Newborn DNA methylation and asthma acquisition across adolescence and early adulthood

Liang Li1, John W. Holloway2, Susan Ewart3, S. Hasan Arshad4,5, Caroline L. Relton6, Wilfried Karmaus1, Hongmei Zhang1*

Affiliations: 1 Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN, USA. 2 Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK. 3 College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA. 4 Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK. 5 The David Hide Asthma and Allergy Research Centre, St Mary’s, Hospital, Parkhurst Road, Newport, Isle of Wight PO30 5TG, UK. 6 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK.

*Correspondence: Hongmei Zhang, Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN, USA. E-mail: hzhang6@memphis.edu

Abstract

Background: Little is known about the association of newborn DNA methylation (DNAm) with asthma acquisition across adolescence and early adult life.

Objective: We aim to identify epigenetic biomarkers in newborns for asthma acquisition during adolescence or young adulthood.

Methods: The Isle of Wight Birth Cohort (IOWBC) (n=1456) data at ages 10, 18, and 26 years were assessed. To screen Cytosine-phosphate-Guanine site (CpGs) potentially associated with asthma acquisition, at the genome-scale, we examined differentially methylated regions (DMR) using dmrff R package and individual CpG sites using linear regression on such associations. For CpGs that passed screening, we examined their enrichment in biological pathways using their mapping genes and tested their associations with asthma acquisitions using logistic regressions. Findings in IOWBC were tested in an independent cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort.

Results: In total, 2636 unique CpGs passed screening, based on which we identified one biological pathway linked to asthma acquisition during adolescence in females (FDR adjusted p-value=0.003 in IOWBC). Via logistic regressions, for females, 4 CpGs were shown to be associated with asthma acquisition during adolescence, and another 4 CpGs with asthma acquisition in young adulthood (FDR adjusted p-value<0.05 in IOWBC).
and these 8 CpGs were replicated in ALSPAC (all p-values<0.05). DNAm at all the identified CpGs was shown to be temporally consistent, and at 6 of the CpGs was associated with expressions of adjacent or mapping genes in females (all p-values<0.05). For males, 622 CpGs were identified in IOWBC (FDR=0.01), but these were not tested in ALSPAC due to small sample sizes.

**Conclusion and clinical relevance:** Eight CpGs on LHX5, IL22RA2, SOX11, CBX4, ACPT, CFAP46, MUC4, and ATP1B2 genes have the potential to serve as candidate epigenetic biomarkers in newborns for asthma acquisition in females during adolescence or young adulthood.

**KEYWORDS:** IOWBC, ALSPAC, DNA methylation, Epigenome-wide, Asthma acquisition.

**INTRODUCTION**

Asthma is the most prevalent non-communicable disease in children and the second most prevalent chronic respiratory disease worldwide, with an estimated 300 million people affected and increasing disease burden in the next few decades. Asthma developmental patterns (or asthma transitions when focusing on status change) involve childhood-onset asthma persistent until adulthood, asthma remission, asthma relapse, and adult-onset asthma. These different developmental patterns are presumably linked to diverse underlying pathophysiological processes. Sex-specific incidence of asthma occurs throughout life. In particular, females have more new-onset asthma during adolescence, while having fewer asthma occurrences during childhood compared to males. However, to our knowledge, only a few studies focused on asthma patterns from childhood to adulthood with a substantial knowledge gap for asthma status transition from pre- to post-adolescence. Furthermore, adulthood-onset asthma is more likely to have a poorer treatment response and worse prognosis compared to childhood-onset asthma, emphasizing the importance of studies contributing to new-onset of asthma during adulthood.

Risk factors associated with childhood-onset asthma are generally different from adulthood-onset asthma. Evidence shows that genetic predisposition plays a stronger role in asthma onset in early childhood with genetic risk factors for adult-onset asthma largely a subset of the genetic risk for childhood-onset asthma but with smaller effect sizes, suggesting a larger role for non-genetic risk factors such as environment in adult-onset asthma. Prospective studies have shown that exposure to perennial allergens or to prenatal smoke are associated with a high risk to childhood-onset asthma. In addition, viral infection and bacterial colonization of the airway also increase the risk of childhood-onset asthma. For onset of asthma in adults, other risk factors such as irritant exposure to environmental pollutants, female sex hormones, upper airway disease, obesity and stress have shown to play a role. Environmental exposures and developmental changes (e.g., puberty) associated with new
onset of asthma in adulthood have been shown to modify the epigenome. DNA methylation (DNAm) is one epigenetic mechanism that is established in utero and modified by environmental exposure and aging, which regulates gene expression and alters pathophysiological process of disease.

Evidence has shown that DNAm patterns in blood-derived DNA are associated with the status of asthma. However, the majority of studies have focused on childhood. Some recent studies showed that DNA methylation in cord blood was associated with asthma status in childhood. Nevertheless, it is unknown whether epigenetic factors at birth were associated with asthma acquisition at different stages of life from childhood to adulthood. Findings from such studies will be beneficial to future studies on predicting asthma acquisition at a much earlier age of life.

This study aimed to identify CpGs in blood at birth where DNAm at those sites was associated with asthma acquisition at different stages of life (from pre-adolescence to young adulthood) using sex-stratified epigenome-wide association studies (EWAS), in two birth cohorts, the Isle of Wight birth cohort (IOWBC; the discovery cohort) and the Avon Longitudinal Study of Parents and Children (ALSPAC; the replication cohort). We hypothesized that newborn DNA methylation was associated with asthma acquisition patterns from pre- to post-adolescence and to adulthood. For the identified CpGs, we further assessed their biological relevance by examining their association with expression of their mapping genes.

METHODS

Discovery cohort - IOWBC

The Isle of Wight Birth Cohort (IOWBC) is a population-based birth cohort established in 1989-1990 on the Isle of Wight, an island off the south coast of England. A total of 1456 eligible children with parental consent were included in the cohort (IOWBC-F1 generation). We focused on subjects in the F1 generation with DNAm data available in newborns and who were asthma free at age 10 years (some of them developed asthma at age 18 years or 26 years). Gene expression data in the F2 generation (children of the IOWBC-F1 women and partners of F1-males).

Asthma acquisition

Asthma information at ages 10, 18, and 26 years was collected using International Study of Asthma and Allergies in Childhood questionnaires (ISAAC) with a detailed assessment of asthma symptoms and treatment. Asthma was defined by the core questions as: “ever had asthma” and either “wheezing or whistling in the chest during the previous 12 months” or “current treatment for asthma”. Two groups of asthma acquisition
were defined based on asthma status at ages 10, 18 and 26 years. Asthma acquisition during adolescence was defined as asthma-free at age 10 and having asthma at ages 18 and 26 years (No-Yes-Yes). Asthma acquisition during young adulthood was defined as asthma-free at ages 10 and 18 years and acquiring asthma by age 26 years (No-No-Yes). A transition pattern involving asthma remission was not considered in the present study, i.e., the pattern of No-Yes-No. Participants without asthma at ages 10, 18, and 26 years (No-No-No) were included as the reference group of no asthma.

**DNA methylation (DNAm) and cell type composition**

In the F1 generation of IOWBC, a peripheral blood sample from a heel prick was collected on Guthrie cards within the first seven days of age (n=796). DNAm was measured using MethylationEPIC BeadChips. Details of DNAm assessment and preprocessing are in Supplemental Material Methods (S1). Base 2 logit-transformed DNAm β values (M-values) were used in all the analyses.

DNA methylation is influenced by the heterogeneity of cell-type compositions in cord blood. To control the influence of cell-type proportions, six cell types, CD4+ T cells, natural killer cells, neutrophil, B cells, monocytes, and eosinophils, were estimated using the R package minfi and included in the analyses as confounders.

**Confounders**

Confounding variables potentially associated with DNAm and asthma status included socioeconomic status (SES), secondhand smoke in childhood, breastfeeding duration, and body mass index (BMI) at age 10 years. SES was ascertained as ‘low’, ‘medium’, and ‘high’ according to the assessment of parental occupation, level of household income, and the number of rooms in the house. Secondhand smoke information at childhood was collected from their parents and either of them with smoking was regarded as secondhand smoke to children. Breastfeeding duration data were obtained at the 1- and 2-year follow-ups.

**Gene expression in IOW Cohort F2 generation**

Gene expression data in the F1 generation at birth was not available. Thus, to assess biological functionality of identified CpGs, DNAm and gene expression in the IOW F2 cohort, both assessed at birth, were utilized. In the F2 generation, RNA was isolated from cord blood samples (n=161) that were collected into PAXgene Bone Marrow RNA kits according to the manufacturer’s instructions (Qiagen, Valencia, CA, USA). RNA quality
was determined using the Agilent 2100 BioAnalyzer system. Gene expression levels were measured using SurePrint G3 Human GE v2 8×60K Agilent Microarray (Agilent Technologies). The details on measurement and quality control of gene expression level are described elsewhere 30.

Replication cohort – the ALSPAC cohort

The Avon Longitudinal Study of Children and Parents (ALSPAC) is a prospective population-based study established in the Bristol area of the United Kingdom during 1990-92 31-33 (details are in the Supplemental Material S2). DNAm in cord blood (n=861) and asthma status at ages 10 (n=7796), 17 (n=5126), and 22 years (n=993), along with related confounders comparable to those in IOWBC, were analyzed 34. Procedures for collection of cord blood samples and DNA sample preparation are described elsewhere 35,36. DNA methylation in cord blood was measured using the Illumina Infinium HumanMethylation 450K BeadChip assay with over 485,000 CpG sites 37. Cell-type proportions for cord blood at birth were estimated using the Houseman method 38 with cord blood reference panel 39.

Statistical analysis

Statistical analyses in IOWBC

All analyses were separated by sex due to different asthma acquisition patterns from pre-adolescence through young adults in males and females 40. Analyses were performed using SAS 9.4 (SAS, Cary, NC, USA) and R packages. To examine whether participants included in the current study reasonably represented the complete study cohort, we evaluated the variables potentially related to asthma acquisition in both sexes. Categorical variables were presented as percentages compared to the complete cohort statistics using a one-sample proportion test. Continuous variables were presented as means and standard deviations and a one-sample t-test was applied to compare the subsamples with the complete cohort. For variables not following normal distribution, median and median absolute deviation (MAD) were present and signed rank tests were applied. A p-value <0.05 was deemed to be statistically significant.

We screened CpGs potentially associated with asthma acquisition from two directions. For both directions, the screening was stratified by sex. In the first direction, a linear regression was applied to each CpG site to assess the association of DNAm (as the dependent variable) with asthma transition status with no asthma at three ages (10, 18, and 26) as the reference group. A CpG with a p-value <0.001 (to be more stringent) was regarded as a candidate CpG. In addition, we applied the dmrff package in R to detect differentially methylated regions (DMRs) with respect to asthma acquisition that has 2 or more CpG sites within 500bp with multiple testing controlled by the Bonferroni approach at experiment-wise
p-value <0.05. CpGs in identified DMRs and CpGs that passed screening via linear regression were included in subsequent analyses for their association with adolescence and young adulthood asthma acquisition using logistic regressions, adjusting for the confounders in both sexes. CpGs in each identified DMR was treated as one CpG unit and the unit’s DNAm was represented by the mean of DNAm at those CpGs. Multiple testing was adjusted by controlling the false discovery rate (FDR) of 0.01, a stringent control of multiple testing in an effort to improve the informativity of identified CpG sites.

Replication analyses in the ALSPAC cohort

CpGs identified in the IOWBC were further tested in the ALSPAC cohort using logistic regression models controlling for comparable covariates. The asthma diagnoses were comparable with those in IOWBC.

Consistency of DNAm over time

The temporal consistency of DNAm would substantiate that asthma acquisition later in life is related to earlier epigenetics at birth. To this end, we applied linear mixed models with repeated measures. Dependent variables were DNAm at age 10 and 18 years, and independent variables were newborn DNAm, age (10 and 18 years), newborn DNAm and age interaction, and sex. In IOWBC, DNAm of newborns (n=796), age 10 (n=330), and age 18 (n=476) was included in the analyses. For ALSPAC, DNAm at birth (n=883), age 7 (n=927), and age 17 (n=688) was analyzed. CpGs with no statistically significant interactions between DNAm and age (p-value >0.05) were regarded as temporally consistent CpGs.

Gene set analysis

To address biological pathways of genes corresponding to the CpGs associated with asthma acquisition in screening via linear regression or DMRs analyses, the gometh function ([https://rdrr.io/bioc/missMethyl/man/gometh.html](https://rdrr.io/bioc/missMethyl/man/gometh.html)) in the missMethyl R package was applied. The multiple testing correction was conducted at the FDR of 0.05 level.

Biological relevance

For CpGs associated with asthma acquisition in both cohorts, we evaluated the association of newborn DNAm in M values with gene expression (in log-scale) in cord blood on genes within 250kb upstream and downstream of the identified CpGs, following Reese et al. Linear regression was applied to assess their relationships. Statistical significance was set at 0.05.
RESULTS

Population characteristics

Cohort members who were asthma free by age 10 years and had newborn DNAm available were included in the study (n=453), of which, 28 developed asthma in adolescence (and persistent in young adulthood; No-Yes-Yes) and 17 participants developed asthma at young adulthood (No-No-Yes) (Table 1). We compared the analytical sample with the cohort members having the defined acquisition patterns (“Whole cohort”; Table 1) on variables potentially related to asthma. No statistically significant difference was found (Table 1). Furthermore, the proportions of acquisition showed no difference between the analytical sample and the whole cohort for asthma acquisition (Table 2). In addition, we assessed maternal status and asthma status at earlier ages (at ages 4, 10, 18 years) between those remaining in the cohort and those lost to follow-up. No differences were found (Table S1).

TABLE 1. Asthma acquisition between the sub-sample and cohort and between the two sexes.

<table>
<thead>
<tr>
<th></th>
<th>Asthma acquisition during adolescence</th>
<th>p-value</th>
<th>Asthma acquisition during young adulthood</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whole cohort*</td>
<td>Analytical sample*</td>
<td>Whole cohort*</td>
<td>Analytical sample*</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (3.96)</td>
<td>9 (4.81)</td>
<td>14 (4.26)</td>
<td>8 (4.30)</td>
</tr>
<tr>
<td>No</td>
<td>315 (96.04)</td>
<td>178 (95.19)</td>
<td>315 (95.74)</td>
<td>178 (95.70)</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Yes</td>
<td>30 (7.41)</td>
<td>19 (7.63)</td>
<td>16 (4.09)</td>
<td>9 (3.77)</td>
</tr>
<tr>
<td>No</td>
<td>375 (92.59)</td>
<td>230 (92.37)</td>
<td>375 (95.91)</td>
<td>230 (96.23)</td>
</tr>
</tbody>
</table>
| p-value        | 0.05                                 | 0.23    |                                          | 0.91    | 0.78

*The analytical samples include subjects with newborn DNAm available as well as asthma status data up to young adulthood. The whole cohort samples include subjects with asthma status data up to young adulthood.

TABLE 2. Characteristics of subjects with available methylation data with their asthma status at ages 10, 18, and 26 years stratified by sex in the IOW cohort.
## Data for asthma acquisition during adolescence vs young adulthood

<table>
<thead>
<tr>
<th>Factors</th>
<th>Whole cohort</th>
<th>Analytical sample</th>
<th>p-value</th>
<th>Whole cohort</th>
<th>Analytical sample</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=328</td>
<td>n=187</td>
<td></td>
<td></td>
<td>n=329</td>
<td>n=186</td>
<td></td>
</tr>
<tr>
<td>SES$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>33 (10.06)</td>
<td>16 (8.56)</td>
<td>0.48</td>
<td>32 (9.73)</td>
<td>15 (8.06)</td>
<td>0.35</td>
</tr>
<tr>
<td>Middle</td>
<td>247 (75.3)</td>
<td>142 (75.94)</td>
<td></td>
<td>248 (75.38)</td>
<td>141 (75.81)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>39 (11.89)</td>
<td>27 (14.44)</td>
<td></td>
<td>39 (11.85)</td>
<td>28 (15.05)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>9 (2.74)</td>
<td>2 (1.07)</td>
<td></td>
<td>10 (3.04)</td>
<td>2 (1.08)</td>
<td></td>
</tr>
<tr>
<td>Secondhand smoke$^b$</td>
<td></td>
<td></td>
<td>0.85</td>
<td></td>
<td></td>
<td>0.88</td>
</tr>
<tr>
<td>Yes</td>
<td>112 (34.15)</td>
<td>68 (36.36)</td>
<td></td>
<td>115 (34.95)</td>
<td>68 (36.56)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>181 (55.18)</td>
<td>101 (54.01)</td>
<td></td>
<td>180 (54.71)</td>
<td>101 (54.3)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>35 (10.67)</td>
<td>18 (9.63)</td>
<td></td>
<td>34 (10.33)</td>
<td>17 (9.14)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI at age 10 years</strong></td>
<td>17.56±2.49</td>
<td>17.67±2.59</td>
<td>0.59</td>
<td>17.53±2.42</td>
<td>17.64±2.47</td>
<td>0.55</td>
</tr>
<tr>
<td>Breastfeeding$^c$ (weeks)</td>
<td>12 (12)</td>
<td>12.5 (12.5)</td>
<td>0.31</td>
<td>12 (11.5)</td>
<td>12 (12)</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=405</td>
<td>n=249</td>
<td></td>
<td></td>
<td>n=391</td>
<td>n=239</td>
<td></td>
</tr>
<tr>
<td>SES$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>39 (9.63)</td>
<td>25 (10.04)</td>
<td>0.63</td>
<td>38 (9.72)</td>
<td>25 (10.46)</td>
<td>0.39</td>
</tr>
<tr>
<td>Middle</td>
<td>300 (74.07)</td>
<td>192 (77.11)</td>
<td></td>
<td>292 (74.68)</td>
<td>186 (77.82)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>59 (14.57)</td>
<td>30 (12.05)</td>
<td></td>
<td>54 (13.81)</td>
<td>27 (11.30)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>7 (1.73)</td>
<td>2 (0.80)</td>
<td></td>
<td>7 (1.79)</td>
<td>1 (0.42)</td>
<td></td>
</tr>
<tr>
<td>Secondhand smoke$^b$</td>
<td></td>
<td></td>
<td>0.20</td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>Yes</td>
<td>144 (35.56)</td>
<td>74 (29.72)</td>
<td></td>
<td>135 (34.53)</td>
<td>68 (28.45)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>212 (52.35)</td>
<td>136 (54.62)</td>
<td></td>
<td>208 (53.2)</td>
<td>133 (55.65)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>49 (12.10)</td>
<td>39 (15.66)</td>
<td></td>
<td>48 (12.28)</td>
<td>38 (15.9)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI at age 10 years</strong></td>
<td>18.58±3.08</td>
<td>18.49±2.97</td>
<td>0.99</td>
<td>18.5±3</td>
<td>18.41±2.87</td>
<td>0.66</td>
</tr>
<tr>
<td>Breastfeeding weeks</td>
<td>11.5 (11.5)</td>
<td>12 (12)</td>
<td>0.56</td>
<td>11.5 (11.5)</td>
<td>12 (12)</td>
<td>0.59</td>
</tr>
</tbody>
</table>
#: SES refers to socioeconomic status in childhood. §: Participants got secondhand smoke before age 10 years. ¥: The length of time for breastfeeding.

Screening CpGs

The screening of CpGs was conducted for males and females separately as patterns of asthma acquisition in adolescence differ between males and females. We first screened CpGs using linear regression with DNAm at each CpG site as the dependent variable and asthma acquisition status as the independent variable. In total, 900 unique CpGs (463 CpGs for males) for their association with asthma acquisition during adolescence and 1738 unique CpGs (1117 CpGs for males and 624 in females with 3 common CpGs) for asthma acquisition during young adulthood passed screening (Figure 1).

Second, we used DMR-based screening. We identified one statistically significant DMR (Bonferroni adjusted p-value=0.042) showing association with young adulthood asthma acquisition in females. This DMR (chromosome 10: 3143354-3143596) spanned two CpG sites (cg09875661, cg16519433) that are annotated to the PFKP gene. These two CpGs were treated as one CpG unit. CpG cg09875661 was among the 1738 candidate CpGs identified through regression. Thus, for young adulthood asthma acquisition, 1738 CpGs/units were included in the full model of logistic regression.

Discovered CpGs in IOWBC

Logistic regression models were applied to evaluate the association of DNAm of the candidate CpGs (including the CpG unit) at birth with each of the two types of asthma acquisition patterns. For males, 174 and 448 CpGs showed statistically significant association with asthma acquisition during adolescence and young adulthood, respectively, after adjusting for multiple testing by controlling the FDR of 0.01 (Figure 1; Table S2-S3). For females, at 205 CpGs and at 255 CpGs plus the CpG unit, DNAm was associated with asthma acquisition during adolescence and young adulthood, respectively, at FDR=0.01. These CpGs identified in IOWBC were further tested in the ALSPAC cohort. No overlap in the identified CpGs was found between males and females.
Newborn DNA methylation and asthma transition at age 10, 18, and 26 years
(551,710 CpGs)

Screening: Linear regression and dmrff

Adolescence asthma acquisition (Males: 463 CpGs; Females: 437 CpGs)
Young adulthood asthma acquisition (Males: 1117 CpGs; Females: 625 CpGs)

Statistical significance in logistic regression

CpGs available in ALSPAC

Adolescence asthma acquisition (Males:174 CpGs; Females: 205 CpGs)
Young adulthood asthma acquisition (Males: 448 CpGs; Females: 256 CpGs$)

Logistic regression in females

Adolescence and young adulthood asthma acquisition: 54 and 66 CpGs had the consistent
direction of the association, respectively
Adolescence and young adulthood asthma acquisition: 4 and 4 CpGs
were replicated, respectively

* The data in males were not further analyzed in ALSPAC due to small sample sizes. For females, n=196 with 6
having adolescence asthma acquisition and n=194 with 4 having young adulthood asthma acquisition. ² 255
CpG and 1 CpG unit.

FIGURE 1. Flow chart of statistical analyses identifying the association of newborn DNAm with asthma
acquisition.
Replication in ALSPAC

In ALSPAC, for females, 10 participants experienced asthma acquisition during adolescence or during young adulthood, and 190 participants were without asthma at three-time points. For males, the data were not further analyzed due to no participant with asthma acquisition during adolescence and only one young adulthood asthma acquisition participant. For females, of the 205 CpG identified in IOWBC associated with asthma acquisition during adolescence, DNAm at 108 CpGs was available in ALSPAC. Similarly, DNAm at 141 (including the CpG unit) of the 256 discovered CpGs/units for asthma acquisition during young adulthood was available in ALSPAC (Figure 1). For asthma acquisition during adolescence, of the 108 CpGs examined, 54 (50%) CpGs showed consistent directions of association with those in IOWBC (Table S4, Figure 2). Of the 54 CpGs, associations at 4 CpGs were statistically significant (p<0.05) (Table 3, Figure 2); an increase in DNAm at 3 of the 4 CpGs was associated with decreased odds of adolescence asthma acquisition. For young adulthood asthma acquisition, at 66 (47%) of the 141 CpGs/units, consistent directions of association were observed between IOWBC and ALSPAC (Table S5, Figure S1) and at 4 of the 66 CpGs/units statistical significance was observed (p<0.05) (Table 3, Figure S1). At one of these 4 CpGs, increased DNAm was associated with decreased odds of young adulthood asthma acquisition.

TABLE 3. CpG sites with DNAm showing associations with adolescence asthma acquisition and young adulthood asthma acquisition in females that are consistent between the discovery cohort, IOWBC and the replication cohort, ALSPAC, with p-value<0.05. Complete results are in Table S4 and S5.

<table>
<thead>
<tr>
<th>Asthma acquisition</th>
<th>CpG name</th>
<th>Chr.</th>
<th>Gene name</th>
<th>Region</th>
<th>IOWBC Log-odds ratio a</th>
<th>Raw p-value</th>
<th>FDR p-value</th>
<th>ALSPAC Log-odds ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>In adolescence</td>
<td>cg11586330</td>
<td>12</td>
<td>LHX5</td>
<td>3'UTR</td>
<td>-2.06</td>
<td>0.0018</td>
<td>0.0076</td>
<td>-1.91</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>cg23507945</td>
<td>6</td>
<td>IL22RA2</td>
<td>Body</td>
<td>-0.35</td>
<td>0.0036</td>
<td>0.0089</td>
<td>-1.19</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>cg04771413</td>
<td>2</td>
<td>SOX11</td>
<td>Intergenic</td>
<td>-2.44</td>
<td>0.0016</td>
<td>0.0074</td>
<td>-1.34</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>cg00646381</td>
<td>17</td>
<td>CBX4</td>
<td>Intergenic</td>
<td>3.47</td>
<td>0.0027</td>
<td>0.0079</td>
<td>2.22</td>
<td>0.018</td>
</tr>
<tr>
<td>In young adulthood</td>
<td>cg15782391</td>
<td>19</td>
<td>ACPT</td>
<td>TSS1500</td>
<td>5.32</td>
<td>0.0012</td>
<td>0.0128</td>
<td>2.35</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>cg12319143</td>
<td>10</td>
<td>CFAP46</td>
<td>Intergenic</td>
<td>-0.96</td>
<td>0.0019</td>
<td>0.0128</td>
<td>-0.59</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>cg19983801</td>
<td>3</td>
<td>MUC4</td>
<td>5'UTR;1stExon</td>
<td>3.62</td>
<td>0.0006</td>
<td>0.0128</td>
<td>1.73</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>cg02705837</td>
<td>17</td>
<td>ATP1B2</td>
<td>1stExon;5'UTR</td>
<td>1.79</td>
<td>0.0025</td>
<td>0.0128</td>
<td>2.48</td>
<td>0.048</td>
</tr>
</tbody>
</table>
# Chromosome number. &: Log-odds ratios assessing the associations of DNAm level at birth with asthma acquisition, with no asthma at age 10, 18, 26 years as the reference.

**Consistency of DNAm over time**

Of the 8 CpGs replicated in ALSPAC (p< 0.05), cg23507945 was not available at ages 10 and 18 in IOWBC and was not available at ages 7 and 17 in ALSPAC. Results from the linear mixed models showed that the interaction between time and newborn DNAm was not statistically significant (all p-values>0.05; Table S6, Figure S2-S8), indicating DNAm at these CpG sites was likely to be consistent from birth to age 18 years.

**Gene set analysis**

Gene set analysis was applied to CpGs that survived screening (2636 unique CpGs) (Table S7-S11). One biological pathway, KEGG pathway Malaria, linked to adolescence asthma acquisition in females was identified (FDR adjusted p-value=0.003) (Table S12), and seven genes, COMP (chromosome [Chr.] 19), IL6 (Chr. 7), LRP1 (Chr. 12), MET (Chr. 7), TGFβ2 (Chr. 1), THBS4 (Chr. 5), and TNF (Chr. 6) on our candidate gene lists were involved in this pathway. These seven genes do not overlap with the genes identified (Table 3, Table S13).

**DNAm and gene expression**

For asthma acquisition during adolescence, DNAm at 2 of the 4 identified CpGs (cg00646381 and cg23507945; all p-values <0.05) was associated with expression of their neighboring genes (Table 4 and Table S14). For asthma acquisition during young adulthood, DNAm at all the 4 identified CpGs (all p-values <0.05) was associated with expression of 8 neighboring genes (Table 4 and Table S15).

**TABLE 4. Association of DNAm with the gene expression on the genes within 250kb upstream and downstream of the CpGs at birth in F2-generation.** The complete results on the association of DNAm at the identified CpGs with their neighboring genes are in Tables S6a and S6b.

<table>
<thead>
<tr>
<th>Asthma acquisition</th>
<th>CpG name</th>
<th>Gene name</th>
<th>Gene probe</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adolescence asthma acquisition</td>
<td>cg23507945</td>
<td>NHEG1</td>
<td>NHEG1_A_21_P0000551</td>
<td>-0.14</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>cg00646381</td>
<td>TBC1D16</td>
<td>TBC1D16_A_33_P3241937</td>
<td>-0.51</td>
<td>0.031</td>
</tr>
<tr>
<td>Young adulthood asthma acquisition</td>
<td>cg15782391</td>
<td>KLK6</td>
<td>KLK6_A_24_P236935</td>
<td>-0.80</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>cg12319143</td>
<td>GPR123</td>
<td>Inc_GPR123_1_A_21_P00006942</td>
<td>-0.15</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>cg19983801</td>
<td>LINC00969</td>
<td>LINC00969_A_19_P00316857</td>
<td>-0.15</td>
<td>0.017</td>
</tr>
</tbody>
</table>
### DISCUSSION

Previous studies have been focusing on the connection of newborn DNAm with childhood asthma \(^{20-24}\). To our knowledge, this is the first epigenome-scale association study on newborn DNAm in whole blood with asthma acquisition at different stages of life. Via single CpG site screening and screening via DMRs, we first identified potentially informative CpGs with respect to asthma acquisition during adolescence or young adulthood in IOWBC, and these CpGs were in-depth tested for such associations using logistic regressions with confounders adjusted. Four CpGs unique to each period discovered in females in IOWBC were replicated in ALSPAC. Our assessment indicated that DNAm at these eight CpG sites was consistent from birth to age 18 years and DNAm at six of the eight CpGs were associated with expressions of their neighboring genes.

Findings in our screening process deserve discussion. One DMR composed of two CpGs related to young adulthood asthma acquisition in females was detected. However, one of the two CpGs, cg16519433, did not show an association with asthma transition by itself. This observation suggests that methylation in multiple adjacent CpGs might have a joint biological association with the diseases \(^{43,44}\), and future studies may focus on investigating CpGs by regions rather than individually. Furthermore, evidence has shown that sex-specific acquisition of asthma occurs during puberty \(^{45}\), and our previous study showed that the association of DNAm changes with asthma acquisition from pre- to post-adolescence was different between males and females \(^{46}\). These findings support our sex-specific screening results and our sex-stratified study design.
In addition, among the genes with DNAm at CpGs that passed screening for asthma acquisition during adolescence in females, seven were enriched in a KEGG pathway, Malaria, and six of the seven genes have been linked to asthma in previous studies. The pathway Malaria is related to immune and inflammatory response to malaria infection.

For the identified CpGs, our findings on temporal consistency in DNAm strengthen the potential of those CpGs as epigenetic markers in newborns for asthma acquisition later in life. This finding will benefit future effort to detect subjects within days after birth who are at a high risk of asthma acquisition, making primary prevention possible. Temporal consistency in DNAm does not claim biological stability from birth to age 18 years, rather a consistent DNAm changes between birth and age 10 and between birth and age 18. It is worth noting that the CpGs identified in our study do not overlap with those from cross-sectional studies by Xu et al., although we observed a nice agreement if the same study design is applied to our data (Supplemental Material S3). With a longitudinal study design as in the present study, the time-order between DNAm and asthma acquisition is clearer compared to cross-sectional studies, which might explain the unique findings between our study and Xu et al. It is interesting to note that no common CpGs or genes were found between the markers for asthma acquisition during adolescence and those during young adulthood.

Asthma is a complex and heterogeneous disease due to various immune pathways which determine inflammatory pathways. Five genes (IL22RA2, CBX4, CFAP46, ATP1B2, and SOX11), of the 8 genes to which the 8 identified CpGs were mapped, have been shown to be related to asthma or airway diseases. Of these five genes, IL22RA2 was linked to various immune-mediated diseases. This gene is related to two pathways, Innate Immune System and Immune response Antigen presentation by MHC class II, which play a critical role in asthma pathogenesis. The other four genes, CBX4, CFAP46, ATP1B2, and SOX11, have been reported to be related to airway and lung diseases via regulating inflammation response. The relation of biological pathways involving the other three genes with asthma is unclear, but SOX11, a member of the SOX C group, regulates interleukin 13 (IL-13) signaling in lung fibroblasts and allergic disease through IL-13/STAT6/SOX11 pathway. Also, among the 8 genes (Table 3), MUC4 encodes a membrane bound mucin that plays an important role in cell proliferation and differentiation of epithelial cells by inducing specific phosphorylation of ERBB2 and whose expression has been shown to be modified by oral corticosteroids.

Importantly, among all the eight CpGs identified in our study based on two independent cohorts, most of them showed an association with expression of their neighboring genes. In particular, cg02705837 is located in the 1st Exon and 5' UTR of gene ATP1B2, a Wnt signaling gene shown to be involved in the pathogenesis of impairment in individuals with asthma. Unlike the other identified CpGs, DNAm at this CpG was
associated with the expression of multiple neighboring genes, SHBG, TP53, POLR2A, EIF4A1, and SPEM1, and both SHBG and TP53 have found to be associated with asthma\textsuperscript{64,65}. Of particular interest, in an earlier study by Yuan et al., DNAm of gene TP53, at site chr17:7591672 close to the location of cg02705837 (chr17: 7554666), is suggested as a peripheral blood biomarker to predict late-onset asthma\textsuperscript{65}. In our study, we showed that newborn DNAm of cg02705837 was associated with young adulthood asthma acquisition and with expression of its neighboring gene TP53. The comparison of our findings with that in Yuan et al. suggests that cg02705837, although not located on gene TP53, has a potential to regulate activity of its neighboring gene and, further, cg02705837 is likely to be a stable epigenetic marker for asthma acquisition.

The strengths of our study include the time-ordered study design with newborn DNAm and postnatal asthma acquisition with multi-layered analyses protocols, carefully designed screening process in the detection of candidate CpGs, and the utilization of an independent replication cohort. With this careful design, the findings tend to be conservative and the identified CpGs are likely to be informative with a strong potential of replicability. Some limitations did exist in this study. The sources of DNAm assessment are different between the two cohorts; DNAm was measured in heel prick blood of newborns in IOWBC and in cord blood in ALSPAC. Since our analyses focused on the differences in DNAm between subjects with asthma and no-asthma subjects rather than linear associations between continuous variables, a recent assessment showed that under this context, DNAm between the two sources had a high agreement overall\textsuperscript{66}. In addition, DNAm in both cohorts was based on blood cells and may not reflect DNAm in airway tissues. However, most of the genes corresponding to CpGs identified in our study are expressed in respiratory tissue\textsuperscript{47-52,57,58,60,62}. As a related limitation, our biological relevance assessment of the identified CpGs was carried out in the offspring of IOWBC. Since generation-specific factors might have confounded the underlying associations, interpretation of the CpGs showing potential biological relevance needs to be cautious. It is worth noting that, although they seemed not to be a concern in our study, household smoking during pregnancy, active smoking during pregnancy, season of birth, birth order, maternal asthma history, and maternal BMI, along with other prenatal factors, could potentially confound the findings. Due to rather small sample sizes of asthma acquisition in males in the ALSPAC cohort, we did not further test the IOWBC-discovered CpGs in males in ALSPAC. A larger scale study will be needed to confirm the viability of the discovered CpGs. Nevertheless, the findings from both cohorts indicate a potential of newborn epigenetic markers for asthma acquisition at later ages, and such markers are likely to be different at different stages of life.

**CONCLUSION**
In female participants, DNAm in newborns or at birth is associated with asthma acquisition after puberty. Our findings suggest that adolescent and young adulthood asthma acquisition possibly do not share common epigenetic markers.

Acknowledgements: We are indebted to all participants and their families who are followed up in the Isle of Wight 1989 birth cohort and the Avon Longitudinal Study of Parents and Children, which include interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, midwives, and nurses. We also appreciate the hard efforts of these two research teams in maintaining the cohorts and collecting data. We also thanks to the High Performance Computing facility at the University of Memphis.

Author contributions: The original concept and design were initiated by Hongmei Zhang and Wilfried Karmaus. Data analyses were performed by: Liang Li and Hongmei Zhang. Data interpretation and manuscript writing was led by Liang Li and Hongmei Zhang. Critical comments on the concept and the manuscript and final approval of the manuscript were given by all the authors Liang Li, John W. Holloway, Susan Ewart, S. Hasan Arshad, Caroline L. Relton, Wilfried Karmaus, Hongmei Zhang.

Conflict of interest: All authors are nothing to disclose.

Support statement: This work was supported by the National Institutes of Health research funds R01AI121226 (MPI: H Zhang and JW Holloway). DNAm analyses of F1 newborns and gene expression analyses of F2-newborns were funded by National Institute of Allergy and Infectious Diseases [R01 AI091905] and National Heart, Lung, and Blood Institute [R01 HL132321] (PI: W Karmaus). The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). CR is supported by the Medical Research Council and University of Bristol Integrative Epidemiology Unit (MC_UU_00011/5/MRC). The sponsor had no role in study design, data collection and analysis, or the preparation of the manuscript. This publication is the work of the authors and Hongmei Zhang will serve as guarantor for the contents of this paper.

Data availability: The data used in the study are available on request from the corresponding author with justification due to privacy and ethical restrictions.
Ethical approval

For IOWBC, local/national ethics committees approved the recruitment of the birth cohort between January 1989 and February 1990 and subsequent follow-up assessments. For ALSPAC, the ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.
FIGURE 2. Circos plot of mapping genes of CpGs survived screening via linear regression or identified by DMR and those with consistent directions in two cohorts for adolescence asthma acquisition in females (A circos plot was drawn on the website ([https://yimingyu.shinyapps.io/shinycircos](https://yimingyu.shinyapps.io/shinycircos)), indicating the location of genes corresponding to the CpGs. The outer track is for the location of mapping genes survived screening or identified by DMR, and the inner track is for the location of mapping genes corresponding to identified CpG with consistent directions between the two cohorts). On outer track, the black points represent -log10 transformed p-values of CpGs that survived screening via linear regression or identified via DMR. On the inner track, the blue unfilled circles and red filled squares denote -log10 transformed p-values of association with adolescence asthma acquisition for those with consistent direction of associations between the two cohorts (but in ALSPAC the effects were statistically insignificant) and those replicated in ALSPAC (in both cohorts, the directions were consistent and the effects were statistically significant), respectively. Chromosome numbers are shown on outermost circle. On outer track, the black points represent -log10 transformed p-values of CpGs that survived screening via linear regression or identified via DMR. On the inner track, the blue unfilled circles and red filled squares denote -log10 transformed p-values of association with adolescence asthma acquisition for those with consistent direction of associations between the two cohorts (but in ALSPAC the effects were statistically insignificant) and those replicated in ALSPAC (in both cohorts, the directions were consistent and the effects were statistically significant), respectively. Chromosome numbers are shown on outermost circle.


40. Vink NM, Postma DS, Schouten JP, Rosmalen JGM, Boezen HM. Gender differences in asthma development and remission during transition through puberty: The TRacking Adolescents’
Individual Lives Survey (TRAILS) study. *Journal of Allergy and Clinical Immunology.* 2010;126(3):498-504.e496.
45. Withers Al, Green R. Transition for Adolescents and Young Adults With Asthma. *Front Pediatr.* 2019;7:301-301.


