DNA methylation at birth is associated with lung function development till age 26 years

Nandini Mukherjee[1], Ryan Arathimos[2,3,4,5], Su Chen[6], Parnian Kheirkhah Rahimabad[1], Luhang Han[6], Hongmei Zhang[1], John W. Holloway[7,8], Caroline Relton[2,3], A. John Henderson[3,7], Syed Hasan Arshad[7,9,10] , Susan Ewart[11], Wilfried Karmaus[1]

1 Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN, USA
2 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
3 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
4 Social Genetic & Developmental Psychiatry Centre, Kings College London, London, UK
5 NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London, UK
6 Department of Mathematical Sciences, The University of Memphis, Memphis, TN, USA
7 Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK
8 Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK
9 The David Hide Asthma and Allergy Research Centre, Isle of Wight, UK
10 NIHR Southampton Biomedical Research Centre, University Hospital Southampton, UK
11 College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA

Correspondence
Wilfried Karmaus, MD, MPH
School of Public Health, University of Memphis,
301 Robison Hall, 3825 DeSoto Avenue
Memphis, TN 38152
Email: karmaus1@memphis.edu
Tel: +1 (901) 678-2491; Fax: +1(901) 678-1715

Word count Abstract: 250; Body: 2999

Take home message:

In two population-based cohort studies differentially methylated genomic sites at birth associated with lung function from age 10 to 26 years were discovered and replicated. These sites were located on genes involved in lung morphogenesis.

ABSTRACT

Little is known about whether DNA methylation (DNAm) of cytosine-phosphate-guanine (CpG) sites at birth predicts patterns of lung function development. We used heel prick DNAm from the F1-generation of Isle of Wight birth cohort (IOWBC-F1) for discovery of CpGs associated with lung function trajectories (Forced Expiratory Volume, Forced Vital Capacity, their ratio, and Forced Expiratory Flow at 25-75%) over the first 26 years, stratified by sex. We replicated the findings in the Avon Longitudinal Study of Parents and Children (ALSPAC) using cord blood DNAm.

Epigenome-wide screening was applied to identify CpGs associated with lung function trajectories in 396 boys, and 390 girls of IOWBC-F1. Replication in ALSPAC focused on lung function at ages 8, 15 and 24 years. Statistically significantly replicated CpGs were investigated for consistency in direction of association between cohorts, stability of DNAm over time in IOWBC-F1, relevant biological processes, and for association with gene expression (n=161) in IOWBC F2-generation (IOWBC-F2).

Differential DNAm of 8 CpGs on genes *GLUL, MYCN, HLX, LHX1, COBL, COL18A1, STRA6,* and *WNT11* involved in developmental processes, were significantly associated with lung function in the same direction in IOWBC-F1 and ALSPAC, and showed stable patterns at birth, age 10 and 18 years between high and low lung function trajectories in IOWBC-F1. CpGs on *LHX1* and *COL18A1* were linked to gene expression in IOWBC-F2.

In two large cohorts, novel DNAm at birth were associated with patterns of lung function in adolescence and early adulthood providing possible targets for preventative interventions against adverse pulmonary function development.

Abstract word count 250

Keywords DNA-methylation pattern, epigenome-wide, lung function trajectories, prospective, cohort study

Introduction

Lung function parameters are important indicators of pulmonary performance [1] that usually increase from childhood through puberty, level off in early adulthood, and then decline slowly with age [2]. Some individuals however, exhibit persistently low lung function or an initial growth followed by accelerated decline [3, 4]. Reduced childhood lung function may result from adverse *in utero* conditions that mal-adapts the developing fetus to survive in postnatal environments, eventually leading to chronic lung diseases [3-7].

DNA methylation (DNAm) is an epigenetic marker that may retain memories of developmental response of the fetus to *in utero* stimuli [8]. It involves addition or removal of methyl groups to cytosine bases in the cytosine-phosphate-guanine (CpG) sites on the DNA [9] leading to altered gene expression or mRNA splicing [10]. During mammalian embryonic development DNAm undergoes comprehensive erasure after fertilization [11], which is re-established after implantation [11]. Fetal lung development initiates approximately four weeks after fertilization, during which pluripotent embryonic cells differentiate into specialized cell lineages while carrying modified DNAm resulting from *in utero* exposures [12]. These modifications in DNAm may eventually induce structural and functional alterations in the developing lung or alter immune related gene activities, that may increase susceptibility to postnatal stressors potentially leading to reduced lung growth in childhood [13].

Such modifications may be detectable in DNAm measured at birth. Few studies have assessed whether DNAm at birth predicts later lung function [14, 15] at specific time points in life. While lung function at single time points are informative, it does not

reflect the developmental pattern of lung function over time. Lung function trajectories, on the other hand comprise of distinct groups of individuals who develop specific patterns of lung function over time, and are important for early prediction of decline in lung health [3, 4, 16].

No study has yet identified DNAm at birth linked to lung function trajectories from childhood to adulthood. In this sex-stratified epigenome-wide association study (EWAS) we aimed to identify CpGs from heel prick blood that predict lung function trajectories covering ages 10, 18 and 26 years in the F1-generation of Isle of Wight birth cohort (IOWBC-F1), UK. Trajectories of Forced Expiratory Volume (FEV_1), Forced Vital Capacity (FVC), their ratio (FEV_1/FVC), and Forced Expiratory Flow at 25-75% ($FEF_{25-75\%}$) separately in boys and girls of IOWBC-F1, were determined using a group-based method described elsewhere [17]. CpGs linked to lung function trajectories in IOWBC-F1 discovery cohort were tested for replication in Avon Longitudinal Study of Parents and Children (ALSPAC) using cord blood DNAm and lung function measurements at ages 8, 15 or 24 years.

Methods

For data collection and biological assays in IOW and ALSPAC, see online data supplement.

Statistical analyses

An unselected cohort of children born on the Isle of Wight between January 1989 and February 1990 consist of the IOWBC F1-generation (IOWBC-F1). These participants were followed up six times including pregnancy, and their children are being enrolled in IOWBC F2-generation (IOWBC-F2). We conducted an epigenome-wide screening of DNAm at birth associated with lung function trajectories in IOWBC-F1. After quality control, preprocessing, and excluding CpGs with probe-SNPs within ten base pairs with minor allele frequency > 0.007, we analyzed 551,710 CpGs. Lung function trajectories were determined in IOWBC-F1 separately in boys and girls, using unsupervised group-based analyses [18] explained in detail elsewhere [17]. Briefly, we identified two distinct trajectories of FVC, FEV1, FEV_1/FVC and three for FEF_{25-75} showing different lung function at ages 10, 18 and 26 years. We combined mid and low FEF_{25-75} trajectories to consistently retain two trajectories for all lung function outcomes. Lung function trajectory information was available for 577 and 580 boys and girls, respectively, of which heel prick DNAm on Guthrie cards was assessed in 396 boys and 390 girls.

The hierarchical flow of analyses is shown in Figure 1. We used logit transformed β values (M-values, approximated by log2(β /(1-β)) to analyze the association of DNAm with each lung function trajectory separately in boys (n=396) and girls (n=390). In the first screening step, we identified informative CpGs at birth (Illumina 850K) by applying linear regressions within a training and testing approach with the ‘ttscreening’ package

in R [19], adjusting for estimated cell composition [20]. CpGs associated with respective trajectories (p-value≤0.05) in both training and testing data for at least 50% of the iterations were selected. We then determined the risk of being in low trajectory using log-linear models with the identified CpGs as predictors adjusting for cell types, maternal, and paternal asthma, socioeconomic status, and maternal smoking controlling for false discovery rate (FDR) at 0.05.

We replicated our findings in ALSPAC using cord blood DNAm (Illumina 450K) and FVC, FEV1, FEV1/FVC and FEF25-75% measurements at ages 8, 15, or 24 years. Among the discovered CpGs in IOWBC-F1 (Illumina 850K), we tested for replication in ALSPAC (Illumina 450K) only those CpGs also available in 450K data (Figure 1). We analyzed the association between DNAm at birth and lung function at each age separately, using linear regression with the lm() package in R, stratified by sex. The statistical models were adjusted for active or secondhand smoking, height at respective ages and estimated cell composition. Batch effects were adjusted using surrogate variables [21, 22]. Since we applied FDR correction in the discovery phase to eliminate false positives and identify candidates for an independent replication analyses in ALSPAC, we did not consider adjusting for multiple testing again during replication. CpGs identified in IOWBC-F1 that were significantly associated (α≤0.05) in ALSPAC with the respective lung functions at least at one of the ages of 8, 15, or 24 years and showed the same direction association as IOWBC-F1 were deemed to be successfully replicated.

To identify relevant signals, we examined stability of DNAm over time of the successfully replicated DNAm with respect to lung function trajectories in IOWBC-F1.

DNAm was available at birth (n=396 in boys; n= 390 in girls), age 10 (n=128 in boys; n= 93 in girls), and age 18 years (n= 153 in boys; n= 161 in girls). Using repeated measures of DNAm, we analyzed the interaction of time and lung function trajectories on DNAm. CpGs with non-significant interactions (p-value>=0.1) were considered stable. Biological functions of genes corresponding to the stable CpGs were identified using ToppFun [23]. Correlations of cord blood DNAm with gene expression was assessed in IOWBC-F2 (n=161).

Results

Characteristics of study populations of IOWBC-F1 and ALSPAC are provided in Table 1. Boys had a lower proportion of paternal asthma, while girls had a higher proportion of maternal asthma in ALSPAC compared to IOWBC-F1.

We tested 551,710 CpGs (Illumina 850K) in IOWBC-F1 and identified 158-550 CpGs in boys and 95-446 CpGs in girls, respectively to be associated with one of the four low lung function trajectories (Figure 1). Potential confounders (maternal and paternal asthma, socioeconomic status and maternal smoking) did not change the risk ratios by more than 10% (Supplementary Table S1), and thus were excluded from the models. These associations were replicated in ALSPAC (Illumina 450K) using cord blood DNAm and lung function at ages 8, 15, or 24 years, restricting to 47-58% of the discovered CpGs also present in Illumina 450K data (Figure 1). We found 68 CpGs in boys and 58 CpGs in girls to be significantly associated ($\alpha \leq 0.05$) with lung function measures in ALSPAC (Supplementary Table S2). 31 of these 68 CpGs (46%) in boys, and 33 of these 58 CpGs (57%) in girls had the same direction of association in both cohorts (Figure 1 and Supplementary Table S2).

The stability of DNAm measured at birth, and ages 10 and 18 years was evaluated by the assessing the interaction of age and lung function trajectories using repeated measures analyses. Ten of the 31 CpGs (32%) in boys, and fifteen out of 33 CpGs (45%) in girls to have a non-significant interaction terms (p-value<=0.1) indicating no significant change of DNAm over time between trajectories (Supplementary Table S3 and Supplementary Figure S1).

Functional annotation of the genes of ten CpGs in boys and fifteen CpGs in girls using ToppFun [23] revealed eight significant biological processes corresponding to genes of five CpGs in boys and three CpGs in girls namely, *GLUL, MYCN, HLX, LHX1, COBL, COL18A1, STRA6*, and *WNT11.* The biological processes were tube morphogenesis, digestive tract development, paramesonephric and mesonephric duct development, embryonic organ development, female genitalia development, somite and notochord development (Supplementary table S4).

Notably, these five CpGs in boys and three CpGs in girls are associated with lung function in the same direction in IOWBC-F1 and ALSPAC, stable over time in IOWBC-F1, and located on genes with relevant biological functions. In boys, three CpGs were associated with FEF25-75% and two CpGs were associated with FEV1. In girls, two CpGs were linked to FEF25-75% whereas one was linked to FEV1/FVC (Figure 2a). In IOWBC-F1, one unit increase in DNAm (logit-transformed β values) was linked to lower risk of being in the low lung function trajectory for seven CpGs, while a similar increase in DNAm was linked to higher risks for one CpG (Figure 2a). The stable pattern of these CpGs over time in IOWBC-F1 and biological processes associated with their genes are shown in Figure 2a and 2b, respectively.

Additionally, we investigated the association between methylation and gene expression in IOWBC-F2 cord blood samples (n=161). DNAm and gene transcripts were available for five of the above identified eight CpGs. Two CpGs, located in the body region of the genes were significantly correlated with their gene expression (rho=-0.2 p-value=0.008 for cg01885814 (*COL18A1)*; rho=-0.16, p-value=0.04 for cg07103093 (*LHX1*)).

Discussion

This is the first study to identify differentially methylated CpGs at birth that predict the risk of having persistently low lung function from age 10 to age 26 years in two large prospective birth cohorts, IOWBC-F1, and ALSPAC (Figure 1). 31 and 33 CpGs in boys and girls, respectively were significantly associated with lung function the two cohorts in the same direction (Figure 1). In addition, ten out of 31 CpGs in boys, and fifteen out of 33 CpGs in girls showed stable DNAm at birth, age 10 years and 18 years (Supplementary Table S3 and Supplementary Figure S1). Among these, five CpGs in boys and three CpGs in girls), eight belonged to genes involved in embryonic organ development and tube morphogenesis (Supplementary Table S4). In IOWBC-F2, cord blood DNAm of two CpGs were correlated with gene expression.

Well documented evidence regarding sex disparities in lung development [24] and DNAm [25] necessitated separate investigations in boys and girls. Boys and girls in IOWBC-F1 have a different probability of belonging to high or low lung function trajectories [17]. Sex differences in lung development are initiated in embryological stages [26] and continue in adolescence accompanied by sex hormone changes [27, 28]. In parallel, body height and weight that vary between sexes at a given age, also contribute to sex differences in lung function [29]. Sexual dimorphisms in childhood asthma and related immune responses are recognized [26, 28, 30-32]. Sex-differences in DNAm are frequent and stable throughout childhood [25] and known to modify health risks [33, 34].

We prospectively assessed the role of DNAm at birth in predicting lung function from childhood to early adulthood stratified by sex. The hierarchical analyses approach sequentially eliminates non-informative CpGs to select the most relevant signals that

pass the following criteria: a) significant association with lung function trajectories from ages 10 to 26 years in IOWBC-F1 (Discovery phase), b) significant association with lung function at ages 8, 15 or 24 years in ALSPAC in the same direction as IOWBC-F1 (Replication phase), c) stability of DNAm over time in high and low lung function trajectory in IOWBC-F1, d) location on genes with biological role in lung development, e) correlation with gene expression in IOWBC-F2.

The IOWBC-F1 the trajectories provided two groups of participants with specific patterns of lung function development, while in ALSPAC each time specific lung function measured repeatedly in the same individual, represents incomplete information on the developmental pattern. CpGs associated in the same direction with both trajectories and individual lung functions in the two cohorts, identified using two different approaches strengthens the validity of replication.

In the current EWAS, we measured DNAm from heel prick blood in IOWBC-F1 and cord blood in ALSPAC. We have previously shown DNAm to largely agree between heel prick and cord blood [35]. Nevertheless, consistent associations of DNAm from two different blood sources with the same lung function in same direction reinforces the systemic role of these CpGs towards lung function.

Dynamic changes in DNAm over time are common and may reflect the influence of post-natal environmental factors [36, 37], but do not explain the sole contribution of '*in utero*' or genetic factors in the origin of lung function development. To identify latter processes, we considered only those DNAm at birth that covary with lung function from age 10 to 26 years in IOWBC-F1 and remain 'temporally stable'. In other words, these identified DNAm at birth do not change substantially over time between the lung

function trajectories. (Figure 2). However, future studies should explore the *'in utero*' or genetic factors that determine the such patterns of DNAm.

The CpGs linked to lung function in both cohorts in the same direction and showing stable patterns over time in IOWBC-F1 produced statistically significant GO terms, namely, development of embryonic organs, tube, paramesonephric and mesonephric ducts, digestive tract, somite, notochord and female genitalia (Supplementary Table S4) corresponding to eight genes. All eight genes were enriched in the category of tube development that involves intricate branching morphogenesis during embryogenesis of complex tubular organs such as lungs, trachea, kidney, digestive tract, and urinary-genital system [38, 39]. Different subsets of these eight genes were linked to the remaining GO terms. Importance of these biological processes in lung development described in Supplementary Results further substantiates the etiological importance of these DNAm linked to lung function in two cohorts, stable, and located on biologically relevant genes.

Importantly, cord blood DNAm of two of the above eight CpGs, cg01885814 (*COL18A1*), and cg07103093 (*LHX1*) were correlated with their gene expression in IOWBC-F2 that comprises of children of F1-mothers in IOWBC-F1. Same DNAm at birth linked to lung function in IOWBC-F1, and to gene expression in IOWBC-F2 indicates possible mechanistic impact of DNAM on lung function via altered gene expression.

Two recently published epigenetic meta-analyses by Dekker *et al* [15], and Imboden *et al* [40], have linked DNAm and lung function at specific time points. However, there is considerable heterogeneity compared to current EWAS in the design,

analyses, and time points of lung function and DNAm assessments. Unlike the current EWAS, both prior studies [15, 40] analyzed boys and girls together. Imboden *et al* [40] performed cross-sectional analyses of DNAm and lung function, excluding $FEF_{25-75\%,}$ measured during mid- to late adulthood Dekker *et al* performed a prospective EWAS linking differentially methylated regions (DMRs) at birth to childhood lung function and only assessed consistency of these associations with later lung function in other cohorts [15]. Neither studies consider patterns of lung function development over time and stability of DNAm. Despite these differences we found DNAm of the genes *DFNB31, FBXO2, AMPD3* to be linked to lung function in these studies and the current EWAS, indicating their importance towards lung function.

We observed ~54% (37 of the 68 CpGs) in boys and ~43% (25 of 58 CpGs) in girls with discordant direction of associations of DNAm with lung function between ALSPAC and IOWBC-F1. Contrary to prior studies [15, 40], we did not consider directionally discordant CpGs as replicated, since they significantly increased lung function in one cohort and decreased in another. We performed additional assessments to explain the disagreements. First, we removed cell types from statistical models in IOWBC-F1, to explore the impact of cell type induced multicollinearity [41], however, it did not change the directionality. Second, we stratified the analysis in IOWBC-F1 by parental and offspring characteristics (details in Supplementary results) and compared the effects in each stratum with results in ALSPAC. The underlying rationale is that distribution of risk factors could be different between IOWBC-F1 and ALSPAC. For instance, in ALSPAC compared to IOWBC a higher proportion of girls have mothers with asthma, whereas a lower proportion of boys have fathers with asthma. Thus, these

risk factors may have influenced the distribution of DNAm and lung function in offspring. Such different distributions of risk factors between discovery and replication cohorts can result in effects of opposite directions. However, differences between those with and without maternal/paternal asthma could be captured in stratified risk estimation. To this end, we found that in specific strata of paternal, and maternal history of asthma, low birth weight, and asthma, and eczema at age 4 years, the direction of risks in IOWBC-F1 were comparable to ALSPAC, reducing directional discordance from ~54% (37 of the 68 CpGs) to ~28% (19 out of 68 CpGs) in boys, and ~43% (25 of 58 CpGs) to ~17% (10 of 58 CpGs) in girls (Figure 3, Supplementary Table S5(a) and S5(b)). Nevertheless, this observation needs further validation from other studies.

One limitation in the current EWAS is DNAm of a higher number of CpGs were measured in IOWBC-F1 (Illumina 850K) compared to ALSPAC (Illumina 450K), restricting the replication to only 55% of the CpGs identified in IOWBC-F1 that were available in ALSPAC. We also could not explore in IOWBC-F2 whether effects of DNAm on lung function were mediated via gene expression due to few children with lung function currently being enrolled. We identified heel prick or cord blood DNAm, that reflects prenatal effects but are not tissue specific. However, according to the online LungMAP Consortium data repository https://lungmap.net/, all eight biologically relevant genes were expressed in human lungs in early life and adulthood (Supplementary Figure S2) substantiating the potential of these CpGs as robust biomarkers of lung function development. This is the first epigenome-wide association study linking DNAm at birth with lung function trajectories in the first 26 years of life. Future studies need to

identify risk factors affecting these DNAm to develop epigenetic interventions to prevent early decline of lung function.

Acknowledgements

The authors gratefully acknowledge the participation and cooperation of the children and parents of Isle of Wight and appreciate the hard work of Mrs. Sharon Matthews and the Isle of Wight research team in collecting data and Nikki Graham for technical support. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the methylation data.

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We would like to thank Dr. Gemma Sharp for her contribution towards the quality control of the ARIES data.

Funding

This work has been supported by National Institutes of Health/ National Institute of Allergy and Infectious Diseases [R01 AI091905 and R01HL132321 to W.K., R01 AI121226 to H.Z. and J.W.H.]; National Asthma Campaign, UK [364]; and the National Institutes of Health / National Heart, Lung, and Blood Institute [R01 HL082925 to S.H.A.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, USA, and National Asthma Campaign, UK.

The Accessible Resource for Integrated Epigenomics Studies (ARIES) which generated large scale methylation data was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1). Additional epigenetic profiling on the ALSPAC cohort was supported by the UK Medical Research Council Integrative Epidemiology Unit and the University of Bristol (MC_UU_12013_1, MC_UU_12013_2, MC_UU_12013_5 and MC_UU_12013_8), the Wellcome Trust (WT088806) and the United States National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK10324).The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC.

References:

1. Fabbri LM, Romagnoli M, Corbetta L, Casoni G, Busljetic K, Turato G, Ligabue G, Ciaccia A, Saetta M, Papi A. Differences in airway inflammation in patients with fixed airflow obstruction due to asthma or chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2003: 167(3): 418-424.
2. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, Cho MH, Croteau-Chonka DC, Castaldi PJ, Jain G, Sanyal A, Zhan Y, Lajoie BR, Dekker J, Stamatoyannopoulos J, Covar RA, Zeiger RS, Adkinson NF, Williams PV, Kelly HW, Grasemann H, Vonk JM, Koppelman GH, Postma DS, Raby BA, Houston I, Lu Q, Fuhlbrigge AL, Tantisira KG, Silverman EK, Tonascia J, Weiss ST, Strunk RC. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med* 2016: 374(19): 1842-1852.
3. Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, Bowatte G, Gurrin L, Johns DP, Thompson BR, Hamilton GS, Frith PA, James AL, Thomas PS, Jarvis D, Svanes C, Russell M, Morrison SC, Feather I, Allen KJ, Wood-Baker R, Hopper J, Giles GG, Abramson MJ, Walters EH, Matheson MC, Dharmage SC. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med* 2018: 6(7): 535-544.
4. Berry CE, Billheimer D, Jenkins IC, Lu ZJ, Stern DA, Gerald LB, Carr TF, Guerra S, Morgan WJ, Wright AL, Martinez FD. A Distinct Low Lung Function Trajectory from Childhood to the Fourth Decade of Life. *Am J Respir Crit Care Med* 2016: 194(5): 607-612.
5. Majewska R, Pac A, Mroz E, Spengler J, Camann D, Mrozek-Budzyn D, Sowa A, Jacek R, Wheelock K, Perera FP. Lung function growth trajectories in non-asthmatic children aged 4-9 in relation to prenatal exposure to airborne particulate matter and polycyclic aromatic hydrocarbons - Krakow birth cohort study. *Environ Res* 2018: 166: 150-157.
6. Mandy M, Nyirenda M. Developmental Origins of Health and Disease: the relevance to developing nations. *Int Health* 2018: 10(2): 66-70.
7. Tee AK. Chronic Obstructive Pulmonary Disease (COPD): "Not a Cigarette Only Pulmonary Disease". *Ann Acad Med Singapore* 2017: 46(11): 415-416.
8. Bianco-Miotto T, Craig JM, Gasser YP, van Dijk SJ, Ozanne SE. Epigenetics and DOHaD: from basics to birth and beyond. *J Dev Orig Health Dis* 2017: 8(5): 513-519.
9. Chen ZX, Riggs AD. DNA methylation and demethylation in mammals. *J Biol Chem* 2011: 286(21): 18347-18353.
10. Shenker N, Flanagan JM. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. *Br J Cancer* 2012: 106(2): 248-253.
11. Wadhwa PD, Buss C, Entringer S, Swanson JM. Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med* 2009: 27(5): 358-368.
12. Durham AL, Adcock IM. Basic science: Epigenetic programming and the respiratory system. *Breathe* 2013: 9(4): 278-288.
13. Joss-Moore LA, Albertine KH, Lane RH. Epigenetics and the developmental origins of lung disease. *Mol Genet Metab* 2011: 104(1-2): 61-66.
14. Gao L, Urman R, Millstein J, Siegmund KD, Dubeau L, Breton CV. Association between AXL promoter methylation and lung function growth during adolescence. *Epigenetics* 2018: 1-12.
15. den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, Rifas-Shiman SL, Ruiz-Arenas C, Amin N, Bustamante M, DeMeo DL, Henderson AJ, Howe CG, Hivert MF, Ikram MA, de Jongste JC, Lahousse L, Mandaviya PR, van Meurs JB, Pinart M, Sharp GC, Stolk L, Uitterlinden AG, Anto JM, Litonjua AA, Breton CV, Brusselle GG, Sunyer J, Smith GD, Relton CL, Jaddoe VWV, Duijts L. Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. *Eur Respir J* 2019: 53(4).

16. Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, Guerra S, Marott JL, Martinez FD, Martinez-Camblor P, Meek P, Owen CA, Petersen H, Pinto-Plata V, Schnohr P, Sood A, Soriano JB, Tesfaigzi Y, Vestbo J. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2015: 373(2): 111-122.
17. Karmaus W, Mukherjee N, Janjanam VD, Chen S, Zhang H, Roberts G, Kurukulaaratchy RJ, Arshad H. Distinctive lung function trajectories from age 10 to 26 years in men and women and associated early life risk factors - a birth cohort study. *Respir Res* 2019: 20(1): 98.
18. Nagin DS, Jones BL, Passos VL, Tremblay RE. Group-based multi-trajectory modeling. *Statistical Methods in Medical Research* 2018: 27(7): 2015-2023.
19. Ray MA, Tong X, Lockett GA, Zhang H, Karmaus WJ. An Efficient Approach to Screening Epigenome-Wide Data. *Biomed Res Int* 2016: 2016: 2615348.
20. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012: 13: 86.
21. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007: 3.
22. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012: 28(6): 882-883.
23. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009: 37(Web Server issue): W305-311.
24. Boezen HM, Jansen DF, Postma DS. Sex and gender differences in lung development and their clinical significance. *Clin Chest Med* 2004: 25(2): 237-245.
25. Suderman M, Simpkin A, Sharp G, Gaunt T, Lyttleton O, McArdle1 W, Ring, Susan , Smith GD, Relton C. Sex-associated autosomal DNA methylation differences are wide-spread and stable throughout childhood. *bioRxiv preprint* 2017.
26. LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe (Sheff)* 2018: 14(2): 131-140.
27. DeBoer MD, Phillips BR, Mauger DT, Zein J, Erzurum SC, Fitzpatrick AM, Gaston BM, Myers R, Ross KR, Chmiel J, Lee MJ, Fahy JV, Peters M, Ly NP, Wenzel SE, Fajt ML, Holguin F, Moore WC, Peters SP, Meyers D, Bleecker ER, Castro M, Coverstone AM, Bacharier LB, Jarjour NN, Sorkness RL, Ramratnam S, Irani AM, Israel E, Levy B, Phipatanakul W, Gaffin JM, Gerald Teague W. Effects of endogenous sex hormones on lung function and symptom control in adolescents with asthma. *BMC Pulm Med* 2018: 18(1): 58.
28. Soto-Ramirez N, Ziyab AH, Karmaus W, Zhang H, Kurukulaaratchy RJ, Ewart S, Arshad SH. Epidemiologic methods of assessing asthma and wheezing episodes in longitudinal studies: measures of change and stability. *J Epidemiol* 2013: 23(6): 399-410.
29. Park JE, Chung JH, Lee KH, Shin KC. The effect of body composition on pulmonary function. *Tuberc Respir Dis (Seoul)* 2012: 72(5): 433-440.
30. Postma DS. Gender differences in asthma development and progression. *Gend Med* 2007: 4 Suppl B: S133-146.
31. Jones BG, Penkert RR, Surman SL, Sealy RE, Pelletier S, Xu B, Neale G, Maul RW, Gearhart PJ, Hurwitz JL. Matters of life and death: How estrogen and estrogen receptor binding to the immunoglobulin heavy chain locus may influence outcomes of infection, allergy, and autoimmune disease. *Cell Immunol* 2019: 346: 103996.
32. Melen E, Bergstrom A, Kull I, Almqvist C, Andersson N, Asarnoj A, Borres MP, Georgellis A, Pershagen G, Westman M, van Hage M, Ballardini N. Male sex is strongly associated with IgE-sensitization to airborne but not food allergens: results up to age 24 years from the BAMSE birth cohort. *Clin Transl Allergy* 2020: 10: 15.

33. Asllanaj E, Zhang X, Ochoa Rosales C, Nano J, Bramer WM, Portilla-Fernandez E, Braun KVE, Gonzalez-Jaramillo V, Ahrens W, Ikram A, Ghanbari M, Voortman T, Franco OH, Muka T, Glisic M. Sexually dimorphic DNA-methylation in cardiometabolic health: A systematic review. *Maturitas* 2020: 135: 6-26.
34. Kim S, Cho YH, Won S, Ku JL, Moon HB, Park J, Choi G, Kim S, Choi K. Maternal exposures to persistent organic pollutants are associated with DNA methylation of thyroid hormone-related genes in placenta differently by infant sex. *Environ Int* 2019: 130: 104956.
35. Jiang Y, Wei J, Zhang H, Ewart S, Rezwan FI, Holloway JW, Arshad H, Karmaus W. Epigenome wide comparison of DNA methylation profile between paired umbilical cord blood and neonatal blood on Guthrie cards. *Epigenetics* 2020: 15(5): 454-461.
36. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008: 299(24): 2877-2883.
37. Urdinguio RG, Torro MI, Bayon GF, Alvarez-Pitti J, Fernandez AF, Redon P, Fraga MF, Lurbe E. Longitudinal study of DNA methylation during the first 5 years of life. *J Transl Med* 2016: 14(1): 160.
38. Lubarsky B, Krasnow MA. Tube morphogenesis: making and shaping biological tubes. *Cell* 2003: 112(1): 19-28.
39. Nelson WJ. Tube morphogenesis: closure, but many openings remain. *Trends Cell Biol* 2003: 13(12): 615-621.
40. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, Beckmeyer-Borowko A, Harris SE, Starr JM, Deary IJ, Flexeder C, Waldenberger M, Peters A, Schulz H, Chen S, Sunny SK, Karmaus WJJ, Jiang Y, Erhart G, Kronenberg F, Arathimos R, Sharp GC, Henderson AJ, Fu Y, Piirila P, Pietilainen KH, Ollikainen M, Johansson A, Gyllensten U, de Vries M, van der Plaat DA, de Jong K, Boezen HM, Hall IP, Tobin MD, Jarvelin MR, Holloway JW, Jarvis D, Probst-Hensch NM. Epigenome-wide association study of lung function level and its change. *Eur Respir J* 2019.
41. Barton SJ, Melton PE, Titcombe P, Murray R, Rauschert S, Lillycrop KA, Huang R-C, Holbrook JD, Godfrey KM. In Epigenomic Studies, Including Cell-Type Adjustments in Regression Models Can Introduce Multicollinearity, Resulting in Apparent Reversal of Direction of Association. *Frontiers in Genetics* 2019: 10(816).

Footnotes

Posthumous authorship has been given to Professor A. John Henderson for his critical contribution to the proposed analyses and the data acquisition in ALSPAC.

Tables

Table 1 (a): Characteristics of boys with lung function and DNA methylation in IOWBC-F1 and ALSPAC cohort

	IOWBC-F1 ¥			ALSPAC		
n	age 10 years (n=392)	age 18 years (n=367)	age 26 years (n=274)	age 8 years (n=357)	age 15 years (n=262)	age24 years (n=155)
Outcome	Median (min, max)					
FEV1 (Liters)	2.03 (1.41, 3.02)	4.52 (2.8, 6.7)	4.59 (2.64, 6.65)	1.7 (0.7,2.4)	3.7 (1.9,6.1)	4.5 (2.3,6.1)
FVC (Liters)	2.32 (1.47, 3.67)	5.34 (3.45, 7.09)	5.8 (3.42, 8.07)	2.0 (1.1,3.1)	4.2 (2.2,7.1)	5.5 (3.2,7.7)
FEV1/FVC	0.88 (0.71, 1.00)	0.87 (0.61, 1.0)	0.79 (0.58, 0.9)	0.8 (0.5,0.9)	0.8 (0.6,1)	0.8 (0.6,0.9)
FEF25-75% (Liters)	2.36 (1.13, 4.35)	4.93 (2.13, 9.6)	4.15 (1.45, 7.19)	2.1 (0.3,3.7)	4.3 (1.6,8.1)	4.5 (1.8,7.8)
Height (cm)	139.2 (122.9, 161.6)	177.5 (152.0, 195.0)	179.5 (153.5, 196.0)	133.6 (115.4,157.4)	175.0 (147.4,198)	180.9 (162,198)
	n (%)					
Concurrent smoking*	250 (46.4)	241 (65.6)	125 (45.6)	78 (21.8)	141 (53.8)	107 (69.0)
Maternal asthma (Yes)	45 (12)	41 (11.17)	32 (11.7)	44 (12.3)	28 (10.6)	21 (13.5)
Paternal asthma (Yes)	40 (10.2)	40 (10.90)	27 (9.85)	16 (4.5)	13 (4.9)	9 (5.81)

Table 1 (b): Characteristics of girls with lung function and DNA methylation in IOWBC-F1 and ALSPAC cohort

	IOWBC-F1 £			ALSPAC		
n	age 10 years (n=387)	age 18 years (n=377)	age 26 years (n=332)	age 8 years (n=351)	age 15 years (n=314)	age24 years (n=257)
Outcome	Median (min, max)					
FEV1 (Liters)	1.98 (1.21, 3.03)	3.49 (1.4, 4.8)	3.42 (2.24, 4.60)	1.6 (1.03,2.5)	3.0 (1.2,4.5)	3.3 (2.23,4.9)
FVC (Liters)	2.21 (1.3, 3.24)	3.9 (2.3, 5.8)	4.25 (3.02, 6.64)	1.8 (1.1,2.9)	3.3 (1.9,5.2)	3.9 (2.45,5.64)
FEV1/FVC	0.9 (0.64, 1.0)	0.89 (0.58, 1.00)	0.82 (0.61, 0.99)	0.89 (0.6,1)	0.9 (0.48,1)	0.84 (0.5,1)
FEF25-75% (Liters)	2.48 (0.94, 4.42)	4.04 (0.77, 5.98)	3.4 (1.2, 6.05)	2.1 (0.2,3.5)	3.7 (0.13,6.27)	3.4 (0.1,6.1)
Height (cm)	138.4 (122.5, 158.5)	164.0 (139.0, 181.0)	165.5 (150.0, 180.5)	132.4 (114.9,152.1)	165.5 (146,183.5)	167.1 (153.3,182.7)
	n (%)					
Concurrent smoking *	154 (39.8)	250 (66.3)	135 (40.6)	85 (24.2)	210 (66.8)	176 (68.52)
Maternal asthma (Yes)	39 (10.08)	38 (10.1)	34 (10.2)	62 (17.6)	54 (17.2)	45 (17.5)
Paternal asthma (Yes)	31 (8.01)	29 (7.7)	27 (8.13)	21 (5.9)	18 (5.7)	18 (7.0)

*Concurrent smoking refers to either active or secondhand smoking (maternal/paternal/other smoking/outside home smoking) or both.

¥ Among the 396 boys included in analyses, 392, 367, and 274 have lung function measured at age 10, 18 and age 26 years, respectively

£ Among the 390 girls included in analyses, 387, 377, and 332 have lung function measured at age 10, 18, and age 26 years, respectively.

Hierarchical steps	Number of CpGs in Boys					Number of CpGs in Girls				
Discovery in IOWBC-F1	FEV1	FVC	FEV1/ FVC	FEF*	Total	FEV1	FVC	FEV1/ FVC	FEF*	Total
CpGs tested from heel prick DNA methylation (Illumina 850K)	551,710 CpGs tested for each outcome					551,710 CpGs tested for each outcome				
CpGs associated with lung function trajectories identified using 'ttscreening'	275	158	161	550	**1144**	95	147	178	446	**866**
CpGs passed FDR after adjusting for confounders using log-linear regression to obtain risk ratios for lung function trajectories	275	158	161	550	**1144**	95	147	178	446	**866**
Replication in ALSPAC**										
CpGs available in cord blood DNA methylation (Illumina 450K)	146	91	77	304	**608**	45	80	103	254	**482**
CpGs significantly associated with lung function (α <=0.05) at ages 8, 15, and 24 years.	21	10	5	32	**68**	8	4	15	31	**58**
CpGs significantly associated in ALSPAC (α <=0.05) in same direction as IOWBC.	14	5	1	11	**31**	3	1	8	21	**33**
Additional assessments in IOWBC										
Stability of DNA methylation at birth, age 10, and age 18 years in **IOWBC-F1**¥	5	1	1	3	**10**	0	0	4	11	**15**
CpGs on biologically relevant genes#	2	0	0	3	**5**	0	0	1	2	**3**
DNA methylation of CpGs correlated with gene expression in **IOWBC-F2**	1	0	0	1	**2**	0	0	0	0	**0**

Figure 1: Hierarchical assessments to identify potential CpGs predicting lung function.

* FEF refers to FEF 25-75% **Only the CpGs available in Illumina 450K array were tested in ALSPAC ¥Stability of DNAm over time is determined by assessing the interaction of age and lung function trajectory on repeated measures of DNAm at birth, age 10 years, and age 18 years # Biological relevance of the genes was assessed using ToppFun application of the ToppGene Suite

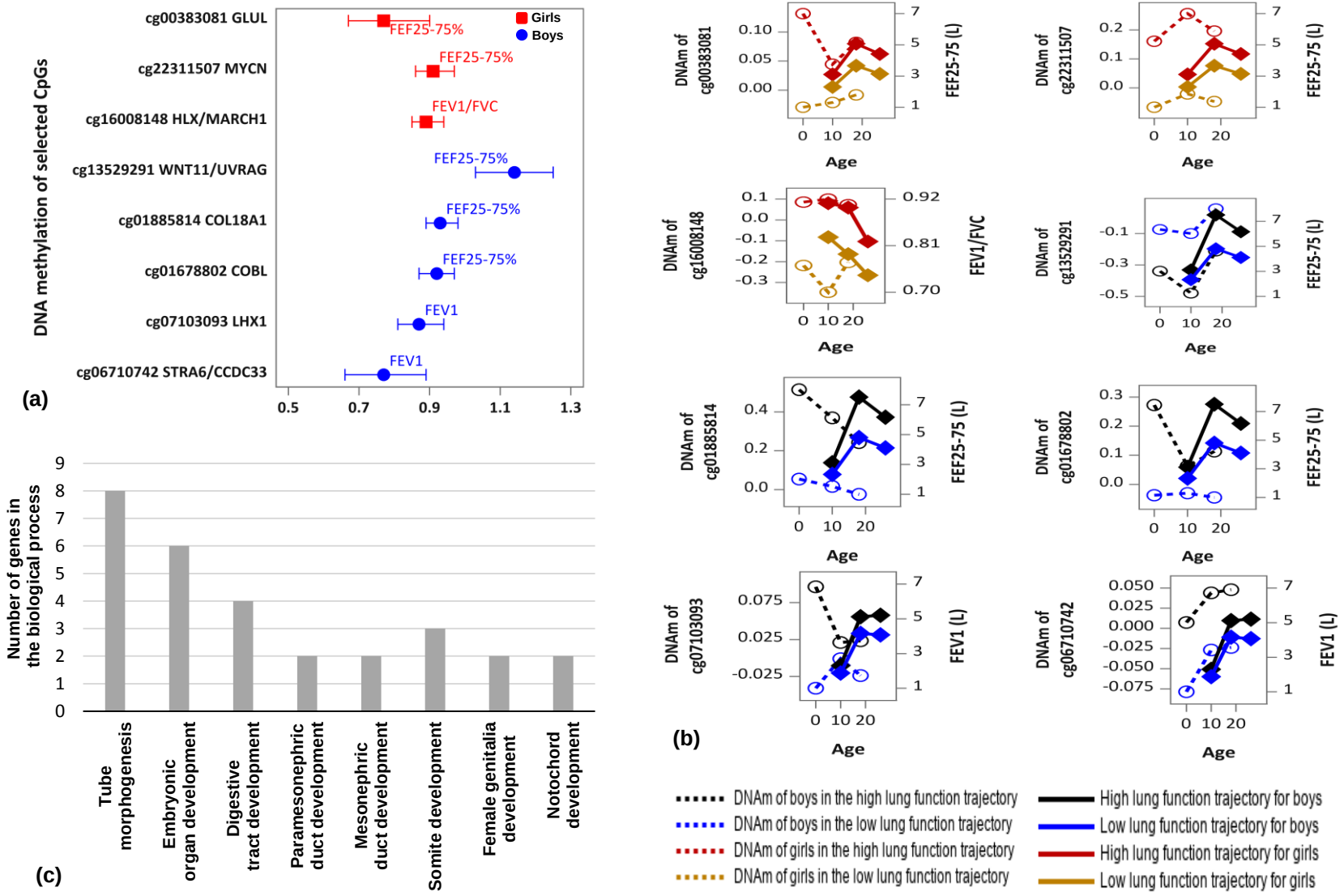


Figure 2: Attributes of eight important CpGs significantly associated with lung function in the same direction in IOWBC-F1 and ALSPAC in the same direction, stable, and enriched in relevant biological processes (a) Risk ratios (95% C.I.) for lower lung function trajectory corresponding to each CpG in Isle of Wight cohort F1 generation; (b) The pattern of consistently different DNAm at birth, age 10 and 18 years in participants belonging to the high and low trajectories of lung function in IOWBC in boys (black and blue lines) and girls (red and orange lines). The left-hand Y-axis represents the residual DNAm after regressing out the effect of cell types for each time point and the right-hand Y-axis represents the lung function levels (in Liters). Stability of DNAm is determined by assessing the interaction of age and lung function trajectory on repeated measures of DNAm at birth, age 10 years, and age 18 years Gene names for the CpGs are: cg00383081 (GLUL), cg22311507 (MYCN), cg16008148 (HLX/MARC1), cg13529291 (WNT11/UVRAG), cg01885814 (COL18A1), cg01678802 (COBL), cg07103093 (LHX1), cg06710742 (STRA6/CCDC33); (c) Biological process involving the genes of the eight CpGs.

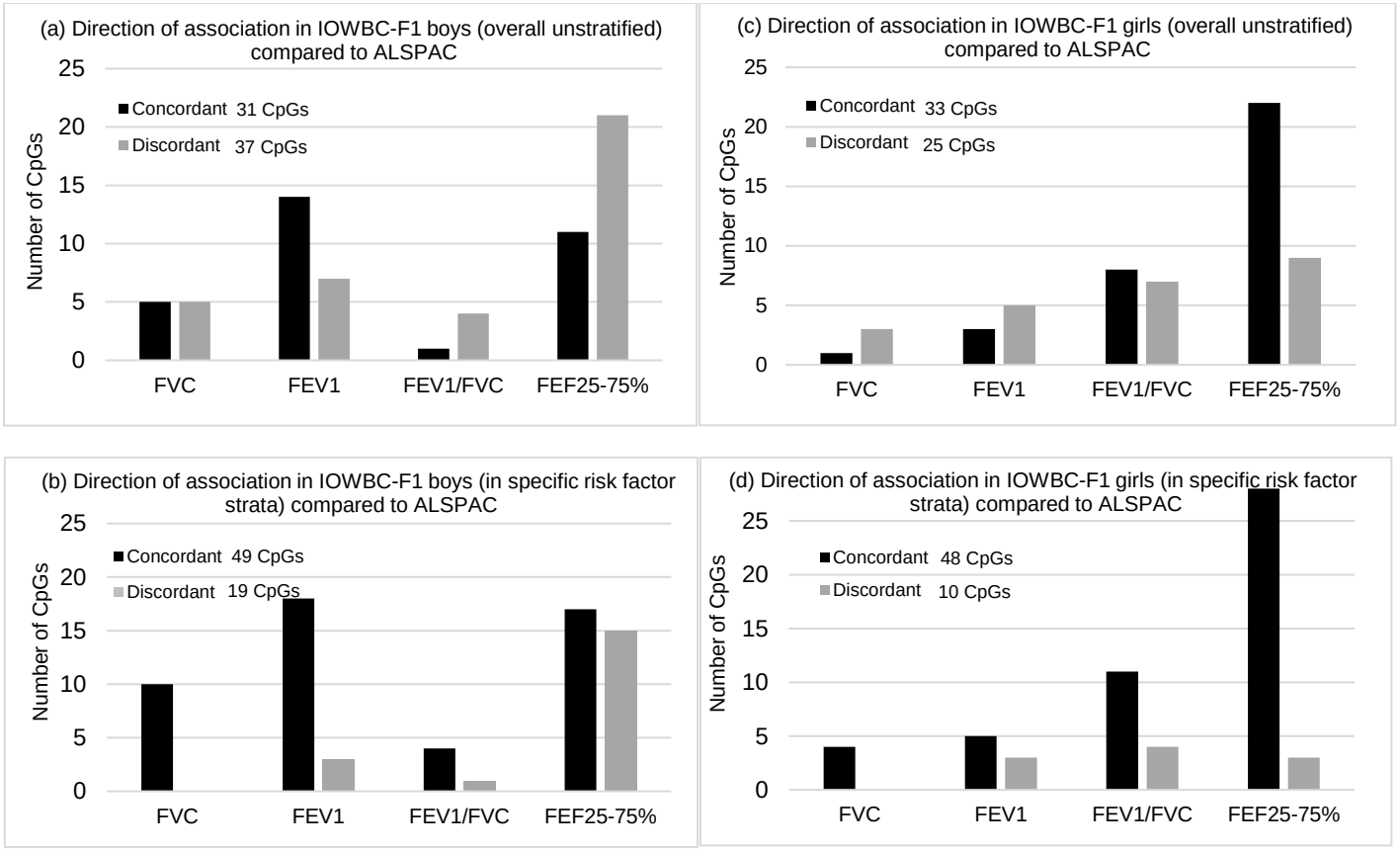


Figure 3: Agreement of direction of association of DNAm with lung function comparing overall sample of boys and girls in IOWBC-F1 to ALSPAC (a and c), and comparing risk factor strata of boys and girls in IOWBC-F1 to ALSPAC (b and d). Black and grey bars indicate the number of CpGs in IOWBC-F1 that have concordant and discordant direction of association, respectively compared to ALSPAC. Associations in IOWBC-F1 estimated in (a) overall unstratified sample of boys: 54% (37/68) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC (b) in risk factor strata of boys: 28% (19/68) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC (c) overall unstratified sample of girls: 43% (25/58) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC (d) in risk factor strata of girls: 17% (10/58) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC. The risk factors used for stratification were prenatal paternal, and maternal history of asthma, low birth weight, asthma, and eczema at age 4 years.

Online data supplement

DNA methylation at birth associated with lung function development in the first 26 years of life.

Nandini Mukherjee, Ryan Arathimos, Su Chen, Parnian Kheirkhah Rahimabad, Luhang Han, Hongmei Zhang, John W. Holloway, Caroline Relton, A. John Henderson, Syed Hasan Arshad, Wilfried Karmaus

Methods

Study population of Discovery cohort: Isle of Wight birth cohort (IOWBC)

The F1 generation of Isle of Wight population-based birth cohort (IOWBC-F1) was established in 1989 in the UK to prospectively study natural history of allergic conditions. 1,536 F1-children were born on the IOW, from January 1989 to February 1990, among which, 1,456 F0-mother-F1-child pairs were enrolled into the cohort study after excluding adoptions, perinatal deaths, and refusals. F1 children were followed up at 1, 2, 4, 10, 18 and 26 years of age with a high (80-90%) follow up rate [1, 2]. The local research ethics committee (South Central - Hampshire B Research Ethics Committee) and the Institutional Review Board of the University of Memphis approved the study. Written parental or child's consent (at age 18 years and later) was provided by all participants at recruitment and each follow-up. The IOW birth cohort has been described in detail elsewhere [2-5]. IOWBC is a dynamic cohort and the number of participants included at each age for lung function assessment is demonstrated in prior publication on lung function trajectory development [6]. Among the 577 and 580 boys and girls with lung function trajectory information, DNAm was available for 396 boys and 390 girls collected on Guthrie cards within a week after birth. Among these participants with complete information of lung function and DNAm, there was very low (<1%) missingness in any other covariates used.

The F2 generation of Isle of Wight population-based birth cohort (IOWBC-F2) are the children of the IOWBC-F1 women who were followed up during pregnancy. These children are currently being enrolled and followed up. Cord blood DNAm and gene expression was assessed in 161 children at birth [2].

Phenotypes

Lung function outcomes were assessed using the KoKo spirometry software package on a portable desktop device (PDS Instrumentation, Louisville, KY, USA) [7]. Tests were performed following the guidelines of the American Thoracic Society and European Respiratory Society [8]. Lung function outcomes were assessed at ages 10, 18, and 26 years. Participants were

required to be free from respiratory infection for 14 days and not administering any systemic oral steroids. It was also necessary that they did not take any short-acting beta-2 agonist for at least 6 hours, long-acting beta-2 agonist for 12 hours and caffeine for at least 4 hours prior to testing. All the measurements were made with the participants in the standing position without using a nose clip. The best of three consecutive expiratory maneuvers of forced vital capacity (FVC), forced expiratory volume 1 (FEV1), the FEV1/FVC ratio, and forced expiratory flow 25-75% (FEF25-75%) were taken as final values. At birth, weight was measured and maternal asthma and paternal history asthma were assessed using questionnaires. At each follow up height was measured and concurrent smoking was ascertained using self-reported questionnaire data. Eczema at age 4 years was defined as chronic or chronically relapsing, itchy dermatitis lasting more than 6 weeks with characteristic morphology and distribution [9] , according to the Hanifin and Rajka criteria [10]. Asthma at age 4 years was determined by a medical investigator based on wheeze over the last 12 months and treatment given for asthmatic conditions or related symptoms.

DNA methylation data in IOWBC

In the F_1 generation, DNA was isolated from dried blood spots on Guthrie cards of 724 neonates using a method based on the procedure described by Beyan *et al* [11]. In the F_2 generation, DNA was extracted from cord blood from a subsample of 193 subjects. DNA concentration was determined by Qubit quantitation. One microgram of DNA was bisulfite-treated for cytosine to thymine conversion using the EZ 96-DNA methylation kit (Zymo Research, CA, USA), following the manufacturer's standard protocol.

In the F1 generation, epigenome-scale DNA methylation (DNAm) was assessed using the Illumina Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates > 850,000 CpGs associated with over 24,000 genes. The estimateCellCounts() function from Minfi package [12] was used, using the reference panel from Houseman et al.

2016, to generate cell type proportions. In the F2 generation, epigenome-scale DNA methylation was assessed using the Illumina Infinium HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates >484,000 CpG sites, and the Illumina Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates > 850,000 CpGs associated with over 24,000 genes. Measurement error is unlikely since the Illumina Infinium HumanMethylation450 and 850 beadchip array are known to have high reliability and reproducibility [13, 14]. Cell types were estimated using the Bakulski reference panel in the *minfi* package [15].

Preprocessing and quality control of the DNA methylation data in IOW birth cohort

The CPACOR [16] pipeline was used for quality control (QC) and pre-processing the quantile normalized beta values from the samples. ComBat [17] was applied to remove batch effects. CpG sites with probe-SNPs within ten base pairs and with minor allele frequency (MAF) greater than 0.007 (which represented about 10 subjects in expectation in the complete study cohort) were excluded. This resulted finally in 551,710 CpGs from 796 participants. White blood cell counts were generated from using the estimateCellCounts() function from Minfi package [12], with the reference panel from Houseman et al. 2016 [18]. Sites on sex chromosomes were excluded due to sex-specific differences of X- and Y-chromosomes.

Methylation levels of each CpG were recorded as beta (β) values that range between zero and one. Beta value is the proportion of methylated (M) over methylated (M) plus unmethylated (U) probes (β=M/[c+M+U], with c being a constant to prevent diving by zero. For analyses purposes, M-values i.e., logit-transformed β values (M-values, approximated by log2(β /(1-β)) are preferred because β values close to 0 or 1 tend to suffer from severe heteroscedasticity [19].

Gene expression in IOWBC-F2

To measure the gene expression in IOWBC-F2 RNA was isolated extracted from 161 cord blood samples collected into PAXgene Bone Marrow RNA kits. Quality of RNA samples was checked with the Agilent 2100 BioAnalyzer system. Gene expression was assessed using Agilent SurePrint G3 Human Gene Expression 8x60k v2 microarray kits, using one color (Cy3) analysis with spike in controls. Array content was then sourced from RefSeq, Ensembl, UniGene, and GenBank databases that provided full coverage of the human transcriptome in 50,599 biological features including replicate probes and control probes. The oligos are 60 nucleotides in length, and each transcript is tagged at least once and may have multiple tagging oligos for genes with documented splice variants. Data were analyzed for QC indices with Agilent GeneSpring software.

Study population of Replication cohort: ALSPAC

The Avon Longitudinal Study of Parents and Children (ALSPAC) recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery between April 1, 1991 and December 31, 1992 [20, 21]. These initial pregnancies contributed to 14,676 newborns, resulting in 14,062 live births; 13,988 children who were alive at age 1 year. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes), data became available for more than the 14,541 pregnancies mentioned above.

The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 18 is 706 (452 and 254 recruited during Phases II and III respectively), resulting in an additional 713 children being enrolled. The phases of enrolment are described in more detail in the cohort

profile paper [22]. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees."

Phenotypes

Lung function outcomes were assessed by spirometry at ages 8.5, 15, and 24 years using Vitalograph Spirotrac (Maids Moreton, United Kingdom) software. All measurements were carried out by trained fieldworkers in a dedicated research clinic following the guidelines of the American Thoracic Society and European Respiratory Society and all flow volume curves were read post hoc by a respiratory physician for quality assurance. Measures were made at 8.5 years without bronchodilator and at 15 and 24 years before and 10-15 minutes after 400 mcg inhaled albuterol. All measurements were made in the seated position with a nose clip. The best of three technically acceptable and repeatable measurements of forced vital capacity (FVC), forced expiratory volume 1 (FEV1), the FEV1/FVC ratio, and forced expiratory flow 25-75% (FEF25-75%) were used in the analysis. Height was measured in clinic at the same ages as lung function.

DNA methylation data

DNAm data for a subset of approximately 1000 mother-child pairs is available under ARIES, the Accessible Resource for Integrated Epigenomics Studies. DNAm was assayed using the Illumina Infinium HumanMethylation 450k BeadChip platform [23]. Cord blood samples (whole blood or buffy coats) were collected according to standard procedures, spun and frozen at -80˚C. DNAm analyses and data pre-processing were performed at the University of Bristol as part of the ARIES project [24] (ariesepigenomics.org.uk). Cord blood samples (n=746) were supplemented with samples from blood spots (N=168). Following extraction, DNA was bisulfite converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA). Following conversion, the genome-wide methylation status of over 485,000 CpG sites was measured using the Illumina Infinium® HumanMethylation450k BeadChip assay according to the standard protocol. The arrays were scanned using an Illumina iScan and initial quality review was assessed using

GenomeStudio (version 2011.1). The level of methylation is expressed as a "Beta" value (β-value), ranging from 0 (no cytosine methylation) to 1 (complete cytosine methylation). Cell types were estimated using the Houseman approach with the cord blood reference data set [15].

Preprocessing and quality control of the DNA methylation data in ALSPAC

Samples from all time-points in ARIES were distributed across slides using a semi-random approach (sampling criteria were in place to ensure that all time-points were represented on each array) to minimize the possibility of confounding by batch effects. Samples failing quality control (average probe detection p-value ≥ 0.01) were repeated. As an additional quality control step genotype probes on the HumanMethylation450k were compared between samples from the same individual and against SNP-chip data to identify and remove any sample mismatches.

Microarray quality control and normalization was completed using meffil [25] in R version 3.2.0. This included checking for sample genotype mismatches, sex mismatches, probe signal outliers, dye bias, and low probe detection rates. The profiles passing quality control were normalized using functional normalization as implemented in the meffil R package. Probe intensity quantiles were adjusted using the top 10 control probe principal components as fixed effects and sample slide as a random effect.

Sites on sex chromosomes were excluded to reduce complexity due to sex-specific differences and X-chromosome inactivation by DNA methylation in females. Finally, probes showing a detection P-value >0.05 for >5% samples were excluded. The sample was restricted to singletons only. For all analyses, beta-values were converted to M-values [19] using logit transformation as described above. Before transformation to M-values the range of beta-values was shrunk from 0 - 1 to 0.005 - 0.995 using the following formula in order to avoid errors caused by logit transformation of methylation values of zero:

$$\beta_{reduced} = ((\beta_{original} - 0.5) * 0.99) + 0.5$$

Supplementary Results

Comparison of association of DNAM with lung function trajectories with and without confounders in IOWBC-F1

To assess the effect of confounders in the association of DNAm and lung function trajectories in IOWBC-F1, we used backward selection process to iteratively eliminate confounders that do not change the risk ratios by >10%, thus considering the cumulative effect of confounding. We adjusted for maternal and paternal asthma, socioeconomic status, maternal smoking, and cell types. Since none of them changed the risk ratios by more than 10%, we excluded them from the model. We retained cell types, as are they directly influence the DNAm measures. The risk ratios with and without confounders for all the CpGs that were significantly replicated in ALSPAC (68 CpGs in boys and 58 CpGs in girls) are compared in Supplementary Table S1.

Identifying CpGs with concordant and discordant directions of associations between IOWBC-F1 and ALSPAC

We identified 68 CpGs in boys and 58 CpGs in girls to be significantly associated with lung function in both IOWBC-F1 and ALSPAC. We compared the direction of associations between the cohorts. In IOWBC-F1 we used trajectories of lung function as outcome and determined the relative risk of being in the low lung function trajectory. In contrast, in ALSPAC, we used individual lung function measures and determined the regression coefficient. Consequently, to compare the direction of association we used the following rule, (a) direction of association between the cohorts is concordant with risk ratio <1 in IOWBC-F1 and positive regression coefficient in ALSPAC or, risk ratio >1 in IOWBC-F1 and negative regression coefficient in ALSPAC indicates; (b) direction of association between the cohorts is discordant with Risk Ratios <1 in IOWBC-F1 and negative regression coefficient in ALSPAC or, risk ratio >1 in IOWBC-F1 and positive regression coefficient in ALSPAC. Following this method, we identified 31 of these 68 CpGs (~46%) in boys, and 33 of these 58 CpGs (~ 57%) in girls to have the

concordant direction of association in both the cohorts. The risk ratios and estimates in IOWBC-F1 and ALSPAC, respectively, for all 68 CpGs are compared in the Supplementary Table S2 (a) for boys and Supplementary Table S2(b) girls with directions of effect denoted in a separate column.

Stability of CpGs over time in IOWBC-F1

Stability of DNAm is one such criteria that highlights the long-term epigenetic modifications of *in utero* or genetic effects on lung function substantiating that later lung health is determined during fetal growth. To this end, we assessed the stability of the CpGs in IOWBC-F1 that were significantly replicated in the same direction in the two cohorts (33 in boys and 31 in girls). DNAm was available at birth (n=396 in boys; n= 390 in girls), age 10 (n=128 in boys; n= 93 in girls), and age 18 years (n= 153 in boys; n= 161 in girls). We performed a repeated measurement analysis to assess whether DNAm changes over time between the lung function trajectories. Since change in cell types over time can affect the DNAm, we regressed the cell types and calculated the residuals DNAm at each time point for the analyses. In the statistical models, DNAm (cell type residuals) was the outcome and interactions were assessed between age and lung function trajectories. CpGs with non-significant interactions with p-values >=0.1 were considered stable indicting that DNAm does not change between the trajectories over the ages. Ten CpGs in boys and fifteen in girls were identified to be stable (Supplementary Table S3). The patterns of mean DNAm these CpGs over time with respect to the lung function trajectories is shown in Supplementary Figure S1. Most CpGs show clear separation of DNAm over time between the trajectories, i.e., their DNAm does not significantly change over time between high and low trajectories. For some CpGs, the separation of DNAm at age ten years is not clear between the trajectories, even though the interaction p-value is less than 0.1, most likely due to lowest sample size with DNAm at age 10 years.

Biological functionality of the genes

We used the functional annotation tool ToppFun belonging to the ToppGene suite to identify important biological processes. Thirty-one gene names corresponding to the twenty-five CpGs showing no difference over time between the high and low trajectories in IOWBC-F1 were queried in ToppFun. We found eight genes to be significantly enriched for GO terms related to developmental processes, namely, tube morphogenesis, embryonic organ development, digestive tract development, paramesonephric duct development, mesonephric duct development, somite development, female genitalia development, and notochord development. All eight genes were enriched in the category of tube development, while different subsets of these genes were enriched in the remaining categories. The remaining categories of the enriched GO terms also belong under the overarching process of tube morphogenesis (Supplementary Table S4). We demonstrated how these biological processes and genes are relevant in development and morphogenesis of the lung in Supplementary Table S4.

Expression of the biologically relevant genes in human lung tissues

In IOWBC-F2, we assessed the correlation of gene expression and DNAm in cord blood for two CpGs, namely, cg01885814 (*COL18A1*) and cg07103093 (*LHX1*). While this provides us an insight regarding the role of DNAm in altering the gene expression in cord blood, which can reveal effect of maternal or prenatal exposures, we still lacked information about gene expression in lung tissues. To this end, we searched in an online database of the LungMAp consortium https://lungmap.net, which is a repository of high-quality molecular data related to lung development in humans and mice., including transcriptomic data. Briefly, this database included cell specific RNA sequencing data from 151 BRINDL repository samples from 26 human donor lungs. RNA was extracted from endothelial, epithelial, mesenchymal, and immune cells from the right upper and middle lung lobes and gene expression was assessed using standard procedures [26]. All eight biologically relevant gene identified in IOWBC-F1 were found

to be expressed in these samples that ranged in age from full term newborns to adults (Supplementary Figure S2). Even though in a limited sample size, given that human lung tissue hard to obtain, this data supports that the eight important genes identified the current EWAS are expressed in fetal and adult lung tissue.

Comparison of directionality of effects between IOWBC-F1 and ALSPAC

We found around 54% CpGs in boys and 43% CpGs in girls to have discordant direction of association of DNAm and lung function between ALSPAC and IOWBC-F1. Such a disagreements has been identified in multiple other metaanalyses [27, 28]. First, we tested a recent suggestion that the direction of associations between predictors and outcome can be reversed by removing cell types from the model since multicollinearity between cell types and CpGs in the statistical model [29] may impact the direction of effects. However, in IOWBC-F1, removing cell types from the model did not affect the directionality. Second, we then explored whether direction of association of DNAm with lung function varies systematically between different subgroups of the population exposed to specific risk factors. To this end, we stratified the association in IOWBC-F1 by separately by several risk factors such as paternal history of asthma, maternal history of asthma, low birth weight, asthma at age 4, and eczema at age 4 years. We then compared the direction of association of DNAm with lung function in ALSPAC with that in the two strata in IOWBC-F1; one where the risk factor is present, and the other where the risk factor is absent. We found that in strata where risk factor is present the direction of association agreed between ALSPAC and IOWBC-F1. For instance, in boys of ALSPAC, higher DNAm of cg01699600 (*FAM38A*) was associated with lower FVC levels. In contrast, in the unstratified analyses in IOWBC-F1 boys (n=396), higher the DNAm of cg01699600 was associated with higher FVC (indicated by a lower risk ratio of being in low FVC trajectory in Supplementary Table S5a under “original risk ratio in IOWBC-F1”). Upon stratification in

IOWBC-F1 by paternal asthma, we observed that in boys with positive history of paternal asthma (n=41), higher DNAm of CpG cg01699600 was associated with lower FVC, agreeing to that of ALSPAC ("Paternal asthma "Yes" in Supplementary Table S5a). The associations were not statistically significant possibly due to small sample sizes in specific strata. Similarly, upon stratification by paternal history of asthma, and maternal history of asthma, low birth weight, asthma and eczema at 4 years we observed agreement in the direction of association between ALSPAC and IOWBC-F1 in the strata where risk factors were present. We were able to reduce the disagreement from ~54% (37 of the 68 CpGs) to ~28% (19 out of 68 CpGs) in boys, and ~43% (25 of 58 CpGs) to ~17% (10 of 58 CpGs) in girls (Figure 3, and Supplementary Table S5a and S5b). These findings indicate that different distribution of these risk factors may play a role in determining the association between CpGs and lung function. A third explanation of directional discordance could be the influence of genetic polymorphisms at or neighboring locus as the directionally discordant CpGs i.e., methQTLs (methylation quantitative trait loci). Since several CpGs are influenced by genetic variants [30], there is a possibility that differences in proportion of genotypes of these genetic variants in ALSAPC compared to IOWBC-F1 could result discordance in direction for associations between DNA-methylation and lung function between the cohorts.

Supplementary Table S1: Comparison of risk ratios determined by with and without adjusting for confounders in boys and girls in IOWBC F1 generation.

				Without confounders*				With confounders#			
Sex	Outcome	Ilmnid	Gene name	Risk ratio	95% LCI	95% UCI	p-value	Risk ratio	95% LCI	95% UCI	p-value
Boys	FEF25-75%	cg01678802	COBL	0.92	0.87	0.97	0.003	0.91	0.86	0.97	0.002
Boys	FEF25-75%	cg01885814	COL18A1	0.93	0.89	0.98	0.003	0.93	0.89	0.98	0.005
Boys	FEF25-75%	cg13529291	WNT11; UVRAG	1.14	1.03	1.25	0.01	1.15	1.04	1.27	0.007
Boys	FEF25-75%	cg22902534	AP2M1	0.83	0.74	0.94	0.003	0.83	0.73	0.93	0.002
Boys	FEF25-75%	cg20246851	SEMA4B	0.85	0.76	0.95	0.003	0.84	0.76	0.94	0.003
Boys	FEF25-75%	cg16465128	SEPT1	0.88	0.81	0.95	0.001	0.88	0.81	0.95	0.0009
Boys	FEF25-75%	cg27582059	PCDHB16;PCDHB8	0.90	0.82	0.98	0.01	0.90	0.82	0.98	0.013
Boys	FEF25-75%	cg14281210	MARCH8	0.92	0.88	0.97	0.002	0.92	0.88	0.97	0.002
Boys	FEF25-75%	cg06549530	CHST15; CPXM2	0.93	0.89	0.98	0.002	0.93	0.89	0.97	0.002
Boys	FEF25-75%	cg26606286	HAND2;HAND2-AS1	1.11	1.03	1.18	0.003	1.11	1.04	1.19	0.003
Boys	FEF25-75%	cg17366410	AZI1	1.15	1.05	1.25	0.002	1.15	1.06	1.26	0.001
Boys	FEF25-75%	cg07481273	C14orf109	0.87	0.80	0.96	0.004	0.87	0.79	0.95	0.003
Boys	FEF25-75%	cg05699739	ESD; HTR2A	0.88	0.80	0.97	0.01	0.88	0.80	0.97	0.007
Boys	FEF25-75%	cg25025968	RRM1	0.89	0.80	0.98	0.01	0.87	0.79	0.97	0.009
Boys	FEF25-75%	cg26829088	C19orf29	0.89	0.83	0.95	0.001	0.88	0.82	0.95	0.0007
Boys	FEF25-75%	cg01443538	MAP3K10	0.89	0.82	0.96	0.003	0.88	0.81	0.96	0.002
Boys	FEF25-75%	cg22469836	ABCF3	0.89	0.83	0.97	0.005	0.89	0.82	0.97	0.005
Boys	FEF25-75%	cg00282510	FOXE1; TRMO	0.90	0.84	0.96	0.0009	0.90	0.84	0.96	0.001
Boys	FEF25-75%	cg00566297	RGS7; FH	0.90	0.84	0.97	0.004	0.89	0.83	0.96	0.003
Boys	FEF25-75%	cg10122932	MCM7	0.90	0.84	0.96	0.002	0.90	0.84	0.96	0.001
Boys	FEF25-75%	cg03724260	ZBTB4;POLR2A	0.91	0.85	0.96	0.001	0.90	0.85	0.96	0.0009
Boys	FEF25-75%	cg06627361	KDM5A	0.91	0.85	0.96	0.002	0.90	0.85	0.96	0.002
Boys	FEF25-75%	cg09537551	OSR2; VPS13B	0.91	0.85	0.96	0.001	0.90	0.85	0.96	0.0008
Boys	FEF25-75%	cg03962214	CASZ1	0.92	0.86	0.97	0.003	0.91	0.86	0.97	0.002
Boys	FEF25-75%	cg01934962	AKAP10	0.92	0.87	0.97	0.001	0.92	0.87	0.97	0.001
Boys	FEF25-75%	cg19471553	TMEM106A; NBR1	0.92	0.87	0.97	0.004	0.92	0.87	0.97	0.004
Boys	FEF25-75%	cg16848490	PHLDB2	0.93	0.88	0.97	0.003	0.93	0.88	0.97	0.002

Supplementary Table S1 continued: Comparison of risk ratios determined by with and without adjusting for confounders in boys and girls in IOWBC F1 generation.

Sex	Outcome	Ilmnid	Gene name	Without confounders*				With confounders#			
				Risk ratio	95% LCI	95% UCI	p-value	Risk ratio	95% LCI	95% UCI	p-value
Boys	FEF25-75%	cg24529269	AEN	1.07	1.02	1.11	0.003	1.07	1.02	1.11	0.002
Boys	FEF25-75%	cg02482718	AJAP1	1.07	1.01	1.13	0.013	1.07	1.01	1.13	0.014
Boys	FEF25-75%	cg08959305	TMEM97	1.08	1.02	1.14	0.009	1.08	1.02	1.14	0.01
Boys	FEF25-75%	cg26784348	SLC35D1	1.11	1.03	1.20	0.008	1.12	1.03	1.21	0.006
Boys	FEF25-75%	cg09738214	ATP6V0E1	1.18	1.06	1.31	0.0021	1.18	1.06	1.32	0.002
Boys	FEV1	cg06710742	STRA6; CCDC33	0.77	0.66	0.89	0.0005	0.78	0.67	0.91	0.001
Boys	FEV1	cg17330460	PUS3	0.85	0.79	0.92	4E-05	0.85	0.79	0.92	7E-05
Boys	FEV1	cg07103093	LHX1	0.87	0.81	0.94	0.0004	0.86	0.80	0.94	0.0003
Boys	FEV1	cg21437345	B4GALNT2	1.19	1.09	1.30	0.0001	1.16	1.06	1.28	0.002
Boys	FEV1	cg16149007	KLHDC10; MEM209	1.21	1.10	1.33	6E-05	1.19	1.09	1.32	0.0003
Boys	FEV1	cg21505334	CEACAM5	1.25	1.14	1.38	4E-06	1.25	1.14	1.38	6E-06
Boys	FEV1	cg08207604	DEF6	0.80	0.72	0.89	7E-05	0.80	0.72	0.90	0.0002
Boys	FEV1	cg05831188	SDK1	0.80	0.72	0.89	5E-05	0.81	0.73	0.90	0.0002
Boys	FEV1	cg14836450	RAPGEF4;LOC91149	0.84	0.77	0.91	6E-05	0.84	0.77	0.91	6E-05
Boys	FEV1	cg09771049	KPNA2	0.87	0.81	0.93	6E-05	0.87	0.81	0.93	4E-05
Boys	FEV1	cg05899183	FAM96B; CES2	0.90	0.85	0.95	0.0001	0.90	0.85	0.95	0.0002
Boys	FEV1	cg21899743	MYOM2	1.14	1.06	1.22	0.0002	1.14	1.06	1.22	0.0005
Boys	FEV1	cg00647165	WIZ	1.18	1.09	1.28	7E-05	1.17	1.08	1.27	0.0003
Boys	FEV1	cg11659361	FBXO2; FBXO44	1.24	1.12	1.36	2E-05	1.24	1.12	1.37	3E-05
Boys	FEV1	cg14706297	ZFP36L2	1.12	1.06	1.18	3E-05	1.13	1.07	1.20	2E-05
Boys	FEV1	cg07460095	UBA5; ACAD11	1.26	1.12	1.41	8E-05	1.25	1.11	1.40	0.0002
Boys	FEV1	cg10236596	CENPF; PTPN14	0.79	0.72	0.88	7E-06	0.79	0.71	0.87	3E-06
Boys	FEV1	cg00280220	SMARCD2; TCAM1P	0.83	0.76	0.91	0.0001	0.83	0.75	0.91	0.0002
Boys	FEV1	cg14140717	NELF	0.88	0.83	0.93	9E-06	0.88	0.83	0.93	2E-05
Boys	FEV1	cg03110267	C18orf25	0.91	0.88	0.95	2E-05	0.91	0.88	0.95	3E-05
Boys	FEV1	cg21200949	SEC23B	1.23	1.12	1.35	2E-05	1.22	1.11	1.34	5E-05

Supplementary Table S1 continued: Comparison of risk ratios determined by with and without adjusting for confounders in boys and girls in IOWBC F1 generation.

Sex	Outcome	Ilmnid	Gene name	Without confounders*				With confounders#			
				Risk ratio	95% LCI	95% UCI	p-value	Risk ratio	95% LCI	95% UCI	p-value
Boys	FEV1byFVC	cg16750801	PTPRE	0.84	0.79	0.90	2E-07	0.82	0.77	0.88	7E-09
Boys	FEV1byFVC	cg16016281	TRIO; DNAH5	0.79	0.70	0.89	0.0002	0.77	0.69	0.88	5E-05
Boys	FEV1byFVC	cg00983520	CPT1B	0.86	0.79	0.93	0.0002	0.86	0.80	0.94	0.0003
Boys	FEV1byFVC	cg19836471	DLG5	0.85	0.79	0.93	0.0001	0.85	0.79	0.93	0.0001
Boys	FEV1byFVC	cg07071157	ATG16L2	1.19	1.11	1.28	1E-06	1.19	1.11	1.29	3E-06
Boys	FVC	cg13904267	DFNB31	0.88	0.82	0.94	0.0002	0.88	0.82	0.94	0.0002
Boys	FVC	cg06639763	cg01678802	1.22	1.10	1.35	0.0001	1.23	1.11	1.36	8E-05
Boys	FVC	cg03224209	ESRRG	0.81	0.73	0.89	3E-05	0.81	0.73	0.90	9E-05
Boys	FVC	cg09771049	KPNA2	0.87	0.81	0.94	0.0001	0.87	0.81	0.93	0.0001
Boys	FVC	cg03918756	TRAPPC9	1.20	1.10	1.31	5E-05	1.19	1.09	1.31	0.0001
Boys	FVC	cg27189973	FAM38A	0.83	0.76	0.90	9E-06	0.83	0.76	0.91	4E-05
Boys	FVC	cg01699600	FAM38A	0.86	0.81	0.92	3E-06	0.86	0.81	0.92	2E-05
Boys	FVC	cg25611736	KIF26A	1.34	1.16	1.55	7E-05	1.36	1.16	1.58	9E-05
Boys	FVC	cg19616339	GKN2	0.78	0.70	0.88	5E-05	0.78	0.70	0.89	8E-05
Boys	FVC	cg14266217	STARD3NL	1.15	1.08	1.24	5E-05	1.15	1.07	1.23	0.0002
Girls	FEF25-75%	cg07380056	GALNT13	0.67	0.52	0.85	0.0014	0.65	0.50	0.85	0.001
Girls	FEF25-75%	cg00383081	GLUL	0.77	0.67	0.90	0.0008	0.77	0.66	0.89	0.0004
Girls	FEF25-75%	cg14846324	BRUNOL6	0.79	0.68	0.92	0.003	0.80	0.69	0.94	0.005
Girls	FEF25-75%	cg15485560	KIZ	0.80	0.69	0.94	0.005	0.78	0.67	0.91	0.001
Girls	FEF25-75%	cg05226043	TRIM26	0.80	0.70	0.92	0.001	0.79	0.69	0.91	0.0007
Girls	FEF25-75%	cg22311507	MYCN	0.91	0.86	0.97	0.002	0.91	0.86	0.97	0.002
Girls	FEF25-75%	cg12635120	SPPL3; HNF1A-AS1	1.31	1.08	1.58	0.007	1.29	1.06	1.57	0.012
Girls	FEF25-75%	cg20697424	NKIRAS2; DNAJC7	1.42	1.19	1.68	7E-05	1.37	1.16	1.63	0.0003
Girls	FEF25-75%	cg08591299	ANKRD11; ZNF778	0.76	0.66	0.87	1E-04	0.78	0.68	0.89	0.0003
Girls	FEF25-75%	cg23937993	AMPD3	0.76	0.64	0.89	0.001	0.74	0.63	0.88	0.0005
Girls	FEF25-75%	cg12629349	EML6	0.76	0.64	0.90	0.002	0.75	0.63	0.90	0.001

Supplementary table S1 continued: Comparison of risk ratios determined by with and without adjusting for confounders in boys and girls in IOWBC F1 generation.

Sex	Outcome	Ilmnid	Gene name	Without confounders*				With confounders#			
				Risk ratio	95% LCI	95% UCI	p-value	Risk ratio	95% LCI	95% UCI	p-value
Girls	FEF25-75%	cg10368052	VPS52	0.76	0.65	0.90	0.001	0.79	0.67	0.93	0.005
Girls	FEF25-75%	cg24656492	SUPV3L1	0.77	0.66	0.90	0.001	0.77	0.66	0.91	0.001
Girls	FEF25-75%	cg23817336	TCP11; SCUBE3	0.79	0.69	0.91	0.001	0.79	0.69	0.92	0.001
Girls	FEF25-75%	cg22514863	PAFAH1B3	0.80	0.70	0.92	0.001	0.81	0.70	0.92	0.002
Girls	FEF25-75%	cg10179911	RPRM	0.84	0.76	0.94	0.002	0.85	0.76	0.95	0.006
Girls	FEF25-75%	cg22176566	SLC2A1	1.13	1.05	1.22	0.001	1.12	1.04	1.21	0.002
Girls	FEF25-75%	cg03341334	DUSP23	1.14	1.06	1.23	0.0005	1.15	1.07	1.24	0.0002
Girls	FEF25-75%	cg23395902	PCGF5	1.19	1.09	1.31	0.0001	1.17	1.07	1.28	0.0005
Girls	FEF25-75%	cg09845296	TRIO; DNAH5	1.20	1.10	1.31	7E-05	1.18	1.08	1.30	0.0002
Girls	FEF25-75%	cg23182674	FAM71B	1.22	1.09	1.37	0.0007	1.25	1.11	1.41	0.0002
Girls	FEF25-75%	cg06338552	ZC3H7B	1.22	1.06	1.41	0.006	1.26	1.08	1.47	0.002
Girls	FEF25-75%	cg21092296	LOC90110	1.24	1.00	1.54	0.05	1.24	0.99	1.55	0.06
Girls	FEF25-75%	cg25967904	TRMU	1.26	1.08	1.46	0.003	1.26	1.08	1.46	0.002
Girls	FEF25-75%	cg13392687	AMOTL1	1.26	1.10	1.44	0.0008	1.24	1.08	1.42	0.002
Girls	FEF25-75%	cg18435928	JAKMIP3	1.31	1.12	1.52	0.0005	1.31	1.13	1.53	0.0005
Girls	FEF25-75%	cg20934191	DKFZ	1.31	1.12	1.53	0.0006	1.31	1.13	1.53	0.0005
Girls	FEF25-75%	cg10887937	ALDH2	1.37	1.17	1.60	8E-05	1.39	1.19	1.63	5E-05
Girls	FEF25-75%	cg19244662	SLC14A1	1.43	1.16	1.76	0.0009	1.46	1.18	1.81	0.0004
Girls	FEF25-75%	cg03795245	GPR162; CD4	1.46	1.19	1.77	0.0002	1.44	1.18	1.77	0.0003
Girls	FEF25-75%	cg15738933	SEC14L1	1.51	1.22	1.88	0.0002	1.47	1.17	1.84	0.0009
Girls	FEV1	cg20825472	UST	0.82	0.75	0.89	3E-06	0.83	0.76	0.90	2E-05
Girls	FEV1	cg14059822	IRX2	0.82	0.74	0.91	0.0001	0.83	0.75	0.91	0.0002
Girls	FEV1	cg12967384	FBRSL1	0.84	0.76	0.91	6E-05	0.83	0.76	0.91	5E-05
Girls	FEV1	cg02637537	PDLIM2	0.84	0.78	0.92	8E-05	0.85	0.78	0.92	0.0001
Girls	FEV1	cg00175344	POU4F1; RNF219	0.87	0.82	0.92	6E-06	0.86	0.81	0.92	2E-06
Girls	FEV1	cg22843625	C19orf63; FAM71E1	1.12	1.06	1.17	7E-06	1.12	1.07	1.18	1E-06

Supplementary table S1 continued: Comparison of risk ratios determined by with and without adjusting for confounders in boys and girls in IOWBC F1 generation.

Sex	Outcome	Ilmnid	Gene name	Without confounders*				With confounders#			
				Risk ratio	95% LCI	95% UCI	p-value	Risk ratio	95% LCI	95% UCI	p-value
Girls	FEV1	cg16472998	UNK	1.13	1.07	1.20	4E-05	1.13	1.06	1.20	0.0001
Girls	FEV1	cg22665383	OR4C16	1.22	1.12	1.34	1E-05	1.21	1.10	1.32	4E-05
Girls	FEV1byFVC	cg03433758	CRTC3	0.85	0.79	0.92	2E-05	0.85	0.79	0.92	2E-05
Girls	FEV1byFVC	cg07052251	COL6A1; PCBP3	0.87	0.81	0.93	3E-05	0.86	0.81	0.92	1E-05
Girls	FEV1byFVC	cg16008148	HLX; MARC1	0.89	0.85	0.94	4E-05	0.89	0.84	0.94	2E-05
Girls	FEV1byFVC	cg08879910	HLA-J; NCRNA00171	1.11	1.06	1.17	6E-05	1.11	1.05	1.17	0.0001
Girls	FEV1byFVC	cg11301337	PTPRN2	1.20	1.10	1.30	3E-05	1.20	1.10	1.30	3E-05
Girls	FEV1byFVC	cg00964751	C17orf96	0.77	0.68	0.87	6E-05	0.76	0.67	0.86	3E-05
Girls	FEV1byFVC	cg13687497	RXRA	0.77	0.69	0.87	3E-05	0.76	0.68	0.86	9E-06
Girls	FEV1byFVC	cg06594008	MKL1	0.81	0.74	0.89	9E-06	0.82	0.75	0.90	2E-05
Girls	FEV1byFVC	cg23221090	NRBF2	0.83	0.76	0.91	4E-05	0.84	0.76	0.92	0.0001
Girls	FEV1byFVC	cg17636541	NRM	0.83	0.76	0.90	2E-05	0.83	0.76	0.90	7E-06
Girls	FEV1byFVC	cg11294750	GPC6	0.85	0.79	0.91	3E-06	0.84	0.79	0.90	1E-06
Girls	FEV1byFVC	cg03869608	PDZRN3	0.86	0.79	0.93	0.0002	0.86	0.80	0.94	0.0005
Girls	FEV1byFVC	cg02308192	NEK9	0.88	0.83	0.93	2E-06	0.87	0.82	0.92	1E-06
Girls	FEV1byFVC	cg11123440	GATA4; NEIL2	0.93	0.90	0.96	8E-06	0.93	0.90	0.96	9E-06
Girls	FEV1byFVC	cg12754671	NDUFS2	1.18	1.10	1.27	5E-06	1.20	1.11	1.28	1E-06
Girls	FVC	cg11418007	GRINL1A; GCOM1	1.17	1.08	1.28	0.0002	1.17	1.07	1.27	0.0003
Girls	FVC	cg00015603	MYEOV2	0.82	0.75	0.90	4E-05	0.82	0.75	0.90	2E-05
Girls	FVC	cg00175344	POU4F1; RNF219	0.88	0.82	0.94	6E-05	0.88	0.83	0.94	0.0001
Girls	FVC	cg25080348	ASPH	1.34	1.16	1.55	9E-05	1.31	1.14	1.52	0.0002

* Modelling the association if DNAm with lung function trajectories adjusting only for cell types

Modelling the association if DNAm with lung function trajectories adjusting for cell types, maternal, paternal asthma, socioeconomic status, and maternal smoking

Supplementary Table S2 (a): CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			Consistency of direction of association between IOWBC-F1 and ALSPAC
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	
Outcome	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FVC	cg13904267	*DFNB31*	0.88 (0.82, 0.94)			0.49 (4.76×10^{-2})	Yes
	cg03224209	*ESRRG*	0.81 (0.73, 0.89)		0.23 (2.52×10^{-2})	0.51 (9.89×10^{-5})	Yes
	cg01699600	*FAM38A*	0.86 (0.81, 0.92)	-0.04 (3.09×10^{-2})			No
	cg27189973	*FAM38A*	0.83 (0.76, 0.90)	-0.06 (2.84×10^{-2})			No
	cg25611736	*KIF26A*	1.34 (1.16, 1.55)	0.08 (3.10×10^{-2})			No
	cg09771049	*KPNA2*	0.87 (0.81, 0.94)		0.23 (1.88×10^{-03})		Yes
	cg14266217	*STARD3NL*	1.15 (1.08, 1.24)			0.39 (3.83×10^{-2})	No
	cg03918756	*TRAPPC9*	1.20 (1.10, 1.31)			-0.31 (1.23×10^{-2})	Yes
	cg06639763	*FBXO31*	1.22 (1.10, 1.35)		-0.21 (4.40×10^{-2})		Yes
	cg19616339	*GKN2; GKN1*	0.78 (0.70, 0.88)	-0.06 (4.08×10^{-2})			No
FEV1/ FVC	cg07071157	*ATG16L2*	1.19 (1.11, 1.28)		0.04 (8.00×10^{-3})		No
	cg00983520	*CPT1B*	0.86 (0.79, 0.93)		-0.01 (4.78×10^{-2})		No

Supplementary Table S2 (a) continued: CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			Consistency of direction of association between IOWBC-F1 and ALSPAC
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
Outcome				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	
	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEV1/ FVC	cg19836471	*DLG5*	0.85 (0.79, 0.93)	-0.01 (3.06×10^{-2})			No
	cg16750801	*PTPRE*	0.84 (0.79, 0.90)		0.01 (4.32×10^{-2})		Yes
	cg16016281	*TRIO; DNAH5*	0.79 (0.70, 0.89)	-0.01 (4.34×10^{-2})			No
FEV1	cg21437345	*B4GALNT2*	1.19 (1.09, 1.30)		-0.14 (3.20×10^{-2})		Yes
	cg03110267	*C18orf25*	0.91 (0.88, 0.95)			-0.75 (3.88×10^{-2})	No
	cg21505334	*CEACAM5*	1.25 (1.14, 1.38)			-0.24 (4.13×10^{-2})	Yes
	cg08207604	*DEF6*	0.80 (0.72, 0.89)		0.18 (3.73×10^{-2})	0.24 (1.79×10^{-2})	Yes
	cg05899183	*FAM96B; CES2*	0.90 (0.85, 0.95)		0.30 (4.72×10^{-2})		Yes
	cg09771049	*KPNA2*	0.87 (0.81, 0.93)		0.17 (9.75×10^{-3})		Yes
	cg07103093	*LHX1*	0.87 (0.81, 0.94)			0.22 (2.69×10^{-2})	Yes
	cg21899743	*MYOM2*	1.14 (1.06, 1.22)		-0.17 (1.76×10^{-2})		Yes

Supplementary Table S2 (a) continued: CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			Consistency of direction of association between IOWBC-F1 and ALSPAC
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	
Outcome	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEV1	cg14140717	*NELF*	0.88 (0.83, 0.93)	-0.03 (3.28×10^{-2})			No
	cg17330460	*PUS3*	0.85 (0.79, 0.92)			0.29 (3.14×10^{-2})	Yes
	cg14836450	*RAPGEF4; LOC91149*	0.84 (0.77, 0.91)		0.19 (4.59×10^{-3})		Yes
	cg05831188	*SDK1*	0.80 (0.72, 0.89)			0.18 (3.17×10^{-2})	Yes
	cg07460095	*UBA5; ACAD11*	1.26 (1.12, 1.41)	0.09 (2.83×10^{-2})			No
	cg00647165	*WIZ*	1.18 (1.09, 1.28)	-0.04 (1.84×10^{-2})			Yes
	cg00280220	*SMARCD2; TCAM1P*	0.83 (0.76, 0.91)		-0.19 (9.68×10^{-3})		No
	cg06710742	*STRA6; CCDC33*	0.77 (0.66, 0.89)		0.29 (9.19×10^{-3})		Yes
	cg10236596	*CENPF; PTPN14*	0.79 (0.72, 0.88)		-0.16 (4.76×10^{-2})		
	cg11659361	*FBXO2; FBXO44*	1.24 (1.12, 1.36)			-0.25 (3.48×10^{-2})	Yes
	cg14706297	*ZFP36L2*	1.12 (1.06, 1.18)			0.14 (2.97×10^{-2})	No

Supplementary Table S2(a) continued: CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEV1	cg16149007	*KLHDC10; TMEM209*	1.21 (1.10, 1.33)	-0.03 (1.97×10^{-2})			Yes
	cg21200949	*SEC23B*	1.23 (1.12, 1.35)			0.21 (4.74×10^{-02})	No
FEF25-75%	cg16465128	*SEPT1*	0.88 (0.81,0.95)		0.30 (2.67×10^{-2})		Yes
	cg22469836	*ABCF3*	0.89 (0.83,0.97)	-0.24 (5.02×10^{-3})			No
	cg24529269	*AEN*	1.07 (1.02,1.11)		1.80 (1.42×10^{-3})		No
	cg02482718	*AJAP1*	1.07 (1.01,1.13)			0.52 (4.08×10^{-2})	No
	cg01934962	*AKAP10*	0.92 (0.87,0.97)	-0.34 (5.22×10^{-3})			No
	cg22902534	*AP2M1*	0.83 (0.74,0.94)		0.59 (4.46×10^{-2})		Yes
	cg09738214	*ATP6V0E1*	1.18 (1.06,1.31)			0.42 (4.33×10^{-2})	No
	cg17366410	*AZI1*	1.15 (1.05,1.25)		-0.27 (3.78×10^{-2})	-0.54 (1.71×10^{-3})	Yes
	cg07481273	*C14orf109*	0.87 (0.80,0.96)	-0.10 (1.19×10^{-2})			No
	cg26829088	*C19orf29*	0.89 (0.83,0.95)	-0.18 (1.27×10^{-2})			No

Supplementary Table S2(a) continued: CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEF25-75%	cg03962214	*CASZ1*	0.92 (0.86,0.97)	-0.14 (9.73×10^{-3})			No
	cg01885814	*COL18A1*	0.93 (0.89,0.98)	0.12 (1.24×10^{-2})			Yes
	cg06627361	*KDM5A*	0.91 (0.85,0.96)	-0.16 (3.41×10^{-2})			No
	cg01443538	*MAP3K10*	0.89 (0.82,0.96)			-0.77 (2.68×10^{-2})	No
	cg14281210	*8-Mar*	0.92 (0.88,0.97)		0.58 (3.31×10^{-2})		Yes
	cg10122932	*MCM7; AP4M*	0.90 (0.84,0.96)	-0.13 (1.31×10^{-2})			No
	cg27582059	*PCDHB16; PCDHB8*	0.90 (0.82,0.98)		0.21 (2.57×10^{-2})		Yes
	cg16848490	*PHLDB2*	0.93 (0.88,0.97)	-0.36 (2.06×10^{-2})			No
	cg25025968	*RRM1*	0.89 (0.80,0.98)	-0.17 (1.43×10^{-2})			No
	cg20246851	*SEMA4B*	0.85 (0.76,0.95)		0.44 (2.52×10^{-3})		Yes
	cg26784348	*SLC35D1*	1.11 (1.03,1.20)		0.66 (4.99×10^{-2})		No
	cg08959305	*TMEM97*	1.08 (1.02,1.14)		0.81 (4.14×10^{-2})		No

Supplementary Table S2(a) continued: CpG sites at birth associated with FEV1, FVC, FEV1/FVC and FEF25-75% with lung function trajectories in in IOWBC-F1 and replicate din ALSPAC with lung function measures at age 8, 15, and 24 years in boys

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			Consistency of direction of association between IOWBC-F1 and ALSPAC
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
Outcome	Ilmnid	Gene Name	Risk Ratio (95%CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEF25-75%	cg03724260	*ZBTB4; POLR2A*	0.91 (0.85,0.96)	-0.15 (2.17×10^{-2})			No
	cg00282510	*FOXE1; TRMO*	0.90 (0.84,0.96)	-0.11 (4.63×10^{-2})			No
	cg00566297	*RGS7; FH*	0.90 (0.84,0.97)			-0.55 (4.04×10^{-2})	No
	cg01678802	*COBL*	0.92 (0.87,0.97)	0.16 (1.99×10^{-3})			Yes
	cg05699739	*ESD; HTR2A*	0.88 (0.80,0.97)	-0.15 (5.77×10^{-3})			No
	cg06549530	*CHST15; CPXM2*	0.93 (0.89,0.98)		0.27 (2.41×10^{-02})		Yes
	cg09537551	*OSR2; VPS13B*	0.91 (0.85,0.96)	-0.16 (1.69×10^{-2})			No
	cg13529291	*WNT11; UVRAG*	1.14 (1.03,1.25)			-1.15 (2.41×10^{-3})	Yes
	cg19471553	*TMEM106A; NBR1*	0.92 (0.87,0.97)	-0.15 (3.93×10^{-2})			No
	cg26606286	*HAND2; HAND2-AS1*	1.11 (1.03,1.18)	-0.09 (2.53×10^{-2})			Yes

Supplementary Table S2 (b): CpG sites at birth in IOWBC-F1 associated with FEV1, FVC, FEV1/FVC and FEF25-75% trajectories and successfully replicated in ALSPAC with lung function at age 8 years, 15 years and 24 years in girls.

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95% CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FVC	cg00015603	*MYEOV2*	0.82 (0.75, 0.90)		0.13 (2.33×10^{-2})		Yes
	cg00175344	*POU4F1; RNF219*	0.87 (0.82, 0.93)	-0.1 (7.80×10^{-3})			No
	cg11418007	*GRINL1A; GCOM1;*	1.17 (1.07, 1.27)		0.18 (4.52×10^{-2})		No
	cg25080348	*ASPH*	1.34 (1.15, 1.54)	0.05 (4.84×10^{-2})			No
FEV1/ FVC	cg00964751	*C17orf96*	0.77 (0.68, 0.87)		-0.04 (5.43×10^{-3})		No
	cg02308192	*NEK9*	0.88 (0.83, 0.93)	0.06 (8.54×10^{-3})			Yes
	cg03433758	*CRTC3*	0.85 (0.79, 0.92)		0.05 (4.19×10^{-2})		Yes
	cg03869608	*PDZRN3*	0.86 (0.79, 0.93)	0.02 (1.17×10^{-2})			Yes
	cg06594008	*MKL1*	0.81 (0.74, 0.89)		0.02 (1.88×10^{-2})		Yes
	cg07052251	*COL6A1; PCBP3*	0.87 (0.81, 0.93)			-0.02 (3.66×10^{-2})	No
	cg08879910	*HLA-J; NCRNA00171*	1.11 (1.06, 1.17)			0.03 (1.81×10^{-2})	No
	cg11123440	*GATA4; NEIL2*	0.93 (0.90, 0.96)	0.01 (4.92×10^{-2})			Yes

Supplementary Table S2 (b) continued: CpG sites at birth in IOWBC-F1 associated with FEV1, FVC, FEV1/FVC and FEF25-75% trajectories and successfully replicated in ALSPAC with lung function at age 8 years, 15 years and 24 years in girls.

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95% CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEV1/ FVC	cg11294750	*GPC6*	0.85 (0.79, 0.91)			-0.02 (1.73×10^{-2})	No
	cg11301337	*PTPRN2*	1.20 (1.10, 1.30)			0.02 (1.53×10^{-2})	No
	cg12754671	*NDUFS2*	1.18 (1.10, 1.27)	-0.03 (3.51×10^{-2})			Yes
	cg13687497	*RXRA*	0.77 (0.69, 0.87)			-0.02 (7.54×10^{-3})	No
	cg16008148	*HLX;MARC1*	0.89 (0.85, 0.94)	0.01 (3.13×10^{-2})			Yes
	cg17636541	*NRM*	0.83 (0.76, 0.90)	0.02 (2.07×10^{-2})			Yes
	cg23221090	*NRBF2*	0.83 (0.76, 0.91)		-0.02 (9.96×10^{-3})		No
FEV1	cg00175344	*POU4F1; RNF219*	0.87 (0.81, 0.92)	-0.07 (2.20×10^{-2})		-0.15 ($4.51\times10^{-2)}$	No
	cg02637537	*PDLIM2*	0.84 (0.77, 0.91)	0.05 (1.24×10^{-2})		0.10 (3.31×10^{-2})	Yes
	cg12967384	*FBRSL1*	0.83 (0.76, 0.91)		-0.16 (3.34×10^{-3})		No
	cg14059822	*IRX2*	0.82 (0.74, 0.90)		0.17 (1.72×10^{-3})		Yes

Supplementary Table S2 (b) continued: CpG sites at birth in IOWBC-F1 associated with FEV1, FVC, FEV1/FVC and FEF25-75% trajectories and successfully replicated in ALSPAC with lung function at age 8 years, 15 years and 24 years in girls.

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
			Association of DNAm at birth with lung function trajectories at age 10, 18 and 26 years	Association of DNAm at birth with lung function measurements at specific ages			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95% CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
FEV1	cg16472998	*UNK*	1.13 (1.06, 1.2)			0.46 (1.05×10^{-3})	No
	cg20825472	*UST*	0.81 (0.75, 0.88)	0.05 (4.44×10^{-2})			Yes
	cg22665383	*OR4C16*	1.22 (1.11, 1.33)			0.13 (2.76×10^{-2})	No
	cg22843625	*C19orf63; FAM71E1*	1.11 (1.06, 1.17)			0.38 (1.77×10^{-3})	No
$FEF_{25-75\%}$	cg00383081	*GLUL*	0.77 (0.66, 0.89)		0.51 (5.18×10^{-3})		Yes
	cg03341334	*DUSP23*	1.14 (1.06, 1.22)		-0.91 (2.36×10^{-2})		Yes
	cg03795245	*GPR162; CD4*	1.45 (1.19, 1.77)	-0.16 (3.87×10^{-2})			Yes
	cg05226043	*TRIM26*	0.80 (0.70, 0.91)		0.34 (2.90×10^{-2})		Yes
	cg06338552	*ZC3H7B*	1.22 (1.06, 1.41)	-0.18 (1.59×10^{-2})	-0.32 (2.75×10^{-2})		Yes
	cg07380056	*GALNT13*	0.66 (0.5, 0.85)	0.14 (3.80×10^{-2})			Yes
	cg08591299	*ANKRD11; ZNF778*	0.75 (0.65, 0.87)		-0.26 (1.24×10^{-2})		No

Supplementary Table S2 (b) continued: CpG sites at birth in IOWBC-F1 associated with FEV1, FVC, FEV1/FVC and FEF25-75% trajectories and successfully replicated in ALSPAC with lung function at age 8 years, 15 years and 24 years in girls.

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95% CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
$FEF_{25-75\%}$	cg09845296	*TRIO; DNAH5*	1.20 (1.09, 1.31)			-0.25 (4.38×10^{-2})	Yes
	cg10179911	*RPRM*	0.84 (0.75, 0.94)			0.59 (2.03×10^{-2})	Yes
	cg10368052	*VPS52*	0.76 (0.64, 0.89)		0.20 (2.30×10^{-2})		Yes
	cg10887937	*ALDH2*	1.37 (1.17, 1.60)		0.23 (4.44×10^{-2})		No
	cg12629349	*EML6*	0.75 (0.64, 0.90)		0.53 (3.24×10^{-3})		Yes
	cg12635120	*SPPL3; HNF1A-AS1*	1.30 (1.07, 1.58)			-0.26 (3.24×10^{-2})	Yes
	cg13392687	*AMOTL1*	1.26 (1.10, 1.44)		0.37 (4.36×10^{-2})		No
	cg14846324	*BRUNOL6*	0.79 (0.68, 0.92)		0.35 (1.64×10^{-2})		Yes
	cg15485560	*KIZ*	0.80 (0.68, 0.93)		0.26 (1.11×10^{-2})		Yes
	cg15738933	*SEC14L1;SEC14L1;SEC*	1.51 (1.21, 1.88)		-0.31 (2.87×10^{-2})	-0.39 (1.10×10^{-2})	Yes
	cg18435928	*JAKMIP3*	1.30 (1.12, 1.51)	-0.09 (4.34×10^{-2})			Yes
	cg19244662	*SLC14A1*	1.42 (1.16, 1.76)	-0.11 (3.10×10^{-2})			Yes
	cg20697424	*NKIRAS2; DNAJC7*	1.41 (1.19, 1.67)		-0.19 (3.31×10^{-2})		Yes
	cg20934191	*DKFZp451B082*	1.31 (1.12, 1.53)		0.24 (2.16×10^{-2})		No

Supplementary Table S2 (b) continued: CpG sites at birth in IOWBC-F1 associated with FEV1, FVC, FEV1/FVC and FEF25-75% trajectories and successfully replicated in ALSPAC with lung function at age 8 years, 15 years and 24 years in girls.

			Discovery cohort: IOWBC-F1	Replication cohort: ALSAPC			
				Lung function at age 8 years	Lung function at age 15 years	Lung function at age 24 years	Consistency of direction of association between IOWBC-F1 and ALSPAC
Outcome	Ilmnid	Gene Name	Risk Ratio (95% CI) (Low vs. high trajectory)	Estimate (P-value)	Estimate P-value)	Estimate (P-value)	
$FEF_{25-75\%}$	cg21092296	*LOC90110*	1.24 (0.99, 1.54)		0.23 (4.53×10^{-2})		No
	cg22176566	*SLC2A1*	1.13 (1.05, 1.21)	0.42 (1.98×10^{-2})			No
	cg22311507	*MYCN*	0.91 (0.85, 0.96)	0.45 ($1.44\times10^{-2)}$	0.79 (4.66×10^{-2})	1.18 (3.44×10^{-3})	Yes
	cg22514863	*PAFAH1B3*	0.80 (0.70, 0.92)	0.24 (4.87×10^{-3})	0.34 (3.82×10^{-2})		Yes
	cg23182674	*FAM71B*	1.22 (1.08, 1.37)		0.37 (9.24×10^{-3})		No
	cg23395902	*PCGF5*	1.19 (1.09, 1.31)			0.54 (1.75×10^{-2})	No
	cg23817336	*TCP11; SCUBE3*	0.79 (0.68, 0.91)		0.40 (3.40×10^{-2})		Yes
	cg23937993	*AMPD3*	0.75 (0.64, 0.89)			0.37 (2.46×10^{-2})	Yes
	cg24656492	*SUPV3L1*	0.77 (0.66, 0.89)	0.15 (1.92×10^{-2})	0.35 (1.40×10^{-2})		Yes
	cg25967904	*TRMU*	1.25 (1.08, 1.46)		0.32 (8.23×10^{-3})		No

Supplementary Table S3: Assessing stability of DNAm at birth, age 10 years, and age 18 years between high and low lung function trajectories in IOWBC-F1. ($DNAm_{ij} = Age_j +$ lung function $trajectory_i + Age_j$* Lung function $trajectory_i + Subject_i + error_{ij}$ where i is subject and j is age of DNAm measurement).

Sex	Outcome	Ilmnid¥	Gene name	Type III sum of squares p-value for interaction term*
Boys	FEV1	cg00647165	*WIZ*	0.03
Boys	FEV1	cg05831188	*SDK1*	0.06
Boys	FEV1	cg05899183	*FAM96B; CES2*	0.03
Boys	FEV1	cg06710742	*STRA6; CCDC33*	0.91
Boys	FEV1	cg07103093	*LHX1*	0.19
Boys	FEV1	cg08207604	*DEF6*	0.07
Boys	FEV1	cg09771049	*KPNA2*	0.07
Boys	FEV1	cg11659361	*FBXO2; FBXO44*	0.03
Boys	FEV1	cg14836450	*RAPGEF4; LOC91149*	0.005
Boys	FEV1	cg16149007	*KLHDC10; MEM209*	0.02
Boys	FEV1	cg17330460	*PUS3*	0.18
Boys	FEV1	cg21437345	*B4GALNT2*	0.24
Boys	FEV1	cg21505334	*CEACAM5*	0.16
Boys	FEV1	cg21899743	*MYOM2*	0.07
Boys	FEV1/FVC	cg16750801	*PTPRE*	0.10
Boys	FVC	cg03224209	*ESRRG*	0.02
Boys	FVC	cg03918756	*TRAPPC9*	0.00
Boys	FVC	cg06639763	*FBXO31*	0.02
Boys	FVC	cg09771049	*KPNA2*	0.11
Boys	FVC	cg13904267	*DFNB31*	0.05
Boys	FEF2575%	cg01678802	*COBL*	0.33
Boys	FEF2575%	cg01885814	*COL18A1*	0.58
Boys	FEF2575%	cg06549530	*CHST15; CPXM2*	--#
Boys	FEF2575%	cg13529291	*WNT11; UVRAG*	0.79
Boys	FEF2575%	cg14281210	*MARCH8*	0.09
Boys	FEF2575%	cg16465128	*SEPT1*	0.001
Boys	FEF2575%	cg17366410	*AZI1*	0.0002
Boys	FEF2575%	cg20246851	*SEMA4B*	0.004
Boys	FEF2575%	cg22902534	*AP2M1*	0.01
Boys	FEF2575%	cg26606286	*HAND2; HAND2-AS1*	--#
Boys	FEF2575%	cg27582059	*PCDHB16*	0.06
Girls	FVC	cg00015603	*MYEOV2*	0.0002
Girls	FEV1/FVC	cg02308192	*NEK9*	0.12
Girls	FEV1/FVC	cg03433758	*CRTC3*	0.40
Girls	FEV1/FVC	cg03869608	*PDZRN3*	0.002
Girls	FEV1/FVC	cg06594008	*MKL1*	0.05
Girls	FEV1/FVC	cg11123440	*GATA4; NEIL2*	0.04
Girls	FEV1/FVC	cg12754671	*NDUFS2*	0.05

¥Only the CpGs associated with lung function in the same direction in both IOWBC-F1 and ALSPAC are tested here. *Non-significant interaction p-value indicates that DNAm does not change over time between the participants of the high and low lung function trajectories. #DNAm of this CpG is not available at all three time points.

Supplementary Table S3 continued: Assessing stability of DNAm at birth, age 10 years, and age 18 years between high and low lung function trajectories in IOWBC-F1 (DNAm~Age + lung function trajectories + age*lung function).

Sex	Outcome	Ilmnid[¥]	Gene name	Type III sum of squares p-value for the interaction term*
Girls	FEV1/FVC	cg16008148	*HLX; MARCH1*	0.63
Girls	FEV1/FVC	cg17636541	*NRM*	0.35
Girls	FEV1	cg02637537	*PDLIM2*	0.05
Girls	FEV1	cg14059822	*IRX2*	0.06
Girls	FEV1	cg20825472	*UST*	0.004
Girls	FEF25-75%	cg00383081	*GLUL*	0.45
Girls	FEF25-75%	cg03341334	*DUSP23*	0.08
Girls	FEF25-75%	cg05226043	*TRIM26*	0.32
Girls	FEF25-75%	cg06338552	*ZC3H7B*	0.03
Girls	FEF25-75%	cg07380056	*GALNT13*	0.73
Girls	FEF25-75%	cg09845296	*TRIO; DNAH5*	--[#]
Girls	FEF25-75%	cg10179911	*RPRM*	9.87E-05
Girls	FEF25-75%	cg10368052	*VPS52*	0.07
Girls	FEF25-75%	cg12629349	*EML6*	0.21
Girls	FEF25-75%	cg12635120	*SPPL3; HNF1A-AS1*	0.54
Girls	FEF25-75%	cg14846324	*BRUNOL6*	0.39
Girls	FEF25-75%	cg15485560	*KIZ*	0.54
Girls	FEF25-75%	cg15738933	*SEC14L1*	0.09
Girls	FEF25-75%	cg18435928	*JAKMIP3*	--[#]
Girls	FEF25-75%	cg19244662	*SLC14A1*	0.69
Girls	FEF25-75%	cg20697424	*NKIRAS2; DNAJC7*	0.17
Girls	FEF25-75%	cg22311507	*MYCN*	0.97
Girls	FEF25-75%	cg22514863	*PAFAH1B3*	0.00
Girls	FEF25-75%	cg23817336	*TCP11; SCUBE3*	0.13
Girls	FEF25-75%	cg23937993	*AMPD3*	0.05
Girls	FEF25-75%	cg24656492	*SUPV3L1*	0.07

[¥]Only the CpGs associated with lung function in the same direction in both IOWBC-F1 and ALSPAC are tested here. *Non-significant interaction p-value indicates that DNAm does not change over time between the participants of the high and low lung function trajectories. [#]DNAm of this CpG is not available at all three time points.

Supplementary table S4: Gene functional enrichment analysis performed using ToppFun algorithm

ID	Biological process	p-Value	FDR p-value	Gene names in input	Relation to respiratory system
GO:0035239	Tube morphogenesis	6.25×10^{-05}	1.34×10^{-02}	GLUL, MYCN, HLX, LHX1, COBL, COL18A1, STRA6, WNT11	During early embryogenesis lung develops from epithelial tubes that extensively branches to form a complex network of alveolar ducts and saccules to maximize the surface area for efficient gas exchange [31, 32].
GO:0048565	Digestive tract development	4.61×10^{-05}	1.34×10^{-02}	HLX, COBL, STRA6, WNT11	The digestive system also shares a common embryonic origin with lungs, as the ventral component of foregut tube branches compartmenting into the trachea [33].
GO:0030903	Notochord development	4.23×10^{-04}	3.70×10^{-02}	COBL, WNT11	The notochord plays a key role in the normal development of the esophagus and trachea from the foregut during embryogenesis[34].
GO:0061205	Paramesonephric duct development	1.86×10^{-05}	1.34×10^{-02}	LHX1, STRA6	Paramesonephric and mesonephric ducts are closely associated with urogenitalia development and these ducts require extensive branching of epithelial tubes during development like mammalian respiratory systems [35, 36].
GO:0072177	Mesonephric duct development	8.32×10^{-05}	1.34×10^{-02}	LHX1, WNT11	
GO:0030540	Female genitalia development	3.14×10^{-04}	2.97×10^{-02}	LHX1, STRA6	
GO:0048568	Embryonic organ development	5.73×10^{-05}	1.34×10^{-02}	HLX, LHX1, MYCN, STRA6, WNT11, COBL	These genes involved in processes related to lung morphogenesis or patterning HLX [37], LHX1 [38], MYCN [39], STRA6 [40], WNT11 [41], COBL [42].
GO:0061053	Somite development	3.15×10^{-04}	2.97×10^{-02}	LHX1, COBL, WNT11	The underlying processes that lead to somitogenesis are essentially similar to that of branch budding and extension during lung morphogenesis [43].

Supplementary Table S5 (a): Comparison of original risk ratios of directionally disagreeing CpGs with that in specific strata of risk factors in boys in IOWBC-F1

				Risk ratios in IOWBC-F1 in strata									
Outcome	Ilmnid	Estimate in ALSPAC£	Original Risk ratio in IOWBC-F1¥	Paternal asthma "Yes" * n=41	Paternal asthma "Yes" # n=41	Maternal asthma "Yes" * n=46	Maternal asthma "Yes" # n=46	Low birth weight (<=2.7kg) "Yes" * n=22	Low birth weight (<=2.7kg) "Yes" # n=22	Asthma at age 4 years "Yes" * n=61	Asthma at age 4 years "Yes" # n=61	Eczema at age 4 years "Yes" * n=45	Eczema at age 4 years "Yes" # n=45
FVC	cg01699600	-0.04	0.86	1.16	1.16				1.04				
FVC	cg14266217	0.39	1.15							0.87	0.88		
FVC	cg19616339	-0.07	0.78						1.07				
FVC	cg25611736	0.08	1.34						0.89				
FVC	cg27189973	-0.06	0.83					1.03	1.13				
FEV1	cg07460095	0.09	1.26				0.99					0.92	0.92
FEV1	cg10236596	-0.16	0.79			1.02	1.03	1.28	1.31				
FEV1	cg14706297	0.14	1.12					0.93					
FEV1	cg21200949	0.21	1.23						0.98				
FEV1/FVC	cg00983520	-0.01	0.86	1.10									
FEV1/FVC	cg07071157	0.04	1.19		0.98	1.00	0.97						
FEV1/FVC	cg16016281	-0.01	0.79			1.34	1.36						
$FEF_{25-75\%}$	cg01934962	-0.34	0.92								1.01		
$FEF_{25-75\%}$	cg03962214	-0.14	0.92	1.01								1.01	1.02
$FEF_{25-75\%}$	cg09738214	0.41	1.18								0.98		
$FEF_{25-75\%}$	cg19471553	-0.15	0.92								1.01		
$FEF_{25-75\%}$	cg24529269	1.8	1.07							0.99			
$FEF_{25-75\%}$	cg26829088	-0.18	0.89								1.00		

£ Estimate from the association of DNAm of CpG with lung function in ALSPAC using linear regression.

¥Original (un-stratified association) risk ratios of being in the low lung function trajectory with increasing DNAm of specified CpG in IOWBC.

*Adjusted for cell types.

Not adjusted for cell types.

Supplementary Table S5 (b): Comparison of original risk ratios of directionally disagreeing CpGs with that in specific strata of risk factors in girls in IOWBC-F1

Outcome	Ilmnid	Estimate in ALSPAC[£]	Original Risk ratio in IOWBC-F1[¥]	Paternal asthma "Yes" * n= 54	Paternal asthma "Yes" # n=54	Maternal asthma "Yes" * n=38	Maternal asthma "Yes" # n=38	Low birth weight (<=2.7kg) "Yes" * n=30	Low birth weight (<=2.7kg) "Yes" # n=30	Asthma at age 4 years "Yes" * n=54	Asthma at age 4 years "Yes" # n=54	Eczema at age 4 years "Yes" * n=41	Eczema at age 4 years "Yes" # n=41
FVC	cg00175344	-0.16	0.88					1.11				1.16	1.13
FVC	cg11418007	0.18	1.17									0.92	
FVC	cg25080348	0.05	1.34				0.93			0.92	0.89		
FEV1	cg00175344	-0.15	0.87					1.16				1.22	1.17
FEV1	cg12967384	-0.16	0.84	1.13				1.02		1.01			
FEV1/FVC	cg07052251	-0.02	0.87									1.17	
FEV1/FVC	cg08879910	0.03	1.11									0.95	0.97
FEV1/FVC	cg11301337	0.02	1.20					0.87					
$FEF_{25-75\%}$	cg10887937	0.23	1.37					0.91	0.94				
$FEF_{25-75\%}$	cg13392687	0.4	1.26	0.83						0.99			
$FEF_{25-75\%}$	cg20934191	0.24	1.31				0.83						
$FEF_{25-75\%}$	cg21092296	0.23	1.24			0.61	0.55	0.88	0.90	0.91		0.92	
$FEF_{25-75\%}$	cg22176566	0.42	1.13							0.89		0.95	
$FEF_{25-75\%}$	cg25967904	0.32	1.26	0.93									

[£] Estimate from the association of DNAm of CpG with lung function in ALSPAC using linear regression.

[¥]Original (un-stratified association) risk ratios of being in the low lung function trajectory with increasing DNAm of specified CpG in IOWBC.

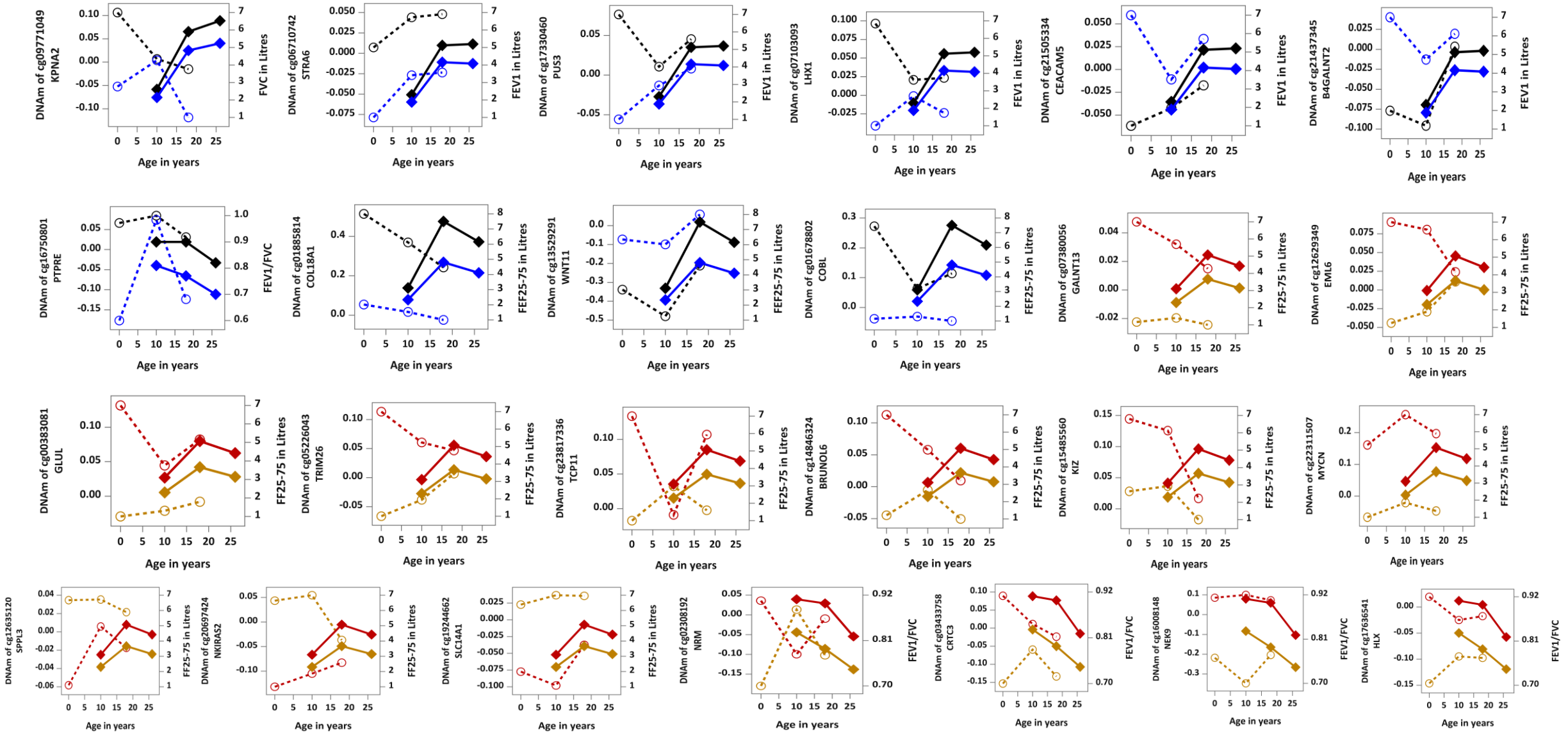
*Adjusted for cell types.

[#] Not adjusted for cell types.

References

1. Soto-Ramirez, N., et al., *The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years.* Clin Epigenetics, 2013. **5**(1): p. 1.
2. Arshad, S.H., et al., *Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC).* Int J Epidemiol, 2018. **47**(4): p. 1043-1044i.
3. Arshad, S.H. and D.W. Hide, *Effect of environmental factors on the development of allergic disorders in infancy.* J Allergy Clin Immunol, 1992. **90**(2): p. 235-41.
4. Arshad, S.H., M. Stevens, and D.W. Hide, *The effect of genetic and environmental factors on the prevalence of allergic disorders at the age of two years.* Clin Exp Allergy, 1993. **23**(6): p. 504-11.
5. Scott, M., et al., *Influence of atopy and asthma on exhaled nitric oxide in an unselected birth cohort study.* Thorax, 2010. **65**(3): p. 258-62.
6. Karmaus, W., et al., *Distinctive lung function trajectories from age 10 to 26 years in men and women and associated early life risk factors - a birth cohort study.* Respir Res, 2019. **20**(1): p. 98.
7. *Koko spirometer & koko digidoser windows operations guide. .* 2002.
8. *Standardization of Spirometry, 1994 Update. American Thoracic Society.* Am J Respir Crit Care Med, 1995. **152**(3): p. 1107-36.
9. Arshad, S.H., et al., *Polymorphisms in the interleukin 13 and GATA binding protein 3 genes and the development of eczema during childhood.* Br J Dermatol, 2008. **158**(6): p. 1315-22.
10. Hanifin, J., Rajka G. , *Diagnostic features of atopic eczema. .* Acta Derm Venereol (Stockh), 1980. **92**: p. 44–7.
11. Beyan, H., et al., *Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans.* Genome Res, 2012. **22**(11): p. 2138-45.
12. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.
13. Pidsley, R., et al., *Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.* Genome Biol, 2016. **17**(1): p. 208.
14. Roessler, J., et al., *Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc.* BMC Res Notes, 2012. **5**: p. 210.
15. Bakulski, K.M., et al., *DNA methylation of cord blood cell types: Applications for mixed cell birth studies.* Epigenetics, 2016. **11**(5): p. 354-62.
16. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies.* Genome Biol, 2015. **16**: p. 37.
17. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
18. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
19. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.
20. Boyd, A., et al., *Cohort profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children.* Int J Epidemiol., 2013. **42**.
21. Fraser, A., et al., *Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.* Int J Epidemiol., 2013. **42**.
22. Fraser, A., et al., *Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.* Int J Epidemiol, 2013. **42**(1): p. 97-110.

23. Relton, C.L., et al., *Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).* International Journal of Epidemiology, 2015. **44**(4): p. 1181-1190.
24. Relton, C.L., et al., *Data resource profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).* Int J Epidemiol., 2015. **44**.
25. Min, J.L., et al., *Meffil: efficient normalization and analysis of very large DNA methylation datasets.* Bioinformatics, 2018: p. bty476-bty476.
26. Bandyopadhyay, G., et al., *Dissociation, cellular isolation, and initial molecular characterization of neonatal and pediatric human lung tissues.* Am J Physiol Lung Cell Mol Physiol, 2018. **315**(4): p. L576-L583.
27. den Dekker, H.T., et al., *Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course.* Eur Respir J, 2019. **53**(4).
28. Imboden, M., et al., *Epigenome-wide association study of lung function level and its change.* Eur Respir J, 2019.
29. Barton, S.J., et al., *In Epigenomic Studies, Including Cell-Type Adjustments in Regression Models Can Introduce Multicollinearity, Resulting in Apparent Reversal of Direction of Association.* Frontiers in Genetics, 2019. **10**(816).
30. Do, C., et al., *Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era.* Genome Biol, 2017. **18**(1): p. 120.
31. Ochoa-Espinosa, A. and M. Affolter, *Branching morphogenesis: from cells to organs and back.* Cold Spring Harb Perspect Biol, 2012. **4**(10).
32. Horowitz, A. and M. Simons, *Branching morphogenesis.* Circ Res, 2008. **103**(8): p. 784-95.
33. Ioannides, A.S., et al., *Foregut separation and tracheo-oesophageal malformations: the role of tracheal outgrowth, dorso-ventral patterning and programmed cell death.* Dev Biol, 2010. **337**(2): p. 351-62.
34. Katz J, M.A., Basit H., *Embryology, Esophagus.* Treasure Island (FL): StatPearls Publishing, 2020. **In: StatPearls [Internet].**
35. Orvis, G.D. and R.R. Behringer, *Cellular mechanisms of Mullerian duct formation in the mouse.* Dev Biol, 2007. **306**(2): p. 493-504.
36. Qi, B.Q. and S.W. Beasley, *Relationship of the notochord to foregut development in the fetal rat model of esophageal atresia.* J Pediatr Surg, 1999. **34**(11): p. 1593-8.
37. Hentsch, B., et al., *Hlx homeo box gene is essential for an inductive tissue interaction that drives expansion of embryonic liver and gut.* Genes Dev, 1996. **10**(1): p. 70-9.
38. Herriges, J.C., et al., *Genome-scale study of transcription factor expression in the branching mouse lung.* Dev Dyn, 2012. **241**(9): p. 1432-53.
39. Okubo, T., et al., *Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation.* Development, 2005. **132**(6): p. 1363-74.
40. Pasutto, F., et al., *Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation.* Am J Hum Genet, 2007. **80**(3): p. 550-60.
41. Yin, Y., et al., *An FGF-WNT gene regulatory network controls lung mesenchyme development.* Dev Biol, 2008. **319**(2): p. 426-36.
42. Ravanelli, A.M. and J. Klingensmith, *The actin nucleator Cordon-bleu is required for development of motile cilia in zebrafish.* Dev Biol, 2011. **350**(1): p. 101-11.
43. Warburton, D., et al., *Lung organogenesis.* Curr Top Dev Biol, 2010. **90**: p. 73-158.



Supplementary Figure S1: The pattern of DNAM for 25 stable CpGs at birth, age 10 and 18 years in participants belonging to the high and low trajectories of lung function in IOWBC in boys (black and blue lines) and girls (red and orange lines). The left-hand Y-axis represents the residual DNAm after regressing out the effect of cell types for each time point and the right-hand Y-axis represents the lung function levels (in Liters).

Supplementary Figure S2: Mean expression of genes in specific lung cells obtained from online LungMap data repository (https://lungmap.net) at specific ages in human donor lung tissue from the BRINDL repository samples (PubMed PMID: 2997510). These genes are corresponding to the CpGs that were linked to lung function in IOWBC-F1 and ALSAPC in the same direction, stable over time in IOWBC-F1, and located on biologically relevant genes.