1 **Persistent DNA methylation changes associated with prenatal NO₂ exposure in a**

2 **Canadian prospective birth study**

3

4 **Authors**:

5 Samantha Lee[1], Hind Sbihi[2,3], Julia L. MacIsaac[4], Padmaja Subbarao[5], Piushkumar J.

6 Mandhane[6], Theo J. Moraes[5], Stuart E. Turvey[7], Qingling Duan[8,9], Amirthagowri

7 Ambalavanan[9], Michael Brauer[3], Jeffrey Brook[10], Michael S. Kobor[4], Meaghan J. Jones[1]

8

9 **Affiliations**:

10 1. Department of Biochemistry and Medical Genetics, University of Manitoba, and
11   Children's Hospital Research Institute of Manitoba
12 2. British Columbia Centre for Disease Control
13 3. School of Population and Public Health, University of British Columbia
14 4. Department of Medical Genetics, University of British Columbia, and BC
15   Children's Hospital Research Institute
16 5. Department of Pediatrics, University of Toronto
17 6. Department of Pediatrics, University of Alberta
18 7. Department of Pediatrics, University of British Columbia
19 8. School of Computing, Queen's University
20 9. Department of Biomedical and Molecular Sciences, Queen's University
21 10. School of Public Health, University of Toronto

22

23

24

25

26

27

28

29

30

1

Abstract

Background: Accumulating evidence suggests prenatal air pollution exposure alters DNA methylation (DNAm), which could go on to affect long-term health. However, it remains unclear whether prenatal DNAm alterations persist through early life. Identifying DNAm changes that persist from birth into childhood would provide greater insight into the molecular mechanisms that most likely contribute to the association of prenatal air pollution exposure with health outcomes such as atopic disease.

Objectives: This study investigated the persistence of DNAm changes associated with prenatal NO_2 exposure (a surrogate measure of traffic-related air pollution) at age one to begin characterizing which DNAm changes most likely to contribute to atopic disease.

Methods: We used an atopy-enriched subset of CHILD study participants (N=145) to identify individual and regional cord blood DNAm differences associated with prenatal NO_2, followed by an investigation of persistence in age one peripheral blood. As we had repeated DNAm measures, we also isolated postnatal-specific DNAm changes and examined their association with NO_2 exposure in the first year of life. MANOVA tests were used to examine the association between DNAm changes associated with NO_2 and child wheeze and atopy.

Results: We identified 24 regions of altered cord blood DNAm, with several annotated to *HOX* genes. Two regions annotated to *MPDU1* and *C5orf63* were significantly associated with age one wheeze. Further, we found the effect of prenatal NO_2 exposure across CpGs within all altered regions remained similar at age one. A single region of postnatal-specific DNAm annotated to *HOXB6* was associated with year one NO_2 and age one atopy.

Discussion: Regional cord blood DNAm changes associated with prenatal NO_2 exposure persist through at least the first year of life, and some of these changes are associated with

2

3

61   age one wheeze. The early-postnatal period remains a sensitive window to DNAm

62   perturbations that may also influence child health.

63 **Introduction**

64   Prenatal and early-life air pollution exposure are associated with an increased risk

65 of childhood asthma, which is characterized by wheezing, difficulty breathing, chest

66 tightness, and bronchial hyper-responsiveness[1,2]. Often, childhood asthma is preceded by

67 asthma-like symptoms and immune disorders, including wheeze and allergies (atopy), in

68 what is known as the atopic march[3]. This progression involves both genetic and

69 environmental factors and is driven by type 2 inflammation[4]. While the exact

70 mechanisms connecting prenatal air pollution exposure to childhood asthma and atopic

71 disease remain unclear, accumulating evidence suggests that biological embedding of

72 increased oxidative stress and concomitant inflammatory signalling may play a role[5].

73 Biological embedding refers to the notion that early-life experiences and exposures can

74 "get under the skin" to influence health and development[6]. At the cellular level, this

75 process may involve epigenetic alterations that regulate DNA packaging and expression

76 and impact how cells respond to future stimuli[6]. One example of epigenetic modification

77 is DNA methylation (DNAm), which refers to the addition of a methyl group on the 5'

78 position of a cytosine in a cytosine-guanine pair (termed "CpG"). Though DNAm

79 patterns are established early in life, they remain malleable in response to environmental

80 exposures. While some of these DNAm perturbations exist only transiently, others persist

81 through development[7]. Given that persistent DNAm alterations are more likely to affect

82 long-term cell function than transient changes, persistent DNAm changes should be of

83 most interest as they are more likely to contribute to development of child asthma

84 following prenatal air pollution exposure than transient DNAm changes[7].

85   Despite the notion that persistent DNAm alterations are more likely to contribute

86 to clinically-relevant health outcomes, most studies examining prenatal air pollution

87 exposure only identify DNAm changes at a single time point, typically at birth. These

88 studies suggest that prenatal air pollution exposure at least transiently alters cord blood

89 DNAm in cellular pathways related to oxidative stress and inflammation. For example, a

90 meta-analysis of prenatal NO_2 exposure across several European cohorts revealed

91 significant cord blood DNAm alterations in *LONP1* and *SLC25A28*, which are related to

92 mitochondrial oxidative stress, as well as in *PVLAP,* which plays a role in leukocyte

93 trafficking[8]. In the EARLI study, prenatal NO_2 exposure is associated with lower *RNF39*

94   DNAm[9]. The *RNF39* protein is part of Major Histocompatibility Complex (MHC) class I,

95   thus it is biologically plausible that altered *RNF39* DNAm may be associated with

96   changes in immune function[10], including asthma. More recently, DNAm at *CASP7* was

97   associated with prenatal NO_2 exposure across all trimesters in a Korean cohort[11]. The

98   *CASP7* gene is a critical mediator of mitochondria-induced apoptosis, thus this finding is

99   consistent with previous observations of altered DNAm in mitochondrial oxidative stress

100   pathways[8,12]. Finally, prenatal NO_2 exposure is associated with lower LINE DNAm at

101   age 11 and is also correlated with higher blood pressure at age 11, suggesting that

102   persistent changes in LINE DNAm induced by prenatal NO_2 exposure may affect long-

103   term child cardio-respiratory outcomes[13]. However, a major limitation of that study was

104   the inability to discern the additional contribution of postnatal air pollution exposure on

105   DNAm and blood pressure.

106   The postnatal period, especially the first year of life, is considered part of the

107   developmental window that is especially sensitive to biological embedding as DNAm

108   patterns, lungs, and the innate immune system are still developing, and there is an

109   increasing diversity of exposures and experiences[14–16]. In an epidemiological cross-

110   sectional analysis prenatal and postnatal passive cigarette smoke exposures were shown

111   to have independent effects on the risk of negative respiratory health outcomes,

112   supporting the notion that the postnatal period remains a sensitive window to exposures[17].

113   While several studies have identified positive correlations between postnatal air pollution

114   exposure and child asthma, the role of postnatal-specific DNAm changes in this

115   relationship has not been investigated[18–20]. This knowledge would help clarify the

116   underlying molecular mechanisms connecting postnatal air pollution exposure to

117   respiratory health and provide insight into how they differ from prenatal mechanisms.

118   However, it is not possible to directly measure postnatal-specific changes in DNAm as

119   DNAm measured in peripheral blood represents changes occurring across the prenatal

120   and postnatal periods. Instead, postnatal-specific DNAm changes must be isolated *in*

121   *silico* by subtracting measures of cord blood DNAm from measures of peripheral blood

122   DNAm.

123   The purpose of this study was to examine the persistence of cord blood DNAm

124   changes associated with prenatal NO_2 exposure to identify the biological pathways that

125   most likely to contribute to pollution-associated childhood asthma. To accomplish this

126   goal, we first identified NO_2-associated DNAm changes at individual sites and local

127   regions using hypothesis-driven and discovery-based approaches. Then, we used linear

128   mixed models to examine whether the effect of prenatal NO_2 exposure on DNAm

129   remained similar at birth and age one. As we had repeated measures of participant DNAm

130   across time points, we were able to subtract cord blood DNAm measures from DNAm

131   measured at age one, allowing us to isolate and examine postnatal-specific changes in

132   DNAm occurring during the first year of life. This analysis is unique because it helps to

133   demonstrate that the first year of life remains a critical window for the biological

134   embedding of early air pollution exposures in addition to and possibly independent from

135   the prenatal period. Finally, we examined the correlation between significant prenatal

136   DNAm and postnatal-specific DNAm changes with predictors of childhood asthma at age

137   one to provide greater insight into the affected pathways that most likely to impact child

138   health outcomes.

139

140

141   **Methods**

142

143   *Study population*

144   The CHILD study (CHILD) is a prospective birth cohort designed to investigate

145   the developmental origins of childhood allergic disease, including asthma[21,22]. Between

146   2008 and 2012 a total of 3624 eligible mothers were recruited from four major Canadian

147   cities: Vancouver, Edmonton, Winnipeg/Winkler, and Toronto. Eligible participants had

148   to be ≥ 18 years of age (19 years in Vancouver), communicate in English, reside within

149   reasonable proximity to a study centre and plan to give birth there, intend to continue

150   living near a study centre, be willing to provide informed consent, and provide

151   information for two alternate contacts[21]. Children born before 35 weeks gestation, born

152   with major congenital abnormalities or respiratory distress syndrome, conceived by *in*

153   *vitro* fertilization, resulting from multiple births, and/or who spend less than 80% of

154   nights in the index home were excluded from the study[21]. Each study centre obtained

155   ethics approval from their research ethics board, and CHILD was approved by the

156   Hamilton Integrated Ethics Board (certificate number 07-2929). This study uses data

157   from an atopy-enriched subset of 145 CHILD participants and was approved by the

158   University of Manitoba Research Ethics Board (HS22880)[23].

159

160   *Measurement of maternal characteristics*

161   We included the *a priori* selected maternal variables education length (a surrogate

162   for socioeconomic status) and smoking in our linear regressions to account for biological

163   variation. Maternal characteristics were assessed by CHILD study health professionals at

164   several time points throughout the prenatal period and first year of life[24]. Maternal

165   education length was self-reported at initial intake and used a surrogate measure of

166   socioeconomic status in all models. Prenatal maternal smoking was defined at 18 weeks

167   gestation as an affirmative response to the question "At the present time, how often do

168   you smoke" or to the question "During this pregnancy, did you cut down (but not

169   completely stop) smoking?" or by any non-missing response to the questions "During this

170   pregnancy, did you completely stop smoking?", "At what week did you stop smoking?",

171   and "At what week of pregnancy did you cut down smoking". Postnatal maternal

172   smoking was defined at the end of year one as an affirmative response to the question "At

173   the present time do you smoke?".

174

175   *Air pollution exposure*

176   Average prenatal exposure to outdoor NO_2 was estimated separately for

177   Vancouver, Edmonton, Winnipeg, and Toronto as previously described[25–27]. Briefly, city-

178   specific land use regression (LUR) models were used to estimate outdoor NO_2

179   concentration at the residential location of each participant for the prenatal period

180   (estimated time of conception until birth) and the first year of life (birth until age

181   one)[22,28]. All models included information on land use, roads and traffic, population,

182   physical geography, and meteorology; consequently, in addition to being related to

183   combustion, NO_2 is typically considered an indicator of traffic-related air pollution

184   (TRAP)[28]. Thus, for CHILD subjects whose NO_2 exposures are estimated to be high, NO_2

185   exposure broadly represents exposure to a mixture of combustion-related emissions,

186   including a range of particles and gases with varying potential toxicities. For participants

187   that reported a change in primary residence during the study period (N=23), outdoor NO_2

188   estimates were computed using a time-weighted average of exposure at each reported

189   residence[28]. All outdoor NO_2 estimates were temporally adjusted on a biweekly basis

190   using local fixed-site ambient monitoring data collected by the National Air Pollution

191   Surveillance (NAPS) network[28]. Participants with neither prenatal or year one outdoor

192   NO_2 estimates were omitted from the study (N = 2 participants)[28].

193

194   *DNA methylation measurement and quality control*

195   At each delivery site umbilical cord blood was collected into heparinized vacuum

196   tubes. For each individual, heparinized blood samples were combined in a single 50 mL

197   conical tube, which was processed to obtain cord blood mononuclear cells (CBMCs)[21].

198   CBMC pellets were resuspended in Buffer RLT-Plus (600 uL per mL; Qiagen) and stored

199   at -80C for later nucleic acid extraction. Two peripheral blood aliquots (1.5 mL to 3 mL

200   each) collected at age one were subjected to the same procedures to obtain peripheral

201   blood mononuclear cells (PBMCs)[21]. CBMC and PBMC samples were used to generate

202   DNAm data as previously described[21]. Briefly, 750 ng of extracted DNA (Qiagen

203   DNAeasy kit) was bisulfite converted (Zymo EZ DNAm Gold kit), followed by

204   application of 160 ng of bisulfite-converted DNA to the Illumina HumanMethylation

205   450k array. All nucleic acid procedures followed the manufacturer's instructions.

206   Preprocessing and quality control of DNAm data was performed in *minfi*, except

207   where noted[29]. Sample and probe quality were assessed using detection p-values. One

208   cord blood sample with 20751 bad probes (detection p-value > 0.01) was removed prior

209   to normalization. Background correction and dye-bias normalization were accomplished

210   using *preprocessNoob*, followed by probe bias correction using *BMIQ* from the

211   *wateRmelon* package[30,31]. Next, we removed 8,482 probes with a detection P > 0.01

212   and/or with fewer than 3 beads contributing to signal, 15,333 probes with a known single

213   nucleotide polymorphism (SNP), 10,860 probes localized to sex-chromosomes, and

214   26,183 cross-reactive probes leaving 424,644 of the initial 485,512 probes[32]. Quality

215   control was performed to confirm sample labelling (by examining participant expected

216   sex) and replicate quality. All sample replicates (n=3 cord blood samples with 3-4

217   replicates each) were of good quality based on strong and distinct methylated and

218   unmethylated β-value peaks. Thus, one replicate from each set was randomly selected to

219   include in DNAm analysis.

220   Measures of DNAm are quantified as the ratio of methylated probe intensities to

221   intensities of all measured probes for a given CpG, known as β-values, or as M-values,

222   which are log-transformed β-values[33]. Both β-values and M-values are commonly used in

223   studies investigating epigenome-wide DNAm changes. M-values are often regarded as

224   more statistically valid as they are homoscedastic, but they are not biologically

225   interpretable[33]. In contrast, the 0-1 scale of β-values corresponds to the methylation

226   percentage of a given CpG site[33]. Additionally, as β-values are linear they can be added

227   or subtracted from one another without distorting DNAm measurements. Therefore, we

228   used DNAm β-values in this study as they were compatible with our analyses of

229   examining the occurrence of DNAm changes in the first year of life.

230   Specifically, we were interested in examining the effects of NO_2 exposure during

231   the first year of life on postnatal DNAm alterations; however, it is not possible to

232   measure postnatal-specific DNAm changes directly. This is because peripheral blood

233   DNAm interrogated in childhood represents DNAm changes occurring across both the

234   prenatal and postnatal periods. To isolate postnatal-specific DNAm changes we used a

235   novel method wherein we subtracted cord blood β-values from age one blood β-values

236   and used the resulting differences in β-values in our analyses, termed "postnatal-specific"

237   β-values. Postnatal-specific β-values represent DNAm changes occurring specifically in

238   the first year of life.

239

240   ***Estimation of cord blood monocyte proportions***

241   DNAm is cell type specific and differences in cell type composition between

242   samples may lead to spurious findings unless accounted for in linear models. Each cell

243   type has a unique DNAm signature; therefore, DNAm patterns can be used to estimate

244   cell type proportions in blood samples using a reference-based deconvolution method[34].

245   As cell types were not directly measured (e.g. by flow cytometry) for most CHILD

246   participants, we estimated cord blood cell type proportions using a reference-based

247   method[34]. Specifically, we used the *FlowSorted.CordBloodCombined.450k* cord blood

248   reference set and the *estimateCellCounts2* function from the *FlowSorted.Blood.EPIC*

249   *package* to estimate the relative proportions of 6 cell types (CD4 T-cells, CD8 T-cells, B-

250   cells, NK cells, monocytes, and nucleated red blood cells) in CBMC samples either using

251   default cord blood probe selection methods ("any" top 100 probes and "both" the 50 top

252   hypomethylated and 50 top hypermethylated probes) or using a predefined set of *IDOL*

253   probes selected to optimize cords blood deconvolution[35–37]. We based our probe selection

254   method choice on whether the probe selection parameter produced cell types estimates

255   within expected proportions and whether cell type estimates differed between cord blood

256   between probe selection parameters based on Pearson correlation coefficients (r) of direct

257   comparisons (**Supplementary Figure 1**). Based on these comparisons, we proceeded to

258   estimate CBMC proportions (**Supplementary File 1**) using the default "any" parameter

259   of *estimateCellCounts2* for deconvolution of cord blood.

260   The compositional properties of cell type estimates require they sum to 100% and

261   can introduce multicollinearity into linear regressions[38]. To remove constraints imposed

262   by cell type composition, we performed isometric-log ratio (ILR) transformations

263   followed by robust principal component analysis (rPCA) on predicted CBMC

264   proportions. We subsequently included the first four principal components (PCs) in linear

265 regressions to account for ~90% of the variance due to differences in estimate cord blood

266 cell type proportions and removed 565 unique probes used to estimate cord blood cell

267 proportions (**Supplementary File 2**) from the total number of cord blood DNAm probes,

268 leaving 424079 probes remaining in our epigenome-wide analysis of prenatal NO_2

269 exposure

270

271 *Estimation of age one peripheral blood monocyte proportions*

272   There is no existing reference set for the estimation of cell proportions in age one

273 blood. Thus, we conducted a preliminary analysis comparing cord blood and adult blood

274 deconvolution procedures to determine which reference set (cord or adult) and probe

275 selection parameters ("any", "both", or "IDOL") are appropriate for age one PBMC

276 prediction[36,37]. We based our selection on whether each combination of reference set and

277 probe selection parameter produced cell types estimates within expected proportions and

278 whether cell type estimates differed between cord blood or adult blood reference sets

279 based on Pearson correlation coefficients (r) of direct comparisons (**Supplementary**

280 **Figure 2**). Cord blood and adult blood reference sets using the "any" probe selection

281 parameter performed similarly (r≥0.92). Therefore, we proceeded with the

282 *FlowSorted.CordBloodCombined.450k*[35] reference set using the *estimateCellCounts2*

283 function with the default cord blood probe selection parameter "any" to estimate the

284 relative proportions of 5 cell types (CD4 T-cells, CD4 T-cells, B-cells, NK cells,

285 monocytes) in age one PBMC samples (**Supplementary File 3**) for consistency with our

286 CBMC deconvolution procedure. We performed ILR-transformation followed by rPCA

287 on estimated PBMC proportions and used the first three PCs in linear regressions to

288 account for >90% of variation related to differences in estimated cell type proportions.

289 We removed the 565 unique probes used to estimate age one blood cell proportions

290 (**Supplementary File 4**) from the total number of probes measuring age one blood

291 DNAm for a total of 424079 remaining probes.

292

293 *Variation due to cell type in postnatal-specific measures of DNA methylation*

294   The analysis of postnatal-specific DNAm changes requires consideration of

295 variation in both CBMC and PBMC cell types. In this case, ILR-transformed CBMC and

11

296 PBMC cell types were combined prior to rPCA and analyzed together. We included the

297 first six PCs accounting for ~90% of the variance due to combined CBMC and PBMC

298 cell types in our analyses examining postnatal-specific DNAm changes. We removed the

299 565 unique probes shared between the probes used to estimate cord blood and age one

300 blood cell proportions from the total number of probes passing quality control for a total

301 of 424079 remaining probes interrogating postnatal-specific DNAm.

302

303 *Variation due to cell type in investigations of DNA methylation persistence*

304 We previously generated cord blood and peripheral blood cell-type PCs separately

305 to correct for differences in cell proportions at birth and age one. The top cord bloods PCs

306 do not necessarily capture variation due to the same cell types as the top age one blood

307 PCs, and thus it is not appropriate to model coefficients across cord blood PCs and age

308 one blood PCs in linear mixed models using to investigate DNAm persistence. Therefore,

309 we combined estimates of cord blood cell proportions with estimates of age one

310 peripheral blood cell proportions in a row-wise manner. To facilitate combination of

311 these data, we created a new variable in the estimates of age one peripheral blood cell

312 proportions for nRBCs and set this value to zero as this cell type is not normally present

313 in healthy children at age one[39]. Combining estimates of cord blood and peripheral blood

314 cell proportion in this manner ensures the resulting cell type PCs reflect variation due to

315 the same cell types across both time points. Afterwards, we conducted ILR and rPCA on

316 the row-wise combined estimates of cell proportions and included the first three PCs

317 accounting for >90% of the variance in our linear mixed models examining the

318 persistence of cord blood DNAm at age one.

319

320 *Genotyping and population substructure*

321 A total of 2967 CHILD participants were genotyped using the Illumina

322 HumanCoreExome BeadChip, including the 145 participants used in this study. Quality

323 control and imputation were performed on all genotyped participants together before

324 selection of the subset of participants used in this study. Briefly, cord blood was collected

325 at birth as described above. The QiaSymphony automated large sample nucleic acid

326 purification system was used to extract DNA for genotyping (Genetic and Molecular

327   Epidemiology Laboratory, McMaster University)[21]. Quality control was carried out at the

328   subject and SNP levels using Plink (v1.9)[40]. Specifically, we omitted 39 participants with

329   missing genotypes >10% and excess heterozygosity (±2 standard deviations), 56

330   participants with discordant sex, and 37 participants that were likely third-degree

331   relatives as identified by pairwise Identify-By-Descent (IBD coefficient 0.185)[41]. At the

332   SNP level QC, we selected for the autosomal variants, and omitted variants exceeding

333   missingness of >0.05. A total of 2835 subjects remained with 515,033 variants passing

334   QC. Imputations were performed using the HRC (r1.1 2016) reference panel

335   implemented through the Michigan Imputation server[42]. Chromosomes were phased

336   using ShapeIT (v2.r790), then SNPs were imputed via the Minimac algorithm[43,44]. This

337   resulted in 28 million SNPs, which included 5.4 million common SNPs (MAF >0.05) that

338   passed the imputation quality score R2>0.3[44]. Quality control measures as above were

339   applied to the imputed markers and no markers or subjects were excluded post-

340   imputation. Afterwards, we performed a PCA of CHILD study genotype data (N=2835)

341   to assess population substructure. We estimated the eigenvectors of CHILD participants

342   using the EIGENSTRAT–SMARTPCA tool[45–47]. We estimated that the first 10

343   eigenvectors account for >99.9% of the total variance among subjects. We included the

344   first three genotyping PCs in our linear regressions to account for differences related to

345   genetic ancestry as the first three genotyping PCs resulted in clear separation of

346   participants used in this study (**Supplementary Figure 3**).

347

348   ***Candidate gene analyses of prenatal and year one NO₂ exposure***

349   The underlying mechanisms of prenatal air pollution exposure include increased

350   oxidative stress and inflammatory signalling and the hypothesized biological embedding

351   of these events in DNAm may contribute to atopic disease in childhood[48]. To investigate

352   this notion, we examined prenatal and postnatal-specific DNAm changes associated with

353   prenatal and year one NO_2 exposure, respectively, across a list of 59 *a priori*-determined

354   candidate genes hypothesized to participate in the cellular pathways activated by air

355   pollution exposure or that were previously identified in epigenome-wide association

356   studies of prenatal air pollution (**Supplementary Table 1**)[8,9,49–53,53–59,59–66]. We also

357   examined whether two previously identified DNAm changes within the genes *LONP1*

13

358   and *SLC25A2* could be replicated in our analysis[8]. Candidate gene analyses were

359   conducted using multivariable linear models using the *lm* function. Our candidate gene

360   analysis of prenatal NO_2 exposure included the following covariates: sex, gestational age,

361   prenatal maternal smoking, maternal education length, cell type, and genetic ancestry. We

362   additionally included 17 surrogate variables (SVs) to account for technical variation, as

363   determined by the *SVA* package (**Supplementary Figure 4**)[67]. When examining

364   postnatal-specific DNAm changes in our candidate genes, we included variables to adjust

365   for sex, year one maternal smoking, maternal education length, cell type, genetic

366   ancestry, and technical variation using 15 SVs as determined by the SVA package

367   (**Supplementary Figure 5**). We did not include study site as a covariate for either

368   model, or in any of the models below, as most of the variation in prenatal and year one

369   NO_2 estimates occurred between study sites (i.e. including study site would limit our

370   power to detect small DNAm changes; **Supplementary Figures 6-8**). The prenatal and

371   postnatal-specific candidate gene analyses were corrected for multiple comparisons using

372   the Bonferroni method. Significant DNAm changes within candidate genes were defined

373   as having an adjusted p<0.05 and displaying an absolute effect size >0.025% DNAm.

374

375   ***Epigenome-wide analyses of prenatal and year one NO_2 exposure***

376   We conducted epigenome-wide analyses to identify novel cord blood and

377   postnatal-specific DNAm changes associated with prenatal and year one NO_2 exposure,

378   respectively. Both analyses used the *limma* R package to conduct multivariable linear

379   regressions on DNAm β-values. Models investigating effects of prenatal NO_2 exposure

380   on cord blood DNAm included covariates to adjust for sex, gestational age, prenatal

381   maternal smoking, maternal education length, cell type, and genetic ancestry, and

382   technical variation using 17 SVs as suggested by the SVA package (**Supplementary**

383   **Figure 4**)[67]. Models examining the effect of year one NO_2 exposure on postnatal specific

384   DNAm included covariates to adjust for sex, year one maternal smoking, maternal

385   education length, cell type, genetic ancestry, and technical variation using 15 SVs as

386   determined by the *SVA* package (**Supplementary Figure 5**)[67]. For each analysis the

387   resulting p-values were adjusted for false discovery using the Benjamini-Hochberg

388   method. Individual CpGs were considered significant if they surpassed FDR p<0.05 and

14

389   exhibited an absolute effect size >0.025% DNAm.

390

391   ***Identification and characterization of differentially methylated regions associated with***

392   ***prenatal and year one NO$_2$ exposure***

393   Differentially methylated regions (DMRs) refer to adjacent CpGs that display

394   altered DNAm between phenotypes or exposures[68]. Region-based analyses are

395   advantageous as they are typically more statistically powerful than epigenome-wide

396   analyses as they can leverage information from adjacent CpGs. Additionally, DMRs can

397   offer greater insight into biologically-relevant DNAm changes, as regions of altered

398   DNAm are more likely to affect cell function than individual altered CpGs. We used the

399   *combp* function from the *ENmix* R package to identify cord blood and postnatal-specific

400   DMRs associated with prenatal and year one NO$_2$ exposure, respectively[69,70]. The *combp*

401   method spatially adjusts p-values obtained from an epigenome-wide investigation to

402   identify DMRs[69]. We considered *combp* DMRs to be statistically significant if they

403   surpassed FDR p<0.05 and exhibited a mean absolute effect size >0.025% DNAm.

404   We were interested in whether observed DNAm changes represented gene-by-

405   environment effects, but lacked power to test whether identified CpGs represented

406   methylation trait quantitative loci (mQTLs)[71]. Instead, we cross-checked all CpGs

407   contained with cord blood (N=151 CpGs) and postnatal-specific (N=6 CpG) DMRs with

408   mQTLdb, an online database of methylation quantitative trait loci (mQTLs)[72]. We

409   compared cord blood DMR CpGs identified in this study with mQTLs surpassing p<1E-

410   14 and exhibiting an effect size >0.02% (**Supplementary File 5**)[72]. Postnatal-specific

411   DMR CpGs represent changes that occur in the first year of life; however, the average

412   age of childhood mQTLs reported on mQTLdb is 7.5 ± 0.2 years, which represents a

413   different developmental stage[72]. Therefore, we compared our postnatal-specific DMR

414   CpGs with mQTLs reported at both birth and childhood on mQTLdb (**Supplementary**

415   **Files 6 and 7**).

416   Gene ontology analysis can provide insight into the biological pathways that are

417   affected by regional changes in DNAm. We performed gene ontology analysis on cord

418   blood DMRs using *goRegion* from the *missMethyl* R package with default settings[73]. We

419   considered gene ontology terms to be significant when FDR p<0.05. We did not perform

15

420   gene ontology analysis on postnatal-specific DMRs as we identified only one DMR at

421   this time point.

422

423   ***Persistence of individual and regional cord blood DNA methylation changes at age one***

424   Previous studies have not examined whether cord blood DNAm changes

425   associated with prenatal air pollution exposures persist through early childhood

426   development. Persistent DNAm changes are of interest because they are more likely to

427   impact child health than transient changes[7]. Therefore, we aimed to conduct a novel

428   analysis examining whether DNAm changes present at birth persisted at age one in

429   CHILD participants. To determine persistence of cord blood DNAm changes at age one,

430   we used linear mixed effects models to examine whether the coefficient of prenatal NO_2

431   exposure significantly differed in cord blood at birth and peripheral blood at age one). We

432   applied these linear mixed models to the top 100 CpGs identified in our epigenome-wide

433   analysis of prenatal NO_2 exposure that exhibited an absolute effect size >0.025% DNAm

434   and FDR p <0.05. In a separate analysis we applied linear mixed models to all individual

435   CpGs (N= 151) contained within cord blood DMRs associated with prenatal NO_2

436   exposure. In both analyses, linear models were run using the *lme4* package. Specifically,

437   linear mixed models included fixed effects to account for prenatal NO_2 exposure, time

438   point (birth or age one), biological sex, gestational age, maternal education status,

439   maternal smoking status, and cell type PCs. Additionally, we included SVs to account for

440   technical variation across both the birth and age one time points. We combined cord

441   blood β-values and age one β-values in a row-wise manner before conducting surrogate

442   variable analysis to generate SVs that capture variation due to the same measured and

443   unmeasured variables across both time points and included the first 16 SVs in our linear

444   mixed models (**Supplementary Figure 9**). Linear mixed models also included interaction

445   terms between time and NO_2 exposure, and time and gestational age to reflect that the

446   effect of these variables likely changes over time. Finally, a random intercept term was

447   included to account for repeated measures across participants. We considered the effect

448   of prenatal NO_2 exposure on DNAm to persist if the p-value of interaction term between

449   time and NO_2 failed to reach significance at p<0.05 after Bonferroni adjustment.

450

16

Global DNA methylation analyses

Global DNAm refers to the average methylation level across a genome. Often, DNAm of long interspersed elements (LINEs) is used as a surrogate measure of global DNAm since LINEs are tightly regulated by DNAm and sensitive to environmental exposures[13,74,75]. Here, we examined both average DNAm across all measured CpGs in a sample as well as average DNAm across all probes mapping to L1 LINEs using the Data Integrator function of the University of California Santa Cruz with annotation from Repeat Masker and the ENCODE HAIB Methyl450 track[76–78]. We identified 7302 probes mapping to L1 LINEs in cord blood and postnatal-specific DNAm, respectively. We assessed the correlation between both measures of "global" DNAm in cord blood and postnatal-specific DNAm using linear regressions with covariates to correct for biological and technical variation as outlined in our candidate gene and epigenome-wide analyses. We defined a significant alteration in either measure of "global" DNAm as $p<0.05$. We did not include an effect size cut-off for biological relevance as even a small change in genome-wide or L1 LINE DNAm during development may have a large impact on cell function and health outcome[13,79].

Linkage of NO_2-associated differentially methylated regions to asthma phenotype

We examined whether any of the identified altered regions of prenatal or postnatal-specific DNAm were associated with predictors of child asthma, including age one wheeze and atopy, to provide greater insight into their biological relevance[3]. We used previously defined classifications of atopy and wheeze as primary asthma-related outcomes[3]. Briefly, age one atopy was defined as a positive skin prick test (wheal diameter ≥ 2 mm compared to glycerin control) to any of 10 tested allergens including Alternaria tenuis, cat hair, dog epithelium, Dermatophagoides pteronyssinus, Dermatophagoides farinae, German cockroach, peanut, soybean, egg white, and cow's milk[3]. Age one wheeze was defined as a caregiver reporting child wheezing with or without a cold in the first year of life via questionnaires administered at three, six, and 12 months, or if a CHILD clinician reported wheezing during the age one clinical visit[80]. The relationship between prenatal and postnatal-specific DMRs and age one atopy or wheeze was assessed using MANOVA tests (with the Pillai–Bartlett test statistic) that

482   included covariates to correct for biological and technical variation as specified in our

483   epigenome-wide analyses. Resulting p-values were adjusted for multiple test comparisons

484   using the Bonferroni method with significance defined as $p<0.05$.

485

486   ***Effect of birth month on the relationship between NO_2 exposure and DNA methylation***

487   We hypothesized that the effect of NO_2 exposure on DNAm may vary by birth

488   month as levels of traffic-related air pollution, including NO_2, vary annually with higher

489   levels occurring during colder months[81]. Based on previous findings investigating

490   prenatal air pollution exposure timing and asthma risk, we postulated that participants

491   who were born in a month that exposed them to the highest levels of NO_2 during a critical

492   period of prenatal development (e.g. those born in warmer summer months are

493   hypothesized to be exposed to higher levels of outdoor NO_2 during the first trimester of

494   pregnancy in colder winter months) would exhibit a greater magnitude of DNAm

495   change[82]. To investigate this hypothesis, we conducted a sensitivity analysis to

496   investigate whether the effect of prenatal or year one NO_2 on DNAm varied by birth

497   month at the top 10 CpGs identified in our prenatal and postnatal specific analyses,

498   respectively. Specifically, we re-ran multivariable linear regressions with an interaction

499   term between NO_2 exposure and birth month and used ANOVA likelihood ratio tests to

500   examine whether modelling the effect of prenatal NO2 exposure on DNAm by birth

501   month significantly reduced the residual sum of squares compared to models that did not

502   include birth month. We considered the interaction model fit to be significantly different

503   when $p<0.05$.

Results

A total of 144 participants with DNAm data available at birth and age one remained after preprocessing and normalization. Of these, 128 participants had complete prenatal NO_2 exposure data, with a median exposure of 12.0 parts per billion (ppb; **Table 1**). These 128 participants were included in our analysis of prenatal NO_2 exposure in cord blood DNAm. We used a subset (N=124) of these 128 participants that also had available age one peripheral blood DNAm data to examine the persistence of cord blood DNAm changes at age one. This subset exhibited the same distribution of prenatal NO_2 exposure as the 128 participants included in our analyses of prenatal NO_2 exposure (**Table 1**). In contrast, participants (N=125) that were included in our postnatal analysis were exposed to slightly lower levels of NO_2 air pollution during the first year of life, with a median postnatal NO_2 exposure of 10.2 ppb (**Table 1**). Paired t-tests indicated that the difference in prenatal and year one NO_2 exposure means was statistically significant for participants (N=122) that had estimates available for both time points (p= 1.8e-05; **Supplementary Figure 10**). However, it is important to note that medians of both prenatal and postnatal NO_2 exposure fall within the same air quality standard defined by the Canadian Council of Ministers of the Environment, but above the most current recommendations from the World Health Organization, and are therefore expected to have similar effects on health[83,84]. Participants included in our analyses of cord blood DNAm, persistence of cord blood DNAm changes, and postnatal-specific DNAm changes exhibited similar demographic profiles, including maternal ethnicity and maternal education, to one another and the full CHILD cohort[24]. However, a slightly higher percentage of participants in our analysis of prenatal NO_2 exposure (9.4%; **Table 1**) reported maternal smoking during the prenatal period compared to full CHILD cohort (5.4%)[24].

Table 1. Study population characteristics

	Prenatal DNA methylation changes (N=128)	Persistence of prenatal DNA methylation changes at age one (N=124)	Postnatal-specific DNA methylation changes (N=125)
Prenatal NO_2 (ppb) [25, 50,75 quartiles (min,max)]	7.32 12.01 16.25	6.98 12.01 16.25	

19

	(2.10, 27.71)	(2.10, 27.71)	
Year one NO$_2$ (ppb) [25, 50,75 quartiles (min,max)]			4.69 10.17 14.94 (1.49, 25.93)
Male sex [n(%)]	73 (57%)	72 (58%)	73 (58%)
Gestational age in days [mean±SD]	276 ± 8.0	277 ± 8.0	
Maternal reported white ethnicity [n(%)]	98 (77%)	96 (77%)	
Maternal education years [mean±SD]	16.4 ± 3.2	16.5 ± 3.2	16.4 ± 3.2
Prenatal maternal smoking [n(%)]	12 (9.4%)	12 (9.7%)	
Year one maternal smoking [n(%)]		6 (4.8%)	6 (4.8%)
Cell count estimates [mean ±SD]			
CD4T	0.26 ± 0.087	0.43 ± 0.075	
CD8T	0.15 ± 0.044	0.16 ± 0.049	
Monocytes	0.23 ± 0.010	0.17 ± 0.069	
NK cells	0.11 ± 0.051	0.081 ± 0.037	
B-cells	0.13 ± 0.052	0.19 ± 0.069	
nRBCs	0.17 ± 0.12		

530

531   As NO$_2$ air pollution exposure induces oxidative stress and inflammation, we

532 began our analyses by investigating DNAm across 59 candidate genes encompassing 877

533 CpGs related to these cellular pathways. This *a priori* list included two previously

534 identified genes, *LONP1* and *SLC25A28*, for replication[8]. No CpGs annotated to the

535 candidate genes were significantly associated with prenatal NO$_2$ exposure in cord blood

536 (**Supplementary File 8**), or with year one NO$_2$ exposure in postnatal-specific DNAm

537 (**Supplementary File 9**). Further, we found no associations between cord blood DNAm

538 and prenatal NO$_2$ exposure at cg12283362 (*LONP1*) and cg08973675 (*SLC25A28*), nor

539 did we observe a relationship between year one NO$_2$ exposure and postnatal-specific

540 DNAm at these CpGs (**Figure 1A and B**).

541

542
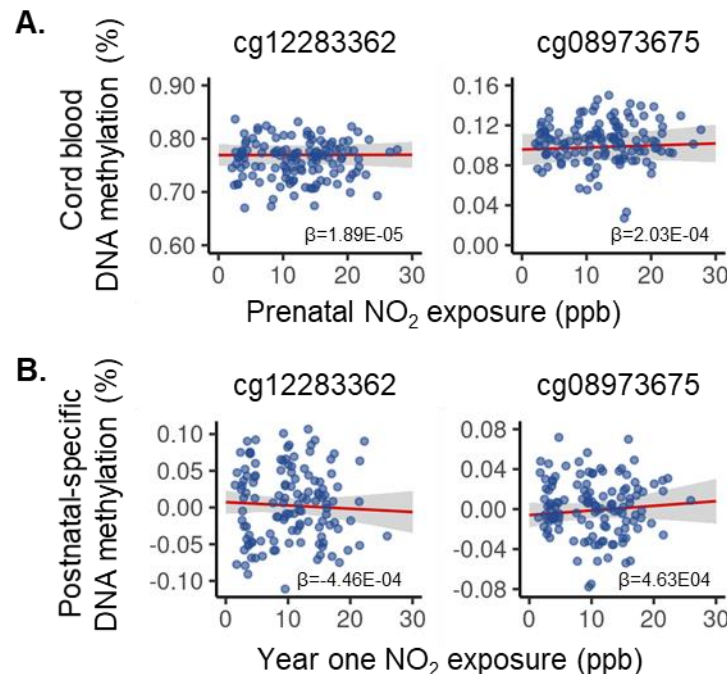
Figure 1. Previously identified cord blood DNA methylation (DNAm) changes in *LONP1* (cg12283362) and *SLC25A28* (cg08973675) do not replicate in CHILD. (**A**) In cord blood, neither DNAm at cg12283362 (adjusted p=1; effect size=4.83e-4; β=1.89e-5) or cg08973675 (adjusted p=1; effect size=5.19E-3; β=2.03E-4) was significantly associated with prenatal NO_2 exposure in CHILD participants (N=128). (**B**) Postnatal-specific measures of DNAm at cg12283362 (adjusted p=1; effect size=-1.10E-2; β=-4.46E-5) and cg08973675 (adjusted p=1; effect size=1.12E-2; β=4.63E-4) were not significantly associated with year one NO_2 exposure in CHILD participants (N=125). Red lines represent the estimated marginal effect of NO_2 exposure on DNAm with a slope equal to coefficient of NO_2 exposure (β). The surrounding shaded area represents the 95% confidence interval of the marginal effect.

554

Epigenome wide analyses can provide insight into novel DNAm changes that may not be captured in a hypothesis-driven candidate gene approach. Therefore, we next investigated epigenome-wide DNAm changes associated with prenatal NO_2 exposure in cord blood (N=128), and with year one NO_2 exposure in postnatal-specific DNAm (N=125). We did not observe any significant DNAm alterations associated with prenatal NO_2 exposure in our epigenome-wide analysis of cord blood at 5% FDR and using a biological cut-off of an absolute effect size >0.025% DNAm (**Supplementary Figure 11; Supplementary File 10**). Similarly, we did not observe any significantly DNAm changes in our epigenome-wide analysis of year one NO_2 exposure and postnatal-specific

564   DNAm at 5% FDR and using an absolute effect size cut-off of >0.025% DNAm

565   (**Supplementary Figure 12; Supplementary File 11**).

566   The small number of available participants and relatively small range of NO_2

567   exposure in our epigenome wide analyses likely limited our power to detect individual

568   cord blood DNAm changes. Analysis of DMRs is often more statistically powerful than

569   epigenome-wide investigations as it can leverage information from adjacent CpGs.

570   Additionally, DMRs are more likely to have a biological effect than individual CpGs.

571   Thus, we examined whether prenatal NO_2 exposure and year one NO_2 exposure were

572   associated with regional DNAm changes in cord blood DNAm and postnatal-specific

573   DNAm measures, respectively. At 5% FDR we identified 24 cord blood DMRs

574   exhibiting an absolute mean effect size of prenatal NO_2 exposure >0.025% DNAm

575   (**Figure 2; Supplementary Table 2, Supplementary Figures 13-36**). These regions

576

577
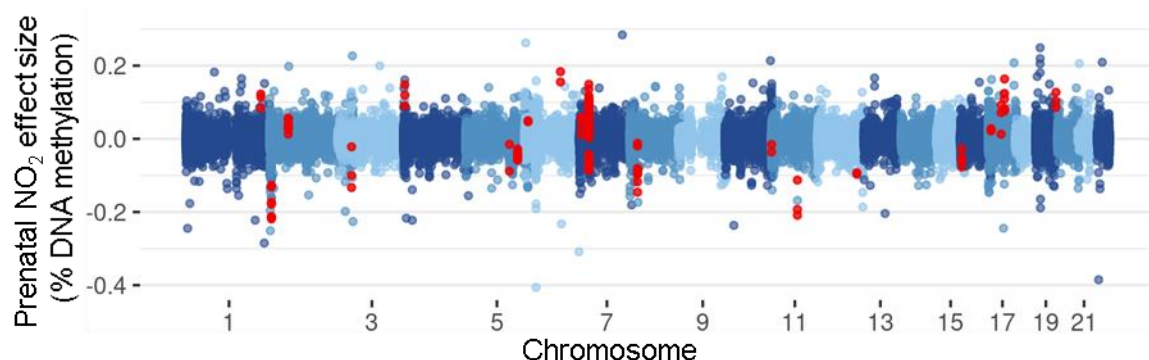
Figure 2. A total of 24 cord blood differentially methylated regions (DMRs) were
significantly associated with prenatal NO_2 exposure. Cord blood DMRs were
identified using *combp* and considered significant if they surpassed 5% FDR and
exhibited a mean absolute effect size of prenatal NO_2 exposure >0.025% DNAm. The
probes (N=151) contained within these DMRs are highlighted in red.

583   were not enriched for gene ontology terms; however, we observed several cord blood

584   DMRs containing CpGs annotated to developmentally relevant genes, such as those in the

585   HOX family (**Supplementary Table 2**). We crossed checked the 151 CpG probes

586   contained within the 24 cord blood DMRs with previously reported mQTLs

587   (**Supplementary File 5**) present at birth in the ARIES cohort to assess gene-by-

588   environment effects[72]. We observed that >50% of CpGs were reported as mQTLs in nine

589   of the 24 cord blood DMRs (**Table 2; Supplementary File 12**). In contrast, we identified

590   only a single postnatal-specific DMR associated with year one NO_2 exposure. This DMR

591   contained 6 CpG probes annotated to *HOXB6* (**Figure 3; Table 3; Supplementary File**

592   **13; Supplementary Figure 37**). We examined whether CpGs within this postnatal-

593

594
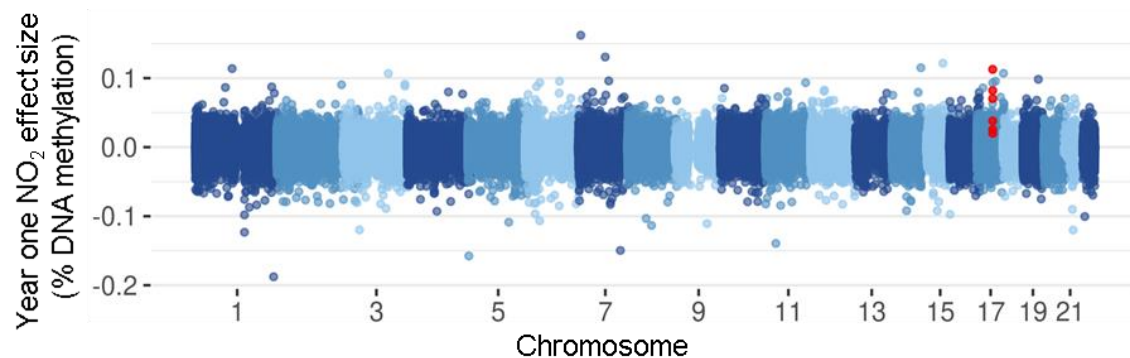
Figure 3. A single postnatal-specific differentially methylated region (DMR) was associated with year one NO2 exposure. Postnatal-specific DMRs were identified using *combp* and considered significant if they surpassed 5% FDR and exhibited a mean absolute effect size of prenatal NO_2 exposure >0.025% DNAm. The probes (N=6) contained within the postnatal-specific DMR we identified are highlighted in red.

602   specific DMR were previously reported as mQTLs at birth or childhood on mQTLdb

603   (**Supplementary Files 6 and 7**). We observed no CpGs within this DMR were

604   previously reported as mQTLs at birth or during childhood using statistical and biological

605   cut-offs of p<1E-14 and an effect size >0.02% to filter CpGs on mQTLdb.

606 **Table 3. Genomic information of the identified postnatal-specific differentially methylated region significantly associated with**

607 **year one NO$_2$ exposure.**

Chr	Start	End	P	FDR	Sidak	Probes	Number of probes that are potential mQTLs	UCSC gene name	UCSC gene group	Mean year one NO$_2$ effect size (% DNAm)	Age one wheeze MANOVA p-value	Age one atopy MANOVA p-value
17	46681111	46681402	3.9E-08	3.9E-08	5.7e-05	6	0	LOC404266; HOXB6	Body;5'UTR	0.058	0.39	0.020

608

609   Persistent DNAm changes are hypothesized to have a greater impact on long term

610 health outcomes than transient DNAm changes[7]. Therefore, we examined whether the

611 cord blood DNAm changes we observed persisted at age one. Specifically, we used linear

612 mixed models to investigate whether the absolute effect size magnitude of prenatal NO_2

613 exposure remained similar between birth and age one at the top 100 cord blood CpGs

614 identified in our epigenome-wide analysis of prenatal NO_2 exposure and at the 151 CpGs

615 contained within significant cord blood DMRs. We investigated individual and regional

616 DNAm changes separately to examine differences in persistence across these

617 perturbations. For discussion purposes we stratified CpGs into those exhibiting lower

618 DNAm at birth and those exhibiting higher DNAm at birth to more readily convey how

619 DNAm is evolving between birth and age one. As linear mixed models were specified in

620 a different manner than multivariable linear models used to investigate epigenome-wide

621 DNAm changes, we first conducted a sensitivity analysis to ensure linear mixed model

622 results were consistent with those from multivariable linear models used in our

623 epigenome-wide analysis and could therefore be used to investigate persistence of cord

624 blood DNAm changes. Specifically, we compared whether effect sizes returned from

625 linear mixed models and multivariable linear models were in the same direction and of

626 similar magnitude. Linear mixed model effect size magnitudes were consistently slightly

627 smaller than those returned from multivariable linear models (**Supplementary File 12**

628 **and 14**). We hypothesize the decrease in effect size magnitude displayed in linear mixed

629 models compared to multivariable linear models is likely due to differences in how we

630 accounted for variation related to estimated cell type proportions and technical effects

631 between the two models. Importantly, we noted that effect sizes always (except for 1

632 CpG) exhibited the same directionality between linear mixed models and multivariable

633 linear models (**Supplementary File 12 and 14**). Based on these findings, we proceeded

634 analyzing the persistence of cord blood DNAm changes using linear mixed models.

635 When examining the persistence of DNAm changes at the top 100 individual CpGs

636 identified in our epigenome-wide analysis of prenatal NO_2 exposure, we found that two

637 CpGs (cg16118259, cg18063278) displaying lower DNAm at birth and two CpGs

638 displaying higher DNAm at birth (cg05686950, cg08033031) exhibited significant

639 decreases in the absolute effect size magnitude of prenatal NO_2 exposure between birth

640 and age one (**Figure 4**; **Supplementary File 14**). In other words, we found that the

641 absolute

642
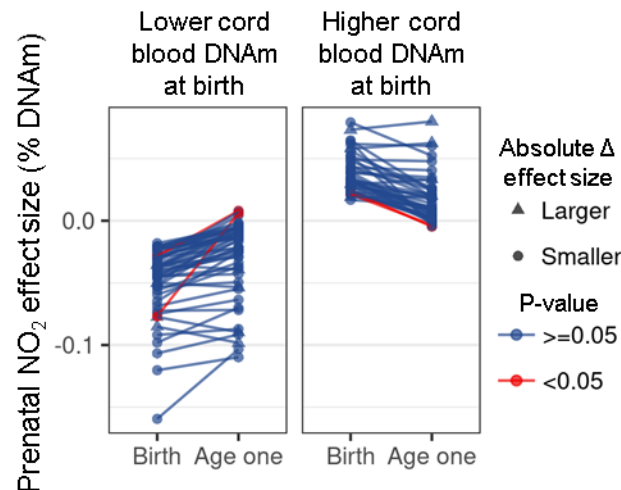
643

644 **Figure 4. The effect size of prenatal NO$_2$ exposure at the top 100 cord blood CpGs**
645 **identified in an epigenome-wide analysis did not significantly differ between birth**
646 **and age one in CHILD participants (N=124).** The CpGs that exhibited lower DNAm at
647 birth are plotted separately from those that displayed higher DNAm at birth to display the
648 change in absolute effect size magnitude of prenatal NO$_2$ exposure between birth and age
649 one more readily. We observed that the absolute effect size magnitude of prenatal NO$_2$
650 exposure decreased between birth and age one at 92 CpGs, with four of these CpGs
651 exhibiting a statistically significant decrease. The remaining 8 CpGs exhibited non-
652 significant increases in the absolute effect size magnitude of prenatal NO$_2$ exposure
653 between birth and age one. The CpGs exhibiting a smaller absolute effect size magnitude
654 of prenatal NO$_2$ exposure at age one are indicated by circles, while those displaying a
655 larger absolute effect size magnitude of prenatal NO$_2$ exposure at age one are indicated
656 by triangles. The CpGs that displayed a significant (Bonferroni p<0.05) change in the
657 absolute effect size of prenatal NO$_2$ exposure between birth and age one are highlighted
658 in red.

659

660 effect size magnitude of prenatal NO$_2$ exposure at these four CpGs was closer to zero at

661 age one than at birth. All four of these CpGs exhibiting an absolute effect size magnitude

662 of prenatal NO$_2$ exposure >0.025% DNAm at birth, but not at age one, suggesting the

663 effect of prenatal NO$_2$ exposure did not persist at these loci (**Figure 4**; **Supplementary**

664 **File 14**). The remaining 96 CpGs from our epigenome-wide analysis exhibited

665 nonsignificant changes in the absolute effect size magnitude of prenatal NO$_2$ exposure

666 between birth and age one. Of these 96 CpGs, 54 exhibited lower DNAm at birth and 42

26

667 displayed higher DNAm at birth. Almost all (N=50) CpGs exhibiting lower DNAm

668 displayed a decrease in the absolute effect size of prenatal NO_2 exposure between birth

669 and age one, with the remaining minority (N=4) displaying an increase absolute effect

670 size magnitude between birth and age one. Similarly, most (N=38) CpGs exhibiting

671 higher DNAm at birth also displayed a decrease in the absolute effect size magnitude of

672 prenatal NO_2 exposure between birth and age one. We noted that 43 of the 54 CpGs

673 exhibiting lower DNAm at birth and 32 of the 42 CpGs exhibiting higher DNAm at birth

674 met the biological cut-off of an absolute effect size magnitude of prenatal NO_2 exposure

675 >0.025% DNAm at the birth time point. Only 21 CpGs displaying lower DNAm at birth

676 and 11 CpGs displaying higher DNAm at birth continued to meet our biological cut-off

677 of >2.5% DNAm at age one, suggesting the effect of prenatal NO_2 exposure persisted at

678 these loci (**Figure 4**; **Supplementary File 14**). Of the 151 CpGs contained within cord

679 blood DMRs, 55 exhibited lower DNAm at birth while the remaining 96 displayed higher

680 DNAm at birth (**Figure 5, Supplementary File 12**). A total of 38 CpGs exhibiting lower

681

682
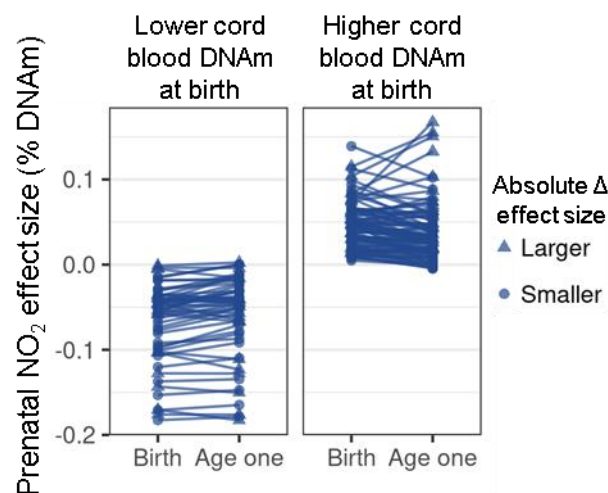
683 **Figure 5. The effect size of prenatal NO_2 exposure on CpG (N=151) DNAm within**
684 **cord blood differentially methylated regions does not significantly differ between**
685 **birth and age one in CHILD participants (N=124).** The CpGs that exhibited lower
686 DNAm at birth are plotted separately from those that displayed higher DNAm at birth to
687 display the change in absolute effect size magnitude of prenatal NO_2 exposure between
688 birth and age one more readily. The 24 identified cord blood DMRs contained a total of
689 151 CpGs. Of these, 109 CpGs exhibited non-significant decreases in the absolute effect
690 size magnitude of prenatal NO_2 exposure between birth and age one. The remaining 42
691 CpGs exhibited non-significant increases in the absolute effect size magnitude of prenatal

27

692   NO$_2$ exposure at age one. The CpGs exhibiting a smaller absolute effect size magnitude
693   of prenatal NO$_2$ exposure at age one compared to birth are indicated by circles, and those
694   displaying a larger absolute effect size of prenatal NO$_2$ exposure at age one compared to
695   birth are indicated by triangles.
696

697   DNAm at birth exhibited a decrease in the absolute effect size magnitude of prenatal NO$_2$

698   exposure between birth and age one. Similarly, the absolute effect size magnitude of

699   prenatal NO$_2$ exposure decreased between birth and age one at 72 of the 96 DMR CpGs

700   exhibiting higher DNAm at birth. These changes failed to reach statistical significance at

701   all 151 DMR CpGs. We noted that 48 of the 55 DMR CpGs exhibiting lower DNAm at

702   birth and 71 of 96 DMR CpGs displaying higher DNAm at birth met the biological cut-

703   off of an absolute effect size magnitude of prenatal NO$_2$ exposure >0.025% DNAm at the

704   birth time point. In contrast to 100 CpGs selected from our epigenome-wide analysis,

705   almost all DMR CpGs meeting biological cut-offs at the birth time continued to meet

706   biological cut-offs at age one. Specifically, 41 of the 48 DMR CpGs exhibiting lower

707   DNAm at birth and 56 of the 71 DMR CpGs exhibiting higher DNAm at birth displayed

708   an absolute effect size magnitude of prenatal NO$_2$ exposure >0.025% DNAm at birth and

709   age one, suggesting the effect of prenatal NO$_2$ exposure persists at these loci until at least

710   age one. Investigating effect size changes across individual CpGs located within DMRs

711   does not clarify which DMRs are persistently or transiently affected by prenatal NO$_2$

712   exposure. Therefore, we also examined which DMRs continued to display a mean

713   absolute effect size magnitude of prenatal NO$_2$ exposure >2.5% DNAm at age one. We

714   noted that 23 of 24 cord blood DMRs displayed a mean absolute effect size magnitude of

715   prenatal NO$_2$ exposure >0.025% at birth based on linear mixed models (**Supplementary**

716   **Table 2**). At age one, 19 DMRs continued to exhibit mean absolute effect size magnitude

717   of prenatal NO$_2$ exposure >2.5% DNAm (**Supplementary Table 2**). The DMRs

718   exhibiting persistent effects of prenatal NO$_2$ exposure included those annotated to

719   developmentally relevant *HOX* genes (**Supplementary Table 2**).

720   Air pollution is hypothesized to reduce global DNAm by potentiating DNA

721   demethylation and altering the availability of enzymes and methyl donors[48]. LINE

722   DNAm can be used as a surrogate measure for whole genome DNAm and previous

723   studies have shown LINE DNAm is affected by air pollution exposure[75]. Therefore, we

724   investigated whether CHILD participants exhibited average genome-wide or LINE-

725   specific DNAm alterations at birth or age one related to prenatal NO_2 exposure or

726   postnatal NO_2 exposure, respectively. Genome-wide DNAm was measured by averaging

727   DNAm across all CpGs for each individual. We obtained LINE-specific DNAm measures

728   by averaging DNAm across all CpGs annotated to L1 LINEs on the Illumina

729   HumanMethylation 450k array. We observed no significant associations between prenatal

730   NO_2 exposure and genome-wide DNAm alterations at birth (**Figure 6A**), nor did we find

731   any correlation between year one NO_2 exposure and genome-wide postnatal-specific

732   DNAm changes (**Figure 6B**). Further, neither prenatal NO_2 exposure (**Figure 6A**) or year

733

734
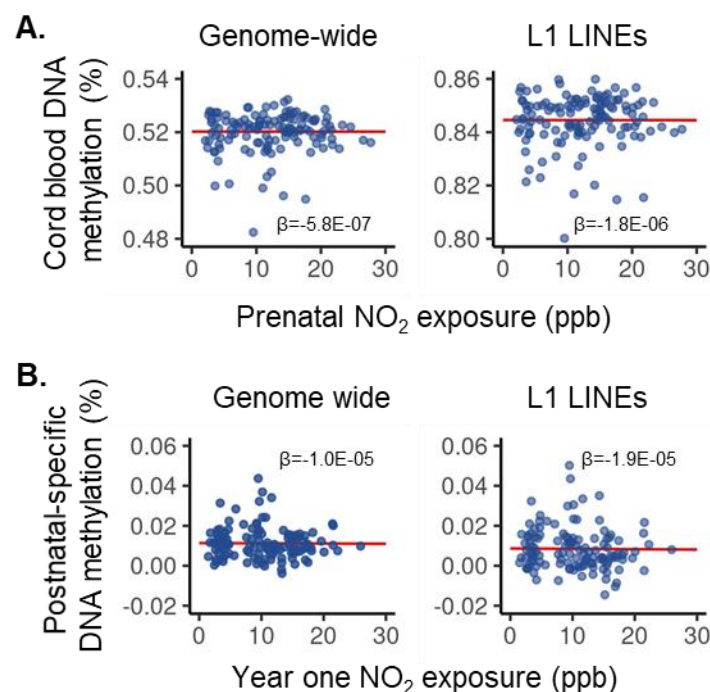
Figure 6. Prenatal and year one NO_2 exposures are not associated with genome average or L1 LINE-specific measures of global DNA methylation in CHILD participants. (**A**) Associations between global measures of cord blood DNAm and prenatal NO_2 exposure. Neither genome-wide DNAm (p=0.94; effect size=-1.5E-05; β=-5.8E-07) or LINE-specific DNAm (p=0.89; effect size=-4.7E-05; β=-1.8E-06) was significantly associated with prenatal NO2 exposure. (**B**) Associations between global measures of postnatal-specific DNAm and year one NO_2 exposure. Neither postnatal-specific changes in genome-wide DNAm (p=0.25; effect size=-4.6E-4; β=-1.0e-05) or LINE-specific DNAm (p=0.33; effect size=-2.5E-4; β=-1.0e-05) were significantly associated with year one NO_2 exposure. Red lines represent the estimated marginal effect of NO_2 exposure on DNAm with a slope equal to coefficient of NO_2 exposure (β). The surrounding shaded area represents the 95% confidence interval of the marginal effect.

747
748   one NO$_2$ exposure (**Figure 6B**) was associated with significant changes to prenatal or

749   postnatal-specific L1 LINE DNAm, respectively.

750   Currently, researchers believe the biological embedding of early life exposures in

751   DNAm patterns contributes to altered health outcomes. Using MANOVA tests we

752   examined the relationship between DNAm across cord blood and postnatal-specific

753   DMRs identified in this study with markers of childhood asthma, including age one atopy

754   and wheeze. In cord blood, DNAm across Chr17:7486551-7486616 (adjusted p = 0.03)

755   and Chr5:126409007-126409062 (adjusted p = 0.002) significantly differed between

756   CHILD participants with (N=33) and without wheeze (N=86; **Figure 7A**). We also

757   identified that DNAm across the Chr19:58879022-58879060 in cord blood

758

759
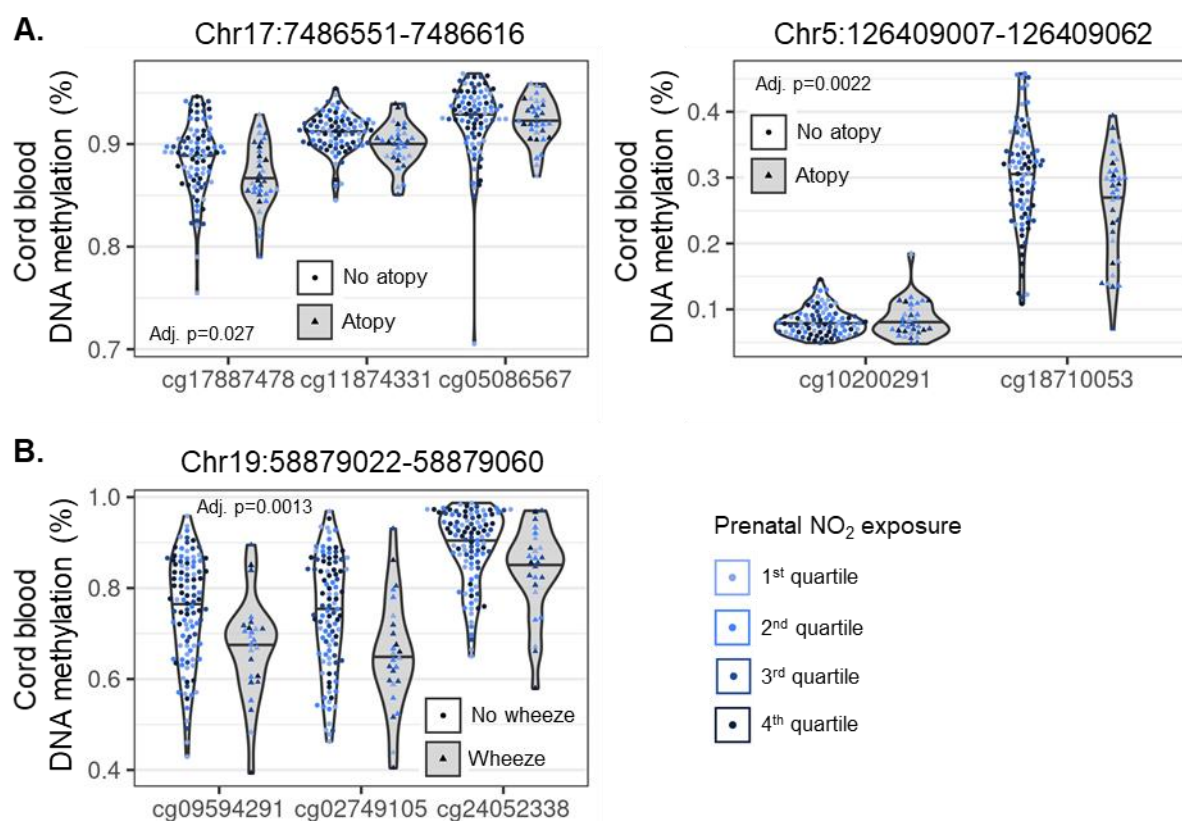
760 **Figure 7. Three cord blood differentially methylated regions (DMRs) are associated**
761 **with age one wheeze or age one atopy in CHILD participants.** (**A**) Two cord blood
762 DMRs are associated with age on wheeze in CHILD participants (N=119). MANOVA
763 tests suggested that DNA methylation (DNAm) across Chr17:7486551-7486616
764 (adjusted p=0.03) and Chr5:126409007-126409062 (adjusted p=0.002) was significantly

765   different between participants with (N=33) and without (N=86) wheeze at age one. (**B**)
766   One cord blood DMR was associated with age on atopy in CHILD participants (N=128).
767   MANOVA tests also suggests that DNAm across Chr19:58879022-58879060 (adjusted
768   p=0.02) was significantly different between CHILD participants with and without atopy
769   at age one. Shaded violin plots with triangular points represent DNAm measures for
770   participants with wheeze (panel A) or atopy (panel B). Unshaded plots with circular
771   points represent DNAm measures for participants without wheeze (panel A) or atopy
772   (panel B). Points are coloured based on quartiles of prenatal NO_2 exposure as specified in
773   Supplementary Table 2 to display the relationship between prenatal NO_2 exposure and
774   altered DNAm most clearly.
775

776   was significantly (adjusted p=0.001) different between CHILD participants with (N=25)

777   and without (N=103) atopy at age one (**Figure 7B**). We next examined whether the

778   postnatal-specific DMR was associated with age one health outcomes. We observed that

779   altered postnatal-specific DNAm across Chr17:76681111-46681402 was significantly

780   (p=0.02) different between participants with (N=23) and without (N=102) age one atopy

781   (**Figure 8**) but did not different between participants with (N=33) and without (N=87)

782   wheeze at age one.

783

784
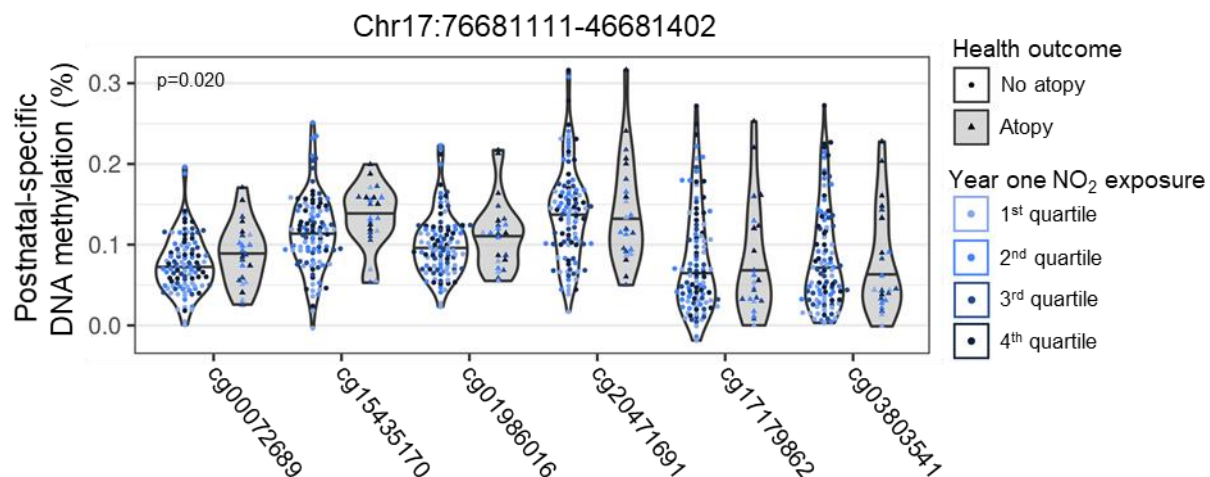
785   **Figure 8. Altered postnatal-specific DNA methylation (DNAm) across**
786   **Chr17:76681111-46681402 is associated with one age atopy in CHILD participants**
787   **(N=125).** MANOVA tests suggested that altered postnatal-specific DNAm across
788   Chr17:76681111-46681402 significantly (p=0.020) differed between CHILD participants
789   with (N=23) and without atopy (N=102). Shaded violin plots with triangular points
790   represent DNAm measures for participants with atopy. Unshaded plots with circular
791   points represent DNAm measures for participants without atopy. Points are coloured
792   based on quartiles of year one NO_2 exposure as specified in Supplementary Table 2 to
793   display the relationship between prenatal NO_2 exposure and altered DNAm most clearly.

794   Levels of ambient NO_2 vary annually in Canada, with the highest levels typically

795 occurring during colder periods[81]. We hypothesized that children exposed to higher levels

796 of air pollution during critical developmental periods may experience larger magnitudes

797 of DNAm alterations. Therefore, we conducted a sensitivity analysis to examine whether

798 effect of NO_2 exposure on DNAm varies by birth month at the top 10 CpGs identified in

799 our prenatal and postnatal-specific epigenome-wide analyses. We found that modelling

800 the effect of prenatal NO_2 exposure on cord blood DNAm by birth month did not

801 significantly improve model fit based on ANOVA likelihood ratio tests (**Supplementary**

802 **Figure 38**). Similarly, ANOVA likelihood tests suggested that modelling the effect of

803 year one NO_2 exposure on postnatal-specific DNAm did not improve model fit

804 (**Supplementary Figure 39**), suggesting the effects of prenatal and year one NO_2

805 exposure on DNAm do not vary by birth month.

Discussion

In the present study, we investigated the persistence of cord blood DNAm changes associated with prenatal NO_2 exposure to identify epigenetic alterations with the greatest potential to influence child health outcome. Additionally, longitudinal measures of DNAm facilitated the investigation of postnatal-specific DNAm changes associated with NO_2 exposure in the first year of life. Our results show that prenatal NO_2 exposure is associated with regional DNAm changes, and that the effect of prenatal NO_2 exposure across CpGs within these regions remains similar at age one. These findings suggest the effects of prenatal NO_2 exposure on DNAm persist through early life and are therefore more likely to affect child respiratory health outcomes. This notion is consistent with our observations that several regional cord blood DMRs were also associated with wheeze or atopy at age one. Additionally, our observations that postnatal-specific DNAm alterations are associated with atopy contribute to our understanding that the postnatal period remains a sensitive window for the biological embedding of air pollution exposure. Together, these findings provide greater insight into the putative molecular pathways underlying the association between prenatal and early life NO_2 exposure and childhood atopic disease.

We observed several cord blood DMRs localized to developmentally relevant genes. Additionally, linear mixed models showed that the effect of prenatal NO_2 exposure on CpG DNAm contained within these DMRs did not significantly vary at age one, and that the majority of DMR CpGs meeting our biological cut-off of a mean absolute effect size of prenatal NO_2 exposure at birth (N=109 CpGs) continued to meet this cut-off at age one (N=98 CpGs). In contrast, only 32 of 75 individual CpGs from our epigenome-wide analysis met the biologically cut-off for prenatal NO2 exposure effect size at birth and age one. Together, this implies the effect of prenatal NO2 exposure is more persistent across altered regions than individual CpGs. We observed persistently altered regions of DNAm in *HOXA4* and *HOXA5*, which are strongly implicated in lung development and also have roles in immune cell differentiation[85,86]. Regional DNAm changes associated with air pollution exposure have previously been reported in *HOXA4* and *HOXA5*. For example, Gruzieva *et al.* (2019) reported a *HOXA4* DMR in cord blood associated with prenatal PM_{10} exposure[87] and Chi *et al.*(2016) reported a *HOX5A* DMR in adult peripheral blood

837   associated with $PM_{2.5}$[76,87]. Further, altered *HOX5A* DNAm reported by Chi *et al.*(2016)

838   was associated with gene expression of *HOX5A*, as well as *HOXA9* and *HOXA10* gene

839   expression, demonstrating that altered *HOX5A* DNAm has functional consequences[76].

840   Both *HOXA4* and *HOXA5* DMRs identified in this study overlap with those previously

841   reported by Gruzieva *et al.* (2019) and Chi *et al.* (2016)[76,87]. It is important to note that in

842   this study we examined the effect of prenatal NO_2 exposure while the aforementioned

843   studies investigated particulate matter exposures. NO_2 exposure is considered a putative

844   causative agent of DNAm change as well as a surrogate measure of traffic-related air

845   pollution, which also contains particulate matter[28]. While it is not possible to disentangle

846   whether the changes we observed are due to the direct of effect of prenatal NO_2 exposure

847   or potentially other air pollutants NO_2 may be correlated with, the similarities between

848   DMRs identified in this study and those observed previously strongly implicate air

849   pollution exposure in altered DNAm of HOX genes. Given the roles HOX genes play in

850   prenatal and early life lung development, persistently altered HOX gene DNAm observed

851   in this study may represent a molecular mechanism linking prenatal air pollution exposure

852   to changes in child lung structure observed in asthma[88].

853   Most studies investigating the impact of early-life NO_2 exposure have focused on

854   the prenatal period based on the notion that disruption of DNAm patterns during

855   epigenetic reprogramming, especially those related to immune function and lung

856   development, will have the largest impact on cell function and long term respiratory

857   health[89,90]. The early postnatal period, especially the first year of life, is often overlooked

858   despite the continued development of the lungs, immune system, and DNAm patterns

859   during this time[14,15,91]. In this study we observed that year one NO_2 exposure was

860   associated with a postnatal-specific DMR annotated to *HOXB6*. Further, we identified

861   that altered DNAm across this region was associated with age one atopy, suggesting that

862   this DNAm change may underlying altered child health outcomes. This is consistent with

863   results from a large cross-sectional epidemiological showing prenatal and postnatal

864   passive cigarette smoke exposures independently affect the risk of negative respiratory

865   health outcomes in childhood[17]. Together, these findings support the first year of life as

866   also a critical window for the biological embedding of air pollution exposure. Of note, we

867   also observed a cord blood DMR localizing to *HOXB6*, although at an upstream location,

34

868   suggesting *HOX6B* DNAm appears to be sensitive to air pollution exposure throughout

869   early-life, and that prenatal and postnatal-specific changes in DNAm may play different

870   roles in the regulation of *HOXB6* expression[92]. The *HOXB6* gene is one of the most

871   highly expressed HOX family members in human fetal and adult lung, playing an

872   important role in early lung development, and also in hematopoiesis[88]. Thus, prenatal and

873   postnatal *HOXB6* DNAm changes may work together to affect lung development across

874   early-life, leading to long-term changes in lung structure and/or function that contribute

875   to the pathology of childhood asthma[93]. Alternatively, changes to *HOXB6* DNAm in

876   hematopoietic stem cells could potentiate the expansion of myeloid precursor cells and

877   their progenitors, leading to changes in cell abundances (such as increased mast cells)

878   observed in asthma[94]. In this study, we investigated DNAm changes under the

879   hypothesis of cellular reprogramming[7]. Additional investigation is needed to determine if

880   *HOXB6* DNAm is affected in a similar manner within other tissues and how the DNAm

881   alterations we observed affect cell function and differentiation.

882   In CHILD study participants, concurrent atopy and wheeze at age one is a

883   stronger predictor of age three asthma than either condition occurring alone[3]. We were

884   limited in our ability to examine associations between DNAm across significant DMRs

885   and atopy and wheeze at age one together due to a small number (N=9 in our analysis of

886   prenatal NO_2 exposure) of participants in our analyses displaying both conditions at age

887   one. Instead, we used MANOVA tests to separately examine the association between

888   DNAm across significant cord blood and postnatal-specific DMRs and age one wheeze or

889   age one atopy. We observed two cord blood DMRs annotated to *MPDU1* and *C5orf63*

890   (*FLJ44606*) that were associated with age one wheeze. The association between *MPDU1*

891   DNAm and age one wheeze represents a novel observation. *MPDU1* codes for an

892   endoplasm reticulum protein that utilizes mannose-P-dichol in the synthesis of lipid-

893   linked oligosaccharides and glycosylphosphatidylinositols, which are used downstream in

894   post-translational modification of proteins[95]. It remains possible that alterations in protein

895   post-translational modifications arising from changes in *MPDU1* DNAm may contribute

896   to childhood wheeze and asthma. For example, glycosylation of the Fc region of IgG1

897   can impact antibody function and decreases in IgG1 glycosylation are associated with

898   immune activation and have been observed in asthmatic compared to non-asthmatic

899   children, as well as in murine models of allergic asthma[96–98]. Two previous studies have

900   reported DMRs in *C5orf63* associated with diesel exhaust inhalation in bronchial

901   epithelial cells and with smoking in peripheral blood samples obtained from a Korean

902   cohort[99,100]. The *C5orf63* DMR observed in this study overlaps with both previously

903   identified *C5orf63* DMRs, implying this DMR represents a robust DNAm change across

904   development stages, tissues, and human populations in response to inhaled pollutant

905   exposure. One of the mechanisms by which inhaled air pollutants induce DNAm changes

906   is through oxidative stress. Given that regional *C5orf63* DNAm changes are robust across

907   studies and that *C5orf63* encodes a glutaredoxin-like protein, these findings support a

908   role for *C5orf63* in oxidative stress following inhaled air pollutant exposure that is

909   regulated by DNAm[101]. To our knowledge, this is the first study to identify an association

910   between *C5orf63* DNAm and childhood wheeze. Increased oxidative stress contributes to

911   the underlying pathology of asthma[102]. The observed changes in *C5orf63* DNAm may

912   represent an underlying molecular mechanism of this pathology. In cord blood, we also

913   observed that altered DNAm in *ZNF837* was associated with age one atopy. The *ZNF837*

914   protein is predicted to help regulate nuclear transcription by RNA polymerase II. There

915   exists a previous report of transcription factor binding sites for AHR and ARNT in the

916   promoter of *ZNF387*[103,104]. These proteins represent transcription factors which enter the

917   nucleus and modulate gene expression following inhaled air pollution exposure,

918   suggesting that the *ZNF837* product plays a role in this pathway[105]. Though the genes

919   which the *ZNF837* protein regulates remain to be elucidated, we envision a role for

920   *ZNF837* in regulating expression and DNA binding of atopy and asthma-relevant gene

921   products and transcription factors, respectively[106]. Replication of these results followed

922   by further experiments is needed to confirm the exact role *ZNF837* plays in gene

923   expression following air pollution exposure.

924   This subset of CHILD study participants was over-sampled for atopic conditions

925   including allergenic sensitization to inhaled or food allergens. Atopy often occurs

926   together with asthma and atopic dermatitis in what is known as the "atopic triad"[107].

927   There is an underlying genetic component to the atopic triad; however, early life

928   environmental exposures such as air pollution exposure increase disease

929   predisposition[107]. It is possible that genetic variants that increase atopy risk may also

930 increase the susceptibility of nearby CpGs to DNAm changes following air pollution

931 exposure, known as methylation trait quantitative loci (mQTLs)[71]. Our small sample size

932 limited our ability to investigate the role of mQTLs in the DNAm changes we observed.

933 When we cross checked cord blood DMR CpGs identified in this study with an online

934 database constructed using data from the European ARIES cohort[72], we found that 55 of

935 these cord blood DMR CpGs represented potential mQTLs with effect sizes ranging from

936 0.02 to 0.27. Interestingly, no CpGs contained within the cord blood DMRs associated

937 with wheeze or atopy were identified as potential mQTLs, suggesting that DNAm within

938 these regions plays a primary role in determining child respiratory outcomes.

939 Investigation of mQTLs in future studies will provide insight to which populations may

940 be most susceptible to DNAm changes and adverse health outcomes following early-life

941 air pollution exposure.

942 The longitudinal design of the CHILD study is a major advantage in investigating

943 persistent DNAm changes as we were able follow DNAm within the same group of

944 individuals through the first year of life. Additionally, this study design allowed us to

945 isolate and examine postnatal-specific DNAm alterations associated with NO_2 exposure.

946 To our knowledge, this is the first time postnatal-specific DNAm changes have been

947 investigated in the context of early-life air pollution exposure. The biological context of

948 our findings in relation to childhood atopic disease is limited by the sample type (blood

949 mononuclear cells) used. White blood cells are of interest as changes to immune cell

950 function are thought to play a role in the underlying pathology of child atopic disease[108].

951 While some DNAm patterns in blood appear to be surrogate markers for less accessible

952 tissues, such as brain[109], the correlation between DNAm in blood and DNAm in tissues

953 related to respiratory health such as lung remains unclear. Thus, it remains to be

954 determined whether the DNAm changes we identified also occur in respiratory tissue

955 following early life NO_2 exposure. This investigation was also limited by the relatively

956 small sample size of CHILD participants and a relatively low variation in NO_2 compared

957 to previous studies[8], likely decreasing our ability to detect additional DNAm alterations.

958 Finally, the CHILD cohort is mainly comprised of white participants with higher

959 socioeconomic status. Replication of our results in larger cohorts with more diverse

960 populations is required before they can be generalized.

Summary

961

962   Together, these findings suggest prenatal exposure to NO_2 air pollution alters

963 regions of DNAm and that DNAm changes across most of these regions persist into early

964 childhood. For the first time we show that the first year of life remains a sensitive period

965 to NO_2 exposure, resulting in postnatal-specific DNAm alterations. Our observations

966 provide further support these changes are occurring in cellular pathways related to early

967 lung development. This knowledge contributes to our understanding of the molecular

968 pathways that may be altered by early life air pollution exposure and affect child health

969 outcome.

970

971

977

978

Conflicts of interest

980 The authors of this study have no conflicts of interest to report.

981

982

Code availability

984 Code used to conduct the analyses reported in this study can be accessed at

985 https://github.com/mjoneslab/CHILD_prenatal_no2_analysis.

986

987

988

989

990

38

References

1. Achakulwisut, P., Brauer, M., Hystad, P. & Anenberg, S. C. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO2 pollution: estimates from global datasets. *Lancet Planet. Health* **3**, e166–e178 (2019).

2. Bacharier, L. B. *et al.* Diagnosis and treatment of asthma in childhood: a PRACTALL consensus report. *Allergy* **63**, 5–34 (2008).

3. Arrieta, M.-C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* (2015) doi:10.1126/scitranslmed.aab2271.

4. Davidson, W. F. *et al.* Report from the National Institute of Allergy and Infectious Diseases workshop on "Atopic dermatitis and the atopic march: Mechanisms and interventions". *J. Allergy Clin. Immunol.* **143**, 894–913 (2019).

5. Isaevska, E. *et al.* Exposure to ambient air pollution in the first 1000 days of life and alterations in the DNA methylome and telomere length in children: A systematic review. *Environ. Res.* **193**, 110504 (2021).

6. Hertzman, C. Putting the concept of biological embedding in historical perspective. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17160–17167 (2012).

7. Lappalainen, T. & Greally, J. M. Associating cellular epigenetic models with human phenotypes. *Nat. Rev. Genet.* **18**, 441–451 (2017).

8. Gruzieva, O. *et al.* Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. *Environ. Health Perspect.* **125**, 104–110 (2017).

9. Ladd-Acosta, C. *et al.* Epigenetic marks of prenatal air pollution exposure found in multiple tissues relevant for child health. *Environ. Int.* **126**, 363–376 (2019).

1014   10. Meyer, M., Gaudieri, S., Rhodes, D. a. & Trowsdale, J. Cluster of TRIM genes in the

1015   human MHC class I region sharing the B30.2 domain. *Tissue Antigens* **61**, 63–71

1016   (2003).

1017   11. Park, J. *et al.* Prenatal Exposure to Traffic-Related Air Pollution and the DNA

1018   Methylation in Cord Blood Cells: MOCEH Study. *Int. J. Environ. Res. Public.*

1019   *Health* **19**, 3292 (2022).

1020   12. Lakhani, S. A. *et al.* Caspases 3 and 7: Key Mediators of Mitochondrial Events of

1021   Apoptosis. *Science* **311**, 847–851 (2006).

1022   13. Breton, C. V. *et al.* Prenatal Air Pollution Exposures, DNA Methyl Transferase

1023   Genotypes, and Associations with Newborn LINE1 and Alu Methylation and

1024   Childhood Blood Pressure and Carotid Intima-Media Thickness in the Children's

1025   Health Study. *Environ. Health Perspect.* **124**, 1905–1912 (2016).

1026   14. Wikenius, E., Moe, V., Smith, L., Heiervang, E. R. & Berglund, A. DNA methylation

1027   changes in infants between 6 and 52 weeks. *Sci. Rep.* **9**, 1–12 (2019).

1028   15. Schittny, J. C. Development of the lung. *Cell Tissue Res.* **367**, 427–444 (2017).

1029   16. Georgountzou, A. & Papadopoulos, N. G. Postnatal Innate Immune Development:

1030   From Birth to Adulthood. *Front. Immunol.* **8**, (2017).

1031   17. Pattenden, S. *et al.* Parental smoking and children's respiratory health: independent

1032   effects of prenatal and postnatal exposure. *Tob. Control* **15**, 294–301 (2006).

1033   18. Lavigne, É. *et al.* Effect modification of perinatal exposure to air pollution and

1034   childhood asthma incidence. *Eur. Respir. J.* **51**, (2018).

1035   19. Pennington, A. F. *et al.* Exposure to Mobile Source Air Pollution in Early-life and

1036   Childhood Asthma Incidence: The Kaiser Air Pollution and Pediatric Asthma Study.

1037   *Epidemiology* **29**, 22–30 (2018).

1038   20. Ranci, ère F., Bougas, N., Viola, M. & Momas, I. Early Exposure to Traffic-Related

1039   Air Pollution, Respiratory Symptoms at 4 Years of Age, and Potential Effect

1040   Modification by Parental Allergy, Stressful Family Events, and Sex: A Prospective

1041   Follow-up Study of the PARIS Birth Cohort. *Environ. Health Perspect.* **125**, 737–

1042   745 (2017).

1043   21. Moraes, T. J. *et al.* The Canadian Healthy Infant Longitudinal Development Birth

1044   Cohort Study: Biological Samples and Biobanking. *Paediatr. Perinat. Epidemiol.* **29**,

1045   84–92 (2015).

1046   22. Takaro, T. K. *et al.* The Canadian Healthy Infant Longitudinal Development

1047   (CHILD) birth cohort study: assessment of environmental exposures. *J. Expo. Sci.*

1048   *Environ. Epidemiol.* **25**, 580–592 (2015).

1049   23. Sbihi, H. *et al.* Prenatal exposure to traffic-related air pollution, the gestational

1050   epigenetic clock, and risk of early-life allergic sensitization. *J. Allergy Clin. Immunol.*

1051   **0**, (2019).

1052   24. Subbarao, P. *et al.* The Canadian Healthy Infant Longitudinal Development (CHILD)

1053   Study: examining developmental origins of allergy and asthma. *Thorax* **70**, 998–1000

1054   (2015).

1055   25. Henderson, S. B., Beckerman, B., Jerrett, M. & Brauer, M. Application of land use

1056   regression to estimate long-term concentrations of traffic-related nitrogen oxides and

1057   fine particulate matter. *Environ. Sci. Technol.* **41**, 2422–2428 (2007).

1058   26. Allen, R. W., Amram, O., Wheeler, A. J. & Brauer, M. The transferability of NO and

1059   NO2 land use regression models between cities and pollutants. *Atmos. Environ.* **45**,

1060   369–378 (2011).

1061   27. Jerrett, M. *et al.* Modeling the intraurban variability of ambient traffic pollution in

1062   Toronto, Canada. *J. Toxicol. Environ. Health A* **70**, 200–212 (2007).

1063   28. Sbihi Hind *et al.* Perinatal Exposure to Traffic-Related Air Pollution and Atopy at 1

1064   Year of Age in a Multi-Center Canadian Birth Cohort Study. *Environ. Health*

1065   *Perspect.* **123**, 902–908 (2015).

1066   29. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the

1067   analysis of Infinium DNA methylation microarrays. *Bioinforma. Oxf. Engl.* **30**,

1068   1363–1369 (2014).

1069   30. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation

1070   array data. *BMC Genomics* **14**, 293 (2013).

1071   31. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K.

1072   D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays.

1073   *Nucleic Acids Res.* **41**, e90 (2013).

1074   32. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the

1075   Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209

1076   (2013).

1077   33. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying

1078   methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

1079   34. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture

1080   distribution. *BMC Bioinformatics* **13**, 86 (2012).

1081   35. Salas, L., Gervin, K. & Jones, M. *FlowSorted.CordBloodCombined.450k: Illumina*

1082   *450k/EPIC data on FACS and MACS umbilical blood cells.* (2019).

1083   36. Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-

1084   blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray.

1085   *Genome Biol.* **19**, 64 (2018).

1086   37. Gervin, K. *et al.* Systematic evaluation and validation of reference and library

1087   selection methods for deconvolution of cord blood DNA methylation data. *Clin.*

1088   *Epigenetics* **11**, 125 (2019).

1089   38. Filzmoser, P., Hron, K. & Templ, M. *Applied Compositional Data Analysis: With*

1090   *Worked Examples in R.* (Springer International Publishing, 2018). doi:10.1007/978-3-

1091   319-96422-5.

1092   39. Perrone, S. *et al.* Nucleated red blood cell count in term and preterm newborns:

1093   reference values at birth. *Arch. Dis. Child. - Fetal Neonatal Ed.* **90**, F174–F175

1094   (2005).

1095   40. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and

1096   richer datasets. *GigaScience* **4**, 7 (2015).

1097   41. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for

1098   genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).

1099   42. O'Connell, J. *et al.* A General Approach for Haplotype Phasing across the Full

1100   Spectrum of Relatedness. *PLOS Genet.* **10**, e1004234 (2014).

1101   43. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype

1102   imputation. *Bioinformatics* **31**, 782–784 (2015).

1103   44. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.*

1104   **48**, 1284–1287 (2016).

1105   45. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74

1106   (2015).

1107   46. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-

1108   wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

1109   47. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies.

1110   *Nat. Protoc.* **5**, 1564–1573 (2010).

1111   48. Rider, C. F. & Carlsten, C. Air pollution and DNA methylation: effects of exposure

1112   in humans. *Clin. Epigenetics* **11**, 131 (2019).

1113   49. Lu, T. & Stark, G. R. NF-κB: regulation by methylation. *Cancer Res.* **75**, 3692–3695

1114   (2015).

1115   50. Bhatt, D. & Ghosh, S. Regulation of the NF-κB-Mediated Transcription of

1116   Inflammatory Genes. *Front. Immunol.* **5**, 71 (2014).

1117   51. Kairisalo, M., Korhonen, L., Blomgren, K. & Lindholm, D. X-linked inhibitor of

1118   apoptosis protein increases mitochondrial antioxidants through NF-kappaB

1119   activation. *Biochem. Biophys. Res. Commun.* **364**, 138–144 (2007).

1120   52. Morgan, M. J. & Liu, Z. Crosstalk of reactive oxygen species and NF-κB signaling.

1121   *Cell Res.* **21**, 103–115 (2011).

1122   53. Janasik, B. *et al.* Effect of Arsenic Exposure on NRF2-KEAP1 Pathway and

1123   Epigenetic Modification. *Biol. Trace Elem. Res.* **185**, 11–19 (2018).

1124   54. Kang, K. A. & Hyun, J. W. Oxidative Stress, Nrf2, and Epigenetic Modification

1125   Contribute to Anticancer Drug Resistance. *Toxicol. Res.* **33**, 1–5 (2017).

1126   55. Muscarella, L. A. *et al.* Regulation of KEAP1 expression by promoter methylation in

1127   malignant gliomas and association with patient's outcome. *Epigenetics* **6**, 317–325

1128   (2011).

1129   56. Raghunath, A. *et al.* Antioxidant response elements: Discovery, classes, regulation

1130   and potential applications. *Redox Biol.* **17**, 297–314 (2018).

1131   57. Klingbeil, E. C., Hew, K. M., Nygaard, U. C. & Nadeau, K. C. Polycyclic aromatic

1132   hydrocarbons, tobacco smoke, and epigenetic remodeling in asthma. *Immunol. Res.*

1133   **58**, 369–373 (2014).

1134   58. Tonelli, C., Chio, I. I. C. & Tuveson, D. A. Transcriptional Regulation by Nrf2.

1135   *Antioxid. Redox Signal.* **29**, 1727–1745 (2017).

1136   59. Wittkopp, S. *et al.* Nrf2-related gene expression and exposure to traffic-related air

1137   pollution in elderly subjects with cardiovascular disease: An exploratory panel study.

1138   *J. Expo. Sci. Environ. Epidemiol.* **26**, 141–149 (2016).

1139   60. del Rosario Olguin, S., Camacho, R. & Espinosa, J. J. Epigenetic regulation of

1140   Cytochrome P450 (CYP1A1) by polycyclic aromatic hydrocarbons. *FASEB J.* **27**,

1141   lb95–lb95 (2013).

1142   61. Murrison, L. B., Brandt, E. B., Myers, J. B. & Hershey, G. K. K. Environmental

1143   exposures and mechanisms in allergy and asthma development. *J. Clin. Invest.* **129**,

1144   1504–1515 (2019).

1145   62. Kim, Y.-M., Kim, Y.-S., Jeon, S. G. & Kim, Y.-K. Immunopathogenesis of Allergic

1146   Asthma: More Than the Th2 Hypothesis. *Allergy Asthma Immunol. Res.* **5**, 189–196

1147   (2013).

1148   63. Pahl, H. L. Activators and target genes of Rel/NF-κB transcription factors. *Oncogene*

1149   **18**, 6853–6866 (1999).

1150   64. Bosnjak, B., Stelzmueller, B., Erb, K. J. & Epstein, M. M. Treatment of allergic

1151   asthma: Modulation of Th2 cells and their responses. *Respir. Res.* **12**, 114 (2011).

1152   65. Christensen, S. *et al.* Prenatal environmental tobacco smoke exposure increases

1153   allergic asthma risk with methylation changes in mice. *Environ. Mol. Mutagen.* **58**,

1154   423–433 (2017).

1155   66. Lee, K. W. K. *et al.* Prenatal Exposure to Maternal Cigarette Smoking and DNA

1156   Methylation: Epigenome-Wide Association in a Discovery Sample of Adolescents

1157   and Replication in an Independent Cohort at Birth through 17 Years of Age. *Environ.*

1158   *Health Perspect.* **123**, 193–199 (2015).

1159   67. Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by

1160   Surrogate Variable Analysis. *PLOS Genet.* **3**, e161 (2007).

1161   68. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-Wide Association

1162   Studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).

1163   69. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: software for

1164   combining, analyzing, grouping and correcting spatially correlated P-values.

1165   *Bioinformatics* **28**, 2986–2988 (2012).

1166   70. Xu, Z., Niu, L., Li, L. & Taylor, J. A. ENmix: a novel background correction method

1167   for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.* **44**, e20 (2016).

1168   71. Hannon, E. *et al.* Characterizing genetic and environmental influences on variable

1169   DNA methylation using monozygotic and dizygotic twins. *PLOS Genet.* **14**,

1170   e1007544 (2018).

1171   72. Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation

1172   across the human life course. *Genome Biol.* **17**, 61 (2016).

1173   73. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing

1174   data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288

1175   (2016).

1176   74. Liu, X. *et al.* Effects of prenatal exposure to air particulate matter on the risk of

1177   preterm birth and roles of maternal and cord blood LINE-1 methylation: A birth

1178   cohort study in Guangzhou, China. *Environ. Int.* **133**, 105177 (2019).

1179   75. Baccarelli, A. *et al.* Rapid DNA Methylation Changes after Exposure to Traffic

1180   Particles. *Am. J. Respir. Crit. Care Med.* **179**, 572–578 (2009).

1181   76. Chi, G. C. *et al.* Long-term outdoor air pollution and DNA methylation in circulating

1182   monocytes: results from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ.*

1183   *Health* **15**, 119 (2016).

1184   77. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006

1185   (2002).

1186   78. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-3.0.* (1996).

1187   79. Glinsky, G. V. Transposable Elements and DNA Methylation Create in Embryonic

1188   Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers

1189   and Noncoding RNAs. *Genome Biol. Evol.* **7**, 1432–1454 (2015).

1190   80. Dai, R. *et al.* Wheeze trajectories: Determinants and outcomes in the CHILD Cohort

1191   Study. *J. Allergy Clin. Immunol.* (2021) doi:10.1016/j.jaci.2021.10.039.

1192   81. Levy, I., Mihele, C., Lu, G., Narayan, J. & Brook, J. R. Evaluating Multipollutant

1193   Exposure and Urban Air Quality: Pollutant Interrelationships, Neighborhood

1194   Variability, and Nitrogen Dioxide as a Proxy Pollutant. *Environ. Health Perspect.*

1195   **122**, 65–72 (2014).

1196   82. Hazlehurst, M. F. *et al.* Maternal exposure to PM2.5 during pregnancy and asthma

1197   risk in early childhood: Consideration of phases of fetal lung development. *Environ.*

1198   *Epidemiol.* **5**, e130 (2021).

1199   83. Canadian Council of Ministers of the Environment. *Guidance Document on*

1200   *Achievement Determination for Canadian Ambient Air Quality Standards for*

1201   *Nitrogen Dioxide*. 26

1202   https://ccme.ca/en/res/gdadforcaaqsfornitrogendioxide_en1.0.pdf (2020).

1203   84. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone,

1204   nitrogen dioxide, sulfur dioxide and carbon monoxide.

1205   https://www.who.int/publications-detail-redirect/9789240034228.

1206   85. Packer, A. I., Mailutha, K. G., Ambrozewicz, L. A. & Wolgemuth, D. J. Regulation

1207   of the Hoxa4 and Hoxa5 genes in the embryonic mouse lung by retinoic acid and

1208   TGFbeta1: implications for lung development and patterning. *Dev. Dyn. Off. Publ.*

1209   *Am. Assoc. Anat.* **217**, 62–74 (2000).

1210   86. Ptaschinski, C., Hrycaj, S. M., Schaller, M. A., Wellik, D. M. & Lukacs, N. W. Hox5

1211   paralogous genes modulate Th2 cell function during chronic allergic inflammation

1212   via regulation of Gata3. *J. Immunol. Baltim. Md 1950* **199**, 501–509 (2017).

1213   87. Gruzieva, O. *et al.* Prenatal Particulate Air Pollution and DNA Methylation in

1214   Newborns: An Epigenome-Wide Meta-Analysis. *Environ. Health Perspect.* **127**,

1215   57012 (2019).

1216   88. Golpon, H. A. *et al.* HOX Genes in Human Lung. *Am. J. Pathol.* **158**, 955–966

1217   (2001).

1218   89. Wright, R. J. & Brunst, K. J. Programming of respiratory health in childhood:

1219   influence of outdoor air pollution. *Curr. Opin. Pediatr.* **25**, 232–239 (2013).

1220   90. Joss-Moore, L. A., Albertine, K. H. & Lane, R. H. Epigenetics and the developmental

1221   origins of lung disease. *Mol. Genet. Metab.* **104**, 61–66 (2011).

1222   91. Simon, A. K., Hollander, G. A. & McMichael, A. Evolution of the immune system in

1223   humans from infancy to old age. *Proc. R. Soc. B Biol. Sci.* **282**, 20143085 (2015).

1224   92. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and

1225   beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).

1226   93. Volpe, M. V., Wang, K. T. W., Nielsen, H. C. & Chinoy, M. R. Unique spatial and

1227   cellular expression patterns of Hoxa5, Hoxb4 and Hoxb6 proteins in normal

1228   developing murine lung are modified in pulmonary hypoplasia. *Birt. Defects Res. A.*

1229   *Clin. Mol. Teratol.* **82**, 571–584 (2008).

1230   94. Lezmi, G. *et al.* Mast cells are associated with exacerbations and eosinophilia in

1231   children with severe asthma. *Eur. Respir. J.* **48**, 1320–1328 (2016).

1232   95. Anand, M. *et al.* Requirement of the Lec35 Gene for All Known Classes of

1233   Monosaccharide-P-Dolichol-dependent Glycosyltransferase Reactions in Mammals.

1234   *Mol. Biol. Cell* **12**, 487–501 (2001).

1235   96. de Jong, S. E. *et al.* IgG1 Fc N-glycan galactosylation as a biomarker for immune

1236   activation. *Sci. Rep.* **6**, 28207 (2016).

1237   97. Jennewein, M. F. & Alter, G. The Immunoregulatory Roles of Antibody

1238   Glycosylation. *Trends Immunol.* **38**, 358–372 (2017).

1239   98. Sodemann, E. B. *et al.* Maternal asthma is associated with persistent changes in

1240   allergic offspring antibody glycosylation. *Clin. Exp. Allergy* **50**, 520–531 (2020).

1241   99. Cardenas, A. *et al.* Controlled human exposures to diesel exhaust: a human

1242   epigenome-wide experiment of target bronchial epithelial cells. *Environ. Epigenetics*

1243   **7**, dvab003 (2021).

1244   100.   Lee, M. K., Hong, Y., Kim, S.-Y., London, S. J. & Kim, W. J. DNA methylation

1245   and smoking in Korean adults: epigenome-wide association study. *Clin. Epigenetics*

1246   **8**, 103 (2016).

1247   101.   Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National

1248   Center for Biotechnology Information; [1988] - Homo sapiens (human) C5orf63

1249   chromosome 5 open reading frame 63.

1250   https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=401207.

1251   102.   Jesenak, M., Zelieskova, M. & Babusikova, E. Oxidative Stress and Bronchial

1252   Asthma in Children—Causes or Consequences? *Front. Pediatr.* **5**, 162 (2017).

1253   103.   ZNF837 Gene - GeneCards. https://www.genecards.org/cgi-

1254   bin/carddisp.pl?gene=ZNF837.

1255   104.   Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease

1256   Genome Sequence Analyses. *Curr. Protoc. Bioinforma.* **54**, 1.30.1-1.30.33 (2016).

1257   105.   Vogel, C. F. A., Van Winkle, L. S., Esser, C. & Haarmann-Stemmann, T. The

1258   aryl hydrocarbon receptor as a target of environmental stressors – Implications for

1259   pollution mediated stress and inflammatory responses. *Redox Biol.* **34**, 101530

1260   (2020).

1261   106.   Ong, C.-T. *et al.* Target Selectivity of Vertebrate Notch Proteins: Collaboration

1262   Between Discrete DOmains and CSL-Binding Site Architecture Determines

1263   Activation Probability*. *J. Biol. Chem.* **281**, 5106–5119 (2006).

1264   107.   Spergel, J. M. Epidemiology of atopic dermatitis and atopic march in children.

1265   *Immunol. Allergy Clin. North Am.* **30**, 269–280 (2010).

1266   108.   Smart, J. M. & Kemp, A. S. Increased Th1 and Th2 allergen-induced cytokine

1267   responses in children with atopic disease. *Clin. Exp. Allergy* **32**, 796–802 (2002).

1268   109.   Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G. & Kobor, M. S. BECon: a

1269   tool for interpreting DNA methylation findings from blood in the context of brain.

1270   *Transl. Psychiatry* **7**, e1187 (2017).

1271

1272

1273

1274   **Supplementary figures**

1275

1276
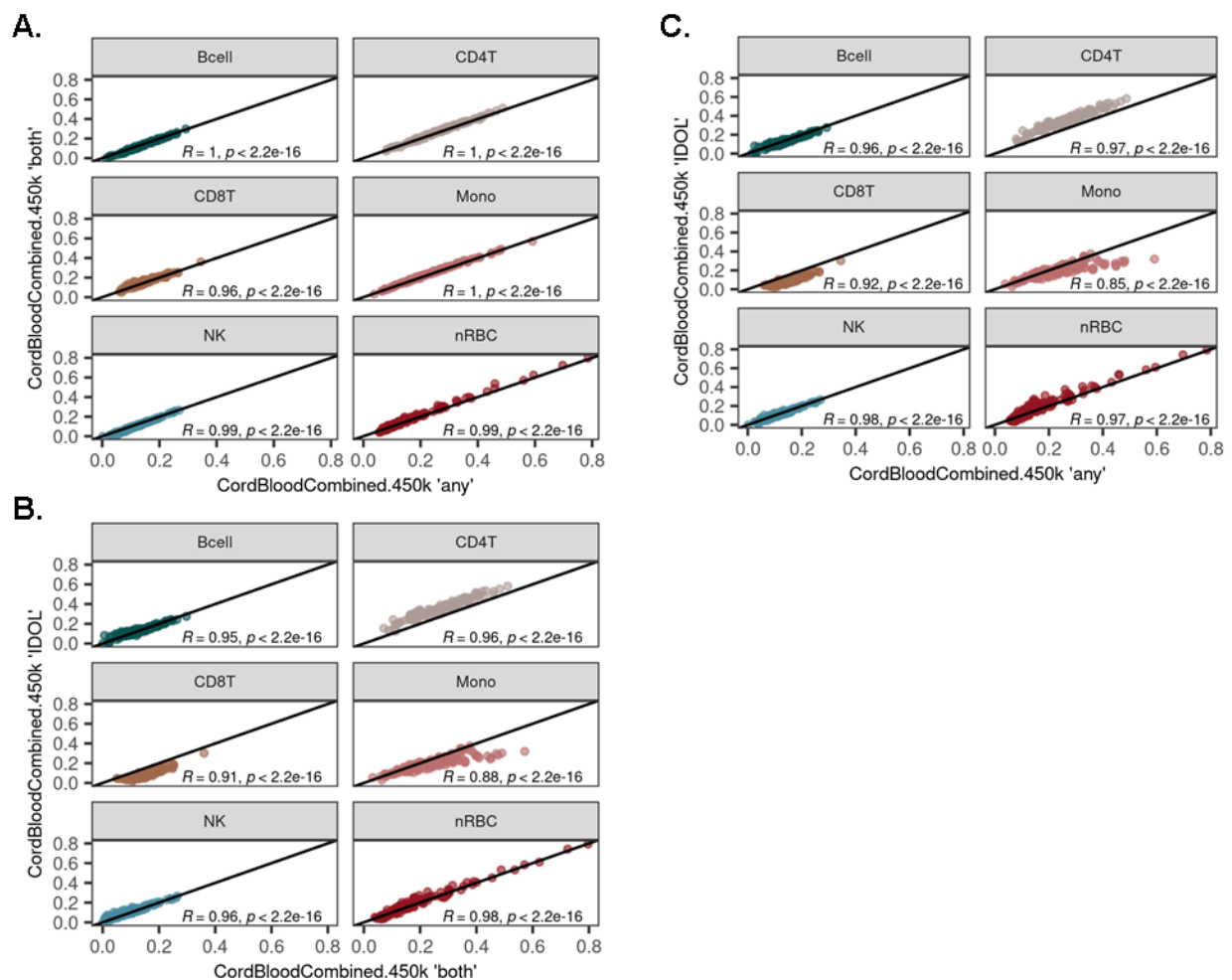
1277   **Supplementary Figure 1. Umbilical cord white blood cell deconvolution methods**
1278   **perform similarly in estimating monocyte cell proportions.** (**A**) Deconvolution of
1279   umbilical cord white blood cells with the FlowSorted.CordBloodCombined.450k package
1280   using "any", which selects the top100 probes with the greatest magnitude of difference
1281   between cell types, or "both", which selects equal numbers of probes with the greatest
1282   magnitude of effect in either the positive or negative direction. (**B**) Deconvolution of
1283   umbilical cord white blood cells with the FlowSorted.CordBloodCombined.450k package
1284   using "any" of the top100 probes, or "IDOL", which using a pre-determined set of
1285   optimized probes for deconvolution of cord blood white cells. (**C**) Deconvolution of
1286   umbilical cord white blood cells with the FlowSorted.CordBloodCombined.450k package
1287   using "both" to select the equal numbers of probes with the greatest magnitude of effect
1288   in the positive (50) or negative (50) direction, or using optimized "IDOL" probes.
1289
1290
1291
1292

52

1293
1294
1295
1296
1297
1298
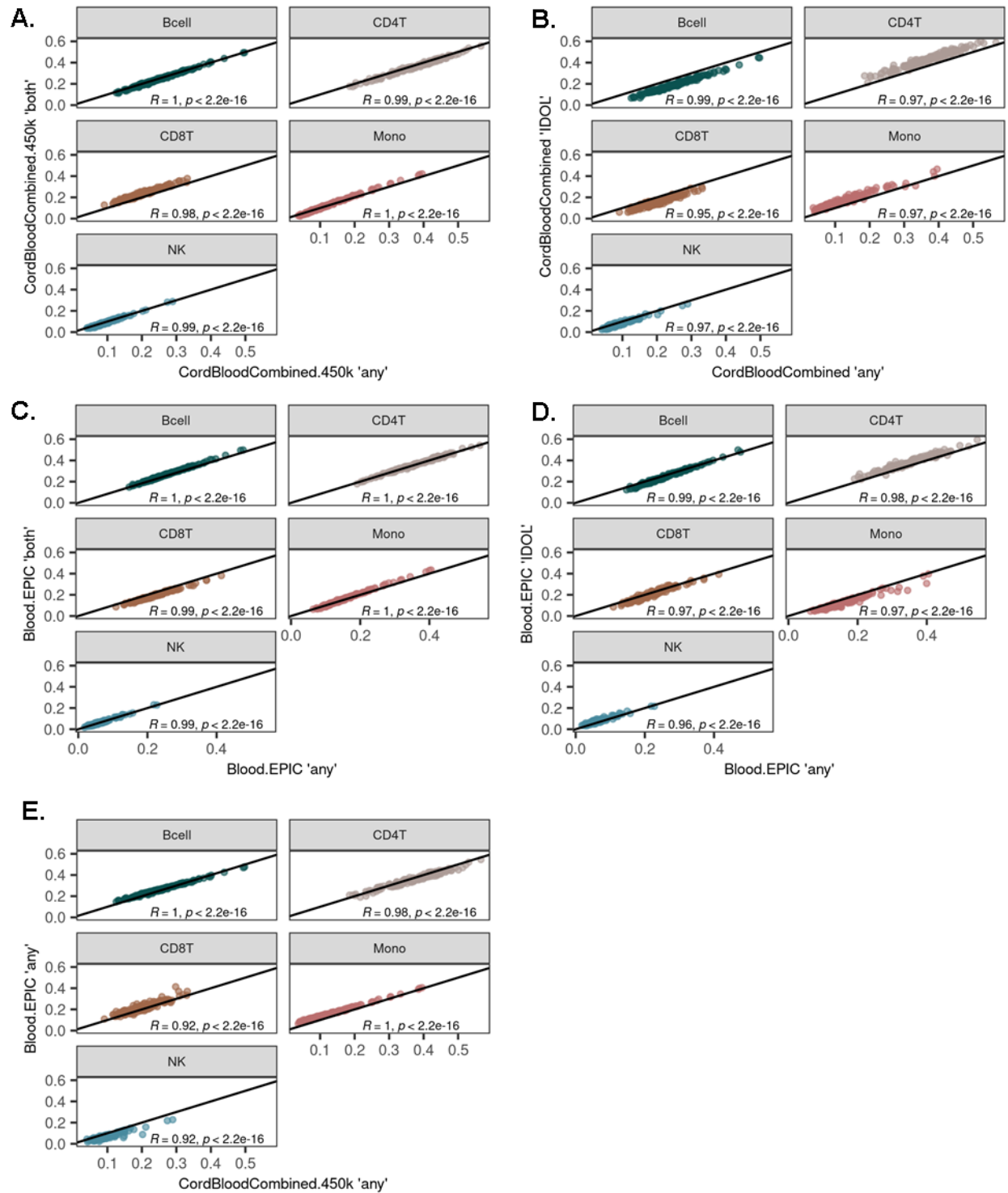
Supplementary Figure 2. Age one one peripheral white blood cell deconvolution methods perform similarly in estimating monocyte cell type proportions. (**A**) Deconvolution of age one peripheral white blood cells using the FlowSorted.CordBloodCombined.450k package (cord blood references set) using "any" probe select, which returns the top100 probes with the greatest magnitude of difference between cell types, or "both" probe select, which returns equal numbers of probes with the greatest magnitude of effect in either the positive or negative direction. (**B**) Deconvolution of age one peripheral white blood cells with the FlowSorted.CordBloodCombined.450k package selecting "any" of the top100 probes, or selecting cord blood "IDOL" probes, which comprises a set of predetermined probes optimized for deconvolution of cord blood white cells. (**C**) Deconvolution of age one peripheral white blood cells with the FlowSorted.Blood.EPIC package (adult peripheral blood reference set) and selecting "any" probes, or "both" probes. (**D**) Deconvolution of age one peripheral white blood cells with the FlowSorted.Blood.EPIC package and selecting "any" probes, or using adult "IDOL" probes, which comprises a set of predetermined probes optimized for deconvolution of adult peripheral white blood cells. (**E**) Deconvolution of age one peripheral white blood cells using the FlowSorted.CordBloodCombined.450k package using "any" probe select, or deconvolution using the FlowSorted.Blood.EPIC package and selecting "any" probes.
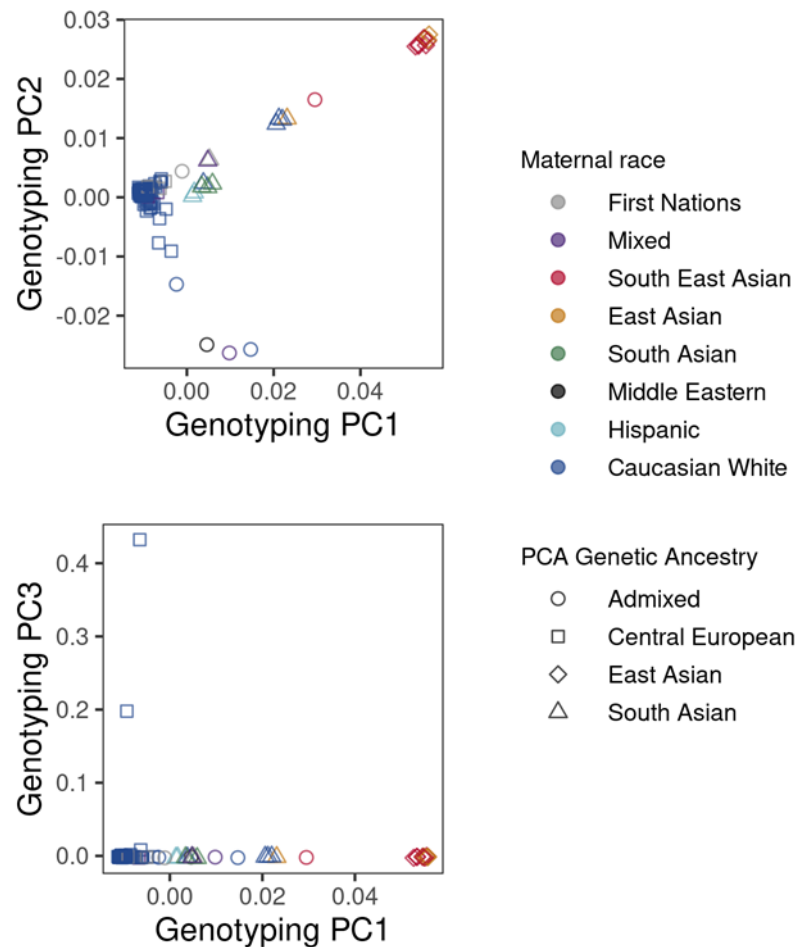
Supplementary Figure 3. The first three genetic ancestry principle components (PCs) separate CHILD participants (N=131) by self-identified race. Point colour represents maternal-self identified race and point shape represents genetic ancestry as determined by principle component analysis.

A.
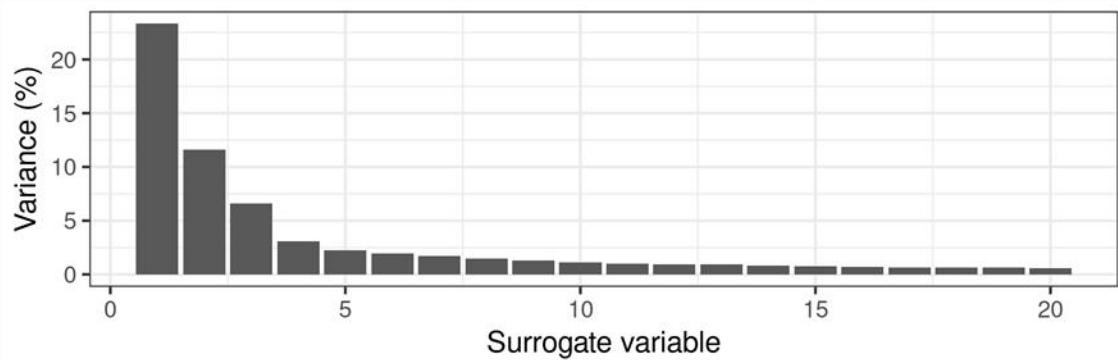
B.
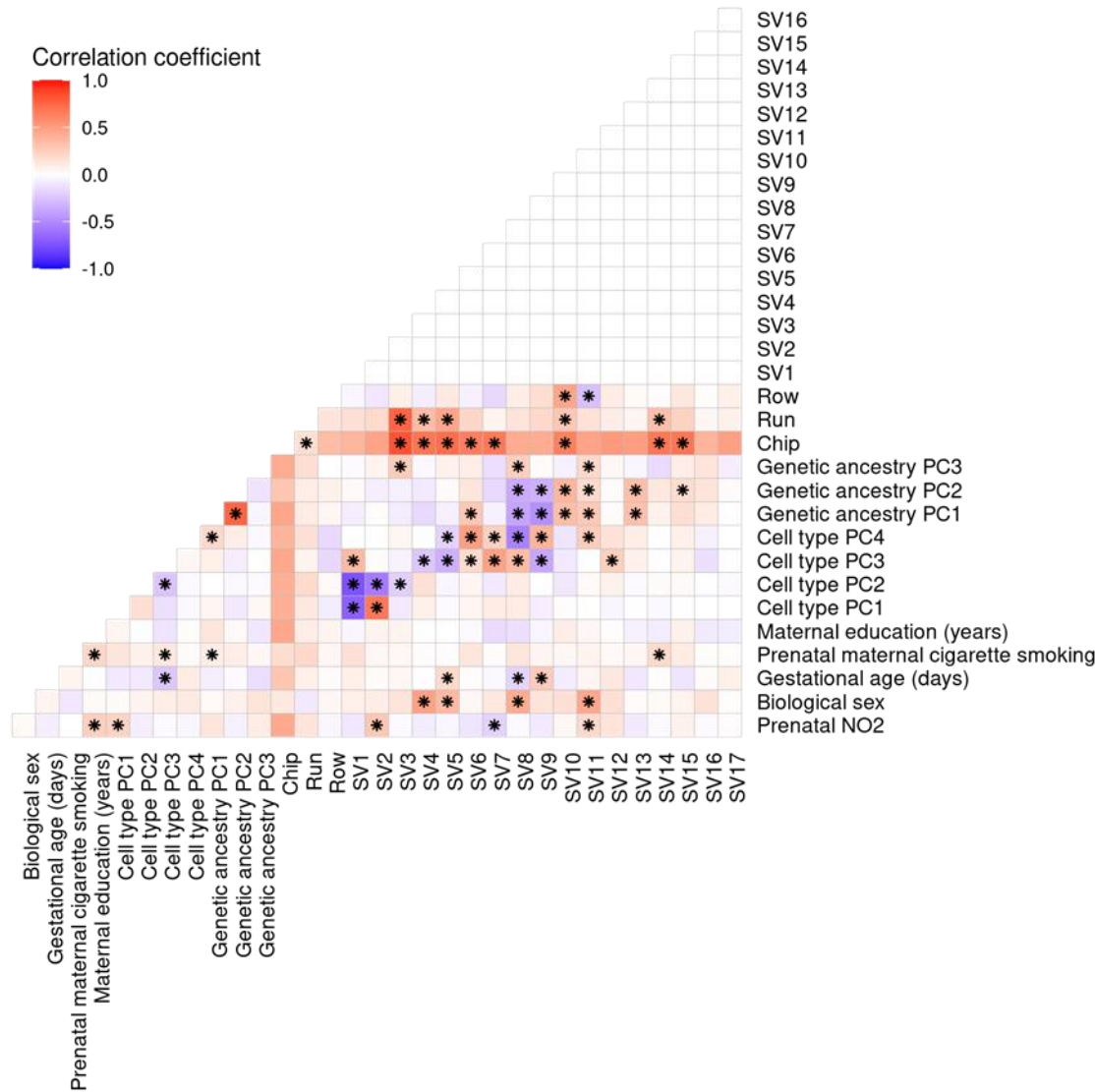
1323
1324
1325
1326

1327 **Supplementary Figure 4. Surrogate variable analysis (SVA) captures technical**
1328 **variation in cord blood DNA methylation (N=128).** (**A**) Variation captured by the first
1329 20 surrogate variables. Surrogate variables capture variation that is not accounted for by
1330 covariates specified in the model of prenatal NO_2 exposure. The first 17 surrogate
1331 variables were included in the prenatal model of NO_2 exposure based on recommendation
1332 of the SVA package. (**B**) Correlations between covariates in the prenatal NO_2 model and
1333 the first 17 surrogate variables returned by SVA. Note that the surrogate variables are
1334 strongly correlated with chip and run. *Represent a significant correlation between two
1335 variables.

1336
1337
1338
1339
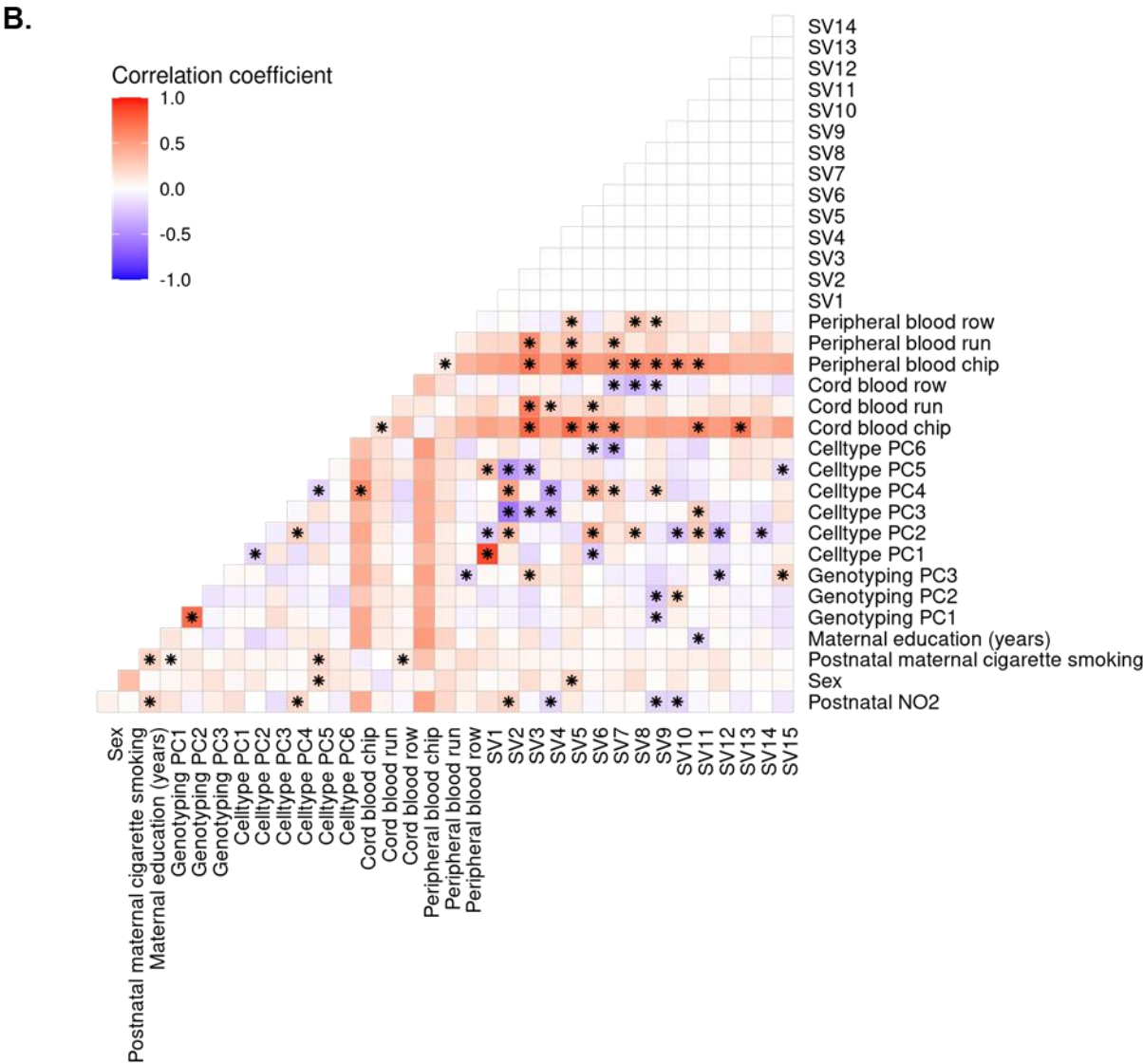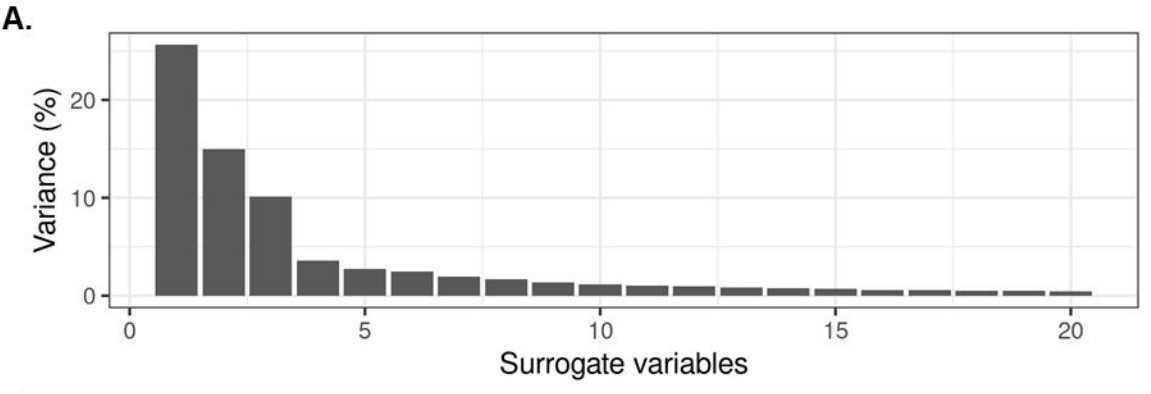
1340 **Supplementary Figure 5. Surrogate variable analysis (SVA) captures technical**
1341 **variation in postnatal-specific DNA methylation (N=125).** (**A**) Variation captured by
1342 the first 20 surrogate variables. Surrogate variables capture variation that is not accounted
1343 for by covariates specified in the model examining postnatal-specific DNA methylation
1344 changes associated with year one NO_2 exposure. The first 15 surrogate variables were
1345 included in models examining year one NO_2 exposure and postnatal-specific DNA
1346 methylation. (**B**) Correlations between covariates included in models examining
1347 postnatal-specific DNA methylation changes associated with year one NO2 exposure and
1348 the first 15 surrogate variables returned by SVA. Note that surrogate variables are
1349 strongly and correlated with chip and run. *Represent a significant correlation between
1350 two variables.
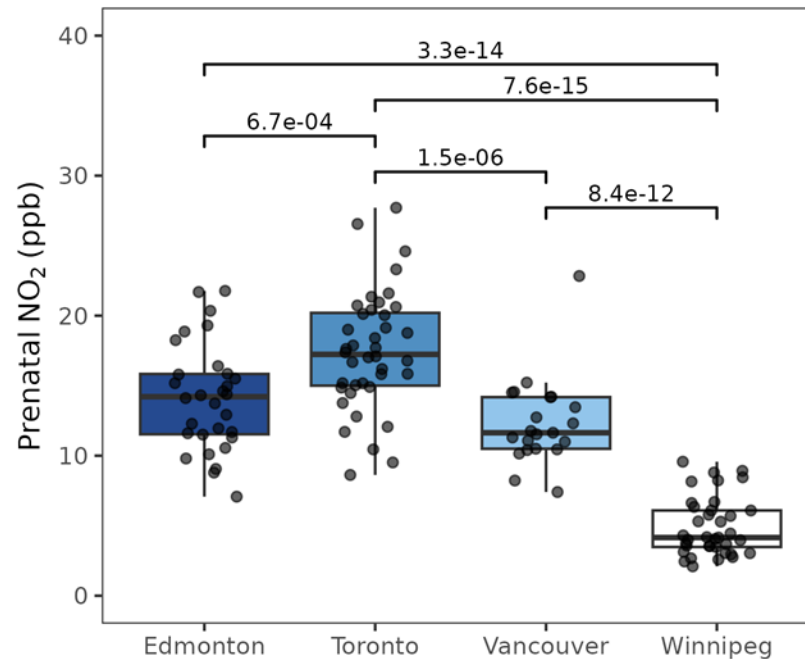
Supplementary Figure 6. Variation in prenatal NO_2 exposure estimates is greater between study sites compared within study sites for participants included in analysis of prenatal DNA methylation changes (N=128). The four CHILD study sites are located in the Canadian cities of Edmonton, Toronto, Vancouver, and Winnipeg. An ANOVA test indicated that prenatal NO_2 exposure estimates were significantly (p = 1.0e-30) different between study sites. Post-hoc Tukey tests using the Honest Significant Difference method revealed that mean prenatal NO_2 exposure estimates were significantly (adjusted $p < 0.05$) different between all study sites except for Edmonton and Toronto. The mean NO_2 exposure estimate across all study sites was 12.2 ± 6.1 ppb. The mean and standard deviation of NO_2 exposure estimates for Edmonton, Toronto, Vancouver, and Winnipeg were 14.1 ± 3.8 ppb, 17.4 ± 4.3 ppb, 12.4 ± 3.2 ppb, and 4.9 ± 2.1 ppb, respectively.
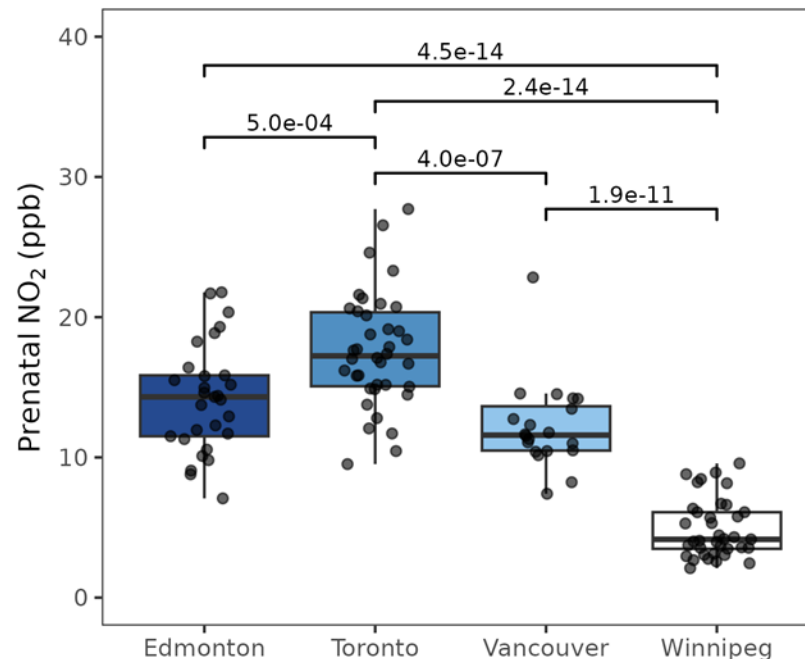
Supplementary Figure 7. Variation in prenatal NO_2 exposure estimates is greater between study sites compared to within study sites for participants included in analysis of persistent DNA methylation changes (N=124). The four CHILD study sites are located in the Canadian cities of Edmonton, Toronto, Vancouver, and Winnipeg. An ANOVA test indicated that prenatal NO_2 exposure estimates were significantly (p = 4.5e-31) different between study sites. Post-hoc Tukey tests using the Honest Significant Difference method revealed that mean prenatal NO_2 exposure estimates were significantly (adjusted p < 0.05) different between all study sites except for Edmonton and Toronto. The mean NO_2 exposure estimate across all study sites was 12.2 ± 6.1 ppb. The mean and standard deviation of NO_2 exposure estimates for Edmonton, Toronto, Vancouver, and Winnipeg were 14.3 ± 3.9 ppb, 17.6 ± 4.1 ppb, 12.2 ± 3.2 ppb, and 4.9 ± 2.1 ppb, respectively.
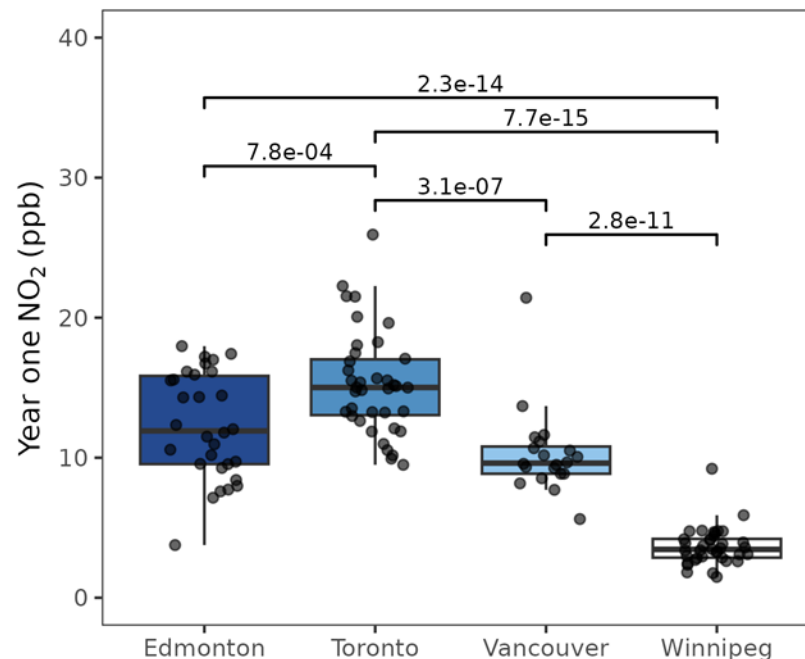
Supplementary Figure 8. Variation in year one NO_2 exposure estimates is greater between study sites compared to within study sites for participants included in analysis of postnatal-specific DNA methylation changes (N=124). The four CHILD study sites are located in the Canadian cities of Edmonton, Toronto, Vancouver, and Winnipeg. An ANOVA test indicated that year one NO_2 exposure estimates were significantly ($p = 3.1e\text{-}31$) different between study sites. Post-hoc Tukey tests using the Honest Significant Difference method revealed that mean prenatal NO_2 exposure estimates were significantly (adjusted $p < 0.05$) different between all study sites except for Edmonton and Toronto. The mean NO_2 exposure estimate across all study sites was 10.3 ± 5.6 ppb. The mean and standard deviation of NO_2 exposure estimates for Edmonton, Toronto, Vancouver, and Winnipeg were 12.3 ± 3.8 ppb, 15.3 ± 3.7 ppb, 10.3 ± 3.2 ppb, and 3.7 ± 1.4 ppb, respectively.

A.

B.

1390
1391

63
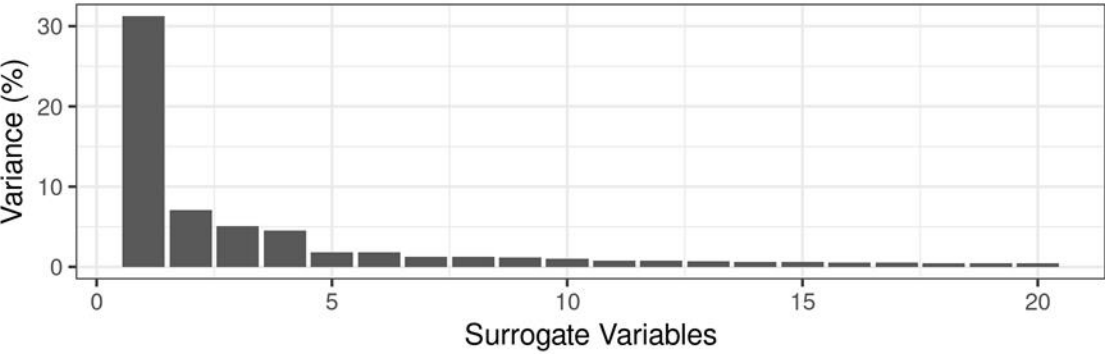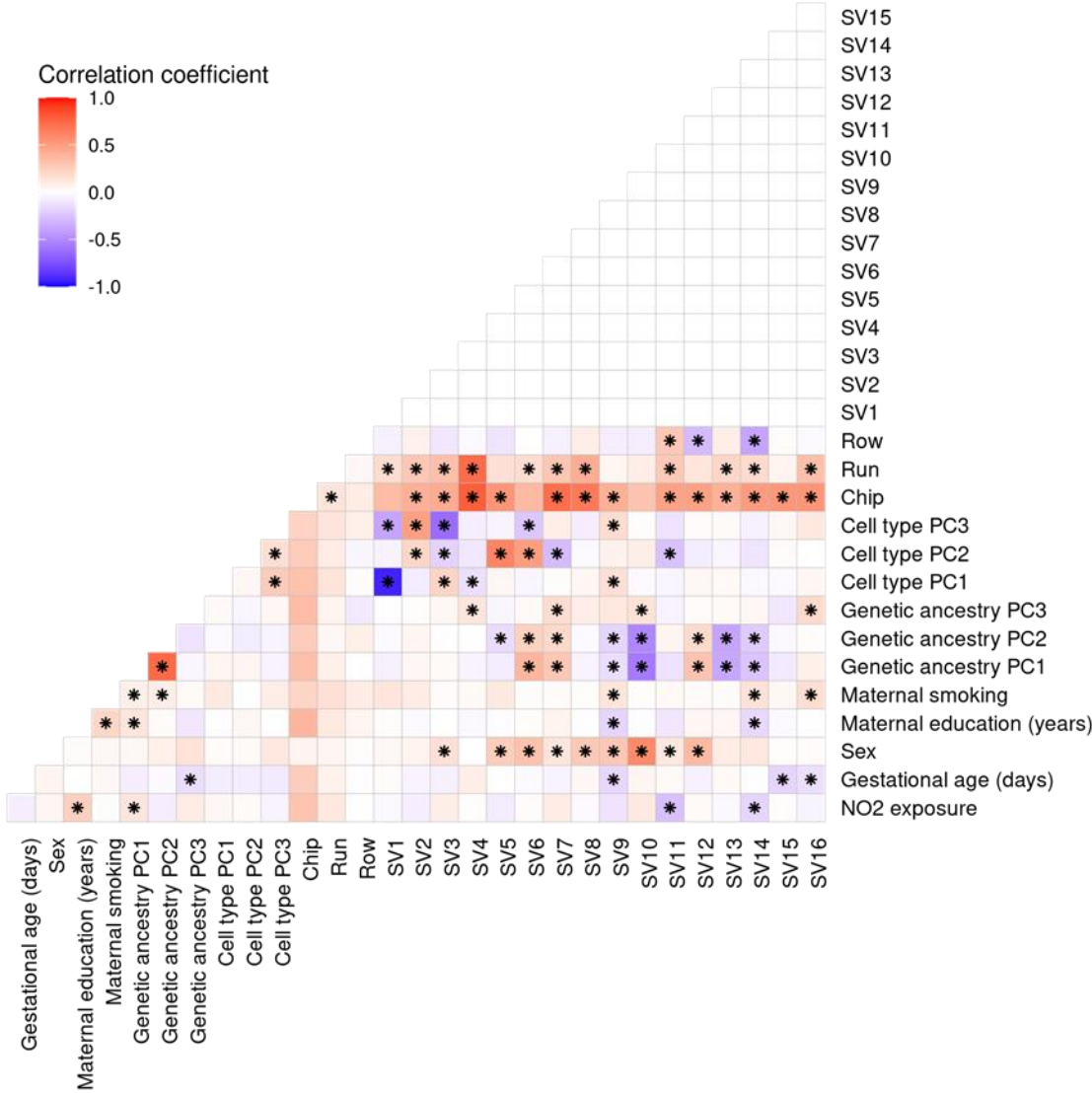
1392 **Supplementary Figure 9. Surrogate variable analysis (SVA) captures technical**
1393 **variation associated across cord blood and peripheral blood DNA methylation used**
1394 **in the investigation of DNA methylation persistence (N=124).** (**A**) Variation captured
1395 by the first 20 surrogate variables. Surrogate variables capture variation that is not
1396 accounted for by covariates specified in the model. The SVA package suggested
1397 inclusion of the first 96 surrogate variables, which would significantly reduce the power
1398 available to detect DNA methylation changes. Therefore, we included the first 16
1399 surrogate variables in linear mixed models examining the persistence of DNA
1400 methylation changes as these surrogate variables accounted for ~60% of the captured
1401 variance, which was similar to variance captured by the first 17 cord blood surrogate
1402 variables (~60%) and the first 15 postnatal-specific surrogate variables (~69%). (**B**)
1403 Correlations between covariates included in linear mixed models investigating the
1404 persistence of cord blood DNA methylation changes and the first 16 surrogate variables
1405 returned by SVA. Note that the surrogate variables are strongly and positively correlated
1406 with chip and run. *Represent a significant correlation between two variables.
1407
1408

1409
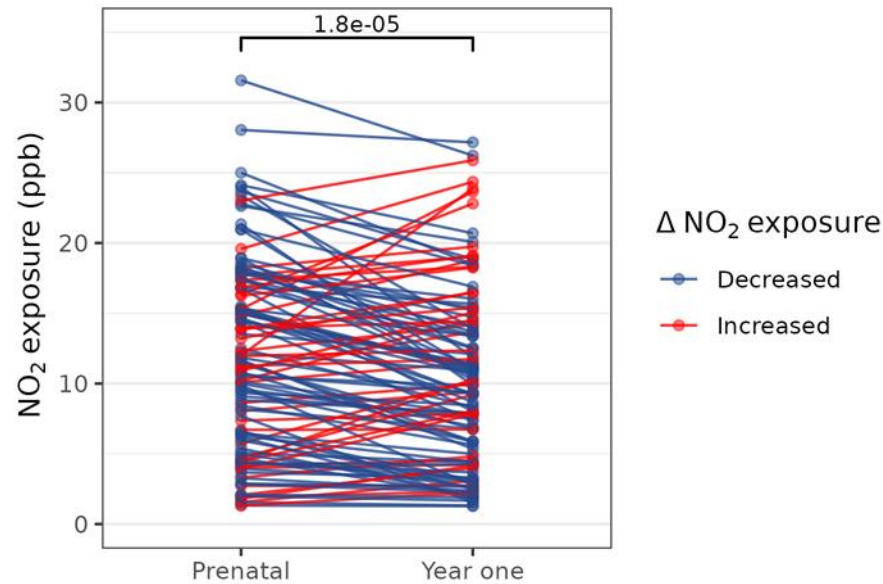
Supplementary Figure 10. Year one NO_2 exposure estimates was slightly lower than prenatal NO_2 exposure estimates for CHILD participants. Paired t-tests indicated this difference was significant (p=1.8e-05) for participants (N=122) that had available both prenatal and year one NO_2 exposure estimates. For the majority of participants (N=82; shown in blue) year one NO_2 exposure estimates were lower than prenatal NO_2 exposure estimates, while year one NO_2 exposure estimates were higher than prenatal NO_2 exposure estimates for the remaining participants (N=40; shown in red).

A.

B.
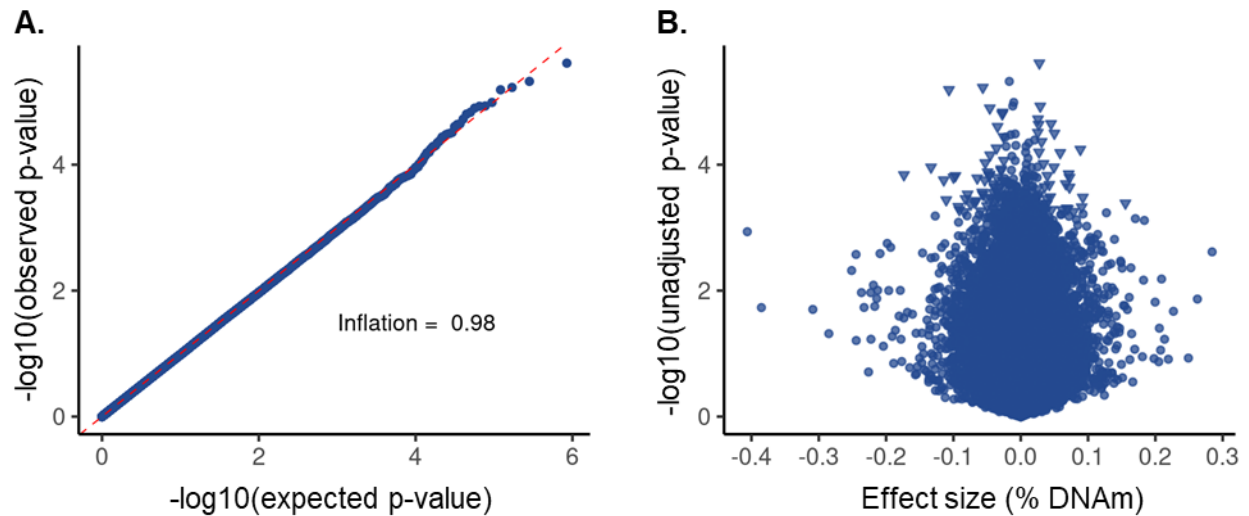
Supplementary Figure 11. No individual CpGs were significant in the epigenome-wide analysis of prenatal NO$_2$ exposure in CHILD participants (N=128). (A) The qqplot did not exhibit significant inflation or deflation suggesting adequate model fit. (**B**) No p-values surpassed our biological cut-off of an absolute effect size of prenatal NO2 exposure >0.025% DNA methylation at a false discovery rate of 5%. Triangular points represent the top 100 CpGs used in our investigation of persistence of individual cord blood DNA methylation changes.
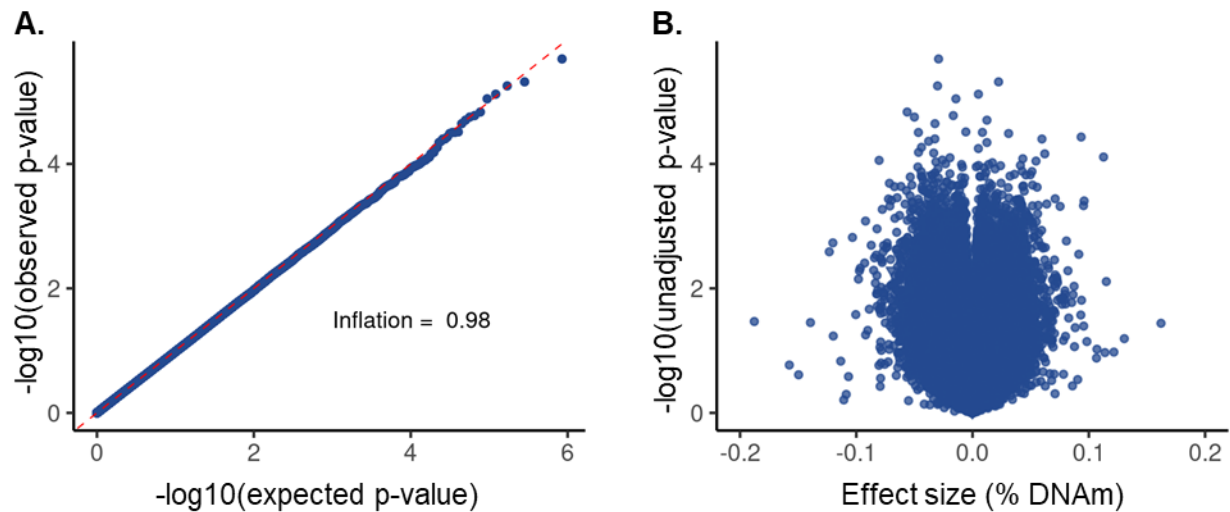
Supplementary Figure 12. No individual CpGs were significant in the postnatal-specific epigenome-wide analysis year one NO_2 exposure in CHILD participants (N=125). (**A**) The qqplot did not exhibit significant inflation or deflation suggesting adequate model fit. (**B**) No p-values surpassed our biological cut-off of an absolute effect size of year one NO_2 exposure >0.025% DNA methylation at a false discovery rate of 5%. Triangular points represent the top 100 CpGs used in our investigation of persistence of individual cord blood DNA methylation changes.

67
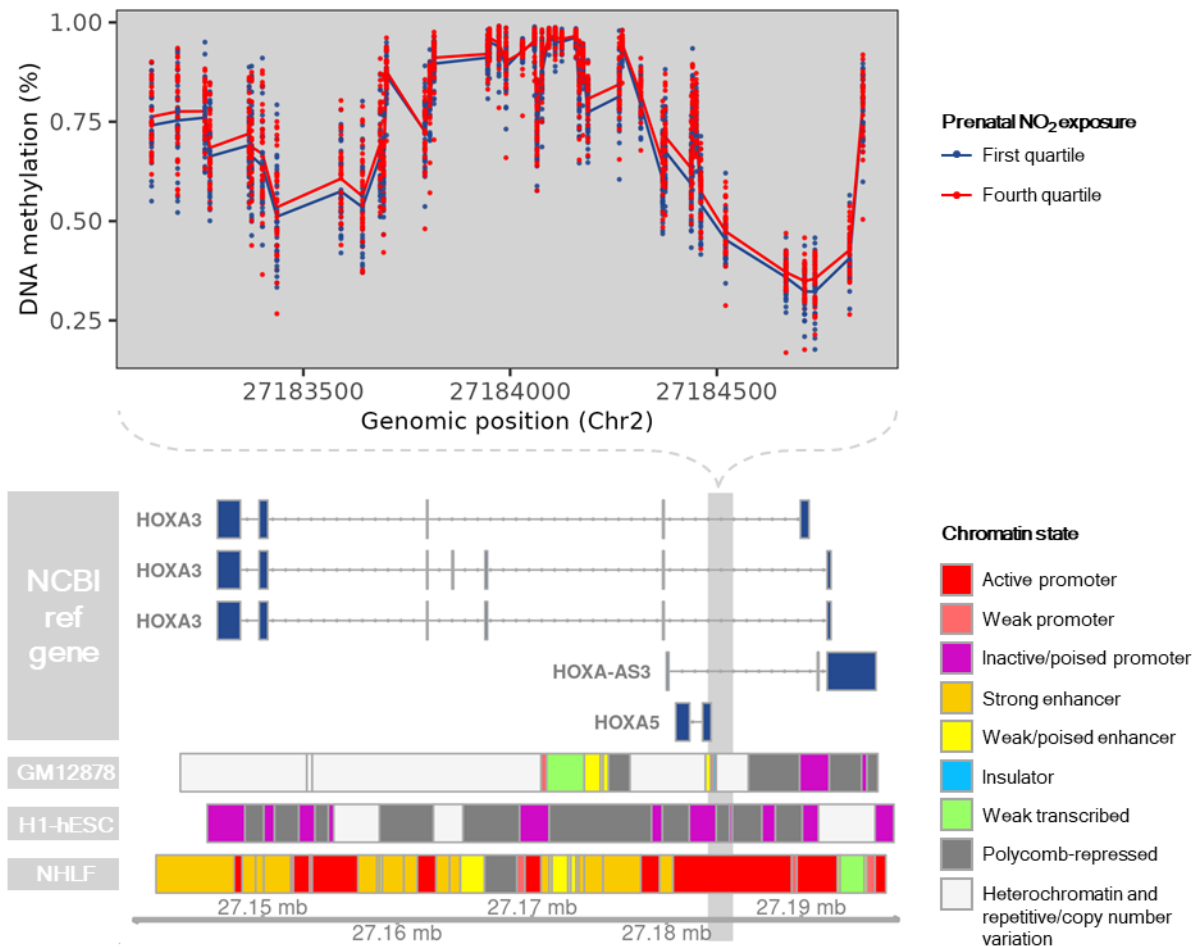
Supplementary Figure 13. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr2: 27183133-27184854. DNA methylation at CpGs contained within Chr2: 27183133-27184854 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack( )* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
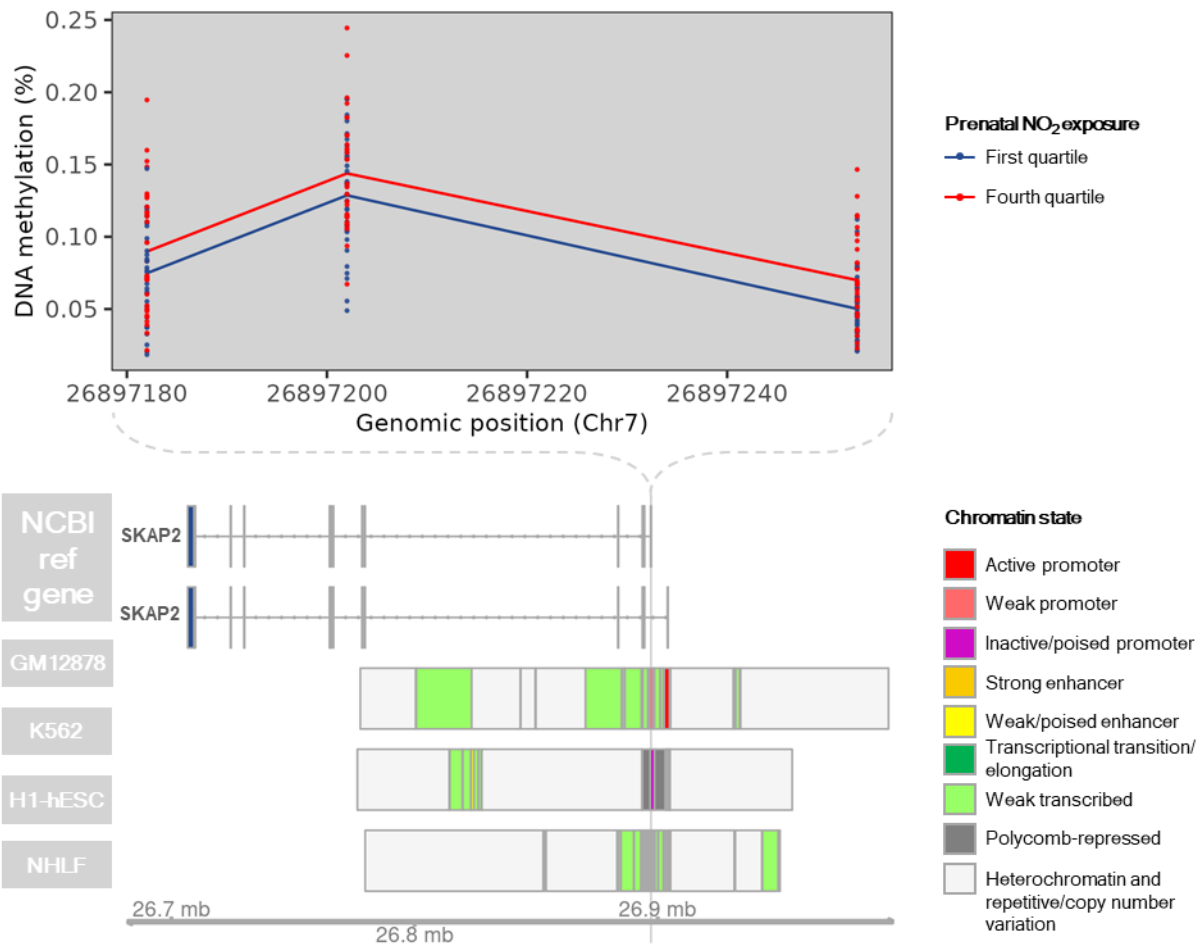
68
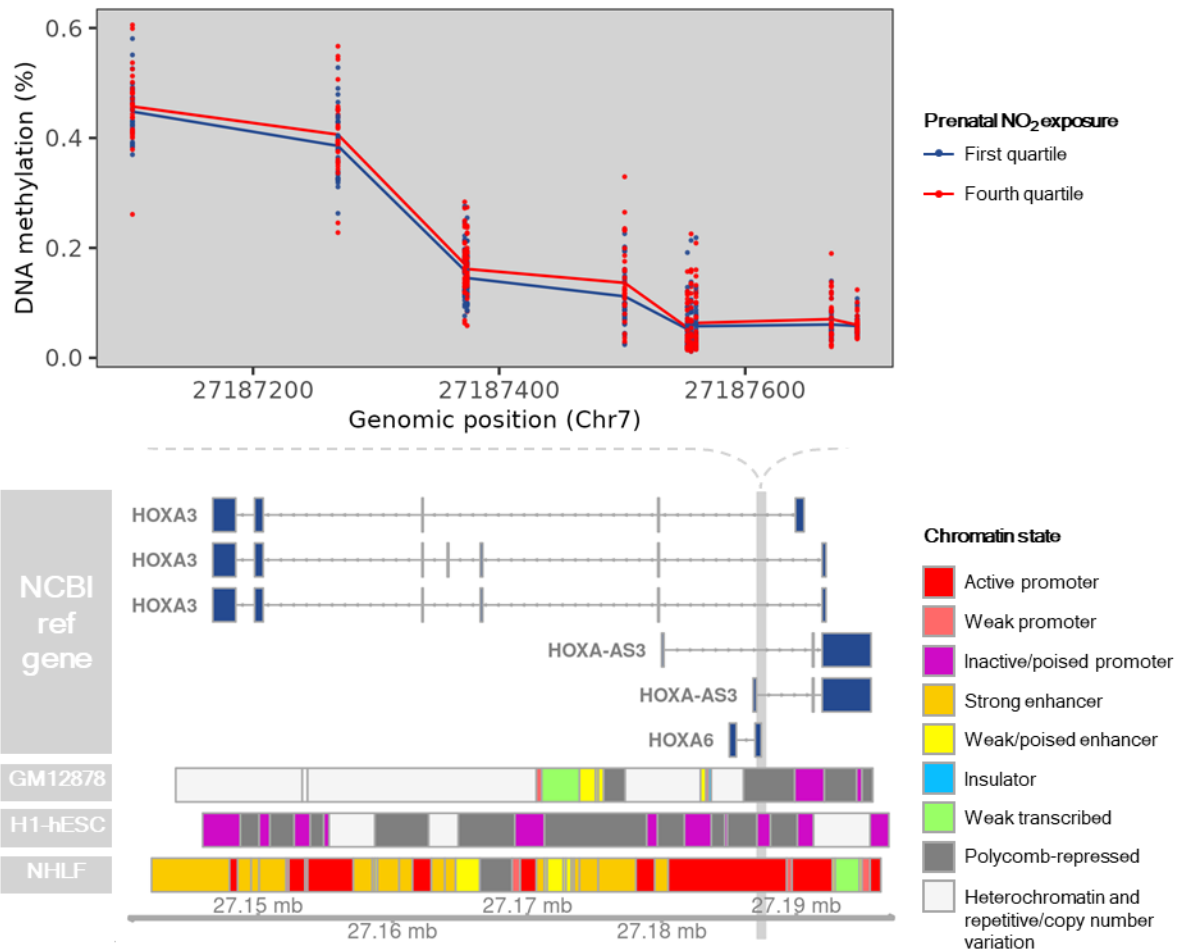
Supplementary Figure 14. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr7: 27187102-27187692. DNA methylation at CpGs contained within Chr7: 27187102-27187692 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
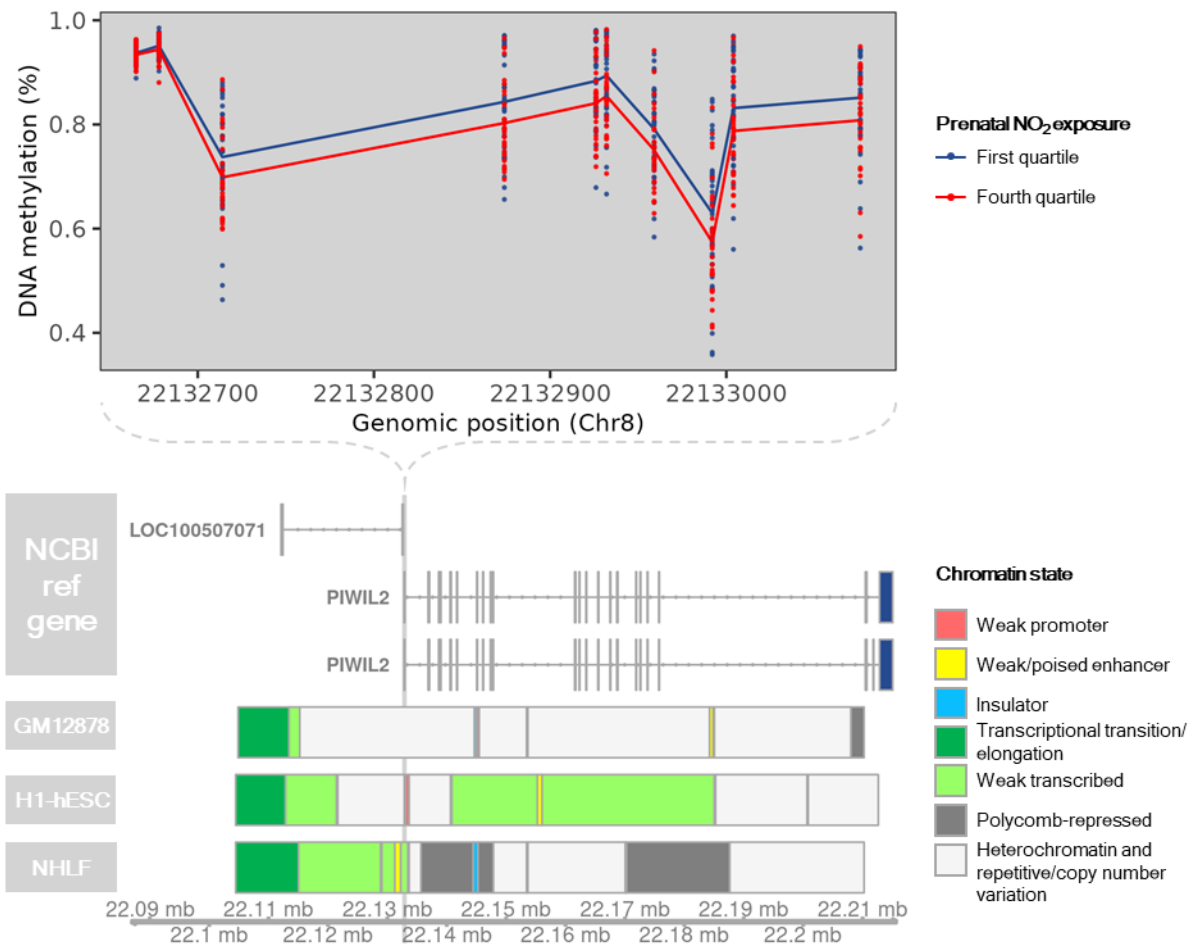
Supplementary Figure 15. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr8: 22132665-22133077. DNA methylation at CpGs contained within Chr8:22132665-22133077 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
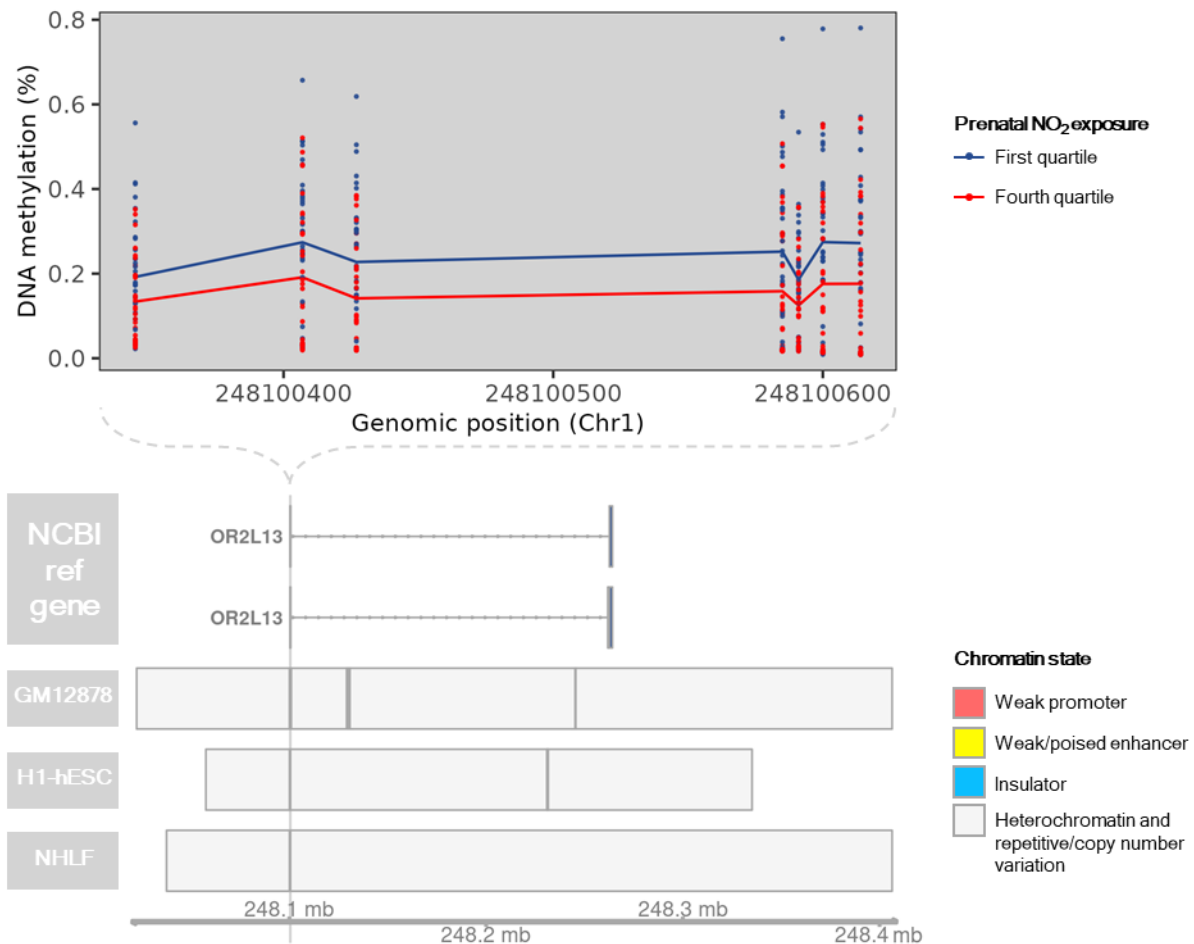
70
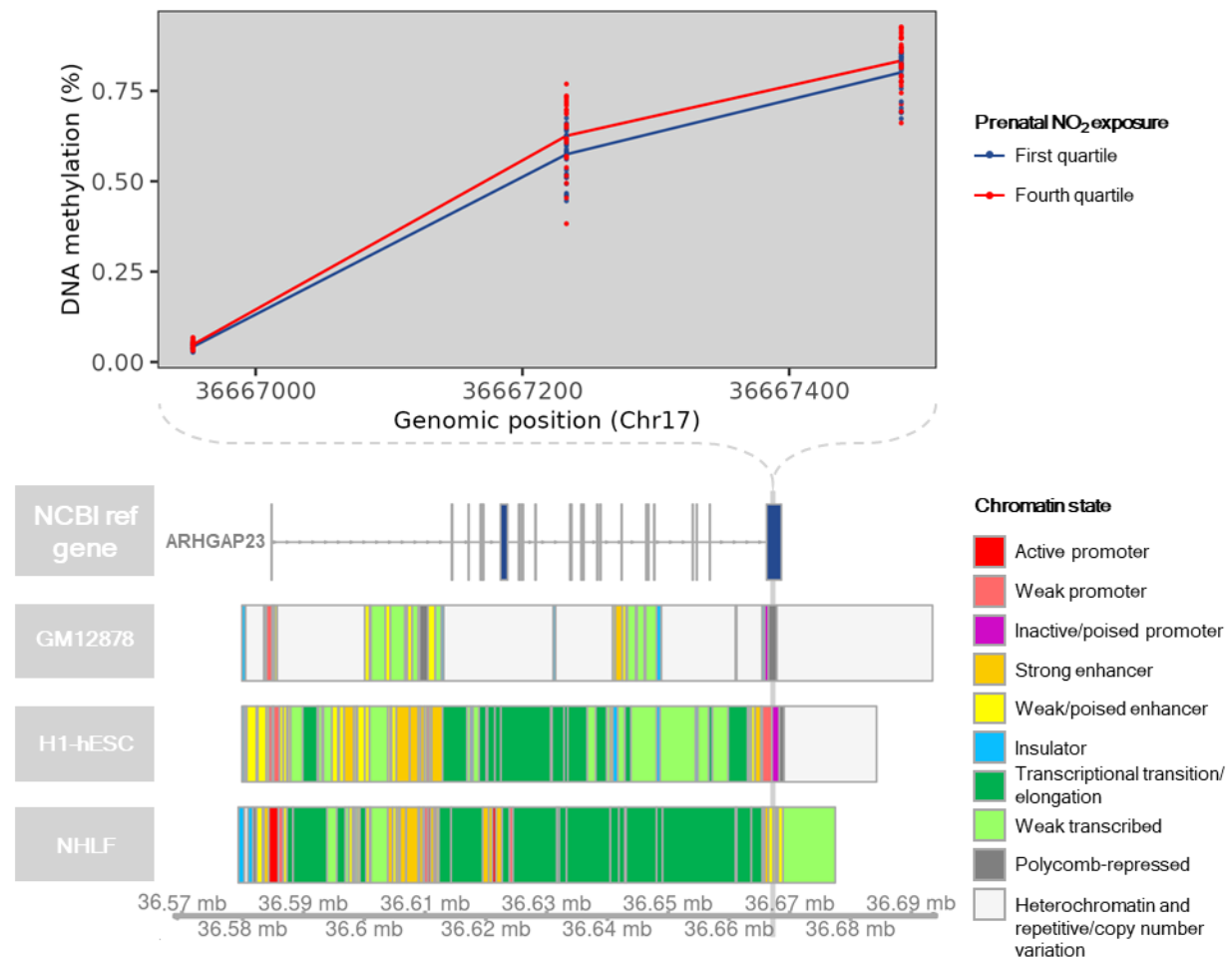
1464
Supplementary Figure 16. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr17:36666953-36667485.
DNA methylation at CpGs contained within Chr17:36666953-36667485 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack( )* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
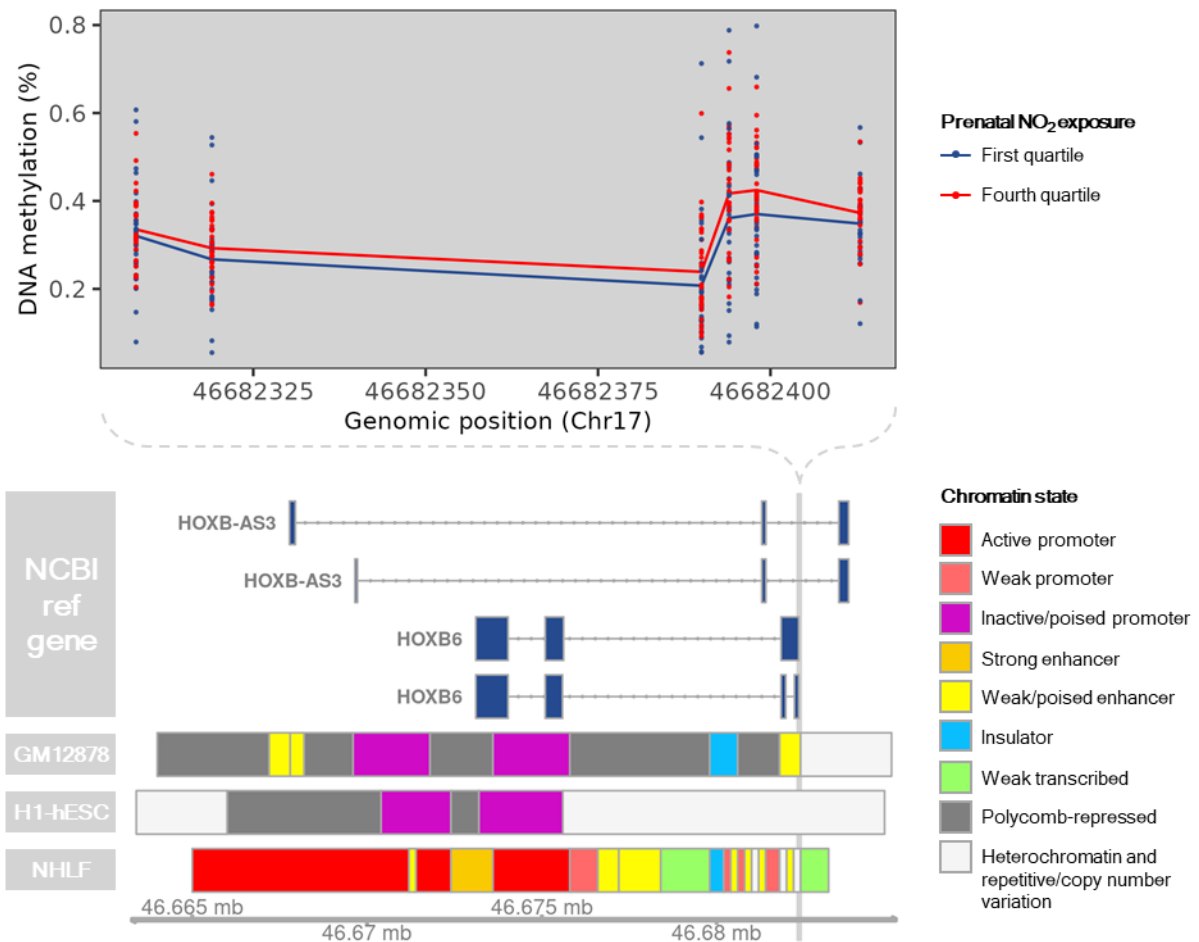
71
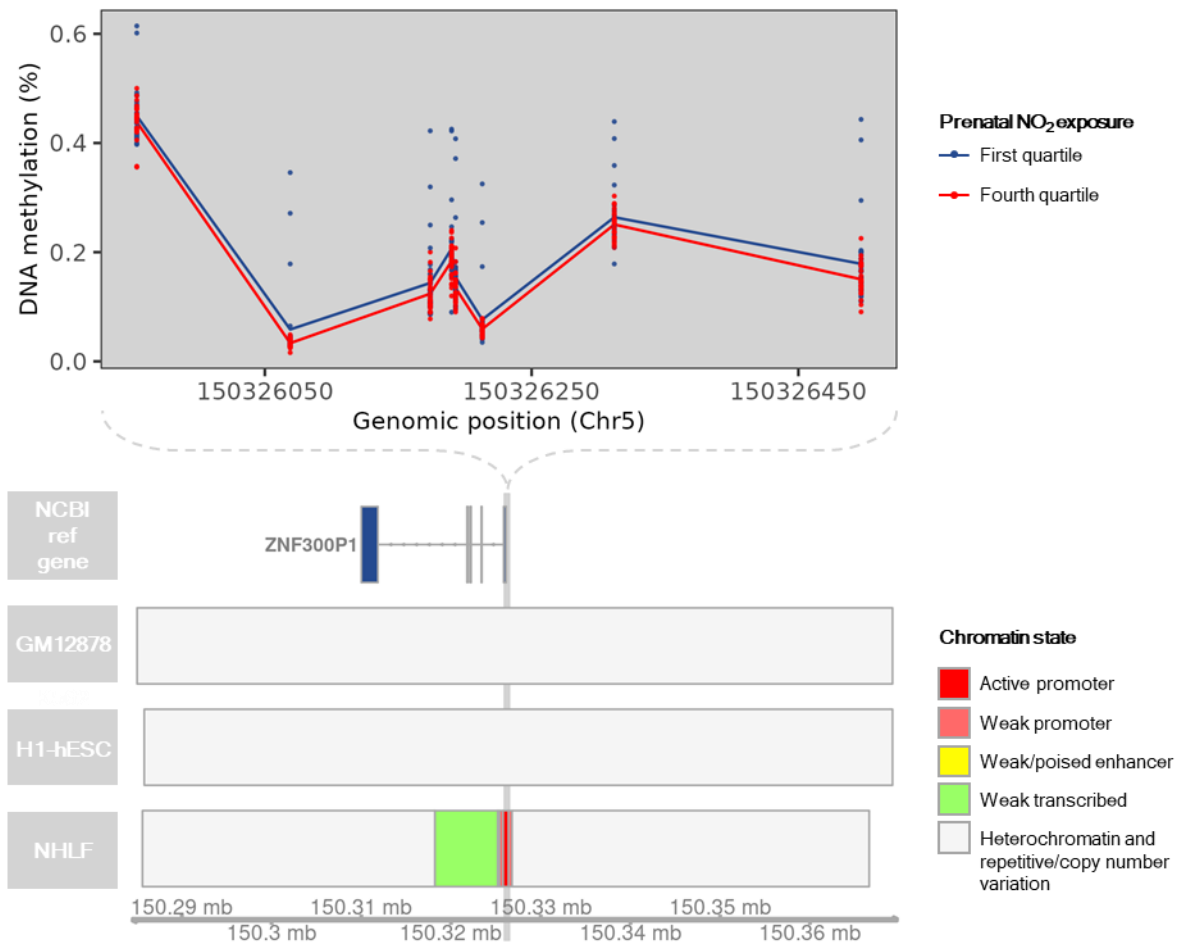
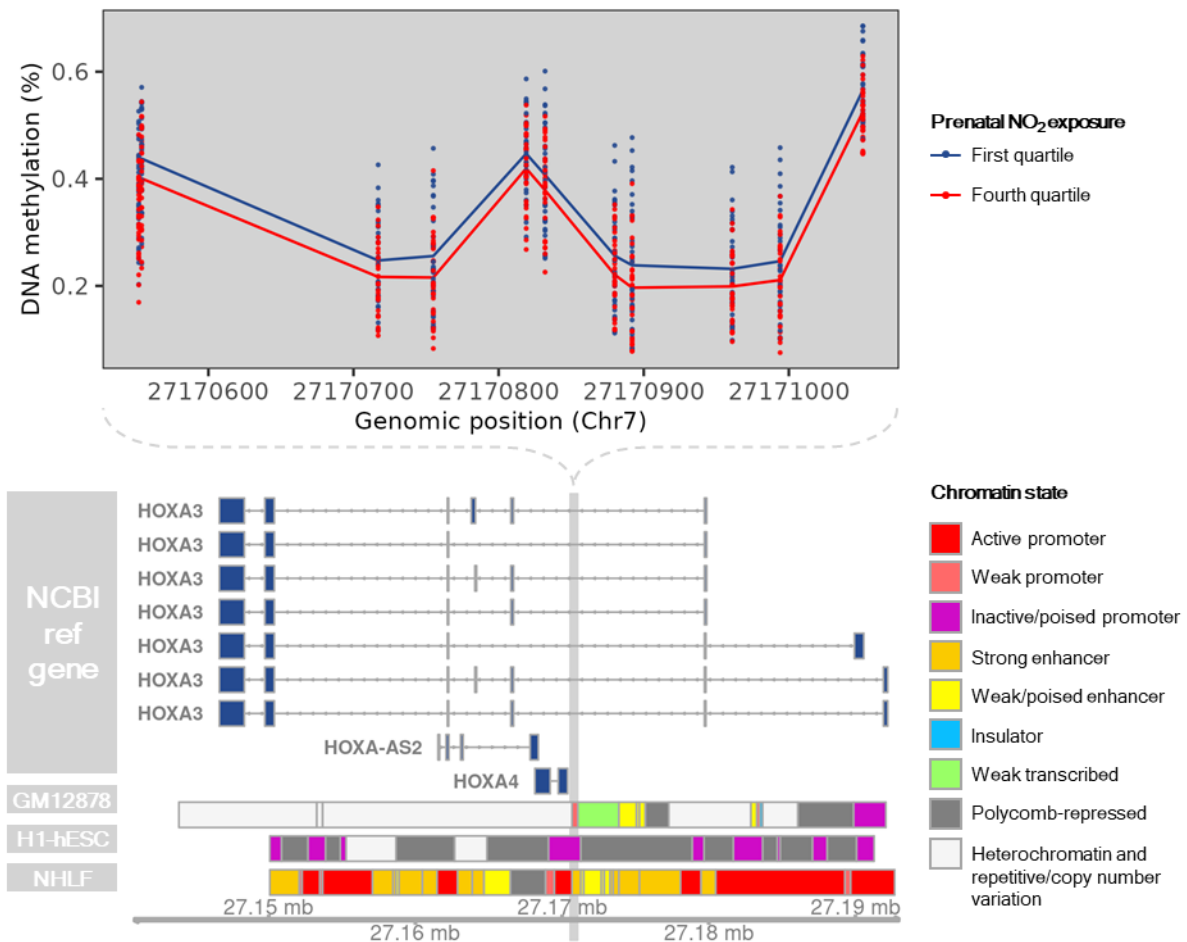**Supplementary Figure 17. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr5:150325954-150326498.** DNA methylation at CpGs contained within Chr5:150325954-150326498 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
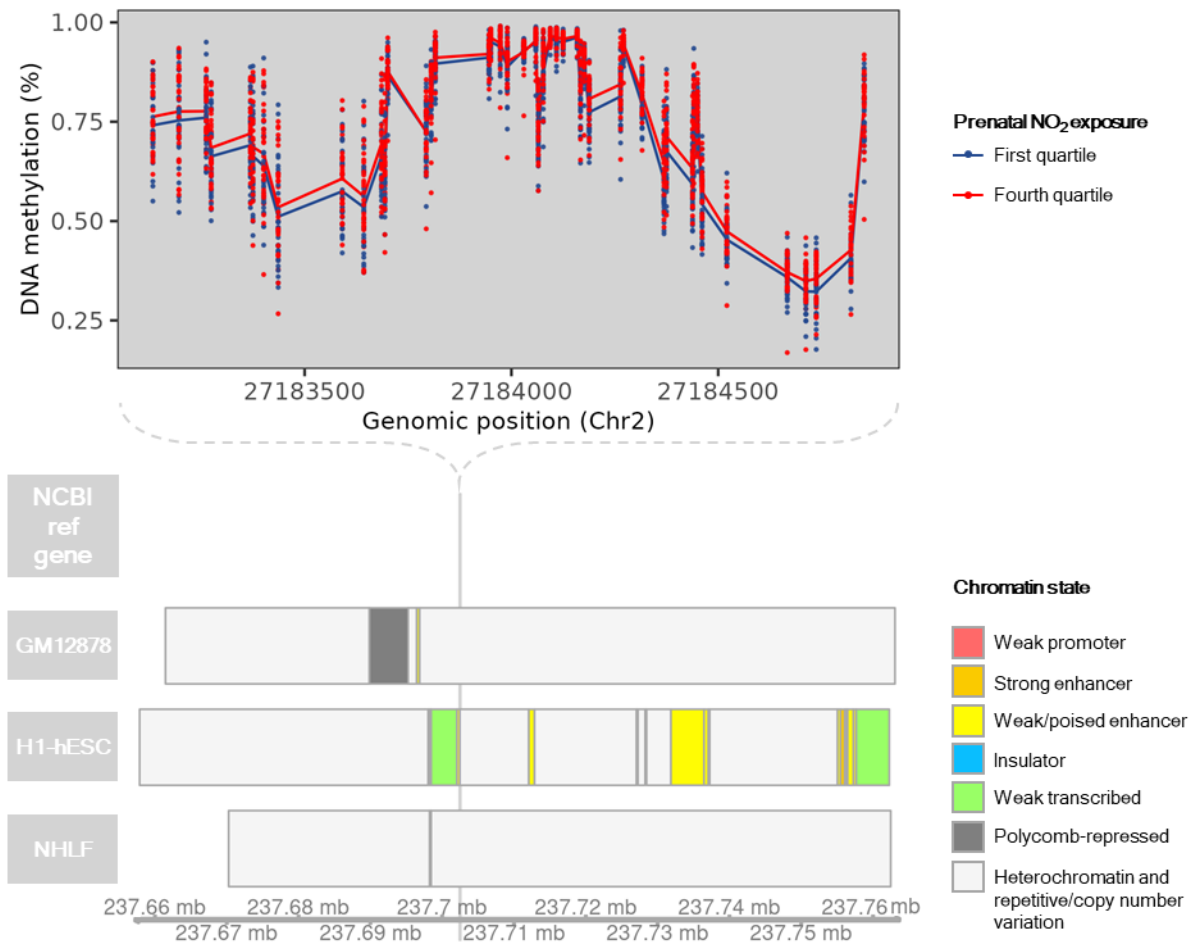
1486
1487
Supplementary Figure 18. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr2:237702390-237702581. DNA methylation at CpGs contained within Chr2:237702390-237702581 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
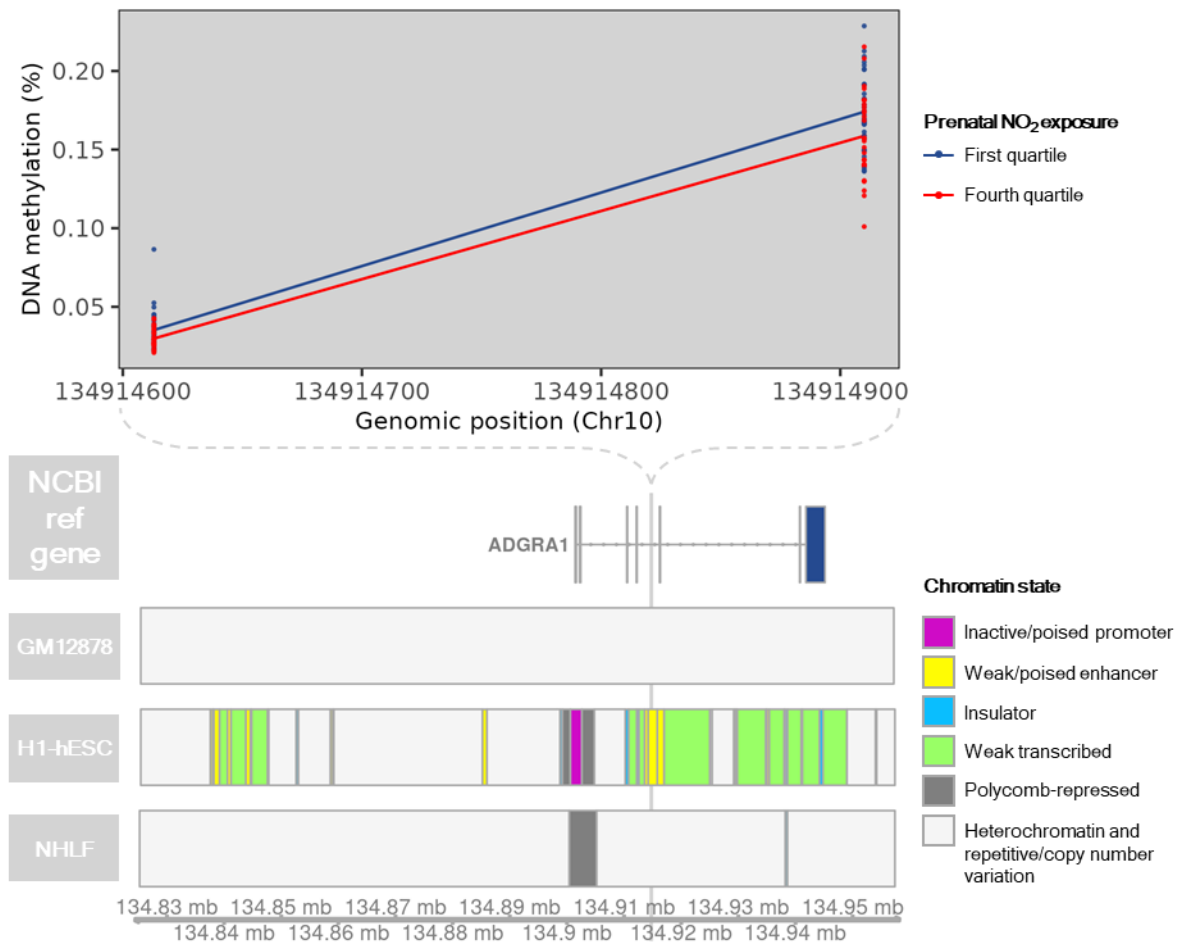
1498
1499 **Supplementary Figure 19. DNA methylation and chromatin state across the cord**
1500 **blood differentially methylated region (DMR) located on Chr7:26897182-26897254.**
1501 DNA methylation at CpGs contained within Chr7:26897182-26897254 is shown for
1502 individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal
1503 NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known
1504 genes and chromatin state were obtained from the University of California Santa Cruz
1505 using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878,
1506 H1-hESC, and NHLF cell lines are displayed based on their relevance to blood
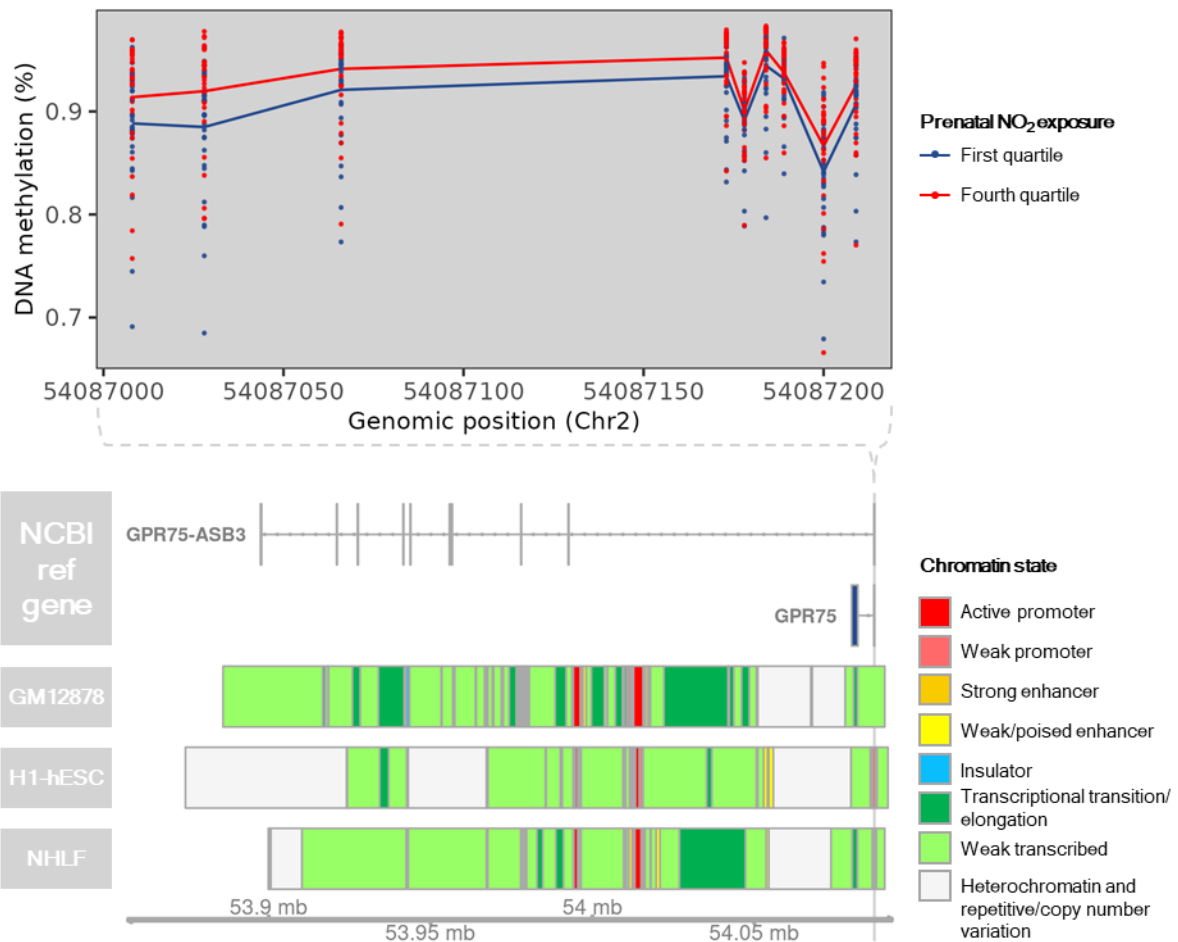1507 monocytes, prenatal development, and lung health and function, respectively.

74

Supplementary Figure 20. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr1:248100345-248100615. DNA methylation at CpGs contained within Chr1:248100345-248100615 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

75

Supplementary Figure 21. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr17:46682308-46682414.
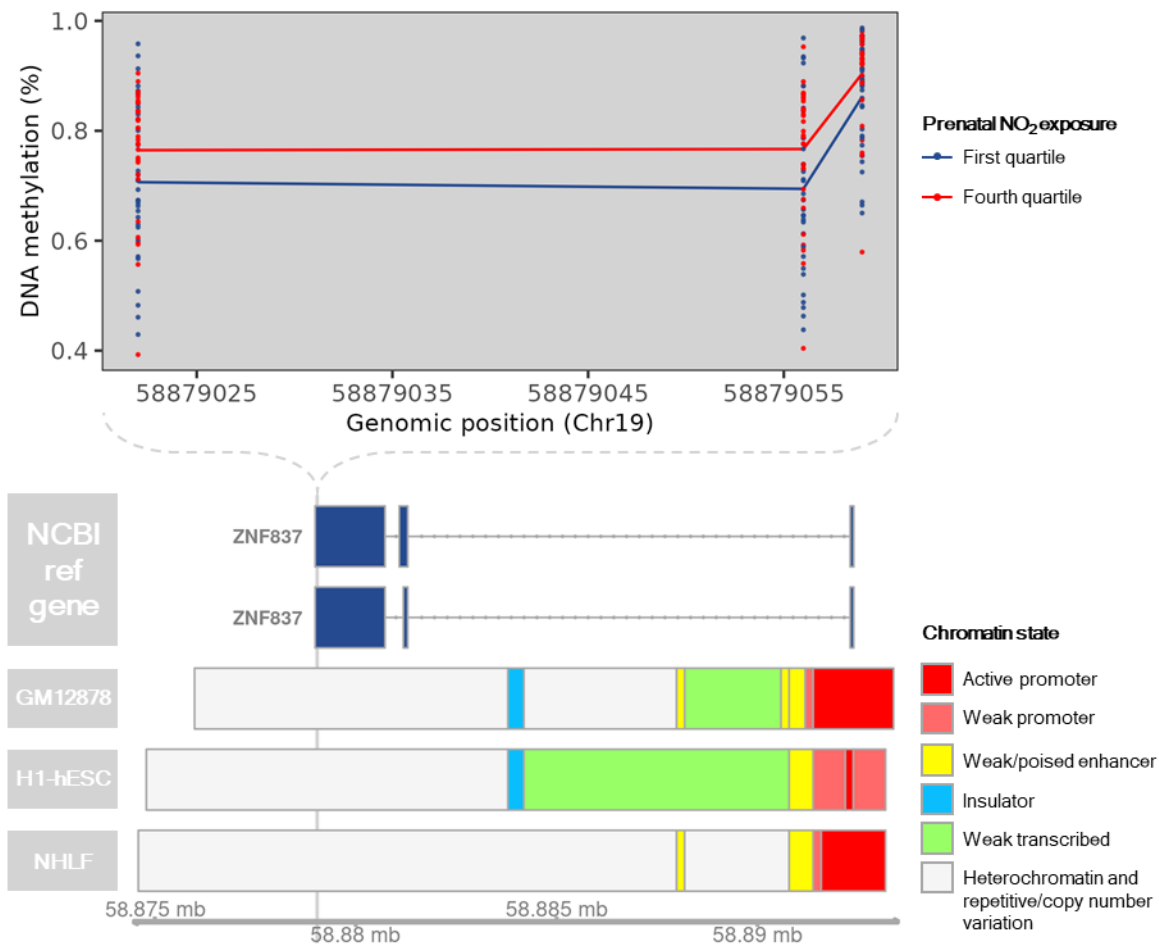DNA methylation at CpGs contained within Chr17:46682308-46682414 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
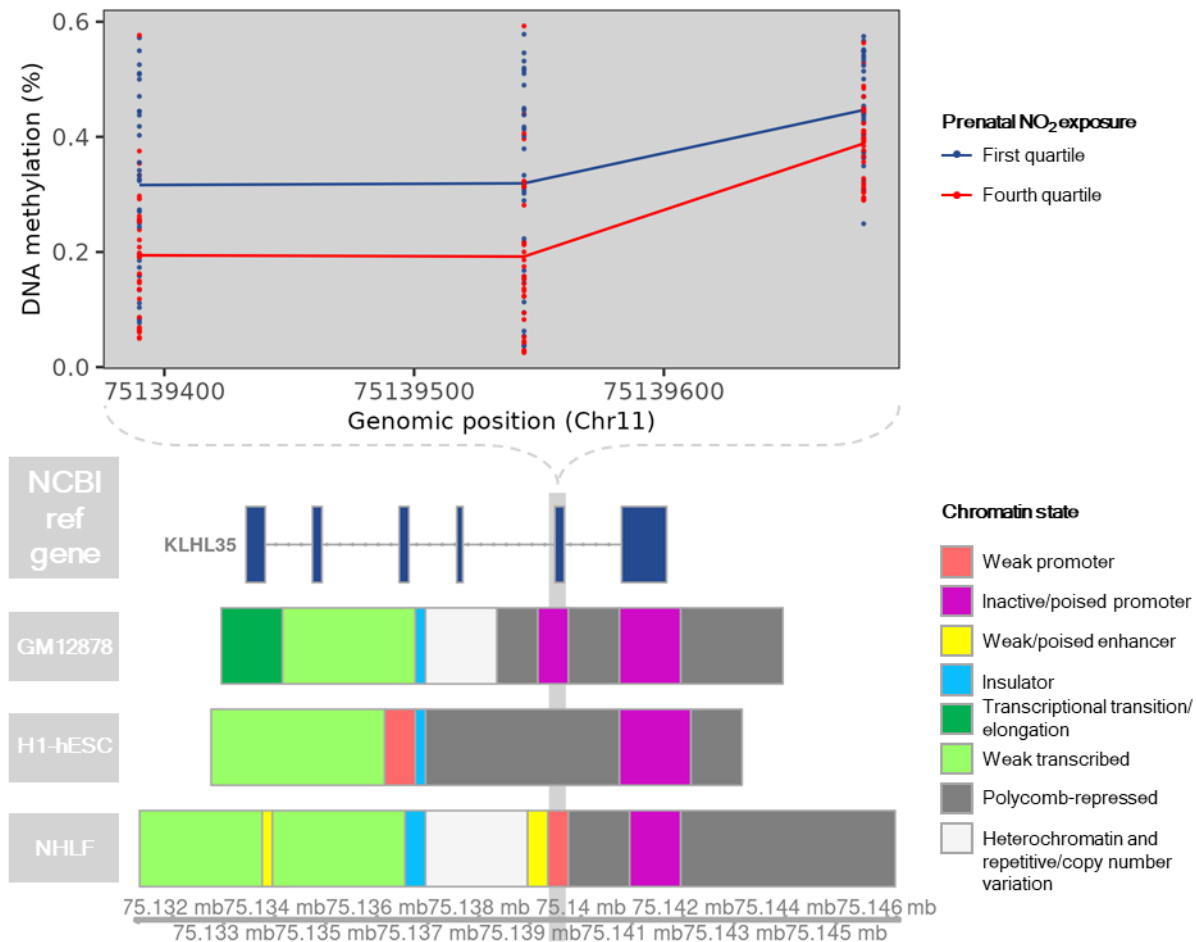
1530
1531
Supplementary Figure 22. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr7:27170552-27171052.
DNA methylation at CpGs contained within Chr7:27170552-27171052 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

77

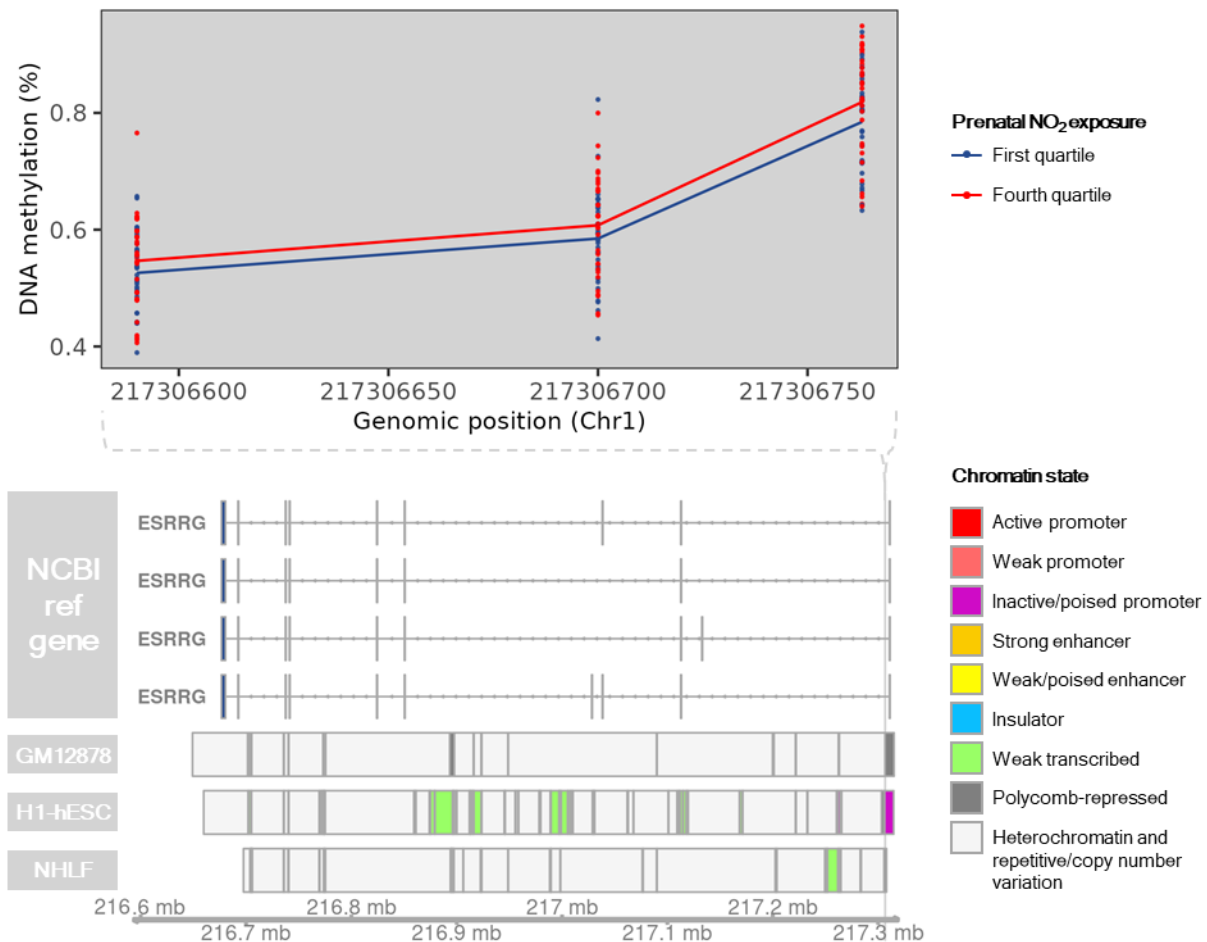**Supplementary Figure 23. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr10:134914613-134914911.** DNA methylation at CpGs contained within Chr10:134914613-134914911 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively. *ADGRA1* is also known as *GPR123*.

1553
1554
Supplementary Figure 24. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr2:54087008-54087210.
DNA methylation at CpGs contained within Chr2:54087008-54087210 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
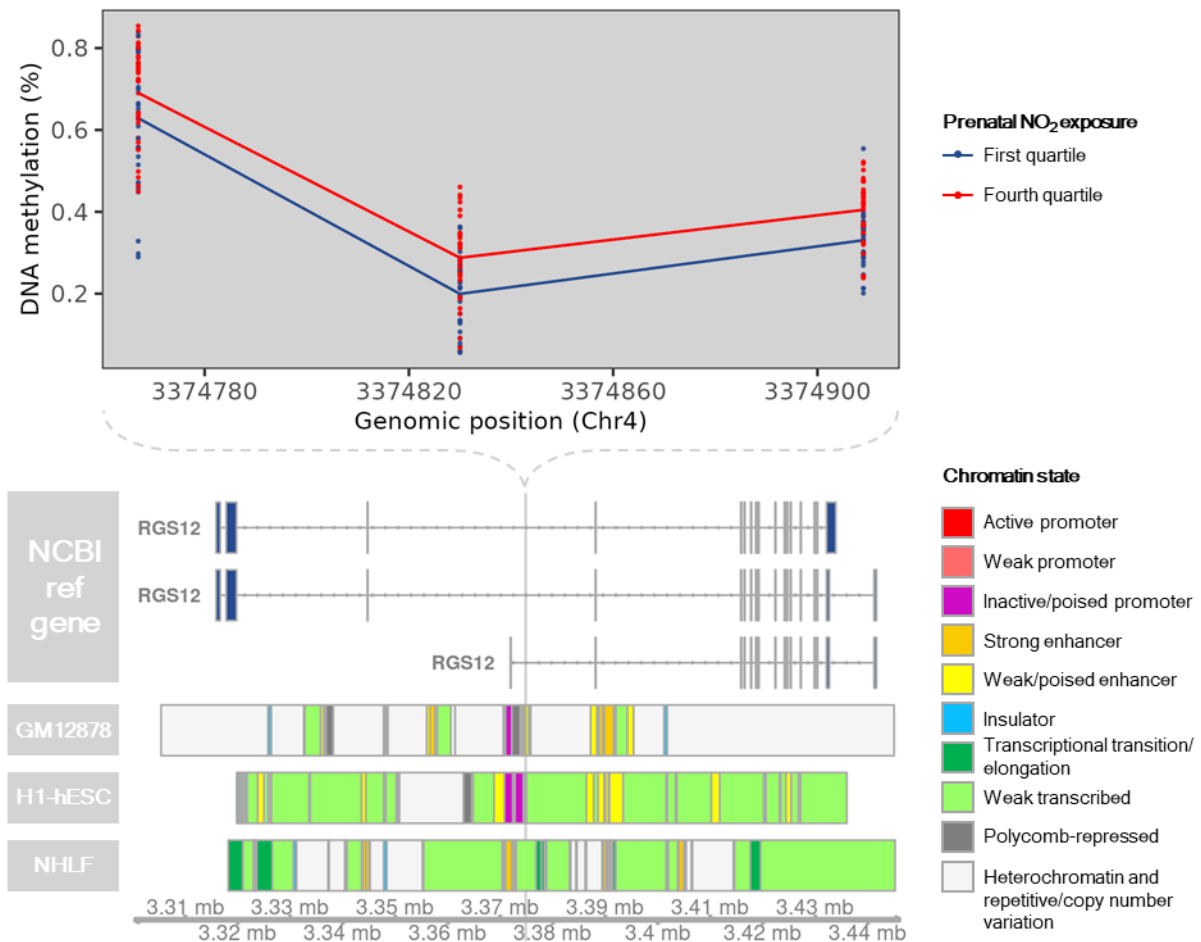
1564
1565
Supplementary Figure 25. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr19:58879022-58879060.
DNA methylation at CpGs contained within Chr19:58879022-58879060 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

80

Supplementary Figure 26. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr11:75139390-75139681.
DNA methylation at CpGs contained within Chr11:75139390-75139681 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
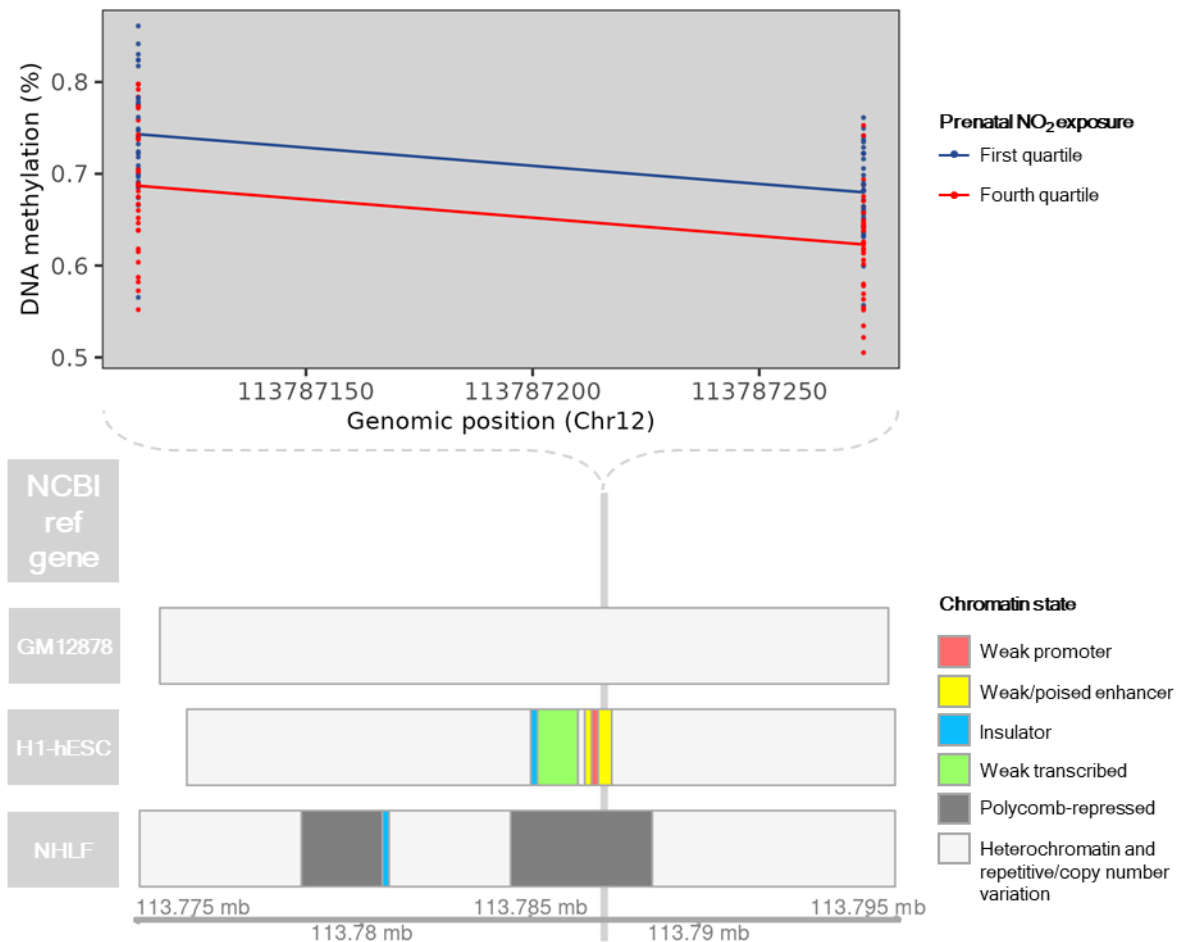
81

Supplementary Figure 27. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr1:217306590-217306764. DNA methylation at CpGs contained within Chr1:217306590-217306764 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
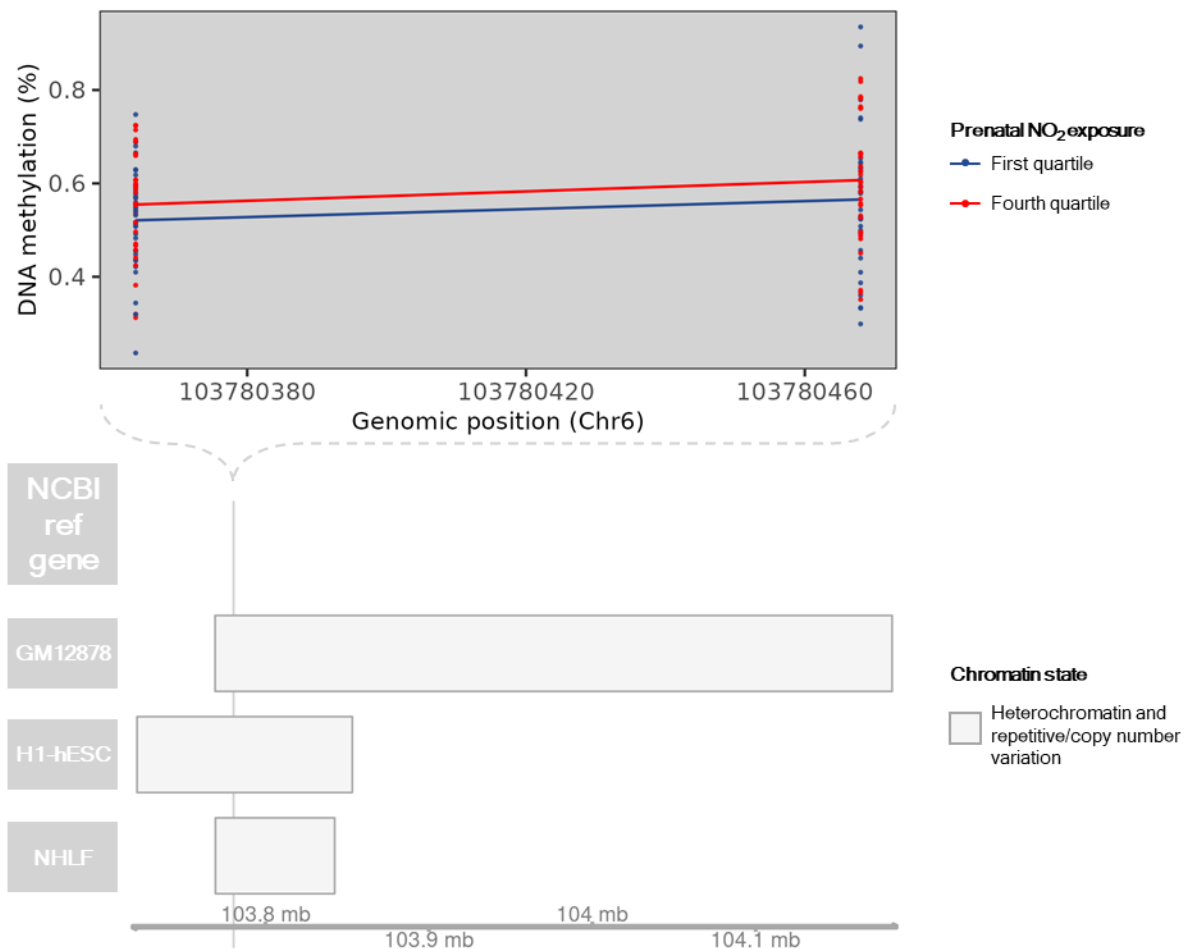
1598
1599
Supplementary Figure 28. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr4:3374767-3374910.
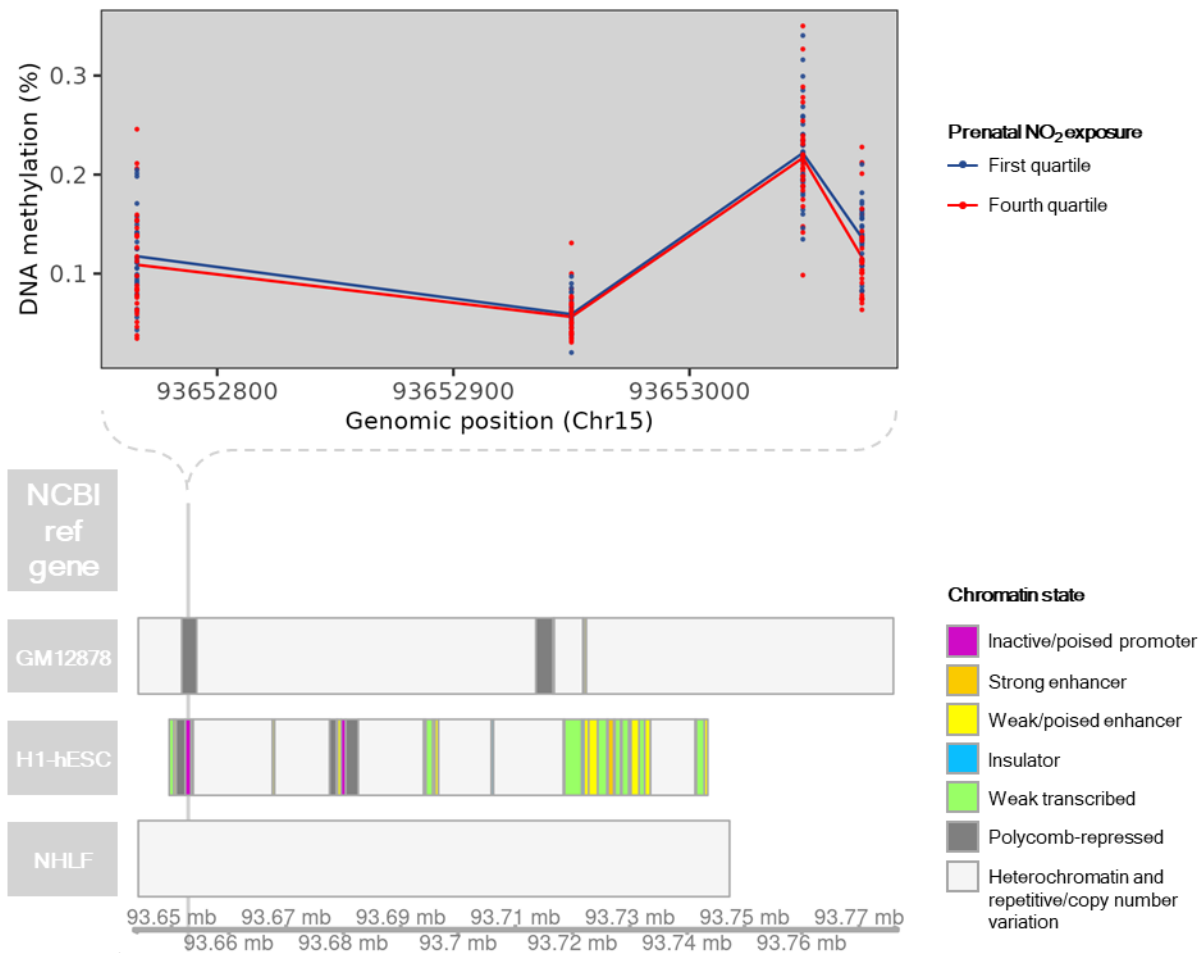DNA methylation at CpGs contained within Chr4:3374767-3374910 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
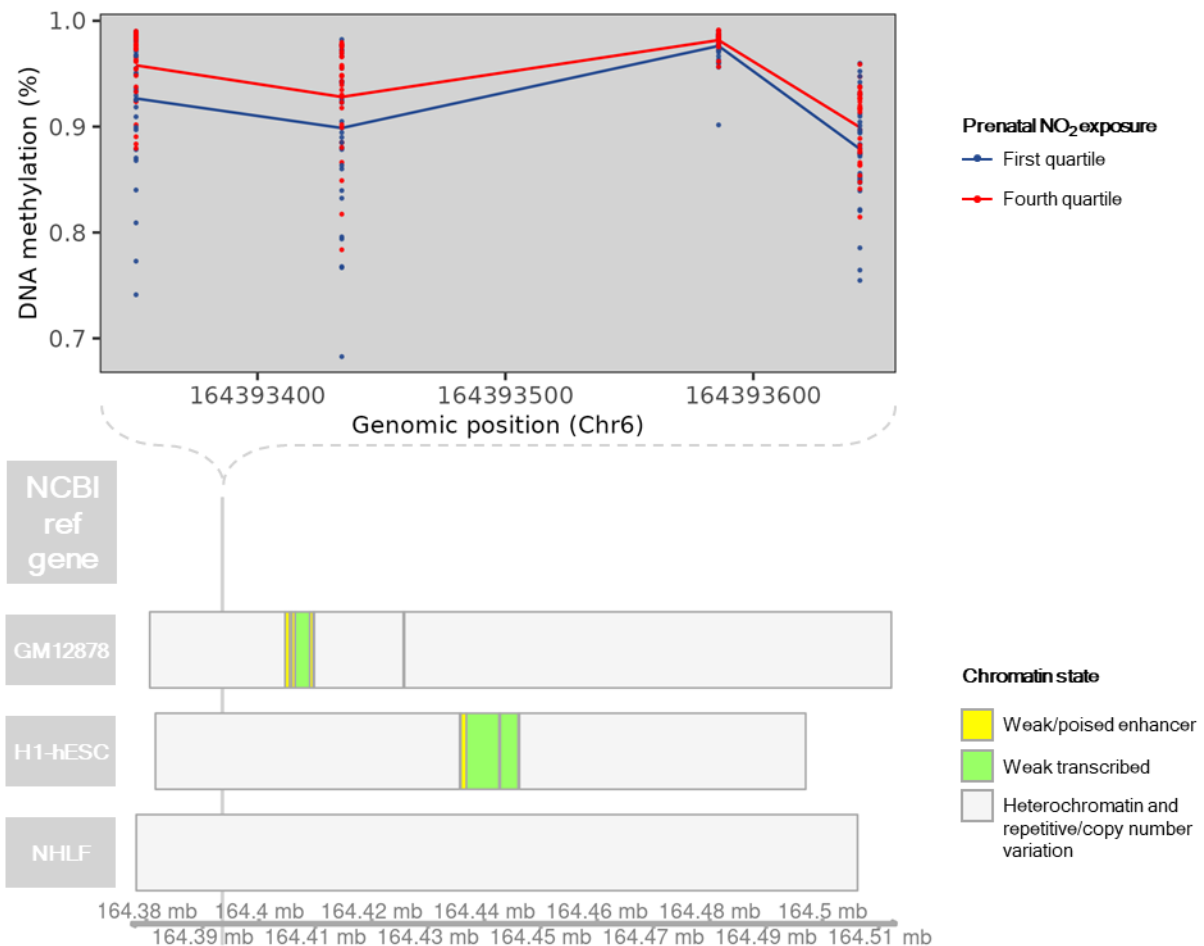
Supplementary Figure 29. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr12:113787113-113787274. DNA methylation at CpGs contained within Chr12:113787113-113787274 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
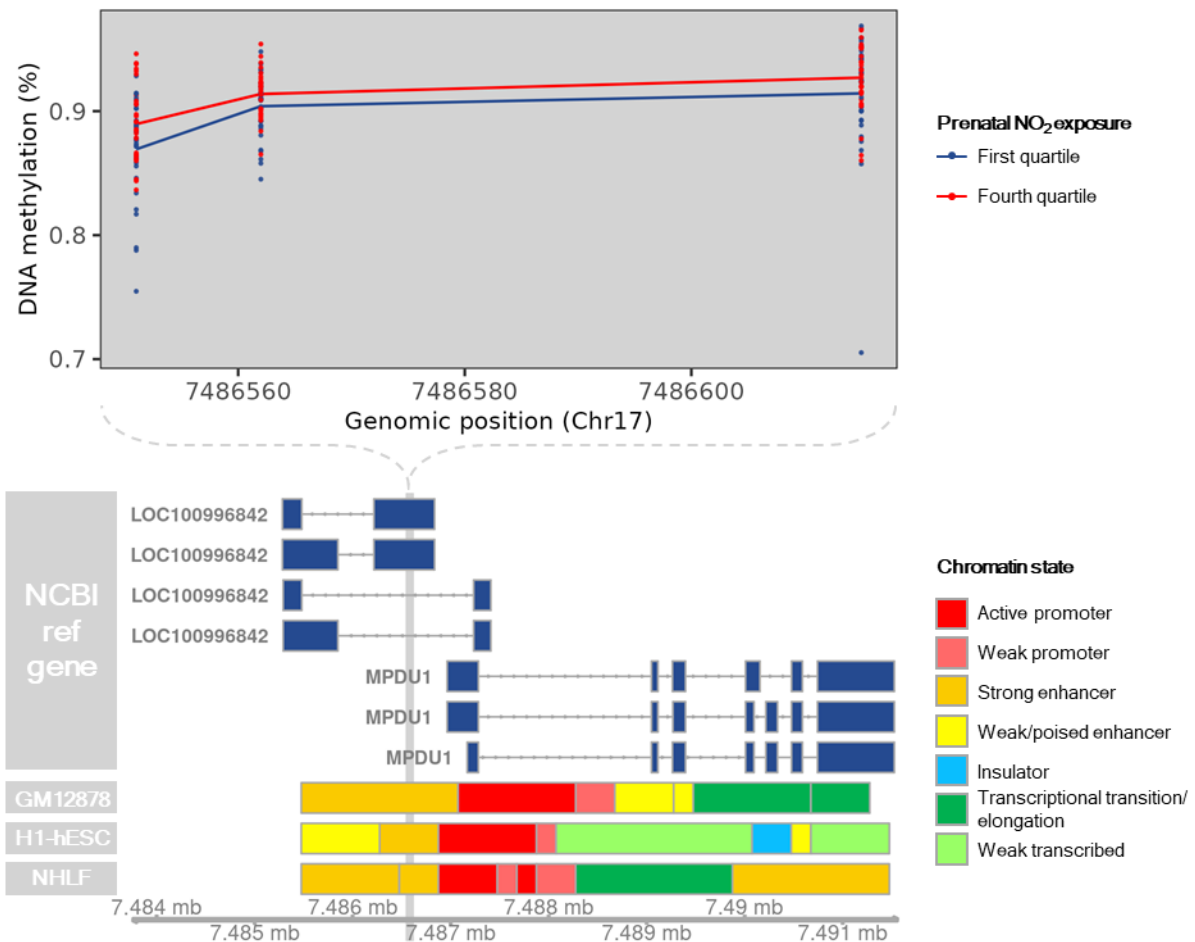
1621
1622
Supplementary Figure 30. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr6:103780364-103780469. DNA methylation at CpGs contained within Chr6:103780364-103780469 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

Supplementary Figure 31. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr15:93652766-93653074. DNA methylation at CpGs contained within Chr15:93652766-93653074 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
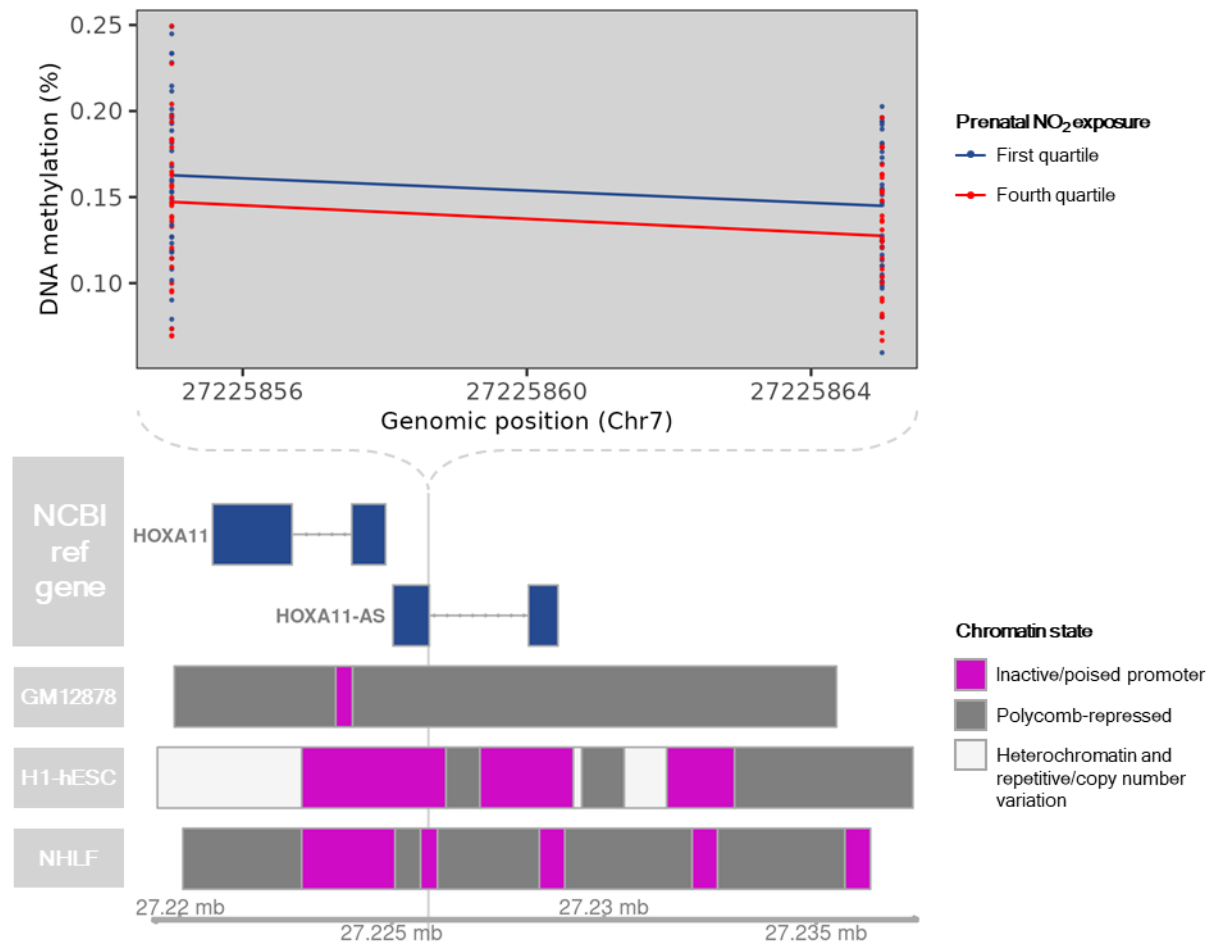
86

Supplementary Figure 32. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr6:164393351-164393644. DNA methylation at CpGs contained within Chr6:164393351-164393644 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
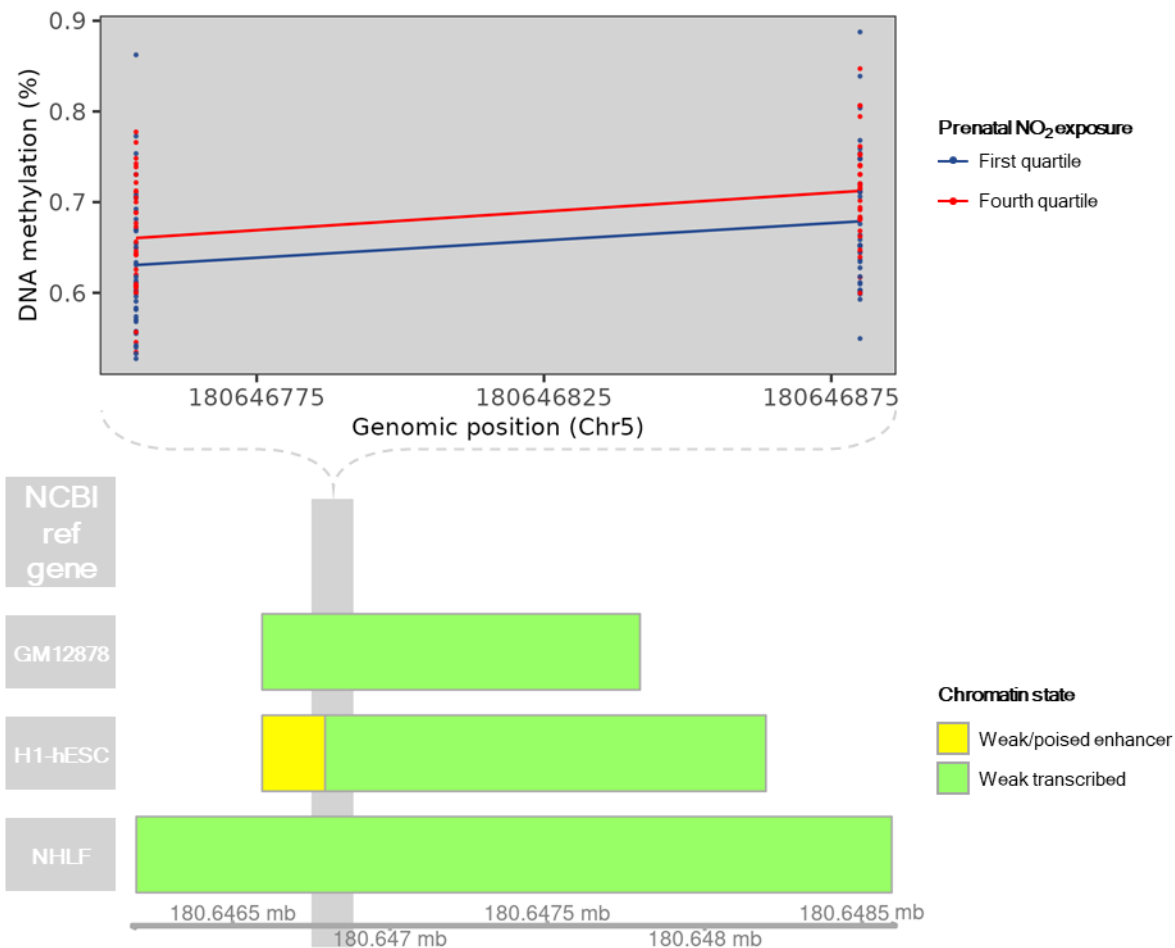
Supplementary Figure 33. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr17:7486551-7486616. DNA methylation at CpGs contained within Chr17:7486551-7486616 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
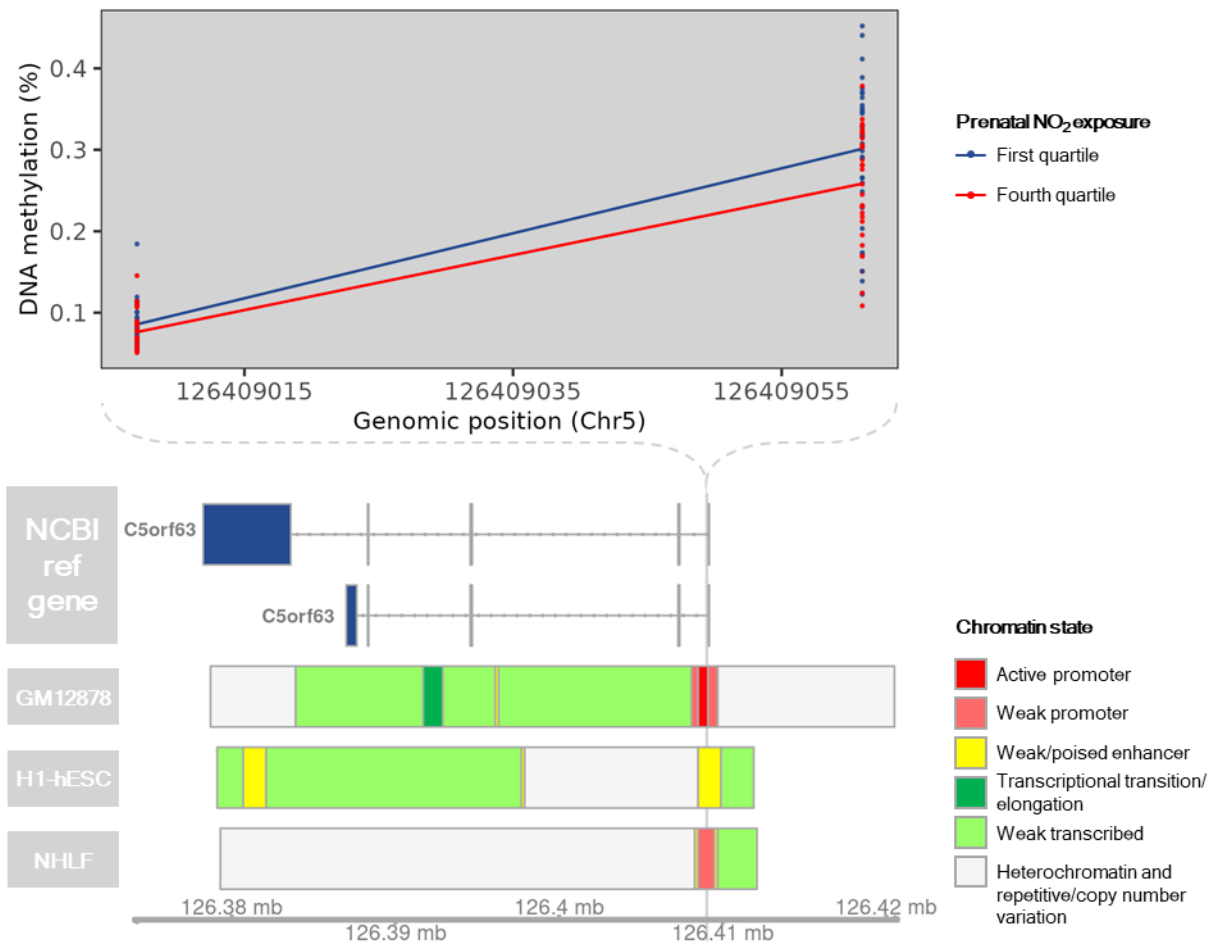
88

Supplementary Figure 34. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr7:27225855-27225866. DNA methylation at CpGs contained within Chr7:27225855-27225866 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

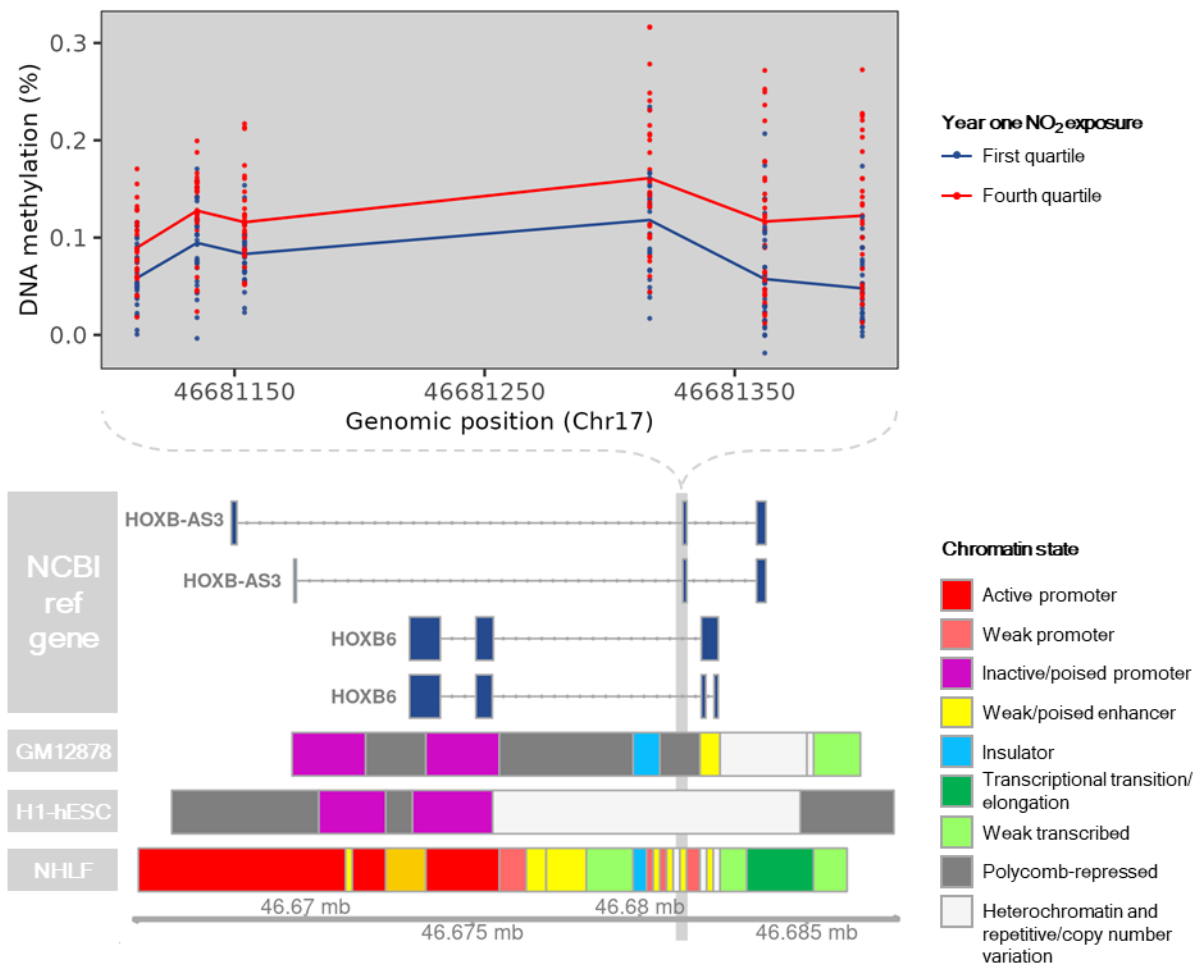**Supplementary Figure 35. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr5:180646754-180646881.** DNA methylation at CpGs contained within Chr5:180646754-180646881 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.

90

Supplementary Figure 36. DNA methylation and chromatin state across the cord blood differentially methylated region (DMR) located on Chr5:126409007-126409062. DNA methylation at CpGs contained within Chr5:126409007-126409062 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
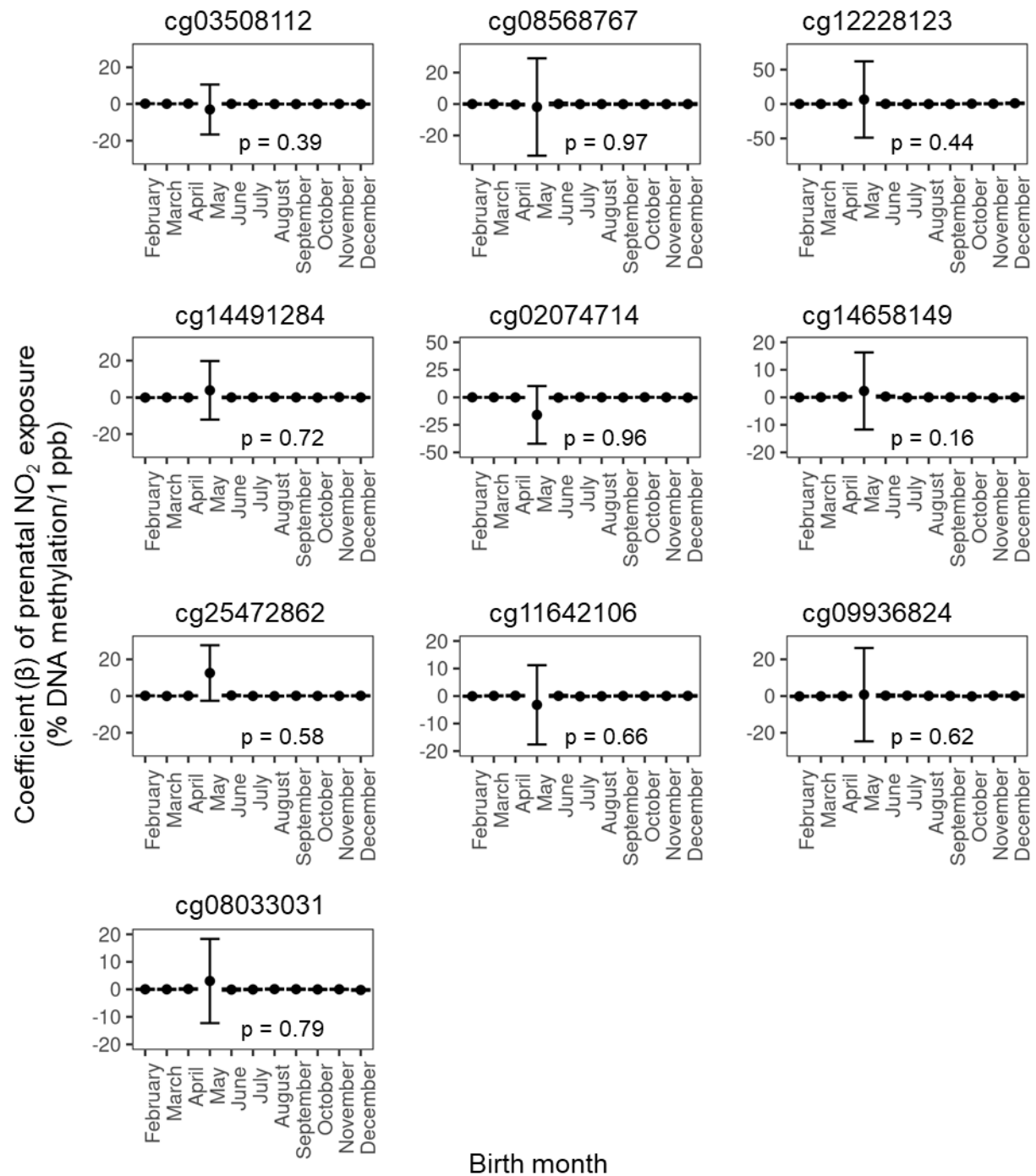
Supplementary Figure 37. DNA methylation and chromatin state across the postnatal-specific differentially methylated region (DMR) located on Chr17:46681111-46681402. DNA methylation at CpGs contained within Chr17:46681111-46681402 is shown for individuals with prenatal NO_2 exposure < 7.3 ppb (first quartile; N=32; blue) or prenatal NO_2 exposure > 16.3 ppb (fourth quartile; N=32; red) is displayed. Annotation of known genes and chromatin state were obtained from the University of California Santa Cruz using the *UcscTrack()* function from the *Gviz* R package. Chromatin states of GM12878, H1-hESC, and NHLF cell lines are displayed based on their relevance to blood monocytes, prenatal development, and lung health and function, respectively.
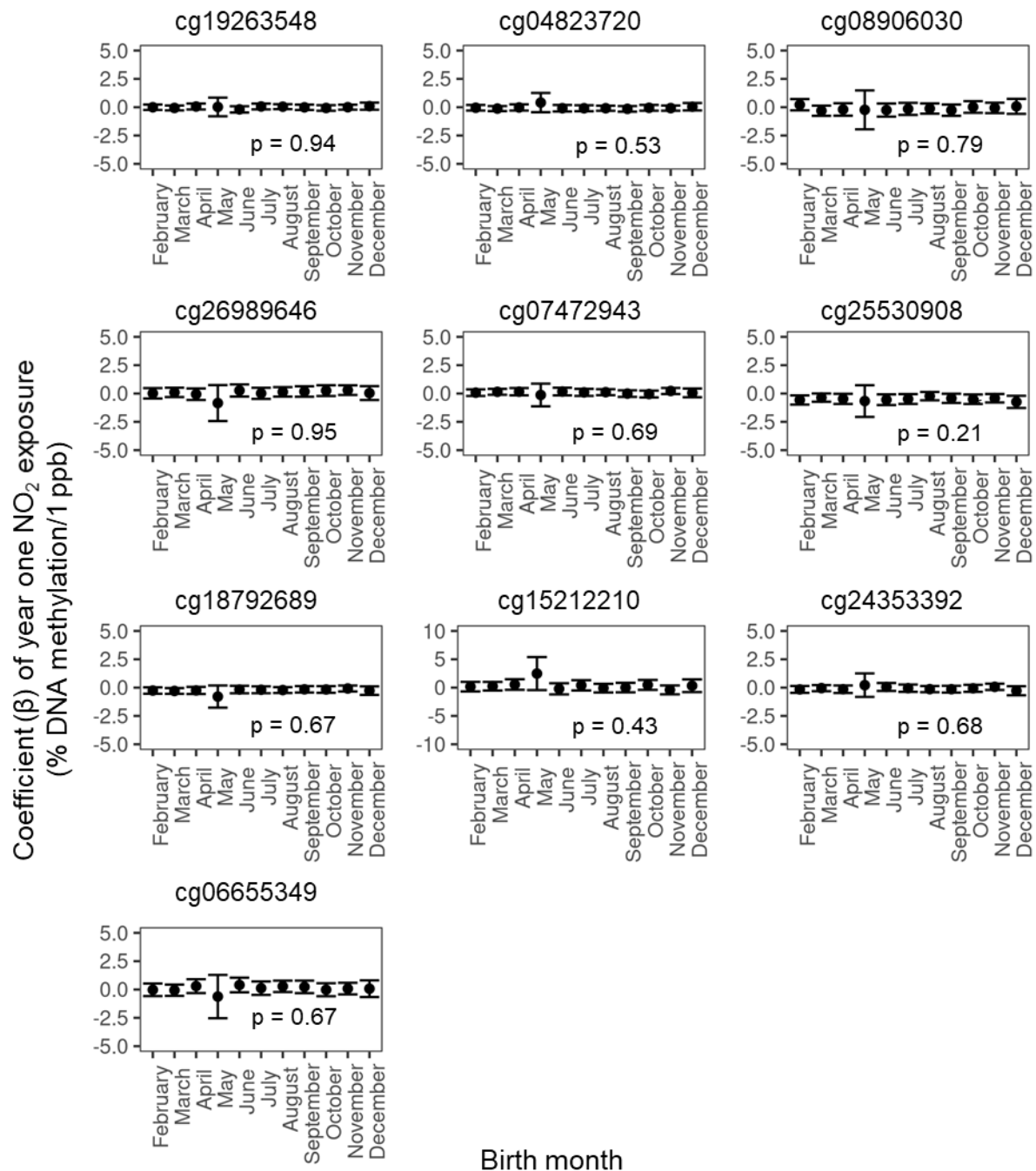
1712
1713
1714
1715
1716
1717
1718
1719
1720

1721 **Supplementary Figure 38. The effect of prenatal NO$_2$ exposure on cord blood DNA**
1722 **methylation (DNAm) does not vary by birth month at the top 10 CpGs identified in**
1723 **the epigenome- wide analysis of cord blood (N=128).** Based on ANOVA likelihood
1724 ratio tests, modelling the effect of prenatal NO$_2$ exposure by birth month did not
1725 significantly improve model fit compared to models that did not include birth month. In
1726 all CpGs investigated, the month of May exhibited greater variation in the coefficient of
1727 prenatal NO$_2$ exposure. This is likely spurious and related to a relatively small (N=3)
1728 number of CHILD participants born in May. The mean number of participants born in
1729 each month is 12 with a standard deviation of ±5.
1730
1731

1732
1733
1734
1735
1736
1737
1738
1739
1740

Supplementary Figure 39. The effect of year one NO$_2$ exposure on postnatal-specific DNA methylation (DNAm) does not vary by birth month at the top 10 CpGs identified in the epigenome-wide analysis of postnatal-specific DNAm changes.
Based on ANOVA likelihood ratio tests, modelling the effect of year one NO$_2$ exposure by birth month did not significantly improve model fit compared to models that did not include birth month. In all CpGs investigated, the month of May exhibited greater variation in the coefficient of year one NO$_2$ exposure. This is likely spurious and related to a relatively small (N=3) number of CHILD participants born in May. The mean number of participants born in each month is 12 with a standard deviation of 5.

1784 **Supplementary Tables**

1785

1786 **Supplementary Table 1. List of candidate genes investigated.**

Gene*	Reference	Rationale
LONP1, **SLC25A28**, **PLVAP**, **CAT**, **TPO**	Gruzieva, O. *et al*. (2017)	Significantly associated with prenatal NO2 exposure
RNF39, **CYP2E1**	Ladd-Acosta, C. *et al*. (2019)	Significantly associated with prenatal NO2 exposure
NFKB1	Lu, T. & Stark (2015) Bhatt, D. & Ghosh (2014)	Regulates pro-inflammatory genes
NRF2	Kang, K. A. & Hyun (2017) Janasik, B. *et al*. (2018)	Regulated by DNA methylation, regulates antioxidant gene expression
SOD1, **SOD2**, **GSTA1**, **GSTA2**, **GSTA3**, **GSTA5**, GSTM1, **GSTM2**, **GSTM3**, **GSTP1**, **GCLC**, **GCLM**, GSR1, GPX1, GPX2, CYP2C11, **CYP7B1**, TRX1, TRX2, **TXN**, **TXNRD1**, **SRXN1**, **MT3**, **SLC27A11**, **NQO1**, **NOS2**, NOX2, **HMOX1**, DDH1, **ALOX5**, ALOX12, G6PD, **FTH1**	Kairisalo, M., Korhonen, L., Blomgren, K. & Lindholm (2007) Morgan, M. J. & Liu, Z (2011) Tonelli, C., Chio, I. I. C. & Tuveson, D. A (2018) Wittkopp, S. *et al*. (2016)	Genes regulated by NRF2
KEAP1	Muscarella, L. A. *et al*. (2011) Janasik, B. *et al*. (2018)	Regulates NRF2
AHR, **ARNT**	Raghunath, A. *et al*. (2018) Klingbeil, E. C., Hew, K. M., Nygaard, U. C. & Nadeau, K. C (2014)	Regulate expression of detoxifying genes
CYP1A1	del Rosario Olguin, S., Camacho, R. & Espinosa, J. J (2013) Wittkopp, S. *et al*. (2016)	Major phase I detoxification enzyme
IL-1, **IL-2**, **IL-3**, **IL-4**, **IL-5**, **IL-6**, **IL-9**, **IL-10**, **IL-12**, **IL-13**, **IL-15**, **IL-17**, **IL-18**, **IL-19**, **IL-21**, **IL-22**, **IL-23**, **IL-25**, **IL-27**, **IL-31**, **IL-33**, **IFNG**, **TNF**, **TGFB**, FOXP3	Murrison, L. B., Brandt, E. B., Myers, J. B. & Hershey, G. K. K (2019) Kim, Y.-M., Kim, Y.-S., Jeon, S. G. & Kim, Y.-K (2013) Bosnjak, B., Stelzmueller, B., Erb, K. J. & Epstein, M. M (2011) Pahl, H. L (1998) Lee, K. W. K. *et al*. (2015) Christensen, S. *et al*. (2017)	Genes involved in TH2 childhood allergic asthma, cytokines related to maturation of T-cells, cytokines related to eosinophilic response and inflammation

1787   ***Bolded genes are annotated on the Illumina HumanMethylation 450k array and**
1788   **were investigated as part of the candidate gene analysis.**

1789 **Supplementary Table 2. Genomic information of cord blood differentially methylated regions (N=24) associated with prenatal**

1790 **NO$_2$ exposure.**

Chr	Start	End	P	FDR	Sidak	Probes	Number of probes that are potential mQTLs	UCSC gene name	UCSC gene group	Mean prenatal NO$_2$ effect size at birth (EWAS)	Mean prenatal NO$_2$ effect size at birth (LME)	Mean prenatal NO$_2$ effect size at age one (LME)	Age one wheeze MANOVA p-value	Adjusted age one wheeze MANOVA p-value	Age one atopy MANOVA p-value	Adjusted age one atopy MANOVA p-value
7	27183133	27184854	1.7E-20	5.3E-19	0	46	13	HOXA5; HOXA-AS3	TSS1500, TSS200, 5'UTR, 1stExon; Body	0.066	0.038	0.027	0.060	1	0. 20	1
7	27187102	27187692	9.9E-16	1.5E-14	7.2E-13	10	0	HOXA6	TSS1500, TSS200, 5'UTR	0.052	0.033	0.022	0.18	1	0.72	1
8	22132665	22133077	4.4E-10	4.5E-09	4.5E-07	10	7	PIWIL2	TSS1500, TSS200, 5'UTR,	-0.084	-0.063	-0.056	0.11	1	0.10	1

								1stExon							
17	36666953	36667485	2.9E-08	2.3E-07	2.3E-05	3	2	ARHGAP23	Body,3'UTR	0.059	0.054	0.044	0.48	1	0.27	1
5	150325954	150326498	4.6E-08	2.8E-07	3.6E-05	8	0	LOC134466	TSS1500, TSS200, 5'UTR	-0.046	-0.042	-0.041	0.11	1	0.55	1
2	237702390	237702581	5.4E-08	2.8E-07	1.2E-04	3	2	Unannotated		-0.085	-0.077	-0.072	0.83	1	0.064	1
7	26897182	26897254	9.1E-08	4.0E-07	5.4E-04	3	0	SKAP2	Body	0.049	0.036	0.036	0.54	1	0.11	1
1	248100345	248100615	1.4E-07	5.6E-07	2.3E-04	7	7	OR2L13	TSS200, 5'UTR, 1stExon	-0.18	-0.15	-0.15	0.58	1	0.15	1
17	46682308	46682414	2.2E-07	7.4E-07	8.6E-04	6	0	LOC404266; HOXB6	5'UTR, 1stExon, Body; TSS200, 5'UTR,	0.11	0.073	0.11	0.34	1	0.0058	0.14

100

									1stExon							
7	27170552	27171052	3.6E-07	1.1E-06	3.1E-04	11	10	HOXA4	TSS1500, TSS200	-0.072	-0.047	-0.030	0.17	1	0.66	1
10	134914613	134914911	3.9E-07	1.1E-06	5.6E-04	2	0	GPR123	Body	-0.026	-0.022	-0.011	0.76	1	0.025	0.60
2	54087008	54087210	4.4E-07	1.1E-06	9.2E-04	9	8	LOC100302652; GPR75	TSS200, 1stExon; 5'UTR, 1stExon	0.039	0.030	0.021	0.31	1	0.0045	0.11
19	58879022	58879060	7.1E-07	1.7E-06	7.9E-03	3	0	ZNF837	3'UTR	0.11	0.11	0.078	0.25	1	5.5E-05	0.0013
11	75139390	75139681	1.2E-06	2.6E-06	1.7E-03	3	0	KLHL35	Body	-0.17	-0.14	-0.13	0.16	1	0.24	1
1	217306590	217306764	1.9E-06	4.0E-06	4.7E-03	3	2	ESRRG	5'UTR	0.11	0.081	0.061	0.46	1	0.018	0.43

4	3374767	3374910	2.1E-06	4.0E-06	6.1E-03	3	2	RGS12	Body	0.12	0.076	0.11	0.043	1	0.0063	0.15
12	113787113	113787274	2.5E-06	4.3E-06	6.5E-03	2	0	Unannotated		-0.095	-0.10	-0.086	0.36	1	0.66	1
6	103780364	103780469	3.5E-06	5.6E-06	1.4E-02	2	2	Unannotated		0.17	0.061	0.066	0.78	1	0.25	1
15	93652766	93653074	6.6E-06	1.0E-05	9.1E-03	4	0	Unannotated		-0.054	-0.039	-0.038	0.19	1	0.60	1
6	164393351	164393644	9.4E-06	1.4E-05	1.4E-02	4	0	Unannotated		0.040	0.034	0.034	0.024	0.58	0.11	1
17	7486551	7486616	6.4E-04	8.7E-04	9.9E-01	3	0	MPDU1	TSS1500	0.025	0.015	0.013	0.0011	0.027	0.016	0.394903
7	27225855	27225866	8.0E-04	1.0E-03	1	2	0	HOXA11AS; HOXA11	Body; TSS1500	-0.045	-0.037	-0.039	0.28	1	0.30	1

| 5 | 180646754 | 180646881 | 1.7E-03 | 2.1E-03 | 1 | 2 | 0 | Unannotated | | 0.048 | 0.049 | 0.012 | 0.12 | 1 | 0.0045 | 0.11 |
| 5 | 126409007 | 126409062 | 2.6E-03 | 3.0E-03 | 1 | 2 | 0 | FLJ44606 (C5orf63) | 5'UTR | -0.052 | -0.046 | -0.028 | 9.1E-05 | 0.0022 | 0.41 | 1 |

1791