Check for updates

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

Epigenetics Insights
Volume 14: 1–9
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25168657211008108
SAGE

Luhang Han[1*], Akhilesh Kaushal[2*], Hongmei Zhang[3], Latha Kadalayil[4], Jiasong Duan[3], John W. Holloway[4,5], Wilfried Karmaus[3], Pratik Banerjee[6], Shih-Fen Tsai[7], Hui-Ju Wen[7], Syed Hasan Arshad[5,8] and Shu-Li Wang[7,9,10]

[1]Department of Mathematical Sciences, University of Memphis, Memphis, TN, USA. [2]School of Medicine, Emory University, Atlanta, GA, USA. [3]Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, Memphis, TN, USA. [4]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK. [5]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. [6]Department of Food Science and Human Nutrition, University of Illinois, Urbana, IL, USA. [7]Division of Environmental Health and Occupational Medicine, National Health Research Institutes, Miaoli. [8]David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Newport, Isle of Wight, UK. [9]School of Public Health, National Defense Medical Center, Taipei. [10]Department 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 ($P < .05$). 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

RECEIVED: March 11, 2020. **ACCEPTED:** November 25, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for the Maternal and Infant Cohort Study in Taiwan was provided by the National Health Research Institutes, Miaoli, Taiwan (Grant No.: EM-105-PP-05), and the Ministry of Science and Technology, Taiwan (MOST104-2314-B-400-001). Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases at National Institutes of Health (NIH), USA, under award numbers R01 AI091905 (PI: Wilfried Karmaus) and R01 AI121226 (MPI: Hongmei Zhang, John Holloway).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Shu-Li Wang, Division of Environmental Health and Occupational Medicine, National Institute of Environmental Health Sciences National Health Research Institutes 35, Keyan Road, Zhunan Town, Miaoli County 35053, Taiwan. Email: slwang@nhri.org.tw

Hongmei Zhang, Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, 3825 DeSoto Avenue, Memphis, TN 38152, USA. Email: hzhang6@memphis.edu

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] 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

*These authors contributed equally to this work.

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

birth. In addition to asthma, study by Ziyab et al found that DNAm 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 DNAm 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.[15] A recent longitudinal study[17] showed that DNAm 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 assocaition of IgE with DNA methylation represents a casual mechanism and to what extent DNAm 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 DNAm 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 may 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 center 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 livebirths 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. DNAm was measured 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 (Gentra 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 DNAm 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 *minfi*[21] 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 probe SNPs within 10 base pairs of the CpG site were removed. We further excluded cross-reactive probes and polymorphic CpGs known in Illumina 450 microarray.[23] After quality control, 314 257 CpG sites were retained for statistical analysis. The pre-processed DNAm data in beta values were transformed to M values, approximated as $\log2\,[\beta/(1-\beta)]$, 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 DNAm

measured from blood samples.[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*.[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.[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.[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[29] and using the gometh function in the R package Limma[30] to further evaluate KEGG terms. The method implemented in gometh 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[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/m^2)	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)[a]	SEX OF THE INFANT		*P*-VALUE[b]
		MALE (N = 38)[a]	FEMALE (N = 26)[a]	
Pregnant women				
Age (years)	28.9 ± 4.8	28.6 ± 4.1	29.5 ± 5.7	.492
Pre-pregnant BMI (kg/m^2)	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

[a]Presented as the mean ± SD or number (percentage).
[b]p-value for difference between male and female newborns using t-test for continuous variables and χ^2 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

Table 3. Distribution of IgE (IU/mL) across 4 different ages in children.

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 5 y (IU/mL)	5.70	7.59	17.95	61.40	102.06	125.13	303.65	524.00
IgE at 8 y (IU/mL)	6.38	7.78	11.00	96.20	167.99	169.00	695.00	921.00
IgE at 11 y (IU/mL)	6.00	6.00	15.25	55.00	159.20	185.25	616.35	946.00

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

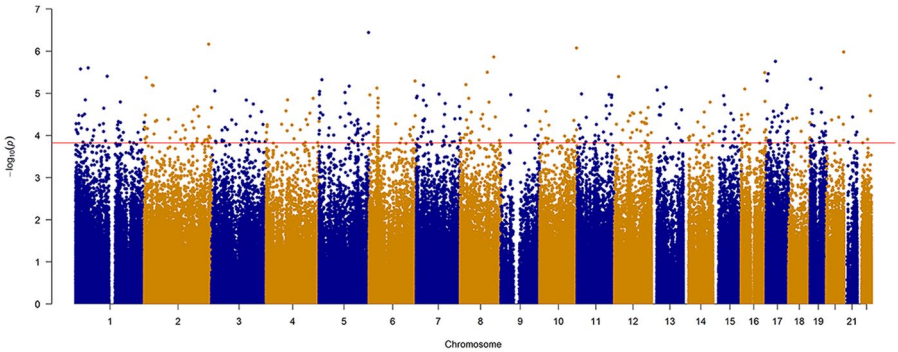

Figure 1. Manhattan plot for the longitudinal association of Genome-wide DNA methylation with $\log_{10}$ Immunoglobulin E (IgE). The horizontal dashed red line corresponds to the significance threshold $P = 7.51 \times 10^{-5}$ (FDR Adjusted P-value $\leqslant .05$). Blue and golden colors are used to differentiate the chromosomes.

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 located in the promoter region (24 CpGs, ~26%) and more than one third in the body region (36 CpGs, ~39%) (Supplemental Figure S2). Most regression coefficients for the 92 CpGs (85%, 78 CpGs) were negative (Supplemental Figure S2). That is, at most of the 92 CpGs, higher DNAm was associated with lower IgE, after adjusting for the effects of cord blood IgE, birth weight, and gender. Among the 92 CpGs, 16 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.

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

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

*Est.: estimated regression coefficients.
$$P_{Raw}$: raw *P*-value, P_{FDR}: *P*-value after controlling FDR of .05.
#CHR: Chromosome.

Discussion

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 was consistent with the findings described in Liang et al.[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,[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).[33] The gene *BMI1* regulates T helper 2 (Th2) cell differentiation by interacting with *GATA3* in T cells.[34] Th2 cells promote the production of IgE by switching the IgM antibody isotype to IgE.[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

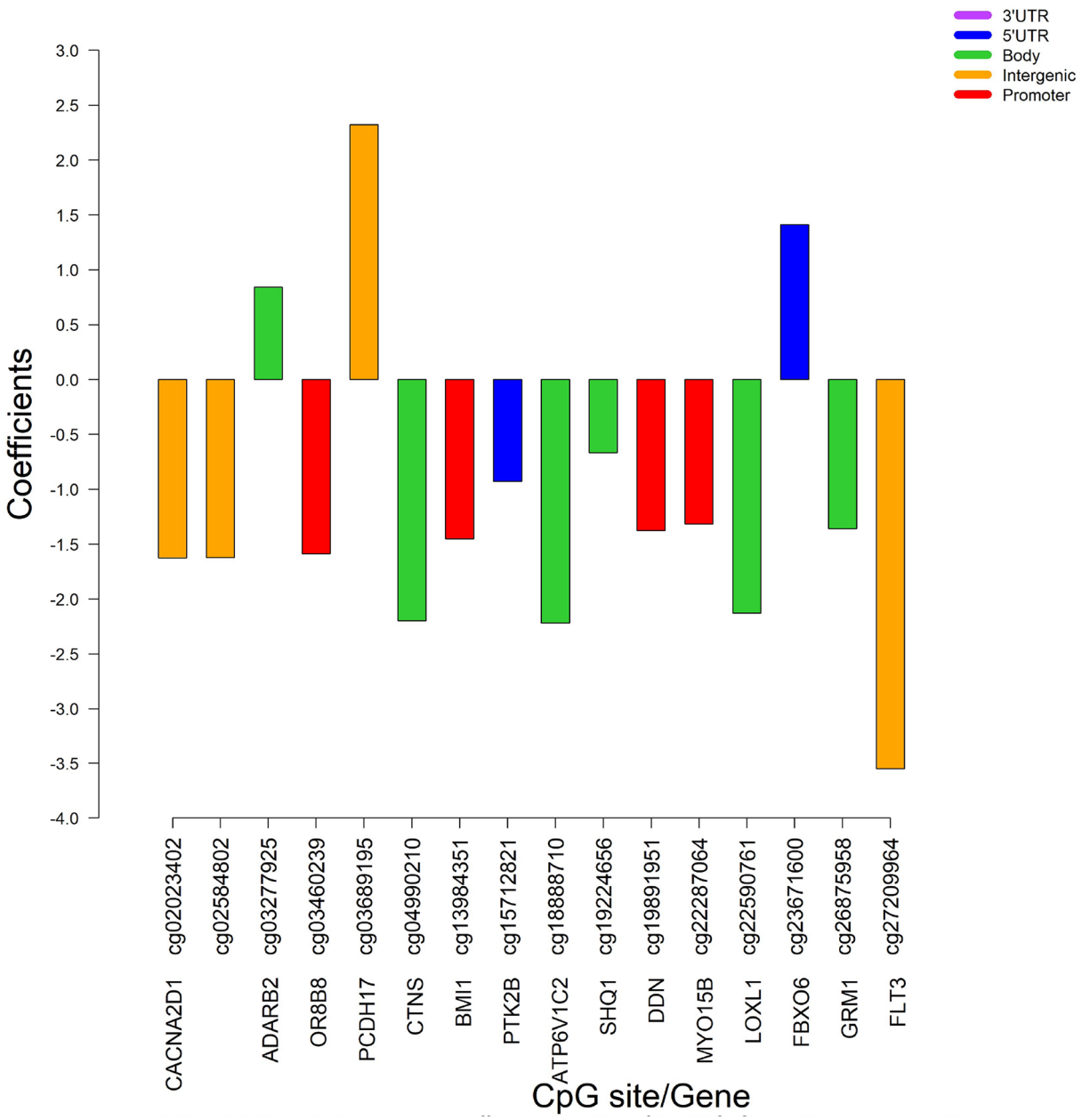

Figure 2. Bar plot displaying regression coefficients, location of the 16 CpGs, and their corresponding genes. Different colors indicate the location of the CpGs on a gene.

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.
$$P_{Raw}$: raw *P*-value.

transplantation.[36,37] A GWAS study has reported chronic obstructive pulmonary disease related loci at or near *ADARB2*.[38] 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[39] and related to recessive distal renal tubular acidosis.[40] *GRM1* is

Table 6. KEGG pathways linked to the 15 genes at statistical significance level of .05.

NAME	PATHWAY	P_{RAW}[$]	HIT COUNT IN GENOME	HIT IN QUERY LIST
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*

The *P*-values included in the table were from gometh function in the R package Limma, and similar *P*-values were obtained in ToppFun.
[$]$P_{Raw}$: 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.[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.[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.[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 ~98% 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 at-birth epigenetic markers for postnatal IgEs, not limited to Asian populations. Furthermore, cell type proportions were inferred based on statistical modeling rather than cell counting and such uncertainty on cellular heterogeneity also contributed to the uncertainty in the identified CpG sites.[44] The small sample sizes (n = 64) in the discovery cohort is another limitation. In this case, the number of CpGs replicated in statistical significance as well as directions of associations was conservative, and thus the identified CpGs were potentially with strong informativity. Since subjects included in the discovery phase were not representative of the whole cohort (n = 299) with respect to pre-pregnant BMI and maternal education, the findings may not represent the original cohort. Thus, generalizability of findings from the current study is limited and should be interpreted with caution. Larger scale studies are warranted to further test the feasibility of the detected CpGs on the association of at birth DNAm with total IgE at later ages.

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 *BMI1*.

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.

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.

Ethics Approval and Consent to Participate

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

ORCID iDs

Luhang Han https://orcid.org/0000-0001-8617-4172

Akhilesh Kaushal https://orcid.org/0000-0002-5137-0461

Hongmei Zhang https://orcid.org/0000-0003-3557-0364

Availability of Data

The data is not available publicly.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Pawankar R, Canonica GW, Holgate ST, Lockey RF, Organization WH. *White Book on Allergy 2011-2012 Executive Summary*. World Allergy Organization, 2011.
2. Tezza G, Mazzei F, Boner A. Epigenetics of allergy. *Early Hum Dev*. 2013;89:S20-S21.
3. Holloway JW, Yang IA, Holgate ST. Genetics of allergic disease. *J Allergy Clin Immunol*. 2010;125:S81-S94.
4. Greer JM, McCombe PA. The role of epigenetic mechanisms and processes in autoimmune disorders. *Biologics*. 2012;6:307-327.
5. Javierre BM, Hernando H, Ballestar E. Environmental triggers and epigenetic deregulation in autoimmune disease. *Discov Med*. 2011;12(67):535-545.
6. Zhang H, Tong X, Holloway JW, et al. The interplay of DNA methylation over time with Th2 pathway genetic variants on asthma risk and temporal asthma transition. *Clin Epigenetics*. 2014;6:8.
7. Lockett GA, Soto-Ramirez N, Ray MA, et al. Association of season of birth with DNA methylation and allergic disease. *Allergy*. 2016;71:1314-1324.
8. Ziyab AH, Karmaus W, Holloway JW, Zhang H, Ewart S, Arshad SH. DNA methylation of the filaggrin gene adds to the risk of eczema associated with loss-of-function variants. *J Eur Acad Dermatol Venereol*. 2013;27:e420-e423.
9. Quraishi BM, Zhang H, Everson TM, et al. Identifying CpG sites associated with eczema via random forest screening of epigenome-scale DNA methylation. *Clin Epigenetics*. 2015;7:68.
10. Ahmad Al, Obaidi AH, Mohamed Al, Samarai AG, Yahya Al, Samarai AK, Al Janabi JM. The predictive value of IgE as biomarker in asthma. *J Asthma*. 2008;45:654-663.
11. Barnetson RS, Wright AL, Benton EC. IgE-mediated allergy in adults with severe atopic eczema. *Clin Exp Allergy*. 1989;19:321-325.
12. Holgate ST, Djukanovic R, Casale T, Bousquet J. Anti-immunoglobulin E treatment with omalizumab in allergic diseases: an update on anti-inflammatory activity and clinical efficacy. *Clin Exp Allergy*. 2005;35:408-416.
13. Chen W, Wang T, Pino-Yanes M, et al. An epigenome-wide association study of total serum IgE in Hispanic children. *J Allergy Clin Immunol*. 2017;140:571-577.
14. Ek WE, Ahsan M, Rask-Andersen M, et al. Epigenome-wide DNA methylation study of IgE concentration in relation to self-reported allergies. *Epigenomics*. 2017;9:407-418.
15. Everson TM, Lyons G, Zhang H, et al. DNA methylation loci associated with atopy and high serum IgE: a genome-wide application of recursive Random Forest feature selection. *Genome Med*. 2015;7:89.
16. Liang L, Willis-Owen SAG, Laprise C, et al. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature*. 2015;520:670-674.
17. Peng C, Cardenas A, Rifas-Shiman SL, et al. Epigenome-wide association study of total serum immunoglobulin E in children: a life course approach. *Clin Epigenetics*. 2018;10:55.
18. Lin L-C, Wang S-L, Chang Y-C, et al. Associations between maternal phthalate exposure and cord sex hormones in human infants. *Chemosphere*. 2011;83:1192-1199.
19. Wang S-L, Su P-H, Jong S-B, Guo YL, Chou W-L, Päpke O. In utero exposure to dioxins and polychlorinated biphenyls and its relations to thyroid function and growth hormone in newborns. *Environ Health Perspect*. 2005;113:1645-1650.
20. Chou WC, Chung YT, Chen HY, et al. Maternal arsenic exposure and DNA damage biomarkers, and the associations with birth outcomes in a general population from Taiwan. *PLoS One*. 2014;9:e86398.
21. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.
22. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol*. 2012;13:R44.
23. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8:203-209.
24. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7:e41361.
25. Koestler DC, Christensen B, Karagas MR, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics*. 2013;8:816-826.
26. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31.
27. Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. *J Allergy Clin Immunol*. 1992;90:235-241.
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)*. 1995;57:289-300.
29. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37:W305-W311.
30. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
31. Liang L, Willis-Owen SAG, Laprise C, Wong KCC, Davies GA, Hudson TJ. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature*. 2015;520:670-674.
32. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1:43-46.
33. Tanaka T, Takada H, Nomura A, Ohga S, Shibata R, Hara T. Distinct gene expression patterns of peripheral blood cells in hyper-IgE syndrome. *Clin Exp Immunol*. 2005;140:524-531.
34. Hosokawa H, Kimura MY, Shinnakasu R, et al. Regulation of Th2 cell development by Polycomb group gene bmi-1 through the stabilization of GATA3. *J Immunol*. 2006;177:7656-7664.
35. Poulsen LK, Hummelshoj L. Triggers of IgE class switching and allergy development. *Ann Med*. 2007;39:440-456.
36. Dierselhuis M, Goulmy E. The relevance of minor histocompatibility antigens in solid organ transplantation. *Curr Opin Organ Transplant*. 2009;14:419-425.
37. Spierings E. Minor histocompatibility antigens: past, present, and future. *Tissue Antigens*. 2014;84:374-360.
38. Cho MH, McDonald ML, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2:214-225.
39. Feng NH, Lin HI, Wang JS, et al. Differential expression of a V-type ATPase C subunit gene, Atp6v1c2, during culture of rat lung type II pneumocytes. *J Biomed Sci*. 2005;12:899-911.
40. Jobst-Schwan T, Klambt V, Tarsio M, et al. Whole exome sequencing identified ATP6V1C2 as a novel candidate gene for recessive distal renal tubular acidosis. *Kidney Int*. 2020;97:567-579.
41. Ege MJ, Strachan DP, Cookson WO, et al. Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol*. 2011;127:138-144, 144 e131-134.
42. Alaskhar Alhamwe B, Khalaila R, Wolf J, et al. Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy Asthma Clin Immunol*. 2018;14:39.
43. Jiang Y, Wei J, Zhang H, et al. Epigenome wide comparison of DNA methylation profile between paired umbilical cord blood and neonatal blood on Guthrie cards. *Epigenetics*. 2020;15(5):454-461.
44. Potaczek DP, Harb H, Michel S, Alhamwe BA, Renz H, Tost J. Epigenetics and allergy: from basic mechanisms to clinical applications. *Epigenomics*. 2017;9:539-571.

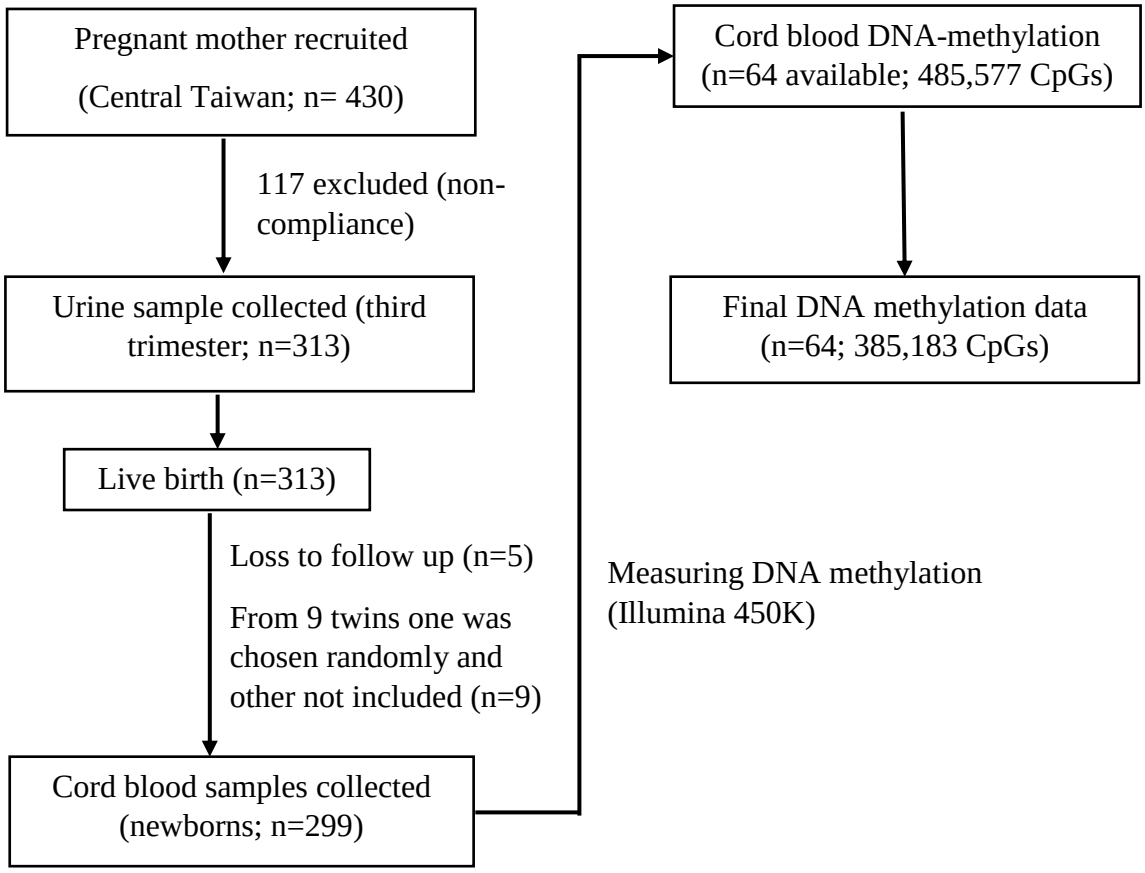

Figure S1. Subject recruitment and preprocessing of DNA methylation data in Taiwanese birth cohort

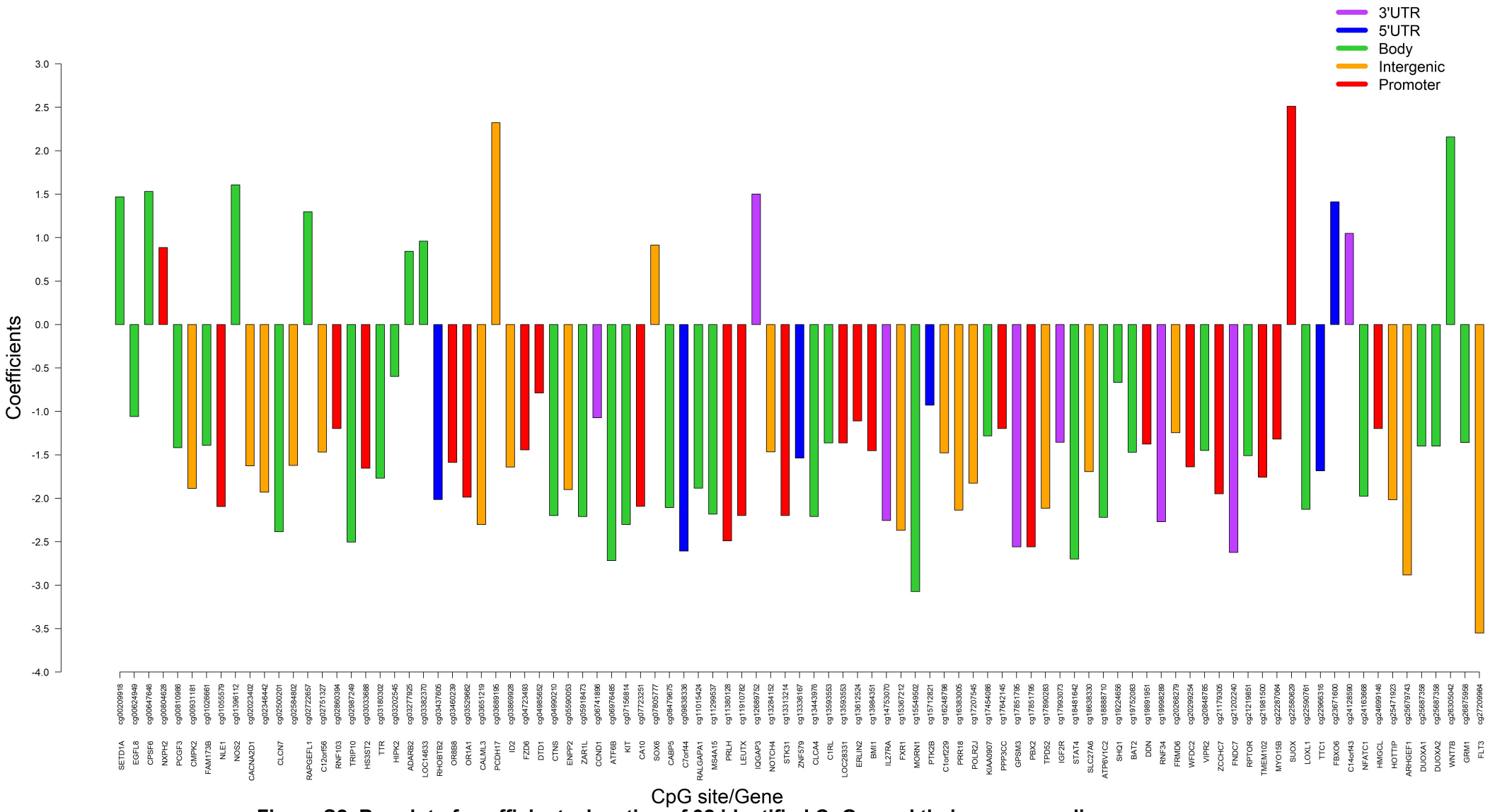

Figure S2. Bar plot of coefficients, location of 92 identified CpGs, and their corresponding genes.

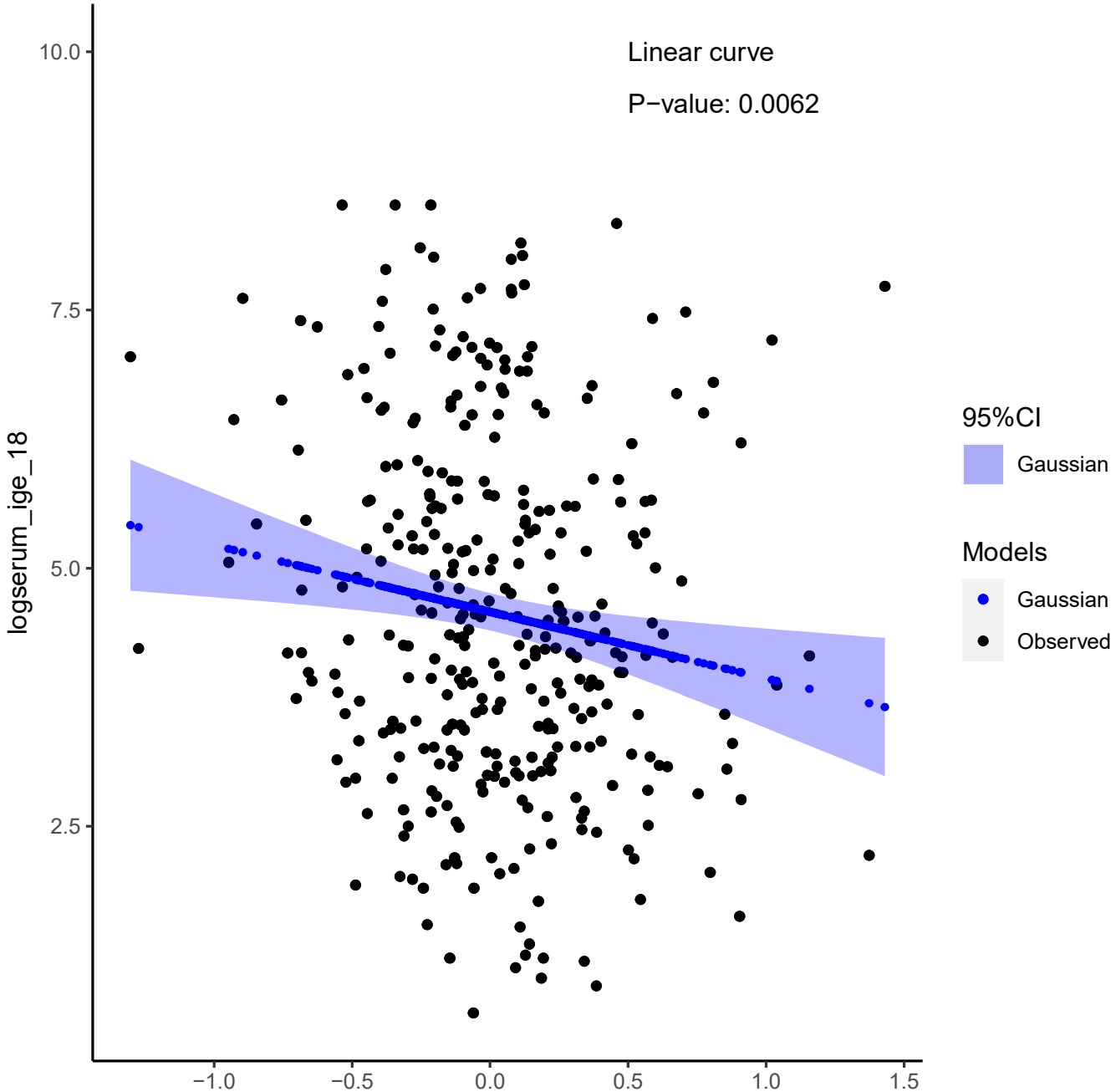
Linear curve
P−value: 0.0062
10.0
7.5
5.0
2.5
logserum_ige_18
−1.0
−0.5
0.0
0.5
1.0
1.5
CpG:cg02584802
95%CI
Gaussian
Models
Gaussian
Observed

Figure S3

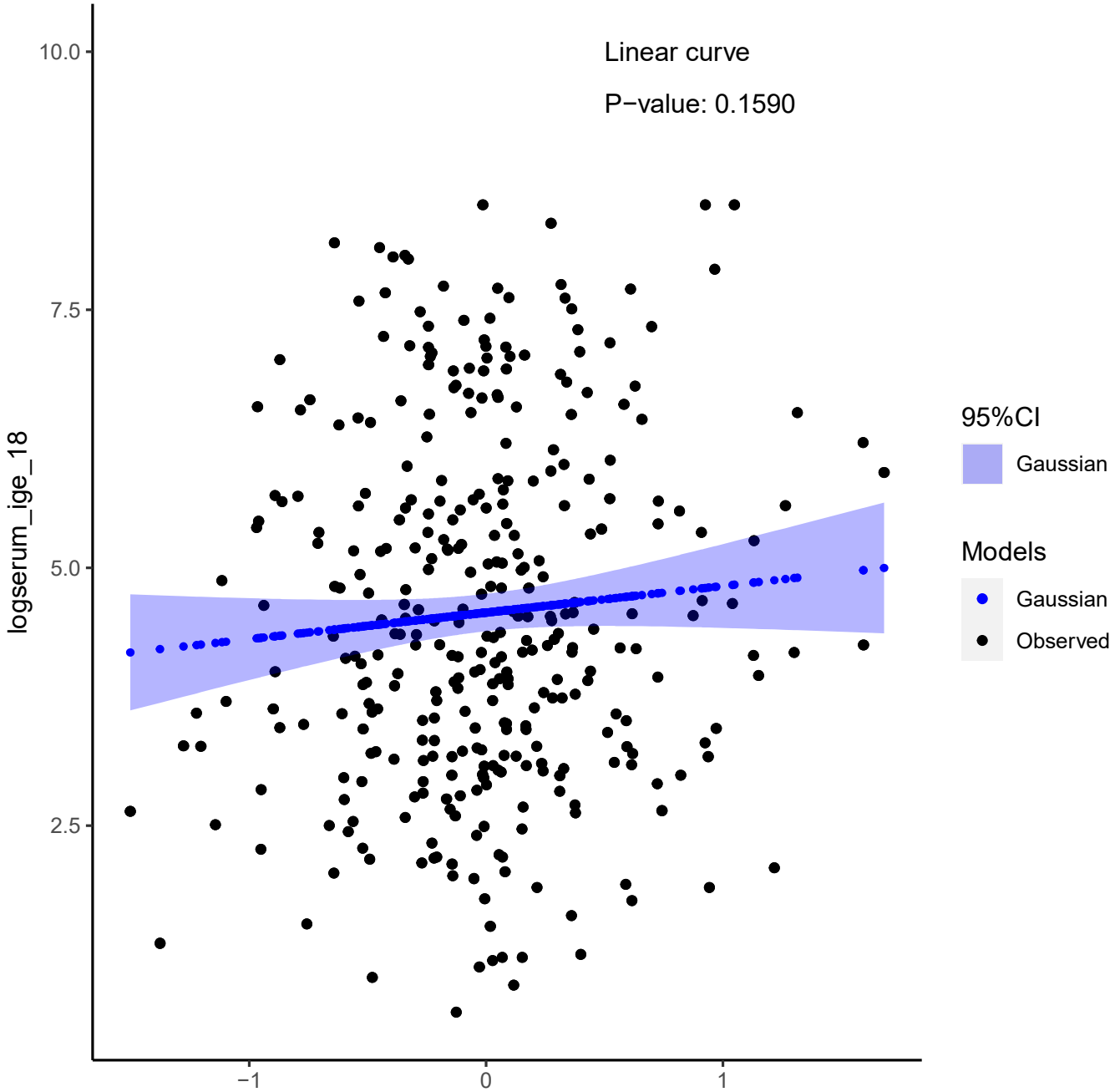
Linear curve
P−value: 0.1590
10.0
7.5
5.0
2.5
logserum_ige_18
−1
0
1
CpG:cg03277925
95%CI
Gaussian
Models
Gaussian
Observed

Figure S4

IoW birth Cohort Description

Isle of Wight cohort consists of the whole population that were born between 1 January 1989 and 28 February 1990 on the Isle of Wight, UK. The local research ethics committee approved the study and of 1536 children born informed consent was obtained for 1456 subjects, detailed description about the cohort is provided elsewhere [1]. This cohort was established to study the natural history of asthma and related allergic disorders with subsequent assessment at 1, 2, 4, 10 and 18 years. In this study n=796 cohort participants were included with IgE and DNA methylation measurement available at birth, age 10 and 18 years. IgE measurement was from cord blood and from peripheral blood of children aged 10- and 18 years, while DNA methylation measurement was from Guthrie card blood and from peripheral blood at age 10 and 18 years.

Assessment of Immunoglobulin E

Total IgE in cord serum was measured using Pharmacia IgE EIA® (Pharmacia Diagnostics AB, Uppsala, Sweden) [2], which is designed to measure IgE between 0.2 to 50 kU/L on 0.1 ml of serum or plasma. IgE at age 10 and age 18 years were determined using PRIST® (Pharmacia Diagnostics AB, Uppsala, Sweden) designed to measure IgE between 2.0 to 1000 kU/L.

DNA-M processing and QA/QC from Guthrie card blood:

DNA from Guthrie cards of n=796 subjects was extracted using a standard salting out procedure [3]. DNA concentration was determined by Qubit quantitation. One microgram of DNA was bisulfite-converted using the EZ 96-DNA methylation kit (Zymo Research, Irvine, CA, USA), following the manufacturer's standard protocol. Epigenome-wide DNA methylation was measured using the Illumina Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates >850,000 CpG sites associated with ~24,000 genes. Multiple identical control samples were assigned to each batch to assess assay variability and samples randomly distributed on arrays to control against batch effects. The

proportion of methylation (β value) calculated for each CpG locus was determined using the methylation module of GenomeStudio software.

Probes not reaching a detection p-value of 10^{-16} in at least 95% of samples were excluded. CpGs on sex chromsomes were also excluded to avoid bias. DNA-M were then pre-processed using the CPACOR pipeline[4] for data from MethylationEPIC [5]. Specifically, the DNA methylation intensity data were quantile normalized using the R computing package, *minfi* [6]. Next, principal components (PCs) inferred based on control probes were used to represent latent variables due to chip-to-chip and technical variation, and the top 15 PCs were included in the analyses to represent latent batch factors [5]. Non-CpG targeting probes with lower methylation than CpGs and 65 control probes, and CpGs on "MULTI" number of chromosomes were excluded. In addition, probes that contained single nucleotide polymorphisms (SNPs) within 10 base pairs of a targeted CpG site with minor allele frequency in at least 0.7% subjects (corresponding to at least 10 subjects in IoW with n=1,456) were excluded due to their influence to the measurement of DNA-M. After pre-processing. A total of 551,710 CpGs were included in subsequent analyses.

Covariates

Child's age, gender, birth weight, Cord blood IgE, top 15 PCs and estimated cell type proportions were included as continuous covariates. Cell type proportions were estimated via the minfi package in R[6,7]. Seven cells are included, CD4T, CD8T, NK, Mono, Bcell, Neu, Eos. In regression models, 6 of the 7 cell types were included as covariates (CD8T cells were excluded) to account for overall cellular heterogeneity.

Statistical Analyses:

For all linear mixed model regressions, we tested the longitudinal relationship between residuals of DNA-M at birth and IgE at 10 and 18 years while adjusting for confounders. DNA-M residual was obtained by regressing DNA-M from Guthrie card on estimated cell proportions from Guthrie card, batch effect as indicator variable, and 15 PCs. The following conceptual model, consistent with the model used in the analyses of data from the Taiwan cohort, was applied to each of the 297 CpG sites identified in the Taiwan cohort and available in the IoW birth cohort (PROC MIXED in SAS, version 9.4).

$log10\ (IgE_{ij}) = Resid_CpG_{ki} + Birth\ Weight_i + Gender_i + CB_IgE_i + Age_{ij}$

Here i = subject (1, 2… 796); j= time point (10, 18); k= k^{th} CpG (1, 2…297);

$Resid_CpG_{ki}$ = DNA-M residual of kth CpG site for i^{th} subject; CB_IgE_i = Cord Blood IgE of ith subject

References:

1. Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. *J Allergy Clin Immunol.* 1992;90(2):235-241.
2. Arshad SH, Twiselton R, Smith J, Hide DW. Influence of genetic and environmental factors on the level of IgE at birth. *Pediatric Allergy and Immunology.* 1992;3(2):79-83.
3. Miller S, Dykes D, Polesky H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215.
4. Lehne B, Drong AW, Loh M, 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;16:37.
5. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome biology.* 2015;16(1):37.
6. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10):1363-1369.
7. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86.