

1

2 MS RUTU PATEL (Orcid ID : 0000-0001-9808-8909)

3 DR HONGMEI ZHANG (Orcid ID : 0000-0003-3557-0364)

4

5

6 Article type : Original Article-Epidemiology of Allergic Disease

7

8

9 Sex-specific associations of asthma acquisition with changes in DNA methylation during adolescence

10 Running title: Asthma acquisition and DNA methylation changes

11 Rutu Patel, MPH, MBBS^{1*}, Farnaz Solatikia, MPH, MS^{1,2*}, Hongmei Zhang, PhD, MS^{1**},

12 Alemayehu Wolde, MS², Latha Kadalayil PhD³, Wilfried Karmaus, MD, Dr.med., MPH¹, Susan

13 Ewart, DVM, PhD, DACVIM⁴, Ryan Arathimos, PhD^{5,6,7,8}, Caroline Relton, PhD^{5,6,7}, Susan Ring,

14 PhD, BSc^{5,6,7}, A. John Henderson⁶, MD, S. Hasan Arshad, MBBS, DM, FRCP^{9,10,11}, John W

15 Holloway, PhD^{3,11}

16

17 ¹Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health,

18 University of Memphis, Memphis, TN, USA.

19 ²Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA.

20 ³Human Development and Health, Faculty of Medicine, University of Southampton, Southampton,

21 UK

22 ⁴College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA.

23 ⁵MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

24 ⁶Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

25 ⁷National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol and

26 University Hospitals Bristol NHS Foundation Trust, Bristol, UK.

27 ⁸Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry Psychology and

28 Neuroscience, King's College London, London, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/CEA.13776](#)

This article is protected by copyright. All rights reserved

29 ⁹Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton,
30 UK.

31 ¹⁰David Hide Asthma and Allergy Research Centre, Isle of Wight, UK.

32 ¹¹NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton,
33 UK

34

35 *: equal contribution, **: corresponding author

36

37 Corresponding author: Dr. Hongmei Zhang

38 Phone: 901.678.4707

39 Email: hzhang6@memphis.edu

40 Fax:901.678.1715

41 Office: 224 Robison Hall, 3825 DeSoto Avenue, Memphis, TN 38152-0001

43 **Abstract (267 words)**

44 **Background-** Underlying biological mechanisms involved in sex differences in asthma status
45 changes from pre- to post-adolescence are unclear. DNA methylation (DNAm) has been shown to be
46 associated with the risk of asthma.

47 **Objective-** We hypothesized that asthma acquisition from pre- to post-adolescence was associated
48 with changes in DNAm during this period at asthma-associated cytosine-phosphate-guanine (CpG)
49 sites and such an association was sex-specific.

50 **Methods-** Subjects from the Isle of Wight birth cohort (IOWBC) with DNAm in blood at ages 10 and
51 18 years (n=124 females, 151 males) were studied. Using a training-testing approach, epigenome-
52 wide CpGs associated with asthma were identified. Logistic regression was used to examine sex-
53 specific associations of DNAm changes with asthma acquisition between ages 10 and 18 at asthma-
54 associated CpGs. The ALSPAC birth cohort was used for independent replication. To assess
55 functional relevance of identified CpGs, association of DNAm with gene expression in blood was
56 assessed.

57 **Results-** We identified 535 CpGs potentially associated with asthma. Significant interaction effects of
58 DNAm changes and sex on asthma acquisition in adolescence were found at 13 of the 535 CpGs in
59 IOWBC ($p\text{-values} < 1.0 \times 10^{-3}$). In the replication cohort, consistent interaction effects were observed at
60 10 of the 13 CpGs. At 7 of these 10 CpGs, opposite DNAm changes across adolescence were
61 observed between sexes in both cohorts. In both cohorts, cg20891917, located on *IFRD1* linked to
62 asthma, shows strong sex-specific effects on asthma transition ($p\text{-values} < 0.01$ in both cohorts).

63 **Conclusion-** Gender-reversal in asthma acquisition is associated with opposite changes in DNAm
64 (males vs females) from pre- to post-adolescence at asthma-associated CpGs. These CpGs are
65 potential biomarkers of sex-specific asthma acquisition in adolescence.

66
67 **Keywords:** asthma acquisition, ALSPAC, DNA methylation, IOWBC, sex-specificity

68 Total word count: 3640 words (Introduction-Discussion)

69 **Introduction (418 words)**

70 Asthma is a common chronic condition that affects approximately 339 million people worldwide [1],
71 causing substantial morbidity, reduced quality of life and substantial health-care costs [2]. Asthma
72 predominantly originates in early childhood [3] with an estimated 1.1 million children affected in the
73 UK [4].

74 There is a male predominance of asthma in early childhood. During adolescence more boys
75 remit asthma than girls, while more girls acquire asthma than boys, which results in gender-reversal
76 of asthma prevalence from pre- to post-adolescence [5-14] with asthma becoming more prevalent and
77 severer among females after puberty [9, 15, 16]. However, the underlying biological mechanisms
78 involved in these sex differences in the natural history of asthma across childhood and adolescence
79 remain unclear.

80 Although the pathogenesis of asthma reflects a combination of inherited susceptibility and
81 environmental exposures, the etiology and biological mechanisms are poorly understood. The
82 increase in prevalence of asthma in recent decades suggests an important role for environmental
83 exposures in the development of asthma in genetically high-risk individuals, and a number of studies
84 have highlighted the potential for a role of epigenetic programming in response to early life
85 environmental exposures in asthma susceptibility [17-21]. One of the most widely studied epigenetic
86 mechanisms is DNA methylation (DNAm) [22, 23]. DNAm at specific cytosine-phosphate-guanine
87 (CpG) sites in DNA from both blood and lung tissue have been found to be associated with asthma
88 [24-38] and related phenotypes such as COPD [39-42].

89 While these studies have established a clear association between DNAm patterns and asthma,
90 they rely on asthma status determined at a single time point. Yet, as previously discussed, asthma
91 phenotype within an individual can be dynamic, new incidence and clinical remission occurring
92 across the life course with gender reversal in asthma prevalence observed in adolescence. In a
93 candidate gene approach, we have previously investigated temporal changes of DNAm at CpG sites in
94 genes encoding proteins in the Th2 pathway and the transition of asthma over adolescence [4]. This
95 study showed that the level of DNAm and the association between specific CpGs (and their

96 interaction with DNA sequence variation) and asthma changes across adolescence. We therefore
97 hypothesized that DNAm changes at specific sites across adolescence might explain the biological
98 basis of sex differences in the natural history of asthma across adolescence and identify biomarkers of
99 asthma acquisition in adolescence that would potentially be beneficial for prediction and prevention.
100 To test this, we have used genome-wide DNAm data to assess sex-specific association of DNAm
101 changes from pre- to post-adolescence with asthma acquisition at asthma-associated CpG sites.

102

103

104

105

106

107

108

109

110

111

112

113 **Methods (1251 words)**

114 **Study Population**

115 The Isle of Wight birth cohort (IOWBC) consists of children born between January 1, 1989 and
116 February 28, 1990 on the Isle of Wight (IoW), United Kingdom [43]. The IOWBC was established to
117 investigate the natural history of allergic diseases among children residing on a semi-rural island near
118 the UK mainland. Of the 1,536 pregnancies in this period, 1,456 parents consented for further
119 follow-up with survey and clinical data collected at 1, 2, 4, 10, and 18 years.

120 **Asthma acquisition**

121 Detailed questionnaires that included the questions from the International Study of Asthma and
122 Allergy in Childhood (ISAAC) were administered to parents/participants at 10 and 18 years. Asthma
123 was defined as “ever had asthma” and “wheezing or whistling in the chest in the last 12 months” or
124 “current treatment for asthma.” This study focuses on new incidence (i.e., acquisition) of asthma
125 between 10 and 18 years defined as no asthma at 10 years but having asthma at 18 years. Subjects that
126 had no asthma at both ages were used as a reference group.

127 Of the 1,053 subjects who did not have asthma at 10 years, 275 subjects had DNAm
128 measurements available from peripheral blood samples at both 10 and 18 years and were included in
129 further analyses.

130 Covariates

131 Information regarding sex, birth weight, maternal and paternal disease status of asthma was assessed
132 based on questionnaire data and hospital records collected at birth. Socio-economic status was defined
133 based on household income, number of rooms and maternal education. Atopic status was assessed at
134 10 and 18 years using skin prick test (SPT) for 11 common allergens (house dust mite, cat dander, dog
135 dander, grass pollen mix, tree pollen mix, *Alternaria alternata*, *Cladosporium herbarium*, cow’s milk,
136 hen’s egg, peanut, and cod), and change of atopic status from 10 to 18 years was recorded. Height and
137 weight were measured at 10 and 18 years, and in cases that a participant did not visit the study center,
138 information was obtained by telephone interviews. Body mass index (BMI) was calculated based on
139 height and weight, and relative changes in height and BMI were calculated for each subject, for
140 instance, relative change in height of a subject is calculated as the difference in height from pre- to
141 post-adolescence divided by their pre-adolescent height.

142 DNA methylation (DNAm)

143 DNAm was measured in peripheral blood with samples collected at 10 and 18 years using either the
144 Infinium HumanMethylation450 BeadChips or MethylationEPIC BeadChips (Illumina, Inc., San
145 Diego, CA). Pre-processing of DNAm was carried out using the CPACOR pipeline [44]. Details of
146 DNAm data generation, quality control, and preprocessing, as well as principal components (PC)

147 analyses detecting latent variables for batch and technical variations are in the Supplemental Material
148 S1. After preprocessing, a total of 442,475 CpGs in common between the two platforms were
149 included in the analyses.

150 Since blood is a mixture of functionally and developmentally distinct cell populations [45],
151 adjusting cell type compositions was needed in analyses to reduce confounding from cell
152 heterogeneity in DNAm measured from blood samples [46]. We estimated cell type proportions using
153 the method proposed by Jaffe and Irizarry [47], adapted from Houseman et al.[48], using the
154 Bioconductor *minfi* package [49]. The estimated cell type proportions of CD4+ T cells, natural killer
155 cells, neutrophil, B cells, monocytes, and eosinophil cells were included in the analyses as
156 confounding factors.

157 Gene Expression

158 Gene expression levels from peripheral blood samples collected at 26 years from IOWBC was
159 determined using paired-end (2*75 bp) RNA sequencing. All samples were sequenced twice using the
160 identical protocol and for each sample the output from both runs were combined. Normalized read
161 count were calculated and their log transformed values were used for data analysis. Details on RNA
162 sequencing, transcript reading, mapping and assembly, and normalization are in the Supplemental
163 Material S2.

164 Statistical analysis

165 To examine whether the subsample (n=275) included in the study reasonably represented the
166 complete cohort (n=1,053), one sample proportion tests and multinomial tests for categorical variables
167 and one sample t-tests for continuous variables were applied.

168 Screening analysis to identify asthma-associated CpGs

169 An R package, *ttScreening*, was implemented to screen for CpGs whose methylation (in M values)
170 was associated with asthma cross-sectionally [50]. The screening method implemented in this
171 package has been shown to perform better than FDR-based and Bonferroni-based methods and has the

172 potential of controlling both types I and II errors [50]. Subjects with DNAm and asthma data at one or
173 both ages were included in the screening. A CpG site showing statistical significance (at the 0.05
174 level) in at least 50% randomly selected training and testing data set pairs was treated as asthma-
175 associated CpGs and included in subsequent analyses. This screening approach was cross-sectional
176 and focused on asthma status rather than asthma transition, to avoid data double dipping, i.e., avoid
177 using the same or a very similar model in screening as well as in final data analyses with the same
178 data.

179 Assessment of DNAm change across adolescence for asthma-associated CpGs

180 CpGs that passed screening were treated as potentially asthma-associated CpGs. At these sites, M
181 values of DNAm at each CpG were regressed on 15 PCs obtained from control probes (Supplemental
182 Material S1) and the 6 cell type proportions [51] to obtain batch and cell-type adjusted DNAm
183 (residuals). This regression analysis was conducted at 10 and 18 years, respectively. At each of the
184 asthma-associated CpGs, the difference in residuals between 10 and 18 years was then calculated for
185 each subject to represent DNAm change from 10 to 18 years.

186 Logistic regressions via R function *glm* with a logit link were applied to evaluate the
187 association of asthma acquisition (with asthma-free as the reference) with DNAm changes
188 (independent variable) adjusted for covariates and confounders potentially associated with asthma:
189 maternal and paternal history of asthma, sex, birth weight, socio-economic status, change of atopic
190 status from 10 to 18 years, and relative changes in height and BMI from 10 to 18 years. Additionally,
191 since multiple studies have demonstrated gender reversal on asthma prevalence from pre- to post-
192 adolescence, interaction effects of CpGs and sex on asthma acquisition were assessed. Multiple
193 testing was adjusted by controlling a false discovery rate (FDR) of 0.05.

194 Replication cohort - the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort

195 CpGs shown to be associated with asthma acquisition in IOWBC were further assessed in an
196 independent cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC) [52]. DNAm
197 data at 7 and 15 or 17 years and asthma acquisition from 7 to 15 years were included in the replication
198 analyses. Details of these data along with information on covariates are presented in the

199 Supplemental Material S3. Please note that the study website contains details of all the data that is
200 available through a fully searchable data dictionary and variable search tool
201 (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). A p -value <0.05 was deemed as being
202 statistically significant.

203 Association between DNAm and gene expression

204 For CpGs with DNAm changes showing consistent associations with asthma acquisition
205 between the two cohorts, we evaluated their biological relevance. Genes annotated to the identified
206 CpGs were extracted from the Illumina's manifest file or SNIPPER (<https://csg.sph.umich.edu/boehnke/snipper/>) version 1.2. We tested the association between DNAm at these CpGs and gene
207 expression in blood at 26 years using linear regressions. Gene expression ($n=136$) was the dependent
208 variable, and DNAm and sex were the independent variables. DNAm at 10 and 18 years were
209 analyzed separately. In addition, to assess sex-specificity of DNAm and expression association, an
210 interaction term of DNAm $\times$ sex was also included in the model. Interaction effects were treated as
211 being statistically significant with p -value <0.05 .

213

214 Results (836 words)

215 In IOWBC, the analytical subsample was representative of the complete cohort with respect to asthma
216 transition status, demographic variables, and other covariates (p -values >0.05 , Table 1). A sex
217 difference in asthma acquisition was observed in the complete cohort; 10.8% of females acquired
218 asthma from 10 to 18 years, as compared to only 7.1% of males (p -value = 0.03).

219 To identify candidate CpGs potentially associated with asthma at 10 and 18 years for each sex,
220 we applied *ttScreening* to 442,475 CpGs, stratified by sex. In total, 265 (220 for males, 45 for
221 females) CpGs and 290 (40 for males, 250 for females) CpGs passed screening at 10 and 18 years,
222 respectively. CpGs that passed screening at either age of males or females (535 CpGs in total;
223 Supplemental Table S1) were treated as asthma-associated CpGs and included in subsequent analyses.

224 For each CpG site that passed screening, whether asthma acquisition was associated with
225 changes in DNAm from 10 to 18 years and whether such an association was sex-specific were tested
226 using logistic regressions. Sex and DNAm changes, and their interaction, along with adjusting factors
227 were included in the model. After controlling FDR at 0.05, statistically significant interaction effects
228 were observed at 13 CpG sites (Table 2, left panel of Figure 1). All the coefficients for the interaction
229 effects between sex and DNAm changes were positive. Combined with the estimates of main effect, a
230 potential gender reversal with respect to the effects of DNAm changes on asthma acquisition was
231 identified. For instance, at CpG site cg03269757, a larger increase in DNAm from 10 to 18 years was
232 associated with an increased risk of acquisition in females (log-OR=3.04), but a decreased risk of
233 acquisition in males (log-OR=-1.34). Such opposite associations between males and females were
234 observed at nine of the 13 CpGs. At the remaining four CpG sites, cg11814087, cg12587133,
235 cg18278943, and cg22484084, the association of DNAm changes with asthma acquisition were much
236 stronger in females with larger effect size (increased risk of acquiring asthma). For example, at
237 cg11814087, the log-OR was 6.72 for females, much higher than the log-OR=0.58 for males
238 (interaction effect p-value 8.85×10^{-4} with 95% CI: 2.29, 10.01).

239 We further tested these 13 CpGs in the ALSPAC cohort. At 10 of the 13 CpG sites, consistent
240 interaction effects with respect to the direction of effects were observed compared to those found in
241 IOWBC, although only one of the 10 CpGs (cg20891917) showed a statistically significant effect
242 (Table 2 and Figure 1). In addition, in the ALSPAC cohort, for eight of these 10 CpGs, the interaction
243 effects were all much stronger than the main effects (Figure 1), the same pattern observed in IOWBC.

244 To explore underlying mechanisms of the observed interaction effects, for each of the 13
245 CpGs, we calculated average DNAm changes between 10 and 18 years in IOWBC, and between 7
246 years and 15 or 17 years in the ALSPAC cohort, for males and females, separately (Figures 2a and
247 2b). In both cohorts, average DNAm changes in non-asthmatic subjects were all around zero at the 13
248 CpGs. However, for subjects in IOWBC who acquired asthma between 10 and 18 years, across all the
249 13 CpGs, the changes in DNAm were opposite between males and females with males showing
250 decrease in DNAm from pre- to post-adolescence (negative differences in males but positive
251 differences in females, Figure 2a). In the ALSPAC cohort, for the first 10 CpGs in Figure 2b, DNAm
252 at these CpGs showed consistent directions of interaction effects with those in IOWBC (Table 2). The

253 pattern of opposite changes in DNAm from age 7 to 15 or 17 years in ALSPAC between males and
254 females were also observed at 7 of these 10 CpGs (Figure 2b).

255 To assess the biological relevance of the 10 CpGs showing consistent sex-specificity between
256 the two cohorts, we evaluated the association of DNAm at these CpGs with expression of their
257 mapped genes and whether such associations were sex-specific. The 10 CpGs were mapped to 10
258 genes (Table 2). We did not have expression data for gene *PTPRV*. Statistically significant interaction
259 effects were observed at five of the nine CpG sites (Table 3) with four genes identified based on age
260 10 DNAm and one based on age 18 DNAm. Combined with the estimates of main effect of DNAm, a
261 potential gender reversal with respect to the association of DNAm with gene expression levels was
262 found at all these five CpG sites (based on opposite signs of the estimated main and interaction
263 regression coefficients). For instance, at CpG site cg11295724, an increase in DNAm at 10 years was
264 associated with increased gene expression levels of *SIRPD* in males, but decreased expression in
265 females. Similar opposite patterns were observed for the CpGs on *CCDC146*, *SLMAP*, and *ZNF385A*.
266 For cg03269757 on *ATL2*, although for both sexes, the regression coefficients were negative (-0.98
267 for males and -0.06 for females), the effect size for males was more than 16 times as that for females
268 (which was close to zero), representing a potential gender reversal effect as well.

269 **Discussion (1135 words)**

270 We examined the sex-specificity on the association of changes in DNAm with asthma acquisition
271 from pre-adolescence to late- (ALSPAC) or post-adolescence (IOWBC) in two birth cohorts. In
272 IOWBC, 13 CpGs mapping to 13 genes were identified that showed statistically significant
273 interaction effects with sex, of which 10 (77%) CpGs showed consistent directions of interaction
274 effects in ALSPAC with one CpG (cg20891917) being statistically significant. In most of these CpGs,
275 the effects of changes in DNAm on asthma acquisition during adolescence were opposite between
276 males and females, showing gender reversal of DNAm effects. Accompanied by the opposite
277 direction of changes in DNAm between males and females at most of the identified CpGs in both
278 cohorts, this suggests that DNAm may represent a mechanism underlying the well-established gender
279 reversal in asthma prevalence across adolescence [15, 16].

280 In addition, assessment of the biological relevance for 9 of the 10 CpGs indicated a potential
281 of epigenetic regulatory functionality on gene activities. DNAm at 5 of the 9 CpGs showed sex-
282 specific associations with gene expression, supporting the gender reversal phenomenon found in the
283 association assessment between DNAm changes and asthma acquisition during adolescence. In
284 addition, the association of gene expression with DNAm was overall stronger in males than in
285 females, indicating a possibility of stronger influence of DNAm on gene activities in males. Previous
286 studies have shown that *ZNF385A* is overexpressed and *SLMAP* is under expressed in asthma cases
287 compared to controls [53], and gene *CCDC146* is also identified as one of the differentially expressed
288 genes in relation to asthma [54], although it is unclear whether such differentiation is different
289 between males and females. The sex-specific association of DNAm with expression of these genes
290 observed in our study may imply different underlying epigenetic regulations on gene activities
291 between males and females. In addition, it will be interesting to examine how changes in DNAm are
292 associated with changes in gene expressions, to further improve our understanding of epigenetic
293 regulatory functionality on asthma acquisition.

294 The strength of this study is the availability of DNAm data and asthma status at two important
295 time points, pre- and post-adolescence, enabling the possibility to examine changes in DNAm and its
296 association with asthma acquisition during adolescence. To our knowledge, this is the first study to
297 examine the epigenetics of asthma acquisition during adolescence regarding sex-differences.

298 For each of the genes annotated to the identified CpGs, we performed a literature search for
299 their possible roles that they played in the risk of asthma. Among the CpGs showing consistent
300 direction of interaction effects, DNAm of gene *ZNF385A* and expression of *IFRD1* have been
301 reported as biomarkers of asthma [55]. Our finding on the association of DNAm with expression of
302 *ZNF385A* further strengthens its relationship with asthma. Lund et al. demonstrated that *NDFIP2*, an
303 IL-4 regulated gene, promoted IFN- γ production by the polarized human Th1 lymphocytes [56]. One
304 of our earlier studies in females showed that genes in the Th2 pathway were likely to contribute to an
305 increased risk of asthma and be associated with the risk of asthma transition [4]. Although in the
306 present study we did not identify genes in the Th2 pathway, Th1 and Th2 cells work tightly and
307 interact with other immune cells by regulating their functions with specific cytokine production,
308 associated with the pathogenesis of asthma. As for gene *ZFR*, single nucleotide polymorphisms in
309 *ZFR* have been shown to be associated with asthma or bronchial hyperresponsiveness [57]. However,
310 none of these studies have mentioned sex-specificity in these associations due to the focus of the
311 study and the methods applied in the study. For instance, the study of Zhang et al. [4] only included
312 females; while in the study of Kurz et al. [57], the focus was to identify single nucleotide
313 polymorphisms associated with asthma or bronchial hyperresponsiveness and the effects of sex were
314 not considered. Furthermore, findings in the literature all focused on the risk of asthma instead of the
315 risk of asthma acquisition, which might also explain the limited findings in the literature supporting
316 the identified genes.

317 At most of the CpGs identified in IOWBC, consistent direction of interaction effects was
318 found in ALSPAC. However, statistical significance was not observed at those CpG sites except for
319 cg20891917. In the ALSPAC cohort, many subjects' DNAm was assessed at 15 years and asthma
320 status change was from 7 years to 15 years. At the age of 15 years, it was likely that children were
321 still in the period of pubertal transition, and thus sex-specificity might not be strong enough to be
322 detected. In addition, we noticed that among the 10 CpG sites showing consistent sex-specificity
323 between the two cohorts, associations in DNAm with expression of genes happened more often with
324 DNAm at age 10 years. We do not have a specific biological explanation for this observation but
325 postulate that this might have been due to larger variations in DNAm data at age 18 compared to
326 DNAm at age 10, and thus we did not have enough power to detect the associations.

327 In this study, the candidate CpGs were identified based on their associations with asthma
328 status at 10 and 18 years separately. With this approach, we were able to focus on asthma related CpG
329 sites, which was the starting point of the study. On the other hand, we might have missed CpGs that
330 were not related to asthma at ages 10 nor 18 years but were related to asthma acquisition from 10 to
331 18 years. However, screening candidate CpGs based on asthma acquisition had the risk of double-
332 dipping the data. That is, the screening and final association analyses would share a similar analytical
333 model applied to the same data, which in general is not encouraged. In addition, screening of CpGs
334 and association analyses were applied to each individual CpG sites. CpG sites might be correlated and
335 jointly impact asthma acquisition. Using our approach, correlated CpGs might have presented an issue
336 that we were unable to address. Approaches analogous to linkage disequilibrium and haplotype
337 identification in genetic studies deserve further investigations both methodologically and
338 experimentally. We also would like to point out that the present study was based on a concurrent
339 analysis (i.e. both DNAm changes and asthma acquisition were in the same period). The focus of the
340 study was on associations rather than causality and this analytical approach does not allow predictions
341 or inferring causality.

342 Nevertheless, the consistency in the results between the two cohorts indicated that the
343 identified CpGs are likely to play a role in the underlying mechanisms of sex-specific asthma
344 acquisition. Furthermore, the sex-specific associations of DNAm at most of these CpGs with
345 expressions of their mapped genes demonstrated their potential of biological relevance and supported
346 our observed sex-specificity related to asthma acquisition. Although future studies are warranted to
347 further examine the credibility of the identified CpGs, these CpGs have the potential to serve as
348 candidate markers in subsequent mechanistic studies on gender reversal of asthma acquisition.

349

350

351

352

353

354

355 **Acknowledgements**

356 This study was supported by the National Institutes of Health research fund R01AI121226 (MPI: H
357 Zhang and JW Holloway). Part of the methylation data generation was supported by R01AI091905
358 (PI: W Karmaus). The Isle of Wight Birth Cohort assessments have been supported by the National
359 Institutes of Health USA (Grant no. R01HL082925, H. Arshad), Asthma UK (Grant no. 364. S.H.
360 Arshad) and the David Hide Asthma and Allergy Research Trust.

361 The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the
362 University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is
363 available on the ALSPAC website ([http://www.bristol.ac.uk/alspac/external/documents/grant-](http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)
364 [acknowledgements.pdf](http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)). Generation of methylation array data was specifically funded by NIH
365 R01AI121226, R01AI091905, BBSRC BBI025751/1 and BB/I025263/1, MRC MC_UU_12013/1,
366 MC_UU_12013/2, MC_UU_12013/8.

367 The authors are thankful to the nurses and staff at the David Hide Asthma & Allergy Research
368 Centre, Isle of Wight, UK, for their help in recruitment and sample collections, and are thankful to all
369 the cohort participants. Our special thanks also go to the High-Performance Computing facility
370 provided by the University of Memphis.

371 For ALSPAC, DNA extraction and generation of Illumina array data was carried out in the
372 Bristol Bioresource Laboratories at the University of Bristol, UK. We are extremely grateful to all the
373 families who took part in this study, the midwives for their help in recruiting them, and the whole
374 ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers,
375 research scientists, volunteers, managers, receptionists and nurses.

376

377 **Data availability statement**

378 The datasets analyzed during the current study are not publicly available but are available from the
379 corresponding author on reasonable request.

380 **Conflicts of interest**

381 The authors declare that they have no conflicts of interest.

382 **Declarations**

383 *Ethics approval and consent to participate*

384 The IoW birth cohort study was approved by Isle of Wight, Portsmouth and Hampshire Local
385 Research Ethics Committee (now known as the National Research Ethics Service, NRES Committee
386 South Central – Hampshire A) (06/Q1701/34) and the IRB at the University of Memphis
387 (FWA00006815). Written informed consent was obtained from parents during in-person visits. For
388 participants assessed by phone interview, consent was documented on the consent form with the name
389 of the person giving consent, and the name and signature of the person taking the form were recorded.
390 For ALSPAC, ethical approval for the study was obtained from the ALSPAC Ethics and Law
391 Committee and the Local Research Ethics Committees and consent for collection of biological
392 samples was provided in accordance with the Human Tissue Act (2004). For age seven years, United
393 Bristol Healthcare Trust: E4168 (ALSPAC Hands on Assessments at Age Seven), Southmead Health
394 Services: 67/98 (Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) - Hands on
395 Assessments at Age Seven) and Frenchay Healthcare Trust: 98/52 (Avon Longitudinal Study of
396 Pregnancy and Childhood (ALSPAC). Hands on Assessments at Age Seven). For age 15 years,
397 Central & South Bristol Research Ethics Committee (UBHT): 06/Q2006/53 Avon Longitudinal Study
398 of Parents and Children (ALSPAC), Hands on Assessments: Teen Focus 3 (Focus 15+), and for age
399 17 years, North Somerset & South Bristol Research Ethics Committee: 08/H0106/9 Avon
400 Longitudinal Study of Parents and Children (ALSPAC), Hands on Assessments: Teen Focus 4 (Focus
401 17+). Full details of ethical approvals (local committees and approval numbers) are available at
402 [http://www.bristol.ac.uk/media-](http://www.bristol.ac.uk/media-library/sites/alspac/documents/governance/Research%20Ethics%20Committee%20approval%20references.pdf)
403 [library/sites/alspac/documents/governance/Research%20Ethics%20Committee%20approval%20refer-](http://www.bristol.ac.uk/media-library/sites/alspac/documents/governance/Research%20Ethics%20Committee%20approval%20references.pdf)
404 [ences.pdf](http://www.bristol.ac.uk/media-library/sites/alspac/documents/governance/Research%20Ethics%20Committee%20approval%20references.pdf)

406 **Statement of author contribution**

407 HZ designed the study, FS, AW, RP and RA analyzed the data, LK interpreted bioinformatics
408 findings, WK, HA, CR, JAH, and JWH supervised the study, SE, JH, and SR provided DNA
409 methylation data, RP and HZ drafted the manuscript, and all authors reviewed the manuscript.

410

411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439

References

1. *The Global Asthma Report 2018: Global Asthma Network*. 2018: Auckland, New Zealand.
2. Eder, W., M.J. Ege, and E. von Mutius, *The asthma epidemic*. N Engl J Med, 2006. **355**(21): p. 2226-35.
3. Martinez, F.D., et al., *Asthma and wheezing in the first six years of life. The Group Health Medical Associates*. N Engl J Med, 1995. **332**(3): p. 133-8.
4. Zhang, H., et al., *The interplay of DNA methylation over time with Th2 pathway genetic variants on asthma risk and temporal asthma transition*. Clin Epigenetics, 2014. **6**(1): p. 8.
5. Arathimos, R., et al., *Sex discordance in asthma and wheeze prevalence in two longitudinal cohorts*. PLoS One, 2017. **12**(4): p. e0176293.
6. Fuseini, H. and D.C. Newcomb, *Mechanisms Driving Gender Differences in Asthma*. Curr Allergy Asthma Rep, 2017. **17**(3): p. 19.
7. Postma, D.S., *Gender differences in asthma development and progression*. Gend Med, 2007. **4 Suppl B**: p. S133-46.
8. Vink, N.M., et al., *Gender differences in asthma development and remission during transition through puberty: the TRacking Adolescents' Individual Lives Survey (TRAILS) study*. J Allergy Clin Immunol, 2010. **126**(3): p. 498-504.e1-6.
9. Koper, I., K. Hufnagl, and R. Ehmann, *Gender aspects and influence of hormones on bronchial asthma - Secondary publication and update*. World Allergy Organ J, 2017. **10**(1): p. 46.
10. Naeem, A. and P. Silveyra, *Sex Differences in Paediatric and Adult Asthma*. Eur Med J (Chelmsf), 2019. **4**(2): p. 27-35.

- 440 11. Zein, J.G. and S.C. Erzurum, *Asthma is Different in Women*. Curr Allergy Asthma Rep, 2015.
441 **15**(6): p. 28.
- 442 12. Han, Y.Y., E. Forno, and J.C. Celedon, *Sex Steroid Hormones and Asthma in a Nationwide*
443 *Study of U.S. Adults*. Am J Respir Crit Care Med, 2019.
- 444 13. Hohmann, C., et al., *Sex-specific incidence of asthma, rhinitis and respiratory multimorbidity*
445 *before and after puberty onset: individual participant meta-analysis of five birth cohorts*
446 *collaborating in MeDALL*. BMJ Open Respiratory Research, 2019. **6**(1): p. e000460.
- 447 14. Soto-Ramírez, N., et al., *Epidemiologic methods of assessing asthma and wheezing episodes*
448 *in longitudinal studies: measures of change and stability*. J Epidemiol, 2013. **23**(6): p. 399-
449 410.
- 450 15. Pignataro, F.S., et al., *Asthma and gender: The female lung*. Pharmacol Res, 2017. **119**: p.
451 384-390.
- 452 16. Osman, M., et al., *Gender-specific presentations for asthma, allergic rhinitis and eczema in*
453 *primary care*. Prim Care Respir J, 2007. **16**(1): p. 28-35.
- 454 17. Begin, P. and K.C. Nadeau, *Epigenetic regulation of asthma and allergic disease*. Allergy
455 Asthma Clin Immunol, 2014. **10**(1): p. 27.
- 456 18. Durham, A.L., C. Wiegman, and I.M. Adcock, *Epigenetics of asthma*. Biochim Biophys Acta,
457 2011. **1810**(11): p. 1103-9.
- 458 19. Ege, M.J., et al., *Exposure to environmental microorganisms and childhood asthma*. N Engl J
459 Med, 2011. **364**(8): p. 701-9.
- 460 20. Stein, M.M., et al., *Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children*.
461 N Engl J Med, 2016. **375**(5): p. 411-421.
- 462 21. Yang, I.V. and D.A. Schwartz, *Epigenetic mechanisms and the development of asthma*. J
463 Allergy Clin Immunol, 2012. **130**(6): p. 1243-55.
- 464 22. Joubert, B.R., et al., *DNA Methylation in Newborns and Maternal Smoking in Pregnancy:*
465 *Genome-wide Consortium Meta-analysis*. Am J Hum Genet, 2016. **98**(4): p. 680-96.
- 466 23. Joubert, B.R., et al., *450K epigenome-wide scan identifies differential DNA methylation in*
467 *newborns related to maternal smoking during pregnancy*. Environ Health Perspect, 2012.
468 **120**(10): p. 1425-31.

- 469 24. DeVries, A., et al., *Epigenome-wide analysis links SMAD3 methylation at birth to asthma in*
470 *children of asthmatic mothers*. J Allergy Clin Immunol, 2017. **140**(2): p. 534-542.
- 471 25. Murphy, T.M., et al., *Methylomic markers of persistent childhood asthma: a longitudinal*
472 *study of asthma-discordant monozygotic twins*. Clin Epigenetics, 2015. **7**: p. 130.
- 473 26. Arathimos, R., et al., *Epigenome-wide association study of asthma and wheeze in childhood*
474 *and adolescence*. Clin Epigenetics, 2017. **9**: p. 112.
- 475 27. Gunawardhana, L.P., et al., *Characteristic DNA methylation profiles in peripheral blood*
476 *monocytes are associated with inflammatory phenotypes of asthma*. Epigenetics, 2014. **9**(9): p.
477 1302-16.
- 478 28. Nicodemus-Johnson, J., et al., *DNA methylation in lung cells is associated with asthma*
479 *endotypes and genetic risk*. JCI Insight, 2016. **1**(20): p. e90151.
- 480 29. Nicodemus-Johnson, J., et al., *Genome-Wide Methylation Study Identifies an IL-13-induced*
481 *Epigenetic Signature in Asthmatic Airways*. Am J Respir Crit Care Med, 2016. **193**(4): p. 376-
482 85.
- 483 30. Rastogi, D., M. Suzuki, and J.M. Grealley, *Differential epigenome-wide DNA methylation*
484 *patterns in childhood obesity-associated asthma*. Sci Rep, 2013. **3**: p. 2164.
- 485 31. Yang, I.V., et al., *DNA methylation and childhood asthma in the inner city*. J Allergy Clin
486 Immunol, 2015. **136**(1): p. 69-80.
- 487 32. Yang, I.V., et al., *The Nasal Methylome: A Key to Understanding Allergic Asthma*. Am J
488 Respir Crit Care Med, 2017. **195**(6): p. 829-831.
- 489 33. Rathod, A., et al., *Interweaving Between Genetic and Epigenetic Studies on Childhood*
490 *Asthma*. Epigenet Insights, 2020. **13**: p. 2516865720923395.
- 491 34. Forno, E., et al., *DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a*
492 *genome-wide study*. Lancet Respir Med, 2019. **7**(4): p. 336-346.
- 493 35. Neophytou, A.M., et al., *In utero tobacco smoke exposure, DNA methylation, and asthma in*
494 *Latino children*. Environ Epidemiol, 2019. **3**(3): p. e048.
- 495 36. Prunicki, M., et al., *Exposure to NO(2), CO, and PM(2.5) is linked to regional DNA*
496 *methylation differences in asthma*. Clin Epigenetics, 2018. **10**: p. 2.

- 497 37. Reese, S.E., et al., *Epigenome-wide meta-analysis of DNA methylation and childhood asthma*.
498 J Allergy Clin Immunol, 2019. **143**(6): p. 2062-2074.
- 499 38. Xu, C.J., et al., *DNA methylation in childhood asthma: an epigenome-wide meta-analysis*.
500 Lancet Respir Med, 2018. **6**(5): p. 379-388.
- 501 39. den Dekker, H.T., et al., *Newborn DNA-methylation, childhood lung function, and the risks of*
502 *asthma and COPD across the life course*. Eur Respir J, 2019. **53**(4).
- 503 40. Machin, M., et al., *Systematic review of lung function and COPD with peripheral blood DNA*
504 *methylation in population based studies*. BMC Pulm Med, 2017. **17**(1): p. 54.
- 505 41. Sundar, I.K., et al., *DNA methylation profiling in peripheral lung tissues of smokers and*
506 *patients with COPD*. Clin Epigenetics, 2017. **9**: p. 38.
- 507 42. Wielscher, M., et al., *Diagnostic Performance of Plasma DNA Methylation Profiles in Lung*
508 *Cancer, Pulmonary Fibrosis and COPD*. EBioMedicine, 2015. **2**(8): p. 929-36.
- 509 43. Arshad, S.H., et al., *Cohort Profile: The Isle Of Wight Whole Population Birth Cohort*
510 *(IOWBC)*. Int J Epidemiol, 2018. **47**(4): p. 1043-1044i.
- 511 44. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450*
512 *BeadChip improves data quality and performance in epigenome-wide association studies*.
513 Genome Biol, 2015. **16**: p. 37.
- 514 45. Reinius, L.E., et al., *Differential DNA methylation in purified human blood cells: implications*
515 *for cell lineage and studies on disease susceptibility*. PLoS One, 2012. **7**(7): p. e41361.
- 516 46. Koestler, D.C., et al., *Blood-based profiles of DNA methylation predict the underlying*
517 *distribution of cell types: a validation analysis*. Epigenetics, 2013. **8**(8): p. 816-26.
- 518 47. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-*
519 *wide association studies*. Genome Biol, 2014. **15**(2).
- 520 48. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture*
521 *distribution*. BMC Bioinformatics, 2012. **13**.
- 522 49. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis*
523 *of Infinium DNA methylation microarrays*. Bioinformatics, 2014. **30**(10): p. 1363-9.
- 524 50. Ray, M.A., et al., *An Efficient Approach to Screening Epigenome-Wide Data*. Biomed Res Int,
525 2016. **2016**: p. 2615348.

- 526 51. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture*
527 *distribution*. BMC Bioinformatics, 2012. **13**: p. 86.
- 528 52. Boyd, A., et al., *Cohort Profile: the 'children of the 90s'--the index offspring of the Avon*
529 *Longitudinal Study of Parents and Children*. Int J Epidemiol, 2013. **42**(1): p. 111-27.
- 530 53. Bunyavanich, S.P., Gaurav; Schadt, Eric, *Nasal Biomarkers of Asthma*, in *International*
531 *Bureau*, W.I.P. Organization, Editor. 2017, Icahn School of Medicine at Mount Sinai: United
532 States of America.
- 533 54. Maghsoudloo, M., et al., *Identification of biomarkers in common chronic lung diseases by co-*
534 *expression networks and drug-target interactions analysis*. Molecular Medicine, 2020. **26**(1):
535 p. 9.
- 536 55. De Boever, P.L., S., *Epigenetic Markers for Respiratory Allergy*, in *International Bureau*,
537 W.I.P. Organization, Editor. 2016, Vito NV.
- 538 56. Lund, R.J., et al., *Genome-Wide Identification of Novel Genes Involved in Early Th1 and Th2*
539 *Cell Differentiation*. The Journal of Immunology, 2007. **178**(6): p. 3648.
- 540 57. Kurz, T., et al., *Fine mapping and positional candidate studies on chromosome 5p13 identify*
541 *multiple asthma susceptibility loci*. J Allergy Clin Immunol, 2006. **118**(2): p. 396-402.
- 542
- 543
- 544

545 **Table 1:** Asthma acquisition and non-asthma participants included in the present study compared to
546 the participants in the complete cohort.

Categorical Variables: N (%)		Females			Males		
		Subsample n=124	Complete cohort n=544	p- value	Subsample n=151	Complete cohort n=509	p- value
Asthma Transition	Acquisition	12 (9.7%)	59 (10.8%)	0.78	14 (9.27%)	36 (7.1%)	0.37
	Non- asthma	112 (90.3%)	485 (89.1%)		137 (90.7%)	473 (92.9%)	
Maternal asthma	Yes	15 (12.1%)	50 (9.3%)	0.35	23 (15.4%)	53 (10.4%)	0.06

	No	109 (87.9%)	490 (90.7%)		126 (84.5%)	453 (89.5%)	
Paternal asthma	Yes	14 (11.5%)	45 (8.4%)	0.27	15 (10.1%)	50 (9.9%)	1
	No	107 (88.4%)	492 (91.6%)		133 (89.8%)	454 (90%)	
Socio-economic status	Mid-High	42 (33.8%)	193 (36.4%)	0.67	55 (36.4%)	197 (40.2%)	0.60
	Low-mid	35 (28.2%)	153 (28.8%)		51 (33.7%)	149 (30.4%)	
	Low-low	47 (37.9%)	184 (34.7%)		45 (29.8%)	143 (29.2%)	
Change of atopic status from 10 to 18 year	Yes-No	4 (3.6%)	7 (2.37%)	0.53	3 (2.5%)	3 (1.4%)	0.43
	No-Yes	20 (18%)	52 (17.6%)		32 (26.8%)	62 (28%)	
	No-No	87 (78.3%)	236 (80%)		84 (70.5%)	156 (70.5%)	
Continuous Variables: Mean (SD)							
Birth weight		3.32 (0.51)	3.33 (0.50)	0.88	3.45 (0.56)	3.50 (0.52)	0.51
Relative change in height from 10 to 18 years		0.18 (0.04)	0.19 (0.04)	0.27	0.28 (0.03)	0.28 (0.04)	0.39
Relative change in BMI from 10 to 18 years		0.29 (0.15)	0.29 (0.16)	0.64	0.29 (0.15)	0.29 (0.15)	0.88

547

548

Table 2: Association of DNAm change with asthma acquisition from pre- to post-adolescence at 13 CpG sites that are sex-specific. Significant interactions between DNAm and sex on asthma acquisition from 10 to 18 years identified in the IOWBC^a were further tested in the ALSPAC cohort^b. Males are in the reference group.

IOWBC							ALSPAC cohort						
CpG site	Est. ^c	Int. ^c		P _{Raw} ^d	P _{FDR} ^d	Gene	CpG	Gene	Chr. ^e	Est. ^c	Int. ^c	P _{Raw} ^d	
cg03269757	-1.34	4.38	95% C.I. ^c 1.93, 6.83	1.88×10 ⁻⁴	0.02	<i>ATL2</i>	islands N_Shore	Body	2	0.03	0.59	0.19	
cg11295724	-0.33	7.82	4.12, 11.52	5.35×10 ⁻⁶	0.003	<i>SIRPD</i>	-	TSS200	20	0.06	0.76	0.22	
cg12587133	0.34	4.85	2.37, 7.32	1.03×10 ⁻⁵	0.003	<i>AGA</i>	-		4	0.01	0.24	0.5	
cg15154628	-0.67	3.23	1.14, 5.31	1.15×10 ⁻³	0.05	<i>CCDC146</i>	-	Body	7	0.24	0.07	0.89	
cg16301989	-0.48	4.42	1.89, 6.93	1.86×10 ⁻⁴	0.02	<i>FUNDC2P2^f</i>	-		2	0.05	0.47	0.26	
cg18278943	0.70	4.55	1.87, 7.24	4.36×10 ⁻⁴	0.03	<i>SLMAP</i>	N_Shelf		3	-0.25	0.82	0.18	
cg20885063	-0.52	4.45	1.86, 7.05	3.78×10 ⁻⁴	0.03	<i>ATPAF2</i>	N_Shelf	Body	17	0.007	0.4	0.36	
cg20891917	-0.43	5.5	2.03, 8.96	9.86×10 ⁻⁴	0.04	<i>IFRD1</i>	-	5'UTR	7	-0.13	1.26	0.01	
cg21163444	-2.22	11.13	4.31, 17.95	8.53×10 ⁻⁴	0.04	<i>ZNF385A</i>	S_Shore	Body	12	0.27	0.14	0.86	
cg22484084	0.29	7.18	2.69, 11.67	1.0×10 ⁻³	0.04	<i>PTPRV</i>	-	TSS1500	1	0.13	0.84	0.39	
cg06492287	-1.91	10.49	4.73, 16.25	7.78×10 ⁻⁵	0.01	<i>SNTG2</i>	-	Body	2	0.002	-0.04	0.96	
cg11770323	-0.7	9.18	4.11, 14.25	7.92×10 ⁻⁵	0.01	<i>NDFIP2</i>	-	Body	13	-0.75	-0.55	0.29	
cg11814087	0.58	6.14	2.29, 10.01	8.85×10 ⁻⁴	0.04	<i>ZFR</i>	-	Body	5	0.11	-0.68	0.05	

^a For the analyses in IoW, logistic regression models were adjusted for birth weight, sex, maternal and paternal disease status of asthma, socio-economic status (SES), change of atopic status, BMI, height from 10 to 18 years.

^b Analyses of ALSPAC used similar covariates: maternal disease status of asthma, sex, birth weight, SES, atopy status at 7 years, and changes of height, and BMI from 7 to 15/17 years.

^c Est. – Estimated main effect; Int. – interaction coefficient of the CpGs with the sex of a child. Interaction effects consistent between the two cohorts are with bold fonts; C.I. – confidence interval

^d P_{Raw} – raw p-value; P_{FDR} – adjusted p-value for multiple testing by controlling a false discovery rate (FDR) of 0.05. All the p-values are for interaction effects.

^e Chr. – Chromosome.

^f Genes closest to the CpG site annotated using the UCSC genome browser.

Table 3: Association of DNAm at 5 CpGs with their mapped genes' expression levels that are sex-specific. Only results on CpGs showing statistically significant interaction effects of DNAm×sex on gene expression were shown. Males are in the reference group. The p-values are for interaction effects.

Gene	CpG	DNAm effect	Sex×DNAm interaction	p-value (Sex×DNAm)	DNAm effect	Sex×DNAm interaction	p-value (Sex×DNAm)
Age 10 years					Age 18 years		
<i>SIRPD</i>	cg11295724	0.89	-1.39	0.022	0.24	-0.42	0.45
<i>CCDC146</i>	cg15154628	-0.40	0.72	0.041	0.44	-0.43	0.39
<i>SLMAP</i>	cg18278943	-1.43	1.90	0.006	-0.28	0.58	0.45
<i>ZNF385A</i>	cg21163444	-0.2	0.72	0.037	0.009	-0.24	0.56
<i>ATL2</i>	cg03269757	-0.66	0.80	0.09	-0.98	0.92	0.040

Figure 1: Effects of changes in methylation on asthma acquisition in adolescence for 13 CpGs in IOWBC (left) and ALSPAC (right) cohorts, stratified by sex-specific (interaction effects with male as the reference group, black bars) and sex nonspecific effects (main effects, gray bars). X-axis is for the regression coefficients (main effects and interaction effects). For interaction effects (black bars) in the IOWBC, 95% confidence intervals are in Table 2.

Figure 2: Scatter plots showing average DNAm changes from pre-adolescence to post-/late-adolescence, stratified by sex. (a) Average DNAm changes between 10 and 18 years in IOWBC. (b) Average DNAm changes between 7 and 15 or 17 years in the ALSPAC cohort. In both panels, left figure is for males, and right for females.

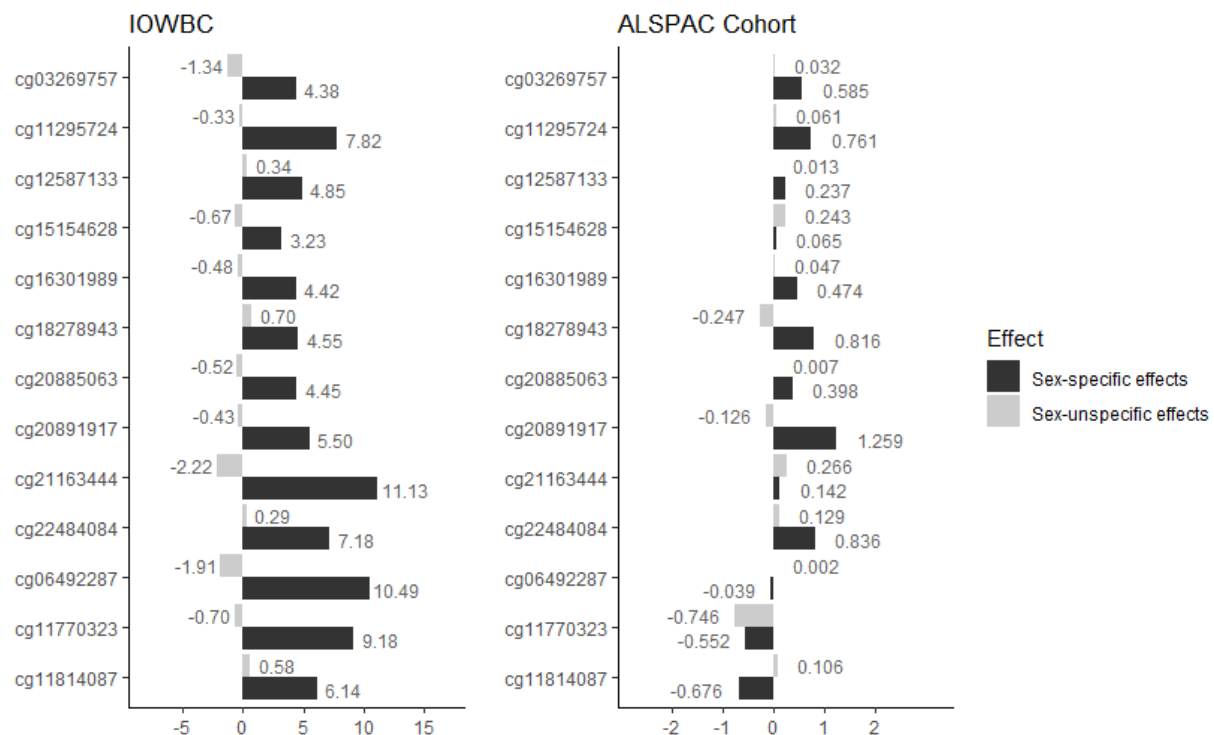


Figure 1: Effects of changes in methylation on asthma acquisition in adolescence for 13 CpGs in IOWBC (left) and ALSPAC (right) cohorts, stratified by sex-specific (interaction effects with male as the reference group, black bars) and sex nonspecific effects (main effects, gray bars). X-axis is for the regression coefficients (main effects and interaction effects). For interaction effects (black bars) in the IOWBC, 95% confidence intervals are in Table 2.

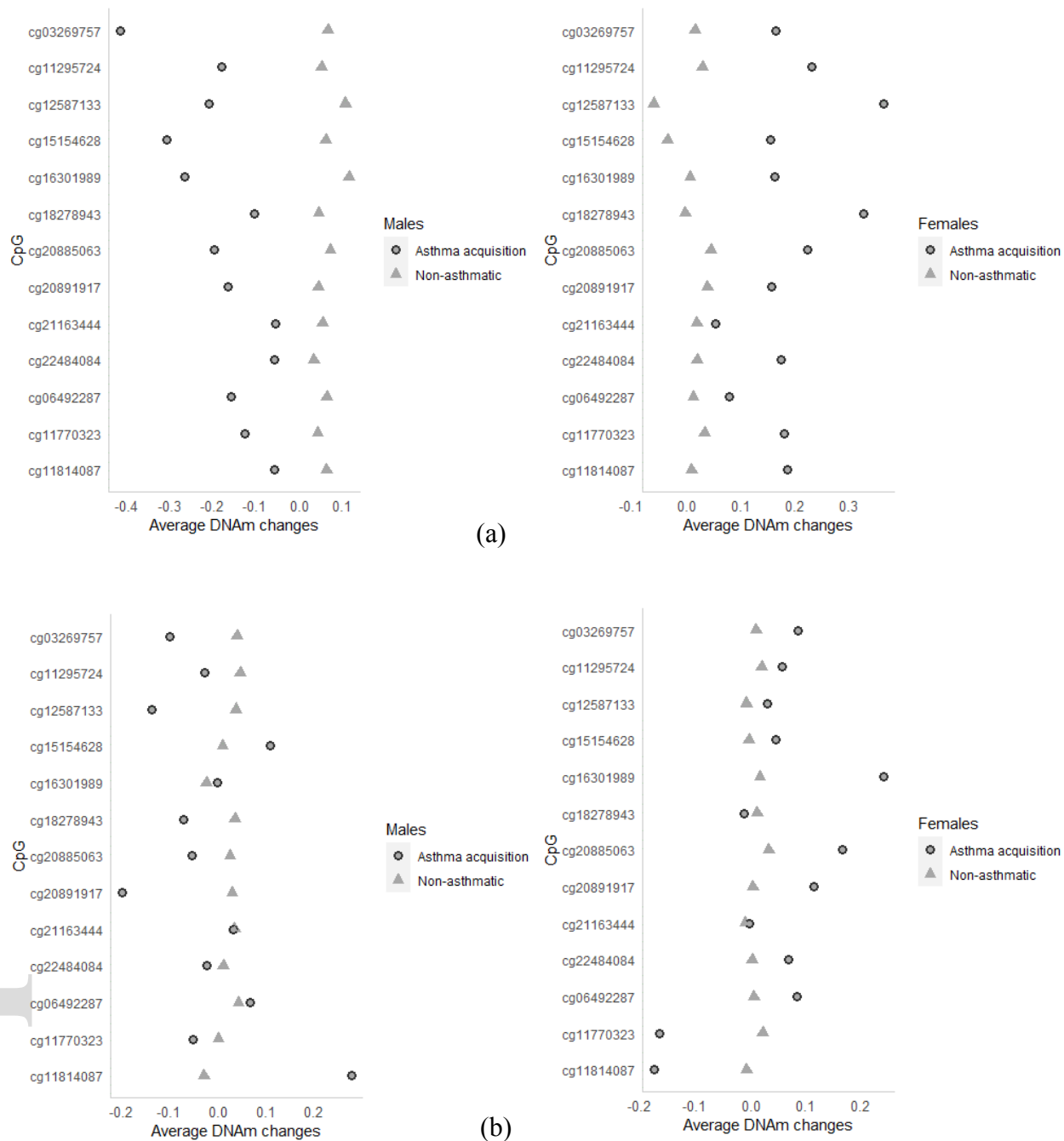


Figure 2: Scatter plots showing average DNAm changes from pre-adolescence to late-adolescence, stratified by sex. (a) Average DNAm changes between 10 and 18 years in IOWBC. (b) Average DNAm changes between 7 and 15 or 17 years in the ALSPAC cohort. In both panels, left figure is for males, and right for females.

Sex-specific associations of asthma acquisition with changes in DNA methylation during
adolescence

Rutu Patel, MPH, MBBS^{1*}, Farnaz Solatikia, MPH, MS^{1,2*}, Hongmei Zhang, PhD, MS^{1**},
Alemayehu Wolde, MS², Latha Kadalayil PhD³, Wilfried Karmaus, MD, Dr.med., MPH¹, Susan
Ewart, DVM, PhD, DACVIM⁴, Ryan Arathimos, PhD^{5,6,7,8}, Caroline Relton, PhD^{5,6,7}, Susan
Ring, PhD, BSc^{5,6,7}, A. John Henderson⁶, MD, S. Hasan Arshad, MBBS, DM, FRCP^{9,10,11}, John
W Holloway, PhD^{3,11}

¹Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health,
University of Memphis, Memphis, TN, USA.

²Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA.

³Human Development and Health, Faculty of Medicine, University of Southampton,
Southampton, UK

⁴College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA.

⁵MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

⁶Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

⁷National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol
and University Hospitals Bristol NHS Foundation Trust, Bristol, UK.

⁸Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry Psychology and
Neuroscience, King's College London, London, UK

⁹Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton,
Southampton, UK.

¹⁰David Hide Asthma and Allergy Research Centre, Isle of Wight, UK.

¹¹NIHR Southampton Biomedical Research Centre, University Hospital Southampton,
Southampton, UK

*: equal contribution, **: corresponding author

Corresponding author: Dr. Hongmei Zhang

Phone: 901.678.4707

Email: hzhang6@memphis.edu

Fax: 901.678.1715

Office: 224 Robison Hall, 3825 DeSoto Avenue, Memphis, TN 38152-0001

Supplemental Material

S1. Quality control and pre-processing genome-wide DNA methylation data

A standard salting out procedure was used to extract DNA from peripheral blood samples collected at ages 10 and 18 years [1]. DNA concentration was determined by PicoGreen dsDNA quantitation (Molecular Probes, INC. OR, USA) or Qubit (ThermoFisher, MA, USA). One microgram of DNA was bisulfite-treated for cytosine to thymine conversion using the EZ 96-DNA Methylation Kit (Zymo Research, Irvine, CA, USA) for each sample. Arrays were processed using a standard protocol with multiple identical control samples assigned to each bisulfite conversion batch to assess assay variability. Methylation level at >484,000 and >850,000 CpG sites was assessed using Infinium HumanMethylation450 BeadChips or Infinium MethylationEPIC BeadChips (Illumina, Inc., San Diego, CA), respectively. Probes not reaching a detection p-value of 10^{-16} in at least 95% of samples were excluded, and the same criterion was applied to the exclusion of a sample. CpGs on the sex chromosomes were excluded from analyses to avoid bias.

DNAm data were preprocessed using the CPACOR pipeline [2]. Specifically, DNAm intensity data were quantile-normalized using the R package, *minfi* [3]. Beta values were calculated representing proportions of intensity of methylated (M) over the sum of methylated and unmethylated (U) probes ($\beta = M/[c + M + U]$, where c is a constant to prevent zero in the denominator if $M + U$ is too small). Beta values close to 0 or 1 tend to suffer from severe heteroscedasticity. Therefore base-2 logit transformed beta values (denoted as M values) were used in analyses [4].

Principal components (PCs) inferred based on control probes to represent latent chip-to-chip and technical variations were inferred. Since DNAm data were from two different platforms (450K and EPIC), PCs were determined based on DNAm at 195 control probes shared between the platforms; the top 15 control probe PCs were used to represent latent batch factors [2]. These 15 PCs were included in subsequent analyses. Probes that contained single nucleotide polymorphisms (SNPs) within 10 base pairs of a targeted CpG site with a minor allele in more than 0.7% subjects (corresponding to at least 10 subjects in IOWBC with $n = 1,456$) were excluded due to their potential influence on DNAm measurement. After preprocessing, a total of 439,635 CpGs in common between the two platforms were included in the analyses.

S2. Genome-wide RNA-seq gene expression data generation

Gene expression levels from peripheral blood samples collected at 26 years from the IoW cohort was determined using paired-end (2*75 bp) RNA sequencing with the Illumina Tru-Seq Stranded mRNA Library Preparation Kit with IDT for Illumina Unique Dual Index (UDI) barcode primers following the manufacturer's recommendations. All samples were sequenced twice using the identical protocol and for each sample the output from both runs were combined. FASTQC were run to assess the quality of the FASTQ files

(<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped against Human Genome (GRch37 version 75) using HISAT2 (v2.1.0) aligner [5]. The alignment files, produced in the Sequence Alignment Map (SAM) format, were converted into the Binary Alignment Map (BAM) format using SAMtools (v1.3.1) [6]. HTseq (v0.11.1) was used to count the number of reads mapped to each gene in the same reference genome used for alignment [7]. Normalized read count FPKM (Fragments Per Kilobase of transcript per Million mapped reads) were calculated using the countToFPKM package (<https://github.com/AAlhendi1707/countToFPKM>) and their log transformed values were used for data analysis.

S3. Screening asthma-associated CpG sites

An R package, *ttScreening*, was implemented to screen for CpGs whose methylation (in M values) was associated with asthma cross-sectionally [8]. This method utilizes training and testing data in robust linear regressions with surrogate variables included to adjust for unknown (latent) factors. Screening was performed separately for each sex at ages 10 and 18 years. Subjects with DNAm and asthma data at one or both ages were included in the screening. The minimum frequency of selecting an informative CpG sites was set at 50%, i.e., a CpG site gained statistical significance in at least 50% of the randomly selected training and testing data set pairs. CpGs that passed the screening at either age for one or both sexes were considered to be potentially associated with asthma and were included as candidate CpGs in subsequent analyses. This approach of screening was cross-sectional and focused on asthma status rather than asthma transition, to avoid data double dipping, i.e., avoiding using the same or a very similar model in screening as well as in final data analyses with the same data.

S4. The replication cohort, ALSPAC

DNAm in the ALSPAC cohort was assessed using the Infinium HumanMethylation450 BeadChip. DNAm data on 1,018 offspring in the ALSPAC cohort were available at ages 7 and 15 or 17 years [9]. The pre-processing of DNAm was performed by correcting for batch effects using the *minfi* package [3] and removing CpGs with detection p-value ≥ 0.01 . Samples were flagged that contained sex-mismatch based on X-chromosome methylation. Estimated cell type proportions of CD4⁺ T cells, natural killer cells, neutrophil, B cells, monocytes, and granulocytes cells were used in the analyses to adjust for cell heterogeneity. DNAm changes between pre- and post-adolescence were calculated in a similar way to that in IOWBC, except that PCs were not utilized in the ALSPAC cohort for batch effect adjustment.

Asthma acquisition was defined as no asthma at age 7 but asthma at age 15 years. Identical logistic regression models were used with comparable covariates (as those in IOWBC) available in ALSPAC, including maternal asthma, sex, birth weight, socio-economic status, and atopy status at 7 years, and relative changes of height and BMI from 7 to 15 or 17 years.

References

1. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells*. Nucleic Acids Res, 1988. **16**(3): p. 1215.
2. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies*. Genome Biol, 2015. **16**: p. 37.
3. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. Bioinformatics, 2014. **30**(10): p. 1363-9.
4. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis*. BMC Bioinformatics, 2010. **11**: p. 587.
5. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements*. Nat Methods, 2015. **12**(4): p. 357-60.
6. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
7. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data*. Bioinformatics, 2015. **31**(2): p. 166-9.

8. Ray, M.A., et al., *An Efficient Approach to Screening Epigenome-Wide Data*. Biomed Res Int, 2016. **2016**: p. 2615348.
9. Relton, C.L., et al., *Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)*. International Journal of Epidemiology, 2015. **44**(4): p. 1181-1190.