DNA methylation at birth is associated with lung function development until 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] and Wilfried Karmaus[1]

Affiliations: [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]Dept 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, Southampton, UK. [11]College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA.

Correspondence: Wilfried Karmaus, School of Public Health, University of Memphis, 301 Robison Hall, 3825 DeSoto Avenue, Memphis, TN 38152, USA. E-mail: karmaus1@memphis.edu

@ERSpublications
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. https://bit.ly/37qtT6Z

Cite this article as: Mukherjee N, Arathimos R, Chen S, *et al*. DNA methylation at birth is associated with lung function development until age 26 years. *Eur Respir J* 2021; 57: 2003505 [https://doi.org/10.1183/13993003.03505-2020].

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 in 1 s, forced vital capacity, their ratio, and forced expiratory flow at 25–75% of forced vital capacity) 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 focussed 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 eight 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, aged 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.

This article has an editorial commentary: https://doi.org/10.1183/13993003.00109-2021

This article has supplementary material available from erj.ersjournals.com

Received: 22 April 2020 | Accepted: 14 Oct 2020

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 fertilisation [11], which is re-established after implantation [11]. Fetal lung development initiates approximately 4 weeks after fertilisation, during which pluripotent embryonic cells differentiate into specialised 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% of FVC ($FEF_{25–75\%}$) 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 analysed 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, FEV_1, 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 $\log2(\beta/(1-\beta))$ to analyse 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 850 K) 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 ($p\text{-value} \leqslant 0.05$) 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.

Hierarchical steps	Number of CpGs in boys					Number of CpGs in girls				
Discovery in IOWBC-F1	FEV_1	FVC	FEV_1/FVC	FEF#	Total	FEV_1	FVC	FEV_1/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 ($\alpha \leq 0.05$) at ages 8, 15 and 24 years	21	10	5	32	68	8	4	15	31	58
CpGs significantly associated in ALSPAC ($\alpha \leq 0.05$) 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 cytosine–phosphate–guanine sites (CpGs) predicting lung function. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; FDR: false discovery rate; IOWBC-F1/F2: F1/F2-generation of Isle of Wight birth cohort. #: FEF refers to forced expiratory flow at 25–75% of forced vital capacity; ¶: only the CpGs available in Illumina 450K array were tested in the Avon Longitudinal Study of Parents and Children (ALSPAC); +: stability of DNA methylation (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.

We replicated our findings in ALSPAC using cord blood DNAm (Illumina 450 K) and FVC, FEV_1, FEV_1/FVC and $FEF_{25-75\%}$ measurements at ages 8, 15, or 24 years. Among the discovered CpGs in IOWBC-F1 (Illumina 850 K), we tested for replication in ALSPAC (Illumina 450 K) only those CpGs also available in 450 K data (figure 1). We analysed 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 ($\alpha \leqslant 0.05$) 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 analysed 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 and 2. 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 850 K) 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

TABLE 1 Characteristics of boys with lung function and DNA methylation data in the F1-generation of Isle of Wight birth cohort (IOWBC-F1) and the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort

	IOWBC-F1[¶]			ALSPAC		
	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)**	**Age 24 years (n=155)**
FEV_1 L	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 L	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)
FEV_1/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)
$FEF_{25–75\%}$ L	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)
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)

Data are presented as median (range) or n (%). FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; $FEF_{25–75\%}$: forced expiratory flow at 25–75% of FVC. [#]: concurrent smoking refers to either active or second-hand 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 26 years, respectively.

change the risk ratios by more than 10% (table S1), and thus were excluded from the models. These associations were replicated in ALSPAC (Illumina 450 K) 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 450 K data (figure 1). We found 68 CpGs in boys and 58 CpGs in girls to be significantly associated ($\alpha \leqslant 0.05$) with lung function measures in ALSPAC (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 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 (table S3 and 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 (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

TABLE 2 Characteristics of girls with lung function and DNA methylation data in F1-generation of Isle of Wight birth cohort (IOWBC-F1) and the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort

	IOWBC-F1[¶]			ALSPAC		
	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)**	**Age 24 years (n=257)**
FEV_1 L	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 L	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)
FEV_1/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)
$FEF_{25–75\%}$ L	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)
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)

Data are presented as median (range) or n (%). FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; $FEF_{25–75\%}$: forced expiratory flow at 25–75% of FVC. [#]: concurrent smoking refers to either active or second-hand smoking (maternal/paternal/other smoking/outside home smoking) or both; [¶]: among the 390 girls included in analyses, 387, 377 and 332 have lung function measured at age 10, 18 and 26 years, respectively.

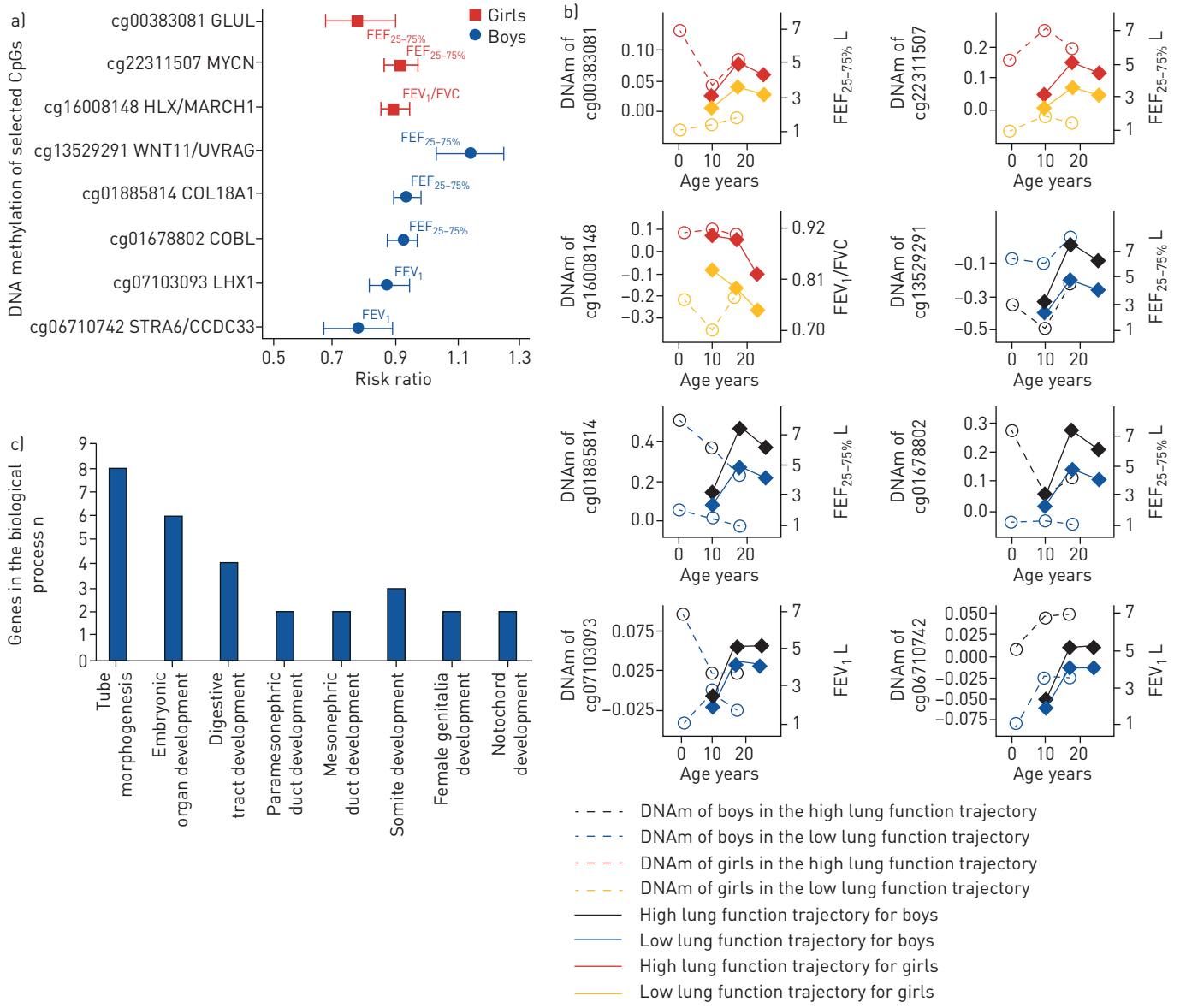

FIGURE 2 Attributes of eight important cytosine–phosphate–guanine sites (CpGs) significantly associated with lung function in the same direction in the F1-generation of Isle of Wight birth cohort (IOWBC-F1) and the Avon Longitudinal Study of Parents and Children (ALSPAC) in the same direction, stable, and enriched in relevant biological processes. a) Risk ratios (95% CI) for lower lung function trajectory corresponding to each CpG in IOWBC-F1; b) The pattern of consistently different DNA methylation (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 abscissa represents the residual DNAm after regressing out the effect of cell types for each time point and the right-hand abscissa represents the lung function levels (in litres). 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.

biological functions. In boys, three CpGs were associated with $FEF_{25–75\%}$ and two CpGs were associated with FEV_1. In girls, two CpGs were linked to $FEF_{25–75\%}$ whereas one was linked to FEV_1/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 figures 2a and b, 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 (table S3 and 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 (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 recognised [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 (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 den 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] analysed boys and girls together.

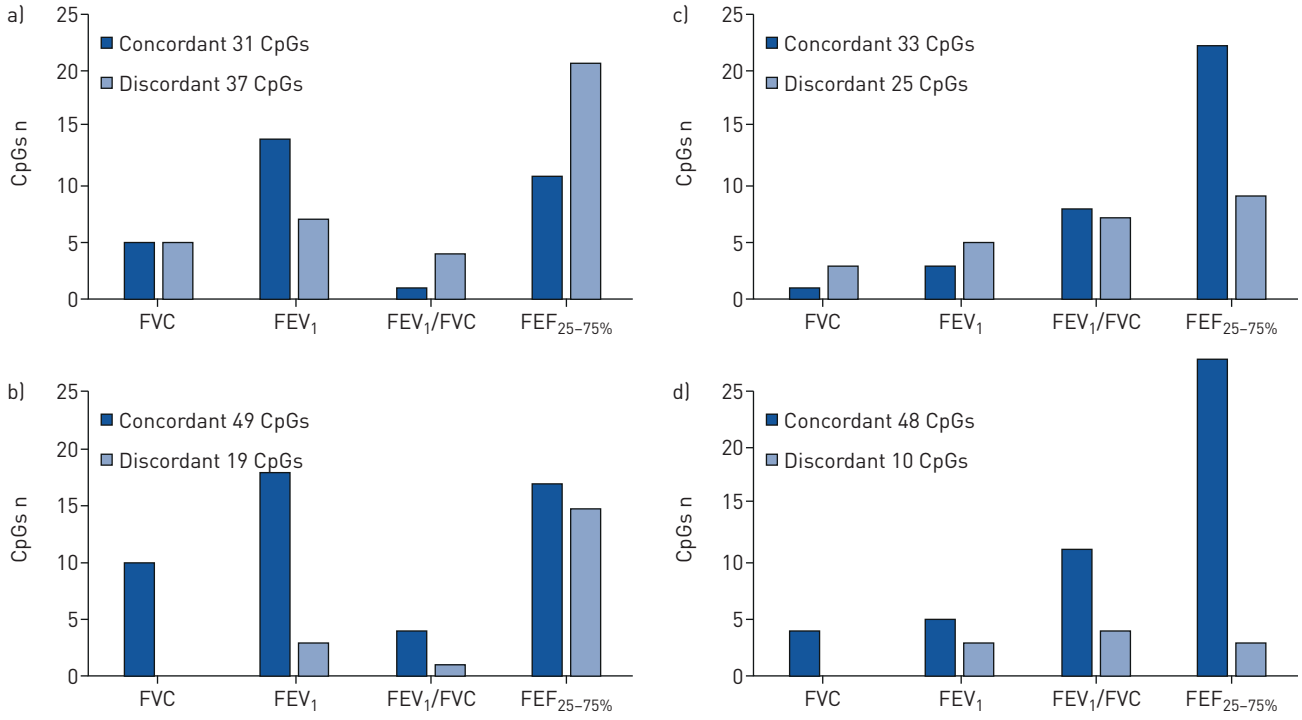

FIGURE 3 Agreement of direction of association of DNA methylation (DNAm) with lung function comparing overall sample of boys and girls in the F1-generation of Isle of Wight birth cohort (IOWBC-F1) to the Avon Longitudinal Study of Parents and Children (ALSPAC) (a and c), and comparing risk factor strata of boys and girls in IOWBC-F1 to ALSPAC (b and d). Dark blue and light blue bars indicate the number of cytosine–phosphate–guanine sites (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 out of 68) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC; (b) in risk factor strata of boys: 28% (19 out of 68) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC; (c) overall unstratified sample of girls: 43% (25 out of 58) of IOWBC-F1 CpGs show discordant directions compared to ALSPAC; (d) in risk factor strata of girls: 17% (10 out of 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. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; $FEF_{25–75\%}$: forced expiratory flow at 25–75% of FVC.

Imboden *et al.* [40] performed cross-sectional analyses of DNAm and lung function, excluding $FEF_{25–75\%}$, measured during mid- to late adulthood. den Dekker *et al.* [15] 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. 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, table S5a and S5b). 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 850 K) compared to ALSPAC (Illumina 450 K), 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 www.lungmap.net/, all eight biologically relevant genes were expressed in human lungs in early life and adulthood (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 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 Gemma Sharp for her contribution towards the quality control of the ARIES data.

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

Conflict of interest: N. Mukherjee has nothing to disclose. R. Arathimos has nothing to disclose. S. Chen has nothing to disclose. P. Kheirkhah Rahimabad has nothing to disclose. L. Han has nothing to disclose. H. Zhang has nothing to disclose. J.W. Holloway reports grants from the National institutes of Health (USA), during the conduct of the study. C. Relton has nothing to disclose. S.H. Arshad has nothing to disclose. S. Ewart has nothing to disclose. W. Karmaus has nothing to disclose.

Support statement: 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. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Fabbri LM, Romagnoli M, Corbetta L, *et al.* 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: 418–424.
2 McGeachie MJ, Yates KP, Zhou X, *et al.* Patterns of growth and decline in lung function in persistent childhood asthma. *N Engl J Med* 2016; 374: 1842–1852.
3 Bui DS, Lodge CJ, Burgess JA, *et al.* 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: 535–544.
4 Berry CE, Billheimer D, Jenkins IC, *et al.* A distinct low lung function trajectory from childhood to the fourth decade of life. *Am J Respir Crit Care Med* 2016; 194: 607–612.
5 Majewska R, Pac A, Mroz E, *et al.* 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: 66–70.
7 Tee AK. Chronic obstructive pulmonary disease (COPD): 'not a cigarette only pulmonary disease'. *Ann Acad Med Singap* 2017; 46: 415–416.
8 Bianco-Miotto T, Craig JM, Gasser YP, *et al.* Epigenetics and DOHaD: from basics to birth and beyond. *J Dev Orig Health Dis* 2017; 8: 513–519.
9 Chen ZX, Riggs AD. DNA methylation and demethylation in mammals. *J Biol Chem* 2011; 286: 18347–18353.
10 Shenker N, Flanagan JM. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. *Br J Cancer* 2012; 106: 248–253.
11 Wadhwa PD, Buss C, Entringer S, *et al.* Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med* 2009; 27: 358–368.
12 Durham AL, Adcock IM. Basic science: Epigenetic programming and the respiratory system. *Breathe* 2013; 9: 278–288.
13 Joss-Moore LA, Albertine KH, Lane RH. Epigenetics and the developmental origins of lung disease. *Mol Genet Metab* 2011; 104: 61–66.

14 Gao L, Urman R, Millstein J, *et al.* Association between AXL promoter methylation and lung function growth during adolescence. *Epigenetics* 2018; 13: 1027–1038.
15 den Dekker HT, Burrows K, Felix JF, *et al.* Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. *Eur Respir J* 2019: 53: 1801795.
16 Lange P, Celli B, Agusti A, *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015; 373: 111–122.
17 Karmaus W, Mukherjee N, Janjanam VD, *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: 98.
18 Nagin DS, Jones BL, Passos VL, *et al.* Group-based multi-trajectory modeling. *Stat Methods Med Res* 2018; 27: 2015–2023.
19 Ray MA, Tong X, Lockett GA, *et al.* An efficient approach to screening epigenome-wide data. *Biomed Res Int* 2016; 2016: 2615348.
20 Houseman EA, Accomando WP, Koestler DC, *et al.* 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: 1724–1735.
22 Leek JT, Johnson WE, Parker HS, *et al.* The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; 28: 882–883.
23 Chen J, Bardes EE, Aronow BJ, *et al.* ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009; 37: W305–W311.
24 Boezen HM, Jansen DF, Postma DS. Sex and gender differences in lung development and their clinical significance. *Clin Chest Med* 2004; 25: 237–245.
25 Suderman M, Simpkin A, Sharp G, *et al.* Sex-associated autosomal DNA methylation differences are wide-spread and stable throughout childhood. *bioRxiv* 2017; preprint [https://doi.org/10.1101/118265].
26 LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe (Sheff)* 2018; 14: 131–140.
27 DeBoer MD, Phillips BR, Mauger DT, *et al.* Effects of endogenous sex hormones on lung function and symptom control in adolescents with asthma. *BMC Pulm Med* 2018; 18: 58.
28 Soto-Ramirez N, Ziyab AH, Karmaus W, *et al.* Epidemiologic methods of assessing asthma and wheezing episodes in longitudinal studies: measures of change and stability. *J Epidemiol* 2013; 23: 399–410.
29 Park JE, Chung JH, Lee KH, *et al.* The effect of body composition on pulmonary function. *Tuberc Respir Dis (Seoul)* 2012; 72: 433–440.
30 Postma DS. Gender differences in asthma development and progression. *Gend Med* 2007; 4 Suppl B: S133–S146.
31 Jones BG, Penkert RR, Surman SL, *et al.* 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, *et al.* 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, *et al.* Sexually dimorphic DNA-methylation in cardiometabolic health: A systematic review. *Maturitas* 2020; 135: 6–26.
34 Kim S, Cho YH, Won S, *et al.* 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, *et al.* Epigenome wide comparison of DNA methylation profile between paired umbilical cord blood and neonatal blood on Guthrie cards. *Epigenetics* 2020; 15: 454–461.
36 Bjornsson HT, Sigurdsson MI, Fallin MD, *et al.* Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008; 299: 2877–2883.
37 Urdinguio RG, Torro MI, Bayon GF, *et al.* Longitudinal study of DNA methylation during the first 5 years of life. *J Transl Med* 2016; 14: 160.
38 Lubarsky B, Krasnow MA. Tube morphogenesis: making and shaping biological tubes. *Cell* 2003; 112: 19–28.
39 Nelson WJ. Tube morphogenesis: closure, but many openings remain. *Trends Cell Biol* 2003; 13: 615–621.
40 Imboden M, Wielscher M, Rezwan FI, *et al.* Epigenome-wide association study of lung function level and its change. *Eur Respir J* 2019; 54: 1900457.
41 Barton SJ, Melton PE, Titcombe P, *et al.* In epigenomic studies, including cell-type adjustments in regression models can introduce multicollinearity, resulting in apparent reversal of direction of association. *Front Genet* 2019; 10: 816.