Sex-specific longitudinal association of DNA methylation with lung function

Shadia Khan Sunny [1], Hongmei Zhang [1], Caroline L. Relton[2], Susan Ring[2,3], Latha Kadalayil[4], Fawaz Mzayek[1], Susan Ewart[5], John W. Holloway [4,6] and S. Hasan Arshad[6,7,8]

[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, University of Bristol, Bristol, UK. [4]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK. [5]Large Animal Clinical Sciences, Michigan State University, East Lansing, MI, USA. [6]NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, UK. [7]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. [8]The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Isle of Wight, UK.

Corresponding author: Hongmei Zhang (hzhang6@memphis.edu)

Abstract

Investigating whether DNA methylation (DNA-M) at an earlier age is associated with lung function at a later age and whether this relationship differs by sex could enable prediction of future lung function deficit.

A training/testing-based technique was used to screen 402714 cytosine-phosphate-guanine dinucleotide sites (CpGs) to assess the longitudinal association of blood-based DNA-M at ages 10 and 18 years with lung function at 18 and 26 years, respectively, in the Isle of Wight birth cohort (IOWBC). Multivariable linear mixed models were applied to the CpGs that passed screening. To detect differentially methylated regions (DMRs), DMR enrichment analysis was conducted. Findings were further examined in the Avon Longitudinal Study of Parents and Children (ALSPAC). Biological relevance of the identified CpGs was assessed using gene expression data.

DNA-M at eight CpGs (five CpGs with forced expiratory volume in 1 s (FEV_1) and three CpGs with FEV_1/ forced vital capacity (FVC)) at an earlier age was associated with lung function at a later age regardless of sex, while at 13 CpGs (five CpGs with FVC, three with FEV_1 and five with FEV_1/FVC), the associations were sex-specific ($p_{FDR}<0.05$) in IOWBC, with consistent directions of association in ALSPAC (IOWBC-ALSPAC consistent CpGs). cg16582803 (*WNT10A*) and cg14083603 (*ZGPAT*) were replicated in ALSPAC for main and sex-specific effects, respectively. Among IOWBC-ALSPAC consistent CpGs, DNA-M at cg01376079 (*SSH3*) and cg07557690 (*TGFBR3*) was associated with gene expression both longitudinally and cross-sectionally. In total, 57 and 170 DMRs were linked to lung function longitudinally in males and females, respectively.

CpGs showing longitudinal associations with lung function have the potential to serve as candidate markers in future studies on lung function deficit prediction.

Introduction

Lung function is pivotal for the diagnosis of respiratory diseases and predicts future disease development [1]. Lung function, specifically forced expiratory volume in 1 s (FEV_1), is inversely correlated with morbidities such as asthma and COPD, and early mortality [2, 3]. The growth of lung function in childhood and adolescence is associated with age and height and the decline in adulthood is related to ageing [3, 4]. In addition, the maximal level of lung function and the age of decline are dependent on sex [3, 5]. Several biological factors determine such sex-dependency including anatomical, immunological, and hormonal factors [5, 6].

The impact of environmental factors on respiratory health and lung function is significant [7]. The importance of the interaction between genetic and environmental factors in determining lung function suggests that other gene regulatory processes [8], such as epigenetic mechanisms, may act as an interface between environmental exposures and genetics [9, 10]. DNA methylation (DNA-M), most commonly the addition of a methyl group onto the 5′ position of the cytosine base at cytosine-phosphate-guanine dinucleotide sites (CpGs), regulates gene expression by recruiting proteins involved in gene repression or by impeding the binding of transcriptional proteins to DNA [11]. Several studies have shown the association of blood-based DNA-M with lung function [12–17] or with related diseases such as asthma [18] and COPD [12–15]. Most existing epigenetic studies on lung function were cross-sectional and focused on older people (>40 years) [12–16]. Cross-sectional designs are subject to reverse causation and create temporal ambiguity. To our knowledge, no existing studies have used repeated measurements of DNA-M, together with longitudinally measured lung function to assess the association of DNA-M with lung function and the stability of these associations over time.

DNA-M changes over time at specific CpGs [19, 20], and such changes have been shown to be sex-specific [20, 21]. Changes in DNA-M can occur in response to biological ageing, but also to environmental exposures [22]. That is, DNA-M at certain CpGs reflects the memory of past exposure as well as significant changes at different stages of life. The association between change in DNA-M and change in lung function has been shown to be different between males and females [23]. However, it is unknown whether DNA-M at an earlier age is associated with lung function at a later age, whether such longitudinal associations are invariant to DNA-M changes over time, and how such associations are different between males and females. A longitudinal design with repeatedly measured DNA-M and lung function data would allow assessment of the stability of time-lagged associations between DNA-M and lung function. As DNA-M has been found to be a potential driver of biological ageing [24], DNA-M biomarkers which have a stable time-lagged association could be useful to predict lung function deficit and detect possible related diseases at an earlier age before the pathology becomes apparent. We hypothesised that DNA-M at specific CpGs in early life is associated with lung function at a later age and such association would be sex-specific. The study was carried out in the Isle of Wight birth cohort (IOWBC) in the UK. To assess generalisability, the findings were further examined in an independent birth cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort in the UK.

Materials and methods

Study subjects and design

IOWBC: discovery cohort

IOWBC is a prospective population-based birth cohort established in 1989. Longitudinal monitoring of allergic diseases, phenotypic measures, genetics and assessments of environmental exposures were conducted at birth and multiple ages from 1 year to 26 years of age. Forced vital capacity (FVC) and FEV_1 were measured at age 10 years (n=980), 18 years (n=838) and 26 years (n=546), and the ratio of FEV_1 over FVC (FEV_1/FVC) was calculated. Genome-wide DNA-M was measured from peripheral blood samples collected at age 10 years (n=330), 18 years (n=476) and 26 years (n=303) from randomly selected subjects for whom DNA was available using the Infinium HumanMethylation450K or EPIC BeadChips (Illumina, Inc., San Diego, CA, USA). After quality control, preprocessing, and excluding probes with single nucleotide polymorphisms, 402 714 CpGs were included in the statistical analyses. RNA-sequencing gene expression data for subjects at age 26 years was available in IOWBC. A detailed description of IOWBC can be found in the online supplement.

ALSPAC: replication cohort

Findings in the IOWBC were further tested in an independent cohort, ALSPAC [25, 26], where DNA-M data at age 7 and 15 years and lung function measurements at age 15 and 24 years were available for replication analyses. Details of these data along with information on covariates are presented in the online supplement. The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

Statistical analyses

To assess whether subjects examined in the study at ages 18 and 26 years reasonably represented those in the complete IOWBC, continuous variables were evaluated using nonparametric one sample sign tests and categorical variables were examined implementing one-sample proportion tests.

Analyses of longitudinal association

Lung function measurements at each age were adjusted by height. DNA-M adjusted for cell types, principle components, and batch effects at each CpG was used (see details in the online supplement). In IOWBC, a two-step analytical approach was used to assess the longitudinal association between DNA-M and lung function at two time-lagged periods: period-1 (10–18 years), the association of DNA-M at age 10 years with lung function at 18 years; and period-2 (18–26 years), the association of DNA-M at age 18 years with lung function at 26 years. In the first step, we filtered out CpGs not potentially associated with lung function in either of the two periods using a screening package, "*ttScreening*" in R 3.3.2 version (detailed in the online supplement) [27, 28]. The screening was applied to each lung function parameter and performed for both time periods, stratified by sex.

In the second step, linear mixed models (LMM) with repeated measures were implemented in period-1 and period-2 in SAS 9.4 (SAS Institute Inc., Cary, NC, USA). Model-1 focused on the main effects of DNA-M. Potential confounders, including birth weight, gestational age, sex, duration of breastfeeding, maternal smoking exposure during pregnancy, recurrent chest infection at ages 1, 2 and 4 years, socioeconomic status, repeated measures of body mass index, smoking status, and paracetamol use at ages 18 and 26 years, were included in model-1. To assess sex-specificity, we further extended model-1 by including DNA-M×sex interaction in model-2. Multiple testing was corrected by controlling false discovery rate (FDR) of 0.05 in both models [29].

Analyses of differentially methylated regions (DMRs)

Regional differential methylation signals among the CpGs that passed screening were examined using DMRcate [30] with default settings of including $\geqslant 2$ significant CpGs that passed screening in a region of $\geqslant 1000$ nucleotides ($p_{FDR} < 0.05$) [30] (detailed in the online supplement).

Replication analysis in ALSPAC

The CpGs identified in IOWBC were further examined in ALSPAC to validate the IOWBC findings. Following a similar approach as that in the IOWBC, *i.e. via* LMMs with repeated measures, the
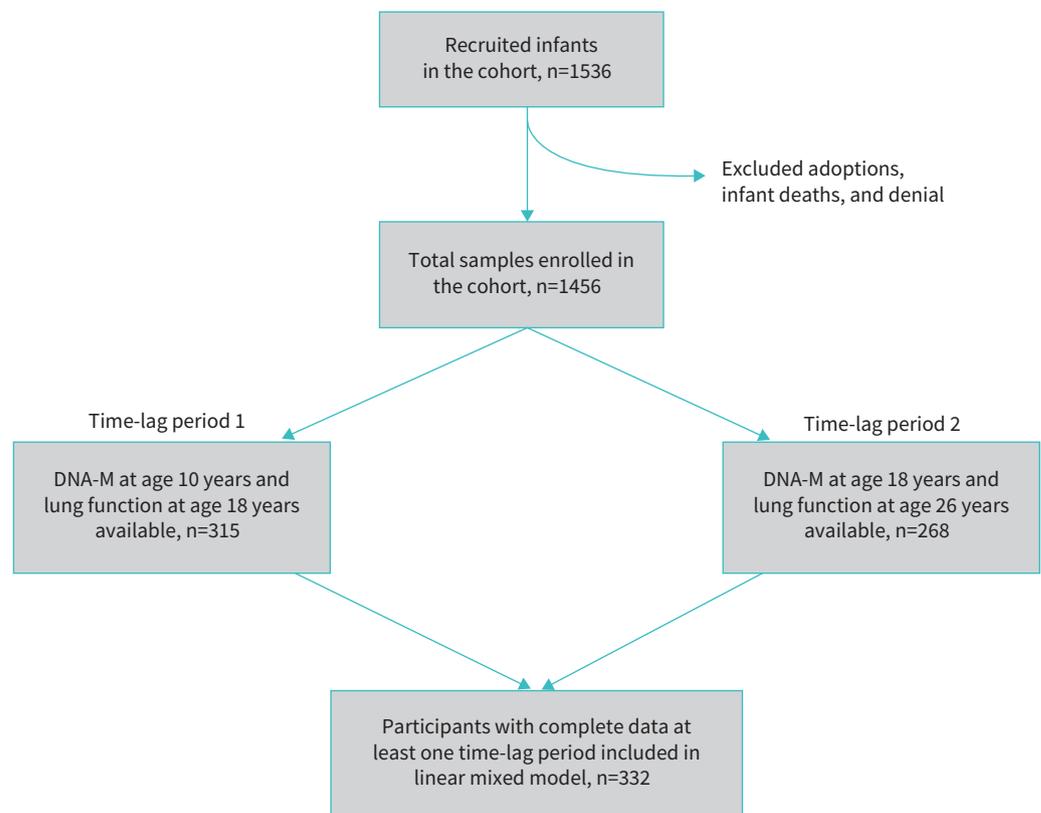

FIGURE 1 Flow chart for final sample determination in the Isle of Wight birth cohort. DNA-M: DNA methylation.

TABLE 1 Comparison of lung function measurements of enrolled participants and participants included in the analyses

Variables	Analytical sample	Enrolled sample	p-value
Age 18 years	**n=315**	**n=839**	
Lung function			
FVC (L)	4.62 (4.04, 5.52)	4.47 (3.93, 5.32)	0.071
FEV_1 (L)	4.11 (3.55, 4.74)	3.91 (3.44, 4.51)	**0.0067**
FEV_1/FVC	0.88 (0.82, 0.93)	0.88 (0.83, 0.92)	0.37
Height (cm)	172.5 (164, 178.5)	171 (164, 178)	0.36
BMI (kg·m^{-2})	22.36 (20.32, 24.84)	22.15 (20.27, 24.81)	0.40
Sex			
Male	181 (57.46)	396 (47.20)	**0.0019**
Female	134 (42.54)	443 (52.80)	
Smoking			
Non-smoker	167 (53.02)	462 (55.07)	0.90
Current smoker	80 (25.40)	205 (24.43)	
Past smoker	64 (20.32)	159 (18.95)	
Missing	4 (1.27)	13 (1.55)	
Age 26 years	**n=268**	**n=547**	
Lung function			
FVC (L)	4.66 (4.13, 5.57)	4.72 (4.14, 5.69)	0.76
FEV_1 (L)	3.70 (3.33, 4.49)	3.76 (3.33, 4.55)	0.18
FEV_1/FVC	0.81 (0.77, 0.84)	0.81 (0.77, 0.85)	0.81
Height (cm)	170 (164.45, 178)	171 (164.4, 178.0)	0.17
BMI (kg·m^{-2})	25.1 (22.30, 29.05)	24.80 (22.0, 28.8)	0.54
Sex			
Male	105 (39.18)	236 (43.14)	0.28
Female	163 (60.82)	311 (56.86)	
Smoking			
Non-smoker	145 (54.10)	288 (52.65)	**0.021**
Current smoker	61 (22.76)	128 (23.40)	
Past smoker	59 (22.01)	101 (18.46)	
Missing	3 (1.12)	30 (5.48)	

Data are presented as median (interquartile range) or n (%), unless otherwise stated. FVC: forced vital capacity; FEV_1: forced expiratory volume in 1 s; BMI: body mass index. Bold font indicates statistical significance.

longitudinal association of DNA-M at age 7 years with lung function at age 15 years, and DNA-M at age 15 years with lung function at age 24 years was examined, controlling the effects of confounders except for two covariates, recurrent chest infection and paracetamol use, which were unavailable.

Gene expression analysis
To assess potential biological relevance of the identified CpGs in model-1 and -2, we examined the association of DNA-M at those CpGs with the expression of their corresponding genes in blood. Linear regressions were applied to two datasets, DNA-M at age 18 years with gene expression at 26 years (longitudinal associations) and DNA-M at age 26 years with gene expression at the same age (cross-sectional associations).

Results
Results of longitudinal association analysis in IOWBC
In total, 332 (172 females) participants were included who had the complete (both DNA-M and lung function) data in at least one of the two periods (figure 1). The analysed sub-samples at age 18 (n=315) and 26 years (n=268) were not statistically different from the enrolled sample with lung function (18 years, n=839; 26 years, n=547) for FVC, FEV_1, and FEV_1/FVC at the corresponding ages, except FEV_1 at age 18 years which was higher in the subsample (table 1). Using *ttScreening*, in total, 194, 207, and 149 CpGs with DNA-M at ages 10 and 18 years were identified as associated with FVC, FEV_1, and FEV_1/FVC at 18 and 26 years, respectively. These CpGs were then included in subsequent analyses (figure 2). In model-1 (main effects of DNA-M), DNA-M at 14 CpGs (three CpGs with FVC, six with FEV_1, and five with FEV_1/FVC at earlier ages) was associated with lung function at later ages longitudinally ($p_{FDR}<0.05$, table S1) after adjusting the confounders. In model-2 (interaction effects of DNA-M×sex), DNA-M at 26
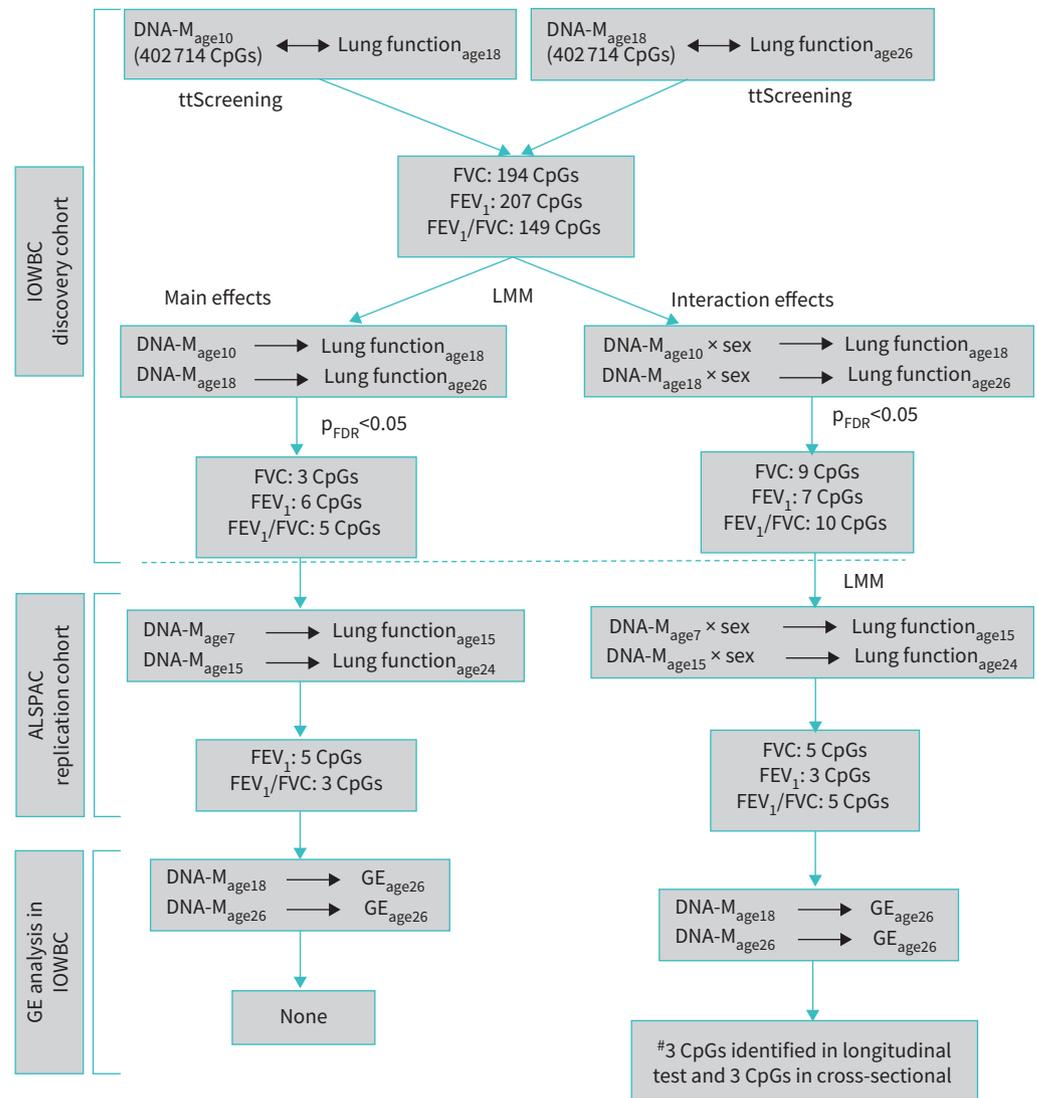
FIGURE 2 Flow chart of statistical analyses and the number of cytosine-phosphate-guanine dinucleotide sites (CpGs) after each step. IOWBC: Isle of Wight birth cohort; ALSPAC: Avon Longitudinal Study of Parents and Children; GE: gene expression; DNA-M: DNA methylation; DNA-M$_{ageX}$: DNA-M at age X years; Lung function$_{ageX}$: lung function at age X years; GE$_{age26}$: gene expression at age 26 years; FDR: false discovery rate; FVC: forced vital capacity; FEV$_1$: forced expiratory volume in 1 s; LMM: linear mixed models. #: 2 CpGs are common between the longitudinal and cross-sectional analysis of DNA-M with GE.

CpGs showed sex-specific associations with lung function (nine CpGs with FVC, seven with FEV$_1$, and 10 with FEV$_1$/FVC; p$_{FDR}$<0.05) (table S2, figure 2). The cg14083603 in *WNT10A* was identified by both model-1 and model-2.

Replication in ALSPAC cohort

In total, 1342 participants (610 males) in ALSPAC had complete data (DNA-M and lung function) in at least one period. Among the 14 CpGs identified in model-1 in IOWBC, five for FEV$_1$ and three for FEV$_1$/FVC showed consistent directions of associations for the main effects, of which the effect of cg16582803 (*WNT10A*) was statistically significant (p=0.034) for FEV$_1$ (table 2). Among the IOWBC-ALSPAC consistent eight CpGs, higher DNA-M at five CpGs (three associated with FEV$_1$ and two with FEV$_1$/FVC), mapped to *ANKRD9*, *WNT10A*, *ZNF727*, *NRN1*, and *DNAJB6*, at earlier ages were associated with lower lung function at later ages. While at the remaining three CpGs, mapped to *HINFP*, *EFNA2* and *C16orf87*, higher DNA-M at earlier ages was associated with higher lung function at later ages (table 2). In model 2, 13 of the 26 CpGs (five CpGs associated with FVC, three with FEV$_1$, and five with FEV$_1$/

TABLE 2 DNA methylation at cytosine-phosphate-guanine dinucleotide sites (CpGs) at earlier ages that showed consistent direction of associations with lung function at later age between the Isle of Wight birth cohort (IOWBC) and the Avon Longitudinal Study of Parents and Children (ALSPAC)

Lung function measure	CpGs	Chromosome number	Gene	Location[#]	IOWBC			ALSPAC	
					Coefficient	p_{RAW}	p_{FDR}	Coefficient	p-value
FEV_1									
	cg10729557	14	*ANKRD9*	Intergenic	−0.18	0.0008	0.035	−2.02	0.22
	cg16582803	**2**	***WNT10A***	**Intergenic**	**−0.16**	**0.0002**	**0.028**	**−0.59**	**0.034**
	cg17315331	11	*HINFP*	TSS200	0.06	0.0003	0.028	1.60	0.84
	cg26174454	19	*EFNA2*	Intergenic	0.23	0.0007	0.035	0.44	0.15
	cg27599129	7	*ZNF727*	Intergenic	−0.23	0.001	0.035	−0.06	0.86
FEV_1/FVC									
	cg18760835	6	*NRN1*	Intergenic	−0.01	0.0004	0.041	−0.22	0.33
	cg21240861	7	*DNAJB6*	TSS200	−0.004	0.0014	0.041	−0.05	0.95
	cg27601198	16	*C16orf87*	Intergenic	0.01	0.0011	0.041	0.08	0.28

Coefficients of IOWBC-ALSPAC consistent CpGs for the association of DNA methylation at earlier age with lung function at a later age. The CpGs that showed the same direction of associations and were significant at 0.05 level are shown in bold font. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; p_{FDR}: p-value after multiple testing correction by controlling false discovery rate; p_{RAW}: p-value without multiple testing adjustment. [#]: genes located at intergenic locations were not found in the Illumina annotation file and were identified using the online tool SNIPPER.

FVC) showed consistent directions of associations for interaction effects with those in IOWBC (table 3) and among these 13 CpGs, cg14083603 (*ZGPAT*) was statistically significant (p=0.0183). For sex-specific analysis in model-2, in males higher DNA-M at eight CpGs at early ages was associated with lower lung function at later ages, while in females higher DNA-M at those CpGs was associated with higher lung function. At the remaining five CpGs, higher DNA-M was associated with higher lung function in males, while in females it was associated with lower lung function (table 3).

Results of gene expression analysis

In a longitudinal assessment of DNA-M at age 18 years with gene expression at 26 years (36 males and 72 females), five identified CpGs in model-1 and 11 in model-2 had the corresponding gene expression data. In longitudinal assessment none of the five CpGs in model-1 were associated with the relevant gene expression. In model-2, amongst the 11 CpGs, DNA-M at cg01376079 (*SSH3*), cg07557690 (*TGFBR3*), and cg15981851 (*AGAP1*) at age 18 years showed significant association with gene expression at age 26 years (table 4). In cross-sectional association of DNA-M at age 26 years with gene expression at 26 years (54 males and 85 females), one CpG in model-1 had corresponding gene expression data but showed no association. In model-2, eight identified CpGs had gene expression data and DNA-M at three CpGs, cg01376079 (*SSH3*), cg07557690 (*TGFBR3*), and cg19736286 (*MSH6*), were shown to be cross-sectionally associated with gene expression, with cg01376079 and cg07557690 also being associated with expression of the corresponding gene in the longitudinal assessment. In both longitudinal and cross-sectional assessment, consistent directions of DNA-M and gene expression associations were found for cg01376079 and cg07557690; higher methylation at cg01376079 was associated with lower expression of *SSH3*, while higher methylation at cg07557690 was associated with higher expression of *TGFBR3* (table 4).

Results of the DMRs analysis

DMR analyses focused on detecting regions showing differential methylation associated with lung function parameters. To potentially improve the power, *via ttScreening*, in males 486, 518, and 461 CpGs and in females 419, 559, and 842 CpGs were selected based on their association with FVC, FEV_1, and FEV_1/FVC, respectively, and were included in the DMR analyses. Using repeated measures of DNA-M and lung function, 17, 24, and 16 statistically significant DMRs in males and 57, 66, and 47 DMRs in females were identified for FVC, FEV_1, and FEV_1/FVC, respectively (p_{FDR}<0.05). The DMRs containing ⩾2 CpGs are presented in table 5 and the complete results in table S3. In total, 132 and 382 CpGs were in the 57 and 170 identified DMRs in males and females, respectively. Four genes were common between the mapped genes of the individually identified CpGs and those of DMRs, namely *TGFBR3*, *WNT10A*, *LY6H*, and *GMIP*.

Discussion

We examined the longitudinal association of genome-wide DNA-M at ages 10 and 18 years with lung function at 18 and 26 years, respectively, using repeated measures from pre-adolescence to post-adolescence

TABLE 3 DNA methylation at cytosine-phosphate-guanine dinucleotide sites (CpGs) at earlier age that showed consistent sex-specific association with lung function at later age between the Isle of Wight birth cohort (IOWBC) and the Avon Longitudinal Study of Parents and Children (ALSPAC)

Lung function measure	CpGs	Sex	Chromosome number	Gene name	Location[#]	IOWBC			ALSPAC	
						Coefficient	p_{RAW}	p_{FDR}	Coefficient	p-value
FVC										
	cg01376079		11	SSH3	TSS1500	0.09	0.38		0.22	0.49
	cg01376079×sex	Male				−0.41	0.0019	0.042	−0.71	0.12
	cg07230380		10	SCD	TSS1500	−0.02	0.25		−4.77	0.62
	cg07230380×sex	Male				0.05	0.0012	0.042	16.21	0.23
	cg07557690		1	TGFBR3	TSS1500	0.30	0.0014		0.21	0.56
	cg07557690×sex	Male				−0.41	0.0017	0.042	−0.40	0.46
	cg14083603		20	*ZGPAT*	Body	−0.49	<0.0001		−0.55	0.76
	cg14083603×sex	**Male**				**0.69**	**0.0007**	**0.042**	**6.22**	**0.018**
	cg23026420		11	PPP2R1B	TSS200	0.17	0.0059		2.96	0.66
	cg23026420×sex	Male				−0.29	0.0019	0.041	−11.86	0.19
FEV₁										
	cg15981851		2	AGAP1	Body	0.07	0.27		0.10	0.77
	cg15981851×sex	Male				−0.36	0.0002	0.015	−0.057	0.91
	cg16582803		2	WNT10A	Intergenic	−0.09	0.77		−0.48	0.21
	cg16582803×sex	Male				−0.29	0.0009	0.032	−0.23	0.67
	cg19736286		2	MSH6	TSS200	0.09	0.41		0.23	0.90
	cg19736286×sex	Male				−0.56	0.0011	0.032	−1.36	0.59
FEV₁/FVC										
	cg02397934		6	H2BC13	Intergenic	−0.01	0.035		−0.094	0.46
	cg02397934×sex	Male				0.03	0.0004	0.024	0.21	0.25
	cg08650125		8	LY6H	Intergenic	0.04	0.0014		0.01	0.73
	cg08650125×sex	Male				−0.06	0.0005	0.0236	−0.029	0.50
	cg09059988		1	HORMAD1	1st Exon; 5′UTR	0.01	0.36		−0.012	0.89
	cg09059988×sex	Male				−0.06	0.0008	0.025	−0.03	0.79
	cg20038169		19	GMIP	1st Exon; 5′UTR	−0.02	<0.0001		−0.59	0.66
	cg20038169×sex	Male				0.02	0.0007	0.025	1.823	0.34
	cg23370466		5	TRIM41	Intergenic	−0.02	0.0045		−0.032	0.43
	cg23370466×sex	Male				0.04	0.0002	0.024	0.040	0.49

Coefficients of IOWBC-ALSPAC consistent CpGs for the sex-specific association of DNA methylation at earlier age with lung function at a later age. The CpGs showed the same direction of associations and were significant at 0.05 level are shown in bold font. FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; p_{FDR}: p-value after multiple testing correction by controlling false discovery rate; p_{RAW}: p-value without multiple testing adjustment; 5′UTR: 5′ untranslated region. [#]: genes located at intergenic locations were not found in the Illumina annotation file and were identified using the online tool SNIPPER.

period at both individual sites and genomic regions. DNA-M at eight CpGs and 13 CpGs at an earlier age was shown to be associated with lung function at a later age for main effects and sex-specific effects, respectively, in the IOWBC, with consistent findings in ALSPAC. Among IOWBC-ALSPAC consistent CpGs, cg16582803 (*WNT10A*) and cg14083603 (*ZGPAT*) were replicated in ALSPAC in terms of direction of associations and were statistical significance for main effect and interaction effects on lung function, respectively. DNA-M at cg01376079 (*SSH3*) and cg07557690 (*TGFBR3*) was associated with gene expression and invariant to longitudinal or cross-sectional assessment. In total, 57 and 170 DMRs at earlier age in relation to lung function at later age were identified in males and females, respectively.

In our study, at a certain proportion of CpGs, the longitudinal associations were shown to be sex-specific. One possible explanation for such an observation might be due to sex-specific changes of DNA-M over time as we have previously observed [20]. Other studies also suggested significant sex differences in patterns of blood-based DNA-M at the genome scale [31]. Although the current study focused on longitudinal association of DNA-M and lung function, the observation on sex-specificity is consistent with our previous findings [23, 32]. In previous studies, the associations of changes in DNA-M with lung function changes [23] and DNA-M with lung function trajectories were found to be different between males and females [32]. Our further analyses indicated that such sex-specificity was time-invariant.

CpGs	Chromosome number	Gene name	Location	DNA-M at age 18 years with gene expression at 26 years		DNA-M at age 26 years with gene expression at 26 years	
				Coefficient	p-value	Coefficient	p-value
Model 1: CpGs identified for main effects							
FEV$_1$							
cg10729557	14	*ANKRD9*	Intergenic	0.10	0.73	−0.053	0.77
cg17315331	11	*HINFP*	TSS200	0.043	0.10		
FEV$_1$/FVC							
cg18760835	6	*NRN1*	Intergenic	−0.072	0.21		
cg21240861	7	*DNAJB6*	TSS200	−0.01	0.62		
cg27601198	16	*C16orf87*	Intergenic	0.021	0.77		
Model 2: CpGs identified for interactions effects							
FVC							
cg01376079	**11**	***SSH3***	**TSS1500**	**−0.17**	**0.011**	**−0.17**	**0.0033**
cg07230380	10	*SCD*	TSS1500	−0.004	0.88		
cg07557690	**1**	***TGFBR3***	**TSS1500**	**1.07**	**2.0×10^{-4}**	**0.10**	**2.7×10^{-7}**
cg14083603	20	*ZGPAT*	Body	−0.058	0.69	0.18	0.20
cg23026420	11	*PPP2R1B*	TSS200	0.001	0.99	0.081	0.36
FEV$_1$							
cg15981851	**2**	***AGAP1***	**Body**	**−0.63**	**0.028**	−0.23	0.29
cg16582803	2	*WNT10A*	Intergenic	−0.20	0.31		
cg19736286	**2**	***MSH6***	**TSS200**	−0.12	0.58	**−0.38**	**0.0165**
FEV$_1$/FVC							
cg09059988	1	*HORMAD1*	1st Exon; 5′UTR	−0.16	0.62	0.18	0.52
cg20038169	19	*GMIP*	1st Exon; 5′UTR	0.021	0.38		
cg23370466	5	*TRIM41*	Intergenic	0.011	0.84	−0.028	0.52

TABLE 4 Association of DNA methylation (DNA-M) with gene expression in the Isle of Wight birth cohort (IOWBC)

Coefficients of IOWBC-ALSPAC (Avon Longitudinal Study of Parents and Children) consistent cytosinephosphate-guanine dinucleotide sites (CpGs) that had available gene expression data for the association of DNA-M with gene expression both longitudinally and cross-sectionally. CpGs that showed significant association at 0.05 level are shown in bold font. FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; 5′UTR: 5′ untranslated region.

The mapped genes of replicated CpGs, such as cg16582803 on *WNT10A* and cg14083603 on *ZGPAT*, have plausible biological relevance to lung function and respiratory diseases. The Wnt/β-catenin pathway is centrally involved in lung development and several lung diseases [33, 34]. In particular, *WNT10A* plays an important role in pathogenesis of idiopathic pulmonary fibrosis (IPF) *via* transforming growth factor (TGF)-β activation [34]. Genetic variation in *ZGPAT* has been shown to be associated with lung function and also the risk of asthma and atopic dermatitis [35–37]. It has been suggested that DNA-M in *ZGPAT* has a causal effect on FEV$_1$, mediated by changes in the expression of *ZGPAT* [37].

Longitudinal association of DNA-M at CpGs/DMRs with lung function measures at a later age may provide insight into the pathogenesis of impaired lung function growth. The association of differential methylation at some of these CpGs with gene expression, such as cg15981851 (*AGAP1*) for time-lagged, cg19736286 (*MSH6*) for cross-sectional assessment, and cg01376079 (*SSH3*) and cg07557690 (*TGFBR3*) for both longitudinal and cross-sectional assessment suggests a functional relevance of these CpGs. cg01376079 (*SSH3*) and cg07557690 (*TGFBR3*) manifest stable effects of DNA-M on gene expression. All the CpGs associated with the gene expression are located at promoter regions, except for cg15981851 (*AGAP1*), which is in the gene body (table 4).

It is important to note the biological relevance of cg07557690, located in the promoter region of gene *TGFBR3* (TGF-β receptor type III). Among the identified CpGs showing associations with gene expression, the association of cg07557690 with expression of *TGFBR3* was the strongest in both effect size and statistical significance. Expression of *TGFBR3* is essential for optimal TGF-signalling during embryonic lung development [38]. TGF-β is also a key regulator of extracellular matrix composition and alveolar epithelial cell and fibroblast function in the lung. Prolonged alterations of TGF-β and its receptors result in compromised gas exchange and lung function, a feature of bronchopulmonary dysplasia, lung fibrosis and COPD [38, 39]. In addition, *TGFBR3* has been suggested to play key roles in the pathogenesis of asthma [40] and COPD susceptibility [39]. *TGFBR3* is also mapped within two lung function associated

TABLE 5 Differentially methylated regions (DMRs; containing ⩾2 cytosine-phosphate-guanine dinucleotide sites (CpGs)) of lung function at later age in relation to DNA methylation at earlier age identified by the DMRcate method

Molecular location of DMR (chromosome: start – end)	CpGs n	Stouffer	Annotated gene
Male			
FVC			
chr4: 2819770-2820479	4	0	SH3BP2
chr8: 11659832-11660733	3	0	FDFT1, RP11
chr3: 194014481-194014745	3	1.30×10^{-198}	CPN2
FEV$_1$			
chr4: 1004525-1004678	3	0	FGFRL1
chr6: 33084825-33085031	3	0	HLA-DPB2[#]
chr15: 30163660-30163825	3	2.00×10^{-203}	TJP1[#]
chr6: 2891973-2892150	3	9.07×10^{-66}	SERPINB9[#]
FEV$_1$/FVC			
chr6: 30038929-30039435	10	0	RNF39[#]
chr19: 1467008-1467032	3	0	APC2[#]
chr6: 33871907-33872861	3	4.70×10^{-188}	MIR1275[#]
Female			
FVC			
chr22: 30476089-30476525	5	7.80×10^{-280}	HORMAD2, CTA
chr6: 88757302-88757392	5	7.46×10^{-74}	SPACA1
chr1: 92352293-92352481	3	0	TGFBR3
chr10: 135191624-135192230	3	0	PAOX, AL360181.1-201
chr11: 2322500-2322808	3	0	TSPAN32, C11orf21
chr20: 25677290-25677582	3	0	ZNF337
chr5: 112824497-112824765	3	7.90×10^{-254}	MCC
chr17: 56744332-56744490	3	9.80×10^{-216}	RNU1
chr10: 34408530-34408654	3	3.50×10^{-139}	PARD3[#]
chr15: 99975310-99975470	3	1.30×10^{-107}	LRRC28[#]
FEV$_1$			
chr5: 1867978-1868693	5	2.75×10^{-84}	IRX4[#]
chr17: 45949743-45949878	4	0	SP6[#]
chr6: 32016257-32017229	4	0	TNXB
chr22: 30476089-30476525	4	6.50×10^{-182}	HORMAD2, CTA
chr11: 18433554-18433745	4	5.93×10^{-39}	LDHC
chr6: 110720918-110721349	4	8.70×10^{-26}	DDO[#]
chr7: 157512397-157513707	3	0	PTPRN2[#]
chr7: 150037890-150038898	3	3.60×10^{-219}	RARRES2
chr6: 32294470-32294577	3	1.30×10^{-190}	HNRNPA1P2
chr20: 62328084-62328427	3	7.10×10^{-181}	TNFRSF6B
chr10: 34408530-34408654	3	2.30×10^{-139}	PARD3[#]
chr15: 99975310-99975470	3	4.30×10^{-108}	LRRC28[#]
chr8: 37605517-37605783	3	6.66×10^{-95}	RP11-109A6.3
chr17: 154420-154671	3	1.37×10^{-88}	RPH3AL
FEV$_1$/FVC			
chr1: 2058230-2059086	3	0	PRKCZ[#]
chr17: 40936570-40937362	3	0	WNK4
chr20: 44829602-44829821	3	0	CDH22[#]
chr2: 113993052-113994035	3	3.10×10^{-156}	PAX8-AS1
chr7: 57471759-57472367	3	1.80×10^{-131}	MIR3147
chr1: 43814764-43815035	3	2.22×10^{-21}	MPL

DMRcate annotates to UCSC RefGene from the Illumina annotation file. FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity. [#]: genes were not found in the Illumina annotation file and were identified using the online tool SNIPPER.

DMRs in this study. Together these results suggest that cg07557690 has potential utility as a biomarker of lung function development. Future in-depth studies of cg07557690 and how it is related to lung function are warranted.

An important strength of this study is its longitudinal design in which DNA-M measurement always precedes the lung function measurement to avoid temporal ambiguity (reverse causation). With repeated measures, longitudinally designed studies potentially gain a higher power to detect change over time and to identify differences between individuals, compared with cross-sectional studies. Moreover, the inclusion of a validation cohort increased the testing power of the identified CpGs. In addition, CpGs showing agreement between the two cohorts have potential generalisability at least in Caucasians.

There are a few limitations to this study. The median value of FEV_1 at age 18 years, the proportion of males and females at age 18 years and smoking status at 26 years were different in the analysed samples to the overall study cohort. At age 26 years, lung function was available for fewer participants compared with age 18 years, leading to a smaller sample size in period-2. This study has Caucasian participants in both cohorts. Although we believe using a replication cohort with the same ethnicity as in the discovery cohort potentially improved the testing power, this design may limit the generalisability of the findings to other populations. In addition, while methylation of several CpGs was shown to be associated with relevant gene expression, this was in mixed cell populations from whole blood and it is not possible to assess cell-type specificity of the relationship, or the relevance to gene expression in the lung. Nevertheless, the identified CpGs have the potential to serve as candidate CpGs for lung function impairment prediction in future studies. Screening for such CpGs in early life may help to identify children at higher risk of reduced lung function at later ages.

References

1   Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet* 2015; 385: 899–909.
2   Vasquez MM, Zhou M, Hu C, *et al.* Low lung function in young adult life is associated with early mortality. *Am J Respir Crit Care Med* 2017; 195: 1399–1401.

3   Kohansal R, Martinez-Camblor P, Agustí A, *et al.* The natural history of chronic airflow obstruction revisited: an analysis of the Framingham Offspring Cohort. *Am J Respir Crit Care Med* 2009; 180: 3–10.

4   Knudson RJ, Lebowitz MD, Holberg CJ, *et al.* Changes in the normal maximal expiratory flow–volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725–734.

5   LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe* 2018; 14: 131–140.

6   Becklake MR, Kauffmann F. Gender differences in airway behaviour over the human life span. *Thorax* 1999; 54: 1119–1138.

7   Schultz ES, Gruzieva O, Bellander T, *et al.* Traffic-related air pollution and lung function in children at 8 years of age: a birth cohort study. *Am J Respir Crit Care Med* 2012; 186: 1286–1291.

8   Maher B. Personal genomes: the case of the missing heritability. *Nature* 2008; 456: 18–21.

9   de Jong K, Vonk JM, Timens W, *et al.* Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. *J Allergy Clin Immunol* 2015; 136: 1664–1672.

10   van der Plaat DA, de Jong K, de Vries M, *et al.* Occupational exposure to pesticides is associated with differential DNA methylation. *Occup Environ Med* 2018; 75: 427–435.

11   Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology* 2013; 38: 23–38.

12   Qiu W, Baccarelli A, Carey VJ, *et al.* Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med* 2012; 185: 373–381.

13   Lepeule J, Baccarelli A, Motta V, *et al.* Gene promoter methylation is associated with lung function in the elderly: the Normative Aging Study. *Epigenetics* 2012; 7: 261–269.

14   Busch R, Qiu W, Lasky-Su J, *et al.* Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. *Respir Res* 2016; 17: 143.

15   Lee MK, Hong Y, Kim SY, *et al.* Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. *Epigenomics* 2017; 9: 971–984.

16   Lange NE, Sordillo J, Tarantini L, *et al.* Alu and LINE-1 methylation and lung function in the normative ageing study. *BMJ Open* 2012; 2: e001231.

17   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.

18   Zhang H, Tong X, Holloway JW, *et al.* The interplay of DNA methylation over time with Th2 pathway genetic variants on asthma risk and temporal asthma transition. *Clin Epigenet* 2014; 6: 8.

19   Florath I, Butterbach K, Muller H, *et al.* Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 2014; 23: 1186–1201.

20   Han L, Zhang H, Kaushal A, *et al.* Changes in DNA methylation from pre- to post-adolescence are associated with pubertal exposures. *Clin Epigenetics* 2019; 11: 176.

21   Mulder RH, Neumann A, Cecil CAM, *et al.* Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence. *Hum Mol Genet* 2021; 30: 119–134.

22   Patil VK, Holloway JW, Zhang H, *et al.* Interaction of prenatal maternal smoking, interleukin 13 genetic variants and DNA methylation influencing airflow and airway reactivity. *Clin Epigenetics* 2013; 5: 22.

23   Sunny SK, Zhang H, Rezwan FI, *et al.* Changes of DNA methylation are associated with changes in lung function during adolescence. *Respir Res* 2020; 21: 80.

24   Pal S, Tyler JK. Epigenetics and aging. *Sci Adv* 2016; 2: e1600584.

25   Boyd A, Golding J, Macleod J, *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: 111–127.

26   Fraser A, Macdonald-Wallis C, Tilling K, *et al.* Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013; 42: 97–110.

27   Li X, Hawkins GA, Ampleford EJ, *et al.* Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J Allergy Clin Immunol* 2013; 132: 313–320.

28   Ray MA, Tong X, Lockett GA, *et al.* An efficient approach to screening epigenome-wide data. *Biomed Res Int* 2016; 2016: 2615348.

29   Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B* 1995; 57: 289–300.

30   Peters TJ, Buckley MJ, Statham AL, *et al. De novo* identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 2015; 8: 6.

31   Singmann P, Shem-Tov D, Wahl S, *et al.* Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin* 2015; 8: 43.

32   Sunny SK, Zhang H, Mzayek F, *et al.* Pre-adolescence DNA methylation is associated with lung function trajectories from pre-adolescence to adulthood. *Clin Epigenetics* 2021; 13: 5.

33   Weng T, Liu L. The role of pleiotrophin and β-catenin in fetal lung development. *Respir Res* 2010; 11: 80.

34   Oda K, Yatera K, Izumi H, *et al.* Profibrotic role of WNT10A *via* TGF-β signaling in idiopathic pulmonary fibrosis. *Respir Res* 2016; 17: 39.

35   Sun L-D, Xiao F-L, Li Y, *et al.* Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat Genet* 2011; 43: 690–694.

36   Wain LV, Shrine N, Artigas MS, *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017; 49: 416.

37   Jamieson E, Korologou-Linden R, Wootton RE, *et al.* Smoking, DNA methylation, and lung function: a mendelian randomization analysis to investigate causal pathways. *Am J Hum Genet* 2020; 106: 315–326.

38   Pozarska A, Niess G, Seeger W, *et al.* A role for the accessory type III transforming growth factor β receptor (Tgfbr3) in lung alveolarisation. *Eur Respir J* 2016; 48: Suppl. 60, PA4026.

39   Morty RE, Königshoff M, Eickelberg O. Transforming growth factor-β signaling across ages: from distorted lung development to chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2009; 6: 607–613.

40   Kim H-K, Jang T-W, Jung M-H, *et al.* Association between genetic variations of the transforming growth factor β receptor type III and asthma in a Korean population. *Exp Mol Med* 2010; 42: 420–427.