerdsiri W, Bornman MSR, Hayes VM, et al. MethylToSNP: identifying SNPs in Illumina DNA methylation array data. Epigenetics Chromatin. 2019;12:79.

ACKNOWLEDGEMENTS

This research was supported by the Australian National Medical and Health Research Council (NHMRC) Project Grants 1066177 (JMF) and 1063960 (JMF), Project Grant 1010781 (PRS) and Investigator Grants 1176716 (PRS) and 1177991 (PBM), and philanthropic support from the Landowskoe Foundation and Mrs Betty Lynch OAM (dec). Collection of samples and phenotyping at the US sites was supported by NIH Collaborative R01s MH068009 (JIN), MH073151, and MH068006 (MGM). Indiana University provided funding for the methylation analysis of participant samples collected at the US sites. SYS was supported by a University International Postgraduate Award from the University of New South Wales, and this work forms part of the first author’s doctoral dissertation. JMF was supported by the Janette Mary O’Neil Research Fellowship. EKS received support from the Intramural Research Program at the National Institute of Mental Health. MGM was supported in part by the HC Prechter Bipolar Research Program at the University of Michigan. DNA of Australian participants was extracted by Genetic Repositories Australia, an Enabling Facility that was supported by NHMRC Enabling Grant 401184. DNA of US participants was extracted by Rutgers University Cell and DNA Repository. The authors wish to acknowledge the use of the services and facilities of the Australian Genome Research Facility (AGRF), for generation of methylation data. We thank Jessica Johnson and Doug Ruderfer for assistance with sample preparation and PsychArray genotyping at Mt Sinai School of Medicine, New York, USA. We thank Leanne M Williams (School of Medicine, Stanford University, CA, USA) and Justine M. Gatt (School of Psychology, University of New South Wales, Sydney, Australia; and Neuroscience Research Australia, Sydney, Australia) for access to genotype data for the independent cohort of controls (PI: Williams; TWIN-E study) for determination of optimal p-value threshold for PRS in the high-risk participants. Recruitment of the TWIN-E cohort was supported by an Australian Research Council Linkage Grant LP1083621 (LMW, JMG, PRS) and genotyping was supported by National Health and Medical Research Council (NHMRC) Project Grant 1037196 (PBM, PRS). This cohort collection utilised Twins Research Australia, a national resource supported by NHMRC Centre of Research Excellence Grant (1079102). Finally, we thank and acknowledge all research participants who gave their time to contribute to the Adolescents at High

S. Hesam-Shariati et al.
Risk for Familial Bipolar Disorder” and “Bipolar Kids and Sibs” research programmes, whose generosity and ongoing support has made this research possible.

AUTHOR CONTRIBUTIONS
The study was conceptualised by SHS, BJO, CT, PRS, and JMF. Analysis methodology was developed by SHS, BJO and JMF, and administrative efforts and resources were provided by JMF, PRS, JIN, and PBM. HJE provided DNA samples from US participants, and DNA samples were received and prepared by SHS and KDP for methylation typing, and by KDP and HJE for genotyping. Formal analysis was conducted by SHS as part of a PhD thesis dissertation, with supervision by JMF and PRS. The PMPS scoring procedure was developed by OJW, with supervision by MJG and JMF. Genotype cleaning, imputation and quality control were performed by CT and JMF, prior to generation of polygenic scores by SHS. Clinical and family environment data were provided by GR, EKS, HCW, MGM, LAH, JIN, and PBM. The original draft was written by SHS, with primary review & editing by BJO and JMF, before being circulated for feedback and editing by all coauthors, each of whom approved the final manuscript. HJE provided particularly helpful critical reading and advice. Funding relevant to the current work was acquired by SHS (UNSW University International Postgraduate Award scholarship; with the support of JMF and PRS), and was supported by funding acquired by JMF, PRS, PBM, MGM, and JIN.

COMPETING INTERESTS
JIN is an investigator for Janssen, with interests unrelated to the current work. MGM has consulted with Janssen Pharmaceuticals and received research support from Janssen, both unrelated to this work. All other authors have no potential or perceived conflicts of interest to declare.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02079-6.

Correspondence and requests for materials should be addressed to Janice M. Fullerton.

Reprints and permission information is available at http://www.nature.com/reprints

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

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/.

© The Author(s) 2022