medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methylome-wide association study of early life stressors and adult mental health reveals a relationship between birth date and cell type composition in blood

David M. Howard*[1,2], Oliver Pain[1], Ryan Arathimos[1,3], Miruna C. Barbu[2], Carmen Amador[4], Rosie M. Walker[5,6], Bradley Jermy[1,3], Mark J. Adams[2], Ian J. Deary[7], David Porteous[5], Archie Campbell[5,8], Patrick F. Sullivan[9,10,11], Kathryn L. Evans[5], Louise Arseneault[1], Naomi R. Wray[12], Michael Meaney[13,14,15], Andrew M. McIntosh[2,7], and Cathryn M. Lewis[1,3]

Affiliations:

[1] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

[2] Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

[3] NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK

[4] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[5] Centre for Genomic and Experimental Medicine, Institute of Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK

[6] Chancellor's Building, 49 Little France Crescent, University of Edinburgh, UK

[7] Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh, UK

[8] Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

[9] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[10] Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

[11] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA

[12] Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia

[13] Douglas Mental Health University Institute, McGill University, Montréal, QC, Canada

[14] Ludmer Centre for Neuroinformatics and Mental Health, McGill University, Montréal, QC, Canada

[15] Singapore Institute for Clinical Sciences, Singapore, Singapore

* Corresponding author: David M. Howard
Social, Genetic and Developmental Psychiatry Centre,
Institute of Psychiatry, Psychology & Neuroscience,
King's College London, UK
+44 (0)20 7848 5433
E-mail: David.Howard@kcl.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

The environment and events that we are exposed to in utero, during birth and in early childhood influence our future physical and mental health. The underlying mechanisms that lead to these outcomes in adulthood are unclear, but long-term changes in epigenetic marks, such as DNA methylation, could act as a mediating factor or biomarker. DNA methylation data was assayed at 713,522 CpG sites from 9,537 participants of the Generation Scotland: Scottish Family Health Study, a family-based cohort with extensive data on genetic, medical, family history and lifestyle information. Methylome-wide association studies of eight early life environment phenotypes and two adult mental health phenotypes were conducted using DNA methylation data collected from adult whole blood samples. Two genes involved with different developmental pathways (*PRICKLE2* and *ABI1*) were annotated to CpG sites associated with preterm birth ($P < 1.27 \times 10^{-9}$). A further two genes important to the development of sensory pathways (*SOBP* and *RPGRIP1*) were annotated to sites associated with low birth weight ($P < 4.35 \times 10^{-8}$). Genes and gene-sets annotated from associated CpGs sites and methylation profile scores were then used to quantify any overlap between the early life environment and mental health traits. However, there was no evidence of any overlap after applying a correction for multiple testing. Time of year of birth was found to be associated with a significant difference in estimated lymphocyte and neutrophil counts. Early life environments influence the risk of developing mental health disorders later in life; however, this study provides no evidence that this is mediated by stable changes to the methylome detectable in peripheral blood.

Introduction

The diathesis-stress model posits that behaviours and psychological disorders are the result of underlying biological factors (diatheses) plus exposure to stressful events or environments. Childhood adversity increases the risk of poorer physical and mental health outcomes in later life [1, 2], with neglect, sexual and emotional abuse, and violence providing greater risk [3-6]. However,

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

multiple studies have reported additional perinatal and early life stressors and environments that also impact on adult mental health. Individuals born preterm or with a low birth weight are more likely to experience problems with attentiveness and hyperactivity, as well as elevated levels of anxiety and depression [7-9]. The time of year of birth has also been shown to influence psychiatric outcomes in later life [10-12], with those born in January having higher risk of schizophrenia and bipolar and those born in June and July having greater risk of depression. Parental factors, including the age of the parents at birth [13] or having an absent parent [14], are reported to increase the risk of depression. Where we live has also been shown to be detrimental to mental health [15] and is likely due to a variety of environmental and social factors [16, 17].

Adverse environments in the gestational and postnatal periods are known to cause long-term alterations to DNA methylation across the genome [18]. These modifications are reported to mediate resilience and responses to stress-related disorders throughout the life course [19]. Epigenomic variation has also been implicated in a range of psychiatric conditions [20], including bipolar disorder [21], schizophrenia [22], and major depressive disorder (MDD) [23]. Much of the epigenetic research seeking to link early life environments and MDD has been performed using small cohort samples sizes (typically less than 1,000) and focussing on candidate regions [24]. However, this approach is likely to be suboptimal as demonstrated by the lack of reproducible results from candidate gene association studies for MDD [25].

The principal aim of the current research was to investigate whether alterations to DNA methylation have the potential to mediate the stress component in the diathesis-stress model. This was achieved by conducting a methylome-wide association study (MWAS) of early life environments and later mental health outcomes in a single study cohort of over 9,500 adults. Linked electronic health records, responses at interview and questionnaire data were used to ascertain early life environments, with MDD and brief resilience scale (BRS) assessed as the mental health outcomes. The associated CpG sites identified by the MWAS of these phenotypes were then annotated to

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

genes and gene-sets. The gene and gene-set overlap between the early life environments and mental health outcomes were then examined. Additionally, methylation profile scores were used to assess the broader methylome-wide overlap between the early life environment and mental health traits. The influence of time of year of birth on estimated blood cell type composition was investigated further in an independent study cohort.

Materials & Methods

Generation Scotland: Scottish Family Health Study (GS:SFHS)

GS:SFHS [26] is a family-based cohort study of 24,080 participants (14,154 female and 9,926 male) aged between 18 and 100 (mean = 47.6 years, standard deviation = 15.4 years). Baseline data was collected between 2006 and 2011 and covered medical, behaviour and lifestyle factors with a subset recontacted (N = 9,618) in 2015 and 2016 with additional phenotypes collected [27]. Clinical information from linked electronic hospital records across the life course was also available. At the baseline appointment, a blood draw was taken from each participant, which has so far been used to obtain DNA methylation data for 9,773 individuals.

Phenotypes

Multiple phenotypes were generated from the baseline and recontact data focused on either early life environments or adult mental health. The early life environments were broadly categorised as either biological (preterm birth, low birth weight, and time of year of birth) or sociodemographic (having a lone parent, having a young parent, urbanicity, and population density). The adult mental health measures were MDD and psychological resilience (measured using the BRS). The demographics of GS:SFHS are provided in Table 1.

Preterm births were categorized as a recorded gestation of period of less than 37 weeks (based on World Health Organization guidance [28]) using the SMR02 - Maternity Inpatient and Day Case linked electronic health records. These records were available for individuals born after 1992. Birth

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

weight for GS:SFHS participants was also obtained from SMR02 electronic health records. The threshold for low birth weight was based on the sex and gestation length adjusted 3rd centile for birth weight reported using a Scottish sample and SMR02 records by Bonellie et al. [29].

The time of year of birth was assessed using two phenotypes. Firstly, a binary phenotype for birth month was generated with those born between April and October inclusive compared to those born during the remaining months. These months were selected based on the review of birth month and depression by Schnittker [12], with increased risk of depression reported for those born April through October. Secondly, the birth date during the year was assessed as a continuous phenotype (y) and calculated as:

$$y = -1 \times cos\ (2\pi(\frac{date + 10}{365}))$$

Where y ranged from -1 for those born on the winter solstice (21st December) to +1 for those born on the summer solstice (21st June). $date$ was the day of birth during the year for each participant (January 1st = 1, January 2nd = 2, etc.). For those born on the 29th February, $date$ = 59.5.

Using the self-reported age of an individual's parents at the time of birth, an individual was classified as having a young parent if either parent was under the age of 21. Where an individual was recorded as having a lone parent (see below), the age of the parent that the individual lived with was used.

An individual was recorded as having a lone parent if either their mother or father was recorded as living in a different country or region to the individual and the other parent at the time of birth of the individual. The regions reflect the 32 council areas of Scotland. Individuals with missing information, or where an individual was reported not to live with either parent, were excluded.

The region that an individual was living in at the time of their birth was also used to define an urban environment phenotype. Individuals living in Edinburgh, Glasgow, Aberdeen, or Dundee were classified as being urban and those living in other regions as non-urban. The year of an individual's birth and the region they were living in were also used to obtain a measure of population density

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

based on population estimates from the National Records of Scotland [30] and was recorded as the number of individuals per square kilometre.

To determine MDD status, the initial screening questions from the Structured Clinical Interview for DSM-IV (SCID) Non-Patient Version [31] were used to identify those individuals that would subsequently complete the mood sections of the SCID. The SCID was administered by nurses trained in its application and further information on the MDD criteria used in GS:SFHS is reported in Fernandez-Pujals et al. [32]. Participants who met the criteria for at least one MDD episode in the mood sections of the SCID were classified as cases and those who did not meet this criteria or did not report MDD symptoms in the initial screening were classified as controls.

The BRS [33] was used to determine a continuous measure of an individual's psychological resilience and ability to 'bounce back' from stressful events. BRS was obtained for a subset (N = 9,505) of the cohort during a recontact of all GS:SFHS participants in 2015 and 2016. BRS was ascertained from the response to six questions using a 5-point Likert scale with further details on the assessment provided by Navrady et al. [27]. Individuals that responded to 5 or more questions were retained (N = 9,354) with missing values imputed using the missForest package [34] applying 500 trees per forest. A principal component analysis using the Psych package [35] was applied to the data to extract the first unrotated principal component which was then scaled to create a continuous measure of BRS.

Table 1. Demographic information for Generation Scotland: Scottish Family Health Study

	Whole sample			Methylation sample		
Item	N	Present / absent (%)	Mean, St. dev.	N	Present / absent (%)	Mean, St. dev.
Female	24,080	14,154 / 9,926 (58.8)		9,537	5,638 / 3,899 (59.1)	
Age	24,080		47.64, 15.41	9,537		49.81, 13.70
Preterm birth	3,134	109 / 3,025 (3.5)		950	31 / 919 (3.3)	
Low birth weight	3,129	103 / 3,026 (3.3)		949	30 / 919 (3.2)	
Birth month	24,080	12,842 / 11,238 (53.3)		9,516	5,459 / 4,057 (57.4)	
Birth date	22,656		0.01, 0.71	9,509		0.01, 0.71
Young parent	23,546	1,241 / 22,305 (5.3)		9,398	534 / 8,864 (5.7)	
Lone parent	22,543	1,333 / 21,210 (5.9)		9,086	580 / 8,506 (6.4)	
Urban environment	19,428	11,739 / 7,689 (60.4)		8,017	5,468 / 2,549 (68.2)	
Population density	19,426		1,949.72, 1,611.32	8,017		2019.44, 1,489.35
Major depressive disorder	21,340	2,766 /18,574 (13.0)		9,481	1,624 / 7,857 (17.1)	
Brief resilience scale	9,354		0.00, 3.80	4,839		-0.19, 3.90

Number of individuals (N) available for each item. The present N, absent N and prevalence (%) are reported for binary items with mean and standard deviation (St. dev.) reported for continuous items.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Regression of adult mental health on early life environments

To determine whether there was an association between early life stressors and the measures of adult mental health a generalized linear mixed model was used from the lme4qtl package [36]. MDD and BRS were assessed separately as the dependent variable for an association with each early life factor in turn as the independent variable. The total genetic value derived from pedigree data (with the variance/covariance structure defined by a kinship matrix using the kinship2 package [37]) was fitted as a random effect, jointly with sex fitted as a fixed effect. Binary traits were fitted as factors and continuous traits were centered and scaled to have a mean of 0 and a standard deviation of 1. A binomial regression with a logit function was used to assess associations with MDD. For a significant association between a mental health trait and an early life environment a Bonferroni correction was applied to the *P*-values within each mental health trait, $P < 6.25 \times 10^{-3}$ ($\alpha = 0.05 / 8$).

Methylation Data

The Infinium MethylationEPIC BeadChip (Illumina Inc.) was used to profile DNA methylation data at 853,307 CpG sites generated in two sets of individuals. There were 5,190 individuals in Set 1 and 4,583 unrelated individuals in Set 2, with no related individuals between the sets. Quality control was then applied to both Sets and full details are provided by Barbu et al. [38] and McCartney et al. [39]. In summary, individuals were excluded if they were outliers based on multi-dimensional scaling or the predicted sex from the methylation data mismatched the recorded sex. CpG sites were excluded based on low beadcount, poor detection *P*-value, or if they were predicted to bind sub-optimally. The log2 ratio of the intensities of methylated probe versus unmethylated probe data was used to generate methylation M-values separately in each Set [40]. In total there were 713,522 CpG sites remaining with 5,087 individuals in Set 1 and 4,450 individuals in Set 2.

After combining the Sets (9,537 individuals), correction was applied for:

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(1) technical variation, where M-values were included as dependent variables in a mixed linear model adjusting for the plate used to profile the DNA methylation data and the date of the individual's blood draw as random effects, jointly with plate position, Set, clinic, year of appointment, day of week of appointment, and the first 10 principal components (from the EPIC array control sites) as fixed effects; and

(2) biological variation by fitting residuals of (1) as dependent variables in a second mixed linear model adjusting for genetic and common family shared environmental contributions (classed as G: common genetic; K: kinship; F: nuclear family; C: couple; and S: sibling, see Xia et al. [41] and Zeng et al. [42] for further information) as random effects, jointly with sex, age, and estimated blood cell type composition (CD8T, CD4T, NK, Bcell, Mono, Gran; obtained using the Houseman algorithm [43] within the Meffil package [44]) as fixed effects.

Methylome-wide association study of early life environments and adult mental health

Two methods (MWAS 1 and MWAS 2) were used to conduct the methylome-wide association study generating two sets of summary statistics for each phenotype. MWAS 1 was conducted using the eBayes function within the limma package [45], which applies a linear mixed model to the data. The M-values, corrected for technical and biological variation, were the dependent variable with each phenotype included separately as the independent variable. Set, smoking status (ever or never), smoking pack years, and the first 20 principal components created from the M-values using the FactoMineR package [46], were fitted as fixed effects.

MWAS 2 was conducted using the software tool OmicS-data-based Complex trait Analysis (OSCA) [47]. First, to account for methylome-wide correlational structure, this tool uses a linear regression analysis to identify groups of lead CpG sites based on the association test statistics. Each phenotype in turn was the dependent variable with the corrected M-values as the independent variable adjusted for the same covariates used in MWAS 1 (Set, smoking status, smoking pack years, and the

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

first 20 principal components). Second, each group of lead sites were fitted as random effects in a Multi-component MLM-based association excluding the target (MOMENT) analysis to assess the effect of each target probe in turn, with the respective phenotype as the dependent variable and the M-values as the independent variable. MOMENT also fits predicted blood cell type composition (basophil, eosinophil, lymphocyte, monocyte, and neutrophil; based on haematological analysis of the Lothian Birth Cohorts 1921 and 1936 (LBC) [48]) as a fixed effect.

A methylome-wide significance threshold was determined using a Bonferroni correction based on the number of CpG sites analysed for each analysis: $P < 7.01 \times 10^{-8}$ ($\alpha = 0.05 / 713{,}522$). To visualise the MWAS output, Miami plots were created using ggplot2 [49] with the *P*-values for MWAS 1 on the $\log_{10}$ scale and for MWAS 2 on the $-\log_{10}$ scale. QQ-plots were created using Haplin [50] with the shaded error representing the 95% confidence interval. Genomic inflation (λ) was calculated for each output as the median of the observed chi-squared distribution of *P*-values divided by the median of the expected chi-squared distribution.

Gene and Gene set analysis

The significant CpG sites ($P < 7.01 \times 10^{-8}$) for each phenotype and each MWAS were annotated to genes based on location. For MWAS 1, the missMethyl package [51] was used to annotate genes according to location. For MWAS 2, the annotations reported by the OSCA package were used. The gometh function within missMethyl was used to analyse the data for enrichment of Gene Ontology and KEGG gene sets. The overlap of the annotated genes and enriched gene sets between the early life environment and adult mental health phenotypes were examined.

Methylation profile scores

MWAS 2 was used to calculate summary statistics for each early life environment phenotype using the individuals in Set 1 (N = 5,190). Multiple *P*-value thresholds ($< 10^{-7}$, $< 10^{-6}$, $< 10^{-5}$, $< 10^{-4}$, $< 10^{-3}$, $< 10^{-2}$, $< 10^{-1}$) were used to identify those sites for inclusion. For each individual in Set 2, an aggregated

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

methylation profile score was calculated by multiplying the CpG effect sizes from Set 1 by their respective M-values in Set 2. There were no related individuals between Set 1 and Set 2. The ability of these profile scores to predict the same phenotype, MDD or BRS, was assessed after adjusting for sex, age, and the first 20 principal components (derived from single nucleotide polymorphism data) as fixed effects. For prediction of binary phenotypes, a coefficient of determination was calculated using Nagelkerke's R^2 [52] using a population prevalence of that equal to the sample. For quantitative phenotypes, a simple linear regression was used with the adjusted R^2 reported. A Bonferroni correction was used to identify significant prediction based on the number of early life phenotypes examined in each case: $P < 6.25 \times 10^{-3}$ ($\alpha = 0.05 / 8$).

Blood cell type composition

A notable difference was observed between the results from MWAS 1 and MWAS 2 for the time of year of birth phenotypes. The inclusion of predicted cell type composition as a fixed effect in MWAS 2 represents a potential source of these differences. To investigate cell type composition further, haematological and DNA methylation data from 1,308 participants of the LBC was obtained [48, 53], mirroring the development of the OSCA package used in MWAS 2. A similar protocol to that described by Zhang et al. [47] was used to identify effect sizes and *P*-values between 214,923 CpG sites and the cell counts of basophils, eosinophils, lymphocytes, monocytes, and neutrophils.

First, the significant lead CpG sites ($P < 7.01 \times 10^{-8}$) identified for birth date for MWAS 1 in GS:SFHS were examined in the LBC for an association with the five cell type compositions using QQ-plots and λ. Lead sites were identified by retaining the site with the lowest *P*-value amongst sites with an r^2 greater than 0.1 within a 1 Mb window. Second, methylation profile scores were created for individuals in the LBC using the effect sizes from MWAS 1 for birth date in GS:SFHS. These profile scores were created using lead CpG sites at seven *P*-values thresholds ($< 10^{-7}$, $< 10^{-6}$, $< 10^{-5}$, $< 10^{-4}$, $< 10^{-3}$, $< 10^{-2}$, $< 10^{-1}$). A simple linear regression was then used to calculate the association between the profile scores for birth date and the five measured cell type compositions, after adjusting for the sex

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and age of participants as fixed effects. A Bonferroni correction based on the five of cell types analysed required a *P*-value < 0.01 ($\alpha = 0.05 / 5$) for an association.

Results

We conducted analyses of early life environments and adult mental health using the Generation Scotland: Scottish Family Health Study cohort (n = 24,080). First, we used regression to assess the association between early life environments and MDD and BRS using phenotypic data. Second, we conducted methylome-wide association analysis to identify CpG sites associated with these phenotypes. Finally, we used the summary statistics from the methylome-wide association analyses to determine the extent of any overlap between phenotypes.

Regression of adult mental health on early life environments

Generalised linear mixed models were used to assess each early life environment in turn and its association with either MDD or BRS (Supplementary Table 1), after adjusting for sex and relatedness between individuals. There were no significant associations between the early life environments and the adult mental health phenotypes after adjusting for multiple testing ($P > 6.25 \times 10^{-3}$). All early life environments, except for preterm birth, marginally increased the risk of developing MDD in adulthood. All early life environments, except for having a lone parent, marginally lowered BRS suggesting a negative effect on psychological resilience. The associations between time of year of birth and BRS were nominally significant ($P < 0.05$) with those born across the summer months scoring lower on the BRS; however, these associations did not remain after correction for multiple testing.

Methylome-wide association study of early life environments and adult mental health

Normalised M-values for 713,522 CpG sites profiled from combined blood samples of 9,537 GS:SFHS individuals (5,087 in Set 1 and 4,450 in Set 2) remained after quality control procedures. The associations between these M-values and the early life environments and mental health phenotypes

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

were calculated using two association study methods (MWAS 1 and MWAS 2). MWAS 1 fits a linear regression model to the data with the M-values as the dependent variable, whereas MWAS 2 fits the phenotype as the dependent variable while accounting for methylome-wide correlational structure and with an additional correction for predicted blood cell type composition.

The significant CpG sites ($P < 7.01 \times 10^{-8}$) from MWAS 1 and MWAS 2 are in Table 2, except those associated with birth month of birth date. The six significant CpG sites identified in MWAS 2 were also significant in MWAS 1 for the same phenotypes. Supplementary Table 2 and Supplementary Table 3 contains the significant sites ($P < 7.01 \times 10^{-8}$) for the birth month and birth date phenotypes, respectively. In MWAS 1, there were 93 significant sites for birth month and 637 significant sites for birth date. However, no CpG sites were significant in MWAS 2 for either birth month or birth date with further analysis covered in the blood cell type composition section.

Table 2. CpG sites associated with phenotypes in either MWAS 1 or MWAS 2

					MWAS 1			MWAS 2		
Phenotype	CpG site	Chr	BP position	Annotated Gene	Effect size	Standard error	*P*-value	Effect size	Standard error	*P*-value
Preterm birth	cg00725333	3	64189256	*PRICKLE2*	-0.2747	0.0376	**6.24×10^{-13}**	-0.0364	0.0060	**1.15×10^{-9}**
	cg17668848	3	108029973	*HHLA2*	-0.2871	0.0396	**9.13×10^{-13}**	-0.0356	0.0058	**8.21×10^{-10}**
	cg24329141	10	27095369	*ABI1*	-0.2368	0.0352	**2.98×10^{-11}**	-0.0345	0.0057	**1.27×10^{-9}**
Low birth weight	cg15582176	1	1183528	*C1QTNF12*	-0.3604	0.0661	**6.53×10^{-8}**	-0.0254	0.0054	2.49×10^{-6}
	cg19909717	6	107924415	*SOBP*	-0.2225	0.0403	**4.35×10^{-8}**	-0.0329	0.0055	**1.59×10^{-9}**
	cg21803443	7	126547786	*GRM8*	-0.3656	0.0618	**4.62×10^{-9}**	-0.0278	0.0056	7.22×10^{-7}
	cg12090821	14	21755658	*RPGRIP1*	-0.0857	0.0152	**2.43×10^{-8}**	-0.0329	0.0058	**1.47×10^{-8}**
	cg05905731	16	85485785		-0.1695	0.0300	**2.23×10^{-8}**	-0.0294	0.0055	7.50×10^{-8}
Young parent	cg00528572	11	92703433	*MTNR1B*	-0.1369	0.0200	**9.04×10^{-12}**	-0.0145	0.0025	**4.47×10^{-9}**
	cg02427109	18	77917459	*PARD6G, PARD6G-AS1*	-0.1163	0.0213	**4.96×10^{-8}**	-0.0101	0.0025	4.12×10^{-5}
Population density	cg06759845	1	156460474	*MEF2D*	5.20×10^{-6}	9.40×10^{-7}	**3.40×10^{-8}**	90.39	17.02	1.10×10^{-7}
	cg12433043	6	100619893		-6.65×10^{-6}	1.20×10^{-6}	**3.37×10^{-8}**	-88.86	16.75	1.13×10^{-7}
	cg03623878	13	113655560	*MCF2L*	9.10×10^{-6}	1.58×10^{-6}	**9.56×10^{-9}**	81.18	16.88	1.51×10^{-6}
	cg08036492	17	13976536	*COX10*	3.09×10^{-6}	5.71×10^{-7}	**6.74×10^{-8}**	87.75	18.29	1.61×10^{-6}
MDD	cg02280719	1	6802222		0.0400	0.0064	**5.99×10^{-10}**	0.0174	0.0043	5.32×10^{-5}
	cg08548783	6	57903690		0.0378	0.0067	**1.37×10^{-8}**	0.0142	0.0043	9.79×10^{-4}

MWAS 1 results are from a methylome-wide association analysis using linear regression and MWAS 2 results are from a methylome-wide association analysis using OmicS-data-based Complex trait Analysis. Bold *P*-values indicate an association after correction for multiple testing ($P < 7.01 \times 10^{-8}$). Sites are ordered by phenotype and then genomic position. Chromosome number (Chr) and base pair (BP) position are based on genome assembly GRCh37 (hg19). Annotation of genes is provided by missMethyl. MDD = Major Depressive Disorder.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Miami plots for preterm birth, low birth weight, having a young parent and MDD are in Figures 1-4, respectively. The remaining Miami plots are in Supplementary Figures 1-6 and QQ-plots and λ for all phenotypes are in Supplementary Figures 7-16.

Genes and gene set analysis

The genes annotated to associated CpG sites for each phenotype from MWAS 1 and MWAS 2 were examined for overlap between early life environment and adult mental health. For the significant sites in MWAS 1, there were three annotated genes for preterm birth, having a young parent and population density, four annotated genes for low birth weight, 20 annotated genes for birth month, and 167 annotated genes for birth date. There were no annotated genes for the mental health phenotypes and therefore no overlap was observed with the early life environment phenotypes.

For significant sites in MWAS 2, there were three annotated genes for preterm birth, two annotated genes for low birth weight, and one annotated gene for having a young parent. As there were no annotated genes for the mental health phenotypes, there was no overlap of annotated genes with the early life environment phenotypes.

There were no enriched Gene Ontology or the KEGG gene sets for any of the phenotypes from either MWAS 1 or MWAS 2.

Methylation profile scores

Set 1 results from MWAS 2 for the early life environment phenotypes were used to calculate profile scores for individuals in Set 2. The utility of these profile scores for predicting the same phenotype (Supplementary Table 4), MDD, or BRS (Table 3) in Set 2 was assessed. There was significant prediction of birth month using birth month profile scores ($R^2 = 0.002$, $P = 5.19 \times 10^{-3}$) and of having a young parent using the profile scores from young parent ($R^2 = 0.013$, $P = 3.46 \times 10^{-5}$). There was nominal prediction of MDD from preterm birth ($R^2 = 0.002$, $P = 0.034$) and birth month ($R^2 = 0.002$, $P = 0.043$), and nominal prediction of BRS from birth month ($R^2 = 0.003$, $P = 0.016$). However, none of

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the profile scores predicting mental health phenotypes were significant after correction for multiple testing correction ($P > 6.25 \times 10^{-3}$).

Table 3. Prediction of major depressive disorder and brief resilience scale in Set 2 using methylation profile scores. Scores were calculated using the CpG site effect sizes from methylome-wide association studies of early life environment phenotypes in Set 1

	Major depressive disorder			Brief resilience scale		
Early life environment	Threshold	R^2	*P*-value	Threshold	R^2	*P*-value
Preterm birth	1×10^{-4}	0.0023	0.034	1×10^{-7}	0.0001	0.74
Low birth weight	1×10^{-5}	0.0005	0.32	1×10^{-6}	0.0009	0.21
Birth month	1×10^{-6}	0.0021	0.043	1×10^{-6}	0.0032	0.016
Birth date	1×10^{-5}	0.0013	0.10	1×10^{-5}	0.0016	0.09
Young parent	1×10^{-3}	0.0006	0.28	1×10^{-5}	0.0010	0.19
Lone parent	1×10^{-5}	0.0005	0.34	1×10^{-4}	0.0001	0.66
Urban environment	1×10^{-5}	0.0004	0.40	1×10^{-2}	0.0008	0.25
Population density	1×10^{-4}	0.0015	0.08	1×10^{-6}	0.0010	0.19

The effect sizes in Set 1 were obtained from methylome-wide association studies using OmicS-data-based Complex trait Analysis. Multiple thresholds based on CpG site *P*-values were examined, with the threshold explaining the greatest phenotypic variance reported. Nagelkerke's R^2 was used to calculate the proportion of phenotypic variance explained on the liability scale for major depressive disorder. An adjusted R^2 from a simple linear regression was reported for brief resilience scale. The *P*-value required for significance after correction for multiple testing was $P < 6.25 \times 10^{-3}$.

Blood cell type composition

In MWAS 1, there were 314 lead CpG sites ($r^2 > 0.1$ within a 1 Mb window) associated with birth date ($P < 7.01 \times 10^{-8}$); however, in MWAS 2 there were no significant CpG sites for birth date. MWAS 2 included predicted blood cell type composition as a fixed effect, using haematological data obtained from the LBC. Access to the LBC was obtained and a MWAS run on five blood cell types. Of the 314 lead CpG sites from MWAS 1 in GS:SFHS, 220 were available in the LBC. The QQ-plots for the observed and expected *P*-values and λ for these 220 sites in LBC are provided in Supplementary Figures 17-21 for each blood cell type. Examination of these plots revealed no inflation of *P*-values for basophil, eosinophil, and monocyte counts. However, there was noticeable inflation for neutrophil (50 out of 220 sites with $P < 0.05$; $\lambda = 2.503$) and lymphocyte (98 out of 220 sites with $P < 0.05$; $\lambda = 5.841$) counts. Similarly, creating methylation profile scores using the MWAS 1 effect sizes for birth date in GS:SFHS (Supplementary Table 5), revealed significant prediction of neutrophil ($P =$

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

7.81×10^{-4}) and lymphocyte counts ($P = 3.45 \times 10^{-7}$) in the LBC. These profile scores captured 0.8% and 1.9% of the phenotypic variance of the neutrophil and lymphocyte counts, respectively.

Discussion

The impact of early life environments on later life is of critical importance across multiple clinical and research domains. We sought to examine eight early life environments experienced around the time of birth and quantify: (1) their impact on mental health in adulthood; (2) whether they led to detectable changes to the methylome in to adulthood; and finally (3) whether there was any shared association on the methylome between early life environments and adult mental health.

Previous studies have reported an effect of extreme preterm (gestation length less than 28 weeks) on mental health [8, 54]. The population-based GS:SFHS cohort had four participants (out of 3,134) born at less than 28 weeks gestation length and therefore a 37-week threshold for a preterm birth was used. Three CpG sites were associated with preterm birth in both the MWAS analyses: cg00725333, cg17668848, and cg24329141. These three sites mapped to *PRICKLE2* (Prickle Planar Cell Polarity Protein 2), *HHLA2* (Human Endogenous Retrovirus-H Long Terminal Repeat-Associating Protein 2), and *ABI1* (Abl Interactor 1) protein coding genes, respectively. Homologs of *PRICKLE2* have been associated with embryonic development in mice [56, 57] and drosophila [58]. In humans, *PRICKLE2* has been associated with myelomeningocele [59], a severe form of spina bifida. Myelomeningocele compromises the development of the spine and spinal cord in the womb and has been associated with preterm birth [60]. *HHLA2* regulates T-cell function [61] with increased expression observed in osteosarcoma tumours [62], however its role in gestation length is unclear. Knockouts of *ABI1* in mice have demonstrated it to be essential for embryonic development, survival [63] and placental development [64]. *PRICKLE2* and *ABI1* have plausible roles in gestation length with altered methylation of proximal CpG sites into adulthood. A methylome-wide association study conducted using preterm infants [65] revealed a different set of associated CpG sites to those observed here in adulthood.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Similar to preterm birth, published studies of birth weight and mental health report that more extreme birth weight thresholds identify stronger associations with poorer mental health [66, 67]. In the current study, a sex and gestational-age adjusted measure of birth weight was used increasing the distinction with the preterm birth phenotype. Five CpG sites were associated with birth weight in MWAS 1 of which two were also significant in MWAS 2. The two sites that were significant in both MWAS annotated to genes important for the development of the sensory systems: *SOBP* (Sine Oculis Binding Protein Homolog) and *RPGRIP1* (Retinitis Pigmentosa GTPase Regulator Interacting Protein). *SOBP* has been implicated in the embryonic development of the mouse cochlea [68]. In a meta-analysis of DNA methylation data in neonates for low birth weight, a CpG site annotated to the *SOBP* gene was also significant after correcting for false discovery rate (but not after Bonferroni correction) [69]. The critical role of *RPGRIP1* in the remodelling of rod photoreceptors has been demonstrated in humans [70]. The impact of low birth weight on the long-term expression of CpG sites requires further investigation, certainly as low birth weight has been associated with both hearing impairment [71] and ophthalmic deficits [72]. cg21803443 on chromosome 7 was significant in MWAS 1 and annotated to the *GRM8* (Glutamate Metabotropic Receptor 8) protein-coding gene. Küpers et al. [69] also identified a CpG site (cg15908975, $P = 4.52 \times 10^{-7}$ in a European meta-analysis) close to *GRM8* associated with birth weight. *GRM8* is involved in the inhibition of the cyclic AMP cascade influencing glutamatergic neurotransmission and had an association with depression ($P = 1.80 \times 10^{-12}$) in a genome-wide association meta-analysis [73]. However, there was no predictive ability of a methylation profile score for low birth weight to predict either MDD or BRS. A previous analysis of birth weight as a continuous trait in GS:SFHS identified one significant CpG site (cg00966482) [74], however this site was not significant in either MWAS 1 ($P = 0.22$) or MWAS 2 ($P = 0.53$).

Schnittker [12] suggests that the role of seasonality of birth on mental health was more prevalent in the early part of the 20th century and was partly attributable to poorer prenatal nutrition across the winter months. However, Disanto et al. [10] analysed post-1950 data from England and reported an

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

effect of time of year of birth on schizophrenia, bipolar disorder and to a lesser extent recurrent MDD. In the current study, there was nominal evidence for an effect of time of year of birth on BRS and nominal prediction of both MDD and BRS using a profile score for birth month calculated from DNA methylation data. The greatest difference between the MWAS 1 and MWAS 2 results was for the time of year of birth phenotypes. Additional prediction into the LBC, using methylation profile scores, demonstrated that this divergence is likely due to differing blood cell type composition (for neutrophils and lymphocytes) between those born in the winter and those born in the summer. Altered gene expression due to seasonality has been reported previously [75], with the time of year of birth also influencing neonatal immune development [76] and thymic output [77]. The current work provides further substantial evidence that time of year of birth impacts blood cell type composition. This impact is detectable throughout the life course and needs to be accounted for in future research on traits influenced by seasonality of birth.

In the present study, there was no effect of having a young parent on mental health. However, a much larger study (2.9M individuals) on Danish participants found an increased risk of mood disorders (International Classification of Disease codes F30 – 39) for those with teenage mothers (incidence risk ratio = 1.35 [95% CI = 1.30 – 1.40]) and those with teenage fathers (incidence risk ratio = 1.20 [95% CI = 1.13 – 1.27]) [13]. The *MTNR1B* (Melatonin Receptor 1B) protein coding gene was annotated to a CpG site that was associated with having a young parent in both MWAS 1 and MWAS 2 and the *PARD6G-AS1* (PARD6G Antisense RNA 1) RNA gene was annotated to an associated CpG site in MWAS 1. *MTNR1B* has been associated with type 2 diabetes [78, 79] and differential expression in *PARD6G-AS1* has been implicated in transient neonatal diabetes [80]. This suggests a putative link between having a young parent and diabetes, but further work is required to validate any association. The methylation profile for having a young parent from Set 1 had significant prediction of having a young parent in Set 2, suggesting replication across Sets.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The lone parent phenotype in the current study did not account for whether the mother or father was absent, custody arrangements, separation during infancy, or the degree of communication with the absent parent. A large study of 184,496 school-aged children found that life satisfaction was decreased when either parent was absent, but where the father was absent the effect was attenuated when accounting for perceived family affluence [81]. A follow-up to this study, incorporating additional cohorts, also reported similar effects [82]. The present study found no effect of having a lone parent on MDD or BRS and there were no significant associations with any of the CpG sites analysed.

Two correlated early life phenotypes, urban environment and population density, were studied to measure the effect of geographical environment at birth. There was no effect of either early life phenotype on MDD or BRS and there were no associated CpG sites. Urbanicity has been associated with depression [83, 84], although no association has been observed in low- and middle- income countries [85] and the United States [86]. Population density has also been shown to have limited impact on depression in Turin, Italy [87]. The contradictory findings and the lack of observable effects here may be due to the multiple factors incorporated in these phenotypes, such as pollution [88] and socio-economic status [89], but may be offset by access to mental health services.

Published DNA methylation analysis of MDD have typically been conducted on relatively small samples. The two CpG sites associated with MDD in the current study did not annotate to known protein coding genes. These two sites were outside of the ten differentially methylated regions for major depression identified by Roberson-Nay et al. [90] (39 major depression cases and 111 controls) and located away from the locations identified for MDD by Oh et al. [91] (103 cases and 97 controls) and Starnawska et al. [92] (724 individuals assessed for depression symptomatology score). A longitudinal methylome-wide association study of MDD (199 cases and 382 in remittance) identified six plausible CpG sites based on function [93], although none of the sites remained significant after applying correction for multiple testing. A further study of profile scores for MDD in

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

GS:SFHS found that prediction of MDD in an independent subset was possible, but was reliant on capturing lifestyle factors associated with MDD [38]. An association meta-analysis (7,948 individuals) identified 20 CpG sites with a suggestive association ($P < 1 \times 10^{-5}$) with depressive symptoms [23], of which one was nominally significant ($P = 0.048$) in a replication cohort of 3,308 individuals. None of these 20 suggestive sites were close to either of the associated sites observed in the current study. To reach replicable findings for CpG sites associated with MDD it is likely that larger sample sizes will be needed as has been demonstrated in genome-wide associations studies [73].

The initial methodology of the current study was to use a single linear regression, MWAS 1. It was only after observing genomic inflation in the signal for the time of year of birth phenotypes that MWAS 2 was also included. MWAS 2 accounts for methylome-wide correlational structure and makes additional adjustment for predicted cell type composition. The non-significance of any of the associated MWAS 1 variants in MWAS 2, suggests cell type composition as the likely cause of the inflation, which was confirmed by our analysis in the LBC. Therefore, correcting M-values for cell type composition using the Houseman algorithm [43] may not be adequate, as there remained a detectable effect of cell type composition in the GS:SFHS data. Furthermore, if we had only run MWAS 2 we would not have observed the effect of time of year on birth on cell type composition. In future, it may be preferable to run multiple regression models on DNA methylation data.

The results reported here are based on a single European population and their applicability to other countries and ancestries is unknown. GS:SFHS is a family-based sample drawn from the general population. Therefore, the more extreme phenotypes that have been analysed in the published literature would not have provided adequate power in the studied cohort. There was also no measure of childhood abuse or neglect collected in GS:SFHS and those phenotypes may provide additional avenues for investigation using a similar methodology to that used here. Finally, the DNA methylation data analysed was obtained from blood and the analysis of other tissue samples may reveal additional associations with the phenotypes examined.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In conclusion, there were plausible CpG sites associated with preterm birth, low birth weight and having a young parent in both MWAS 1 and MWAS 2. Further, one of the more interesting findings was the long-term impact of the time of year of birth on the blood cell type composition for neutrophils and lymphocytes. It was not possible to predict either MDD or BRS from methylation profile scores calculated from early life phenotypes. Although, there was significant predictive ability across the two sets of GS:SFHS data for birth month and having a young parent suggesting the reliability of the results for those phenotypes.

Code and Data availability

Methylome-wide summary statistics calculated from MWAS 1 and MWAS 2 for each of the assessed phenotypes are available from: TBC

Author contributions

D.M.H., A.M.McI., M.M., and C.M.L. conceived the research project. D.M.H, I.J.D, D.P, A.C, L.A, and A.M.McI determined the early life phenotypes used. M.C.B, C.A, R.M.W, and K.L.E applied quality control to the Generation Scotland DNA methylation data. D.M.H ran the analyses. O.P, R.A, M.C.B, B.J., M.J.A., P.F.S., A.M.McI., N.R.W. and C.M.L. provided expertise of methylome-wide association study methodology and statistical analysis. All authors commented on the manuscript.

Conflicts of interest

A.M.McI. has received research support from Eli Lilly and Company, Janssen and the Sackler Trust and speaker fees from Illumina and Janssen. C.M.L is a member of the Myriad Neuroscience R&D SAB.

Acknowledgements

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, including interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/8).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

LBC1921 funding has been received from the UK's Biotechnology and Biological Sciences Research Council (BBSRC) (15/SAG09977, wave 1), a Royal Society-Wolfson Research Merit Award to IJD (wave 2), the Chief Scientist Office (CSO) of the Scottish Government's Health Directorates (CZG/3/2/79, post-wave 1 questionnaire study; CZB/4/505, wave 3; ETM/55, wave 4), and the UK's Medical Research Council (MRC) Centenary Early Career Award to Dr Tom Booth (wave 5). Funding for LBC1936 has been received from Research Into Ageing (Programme grant 251; wave 1), and Age UK (Disconnected Mind Programme grant) and the UK's Medical Research Council (G0701120, wave 2; G1001245, wave 3; MR/M013111/1, wave 4). The Alzheimer Scotland Dementia Research Centre funded LBC1936 dementia ascertainment. BBSRC funded whole-genome sequencing of both cohorts. Wellcome, the University of Edinburgh, the University of Queensland, and Age UK funded DNA methylation analysis in both cohorts.

D.M.H is supported by a Sir Henry Wellcome Postdoctoral Fellowship (Reference 213674/Z/18/Z) and a 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Ref: 27404). A.M.McI is supported by Wellcome Trust (104036/Z/14/Z and 216767/Z/19/Z) and UKRI MRC funding (MC PC 17209 and MR/S035818/1). C.A is supported by the Medical Research Council (MRC) UK (grants MC_PC_U127592696 and MC_PC_U127561128). I.J.D. was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology, which was funded by the Medical Research Council and the Biotechnology and Biological Sciences Research Council (MR/K026992/1). R.E.M is supported by Alzheimer's Research UK major project grant ARUK-PG2017B-10. N.R.W. acknowledges NMHRC grants 1078901 and 1087889. L.A is the Mental Health Leadership Fellow for the ESRC. C.M.L acknowledges MRC grant MR/N015746/1. This investigation represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

1. Kessler, R.C., C.G. Davis, and K.S. Kendler, *Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey.* Psychological Medicine, 1997. **27**(5): p. 1101-1119.
2. Bellis, M.A., H. Lowey, N. Leckenby, K. Hughes, and D. Harrison, *Adverse childhood experiences: retrospective study to determine their impact on adult health behaviours and health outcomes in a UK population.* Journal of Public Health, 2013. **36**(1): p. 81-91.
3. Horwitz, A.V., C.S. Widom, J. McLaughlin, and H.R. White, *The impact of childhood abuse and neglect on adult mental health: A prospective study.* Journal of Health and Social Behavior, 2001. **42**(2): p. 184-201.
4. Molnar, B.E., L.F. Berkman, and S.L. Buka, *Psychopathology, childhood sexual abuse and other childhood adversities: relative links to subsequent suicidal behaviour in the US.* Psychol Med, 2001. **31**(6): p. 965-77.
5. Springer, K.W., J. Sheridan, D. Kuo, and M. Carnes, *Long-term physical and mental health consequences of childhood physical abuse: Results from a large population-based sample of men and women.* Child Abuse & Neglect, 2007. **31**(5): p. 517-530.
6. Morgan, C., C. Gayer-Anderson, S. Beards, K. Hubbard, V. Mondelli, M. Di Forti, R.M. Murray, et al., *Threat, hostility and violence in childhood and later psychotic disorder: population-based case–control study.* The British Journal of Psychiatry, 2020. **217**(4): p. 575-582.
7. Mathewson, K.J., C.H.T. Chow, K.G. Dobson, E.I. Pope, L.A. Schmidt, and R.J. Van Lieshout, *Mental health of extremely low birth weight survivors: A systematic review and meta-analysis.* Psychological Bulletin, 2017. **143**(4): p. 347-383.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

8. Johnson, S. and N. Marlow, *Growing up after extremely preterm birth: Lifespan mental health outcomes.* Seminars in Fetal and Neonatal Medicine, 2014. **19**(2): p. 97-104.
9. Nosarti, C., A. Reichenberg, R.M. Murray, S. Cnattingius, M.P. Lambe, L. Yin, J. MacCabe, et al., *Preterm birth and psychiatric disorders in young adult life.* Archives of General Psychiatry, 2012. **69**(6): p. 610-617.
10. Disanto, G., J.M. Morahan, M.V. Lacey, G.C. DeLuca, G. Giovannoni, G.C. Ebers, and S.V. Ramagopalan, *Seasonal distribution of psychiatric births in England.* PLOS ONE, 2012. **7**(4): p. e34866.
11. Joiner, T.E., J.J. Pfaff, J.G. Acres, and F. Johnson, *Birth month and suicidal and depressive symptoms in Australians born in the Southern vs. the Northern hemisphere.* Psychiatry Research, 2002. **112**(1): p. 89-92.
12. Schnittker, J., *Season of birth and depression in adulthood: Revisiting historical forerunner evidence for in-utero effects.* SSM - Population Health, 2018. **4**: p. 307-316.
13. McGrath, J.J., L. Petersen, E. Agerbo, O. Mors, P.B. Mortensen, and C.B. Pedersen, *A comprehensive assessment of parental age and psychiatric disorders.* JAMA Psychiatry, 2014. **71**(3): p. 301-309.
14. Mäkikyrö, T., A. Sauvola, J. Moring, J. Veijola, P. Nieminen, M.R. Järvelin, and M. Isohanni, *Hospital-treated psychiatric disorders in adults with a single-parent and two-parent family background: a 28-year follow-up of the 1966 Northern Finland Birth Cohort.* Fam Process, 1998. **37**(3): p. 335-44.
15. Tost, H., F.A. Champagne, and A. Meyer-Lindenberg, *Environmental influence in the brain, human welfare and mental health.* Nature Neuroscience, 2015. **18**(10): p. 1421-1431.
16. Gruebner, O., M.A. Rapp, M. Adli, U. Kluge, S. Galea, and A. Heinz, *Cities and mental health.* Deutsches Arzteblatt international, 2017. **114**(8): p. 121-127.
17. Okkels, N., C.B. Kristiansen, P. Munk-Jørgensen, and N. Sartorius, *Urban mental health: challenges and perspectives.* Current Opinion in Psychiatry, 2018. **31**(3).
18. Feil, R. and M.F. Fraga, *Epigenetics and the environment: emerging patterns and implications.* Nature Reviews Genetics, 2012. **13**(2): p. 97-109.
19. Klengel, T., J. Pape, E.B. Binder, and D. Mehta, *The role of DNA methylation in stress-related psychiatric disorders.* Neuropharmacology, 2014. **80**: p. 115-132.
20. Tsankova, N., W. Renthal, A. Kumar, and E.J. Nestler, *Epigenetic regulation in psychiatric disorders.* Nature Reviews Neuroscience, 2007. **8**(5): p. 355-367.
21. Fries, G.R., Q. Li, B. McAlpin, T. Rein, C. Walss-Bass, J.C. Soares, and J. Quevedo, *The role of DNA methylation in the pathophysiology and treatment of bipolar disorder.* Neuroscience & Biobehavioral Reviews, 2016. **68**: p. 474-488.
22. Smigielski, L., V. Jagannath, W. Rössler, S. Walitza, and E. Grünblatt, *Epigenetic mechanisms in schizophrenia and other psychotic disorders: a systematic review of empirical human findings.* Molecular Psychiatry, 2020. **25**(8): p. 1718-1748.
23. Story Jovanova, O., I. Nedeljkovic, D. Spieler, R.M. Walker, C. Liu, M. Luciano, J. Bressler, et al., *DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multiethnic epigenome-wide studies.* JAMA Psychiatry, 2018. **75**(9): p. 949-959.
24. Park, C., J.D. Rosenblat, E. Brietzke, Z. Pan, Y. Lee, B. Cao, H. Zuckerman, et al., *Stress, epigenetics and depression: A systematic review.* Neuroscience & Biobehavioral Reviews, 2019. **102**: p. 139-152.
25. Border, R., E.C. Johnson, L.M. Evans, A. Smolen, N. Berley, P.F. Sullivan, and M.C. Keller, *No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples.* American Journal of Psychiatry, 2019. **176**(5): p. 376-387.
26. Smith, B.H., A. Campbell, P. Linksted, B. Fitzpatrick, C. Jackson, S.M. Kerr, I.J. Deary, et al., *Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its*

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

participants and their potential for genetic research on health and illness. International Journal of Epidemiology, 2013. **42**(3): p. 689-700.

27. Navrady, L.B., M.K. Wolters, D.J. MacIntyre, T.K. Clarke, A.I. Campbell, A.D. Murray, K.L. Evans, et al., *Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS).* International journal of epidemiology, 2018. **47**(1): p. 13-14g.
28. World Health Organization, *Fact sheets: preterm birth*. 2018: https://www.who.int/news-room/fact-sheets/detail/preterm-birth.
29. Bonellie, S., J. Chalmers, R. Gray, I. Greer, S. Jarvis, and C. Williams, *Centile charts for birthweight for gestational age for Scottish singleton births.* BMC Pregnancy and Childbirth, 2008. **8**(1): p. 5.
30. National Records of Scotland, *Population Estimates*. 2019: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates.
31. First, M.B., R.L. Spitzer, Gibbon Miriam., and J.B.W. Williams, *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)* 2002.
32. Fernandez-Pujals, A.M., M.J. Adams, P. Thomson, A.G. McKechanie, D.H.R. Blackwood, B.H. Smith, A.F. Dominiczak, et al., *Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS).* PLoS ONE, 2015. **10**(11): p. e0142197.
33. Smith, B.W., J. Dalen, K. Wiggins, E. Tooley, P. Christopher, and J. Bernard, *The brief resilience scale: Assessing the ability to bounce back.* International Journal of Behavioral Medicine, 2008. **15**(3): p. 194-200.
34. Stekhoven, D.J. and P. Bühlmann, *MissForest—non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2011. **28**(1): p. 112-118.
35. Revelle, W. *psych: Procedures for psychological, psychometric, and personality research.* 2020: Northwestern University, Evanston, Illinois. R package version 2.0.9
36. Ziyatdinov, A., M. Vázquez-Santiago, H. Brunel, A. Martinez-Perez, H. Aschard, and J.M. Soria, *lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals.* BMC Bioinformatics, 2018. **19**(1): p. 68.
37. Sinnwell, J.P., T.M. Therneau, and D.J. Schaid, *The kinship2 R package for pedigree data.* Human Heredity, 2014. **78**(2): p. 91-93.
38. Barbu, M.C., X. Shen, R.M. Walker, D.M. Howard, K.L. Evans, H.C. Whalley, D.J. Porteous, et al., *Epigenetic prediction of major depressive disorder.* Molecular Psychiatry, 2020. **TBC** (https://doi.org/10.1038/s41380-020-0808-3).
39. McCartney, D.L., R.F. Hillary, A.J. Stevenson, S.J. Ritchie, R.M. Walker, Q. Zhang, S.W. Morris, et al., *Epigenetic prediction of complex traits and death.* Genome Biology, 2018. **19**(1): p. 136.
40. Du, P., X. Zhang, C.-C. Huang, N. Jafari, W.A. Kibbe, L. Hou, and S.M. Lin, *Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**(1): p. 587.
41. Xia, C., C. Amador, J. Huffman, H. Trochet, A. Campbell, D. Porteous, S. Generation, et al., *Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation.* PLOS Genetics, 2016. **12**(2): p. e1005804.
42. Zeng, Y., C. Amador, C. Xia, R. Marioni, D. Sproul, R.M. Walker, S.W. Morris, et al., *Parent of origin genetic effects on methylation in humans are common and influence complex trait variation.* Nature Communications, 2019. **10**(1): p. 1383.
43. Houseman, E.A., W.P. Accomando, D.C. Koestler, B.C. Christensen, C.J. Marsit, H.H. Nelson, J.K. Wiencke, et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**(1): p. 86.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

44. Min, J.L., G. Hemani, G. Davey Smith, C. Relton, and M. Suderman, *Meffil: efficient normalization and analysis of very large DNA methylation datasets.* Bioinformatics (Oxford, England), 2018. **34**(23): p. 3983-3989.
45. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Statistical Applications in Genetics and Molecular Biology, 2004. **3**(1).
46. Lê, S., J. Josse, and F. Husson, *FactoMineR: An R package for multivariate analysis.* 2008, 2008. **25**(1): p. 18.
47. Zhang, F., W. Chen, Z. Zhu, Q. Zhang, M.F. Nabais, T. Qi, I.J. Deary, et al., *OSCA: a tool for omic-data-based complex trait analysis.* Genome Biology, 2019. **20**(1): p. 107.
48. Starr, J.M. and I.J. Deary, *Sex differences in blood cell counts in the Lothian Birth Cohort 1921 between 79 and 87 years.* Maturitas, 2011. **69**(4): p. 373-376.
49. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: Springer-Verlag New York.
50. Gjessing, H.K. and R.T. Lie, *Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes.* Annals of Human Genetics, 2006. **70**(3): p. 382-396.
51. Phipson, B., J. Maksimovic, and A. Oshlack, *missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform.* Bioinformatics, 2015. **32**(2): p. 286-288.
52. Lee, S.H., M.E. Goddard, N.R. Wray, and P.M. Visscher, *A better coefficient of determination for genetic profile analysis.* Genetic Epidemiology, 2012. **36**(3): p. 214-224.
53. Taylor, A.M., A. Pattie, and I.J. Deary, *Cohort profile update: the Lothian Birth Cohorts of 1921 and 1936.* International Journal of Epidemiology, 2018. **47**(4): p. 1042-1042r.
54. Van Lieshout, R.J., M.A. Ferro, L.A. Schmidt, M.H. Boyle, S. Saigal, K.M. Morrison, and K.J. Mathewson, *Trajectories of psychopathology in extremely low birth weight survivors from early adolescence to adulthood: a 20-year longitudinal study.* Journal of Child Psychology and Psychiatry, 2018. **59**(11): p. 1192-1200.
55. Carreira-Barbosa, F., M.L. Concha, M. Takeuchi, N. Ueno, S.W. Wilson, and M. Tada, *Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish.* Development, 2003. **130**(17): p. 4037-4046.
56. Okuda, H., S. Miyata, Y. Mori, and M. Tohyama, *Mouse Prickle1 and Prickle2 are expressed in postmitotic neurons and promote neurite outgrowth.* FEBS Letters, 2007. **581**(24): p. 4754-4760.
57. Tao, H., K.-i. Inoue, H. Kiyonari, A.G. Bassuk, J.D. Axelrod, H. Sasaki, S. Aizawa, et al., *Nuclear localization of Prickle2 is required to establish cell polarity during early mouse embryogenesis.* Developmental Biology, 2012. **364**(2): p. 138-148.
58. Jenny, A., R.S. Darken, P.A. Wilson, and M. Mlodzik, *Prickle and Strabismus form a functional complex to generate a correct axis during planar cell polarity signaling.* The EMBO Journal, 2003. **22**(17): p. 4409-4420.
59. Tarui, T., A. Kim, A. Flake, L. McClain, J.D. Stratigis, I. Fried, R. Newman, et al., *Amniotic fluid transcriptomics reflects novel disease mechanisms in fetuses with myelomeningocele.* American Journal of Obstetrics and Gynecology, 2017. **217**(5): p. 587.e1-587.e10.
60. Purisch, S.E., E.A. DeFranco, L.J. Muglia, A.O. Odibo, and D.M. Stamilio, *Preterm birth in pregnancies complicated by major congenital malformations: a population-based study.* American Journal of Obstetrics and Gynecology, 2008. **199**(3): p. 287.e1-287.e8.
61. Zhao, R., J.M. Chinai, S. Buhl, L. Scandiuzzi, A. Ray, H. Jeon, K.C. Ohaegbulam, et al., *HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function.* Proceedings of the National Academy of Sciences, 2013. **110**(24): p. 9879-9884.
62. Koirala, P., M.E. Roth, J. Gill, J.M. Chinai, M.R. Ewart, S. Piperdi, D.S. Geller, et al., *HHLA2, a member of the B7 family, is expressed in human osteosarcoma and is associated with metastases and worse survival.* Scientific Reports, 2016. **6**(1): p. 31154.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

63. Dubielecka, P.M., K.I. Ladwein, X. Xiong, I. Migeotte, A. Chorzalska, K.V. Anderson, J.A. Sawicki, et al., *Essential role for ABI1 in embryonic survival and WAVE2 complex integrity.* Proceedings of the National Academy of Sciences, 2011. **108**(17): p. 7022-7027.
64. Ring, C., M.H. Ginsberg, J. Haling, and A.M. Pendergast, *Abl-interactor-1 (ABI1) has a role in cardiovascular and placental development and is a binding partner of the α4 integrin.* Proceedings of the National Academy of Sciences, 2011. **108**(1): p. 149-154.
65. Sparrow, S., J.R. Manning, J. Cartier, D. Anblagan, M.E. Bastin, C. Piyasena, R. Pataky, et al., *Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function.* Translational Psychiatry, 2016. **6**(1): p. e716-e716.
66. Lund, L.K., T. Vik, S. Lydersen, G.C.C. Løhaugen, J. Skranes, A.-M. Brubakk, and M.S. Indredavik, *Mental health, quality of life and social relations in young adults born with low birth weight.* Health and Quality of Life Outcomes, 2012. **10**(1): p. 146.
67. Husby, I.M., K.M.-T. Stray, A. Olsen, S. Lydersen, M.S. Indredavik, A.-M. Brubakk, J. Skranes, et al., *Long-term follow-up of mental health, health-related quality of life and associations with motor skills in young adults born preterm with very low birth weight.* Health and Quality of Life Outcomes, 2016. **14**(1): p. 56.
68. Chen, Z., M. Montcouquiol, R. Calderon, N.A. Jenkins, N.G. Copeland, M.W. Kelley, and K. Noben-Trauth, *JXC1/SOBP, encoding a nuclear zinc finger protein, is critical for cochlear growth, cell fate, and patterning of the organ of corti.* The Journal of Neuroscience : the official journal of the Society for Neuroscience, 2008. **28**(26): p. 6633-6641.
69. Küpers, L.K., C. Monnereau, G.C. Sharp, P. Yousefi, L.A. Salas, A. Ghantous, C.M. Page, et al., *Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight.* Nature Communications, 2019. **10**(1): p. 1893.
70. Roepman, R., N. Bernoud-Hubac, D.E. Schick, A. Maugeri, W. Berger, H.-H. Ropers, F.P.M. Cremers, et al., *The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors.* Human Molecular Genetics, 2000. **9**(14): p. 2095-2105.
71. Cristobal, R. and J.S. Oghalai, *Hearing loss in children with very low birth weight: current review of epidemiology and pathophysiology.* Archives of Disease in Childhood - Fetal and Neonatal Edition, 2008. **93**(6): p. F462-F468.
72. O'Connor, A.R., C.M. Wilson, and A.R. Fielder, *Ophthalmological problems associated with preterm birth.* Eye, 2007. **21**(10): p. 1254-1260.
73. Howard, D.M., M.J. Adams, T.-K. Clarke, J.D. Hafferty, J. Gibson, M. Shirali, J.R.I. Coleman, et al., *Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions.* Nature Neuroscience, 2019. **22**(3): p. 343-352.
74. Madden, R.A., D.L. McCartney, R.M. Walker, R.F. Hillary, M.L. Bermingham, K. Rawlik, S.W. Morris, et al., *Birth weight associations with DNA methylation differences in an adult population.* Epigenetics, 2020: p. 1-14.
75. Dopico, X.C., M. Evangelou, R.C. Ferreira, H. Guo, M.L. Pekalski, D.J. Smyth, N. Cooper, et al., *Widespread seasonal gene expression reveals annual differences in human immunity and physiology.* Nature Communications, 2015. **6**(1): p. 7000.
76. Thysen, A.H., M.A. Rasmussen, E. Kreiner-Møller, J.M. Larsen, N.V. Følsgaard, K. Bønnelykke, J. Stokholm, et al., *Season of birth shapes neonatal immune function.* Journal of Allergy and Clinical Immunology, 2016. **137**(4): p. 1238-1246.e13.
77. Disanto, G., C.T. Watson, U.C. Meier, G.C. Ebers, G. Giovannoni, and S.V. Ramagopalan, *Month of birth and thymic output.* JAMA Neurology, 2013. **70**(4): p. 527-528.
78. Lyssenko, V., C.L.F. Nagorny, M.R. Erdos, N. Wierup, A. Jonsson, P. Spégel, M. Bugliani, et al., *Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion.* Nature Genetics, 2009. **41**(1): p. 82-88.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

79. Staiger, H., F. Machicao, S.A. Schäfer, K. Kirchhoff, K. Kantartzis, M. Guthoff, G. Silbernagel, et al., *Polymorphisms within the novel type 2 diabetes risk locus MTNR1B determine β-Cell function.* PLOS ONE, 2008. **3**(12): p. e3962.
80. de Sá Machado Araújo, G., R. da Silva Francisco Junior, C. dos Santos Ferreira, P.T. Mozer Rodrigues, D. Terra Machado, T. Louvain de Souza, J. Teixeira de Souza, et al., *Maternal 5mCpG imprints at the PARD6G-AS1 and GCSAML differentially methylated regions are decoupled from parent-of-origin expression effects in multiple human tissues.* Frontiers in Genetics, 2018. **9**(36).
81. Bjarnason, T., P. Bendtsen, A.M. Arnarsson, I. Borup, R.J. Iannotti, P. Löfstedt, I. Haapasalo, et al., *Life satisfaction among children in different family structures: A comparative study of 36 Western societies.* Children & Society, 2012. **26**(1): p. 51-62.
82. Steinbach, A., L. Augustijn, and G. Corkadi, *Joint physical custody and adolescents' life satisfaction in 37 North American and European countries.* Family Process, 2020. **TBC** (https://doi.org/10.1111/famp.12536).
83. Sundquist, K., G. Frank, and J. Sundquist, *Urbanisation and incidence of psychosis and depression: Follow-up study of 4.4 million women and men in Sweden.* British Journal of Psychiatry, 2004. **184**(4): p. 293-298.
84. Hoare, E., F. Jacka, and M. Berk, *The impact of urbanization on mood disorders: an update of recent evidence.* Current Opinion in Psychiatry, 2019. **32**(3): p. 198-203.
85. DeVylder, J.E., I. Kelleher, M. Lalane, H. Oh, B.G. Link, and A. Koyanagi, *Association of urbanicity with psychosis in low- and middle-income countries.* JAMA Psychiatry, 2018. **75**(7): p. 678-686.
86. Kessler, R.C., P. Berglund, O. Demler, R. Jin, D. Koretz, K.R. Merikangas, A.J. Rush, et al., *The epidemiology of major depressive disorder: results from the National Comorbidity Survey replication (NCS-R).* JAMA, 2003. **289**(23): p. 3095-3105.
87. Melis, G., E. Gelormino, G. Marra, E. Ferracin, and G. Costa, *The effects of the urban built environment on mental health: A cohort study in a large Northern Italian city.* International Journal of Environmental Research and Public Health, 2015. **12**(11): p. 14898-14915.
88. Bakolis, I., R. Hammoud, R. Stewart, S. Beevers, D. Dajnak, S. MacCrimmon, M. Broadbent, et al., *Mental health consequences of urban air pollution: prospective population-based longitudinal survey.* Social Psychiatry and Psychiatric Epidemiology, 2020. **TBC** (https://doi.org/10.1007/s00127-020-01966-x).
89. Lorant, V., D. Deliège, W. Eaton, A. Robert, P. Philippot, and M. Ansseau, *Socioeconomic inequalities in depression: a meta-analysis.* American Journal of Epidemiology, 2003. **157**(2): p. 98-112.
90. Roberson-Nay, R., D.M. Lapato, A.R. Wolen, E.E. Lancaster, B.T. Webb, B. Verhulst, J.M. Hettema, et al., *An epigenome-wide association study of early-onset major depression in monozygotic twins.* Translational Psychiatry, 2020. **10**(1): p. 301.
91. Oh, G., S.-C. Wang, M. Pal, Z.F. Chen, T. Khare, M. Tochigi, C. Ng, et al., *DNA modification study of major depressive disorder: beyond locus-by-locus comparisons.* Biological Psychiatry, 2015. **77**(3): p. 246-255.
92. Starnawska, A., Q. Tan, M. Soerensen, M. McGue, O. Mors, A.D. Børglum, K. Christensen, et al., *Epigenome-wide association study of depression symptomatology in elderly monozygotic twins.* Translational Psychiatry, 2019. **9**(1): p. 214.
93. Clark, S.L., M.W. Hattab, R.F. Chan, A.A. Shabalin, L.K.M. Han, M. Zhao, J.H. Smit, et al., *A methylation study of long-term depression risk.* Molecular Psychiatry, 2020. **25**(6): p. 1334-1343.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figures

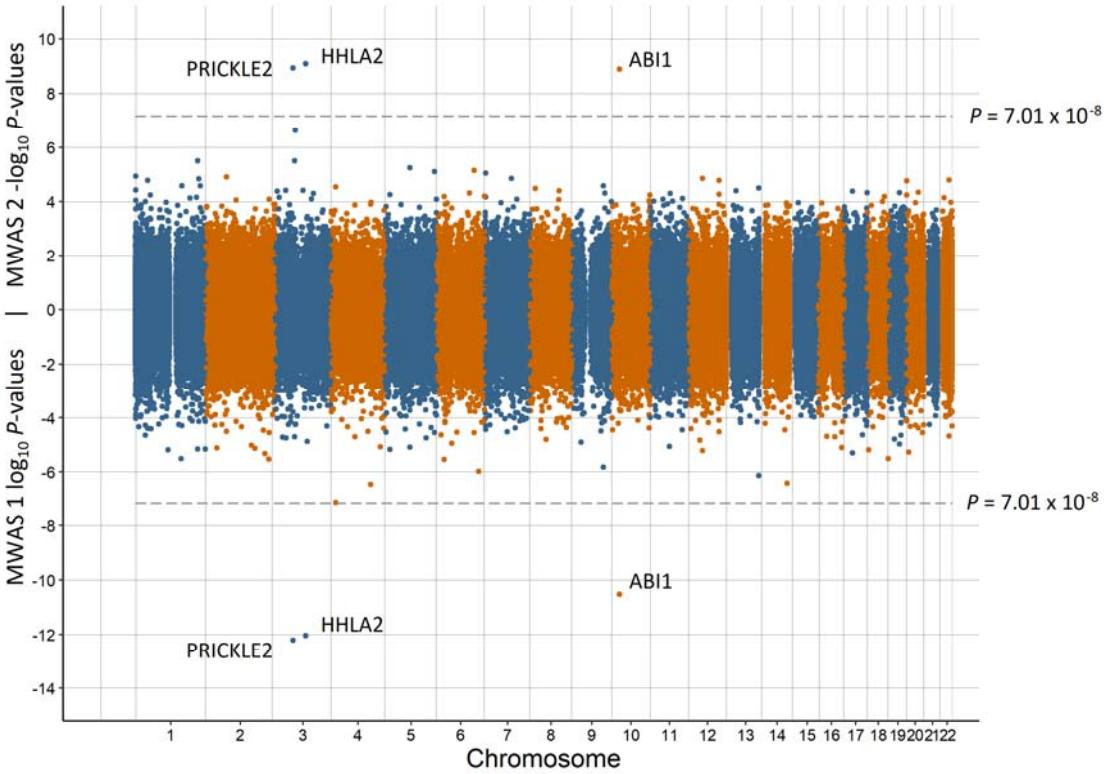

Figure 1. Miami plot of the observed *P*-values of each CpG site for an association with preterm birth

Log_{10} *P*-values are shown for MWAS 1 and $-log_{10}$ *P*-values are shown for MWAS 2. The dotted lines indicate methylome-wide significance ($P = 7.01 \times 10^{-8}$). The annotation of genes for significant sites is reported by missMethyl for MWAS 1 and by OSCA for MWAS 2

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

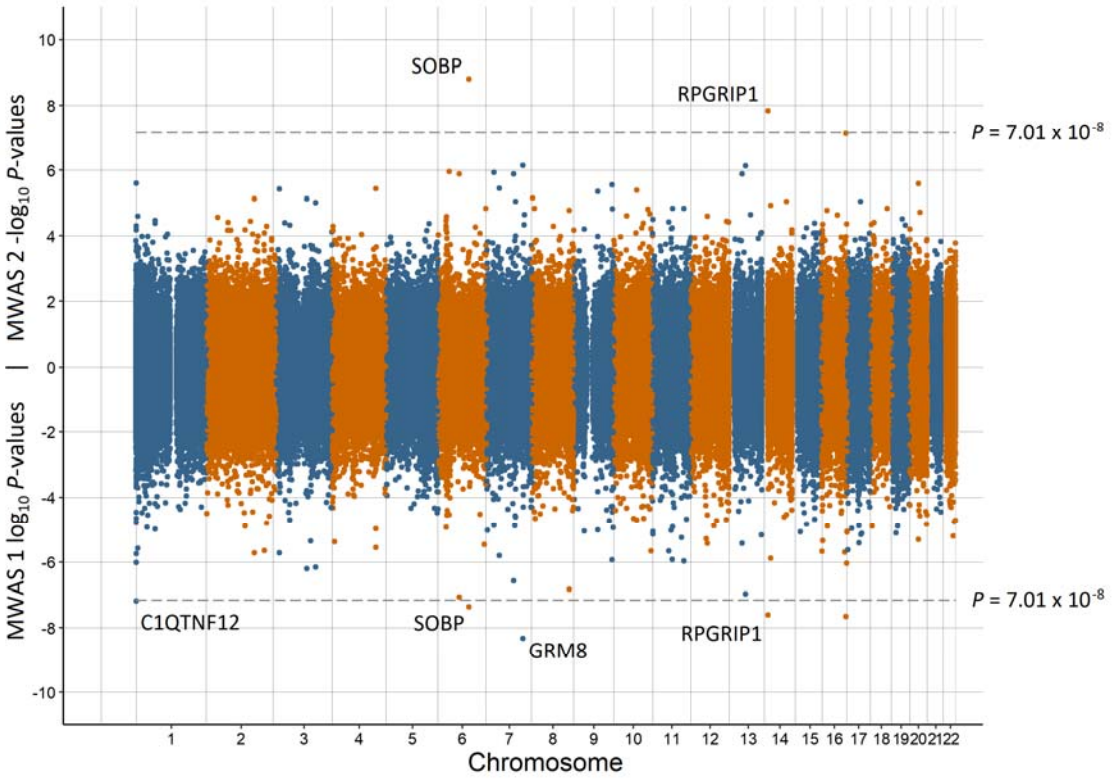

Figure 2. Miami plot of the observed *P*-values of each CpG site for an association with low birth weight

Log_{10} *P*-values are shown for MWAS 1 and $-log_{10}$ *P*-values are shown for MWAS 2. The dotted lines indicate methylome-wide significance ($P = 7.01 \times 10^{-8}$). The annotation of genes for significant sites is reported by missMethyl for MWAS 1 and by OSCA for MWAS 2

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

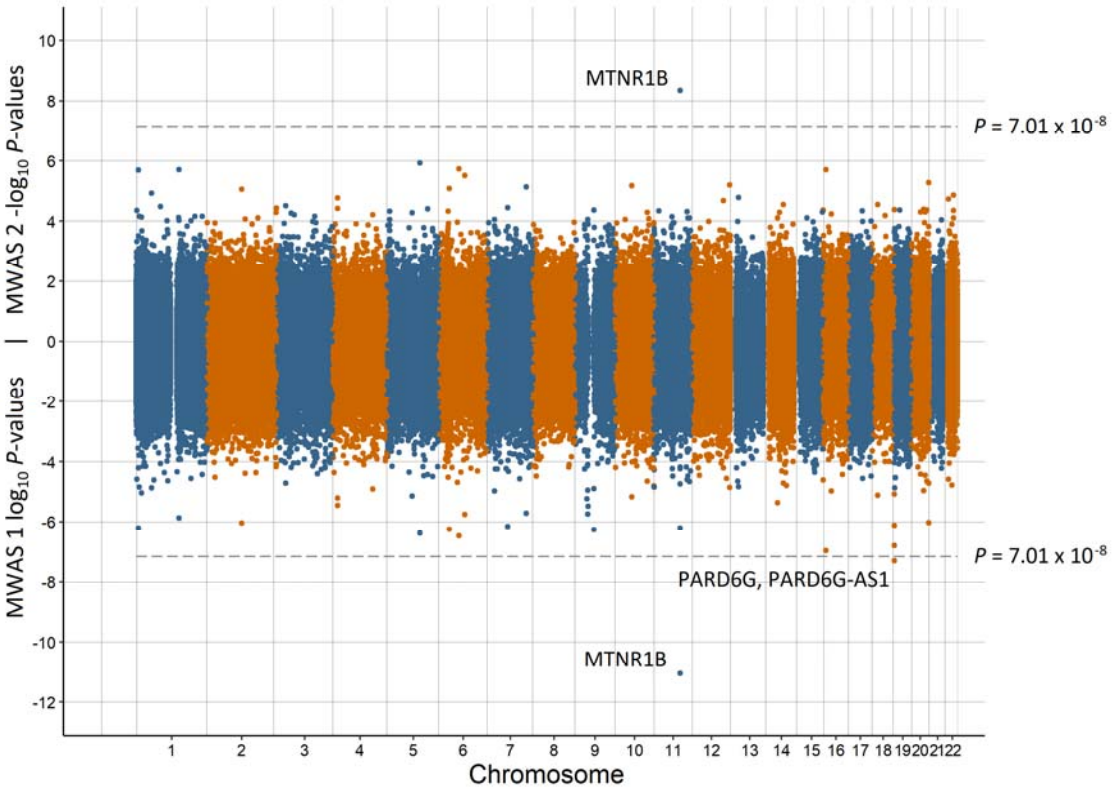

Figure 3. Miami plot of the observed *P*-values of each CpG site for an association with having a young parent

Log_{10} *P*-values are shown for MWAS 1 and $-log_{10}$ *P*-values are shown for MWAS 2. The dotted lines indicate methylome-wide significance ($P = 7.01 \times 10^{-8}$). The annotation of genes for significant sites is reported by missMethyl for MWAS 1 and by OSCA for MWAS 2

medRxiv preprint doi: https://doi.org/10.1101/2021.03.10.21253201; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

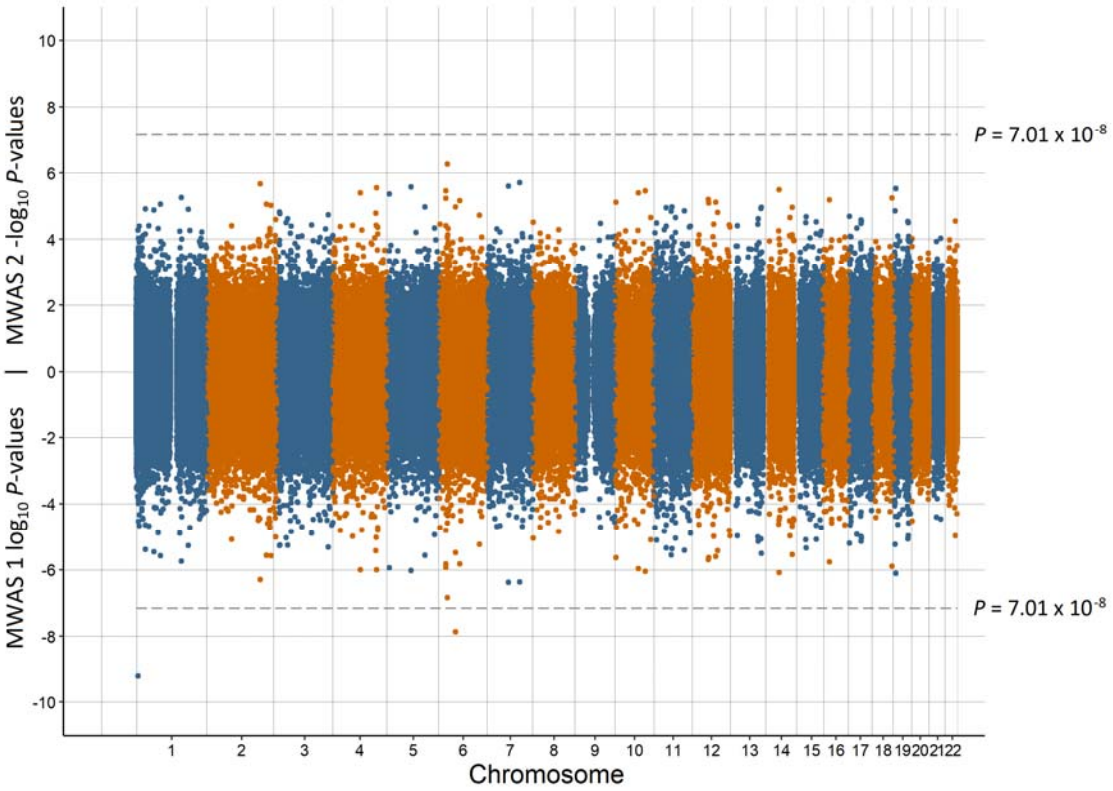

Figure 4. Miami plot of the observed *P*-values of each CpG site for an association with major depressive disorder

Log_{10} *P*-values are shown for MWAS 1 and $-log_{10}$ *P*-values are shown for MWAS 2. The dotted lines indicate methylome-wide significance ($P = 7.01 \times 10^{-8}$). Significant sites did not annotate to any genes.