

1 **Birth history is associated with whole-blood and T-cell methylation patterns in relapse**  
2 **onset multiple sclerosis**

3

4 Maria Pia Campagna<sup>1</sup>

5 Alexandre Xavier<sup>2,3</sup>

6 Jim Stankovich<sup>1</sup>

7 Vicki Maltby<sup>3</sup>

8 Mark Slee<sup>4</sup>

9 Trevor Kilpatrick<sup>5,6</sup>

10 Rodney J Scott<sup>2,3</sup>

11 Helmut Butzkueven<sup>1,7,8</sup>

12 Jeannette Lechner-Scott<sup>3,9\*</sup>

13 Rodney Lea<sup>3,10\*</sup>

14 Vilija Jokubaitis<sup>1,5,6,8\*</sup>

15

16 1. Department of Neuroscience, Monash University, Melbourne, Victoria, Australia

17 2. School of Biomedical Sciences and Pharmacy, University of Newcastle

18 3. University of Newcastle, Hunter Medical Research Institute, Newcastle, New South  
19 Wales, Australia

20 4. College of Medicine and Public Health, Flinders University, Adelaide, Australia

21 5. Department of Medicine, University of Melbourne, Melbourne, Victoria, Australia

22 6. Department of Neurology, Royal Melbourne Hospital, Melbourne, Victoria, Australia

23 7. Neurology Department, Eastern Health, Melbourne, Victoria, Australia

24 8. Neurology Department, Alfred Health, Melbourne, Victoria, Australia

25 9. Neurology Department, Hunter New England, Newcastle, New South Wales, Australia

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

- 26 10. Queensland University of Technology, Brisbane, Queensland, Australia
- 27
- 28 \*Equal senior authors
- 29 Corresponding author: Maria Pia Campagna ([maria.campagna@monash.edu](mailto:maria.campagna@monash.edu))
- 30 Key words: multiple sclerosis, pregnancy, DNA methylation, epigenetics, neural plasticity

31 **Abstract**

32 **Background:** Pregnancy in women with multiple sclerosis (MS) is associated with a  
33 reduction of long-term disability progression. The mechanism that drives this effect is  
34 unknown, but converging evidence suggests a role for epigenetic mechanisms altering  
35 immune and/or central nervous system function. **Objectives:** We aimed to identify whole  
36 blood and immune cell-specific DNA methylation patterns associated with parity in relapse-  
37 onset multiple sclerosis. **Methods:** We compared whole-blood methylation patterns between  
38 96 matched pairs of nulligravida and parous females with MS (n=192). Parity was defined as  
39 at least one term or pre-term birth, and nulligravida was defined as no prior pregnancies.  
40 Methylation was measured with Illumina EPIC arrays, and data was pre-processed and  
41 statistically analysed using the *ChAMP* package. Cell-type proportions were estimated using  
42 the *EpiDISH* package, and cell-specific analysis conducted using linear regression. Gene-set  
43 enrichment analysis (GSEA) was performed with ToppGene API and GOMeth. Methylation  
44 age was calculated with the *methyAge* package. Methylation age acceleration (MAA) was  
45 calculated by regressing methylation age on chronological age. FDR<0.05 was used to assess  
46 significance. **Results:** The median time from last pregnancy to blood collection was 16.66  
47 years (range = 1.45 – 44.42 years). We identified 903 differentially methylated positions  
48 (DMPs) in whole blood; 365 were hypomethylated and 528 were hypermethylated in parous  
49 women. We further identified two differentially methylated regions (DMRs) in *CRYGN* on  
50 Chromosome 7 and an intergenic region on Chromosome 15. There were four and eight cell  
51 type specific DMPs in CD4+ and CD8+ cells, respectively. Differentially methylated genes  
52 were enriched in neuronal plasticity pathways. Parity was associated with reduced MAA by a  
53 mean of 1.44 to 2.27 years using the PhenoAge (p = 0.002) and GrimAge (p = 0.005)  
54 algorithms. **Conclusion:** Whole-blood methylation patterns are associated with birth history  
55 in females with relapse-onset multiple sclerosis. We found enrichment of differentially

- 56 methylated genes encoding neuronal processes and reduced MAA in parous women. These
- 57 methylation changes could mediate the long-term benefit of pregnancy for disease
- 58 progression in multiple sclerosis.

## 59 **Introduction**

60 Multiple sclerosis (MS) is most prevalent in females, with a sex ratio of 3:1<sup>1</sup>. It is frequently  
61 diagnosed between 20-40 years of age, the prime reproductive years for women.

62 Understanding the effect of pregnancy on disease activity and progression is a priority for  
63 women with MS (wwMS) and their care teams.

64

65 Pregnancy has been shown to reduce MS relapse rates and short- and long-term disability  
66 outcomes in wwMS, regardless of the outcome of the pregnancy<sup>2</sup>. With respect to long-term  
67 outcomes, in a study of 2,557 wwMS, a history of childbirth delayed the onset of a clinically  
68 isolated syndrome (CIS, the first demyelinating event indicative of a future MS diagnosis) by  
69 3.4 years<sup>3</sup>. Additionally, in the largest real-world study of 1830 wwMS, one or more  
70 pregnancies after MS onset were associated with lower disability scores after ten years<sup>4</sup>.

71 Notably, the protective effect of pregnancy in this cohort was four-fold greater than that of  
72 first line DMT exposure in the same timeframe<sup>4</sup>. The biological mechanisms underpinning  
73 these long-term effects of pregnancy are not understood. As the effect of pregnancy on age at  
74 CIS onset<sup>3</sup> and disability progression extends for years beyond birth<sup>4</sup>, it cannot be explained  
75 exclusively by transient hormonal and immunological changes during pregnancy.

76

77 Epigenetic mechanisms regulate gene expression in a dynamic and reversible manner. DNA  
78 methylation is a key epigenetic mechanism. The presence or absence of a methyl group on  
79 cytosine-phosphate-guanine (CpG) dinucleotides generally activates or represses gene  
80 transcription, respectively. Epigenetic mechanisms are influenced by life events and  
81 environmental factors, including the multitude of physiological and hormonal changes of  
82 pregnancy. DNA methylation enzymes are specifically influenced by estrogen signalling,  
83 which increases in pregnancy and peaks in the third trimester. Converging evidence outlines a

84 role for DNA methylation in the effect of pregnancy on outcomes in wwMS through altering  
85 immune and central nervous system (CNS) function: 1) estrogen signalling influences DNA  
86 methylation enzymes<sup>5</sup>, 2) pregnancy has been shown to reduce immune epigenetic age in  
87 women without MS<sup>6</sup>, and 3) pregnancy induces changes in the expression of immune-  
88 activation<sup>7</sup> and axon-guidance<sup>8</sup> genes in wwMS for up to 19 years after pregnancy. However,  
89 no epigenome-wide association study (EWAS) of parity in wwMS has been reported to date.

90

91 The objective of this study was to understand the long-term impact of parity on DNA  
92 methylation patterns in women with relapse-onset multiple sclerosis. We first sought to  
93 identify whole-blood and immune cell-specific DNA methylation patterns, across autosomes,  
94 associated with parity. Secondly, we aimed to compare methylation age acceleration (MAA)  
95 between nulligravida and parous wwMS, to determine whether reductions in MAA reported  
96 in health were also evident in an MS cohort.

97

98

## 99 **Materials and Methods**

### 100 Ethics approvals

101 Ethics approval for the collection of demographic, clinical, treatment and pregnancy history  
102 data via the MSBase Registry<sup>9</sup> was obtained from the Alfred Health Human Research Ethics  
103 Committee (528/12), and institutional review boards at all participating centres. Approval for  
104 the collection of genetic data was obtained from the Australian National Mutual Acceptance  
105 Scheme (HREC/13/MH/189). Written informed consent was obtained from participants as  
106 per local laws at each study site.

107

### 108 Clinical data collection

109 This study utilised clinical data from the MSBase Registry, an international, prospective,  
110 observational MS clinical outcomes register. Data are collected in a unified manner, and  
111 include patient demographics, expanded disability status scale (EDSS) scores, relapse,  
112 treatment and pregnancy data, as previously described<sup>9,10</sup>.

113

#### 114 Participant recruitment, parity definitions and sample collection

115 Whole-blood samples were obtained from 1,984 participants. From this cohort, we selected  
116 192 matched participants based on geographical location (Australia), sex (female), birth  
117 history availability (nulligravida or parous) and age (groups age-matched within three years,  
118 **Supplementary Fig. 1**).

119

120 DNA methylation is associated with age<sup>11</sup> and geographical location<sup>11</sup>. Therefore, we  
121 restricted participants to Australians matched by age (within three years). Participants were  
122 also matched by Age-Related Multiple Sclerosis Severity (ARMSS) scores<sup>12</sup> due to non-  
123 negligible differences between nulligravida and parous groups (**Table 1**). Participants were  
124 matched using the *optmatch* package<sup>13</sup> in the R statistical environment.

125

126 The timing of pregnancy effects on methylation patterns remains unclear in wwMS, as does  
127 the impact of pregnancies resulting in miscarriage or termination compared to birth. We  
128 therefore excluded gravida women (i.e., those experiencing a miscarriage or induced abortion  
129 only) and restricted study inclusion to women who had at least one preterm or term birth prior  
130 to the date of blood collection, or those who were nulligravida. We included wwMS from the  
131 Royal Melbourne Hospital (VIC, n=73), Box Hill Hospital (VIC, n=56), John Hunter  
132 Hospital (NSW, n=25), and Flinders Medical Centre (SA, n=38). A total of 96 nulligravida  
133 and 96 parous females with RMS were included in this study (n=192).

134

135 DNA extraction

136 At each site, genomic DNA was extracted from whole blood using standard protocols and  
137 procedures.

138

139 Methylation arrays

140 DNA samples were processed for methylation arrays at the Hunter Medical Research Institute  
141 (NSW). DNA quantity and quality were assessed using Qbit (Invitrogen™, USA) and  
142 TapeStation (Agilent™, USA), respectively. Samples meeting concentration and quality  
143 requirements were bisulfite converted using the EZ-DNA Methylation™ Kit (Zymo)  
144 according to manufacturer guidelines. Converted DNA was hybridised to Illumina  
145 Methylation EPIC BeadChip arrays (EPIC arrays). Samples were randomised based on clinic  
146 site using the *OSAT* R package to avoid batch effects. EPIC arrays were read using an iScan  
147 (Illumina™) and raw Idat files were produced for analysis.

148

149 Genotyping arrays

150 Genomic DNA was sent from participating study sites to the Center for Genome Technology,  
151 John P. Hussman Institute for Human Genomics, University of Miami, for quality assessment  
152 and genotyping. Genotyping was performed in two batches using Illumina Multi-ethnic  
153 genotyping array (MEGA<sup>EX</sup>) arrays. Genotype calling was conducted in GenomeStudio v2.0  
154 (Illumina).

155

156 DNA methylation analysis pipeline

157 Our EWAS analysis was informed by the guidelines described in Campagna et al. (2021)<sup>14</sup>.

158 The *Chip Analysis Methylation Pipeline (ChAMP)* Bioconductor package<sup>15</sup> was used for



159 methylation data pre-processing in the R statistical environment. Raw Idat files were filtered  
160 to exclude low quality samples (failed to successful probe ratio > 0.1), low quality probes  
161 (detection p-value > 0.01, bead count < 3 in  $\geq 5\%$  of samples), non-CpG probes, SNP-related  
162 probes, non-autosomal probes, and multi-hit probes. Additional multi-hit probes were  
163 excluded based on Pidsley (2016) Supplementary Table 1<sup>16</sup>. Beta values were normalised  
164 using the beta-mixture quantile (BMIQ) method<sup>17</sup>. Batch effects at the array and chip level  
165 were identified with singular value decomposition (SVD) analysis<sup>18</sup>, and corrected for using  
166 the *Combat* algorithm<sup>19</sup>.

167

### 168 Primary differential methylation analysis

169 Differential methylation ( $\Delta_{\text{meth}}$ ) between nulligravida and parous groups was identified at the  
170 single CpG level i.e. differentially methylated positions (DMPs), and genomic region level  
171 i.e. differentially methylated regions (DMRs) using the filtered and normalised beta matrix,  
172 as previously described<sup>14</sup>. We used the *ChAMP* function *champ.DMP* to implement an  
173 unadjusted logistic model of methylation level at each probe and parity group. A false  
174 discovery rate (FDR) threshold of 0.05 was used to assess statistical significance for all  
175 analyses. Methylation beta values equate to percentage methylation, and thus going forward  
176 we report methylation differences (effect size) as a percentage (e.g.,  $\Delta_{\text{meth}}$  of 0.01 = 1%).  
177 DMPs with an  $\Delta_{\text{meth}}$  less than 1% were removed to avoid false positives produced from  
178 technical error.

179

180 We identified DMRs using a two-pronged approach. Firstly, with the *DMRcate* R package<sup>20</sup>  
181 using the following parameters: at least three DMPs within 1000bp of the adjacent DMP, a  
182 DMP and DMR threshold of FDR < 0.05. Secondly, using the DMP list to identify at least  
183 three DMPs with an FDR < 0.05 and the same direction of effect, located within 1000bp of

184 each other. The validity of this strategy to identify DMRs in studies with small sample and/or  
185 effect sizes has previously been shown<sup>21,22</sup>.

186

187 As methylation can be cell type specific, immune cell type proportions were estimated to  
188 confirm that differential methylation in whole blood was not driven by differences in cell  
189 type proportions. Immune cell type proportions were estimated using the *EpiDISH* R  
190 package<sup>23</sup>, using methylation M-values and the reference-based *CIBERSORT* algorithm<sup>24</sup>.

191 Subsequently, cell-type specific DMPs (csDMPs) were identified using a modified version of  
192 the cellDMC function of the *EpiDISH* R package<sup>23</sup>. We used linear regression where the  
193 outcome was methylation M-value and the predictors were cell type proportion estimate, and  
194 an interaction term of cell type proportion and parity. A separate model was run for each cell  
195 type. We used a genome-wide threshold of  $p \leq 9 \times 10^{-8}$  to assess statistical significance.

196

### 197 Sensitivity analyses

198 Sensitivity analyses were performed to assess the potential impact of a series of demographic,  
199 clinical, biological, and environmental covariates on the primary methylation analysis.

200 Covariates were selected based on non-negligible differences between groups (Cohen's  $d >$   
201 0.15), or *a priori* selected. They included symptom duration, annualised relapse rate (ARR),  
202 cell type proportion estimates (B cells, CD4+ cells, CD8+ cells, NK cells, monocytes, and  
203 granulocytes), and methylation age acceleration (PhenoAge and GrimAge). Environmental  
204 factors including treatment at blood collection (yes or no) and smoking status at blood  
205 collection (ever or never) were also tested. An FDR threshold of 0.05 was used to assess  
206 statistical significance for all sensitivity analyses.

207

208 Nulligravida and parous participants were matched by age at blood collection and ARMSS  
209 scores (n=192, 96 pairs) and the difference in methylation at each probe ( $\Delta_{\text{meth}}$ ) was  
210 calculated. Subsequently, the correlation between each  $\Delta_{\text{meth}}$  and covariate was tested.  
211 Pearson's correlation tests were used for continuous covariates (ARR, symptom duration,  
212 years follow-up in MSBase and number of available EDSS scores available). ANOVA tests  
213 were used for categorical covariates (treatment and smoking status). For categorical  
214 variables, pairs were required to have the same value for the correlation with methylation to  
215 be tested. Of 96 pairs in total, 40 pairs were on treatment at blood collection and 14 were off  
216 treatment, while eight pairs were 'ever' smokers at blood collection and two were 'never'  
217 smokers. Smoking history was unavailable for the remaining pairs. DMPs were filtered for  
218 2,622 known smoking-associated CpGs identified by Johanes et al. (2016, Supplementary  
219 Table 2)<sup>25</sup> due to the known effect of smoking on the methylome, and limited smoking data  
220 available for this cohort.

221

### 222 Single Nucleotide Variant analysis

223 Quality control was performed with *PLINKv1.9*<sup>26</sup>. Single Nucleotide Variants (SNVs) were  
224 excluded based on low call rate (<95%), low minor allele frequency (MAF < 0.05), violation  
225 of Hardy–Weinberg equilibrium ( $p < 1 \times 10^{-5}$ ), monomorphism and non-autosomal location.  
226 Samples were excluded based on sex inconsistencies, low call rate (<95%) and relatedness  
227 ( $\pi\text{-hat} > 0.05$ ). Relatedness was assessed using Identity by Descent (IBD) analysis in  
228 *PLINKv1.9*, followed by confirmation in *KING*<sup>27</sup>. Principal components (PC) analysis was  
229 implemented in *EIGENSTRAT*<sup>28</sup>. PCs were projected to 1000 Genomes Project<sup>29</sup> data to  
230 assess population stratification effects, and exclude population outliers. Genotypes were then  
231 imputed using Haplotype Reference Consortium<sup>30</sup> on the Michigan Imputation Server

232 (<https://imputationserver.sph.umich.edu/index.html#!>), and converted to genotype calls in  
233 *PLINKv1.9*.

234

### 235 Methylation quantitative trait loci (mQTL) analysis

236 Differential methylation at certain genetic loci may be influenced by the underlying SNVs at  
237 or near that site, known as methylation quantitative trait loci (mQTLs). Therefore, we tested  
238 the relationship between genotype and methylation at CpGs in the identified DMR to  
239 determine whether differential methylation was associated with, or independent of,  
240 underlying genotype.

241

242 We extracted genotypes at SNVs located five kilobases (kb) up and downstream of DMR<sup>Chr21</sup>  
243 boundaries using the *KRIS R* package<sup>31</sup>, and assessed linkage disequilibrium (LD) using  
244 bivariate correlations of genotype frequencies with a significance threshold of  $p < 0.05$ .

245

246 To test if differential methylation within DMRs was driven by genetic effects rather than  
247 parity, we performed a linear regression with methylation as the dependent variable and  
248 genotype and parity as the independent variables.

249

### 250 Multi-factor feature selection

251 Elastic net regression is a form of penalised regression that is useful for uncovering multiple  
252 conjoint effects in datasets with correlated features (e.g., methylation) and a greater number  
253 of features than samples ( $p \gg n$ ). This method can be useful for identifying important  
254 features with greater sensitivity than conventional EWAS analyses. We used machine  
255 learning to build an elastic net regression model to identify CpGs at which methylation was  
256 associated with parity, inputting beta values at approximately 748,000 CpGs. Samples were

257 split into training (n=134) and testing sets (n=58) to reduce overfitting. The model was  
258 trained using a cross-validation resampling method with 10 iterations, with the train function  
259 of the *caret* R package<sup>32</sup>. The optimal alpha value was used in a subsequent k-fold cross-  
260 validation elastic net regression to identify the minimum lambda value; using the *cv.glmnet*  
261 function of the *glmnet* R package<sup>33</sup>. These alpha and lambda values were used in the final  
262 elastic net regression model that was applied to the testing set using the *glmnet* function of  
263 *glmnet* R package<sup>33</sup>. Features (CpGs) identified by the model to be associated with parity  
264 were compared to DMPs and DMRs identified in the primary analysis, as well as mapped to  
265 genes for GSEA performed as described above.

266

#### 267 Gene-set enrichment analysis (GSEA)

268 We used gene-set enrichment analysis (GSEA) to generate hypotheses about the functional  
269 consequence of differentially methylated genes between nulligravida and parous women. All  
270 CpGs that were associated with parity in the primary differential methylation analysis and  
271 elastic net regression were used as input. We conducted GSEA using two methods. Firstly,  
272 the *ToppGene* online application programming interface (API)<sup>34</sup> which takes an FDR ranked  
273 gene list ranked as input, with hypomethylated and hypermethylated genes analysed  
274 separately. Secondly, we used the *GOMeth* function<sup>35</sup> of the *missMethyl* R package<sup>36</sup> to  
275 address probe number and multi-gene bias specific to methylation data from arrays. A list of  
276 DMPs and all CpGs tested were used as input, and both Gene Ontology (GO) and KEGG  
277 pathway collections were tested. We used a Benjamini-Hochberg adjusted p-value ( $FDR_{B\&H}$ )  
278 threshold of 0.05 to assess the statistical significance of enriched gene sets.

279

#### 280 Methylation age analysis

281 Methylation age is the prediction of biological age from methylation levels at a subset of  
282 CpGs (clock CpGs). PhenoAge<sup>37</sup> and GrimAge<sup>38</sup> are the most accurate and widely used  
283 methylation age algorithms, and have been associated with increased risk of various  
284 morbidities and mortality<sup>37–39</sup>.

285

286 We estimated methylation age using the PhenoAge<sup>37</sup> algorithm with the `methyAge` function  
287 of the *ENmix* R package<sup>40</sup>. GrimAge was calculated with the online calculator at  
288 <https://dnamage.genetics.ucla.edu/>. MAA was defined as the residual term from regressing  
289 chronological age on methylation age estimates. For each algorithm, Shapiro-Wilk normality  
290 tests were used to test the normality of the MAA distribution. To test if mean MAA was  
291 significantly different between groups a one tailed t-test was used for the PhenoAge  
292 algorithm, and a Mann-Whitney test for the GrimAge algorithm.

293

## 294 **Results**

### 295 Cohort descriptive statistics

296 This study included 192 females with RMS across four study sites. Participants were  
297 categorised as nulligravida (n=96) or parous (n=96) based on available pregnancy history  
298 data. For parous participants, the median time from last conception to blood collection was  
299 16.66 years (range = 1.45 – 44.42 years, **Table 1**).

300

### 301 Differential methylation analysis – whole blood

302 After methylation data pre-processing, approximately 747,000 (86%) of 867,000 probes  
303 remained for differential methylation analysis (**Supplementary Fig. 2**). Batch effect analysis  
304 identified Plate, Sentrix ID and Sentrix Position as significant sources of technical variation

305 (p < 0.01), which were corrected and reduced to negligible effects using the *Combat*  
306 algorithm<sup>19</sup> (**Supplementary Fig. 2**).

307

308 Whole-blood methylation analysis revealed 903 differentially methylated positions (DMPs)  
309 surpassing genome-wide thresholds (FDR < 0.05 and  $\Delta_{\text{meth}} > 1\%$ , **Table 2** shows the top 10  
310 DMPs by effect size (full list is available in **Supplementary Table 1**. DMPs mapped to 585  
311 genes and 318 unannotated genomic locations. Of the 903 DMPs, 365 (40%) were  
312 hypomethylated and 528 (60%) were hypermethylated in the parous group relative to the  
313 nulligravida group (**Supplementary Fig. 3**).  $\Delta_{\text{meth}}$  ranged from -13.28% to 16.10%. CpG  
314 islands are associated with gene promoter regions, in which methylation is likely to impact  
315 gene transcription. Only 106 (11.7%) of DMPs were in islands, with 173 (19.2%) in shores,  
316 64 (7.1%) in shelves and the majority in open sea regions (560, 62.0%). Of 903 DMPs, five  
317 overlapped with the 10,592 DMPs identified by Mehta et al. (2019, **Supplementary Table**  
318 **2**).

319

320 No differentially methylated regions (DMRs) were identified using the *DMRcate* algorithm at  
321 an FDR threshold of 0.05. Therefore, we identified DMRs from our DMP list, defining a  
322 DMR as a region containing at least three DMPs with the same effect direction and FDR <  
323 0.01, within 1000bp of the adjacent DMP/s<sup>21</sup>. Using this definition, we identified five DMRs  
324 on Chromosomes 7, 15, 17, 18 and 21 (**Table 3**). However, only DMR<sup>Chr7</sup> ( $\Delta_{\text{max}} = 0.029$ , FDR  
325 = 0.021, **Supplementary Fig. 4a**) and DMR<sup>Chr15</sup> ( $\Delta_{\text{max}} = 0.049$ , FDR = 0.015, **Supplementary**  
326 **Fig. 4b**) remained after mQTL analysis (see below). DMR<sup>Chr7</sup> mapped to CRYGN, while  
327 DMR<sup>Chr15</sup> mapped to an intergenic region. Both DMRs were hypermethylated in the parous  
328 group, relative to the nulligravida group.

329

330 Differential methylation analysis – immune cell specific

331 Differential methylation analysis of whole blood may not be sensitive to cell specific DMPs  
332 associated with outcome. Therefore, we estimated and compared the proportion of immune  
333 cell types between groups. There were no significant differences in immune cell type  
334 proportions between nulligravida and parous women (data not shown), and therefore we did  
335 not need to adjust our whole blood analysis for this variable. Statistical deconvolution  
336 revealed four CD4+ (**Table 4**) and eight CD8+ T cell specific DMPs (**Table 4**). All CD4+ T  
337 cell DMPs were hypermethylated in the parous group compared to the nulligravida group,  
338 and only one DMP mapped to a gene (cg14172633, *HMCNI*). In CD8+ T cells, three DMPs  
339 were hypermethylated and five were hypomethylated in the parous group. The DMP  
340 cg25577322 had the largest effect size (estimate = -8.32, SE = 1.45) and mapped to *AHR*.  
341 Seven of the eight DMPs mapped to a gene, and two DMPs mapped to *OR2L13* (cg08944170  
342 and cg20507276).

343

344 Sensitivity analysis

345 Sensitivity analyses revealed no major effects of symptom duration, ARR, cell type  
346 proportion estimates (B cells, CD4+ cells, CD8+ cells, NK cells, monocytes, and  
347 granulocytes), methylation age acceleration (PhenoAge and GrimAge), treatment at blood  
348 collection (yes or no) or smoking status at blood collection (ever or never) on differential  
349 methylation in this cohort, as demonstrated by the lack of association between CpGs and the  
350 covariates tested (data not shown). Therefore, these covariates were not included in the  
351 differential methylation analyses so as not to unnecessarily burden the model and reduce  
352 statistical power. One CpG (cg03708250) showed suggestive association with age at blood  
353 collection (FDR = 0.042) but was not identified as a DMP. Of the 2,622-smoking associated



354 CpGs from Joehanes (2016) and 903 DMPs identified in this study, 25 overlapped and were  
355 removed prior to downstream analyses to avoid confounding (**Supplementary Table 3**).

356

### 357 Methylation quantitative trait loci (mQTL) analyses

358 After quality control and filtering, 183 patients remained for mQTL analysis. SNVs located  
359 within 5kb up/downstream of each DMR were identified for LD. DMR<sup>Chr7</sup> contained five  
360 independent SNVs (**Supplementary Table 4**). Methylation at cg23666844, cg16077872 and  
361 cg17362899 was associated with genotype at 7:151133104:G:C, 7:151135503:C:T,  
362 7:151137301:G:C and 7:151140431:T:C (**Supplementary Table 5**). After accounting for  
363 genotype at these SNVs, methylation at cg23666844, cg16077872 and cg17362899 remained  
364 associated with parity despite the presence of mQTLs (**Supplementary Table 5**).

365

366 DMR<sup>Chr15</sup> contained five independent SNVs (**Supplementary Table 4**). Methylation at  
367 cg26795333 was associated with genotype at all SNVs. Methylation at cg20560283 and  
368 cg17174814 was associated with genotype at 15:67224485:A:G, 15:67224701:G:C,  
369 15:67224979:C:G and 15:67228085:C:T (**Supplementary Table 5**). After accounting for  
370 genotype at these SNVs, methylation at cg26795333, cg20560283 and cg17174814 remained  
371 associated with parity despite the presence of mQTLs (**Supplementary Table 5**).

372

373 DMR<sup>Chr17</sup> contained four independent SNVs (**Supplementary Table 4**). Methylation at  
374 cg22349396 was associated with genotype at 17:67225730:G:C, 17:67226643:C:T and  
375 17:67227383:A:T. Methylation at cg06444025 and cg01726265 was not significantly  
376 associated with genotype at any SNV (data not shown). However, methylation at all CpGs  
377 was no longer associated with parity after accounting for genotype at these four SNVs  
378 (**Supplementary Table 5**).

379

380 DMR<sup>Chr18</sup> contained six independent SNVs (**Supplementary Table 4**). Methylation at  
381 cg23973972 was associated with genotype at 18:67224072:A:C, 18:67226036:G:A,  
382 18:67226036:G:A and 18:67227787:A:C (**Supplementary Table 5**). Methylation at  
383 cg15682262 was associated with genotype at 18:67224072:A:C, 18:67226036:G:A,  
384 18:67226036:G:A and 18:67227787:A:C. Methylation at cg27477494 was associated with  
385 genotype at 18:67226036:G:A, 18:67226036:G:A and 18:67227787:A:C. After accounting  
386 for genotype at these SNVs, methylation at cg23973972, cg15682262 and cg27477494 was  
387 no longer associated with parity (**Supplementary Table 5**).

388

389 DMR<sup>Chr21</sup> contained three independent SNVs: 21:44782007:C:T, 21:44782634:C:A and  
390 21:44782732:A:C (**Supplementary Table 4**). Methylation at cg17577862, cg02260098,  
391 cg25191041 and cg14081667 was not associated with genotype at any SNV (data not shown).  
392 After accounting for genotype at these SNVs, methylation at cg17577862 remained  
393 associated with parity, however, methylation at cg02260098, cg25191041, cg14081667 was  
394 no longer associated with parity (**Supplementary Table 5**).

395

396 After accounting for genotype at independent SNVs located within (plus 5kb up/downstream)  
397 of each DMR, only DMR<sup>Chr7</sup> and DMR<sup>Chr15</sup> contained enough DMPs ( $\geq 3$ ) to be considered  
398 DMRs.

399

#### 400 Multi-factor feature selection

401 Using elastic net regression, we identified a panel of CpGs conferring a conjoint association.  
402 We determined the optimal alpha (0.1) and lambda (0.02) values for our data using a cross

403 validation approach. With an alpha value closer to zero than one, our elastic net regression  
404 resembled a lasso regression more closely than a ridge regression.

405

406 Using these model parameters, our elastic net regression model selected 1556 CpGs  
407 associated with parity (top 10 shown in **Table 5**, full list in **Supplementary Table 6** in our  
408 training dataset (n=134, 70% of cohort). Of these, 322 CpGs (34.5%) were also identified as  
409 DMPs in our primary analysis. The most important CpG in the model, cg08186508 (variable  
410 importance = 100), is located on Chromosome 14 and maps to the *PIGH* gene. However,  
411 cg08186508 was not identified as a DMP in our primary analysis suggesting its effect is in  
412 correlation with other CpGs.

413

#### 414 Gene-set enrichment analysis (GSEA)

415 We conducted GSEA on all differentially methylated genes identified in the primary  
416 methylation analysis and elastic net regression to elucidate potentially small but cumulative  
417 effects of parity on methylation patterns (**Supplementary Table 7**). 609 of 903 DMPs (67%),  
418 and 1208 of the 1556 CpGs (78%) identified in the elastic net regression, mapped to a gene.  
419 In total, 1318 unique genes were used for GSEA.

420

421 ToppGene revealed that differential methylation, regardless of direction of effect, was  
422 primarily enriched in biological processes (**Figure 1a**) and cellular compartments (**Figure**  
423 **1b**) related to neuroplasticity, including neurogenesis ( $n_{\text{genes}} = 178$ ,  $\text{FDR}_{\text{B\&H}} = 2.77 \times 10^{-5}$ ),  
424 neuron projection morphogenesis ( $n_{\text{genes}} = 87$ ,  $\text{FDR}_{\text{B\&H}} = 2.77 \times 10^{-5}$ ) and neuron projection  
425 ( $n_{\text{genes}} = 175$ ,  $\text{FDR}_{\text{B\&H}} = 6.96 \times 10^{-10}$ ). Furthermore, the top enriched molecular functions  
426 related to ion transport including cell adhesion molecule binding ( $n_{\text{genes}} = 70$ ,  $\text{FDR}_{\text{B\&H}} =$   
427  $2.06 \times 10^{-4}$ ) and GTPase regulator activity ( $n_{\text{genes}} = 61$ ,  $\text{FDR}_{\text{B\&H}} = 3.29 \times 10^{-5}$ , **Figure 1c**).

428

429 Hypomethylated genes (n=576) drove the enrichment of neuron development and growth  
430 biological processes and cellular compartments. Alternatively, hypermethylated genes  
431 (n=775) drove the enrichment of signal transduction biological processes and molecular  
432 functions.

433

434 There were no enriched gene ontology terms using GOMeth with an FDR threshold of 0.05.  
435 This suggests that our ToppGene findings could be a result of probe number or multi-gene  
436 bias. However, we used GSEA as an exploratory analysis to generate hypotheses about the  
437 mechanism in which pregnancy impacts MS clinical outcomes and have therefore interpreted  
438 the results with caution.

439

#### 440 Methylation age analysis

441 Methylation Age Acceleration (MAA) measures the disparity between chronological and  
442 biological age as estimated using methylation age algorithms, and can provide insight into an  
443 individual's health and lifespan<sup>39,37,38</sup>. As groups were *a priori* matched by age, there were no  
444 significant differences in chronological age between groups (**Table 1**).

445

446 The correlation between chronological age and methylation age using the PhenoAge and  
447 GrimAge algorithms were 0.77 and 0.91, respectively. We did not find any evidence for  
448 differences in methylation age between groups using the GrimAge algorithm ( $p = 0.854$ ).  
449 However, we did find significant differences in methylation age between groups using the  
450 PhenoAge algorithm ( $p = 0.034$ , **Supplementary Fig. 5**).

451

452 MAA was calculated as the residual term from regressing chronological age on methylation  
453 age. Residual terms were normally distributed for the PhenoAge ( $p = 0.551$ ) algorithm, but  
454 not the GrimAge algorithm ( $p = 3.52 \times 10^{-05}$ ). There were significant differences in MAA  
455 between nulligravida and parous groups using both the PhenoAge (mean difference = 2.27  
456 years,  $p = 0.001$ ) and GrimAge algorithms (mean difference = 1.44 years,  $p = 0.005$ ,  
457 **Figure 2**).

458

459

## 460 **Discussion**

461 Studies have demonstrated an association between pregnancy and reduced disability  
462 accumulation in women with MS (wwMS)<sup>2</sup>, lasting for up to ten years post pregnancy<sup>4</sup>.  
463 Recent studies have identified associations between birth history and methylation patterns in  
464 health<sup>41–43</sup> and MS<sup>8</sup>, as well as negative associations between birth history and methylation  
465 age acceleration in women without MS<sup>6</sup>. No methylome-wide studies to date have examined  
466 associations between methylation patterns and parity, or methylation age acceleration in  
467 wwMS.

468

469 Our primary EWAS of whole blood methylation differences between nulligravida and parous  
470 wwMS identified 903 differentially methylated positions (DMPs) across autosomes. Of the  
471 DMPs identified in our study, five overlapped with those previously identified by Mehta and  
472 colleagues (2019)<sup>8</sup>. This is reasonably explained by differences in cohort size and study  
473 design, where Mehta et al. (2019) sought to identify DMPs in genes that were identified *a*  
474 *priori*<sup>8</sup>, compared to our genome-wide approach. Moreover, they included women with a  
475 history of pregnancy, compared to our study which included only women with a history of  
476 birth. The overlapping DMPs mapped to *PRIC285*, *GRTPI*, *SIM2* and *CCDC90B*, and one

477 intergenic region. *PRIC285* is a nuclear transcriptional co-activator and an interferon effector  
478 gene that is integral to antiviral immune responses<sup>44</sup>. *GRTPI*, is a GTPase activator  
479 associated with platelet counts<sup>45</sup>. Notably, *SIM2* is a transcription factor and master regulator  
480 of central nervous system development and neurogenesis<sup>46</sup>, and a link between neurogenesis,  
481 neuronal reserve and MS outcomes is frequently hypothesised<sup>47</sup>. Further, *CCDC90B* is a  
482 protein coding gene which regulates mitochondrial calcium ion concentrations affecting ATP  
483 production<sup>45</sup>. A link between mitochondrial dysfunction in neurons and MS outcomes has  
484 recently been identified<sup>48</sup>. While we identified a different direction of effect at *CCDC90B*  
485 between our study and Mehta et al. (2019), this may be due differences in sample size or  
486 timing of sample collection. In this study, we confirm that genes related to these processes  
487 are differentially methylated between nulligravida and parous wwMS for up to 44.4 years  
488 post-pregnancy. Single nucleotide variants associated with mitochondrial and CNS function  
489 were recently shown to associate with MS severity outcomes<sup>49</sup>. Our study demonstrates a  
490 putative mechanism by which pregnancy may impact long-term legacy effects on outcomes  
491 in wwMS.

492  
493 In addition to 903 DMPs, we identified two differentially methylation regions (DMRs) on  
494 Chromosome 7 and 15.  $DMR^{Chr7}$  contains three DMPs in the transcript start site (up to  
495 1500bp 5' of 5'UTR) promoter region of *CRYGN* which encodes crystallin gamma N, a  
496 structural protein in eye lenses. The literature on *CRYGN* is limited, and it has not previously  
497 been linked to pregnancy or MS. Our finding requires validation in an independent cohort  
498 and replicated DMP/DMR signals would provide a strong rationale for in vitro functional  
499 studies of gene and protein expression control mediated by the DMR.

500

501 Using statistical deconvolution, we identified four CD4+ T cell specific DMPs (csDMPs).  
502 CD4+ T cells are central to immune regulation and tolerance, and have been strongly linked  
503 to both MS<sup>50</sup> and pregnancy<sup>51</sup>. Multiple studies have reported changes in the epigenetic  
504 patterns of CD4+ T cells during pregnancy in wwMS<sup>7,52,53</sup>. The only gene-associated DMP  
505 was at the transcription start site for *hemicentin 1 (HMCNI)*, a member of the  
506 immunoglobulin superfamily. However, to date *HMCNI* has not been associated with  
507 differential methylation in healthy populations or wwMS during pregnancy<sup>52</sup>. Nor is there  
508 literature linking *HMCNI* to MS risk. Therefore, this finding, together with the association of  
509 differential methylation of *CRYGN* is highly novel and requires independent validation.  
510  
511 We also identified eight CD8+ T cell csDMPs that map to six genes and one intergenic  
512 region. The involvement of CD8+ T cells in MS pathophysiology is well established<sup>50</sup>.  
513 During pregnancy CD8+ T cells are critical for maternal–fetal tolerance and protection  
514 against viruses<sup>54</sup>. The functions and diseases associated with the CD8+ T cell csDMPs  
515 identified in this study suggest they are markers of pregnancy outcomes, rather than genes  
516 implicated in the modulation of MS outcomes due to pregnancy (e.g., *OR2L1*, *HOOK2* and  
517 *CUL2*). Most notably, Aryl Hydrocarbon Receptor (*AHR*, cg25577322) is upregulated in  
518 decidual natural killer cells in women with recurrent spontaneous abortion and was  
519 hypomethylated in parous women in our study<sup>55</sup>. Here, we excluded pregnancies ending in  
520 miscarriage or termination to prevent identifying epigenetic biomarkers of miscarriage or  
521 termination. While it is possible that this signal was driven by unreported terminations and/or  
522 unknown miscarriages, it was identified in peripheral CD8+ T cells only (not whole blood)  
523 and is therefore unlikely to be a marker of recurrent spontaneous abortion in this cohort.  
524 Furthermore, multiple studies have recently correlated *AHR* agonist activity with MS subtype  
525 and prognosis<sup>56,57</sup>. In these studies, *AHR* agonist activity increase was associated with relapse

526 in CIS and RRMS<sup>56</sup>. A decrease in *AHR* agonist activity was associated with RRMS  
527 remission<sup>56</sup> and progressive MS,<sup>57</sup> thus implicating *AHR* in neuroinflammatory processes. In  
528 our study, *AHR* was hypomethylated. Hypomethylation is often, but not always associated  
529 with upregulation of gene expression. Unfortunately, we did not assess gene expression.  
530 However, this provides a plausible mechanism by which pregnancy could modulate disease  
531 outcomes and warrants further investigation.

532

533 We conducted GSEA on 1318 differentially methylated genes identified in the primary  
534 analysis and elastic net regression to generate hypotheses about the functional role of these  
535 genes in long-term MS outcomes. Hypomethylated genes in parous wwMS were enriched in  
536 neuron development and growth biological processes and cellular compartments, and  
537 hypermethylated genes were enriched in signal transduction biological processes and  
538 molecular functions. Mehta et al. (2019) similarly found enrichment of neuronal pathways  
539 including axon guidance in their study of differentially expressed genes between nulliparous  
540 and parous wwMS<sup>8</sup>. While the majority of DMPs (95.9%) in our primary differential  
541 methylation analysis had small effect sizes, the strength of our penalised regression approach  
542 is the ability to reveal small, correlated relationships between features. Taken together, these  
543 findings suggest that methylation impacts neuraxonal maintenance and neurite growth in  
544 parous women in a small but cumulative manner, up to 44.4 years after pregnancy. These  
545 findings are consistent with reports that the brains of women who have children undergo  
546 pronounced morphological changes as a result of pregnancy<sup>58</sup>.

547

548 Ours is the first study to report a reduction in MAA in parous wwMS, compared to age-  
549 matched nulligravida wwMS. We demonstrated that parous women have a reduced mean  
550 MAA of between 1.44 - 2.27 years depending on the algorithm employed. This shows that,



551 as in health, parity is associated with a reduction in MAA in wwMS<sup>6</sup>. GrimAge is the newest  
552 algorithm with robust associations with morbidity and mortality<sup>38</sup>. Furthermore, PhenoAge  
553 acceleration is associated with an increased risk of physical functioning problems<sup>59</sup>. As a  
554 whole, our findings demonstrate slower biological aging in parous wwMS, and potentially a  
555 longer period of health and lifespan<sup>59</sup>.

556

557 Ours is the largest study to date investigating the association between genome-wide  
558 methylation and parity in women with relapse-onset MS. We identified hundreds of  
559 methylation changes associated with parity that may underlie long-term outcomes in wwMS.  
560 Cohort matching by age limited confounding and erroneous associations between methylation  
561 patterns and parity. We aimed to mitigate against confounding by disease severity by  
562 matching for ARMSS scores, therefore allowing us to study the relationship between  
563 methylation patterns and parity specifically. Whether these changes are specific to wwMS or  
564 a broader response to pregnancy remain to be confirmed in future studies. We were  
565 underpowered to adjust for a range of clinical and environmental factors potentially  
566 associated with methylation patterns, including number of births and DMT<sup>60</sup>. Study power  
567 also limited our ability to identify small cell type-specific effects beyond those identified in T  
568 cells. Therefore, our findings require validation in a larger, independent cohort of wwMS. As  
569 ours is a retrospective and cross-sectional study, we were not able to establish a causal link  
570 between pregnancy, methylation pattern changes, and long-term clinical outcomes in wwMS.  
571 We are currently undertaking a longitudinal and prospective study of methylation changes  
572 during and after pregnancy relative to a nulligravida baseline, to investigate temporal and  
573 causal relationships between pregnancy, methylation, and disease outcomes in wwMS. This  
574 could lead to the identification novel therapeutic targets.

575

576 **Conclusion**

577 We investigated the association between whole blood and cell type specific genome-wide  
578 methylation patterns and parity in 192 women with relapse-onset MS. We identified small but  
579 potentially cumulative differences in whole-blood and T-cell methylation patterns in genes  
580 related to neural plasticity, offering a putative molecular mechanism driving the long-term  
581 effect of pregnancy on MS outcomes. We further identified reduced methylation age  
582 acceleration in parous wwMS, demonstrating slower biological aging compared to  
583 nulligravida wwMS. As methylation patterns can be cell type specific, our results suggest a  
584 potential ‘CNS signature’ of methylation in peripheral immune cells, as previously described  
585 in relation to MS progression<sup>61</sup>. This is the first genome-wide methylation study of parity in  
586 wwMS and therefore, validation studies are needed to confirm our findings.

587 **Data availability**

588 Supplemental files contain data supporting the conclusions in this article. Data access  
589 requests with scientifically sound proposals can be made in writing to Dr Viliija Jokubaitis  
590 ([viliija.jokubaitis@monash.edu](mailto:viliija.jokubaitis@monash.edu)). Clinical data from the MSBase Registry: To protect  
591 participant confidentiality, de-identified patient-level data sharing may be possible in  
592 principle but will require permissions/consent from each contributing data controller.

593

594 **Acknowledgements**

595 We thank all the people with MS who participated in this research without whom this work  
596 would not be possible. The authors would like to acknowledge research nurses Ms Jo Baker,  
597 Ms Jodi Haartsen, Ms Sandra Williams, Ms Lisa Taylor for assisting with blood collection  
598 for this study, and Dr Louise Laverick, and Ms Malgorzata Krupa for research assistance.

599

600 **Funding**

601 This study was financially supported by an MSRA Project Grant (18-0424), RMH Home  
602 Lottery Grant (MH2013-055), MSBase Foundation Project Grant, Charity Works for MS  
603 Project Grant, Pennycook Foundation Grant 2018, and a Monash University Project Grant.

604

605 **Author declarations**

606 None.

607

608 **Conflicts of interest**

609 The authors report no conflict of interest.

610

611 **Author contributions**

612 MPC conducted statistical analyses, drafted, and revised the manuscript and aided in study  
613 design and data interpretation. VM aided in sample collection, data interpretation and revised  
614 the manuscript. MS and TK aided in sample collection and revised the manuscript. HB and  
615 JLS aided in sample collection, study design, data interpretation and revised the manuscript.  
616 RS acquired funding, helped develop the concept, aided in data interpretation, and revised the  
617 manuscript. RL and VJ designed and conceptualised the study, acquired study funding, aided  
618 in sample collection and data interpretation, and drafted and revised the manuscript.

619 **References**

- 620 1. Krysko KM, Graves JS, Dobson R, et al. Sex effects across the lifespan in women with  
621 multiple sclerosis. *Ther Adv Neurol Disord* 2020;13:1756286420936166.
- 622 2. Nguyen A-L, Eastaugh A, van der Walt A, Jokubaitis VG. Pregnancy and multiple  
623 sclerosis: Clinical effects across the lifespan. *Autoimmun Rev* 2019;18(10):102360.
- 624 3. Nguyen A-L, Vodehnalova K, Kalincik T, et al. Effect of pregnancy on the onset of  
625 clinically isolated syndrome. *JAMA Neurology* 2020;
- 626 4. Jokubaitis VG, Spelman T, Kalincik T, et al. Predictors of long-term disability accrual in  
627 relapse-onset multiple sclerosis. *Ann. Neurol.* 2016;80(1):89–100.
- 628 5. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. *Adv Protein Chem*  
629 *Struct Biol* 2019;116:135–170.
- 630 6. Ross KM, Carroll J, Horvath S, et al. Immune epigenetic age in pregnancy and 1 year  
631 after birth: Associations with weight change. *American Journal of Reproductive*  
632 *Immunology* 2020;83(5):e13229.
- 633 7. Iannello A, Rolla S, Maglione A, et al. Pregnancy Epigenetic Signature in T Helper 17  
634 and T Regulatory Cells in Multiple Sclerosis. *Front. Immunol.* 2019;9:3075.
- 635 8. Mehta D, Wani S, Wallace L, et al. Cumulative influence of parity-related genomic  
636 changes in multiple sclerosis. *J. Neuroimmunol.* 2019;328:38–49.
- 637 9. Butzkueven H, Chapman J, Cristiano E, et al. MSBase: an international, online registry  
638 and platform for collaborative outcomes research in multiple sclerosis. *Mult. Scler.*  
639 2006;12(6):769–774.

- 640 10. Jokubaitis VG, Skibina O, Alroughani R, et al. The MSBase pregnancy, neonatal  
641 outcomes, and women's health registry. *Ther Adv Neurol Disord*  
642 2021;14:17562864211009104.
- 643 11. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications.  
644 *Nat Rev Genet* 2012;13(2):97–109.
- 645 12. Manouchehrinia A, Westerlind H, Kingwell E, et al. Age Related Multiple Sclerosis  
646 Severity Score: Disability ranked by age. *Mult. Scler.* 2017;23(14):1938–1946.
- 647 13. Hansen BB, Klopfer SO. Optimal Full Matching and Related Designs via Network  
648 Flows. *Journal of Computational and Graphical Statistics* 2006;15(3):609–627.
- 649 14. Campagna MP, Xavier A, Lechner-Scott J, et al. Epigenome-wide association studies:  
650 current knowledge, strategies and recommendations. *Clin Epigenetics* 2021;13(1):214.
- 651 15. Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for  
652 Illumina BeadChips. *Bioinformatics* 2017;33(24):3982–3984.
- 653 16. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina  
654 MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.  
655 *Genome Biol* 2016;17(1):208.
- 656 17. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization  
657 method for correcting probe design bias in Illumina Infinium 450 k DNA methylation  
658 data. *Bioinformatics* 2013;29(2):189–196.
- 659 18. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. An Epigenetic Signature in  
660 Peripheral Blood Predicts Active Ovarian Cancer [Internet]. *PLoS One* 2009;4(12)

- 661 19. Müller C, Schillert A, Röthemeier C, et al. Removing Batch Effects from Longitudinal  
662 Gene Expression - Quantile Normalization Plus ComBat as Best Approach for  
663 Microarray Transcriptome Data [Internet]. PLoS One 2016;11(6)
- 664 20. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially  
665 methylated regions in the human genome. Epigenetics Chromatin 2015;8:6.
- 666 21. Maltby VE, Lea RA, Burnard S, et al. Epigenetic differences at the HTR2A locus in  
667 progressive multiple sclerosis patients. Sci Rep 2020;10(1):22217.
- 668 22. Maltby VE, Lea RA, Sanders KA, et al. Differential methylation at MHC in CD4+ T  
669 cells is associated with multiple sclerosis independently of HLA-DRB1. Clin Epigenet  
670 2017;9(1):71.
- 671 23. Teschendorff A. Epigenetic Dissection of Intra-Sample-Heterogeneity [Internet].  
672 2017.Available from:  
673 <https://www.bioconductor.org/packages/release/bioc/html/EpiDISH.html>
- 674 24. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue  
675 expression profiles. Nature Methods 2015;12(5):453–457.
- 676 25. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. Circ  
677 Cardiovasc Genet 2016;9(5):436–447.
- 678 26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome  
679 Association and Population-Based Linkage Analyses. Am J Hum Genet  
680 2007;81(3):559–575.
- 681 27. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in  
682 genome-wide association studies. Bioinformatics 2010;26(22):2867–2873.

- 683 28. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLOS Genetics  
684 2006;2(12):e190.
- 685 29. A global reference for human genetic variation | Nature [Internet]. [date unknown];[cited  
686 2021 Mar 10 ] Available from: <https://www.nature.com/articles/nature15393>
- 687 30. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for  
688 genotype imputation. Nat Genet 2016;48(10):1279–1283.
- 689 31. Chaichoompu K, Abegaz F, Sissades T, et al. KRIS: Keen and Reliable Interface  
690 Subroutines for Bioinformatic Analysis. 2018.
- 691 32. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical  
692 Software 2008;28(1):1–26.
- 693 33. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models  
694 via Coordinate Descent. J Stat Softw 2010;33(1):1–22.
- 695 34. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment  
696 analysis and candidate gene prioritization. Nucleic Acids Res 2009;37(Web Server  
697 issue):W305–W311.
- 698 35. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide  
699 DNA methylation data [Internet]. Bioinformatics; 2020.
- 700 36. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from  
701 Illumina’s HumanMethylation450 platform. Bioinformatics 2016;32(2):286–288.
- 702 37. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and  
703 healthspan. Aging (Albany NY) 2018;10(4):573–591.



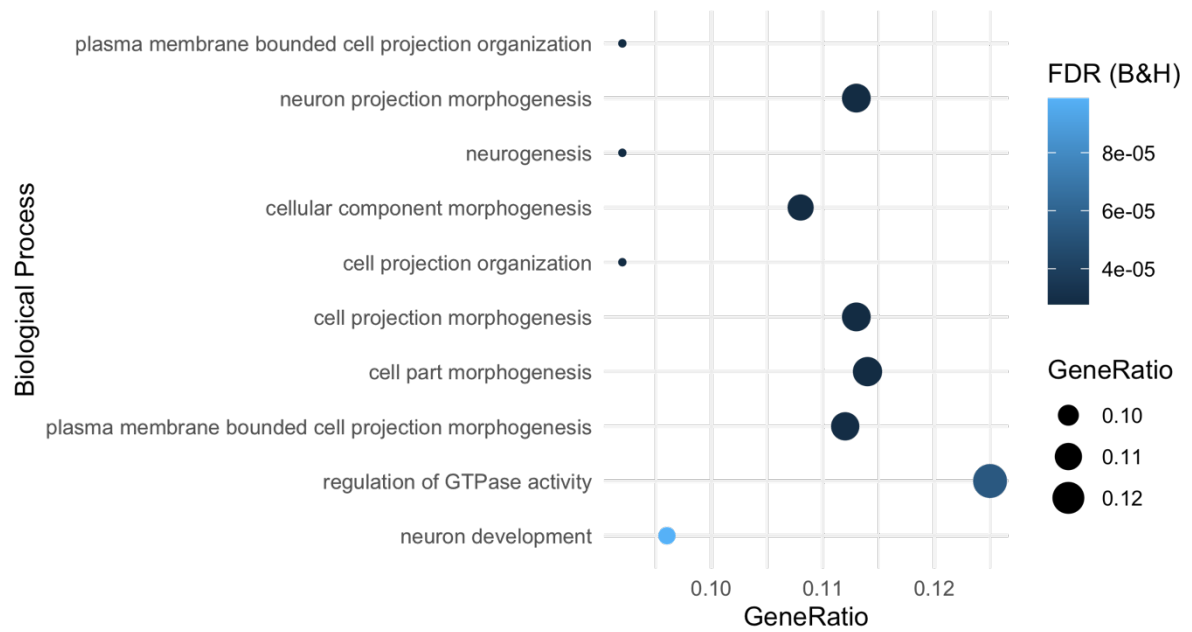
- 704 38. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan  
705 and healthspan. *Aging (Albany NY)* 2019;11(2):303–327.
- 706 39. Gibson J, Russ TC, Clarke T-K, et al. A meta-analysis of genome-wide association  
707 studies of epigenetic age acceleration. *PLoS Genet* 2019;15(11):e1008104.
- 708 40. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for  
709 Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res* 2016;44(3):e20.
- 710 41. Gruzieva O, Merid SK, Chen S, et al. DNA Methylation Trajectories During Pregnancy  
711 [Internet]. *Epigenet Insights* 2019;12
- 712 42. Chen S, Mukherjee N, Janjanam VD, et al. Consistency and Variability of DNA  
713 Methylation in Women During Puberty, Young Adulthood, and Pregnancy [Internet].  
714 *Genet Epigenet* 2017;9
- 715 43. Iqbal S, Lockett GA, Holloway JW, et al. Changes in DNA Methylation from Age 18 to  
716 Pregnancy in Type 1, 2, and 17 T Helper and Regulatory T-Cells Pathway Genes  
717 [Internet]. *Int J Mol Sci* 2018;19(2)
- 718 44. Fairfax BP, Makino S, Radhakrishnan J, et al. GENETICS OF GENE EXPRESSION IN  
719 PRIMARY IMMUNE CELLS IDENTIFIES CELL-SPECIFIC MASTER  
720 REGULATORS AND ROLES OF HLA ALLELES. *Nat Genet* 2012;44(5):502–510.
- 721 45. Chen M-H, Raffield LM, Mousas A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell  
722 Genetics in 746,667 Individuals from 5 Global Populations. *Cell* 2020;182(5):1198-  
723 1213.e14.
- 724 46. Rachidi M, Lopes C, Charron G, et al. Spatial and temporal localization during  
725 embryonic and fetal human development of the transcription factor SIM2 in brain

- 726 regions altered in Down syndrome. *International Journal of Developmental*  
727 *Neuroscience* 2005;23(5):475–484.
- 728 47. Zeydan B, Kantarci OH. Impact of Age on Multiple Sclerosis Disease Activity and  
729 Progression. *Curr Neurol Neurosci Rep* 2020;20(7):24.
- 730 48. de Barcelos IP, Troxell RM, Graves JS. Mitochondrial Dysfunction and Multiple  
731 Sclerosis. *Biology (Basel)* 2019;8(2):37.
- 732 49. Jokubaitis VG, Ibrahim O, Stankovich J, et al. Not all roads lead to the immune system:  
733 The Genetic Basis of Multiple Sclerosis Severity Implicates Central Nervous System  
734 and Mitochondrial Involvement [Internet]. 2022;2022.02.04.22270362.
- 735 50. van Langelaar J, Rijvers L, Smolders J, van Luijn MM. B and T Cells Driving Multiple  
736 Sclerosis: Identity, Mechanisms and Potential Triggers. *Front Immunol* 2020;11:760.
- 737 51. Kieffer TEC, Laskewitz A, Scherjon SA, et al. Memory T Cells in Pregnancy [Internet].  
738 *Frontiers in Immunology* 2019;10
- 739 52. Badam TV, Hellberg S, Mehta RB, et al. CD4+ T-cell DNA methylation changes during  
740 pregnancy significantly correlate with disease-associated methylation changes in  
741 autoimmune diseases. *Epigenetics* 2021;0(0):1–16.
- 742 53. Søndergaard HB, Airas L, Christensen JR, et al. Pregnancy-Induced Changes in  
743 microRNA Expression in Multiple Sclerosis. *Front Immunol* 2021;11:552101.
- 744 54. Hardardottir L, Bazzano MV, Glau L, et al. The New Old CD8+ T Cells in the Immune  
745 Paradox of Pregnancy. *Front Immunol* 2021;12:765730.

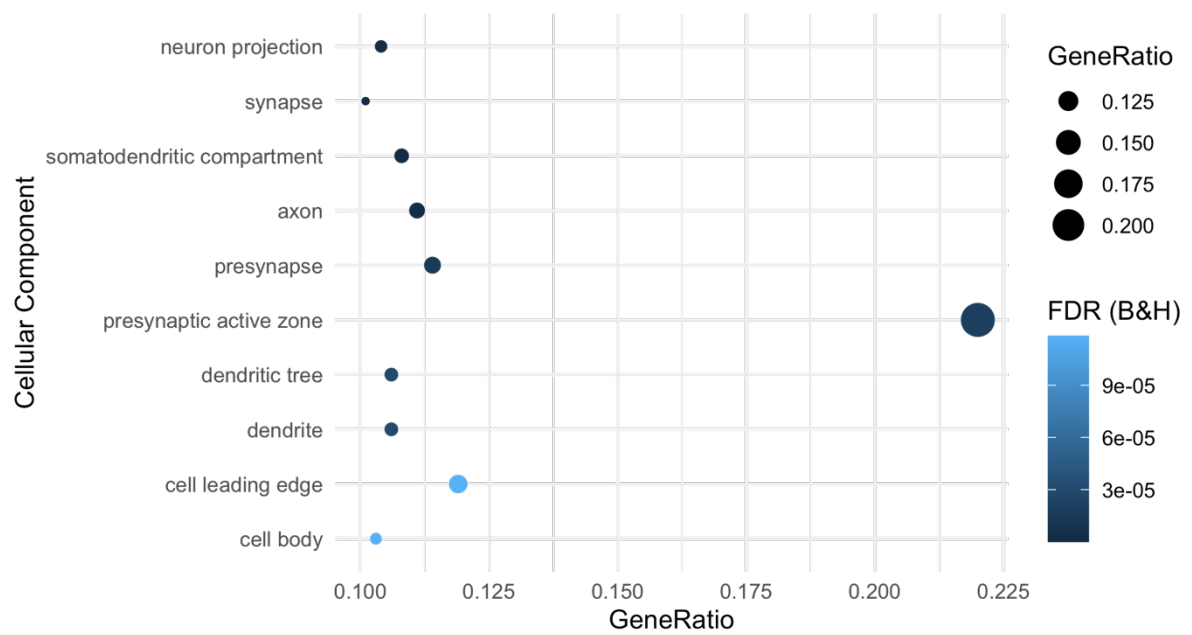
- 746 55. Yang S-L, Tan H-X, Niu T-T, et al. Kynurenine promotes the cytotoxicity of NK cells  
747 through aryl hydrocarbon receptor in early pregnancy. *J Reprod Immunol*  
748 2021;143:103270.
- 749 56. Cirac A, Tsaktanis T, Beyer T, et al. The Aryl Hydrocarbon Receptor-Dependent TGF-  
750  $\alpha$ /VEGF-B Ratio Correlates With Disease Subtype and Prognosis in Multiple Sclerosis.  
751 *Neurol Neuroimmunol Neuroinflamm* 2021;8(5):e1043.
- 752 57. Tsaktanis T, Beyer T, Nirschl L, et al. Aryl Hydrocarbon Receptor Plasma Agonist  
753 Activity Correlates With Disease Activity in Progressive MS [Internet]. *Neurology -*  
754 *Neuroimmunology Neuroinflammation* 2021;8(2)
- 755 58. Hoekzema E, Barba-Müller E, Pozzobon C, et al. Pregnancy leads to long-lasting  
756 changes in human brain structure. *Nat Neurosci* 2017;20(2):287–296.
- 757 59. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and  
758 healthspan. *Aging (Albany NY)* 2018;10(4):573–591.
- 759 60. Maltby VE, Lea RA, Ribbons KA, et al. DNA methylation changes in CD4+ T cells  
760 isolated from multiple sclerosis patients on dimethyl fumarate [Internet]. *Mult Scler J*  
761 *Exp Transl Clin* 2018;4(3)
- 762 61. Ewing E, Kular L, Fernandes SJ, et al. Combining evidence from four immune cell types  
763 identifies DNA methylation patterns that implicate functionally distinct pathways during  
764 Multiple Sclerosis progression. *EBioMedicine* 2019;43:411–423.
- 765

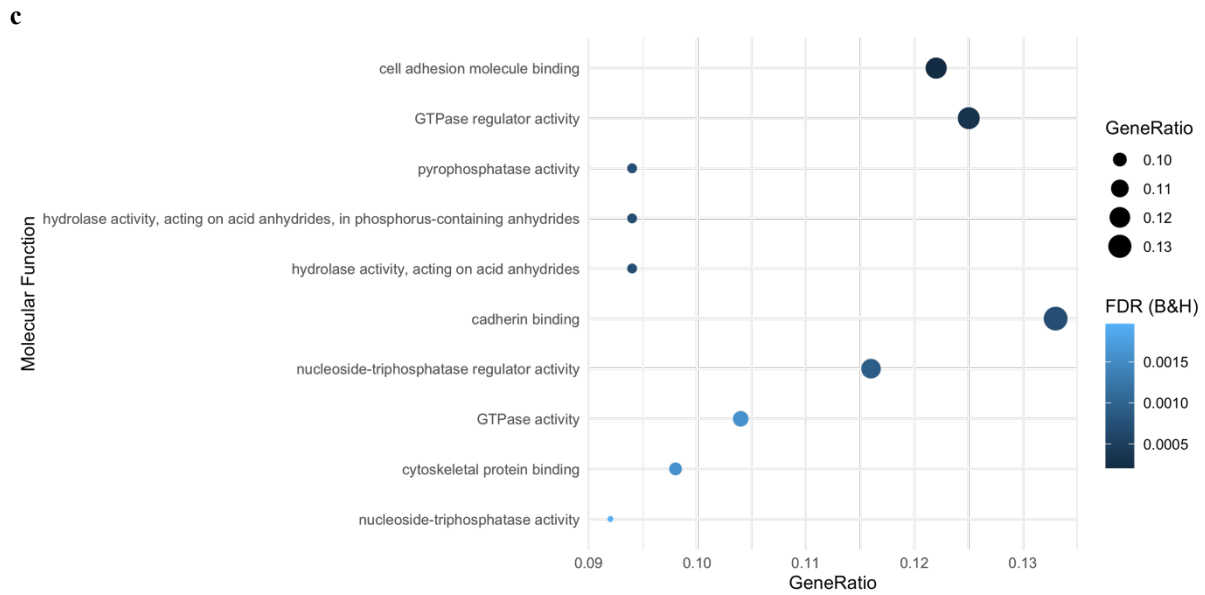
## Figures

**a**

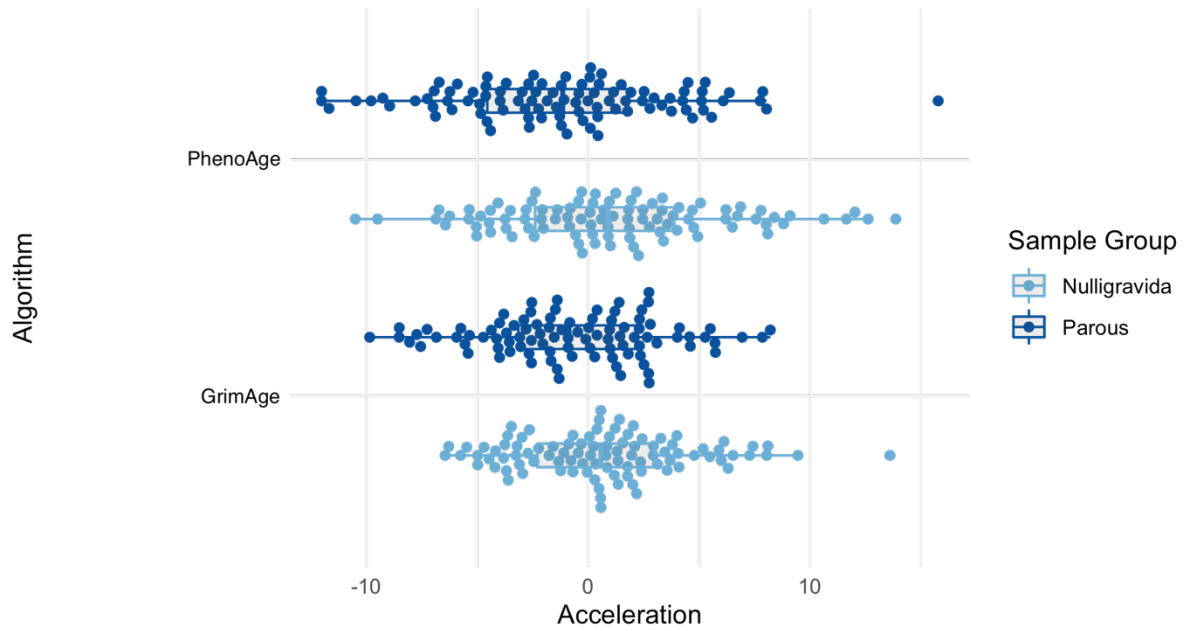


**b**





**Figure 1. Gene Set Enrichment Analysis (GSEA) using ToppGene API.** Input data were genes identified in both the differential methylation analysis and elastic net regression (n=1318). The **a**) ten most significantly enriched biological processes, **b**) ten most significantly enriched cellular compartments, **c**) ten most significantly enriched molecular functions. Gene ratio is the ratio of the number of genes in the query list and the hit count for that gene set in the genome.



**Figure 2. Methylation age acceleration (MAA) by sample group using the PhenoAge and GrimAge algorithms.** There are significant differences in MAA between groups using the PhenoAge and GrimAge algorithms. PhenoAge: nulligravida mean = 1.14 (SE = 0.502), parous mean = -1.14 (SE = 0.504), mean difference = 2.27 years,  $p = 0.001$ . GrimAge: nulligravida mean = 0.720 (SE = 0.749), parous mean = -0.720 (SE = 856), mean difference = 1.44 years,  $p = 0.005$ .

*Abbreviations: SE = standard error*

## Tables

**Table 1. Cohort summary statistics**

Characteristics		Nulligravida (n=96)	Parous (n=96)	All (n=192)	Cohen's d
<b>Time from last pregnancy to blood collection (years)</b>	Median (IQR)	NA	16.66 (9.13, 27.66)	NA	-
	Range	NA	1.45 – 44.42	NA	
<b>ARMSS score</b>	Median (IQR)	6.63 (1.47, 8.73)	7.08 (1.29, 8.22)	6.99 (1.39, 8.37)	0.01
	Range	0.16 – 9.55	0.19 – 9.91	0.16 – 9.91	
<b>Disease course</b>	RRMS	57 (60.0%)	63 (66.3%)	120 (63.2%)	NA
	SPMS	38 (40.0%)	32 (33.7%)	70 (36.8%)	
<b>Sex</b>	Female	124 (100.0%)	96 (100.0%)	220 (100.0%)	NA
	Male	0 (0%)	0 (0%)	0 (0%)	
<b>Age at most recent visit</b>	Median (IQR)	48.3 (40.7, 56.6)	48.6 (39.5, 57.2)	48.9 (40.7, 57.1)	0.03
	Range	27.6 – 70.6	24.2 – 69.8	24.2 – 70.6	
<b>Age at blood collection</b>	Median (IQR)	48.7 (41.2, 57.0)	48.9 (40.3, 57.9)	48.9 (40.7, 57.1)	0.03
	Range	28.3 – 70.6	26.8 – 69.8	26.8 – 70.6	
<b>Follow-up in MSBase (years)</b>	Median (IQR)	6.26 (3.46, 8.91)	6.92 (5.51, 9.42)	6.54 (4.16, 8.99)	0.26
	Range	0.00 – 24.80	0.00 – 19.30	0.00 – 24.80	
<b>Number of EDSS scores assessed</b>	Median (IQR)	7.5 (4.0, 9.0)	8.5 (6.0, 9.0)	8.0 (5.0, 9.0)	0.35
	Range	1.0 – 9.0	1.0 – 9.0	1.0 – 9.0	

<b>Symptom duration (years)</b>	Median (IQR)	15.77 (9.62, 24.60)	14.98 (8.84, 20.65)	15.11 (9.26, 22.15)	0.15
	Range	1.01 – 42.37	0.21 – 41.56	0.21 – 42.37	
<b>ARR in year preceding blood collection</b>	Median (IQR)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.08
	Range	0.0 – 3.0	0.0 – 2.0	0.0 – 3.0	
<b>Number of births prior to blood collection</b>	0	96 (100%)	0 (0%)	96 (50%)	-
	1	0 (0%)	29 (30%)	29 (15%)	
	2	0 (0%)	46 (48%)	46 (24%)	
	3	0 (0%)	14 (15%)	14 (7%)	
	≥4	0 (0%)	7 (7%)	7 (4%)	
	None	37 (38%)	34 (35%)	71 (37%)	
<b>DMT at blood collection</b>	Alemtuzumab	0 (0%)	1 (1%)	1 (1%)	-
	Dimethyl fumarate	2 (2%)	8 (8%)	10 (5%)	
	Fingolimod	21 (22%)	25 (26%)	46 (24%)	
	Glatiramer acetate	7 (7%)	3 (3%)	10 (5%)	
	Interferon beta	12 (13%)	10 (11%)	22 (11%)	
	Natalizumab	17 (18%)	11 (12%)	28 (15%)	
	Teriflunomide	0 (0%)	4 (4%)	4 (2%)	
	Unknown	63 (65.6%)	29 (30.2%)	92 (47.9%)	
<b>Smoking history at blood collection</b>	Ever	16 (16.7%)	39 (40.6%)	55 (28.6%)	-
	Never	17 (17.7%)	28 (29.2%)	45 (23.4%)	
<b>Smoking history at blood collection</b>	Unknown	63 (65.6%)	29 (30.2%)	92 (47.9%)	-
	Unknown	63 (65.6%)	29 (30.2%)	92 (47.9%)	

Abbreviations: ARMSS = Age-Related Multiple Sclerosis Severity Score, IQR = Interquartile Range, EDSS = Expanded Disability Status Scale, ARR = Annualised Relapse Rate, DMT = Disease Modifying Therapy



**Table 2. Top 10 differentially methylated positions (DMPs) by effect size ( $\Delta_{\text{meth}}$ )**

CpG	CHR	MAPINFO	$\Delta_{\text{meth}}$	FDR	Gene	Feature	CGI
cg12036633	15	63758958	-0.161	0.036		IGR	opensea
cg02122327	13	50194322	0.133	0.028		IGR	opensea
cg03885684	2	120770471	0.108	0.014	EPB41L5	TSS200	island
cg24000535	14	91110600	-0.083	0.037	LOC101928909	Body	opensea
cg08166072	2	46213920	0.080	0.017	PRKCE	Body	opensea
cg14248704	5	151470842	0.074	0.014	CTB-12O2.1	Body	opensea
cg10140164	9	75597328	-0.072	0.029		IGR	opensea
cg07723864	13	25670042	-0.069	0.029	PABPC3	TSS1500	island
cg06809965	14	70070333	-0.063	0.024		IGR	opensea
cg19938535	6	25341389	0.063	0.047	LRRC16A	Body	opensea

*Abbreviations: Chr = Chromosome, bp = base pair, CpG = cytosine-phosphate-guanine, CGI = CpG island, max = maximum, IGR = intergenic region, TSS200 = transcript start site (up to 200bp 5' of 5'UTR) promoter region, Body = gene body, TSS1500 = Transcript start site (up to 1500bp 5' of 5'UTR) promoter region*

**Table 3. Differentially methylated region (DMR)**

Chr	Start (bp)	End (bp)	Width	n CpGs	$\Delta_{max}^*$	$\Delta_{mean}^*$	Gene	Feature	CGI	CpG Names
7	151137882	151138295	413	3	0.029	0.023	CRYGN	TSS1500	Shore	cg23666844, cg16077872, cg17362899
15	67228722	67228986	264	3	0.049	0.039		IGR	Open Sea	cg26795333, cg20560283, cg17174814
17	4803684	4804357	673	3	-0.045	-0.042	CHRNE	Body	Island	cg22349396, cg06444025, cg01726265
18	72152075	72152314	239	3	0.046	0.033		IGR	Open sea	cg23973972, cg15682262, cg27477494
21	44782331	44782497	166	4	0.022	0.018	LINC01679	TSS200, IGR	Island	cg17577862, cg02260098, cg25191041, cg14081667

\*Change from nulligravida to parous groups

Abbreviations: Chr = chromosome, bp = base pair, CpG = cytosine-phosphate-guanine, CGI = CpG island, max = maximum, CRYGN = crystallin gamma N, CHRNE = cholinergic receptor nicotinic epsilon subunit, LINC01679 = long intergenic non-coding RNA 1679, TSS1500 = transcript start site (up to 1500bp 5' of 5'UTR) promoter region, TSS200 = transcript start site (up to 200bp 5' of 5'UTR) promoter region, IGR = intergenic region

**Table 4. Cell specific differentially methylated positions (csDMPs)**

CpG	Est.	SE	p	Nulligravida mean*	Parous mean*	Direction of effect**	Chr	bp	Gene	Feature
<i>CD4+ T cells</i>										
cg14172633	-3.64	0.62	1.7x10 <sup>-8</sup>	-3.10	-2.84	Hyper.	1	185703557	<i>HMCN1</i>	TSS200
cg15145296	-3.23	0.56	3.6x10 <sup>-8</sup>	-4.24	-3.98	Hyper.	3	125709740		
cg06032337	-3.17	0.56	5.1x10 <sup>-8</sup>	-2.82	-2.71	Hyper.	6	29648468		
cg06818823	-5.88	1.04	5.8x10 <sup>-8</sup>	-6.55	-5.90	Hyper.	6	46459236		
<i>CD8+ T cells</i>										
cg01858500	-4.27	0.71	1.1x10 <sup>-8</sup>	-3.57	-3.52	Hyper.	17	68202566		
cg08944170	-2.78	0.48	2.3x10 <sup>-8</sup>	-3.38	-3.60	Hypo.	1	248100614	<i>OR2L13</i>	1stExon
cg25577322	-8.32	1.45	4.0x10 <sup>-8</sup>	-6.33	-6.94	Hypo.	7	17338213	<i>AHR</i>	TSS200
cg16402757	-2.14	0.38	4.6x10 <sup>-8</sup>	-2.14	-2.05	Hyper	10	35311004	<i>CUL2</i>	Body
cg03495768	-3.05	0.54	6.1x10 <sup>-8</sup>	-2.98	-3.07	Hypo.	13	100637113	<i>ZIC2</i>	Body
cg04798314	-2.71	0.48	6.7x10 <sup>-8</sup>	-2.35	-2.56	Hypo.	1	246668601	<i>SMYD3</i>	Body
cg11738485	-2.42	0.43	6.8x10 <sup>-8</sup>	-2.79	-2.51	Hyper	19	12877000	<i>HOOK2</i>	Body
cg20507276	-2.68	0.48	8.6x10 <sup>-8</sup>	-3.40	-3.55	Hypo.	1	248100600	<i>OR2L13</i>	1stExon

\*Mean M-values reported as M-values used in cell-specific statistical analyses

\*\*Nulligravida to parous

Abbreviations: CpG = cytosine-phosphate-guanine, Est = estimate (from linear regression), SE = standard error, Hyper. = hypermethylated, Chr = chromosome, bp = base pair, TSS200 = transcript start site (up to 200bp 5' of 5'UTR) promoter region

**Table 5. Top ten CpGs associated with parity as selected by the elastic net model**

<b>CpG</b>	<b>Importance</b>	<b>Chr</b>	<b>Position (bp)</b>	<b>Gene</b>	<b>Feature</b>
<b>cg08186508</b>	100.00	14	68067006	<i>PIGH</i>	5'UTR
<b>cg26506013</b>	99.95	5	133984553	<i>SEC24A</i>	1stExon
<b>cg23841819</b>	99.92	1	204970383	<i>NFASC</i>	Body
<b>cg25485991</b>	99.86	17	8066461	<i>VAMP2</i>	TSS200
<b>cg23367339</b>	99.74	17	36622717	<i>ARHGAP23</i>	Body
<b>cg17070338</b>	99.69	13	111268441	<i>CARKD</i>	Body
<b>cg07360021</b>	99.49	6	151186904	<i>MTHFD1L</i>	1stExon
<b>cg11918372</b>	99.14	2	48132755	<i>FBXO11</i>	5'UTR
<b>cg27573735</b>	99.04	3	82857144		
<b>cg12835012</b>	98.88	4	183795785		

Abbreviations: Chr = Chromosome, bp = base pairs