DNA Methylation at Birth is Associated with Childhood Serum Immunoglobulin E Levels

Luhang Han1*, Akhilesh Kaushal2*, Hongmei Zhang3*, Latha Kadalyal4, Jiasong Duan3, John W. Holloway4,5, Wilfried Karmaus3, Pratik Banerjee6, Shih-Fen Tsai7, Hui-Ju Wen7, Syed Hasan Arshad8,9,10 and Shu-Li Wang7,9,10

1Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA. 2School of Medicine, Emory University, Atlanta, GA, USA. 3Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, Memphis, TN, USA. 4Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK. 5Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. 6Department of Food Science and Human Nutrition, University of Illinois, Urbana, IL, USA. 7Division of Environmental Health and Occupational Medicine, National Health Research Institutes, Miaoli, Taiwan. 8David Hide Asthma and Allergy Research Centre, St Mary’s Hospital, Newport, Isle of Wight, UK. 9School of Public Health, National Defense Medical Center, Taipei. 10Department of Public Health, China Medical University, Taichung.

ABSTRACT: Immunoglobulin E (IgE) is known to play an important role in allergic diseases. Epigenetic traits acquired due to modification of deoxyribonucleic acid (DNA) methylation (DNAm) in early life may have phenotypic consequences through their role in transcriptional regulation with relevance to the developmental origins of diseases including allergy. However, epigenome-scale studies on the longitudinal association of cord blood DNAm with IgE over time are lacking. Our study aimed to examine the association of DNAm at birth with childhood serum IgE levels during early life. Genome-scale DNAm and total serum IgE measured at birth, 5, 8, and 11 years of children in the Taiwan Maternal and Infant Cohort Study were included in the study in the discovery stage. Linear mixed models were implemented to assess the association between cord blood DNAm at ~310K 5′-cytosine-phosphate-guanine-3′ (CpG) sites with repeated IgE measurements, adjusting for cord blood IgE. Identified statistically significant CpGs (at a false discovery rate, FDR, of 0.05) were further tested in an independent replication cohort, the Isle of Wight (IoW) birth cohort. We mapped replicated CpGs to genes and conducted gene ontology analysis using ToppFun to identify significantly enriched pathways and biological processes of the genes. Cord blood DNAm of 273 CpG sites were significantly (FDR = 0.05) associated with IgE levels longitudinally. Among the identified CpGs available in both cohorts (184 CpGs), 92 CpGs (50%) were replicated in the IoW in terms of consistency in direction of associations between DNA methylation and IgE levels later in life, and 16 of the 92 CpGs showed statistically significant associations (P < 0.05). Gene ontology analysis identified 4 pathways (FDR = 0.05). The identified 16 CpG sites had the potential to serve as epigenetic markers associated with later IgE production, beneficial to allergic disease prevention and intervention.

KEYWORDS: DNA methylation, Immunoglobulin E, childhood, cord blood, Taiwan cohort, Isle of Wight cohort, epigenetic, longitudinal

Introduction

The prevalence of allergic diseases, including asthma, continues to increase in children and young adults worldwide. About fifty percent of school children are sensitized to one or more common allergens.1 Susceptibility to allergic disease is, in part, determined by genetic susceptibility to react to allergens.2,3

In addition to genetic factors, gene-environment interactions during critical periods of immune development are important for subsequent development of allergic diseases. Epigenetic variation is postulated to constitute an important mechanism through which these interactions are mediated.4

*These authors contributed equally to this work.

Epigenetic processes regulate gene expression during immune development, and studies suggest disruptions in these processes can modify disease risk in a manner analogous to single nucleotide polymorphisms (SNPs).5 deoxyribonucleic acid (DNA) methylation (DNAm) is one such epigenetic process which is associated with gene silencing and with the patterning of gene expression that determines cell types and function.

The link between DNAm and allergic diseases has been assessed in prior studies. Zhang et al.6 showed that DNAm along with single nucleotide polymorphisms played a role in asthma risk among pre-adolescent children and young adults. Lockett et al.7 found that allergic risk in adulthood is associated with DNAm potentially impacted by a child's season of birth.
birth. In addition to asthma, study by Ziyab et al found that DNA methylation modulates the association between filaggrin gene loss of function variants and eczema,8 another recent study via random forest demonstrated that DNA methylation is linked to the risk of eczema.9

Immunoglobulin E (IgE) plays a major role in allergic diseases such as asthma, eczema, and hay fever.10-12 IgE production leads to type I hypersensitivity, which manifests in various allergic diseases. However, the mechanisms that lead to increased propensity for IgE isotype switching by B cells in atopic individuals is poorly understood. Multiple cross-sectional studies have suggested that DNA methylation is associated with total serum IgE levels.13-16 For instance, an epigenome-scale study using Illumina methylation 27K array identified 5′-cytosine-phosphate-guanine-3′ (CpG) loci from peripheral blood associated with total serum IgE.16 Another cross-sectional study used the peripheral blood of 18-year-old men and women identified associations between CpG loci and serum IgE.17 A recent longitudinal study17 showed that DNA methylation in cord blood and childhood peripheral blood was associated with total serum IgE levels in childhood. However, it is not clear from these studies whether association of IgE with DNA methylation represents a casual mechanism and to what extent DNA methylation at an earlier age can associate with IgE at a later ages. In this study, we tackle this problem by assessing the potential of taking DNA methylation at birth (cord blood) as independent variable for total serum IgE during the course of early life. We hypothesize that DNA methylation at birth has the potential to associate with total serum IgE at a later age in early life. The findings will contribute to an improved understanding to the impact of early life conditions on IgE levels later in life.

Methods
Taiwanese maternal infant and birth cohort description
The discovery cohort is a subset of the Maternal and Infant Cohort Study in Taiwan investigating various prenatal and postnatal factors that might affect child health outcomes.18-20 All pregnant women participating in this study signed informed consent forms explaining the benefits and risks of participation. This study was approved by Human Ethical Committee of the National Health Research Institutes in Taiwan. Pregnant women who received medical care at the local medical centers were invited to join this study between December 2000 and November 2001. Among the 610 women who met the requirement, 430 volunteered to participate in the study (data collection flow chart is in Supplemental Figure S1). Of the 430 pregnant women, 117 were excluded due to non-compliance of providing samples. Urine samples were then collected from the remaining 313 pregnant women during the third trimester (28-38 weeks of gestation). In total 313 live births were reported as noted in our earlier work.20 Out of the 313 live births 9 were twins and one of the twins was randomly selected for subsequent studies. In addition, 5 newborns could not be included due to loss of follow up. This resulted in 299 mother-newborn pairs. The cord blood sample was collected for all the 299 mother-newborn pairs. DNA methylation in 64 cord blood samples that had the required DNA concentration and quality for this epigenome assay.

Assessment of Immunoglobulin E
Measurements of cord blood IgE and total IgE in children at 5, 8, and 11 years were performed using the ADVIA Centaur chemiluminescence immunoassay system (Siemens Healthcare Diagnostics; Deerfield, Illinois, USA). Blood samples (0.5 mL) were centrifuged and the sera stored at −20°C prior to analysis. In the analyses, to normalize the data, we used base-10 logarithm transformation to the original IgE measurements.

DNA isolation and DNA methylation assessment
DNA was isolated from EDTA-treated cord blood buffy coat samples (n = 64) using a commercial kit (Genta Puregene; Qiagen, Hilden, Germany) and was bisulfite converted using the EZ DNA Methylation kit (Zymo Research Corp, Irvine, California, USA). Samples were randomized across several plates and subsequently subjected to epigenome-wide DNA methylation assessment using the Illumina Infinium Human Methylation 450 BeadChip (Illumina, San Diego, California, USA) which simultaneously profiles the methylation status of >485,000 CpG sites with single-nucleotide resolution. DNA isolation and methylation assessment were performed by Genetech Biotech Co., Ltd.

Quality control
Raw DNA methylation data for 485,577 CpG site were pre-processed according to an established pipeline. Probes with detection P > .01 in more than 10% of samples were removed. The function preprocessSWAN in the Bioconductor package minfi21 was used for normalization, background correction and peak correction. The preprocessSWAN uses subset within array normalization (SWAN) to reduce differences in probe chemistry between the 2 types of probe on the arrays.22 CpG sites located on sex chromosomes and annotated by polymorphic CpGs known in Illumina 450 microarray.21 After quality control, 314,257 CpG sites were retained for statistical analysis. The pre-processed DNA methylation data in beta values were transformed to M values, approximated as log2 [β/(1-β)], to ensure a better fit to statistical model assumptions.

Correction for cell mixture proportion
Since blood is a mixture of functionally and developmentally distinct cell populations,24 adjusting for this cell type removes the confounding effects of cell heterogeneity in DNA methylation.
measured from blood samples.\textsuperscript{25} Therefore, to control for the confounding effects of cellular heterogeneity, cell type composition of the blood sample was calculated using function `estimateCellCounts` in the R package `minfi`.\textsuperscript{26} IDAT files from 450K Illumina DNA methylation were used to estimate the proportion of 7 cell types: CD4T-cells, CD8T-cells, natural killer cells (NK), monocytes, B-cells, granulocytes, eosinophils.

**Replication study**

The replication study was conducted using the data from an independent cohort, the Isle of Wight (IoW) birth cohort.\textsuperscript{27} Briefly, the IoW cohort is an ongoing prospective birth cohort consisting of the whole population that were born between 1 January 1989 and 28 February 1990 on the Isle of Wight, UK. The purpose was to study the natural history of asthma and related allergic disorders with subsequent assessment at 1, 2, 4, 10, and 18 years. IgE measurements were obtained from cord blood and peripheral blood at age 10 and 18 years. DNA methylation from Guthrie card blood and peripheral blood at age 10 and 18 years was assessed using the Illumina Infinium HumanMethylationEPIC BeadChip. DNAm in DNA extracted from Guthrie cards (n = 796) was included in the replication study. Details about this cohort, DNA methylation assessment, and related data analyses are described in the supplemental materials.

**Statistical analyses**

To compare the study samples with the complete cohort, we compared features in mothers as well as in offspring. For mothers, we examined the consistency in age, pre-pregnancy body mass index, and education between the study samples and the complete cohort treating the statistics in the complete cohort as the truth. For offspring, we focused on gestational age and serum IgE in the assessment of consistency. For education levels, we used 1-sample proportion test, and for continuous variables, we applied 1 sample t-tests (age, pre-pregnancy BMI, and gestational age) or Wilcoxon signed rank tests (serum IgE, due to its non-normal distribution).

To identify CpG sites whose DNAm could affect IgE levels in children at 5, 8, and 11 years of age, the analysis was performed in 2 stages. In stage 1 we obtained the residuals of DNAm by regressing DNAm of each CpG (314 257 CpGs) on cell proportions and batches of DNAm assessment at different time points. In stage 2 we used linear mixed models with repeated IgE measures to check the association of residuals of DNAm with the IgE in longitudinal setting, while adjusting for cord blood IgE, birth weight and gender of the child. The procedure PROC MIXED in SAS (version 9.4) was used to fit the model. Multiple testing correction was performed by controlling false discovery rate (FDR) of 0.05.\textsuperscript{28} Statistically significant CpG sites were further tested in the IoW cohort, using the same SAS procedure with similar covariates (supplemental file on the IoW birth cohort). For CpG sites showing statistical significance at .05 in the IoW cohort, we further assessed the association of DNAm at birth with allergic asthma status (subjects without asthma is the reference group) at ages 10 and 18 using log-linear models with repeated measures via PROC GENMOD in SAS. In addition to DNAm, age, sex, and sex × DNAm interaction were included in the model as covariates. To examine the potential of sex-specific association between DNAm and allergic asthma, we set the level of statistical significance at .1.

**Pathway analyses**

The genes annotated to the identified CpGs were obtained from the Illumina EPIC array manifest file. When a CpG was annotated to more than 1 gene, all annotated genes were included. For the corresponding genes of CpG sites could not be identified from Illumina array manifest, SNIPPER (https://csg.sph.umich.edu/boehnke/snipper/) version 1.2 was used to identify the closest gene to the CpG based on the number of base pairs. Functional enrichment analysis of Pathways of the resulting genes was then carried out using ToppFun of ToppGene Suite, a gene list enrichment analysis tool\textsuperscript{29} and using the gemeth function in the R package Limma\textsuperscript{30} to further evaluate KEGG terms. The method implemented in gemeth takes into account the number of CpGs of each gene included in the platform used by Illumina for DNA methylation assessment. The method in ToppFun, on the other hand, utilizes hypergeometric distribution without such a consideration.

**Results**

The data were from a birth cohort study examining multiple prenatal and postnatal factors in relation to child health outcomes as part of the nationwide Taiwan Maternal and Infant Cohort Study\textsuperscript{18,19} established in Taiwan in 2000 to 2001. In total, 64 subjects with DNA methylation at the genome-scale in cord blood, child’s gender, batch effect, and birth weight were included in the analyses.

Table 1 compares the characteristics between the 64 samples and available methylation data with those of the whole cohort (n = 299). For the majority of the characteristics the samples were comparable to the whole cohort. However maternal education levels and pre-pregnancy BMI of mothers included in the present analysis were different from those in the whole cohort (P < .0001). The majority of subjects in the methylation subset had higher education and their average BMI was lower than that in the whole cohort. Table 2 shows the characteristics of pregnant women and newborns by sex. Of the 64 newborns, 38 (59.4%) were male. The levels and distribution of cord blood and serum IgE for children at ages 5, 8, and 11 years are shown in Table 3.

Epigenome-scale assessments of statistical associations between log_{10} IgE and residuals of DNAm in cord blood at 314 257 CpG sites were conducted in a longitudinal setting via
Table 1. Comparison of the analytical with the whole cohort.

| VARIABLES                        | STUDY SAMPLE (N=64) | ALL MOTHERS AND THEIR NEWBORN (N=299) | P-VALUE |
|----------------------------------|---------------------|--------------------------------------|---------|
|                                  | MEAN±SD OR NUMBER (%) OR MEDIAN (IQR) | MEAN±SD OR NUMBER (%) OR MEDIAN (IQR) |         |
| Maternal characteristics         |                     |                                      |         |
| Age (years)                      | 28.9 ± 4.8          | 28.3 ± 4.2                           | .32     |
| Pre-pregnant BMI (kg/m²)         | 20.5 ± 2.6          | 25.6 ± 3.9                           | 1.92 × 10⁻²³ |
| Maternal education               |                     |                                      | 5.12 × 10⁻¹⁵ |
| ≤ high school                    | 3 (4.7%)            | 132 (44%)                            |         |
| High school + 2 y                | 22 (34.3%)          | 117 (39%)                            |         |
| ≥ high school + 4 y              | 39 (61%)            | 50 (17%)                             |         |
| Offspring                        |                     |                                      |         |
| Gestational age (weeks)          | 39 ± 1.2            | 39 ± 2.8                             | 1       |
| Immunoglobulin E (IU/mL)         |                     |                                      |         |
| Cord blood IgE                   | 0.22 (0.86)         | 0.25 (0.97)                          | .41     |
| Total IgE at age 5 y             | 61.40 (119.70)      | 54.40 (115.70)                       | 1.00    |
| Total IgE at age 8 y             | 96.20 (158.00)      | 96.25 (153.20)                       | 1.00    |
| Total IgE at age 11 y            | 55.00 (182.00)      | 113.5 (231.00)                       | .26     |

Table 2. Characteristics of mothers and their newborns by newborn gender in Taiwan during 2000 to 2001 (n=64).

| CHARACTERISTICS                        | ALL (N=64) | SEX OF THE INFANT | P-VALUE |
|---------------------------------------|------------|------------------|---------|
|                                      |            | MALE (N=38)      | FEMALE (N=26) |         |
| Pregnant women                        |            |                  |          |
| Age (years)                           | 28.9 ± 4.8 | 28.6 ± 4.1       | 29.5 ± 5.7 | .492    |
| Pre-pregnant BMI (kg/m²)              | 20.5 ± 2.6 | 20.2 ± 2.4       | 21.0 ± 2.9 | .244    |
| Urinary creatinine (mg/dL)            | 63.6 ± 41.7| 70.9 ± 46.0      | 53.0 ± 32.9| .078    |
| Maternal education                    |            |                  | .303     |
| High school + 2 y                     | 25 (39%)   | 13 (34%)         | 12 (48%)  |         |
| ≥ High school + 4 y                   | 39 (61%)   | 25 (66%)         | 14 (52%)  |         |
| Newborns                              |            |                  |          |
| Gestational age (weeks)               | 39 ± 1.2   | 39 ± 1.1         | 39 ± 1.4 | .791    |

*Presented as the mean±SD or number (percentage).

*p-value for difference between male and female newborns using t-test for continuous variables and χ² or Fishers Exact Test for categorical variable.

linear mixed modeling. This analysis was performed in 2 stages. In stage 1 we obtained the residuals of DNAm by regressing DNAm of each CpG (314 257) on cell proportions (CD4T-cells, CD8T-cells, natural killer cells (NK), monocytes, B-cells, granulocytes, and eosinophils) and indicator variables associated with different batches of DNA methylation data. In stage 2 we used linear mixed models to check the association of residuals of DNAm in cord blood with IgE at ages 5, 8, and
Han et al

11 years, adjusting for cord blood IgE, birth weight, and gender of the child. In total, 273 CpG sites showed statistically significant associations, after correcting for multiple testing by controlling FDR of 0.05. Figure 1 shows the Manhattan plot of \( P \)-values for testing on the 314,257 CpG sites, with a red line indicating the \( P \)-value threshold corresponding to FDR of 0.05.28 Supplemental Table S1 lists the 273 CpG sites with their regression coefficients along with \( P \)-values, chromosomes they belong to, their corresponding genes and location on the genes. The 273 CpG sites were mapped to 223 genes (Supplemental Table S1).

The 273 CpG sites identified in our study were further tested in an independent cohort, the Isle of Wight (IoW) birth cohort in the United Kingdom. Of the 273 CpGs, 184 were available in the IoW cohort. Genome-scale DNAm in Guthrie card samples were available in 796 children. Total IgE was measured at ages 10 and 18 years. The analysis was performed in 2 stages similar to the main study and used the same covariates. At half (92 CpGs) of the 184 CpG sites (Supplemental Table S1, Figure S2), longitudinal associations of Guthrie card blood DNAm with IgE over time were consistent with those found in our study in terms of direction of regression coefficients. More than one fourth of the 92 CpGs were statistically significant at the .05 significance level (Table 4; Figure 2). At 13 CpGs (81%) of the 16 CpGs the associations were negative. To exemplify the associations, 2 CpGs showing a high statistical significance in both cohorts, cg02584802 and cg03277925, were selected to illustrate the associations, using age 18 IgE data in the IoW cohort (Supplemental Figures S3 and S4 in the supplemental material).

Our assessment on the potential of clinical relevance of the detected 16 CpGs, based on data in the IoW cohort and via generalized linear mixed models, identified 3 CpGs showing potential sex-specific associations with pre- (age 10 years) and post-adolescence (age 18 years) allergic asthma (Table 5). At 2 CpG sites (cg22287064 and cg03277925), higher DNAm was associated with higher risk of allergic asthma for females, while in males, higher DNAm was associated with lower risk of allergic asthma. In comparison with cg22287064 and cg03277925, at CpG site cg03460239, the association was in the opposite direction (Table 5).

For the 16 CpGs, we further conducted functional annotation analysis using 2 bioinformatics tools, the ToppFun in the ToppGene Suite and gometh function in the R package Limma. These 16 CpGs were mapped to 15 genes. Five pathways were identified by both the ToppFun webtool and the gometh function in the R package Limma (Table 6) at the significance level of .05, although none of them survived multiple testing correction.

| AGE/PERCENTILE          | MIN | 5TH | 25TH | 50TH | MEAN | 75TH | 95TH | MAX  |
|-------------------------|-----|-----|------|------|------|------|------|------|
| Birth (cord blood)      | 0.03| 0.03| 0.06 | 0.22 | 3.19 | 0.89 | 22.90| 61.40|
| IgE at 5y (IU/mL)       | 5.70| 7.59| 17.95| 61.40| 102.06| 125.13| 303.65| 524.00|
| IgE at 8y (IU/mL)       | 6.38| 7.78| 11.00| 96.20| 167.99| 169.00| 695.00| 921.00|
| IgE at 11y (IU/mL)      | 6.00| 6.00| 15.25| 55.00| 159.20| 185.25| 616.35| 946.00|

The low median of IgE at 11y was due to large missingness (68.75% missing values) in the dataset.

![Manhattan plot](image-url)
The overall aim of this study was to identify CpG sites of which DNAm at birth measured in cord blood or Guthrie cards was associated with levels of IgE at later ages. We identified 273 CpG sites in cord blood (at FDR of 0.05) longitudinally associated with IgE in the discovery cohort (n = 64), of which DNAm in Guthrie cards at 184 CpG sites were available in the IoW replication birth cohort (n = 796). Out of the 184 CpG sites, findings at 92 CpGs in the IoW cohort were consistent with those from the discovery cohort in terms of direction of association with IgE. These 92 CpGs were on 93 genes located mostly in body and promoter regions. Among the 92 CpG sites, 16 CpGs were statistically significant at .05, and DNAm at most of these 16 CpG sites were negatively associated with the IgE levels. Such negative associations between DNAm and IgE were consistent with the findings described in Liang et al.\textsuperscript{31} More than half of the 16 CpGs were located either in the body (~38%) or the promoter (~25%) regions of a gene. It is worth noting that multiple testing was implemented in the discovery cohort from Taiwan, but such penalty was not considered in the replication cohort (the IoW cohort). When testing the discovered CpGs in a replication cohort, these CpGs were not randomly selected as in genome-wide studies. Instead, they were targeted CpGs. For situation like this, as noted in Rothman,\textsuperscript{32} forcing multiple testing potentially would increase type II error and was not encouraged.

We did not identify statistically significant KEGG pathways after adjusting for multiple testing, possibly due to the small number of genes involved in the analyses. However, some of the genes to which the 16 CpG sites were mapped to have been shown to be linked to allergic diseases and IgE levels. In particular, MYO15B has been shown to be overexpressed in hyper-immunoglobulin E (IgE) syndrome (HIES).\textsuperscript{33} The gene BMI1 regulates T helper 2 (Th2) cell differentiation by interacting with GATA3 in T cells.\textsuperscript{34} Th2 cells promote the production of IgE by switching the IgM antibody isotype to IgE.\textsuperscript{35} PTK2B was in gene family of Minor histocompatibility antigens. Minor histocompatibility antigens are known to play a role in developing immunological response in solid-organ

### Table 4. Summary of the replicated 16 statistically significant CpGs, regression coefficients and characteristic of corresponding genes.

| CPG          | TAIWAN COHORT |                         |                         |                         |                         |                         |                         |                         |                         |
|--------------|---------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|              | EST.* | $P_{\text{raw}}$ \textsuperscript{$\dagger$} | $P_{\text{FDR}}$ \textsuperscript{$\ddagger$} | EST.* | $P_{\text{raw}}$ \textsuperscript{$\dagger$} | $P_{\text{FDR}}$ \textsuperscript{$\ddagger$} | GENE NAME | LOCATION | CHR# | CPG REGIONS |
| cg02023402   | −1.63   | 1.05 × 10\textsuperscript{−5} | 0.0044 | −0.22   | 0.0324 | CACNA2D1 | Intergenic | 7            |
| cg02584802   | −1.62   | 1.07 × 10\textsuperscript{−5} | 0.0045 | −0.31   | 0.0003 | Intergenic | 11 | S_Shelf |
| cg03277925   | 0.84    | 4.61 × 10\textsuperscript{−5} | 0.0194 | 0.18    | 0.0059 | ADARB2 | Body | 10 |
| cg03460239   | −1.59   | 1.25 × 10\textsuperscript{−5} | 0.0053 | −0.30   | 0.0219 | OR8B8 | Promoter | 11 |
| cg03689195   | 2.32    | 1.18 × 10\textsuperscript{−5} | 0.0497 | 0.26    | 0.0107 | PCDH17 | Intergenic | 13 | Island |
| cg04990210   | −2.20   | 5.04 × 10\textsuperscript{−4} | 0.0021 | −0.19   | 0.0257 | CTNS | Body | 17 |
| cg13984351   | −1.45   | 2.68 × 10\textsuperscript{−4} | 0.0113 | −0.30   | 0.0139 | BMI1 | Promoter | 10 |
| cg15712821   | −0.93   | 6.62 × 10\textsuperscript{−5} | 0.0279 | −0.17   | 0.0314 | PTK2B | 5'UTR | 8 | S_Shelf |
| cg18888710   | −2.22   | 6.86 × 10\textsuperscript{−5} | 0.0289 | −0.21   | 0.0450 | ATP6V1C2 | Body | 2 |
| cg19224656   | −0.67   | 4.26 × 10\textsuperscript{−4} | 0.0179 | −0.06   | 0.0331 | SHQ1 | Body | 3 | Island |
| cg19891951   | −1.38   | 4.58 × 10\textsuperscript{−5} | 0.0193 | −0.17   | 0.0159 | DDN | Promoter | 12 | Island |
| cg22287064   | −1.32   | 8.89 × 10\textsuperscript{−5} | 0.0374 | −0.15   | 0.0210 | MYO15B | Promoter | 17 | Island |
| cg22590761   | −2.13   | 8.28 × 10\textsuperscript{−5} | 0.0348 | −0.27   | 0.0334 | LOXL1 | Body | 15 | Island |
| cg23671600   | 1.41    | 3.32 × 10\textsuperscript{−4} | 0.0140 | 0.11    | 0.0239 | FBXO6 | 5'UTR | 1 | Island |
| cg26875958   | −1.36   | 9.46 × 10\textsuperscript{−5} | 0.0398 | −0.27   | 0.0107 | GRM1 | Body | 6 | N_Shelf |
| cg27209964   | −3.55   | 3.24 × 10\textsuperscript{−5} | 0.0137 | −0.48   | 0.0124 | FLT3 | Intergenic | 13 |

*Est.: estimated regression coefficients. $P_{\text{raw}}$: raw $P$-value, $P_{\text{FDR}}$: $P$-value after controlling FDR of .05.
#CHR: Chromosome.
transplantation. A GWAS study has reported chronic obstructive pulmonary disease related loci at or near ADARB2. Among the genes that the identified 16 CpGs were mapped to, a small portion has previously been reported in the literature with a connection to IgE. The limited concordance might have been due to the discovery cohort being Asian, a population not largely investigated in the related area. However, we may also have detected novel CpGs due to our stringent assessment (discovery and replication). Some genes including ATP6V1C2 (cg18888710) and GRM1 (cg26875958) were also identified in statistically significant KEGG pathways. ATP6V1C2 is found to have differential expression during culture of rat lung type II pneumocytes and related to recessive distal renal tubular acidosis. GRM1 is

![Bar plot displaying regression coefficients, location of the 16 CpGs, and their corresponding genes. Different colors indicate the location of the CpGs on a gene.](image)

Table 5. Effects of DNAm at birth on allergic asthma (no asthma was the reference group) at ages 10 and 18 years at the 16 identified CpGs using data in the IoW cohort.

| CPG SITES     | EST. COEF.* (DNAM) | EST. COEF.* (DNAM × SEX) | P RAW* (DNAM × SEX) |
|---------------|---------------------|--------------------------|---------------------|
| cg22287064    | 0.66                | −1.49                    | 0.0396              |
| cg03277925    | 0.38                | −1.22                    | 0.0481              |
| cg03460239    | −1.21               | 2.43                     | 0.0781              |

The P-values are for the effects of DNAm × sex interaction (female was the reference group) and results with P-values less than .10 are included. Complete results are in Supplemental Table S2 in the Supplemental material.

*Est. coef.: estimated regression coefficients.

PRaw*: raw P-value.
found to have a protective interaction between asthma and farming and farming-related exposures such as contact with cows and straw or contact with straw alone.\textsuperscript{41}

Furthermore, gender reversal of asthma prevalence from pre- to post-adolescence has been demonstrated in multiple studies. The detection of potential sex-specific association of DNAm at birth with allergic asthma pre- and post-adolescence at certain CpG sites sheds some light of the origin of gender reversal and certainly deserves further in-depth investigation. In this article, DNA methylation is our epigenetic focus. Another type of epigenetic modification, histone modification, also greatly contributes to the pathogenesis of allergic diseases.\textsuperscript{42}

The study has some limitations. The DNAm in the discovery cohort was from cord blood while in the replication cohort it was from Guthrie cards. Our recent study showed that DNAm agreement between these 2 sources is moderate.\textsuperscript{43} Hence, the CpGs not replicated in the IoW cohort might have been partially due to the difference in DNAm sources. In addition, the 2 study cohorts were of different ethnicity with Asian children in the discovery cohort and \textasciitilde98\% Caucasian children in the replication cohort, which could have been a major cause of the inconsistent results at some CpGs between the 2 cohorts due to either genetic or environmental exposure (eg, maternal diet) differences between the cohorts. Replicating the findings in a group with the same ethnicity might improve consistency and support reproducibility in Asian populations. On the other hand, because of the differences between the 2 cohorts in terms of region and ethnicity, the 92 CpGs showing consistent findings between the 2 cohorts have the potential to serve as epigenetic markers that associated with later IgE production. Among the identified CpG sites, a higher level of DNA methylation at most sites was associated with lower IgE production. In addition, certain genes corresponding to the identified CpGs were in pathways regulating IgE levels including \textit{BMI1}.

\textbf{Conclusions}

At 16 CpGs, DNA methylation at birth was associated with IgE production longitudinally, suggesting that these CpGs had a potential to serve as epigenetic markers that associated with later IgE production. Among the identified CpG sites, a higher level of DNA methylation at most sites was associated with lower IgE production. In addition, certain genes corresponding to the identified CpGs were in pathways regulating IgE levels including \textit{BMI1}.

\textbf{Acknowledgements}

The authors acknowledge the cooperation of the gynecologists and pediatricians who participated in this study, are thankful to the staff members and nurses at the David Hide Asthma and Allergy Research Centre, Isle of Wight, United Kingdom, and the High Performance Computing facility at the University of Memphis.

\textbf{Author Contributions}

HZ and SW conceived the study. HZ, SW, LH, and AK carried out the project. HZ and WK provided guidance on the analytical and statistical aspects. AK, LH, and JD performed statistical analyses. WK, LK, and JWH provided guidance on epigenome. PB helped with the design of the study. ST prepared DNA. HW prepared the data. HA provided data from the Isle of Wight cohort and advised on clinical aspects related to IgE. LH, AK, and HZ drafted the manuscript. All authors were involved in editing and revising the manuscript.

\textbf{Ethics Approval and Consent to Participate}

The study is approved by the Institutional Review Board (IRB) of the University of Memphis.

\begin{table}
\centering
\caption{KEGG pathways linked to the 15 genes at statistical significance level of .05.}
\begin{tabular}{|l|l|l|l|}
\hline
\textbf{NAME} & \textbf{PATHWAY} & \textbf{P\textsubscript{Raw}}$^5$ & \textbf{HIT COUNT IN GENOME} & \textbf{HIT IN QUERY LIST} \\
\hline
Phospholipase D signaling pathway & path:hsa04072 & 0.0089 & 148 & PTK2B; GRM1 \\
Calcium signaling pathway & path:hsa04020 & 0.0135 & 201 & PTK2B; GRM1 \\
Collecting duct acid secretion & path:hsa04966 & 0.0181 & 27 & ATP6V1C2 \\
Vibrio cholerae infection & path:hsa05110 & 0.0390 & 50 & ATP6V1C2 \\
Epithelial cell signaling in Helicobacter pylori infection & path:hsa05120 & 0.0494 & 70 & ATP6V1C2 \\
\hline
\end{tabular}
\end{table}

The \textit{P}-values included in the table were from gometh function in the R package Limma, and similar \textit{P}-values were obtained in ToppFun. $^5$\textit{P\textsubscript{Raw}}: raw \textit{P}-value.
