

RESEARCH

Open Access



# Changes of DNA methylation are associated with changes in lung function during adolescence

Shadia Khan Sunny<sup>1</sup>, Hongmei Zhang<sup>1\*</sup>, Faisal I. Rezwan<sup>2</sup>, Caroline L. Relton<sup>3</sup>, A. John Henderson<sup>4</sup>, Simon Kebede Merid<sup>5</sup>, Erik Melén<sup>5,6</sup>, Jenny Hallberg<sup>5,6</sup>, S. Hasan Arshad<sup>7,8</sup>, Susan Ewart<sup>9</sup> and John W. Holloway<sup>10</sup>

## Abstract

**Background:** Adolescence is a significant period for the gender-dependent development of lung function. Prior studies have shown that DNA methylation (DNA-M) is associated with lung function and DNA-M at some cytosine-phosphate-guanine dinucleotide sites (CpGs) changes over time. This study examined whether changes of DNA-M at lung-function-related CpGs are associated with changes in lung function during adolescence for each gender, and if so, the biological significance of the detected CpGs.

**Methods:** Genome-scale DNA-M was measured in peripheral blood samples at ages 10 ( $n = 330$ ) and 18 years ( $n = 476$ ) from the Isle of Wight (IOW) birth cohort in United Kingdom, using Illumina Infinium arrays (450 K and EPIC). Spirometry was conducted at both ages. A training and testing method was used to screen 402,714 CpGs for their potential associations with lung function. Linear regressions were applied to assess the association of changes in lung function with changes of DNA-M at those CpGs potentially related to lung function. Adolescence-related and personal and family-related confounders were included in the model. The analyses were stratified by gender. Multiple testing was adjusted by controlling false discovery rate of 0.05. Findings were further examined in two independent birth cohorts, the Avon Longitudinal Study of Children and Parents (ALSPAC) and the Children, Allergy, Milieu, Stockholm, Epidemiology (BAMSE) cohort. Pathway analyses were performed on genes to which the identified CpGs were mapped.

**Results:** For females, 42 CpGs showed statistically significant associations with change in FEV<sub>1</sub>/FVC, but none for change in FEV<sub>1</sub> or FVC. No CpGs were identified for males. In replication analyses, 16 and 21 of the 42 CpGs showed the same direction of associations among the females in the ALSPAC and BAMSE cohorts, respectively, with 11 CpGs overlapping across all the three cohorts. Through pathway analyses, significant biological processes were identified that have previously been related to lung function development.

**Conclusions:** The detected 11 CpGs in all three cohorts have the potential to serve as the candidate epigenetic markers for changes in lung function during adolescence in females.

**Keywords:** Lung function, DNA methylation, Genome-wide, Adolescence, IOW cohort, ALSPAC, BAMSE

\* Correspondence: [h Zhang6@memphis.edu](mailto:h Zhang6@memphis.edu)

<sup>1</sup>Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN 38152, USA  
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The period from childhood to adolescence is associated with rapid somatic growth and incorporates a range of gender-dependent physiological and behavioral changes, including hormonal, height and body mass index (BMI) changes, possible use of oral contraceptives, and possible initiation of nicotine use [1, 2]. This period is also significant for the development of lung function as it represents a phase of dramatic growth from childhood to adolescence to reach a maximal level of lung function in early adulthood [3–5]. Lung function growth is gender-dependent and such dependence is attributable to multiple biological determinants, including dimensional/anatomical (e.g., airway size, somatic growth, lung growth, adolescence growth spurts), immunological, and hormonal determinants such as different phases of the menstrual cycle and common hormonal and metabolic conditions [6–9].

DNA methylation (DNA-M), as a potential marker of past exposure or significant changes in life such as pubertal onset, is an epigenetic mechanism and has been shown to play an important role in human development and health. DNA-M refers to methylation of the 5' position of the cytosine base of cytosine-phosphate-guanine dinucleotide sites (CpG sites or CpGs) in the DNA [10]. It regulates gene function through the modulation of gene expression. Imboden et al. 2019 [11] and others have demonstrated that DNA-M in whole blood is associated with lung function [12–16], risk of asthma [17], and chronic obstructive pulmonary disease (COPD) [12, 13, 15, 16]. When assessing the association of DNA-M with lung function, most previous studies have been cross-sectional with both lung function and DNA-M measured at single time points [12–16], although DNA-M at some CpGs changes over time [18–22]. In our recent genome-wide study, we identified more than 10 K CpGs where DNA-M significantly changes over the adolescence period, and at some CpGs, such changes were gender-dependent [23].

To our knowledge, at CpGs which are potentially associated with lung function parameters such as forced expiratory volume in one second (FEV<sub>1</sub>) and forced vital capacity (FVC), no studies have examined whether and how changes in DNA-M at those CpGs are associated with changes in lung function during adolescence. Such an investigation will improve our understanding of epigenetic mechanisms in lung function development. In addition, DNA-M changes at CpGs shown to be associated with changes in lung function have the potential to predict future lung function changes, which, in the long run, may lead to strategies for the prevention of pulmonary disease. Taken together, we hypothesized that during adolescence, changes of DNA-M at some CpGs are associated with changes in lung function. Given that changes

during adolescence are gender-dependent, we examined this hypothesis separately in males and females. The study was carried out in a birth cohort located on the Isle of Wight (IOW) in the United Kingdom. To assess generalizability, the findings were further examined in two independent birth cohorts, Avon Longitudinal Study of Children and Parents Cohort (ALSPAC) in the United Kingdom and Children, Allergy, Milieu, Stockholm, Epidemiology (BAMSE) in Sweden.

## Methods

### Discovery cohort - IOW cohort

#### Study participants

The IOW cohort is a population-based birth cohort and was established in 1989 on the IOW, United Kingdom. The study was approved by the IOW Local Research Ethics Committee at recruitment initial assessments and further assessments were approved by the National Research Ethics Service, Committee South Central – Southampton B (06/Q1701/34). Informed written consent was obtained from participants or their parents before participating. The study enrolled 1456 eligible children of 1536 born between January 1989 and February 1990 (after exclusion of adoptions, infant deaths, and denial). Details of the birth cohort of 1989 have been described elsewhere [24]. Longitudinal monitoring of diseases and assessments of environmental exposures in this cohort was conducted at birth, and ages 1, 2, 4, 10, 18, and 26 years. In the present study, we focused on data collected at ages 10 ( $n = 1373$ ) and 18 ( $n = 1313$ ) years. In total 320 and 453 participants had both DNA-M and lung function data available at ages 10 and 18 years, respectively, including 301 participants that had data at both time points.

#### Lung function

Spirometric measurements, specifically, FVC and FEV<sub>1</sub> at ages 10 ( $n = 980$ ) and 18 ( $n = 838$ ) years were conducted using a Koko spirometer and software with a portable desktop device (both PDS Instrumentation, Louisville, KY, USA) and the ratio of FEV<sub>1</sub> over FVC (FEV<sub>1</sub>/FVC) was calculated. Spirometry was conducted and evaluated according to the American Thoracic Society (ATS) guidelines [25, 26]. Participants were required to be free of respiratory infection and had not taken oral steroids for two weeks. In addition, participants were instructed to abstain from any  $\beta$ -agonist medication for six hours and caffeine intake for at least 4 h.

#### Measuring DNA methylation (DNA-M)

Peripheral blood samples collected at ages 10 ( $n = 330$ ) and 18 ( $n = 476$ ) years from randomly selected subjects were used for DNA extraction via a standard salting out

procedure [27]. DNA concentration was estimated by Qubit quantitation. For each sample, one microgram DNA was bisulfite-treated for cytosine to thymine conversion using the EZ 96-DNA methylation kit (Zymo Research, Irvine, CA, USA), following the manufacturer's protocol. DNA-M was measured using HumanMethylation450K or HumanMethylationEPIC BeadChips (Illumina, Inc., San Diego, CA, USA). Arrays were processed using a standard protocol as described elsewhere [28], with multiple identical control samples assigned to each bisulfite conversion batch to assess assay variability. DNA samples were randomly distributed on microarrays to control against batch effects. Intensities of methylated and unmethylated sites were measured.

### Preprocessing

Probes not reaching a detection  $p$ -value of  $10^{-16}$  in at least 95% of samples were excluded. CpGs on sex chromosomes were also excluded to avoid potential bias in DNA-M as there are the parent of origin differences in methylation of paternally and maternally inherited X chromosomes [29]. DNA-M data were pre-processed using the "CPACOR" pipeline for data from both platforms [30]. DNA-M intensities were quantile normalized using the R computing package, *minfi* [31]. DNA-M  $\beta$  values for each CpG was calculated as a ratio of methylated (M) over the sum of methylated and unmethylated (U) probes ( $\beta = M/[c + M + U]$ ) interpreted as the percentage of methylation [32], where  $c$  is used as a constant to prevent zero in the denominator. Principal components (PCs) inferred based on control probes were used to represent latent variables due to chip-to-chip and technical (batch) variation. Since DNA-M data were from two different platforms (450 K and EPIC), we determined the PCs based on DNA-M at shared control probes between the two platforms. The 450 K BeadChips contained 220 control probes and the EPIC BeadChips contained 204 control probes, of which 195 overlapped between the two platforms. These 195 shared probes were then used to calculate the control probe PCs, top 15 of which were used to represent latent batch factors [30].

After pre-processing, a total of 473,864 and 847,155 CpGs were available in the 450K and EPIC methylation array data, respectively, and 439,635 overlapping CpGs were identified between the two platforms. CpGs with a single nucleotide polymorphisms (SNP) overlapping the detection probe with minor allele frequency  $\geq 0.7\%$  in Caucasians (corresponding to at least 10 subjects in the IOW cohort with  $n = 1456$ ) within 10 base pairs of the targeted CpGs were excluded due to potential bias that those SNPs brought to the measurement of DNA-M. After excluding probe SNPs, 402,714 CpGs were included in the statistical analyses.

### Confounders

Variables potentially associated with lung function change in addition to DNA-M change in adolescents are considered to be confounders, including changes in height and BMI, age of puberty onset, smoking status, socioeconomic status (SES), exposure to pets, exposure to air pollution, education status, farm exposure, paracetamol (acetaminophen) use, and non-steroidal anti-inflammatory drugs (NSAIDs) use [33–36].

Gender information was collected by questionnaire at each follow-up. Height was measured at 10 and 18 years of age before spirometric assessment. BMI was calculated from height and weight at age 10 and 18 years. Then changes of the height and BMI were calculated from age 10 to 18 years. The minimum age of puberty onset was estimated based on the following questions about the age of initiation of different pubertal changes: growth spurt of male or female, body hair growth of male or female, skin changes of male or female, deepening voice of male, facial hair of male, breast development of female, and initiation of menstruation of female. Smoking status was defined by the questions of current and past personal smoking status at age 18 years. A composite "SES-cluster" variable that accounts for SES broadly defined was used [37]. In order to correctly classify them, family SES were clustered using: (a) British socioeconomic classes (1-6) derived from parental occupation reported at birth; (b) number of children in the index child's bedroom (collected at age 4 years); and (c) family income at age 10 years [37]. This composite variable captures the family social class across the entire study period. Information on exposure to cats, dogs, and other animals was collected at both ages 10 and 18 years via questionnaire. Information on whether the subjects are still in education (yes/no), farm exposure (yes/no), how often health is affected by exposing to air pollution (never/ every day/ once a month/ once a week/ once a year), paracetamol use (frequency of taking paracetamol in a month) and use of NSAIDs (frequency of taking NSAIDs in a month) were collected by questionnaire at age 18 years.

### Replication cohort – the ALSPAC cohort

The Avon Longitudinal Study of Children and Parents (ALSPAC) is a population-based birth cohort study established in 1991 in Avon, United Kingdom, approximately 75 miles from the IOW. Details of the cohort were described elsewhere [38, 39]. Women residing in the South West of England who were pregnant and expecting to deliver between April 1, 1991 and December 31, 1992 were eligible to be recruited. In total, 14,541 pregnant women were eligible for the study, of those 13,761 were included with 10,321 providing DNA from blood samples. Participants were given questionnaires to

gauge information regarding the mother. Written informed consent was obtained for all ALSPAC participants. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Information on environment, lifestyle, and health of the child and family was collected through annual questionnaires since the child's birth. From age 7 years, all participants were invited to an annual research clinic, and thus exposure and other demographic data were available annually from 7 to 17 years. The follow-up cohort was composed of 13,988 children including multiple children from one family. In the replication study, we focused on ages 7 to 8 (7/8) and 15 years. Spirometry (Vitalograph 2120; Vitalograph, Maids Moreton, United Kingdom) was performed at 8 and 15 years of age according to ATS standards [26, 36], the same method as that applied in the IOW cohort. Please note that 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/>).

DNA-M in peripheral blood was assessed using the Infinium HumanMethylation450K BeadChip. The procedure for DNA sample preparation was comparable to that applied in the IOW cohort. DNA-M data of children at ages 7 ( $n = 966$ ) and 15 ( $n = 966$ ) years were available (twin participants were excluded). The pre-processing of DNA-M was performed by adjusting the batch effect, excluding CpGs with detection  $p$ -value  $\geq 0.01$ , and excluding samples that were flagged a sex-mismatch based on X-chromosome methylation [40]. CpGs on sex chromosomes were not included in the analyses. Only fully characterized subjects with DNA-M and lung function at both ages (7/8 years and 15 years) were included in the replication study, which resulted in 691 paired samples.

#### Replication cohort – the BAMSE cohort

The Swedish Children, Allergy, Milieu, Stockholm, Epidemiology (BAMSE) cohort is an unselected, population-based cohort study of children from Stockholm, Sweden. During 1994–1996, a total of 4089 children were recruited at birth from four municipalities in Stockholm County and followed during childhood. The Regional Ethical Review Board, Karolinska Institute in Stockholm, Sweden, approved the baseline study with its follow-up. A thorough description of the cohort, inclusion and enrollment criteria, and procedure of data collection have been described elsewhere [41]. Follow-up questionnaires focusing on the children's respiratory health, allergic diseases and on various exposure factors were collected at 1, 2, 4, 8, and 16 years old after obtaining informed consent from the parents of all participating children. At ages 8 ( $n = 1838$ ) and 16 ( $n = 2063$ )

years, lung function testing was conducted [42]. Maximal expiratory flow volume (MEFV) tests were performed at 8 and 16 years of age using the 2200 Pulmonary Function Laboratory (Sensormedics, Anaheim, CA, USA) and Jaeger MasterScreen-IOS system (Carefusion Technologies, San Diego, CA), respectively [42, 43]. All children performed several MEFV measurements and the maximal values of FVC and FEV<sub>1</sub> were extracted for the analyses. The MEFV curve that passed visual quality inspection, and the two highest FEV<sub>1</sub> and FVC readings were reproducible according to ATS/ European Respiratory Society criteria [26]. FEV<sub>1</sub>/FVC ratios were calculated. Height was measured before lung function testing for each participant.

DNA extracted from peripheral blood samples at ages 8 and 16 years of follow up was used to measure DNA-M [44]. For each sample, 500 ng DNA underwent bisulfite treatment for cytosine to thymine conversion using the EZ 96-DNA methylation kit (Shallow; Zymo Research Corporation, Irvine, CA, USA). DNA-M was assessed using the Illumina Infinium HumanMethylation450K BeadChip (Illumina, Inc.). After data pre-processing and quality control following the standard criteria [45], DNA-M data of 464 and 267 participants were available at ages 8 and 16 years, respectively.

#### Statistical analyses in the IOW cohort

To evaluate whether subjects included in the study reasonably represented those in the complete study cohort, we focused on the assessment of lung function at each age for both genders together and for each gender separately. To compare with the complete cohort, for continuous variables, including lung function, height, and BMI, one-sample t-tests were applied, and for categorical variables, including gender and smoking status, one-sample proportion tests were implemented.

Due to heteroscedasticity of DNA-M measured by  $\beta$  values [32],  $\beta$  values were logit-transformed to M values using  $\log_2(\beta \text{ value}/(1 - \beta \text{ value}))$  [46]. Lung function measurements (FVC, FEV<sub>1</sub>, and FEV<sub>1</sub>/FVC) at each age were adjusted by height and gender by regressing lung functions on these two variables using SAS 9.4 procedure PROC GLM (SAS, Gary, N.C., USA).

In this study, we focused on lung-function-related CpGs. To achieve this goal, we first excluded CpGs which were not potentially associated with lung function. A screening package, *ttScreening* (training and testing screening, R package 3.3.2 version) [47, 48] was applied for this purpose. This method utilizes training and testing data in robust linear regressions with surrogate variables included in the regressions to adjust for unknown effects. For each lung function measure (FVC, FEV<sub>1</sub>, and FEV<sub>1</sub>/FVC), we performed the screening for each gender (males and females) at each age (10 and 18 years).

DNA-M measured in peripheral blood might be potentially influenced by cellular composition of blood samples, different batches for DNA-M measurement, and technical variation in the process of analyzing DNA samples. To adjust the impact of these factors on DNA-M, linear regressions were applied with DNA-M as the outcome variable, and cell type proportions, batch information, and top 15 principal components of the control probes were included as independent variables for age 10 and 18 years. Cell type proportions (CD4+ T cells, CD8+ T cells, natural killer cells, B cells, monocytes, neutrophils, and eosinophils) were inferred from methylation data for each sample using the R computing package *minfi* [31, 49]. After estimating the adjusted DNA-M for each age (10 and 18 years), differences in the adjusted DNA-M between ages 10 and 18 were calculated (DNA-M at age 18 – DNA-M at age 10) and included in subsequent analyses.

Finally, to explore whether the changes of DNA-M over the adolescence period from ages 10 to 18 years were associated with the change in lung function, a linear regression model was fitted for each lung function measure, stratified by gender. Changes in height- and gender-adjusted lung function from 10 to 18 years of age were treated as the outcome variable, and changes of the adjusted DNA-M at each CpG that passed screening were used as an independent variable and potential confounders as described above were included in the model. In all analyses, *p*-values were considered significant at a level of 0.05.

### Replication analyses

CpGs identified in the IOW cohort were further tested in both the ALSPAC and BAMSE cohorts. Comparable analytical methods were applied except for the availability of some covariates. In ALSPAC, pet exposure, exposure to pollution, paracetamol use, and non-steroidal anti-inflammatory drugs use were not available, and in BAMSE, minimum age of puberty onset, pet exposure, exposure to pollution, and paracetamol use were not included in the final model.

### Pathway analysis

For CpGs that showed consistent directions of association in the ALSPAC and BAMSE cohorts, the nearest gene was identified based on Illumina array manifest file and SNIPPER (<https://csg.sph.umich.edu/boehnke/snipper/>) version 1.2. Bioinformatic assessment of the genes was conducted using the online bioinformatics tool ToppFun, available in the ToppGene Suite [50]. Multiple testing was adjusted by controlling the false discovery rate (FDR) of 0.05.

## Results

### Results from the IOW cohort

In total, 320 participants at age 10 years and 453 at age 18 years were included in the analyses for screening in the IOW cohort with available DNA-M and lung function data (Table 1). The mean values of FVC, FEV<sub>1</sub>, FEV<sub>1</sub>/FVC, height, and BMI for subjects in the present study were not significantly different from participants of the whole cohort with lung function at ages 10 (*n* = 980) and 18 (*n* = 838) years (Table 1) and for males and females separately with lung function at ages 10 (males = 488, females = 492) and 18 (males = 395, females = 443) (Table 2). Proportions of subjects who smoke or formerly smoked were also comparable to those in the complete cohort (Tables 1 and 2). One exception is that at age 10 years, a higher proportion of males were included in the present study compared to the whole cohort (Table 1).

To identify candidate CpGs potentially associated with lung function at ages 10 and 18 years, we applied *ttScreening* to the 402,714 CpGs in each gender. Three lung function parameters were considered in the screening process, FVC, FEV<sub>1</sub>, and FEV<sub>1</sub>/FVC. At age 10 years, across all the three lung function parameters, in total 361 distinct CpGs passed screening (157 CpGs for males and 204 CpGs for females), and at age 18 years, 530 distinct CpGs passed screening (274 CpGs for males and 256 CpGs for females). The break-down of the numbers of CpGs that passed screening for each lung function parameter was given in Fig. 1. Combining the CpGs that passed the screening at either time point for each gender and each lung function measurement, in males 431 distinct CpGs (178 CpGs for FVC, 151 for FEV<sub>1</sub>, and 122 for FEV<sub>1</sub>/FVC) and in females 460 distinct CpGs (174 CpGs for FVC, 158 for FEV<sub>1</sub>, and 161 FEV<sub>1</sub>/FVC) were included in the subsequent analyses. There were no common CpGs between the 431 and 460 CpGs identified in males and females.

Linear regression models were applied to assess the association of change in DNA-M at each of the screened CpG with the change of each lung function parameter (FVC, FEV<sub>1</sub>, and FEV<sub>1</sub>/FVC) for males (*n* = 169) and females (*n* = 132) separately. For females, after adjusting for multiple testing by controlling the FDR of 0.05, 42 CpGs showed statistically significant association with FEV<sub>1</sub>/FVC change, but for FEV<sub>1</sub> and FVC, we did not identify any statistically significant CpGs. At these 42 CpGs, a larger increase in DNA-M was associated with a larger decrease in FEV<sub>1</sub>/FVC in females. From childhood to adolescence, generally FEV<sub>1</sub>/FVC is constant or falls linearly with age because FVC has a proportionately greater increase than FEV<sub>1</sub> [51], which supports our findings. For males, no CpG survived multiple testing for any of the three lung function parameters. The 42

**Table 1** Characteristics of subjects with available methylation data with their lung function of the IOW cohort

IOW cohort						
	Sub cohort at age 10: Participants with lung function Mean $\pm$ SD	Study sample at age 10: Participants with lung function and DNA-M Mean $\pm$ SD	<i>P</i> -values	Sub cohort at age 18: Participants with lung function Mean $\pm$ SD	Study sample at age 18: Participants with lung function and DNA-M Mean $\pm$ SD	<i>P</i> -values
Factors	<b>n = 980</b>	<b>n = 320</b>		<b>n = 838</b>	<b>453</b>	
<i>Lung Function parameters</i>						
<i>FEV<sub>1</sub> (L)</i>	2.03 $\pm$ 0.30	2.04 $\pm$ 0.30	0.456	4.01 $\pm$ 0.78	4.05 $\pm$ 0.76	0.226
<i>FVC (L)</i>	2.30 $\pm$ 0.34	2.30 $\pm$ 0.34	0.694	4.61 $\pm$ 0.93	4.66 $\pm$ 0.91	0.283
<i>FEV<sub>1</sub> /FVC</i>	0.89 $\pm$ 0.06	0.89 $\pm$ 0.05	0.162	0.87 $\pm$ 0.07	0.87 $\pm$ 0.07	0.269
<i>Height (cm)</i>	138.92 $\pm$ 6.18	138.06 $\pm$ 6.22	0.685	170.88 $\pm$ 9.17	170.92 $\pm$ 9.08	0.925
<i>BMI</i>	18.16 $\pm$ 3.01	18.08 $\pm$ 2.99	0.649	23.21 $\pm$ 4.33	23.30 $\pm$ 4.26	0.639
	<b>n (%)</b>	<b>n (%)</b>		<b>n (%)</b>	<b>n (%)</b>	
<i>Gender</i>						
<i>Male</i>	488 (49.8)	183 (57.19)	<b>0.010</b>	395 (47.14)	212 (46.80)	0.932
<i>Female</i>	492 (50.2)	137 (42.81)		443 (52.86)	241 (53.20)	
<i>Smoking<sup>a</sup></i>						
<i>Non- smoker</i>	–	–	–	462 (55.13)	248 (54.75)	0.913
<i>Current smoker</i>	–	–		204 (24.34)	107 (23.62)	
<i>Past smoker</i>	–	–		159 (18.97)	93 (20.53)	
<i>Missing</i>	–	–		13 (1.55)	5 (1.10)	

<sup>a</sup>Active smoking at age 10 years in the IOW Cohort was not identified

CpGs identified in females in the IOW cohort were further tested in the ALSPAC and BAMSE cohorts.

### Results from the ALSPAC cohort

In total, 345 female ( $n = 935$ ) participants in the ALSPAC had FEV<sub>1</sub>/FVC measurements and DNA-M measurements at both 7/8 years and 15 years old. Of the 42 CpGs examined, DNA-M changes at 16 CpGs (Table 3) showed consistent associations with FEV<sub>1</sub>/FVC changes (in terms of regression coefficients) compared to those observed in the IOW cohort (Fig. 2, Table 3), although not statistically significant at the 0.05 level. These 16 CpGs were noted as IOW-ALSPAC consistent CpGs. The complete results of this analysis were included in Additional file 1: Table S1.

### Results from the BAMSE cohort

In the BAMSE cohort, 48 female participants had lung function and DNA-M data at ages 8 and 16 years, and DNA-M at 41 of the 42 CpGs were available in these 48 females. At 22 of the 41 CpGs, the associations of DNA-M changes with changes in FEV<sub>1</sub>/FVC were consistent with the findings in the IOW cohort, with one CpG showing statistical significance at 0.05 level (cg14552568) and two CpGs approached significance (cg01082111 and cg10027934,  $p$ -value < 0.1). These 22 CpGs were noted as IOW-BAMSE consistent CpGs, of

which 11 of these IOW-BAMSE consistent CpGs were among the 16 IOW-ALSPAC consistent CpGs. These 11 CpGs were further noted as IOW-ALSPAC-BAMSE consistent CpGs.

### Findings of the biological pathway analysis

Genes to which CpGs showed consistent results in either of the two cohorts (ALSPAC and BAMSE) in terms of the direction of associations mapped to were included in the pathway analyses. The 16 IOW-ALSPAC consistent CpGs were mapped to 16 genes, and 22 genes were identified for the 22 IOW-BAMSE consistent CpGs (Table 3). The selected 16 and 22 genes were further investigated to discover the functional enrichment in the biological process by using the bioinformatics tool ToppFun.

In total, eight biological processes were identified from the FDR adjusted  $p$ -value of 0.05 (Table 4). Eight genes, *CELF4*, *INSIG1*, *PTCH1*, *RPS6KA4*, *ZNF304*, *RARA*, *IKBKB*, and *BANP* to which the IOW-ALSPAC consistent CpGs were mapped, were involved in most of the eight biological processes. The same biological processes were found that involved genes *CELF4*, *INSIG1*, *PTCH1*, *RPS6KA4*, *ZNF304*, *DLX5*, *WWOX*, and *ASH1L* corresponding to the IOW-BAMSE consistent CpGs, although they did not survive multiple testing.

**Table 2** Characteristics of subjects with methylation data and lung function of IOW cohort, stratified by gender

IOW cohort						
	Sub cohort at age 10: Participants with lung function Mean $\pm$ SD	Study sample at age 10: Participants with lung function and DNA-M Mean $\pm$ SD	P- values	Sub cohort at age 18: Participants with lung function Mean $\pm$ SD	Study sample at age 18: Participants with lung function and DNA-M Mean $\pm$ SD	P- values
<b>Factors</b>	<b>n = 488</b>	<b>n = 183</b>		<b>n = 395</b>	<b>n = 212</b>	
<b>Males</b>						
FEV <sub>1</sub> (L)	2.05 $\pm$ 0.30	2.06 $\pm$ 0.29	0.477	4.62 $\pm$ 0.62	4.64 $\pm$ 0.62	0.621
FVC (L)	2.35 $\pm$ 0.34	2.36 $\pm$ 0.33	0.813	5.35 $\pm$ 0.72	5.35 $\pm$ 0.73	0.930
FEV <sub>1</sub> /FVC	0.88 $\pm$ 0.06	0.88 $\pm$ 0.06	0.651	0.87 $\pm$ 0.07	0.87 $\pm$ 0.07	0.904
Height (cm)	139.00 $\pm$ 5.90	138.94 $\pm$ 5.95	0.893	177.83 $\pm$ 6.65	177.56 $\pm$ 6.87	0.568
BMI	17.57 $\pm$ 2.52	17.65 $\pm$ 2.54	0.665	22.51 $\pm$ 3.72	22.64 $\pm$ 3.72	0.609
	<b>n (%)</b>	<b>n (%)</b>		<b>n (%)</b>	<b>n (%)</b>	
<i>Smoking<sup>a</sup></i>						
Non-smoker	–	–	–	222 (56.20)	119 (56.13)	0.813
Current smoker	–	–	–	93 (23.54)	46 (21.70)	
Past smoker	–	–	–	72 (18.23)	44 (20.75)	
Missing	–	–	–	8 (2.03)	3 (1.42)	
<b>Females</b>						
FEV <sub>1</sub> (L)	2.00 $\pm$ 0.29	2.01 $\pm$ 0.29	0.832	3.51 $\pm$ 0.45	3.53 $\pm$ 0.43	0.340
FVC (L)	2.23 $\pm$ 0.33	2.24 $\pm$ 0.34	0.657	4.03 $\pm$ 0.53	4.04 $\pm$ 0.51	0.737
FEV <sub>1</sub> /FVC	0.90 $\pm$ 0.06	0.90 $\pm$ 0.05	0.337	0.88 $\pm$ 0.07	0.87 $\pm$ 0.07	0.563
Height (cm)	139.02 $\pm$ 6.43	139.22 $\pm$ 6.58	0.719	164.68 $\pm$ 6.17	165.08 $\pm$ 6.37	0.331
BMI	18.74 $\pm$ 3.34	18.66 $\pm$ 3.41	0.789	23.84 $\pm$ 4.72	23.89 $\pm$ 4.61	0.870
	<b>n (%)</b>	<b>n (%)</b>		<b>n (%)</b>	<b>n (%)</b>	
<i>Smoking<sup>a</sup></i>						
Non-smoker	–	–	–	240 (54.18)	129 (53.53)	0.979
Current smoker	–	–	–	111 (25.06)	61 (25.31)	
Past smoker	–	–	–	87 (19.64)	49 (20.33)	
Missing	–	–	–	5 (1.13)	2 (0.83)	

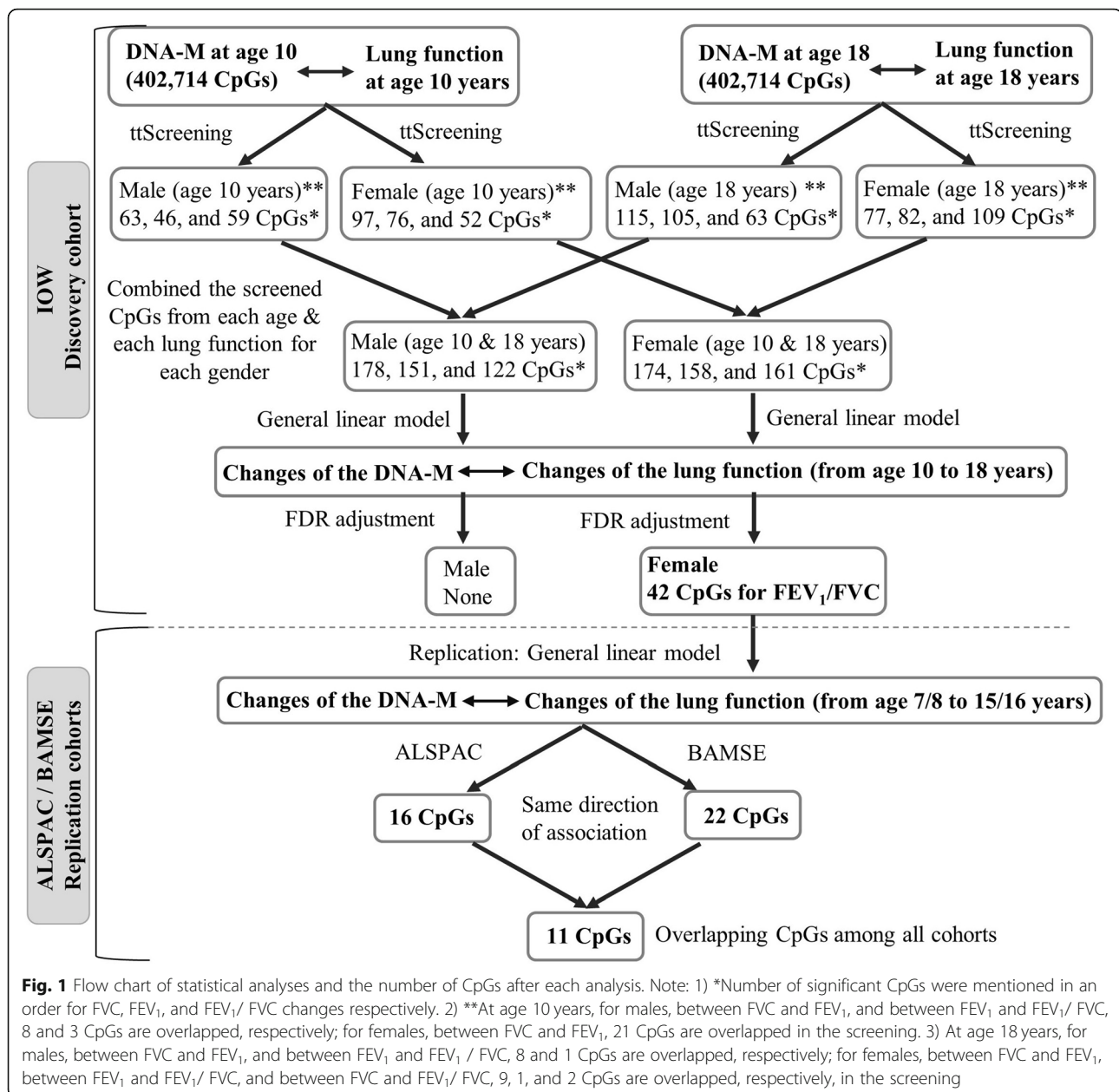
<sup>a</sup>Active smoking at age 10 years in the IOW Cohort was not identified

## Discussion

Limited studies have focused on longitudinal lung function and DNA-M measurements during adolescence, an important period of life that significantly contributes to lung function development [36, 43]. The present study is the first genome-scale exploration of the association of changes of DNA-M with changes in lung function during adolescence, stratified by gender. We showed that DNA-M changes in 11 CpGs were associated with changes in FEV<sub>1</sub>/FVC in females in adolescence, based on findings from the IOW cohort and two independent cohorts. Such associations were not identified in males. It is important to mention that, the final results focused on the direction of associations rather than statistical significance as non-equivalence of statistical significance

and clinical significance has been recognized [52, 53]. We suggest that in replication studies agreement in clinical significance should be more important than statistical significance, although it will be most desirable when an agreement is reached in clinical significance accompanied by statistical significance.

Among the genes involved in the identified biological processes based on the findings in both ALSPAC and BAMSE cohorts, genes *INSIG1*, *PTCH1*, and *PTPRN2* have been shown in a range of studies for their involvement in lung development, lung function, and inflammatory airway diseases such as asthma and COPD [54–60], although most findings were not specifically linked to adolescence. Gene *INSIG1* allied with cg15575249 encodes the protein, insulin induced gene 1, which plays a



significant role in regulating lipogenesis in alveolar types 2 cells consistent with the roles of sterol regulatory element-binding protein (SREBP)/ sterol cleavage-activating protein in lung lipid synthetic pathways [54]. *INSIG1* is primarily involved in epithelial development and surfactant physiology during the perinatal period [55]. The findings in our study further emphasize its importance in the change of lung function in adolescence.

Gene *PTCHI* allied with cg14319249 encodes a member of the patched family of proteins that functions as a receptor and a component of the hedgehog (Hh) signaling pathway [56–58]. The Hh signaling pathway is crucial in embryonic lung development processes, including

the morphogenesis of lung and regulating the interaction between epithelial and mesenchymal cell populations in the airway and alveolar compartments [56–58]. Sonic Hh (one type of Hh signaling) is active in adult lung function [57, 58], but to our knowledge, its relation to lung function changes in adolescence has not been examined before. The link of *PTCHI* with FEV<sub>1</sub>/FVC was also established in a genome-wide association study meta-analysis by the CHARGE consortium [59]. CpGs cg21584493 is mapped to gene *PTPRN2*. In a recent study, differentially methylated region (DMR) annotated to *PTPRN2* genes was identified for the association with lung function and asthma in children [60]. Findings in



**Table 3** CpGs showing consistent associations in females between the IOW and replication cohorts, ALSPAC and BAMSE

CpG Name	Chr.	Gene name	Location	IOW cohort			ALSPAC-cohort		BAMSE cohort	
				Coeff.	$P_{Raw}$ -value	$P_{FDR}$ -value	Coeff.	$P$ -value	Coeff.	$P$ -value
cg08095278	1	<i>ASH1L</i>	TSS1500	-0.008	0.0031	0.0218	-	-	-0.031	0.339
cg13342625	1	<i>WDR65</i>	TSS200	-0.009	0.0016	0.0187	-	-	-0.021	0.663
cg02288301	2	<i>TMEFF2</i>	TSS1500	-0.011	0.0021	0.0201	-	-	-0.020	0.549
<b>cg08366885</b>	<b>2</b>	<b><i>RAPH1</i></b>	<b>5'UTR</b>	<b>-0.006</b>	<b>0.0092</b>	<b>0.0381</b>	<b>-0.001</b>	<b>0.889</b>	<b>-0.005</b>	<b>0.758</b>
<b>cg16710348</b>	<b>3</b>	<b><i>SLC15A2</i></b>	<b>3'UTR</b>	<b>-0.008</b>	<b>0.0114</b>	<b>0.0436</b>	<b>-0.0002</b>	<b>0.975</b>	<b>-0.001</b>	<b>0.980</b>
<b>cg09839318</b>	<b>4</b>	<b><i>GAK</i></b>	<b>Body</b>	<b>-0.007</b>	<b>0.0084</b>	<b>0.0364</b>	<b>-0.002</b>	<b>0.762</b>	<b>-0.035</b>	<b>0.163</b>
cg00930455	7	<i>DLX5</i>	TSS1500	-0.010	0.0055	0.0299	-	-	-0.0003	0.990
<b>cg04132649</b>	<b>7</b>	<b><i>TECPR1</i></b>	<b>Body</b>	<b>-0.008</b>	<b>0.0071</b>	<b>0.0345</b>	<b>-0.003</b>	<b>0.573</b>	<b>-0.008</b>	<b>0.770</b>
cg14552568	7	<i>HTR5A</i>	Intergenic	-0.011	0.0012	0.0187	-	-	-0.091	<b>0.006</b>
<b>cg15575249</b>	<b>7</b>	<b><i>INSIG1</i></b>	<b>Intergenic</b>	<b>-0.008</b>	<b>0.0014</b>	<b>0.0187</b>	<b>-0.005</b>	<b>0.523</b>	<b>-0.007</b>	<b>0.826</b>
<b>cg21584493</b>	<b>7</b>	<b><i>PTPRN2</i></b>	<b>Body</b>	<b>-0.024</b>	<b>0.0033</b>	<b>0.0218</b>	<b>-0.008</b>	<b>0.287</b>	<b>-0.044</b>	<b>0.369</b>
cg09573852	8	<i>IKBKB</i>	Body	-0.006	0.0054	0.0299	-0.005	0.399	-	-
cg23188819	8	<i>FAM160B2</i>	Body	-0.007	0.0030	0.0218	-0.009	0.422	-	-
<b>cg14319249</b>	<b>9</b>	<b><i>PTCH1</i></b>	<b>TSS200</b>	<b>-0.008</b>	<b>0.0021</b>	<b>0.0201</b>	<b>-0.0002</b>	<b>0.987</b>	<b>-0.012</b>	<b>0.684</b>
cg09033333	10	<i>JAKMIP3</i>	Intergenic	-0.007	0.0035	0.0223	-0.0002	0.977	-	-
<b>cg01082111</b>	<b>11</b>	<b><i>RPS6KA4</i></b>	<b>Intergenic</b>	<b>-0.016</b>	<b>0.0002</b>	<b>0.0092</b>	<b>-0.005</b>	<b>0.551</b>	<b>-0.058</b>	<b>0.051</b>
cg07427606	12	<i>MMP17</i>	Body	-0.007	0.0031	0.0218	-	-	-0.022	0.412
cg05312779	15	<i>ANPEP</i>	3'UTR	-0.010	0.0062	0.0315	-	-	-0.037	0.220
cg04575609*	16	<i>BANP</i>	Body	-0.007	0.0021	0.0201	-0.001	0.884	-	-
cg04933438	16	<i>WVVOX</i>	Body	-0.012	0.0002	0.0092	-	-	-0.006	0.849
cg11316510	17	<i>RARA</i>	Body	-0.006	0.0100	0.0403	-0.012	0.290	-	-
<b>cg11493223</b>	<b>17</b>	<b><i>TMC6</i></b>	<b>TSS200</b>	<b>-0.008</b>	<b>0.0027</b>	<b>0.0217</b>	<b>-0.005</b>	<b>0.703</b>	<b>-0.039</b>	<b>0.186</b>
<b>cg13206530</b>	<b>18</b>	<b><i>CELF4</i></b>	<b>Intergenic</b>	<b>-0.011</b>	<b>0.0026</b>	<b>0.0217</b>	<b>-0.0004</b>	<b>0.961</b>	<b>-0.016</b>	<b>0.621</b>
cg00850039	19	<i>ZNF442</i>	TSS200	-0.007	0.0056	0.0299	-	-	-0.013	0.710
<b>cg10157975</b>	<b>19</b>	<b><i>ZNF304</i></b>	<b>TSS1500</b>	<b>-0.006</b>	<b>0.0063</b>	<b>0.0315</b>	<b>-0.020</b>	<b>0.256</b>	<b>-0.003</b>	<b>0.909</b>
cg10027934	22	<i>MAP3K7IP1</i>	Body	-0.014	0.0012	0.0187	-	-	-0.051	0.081
cg27652464	22	<i>FAM19A5</i>	Body	-0.008	0.0022	0.0201	-	-	-0.028	0.219

Note: 1) Regression coefficients were for the associations of changes in DNA-M with FEV<sub>1</sub>/FVC changes in females

2) CpGs with the genes' names in **bold** font were overlapped across all the three cohorts (IOW-ALSPAC-BAMSE consistent CpGs)

3) In BAMSE cohort, DNA-M of \*cg04575609 was excluded at the time of quality control and was not available for the replication analysis

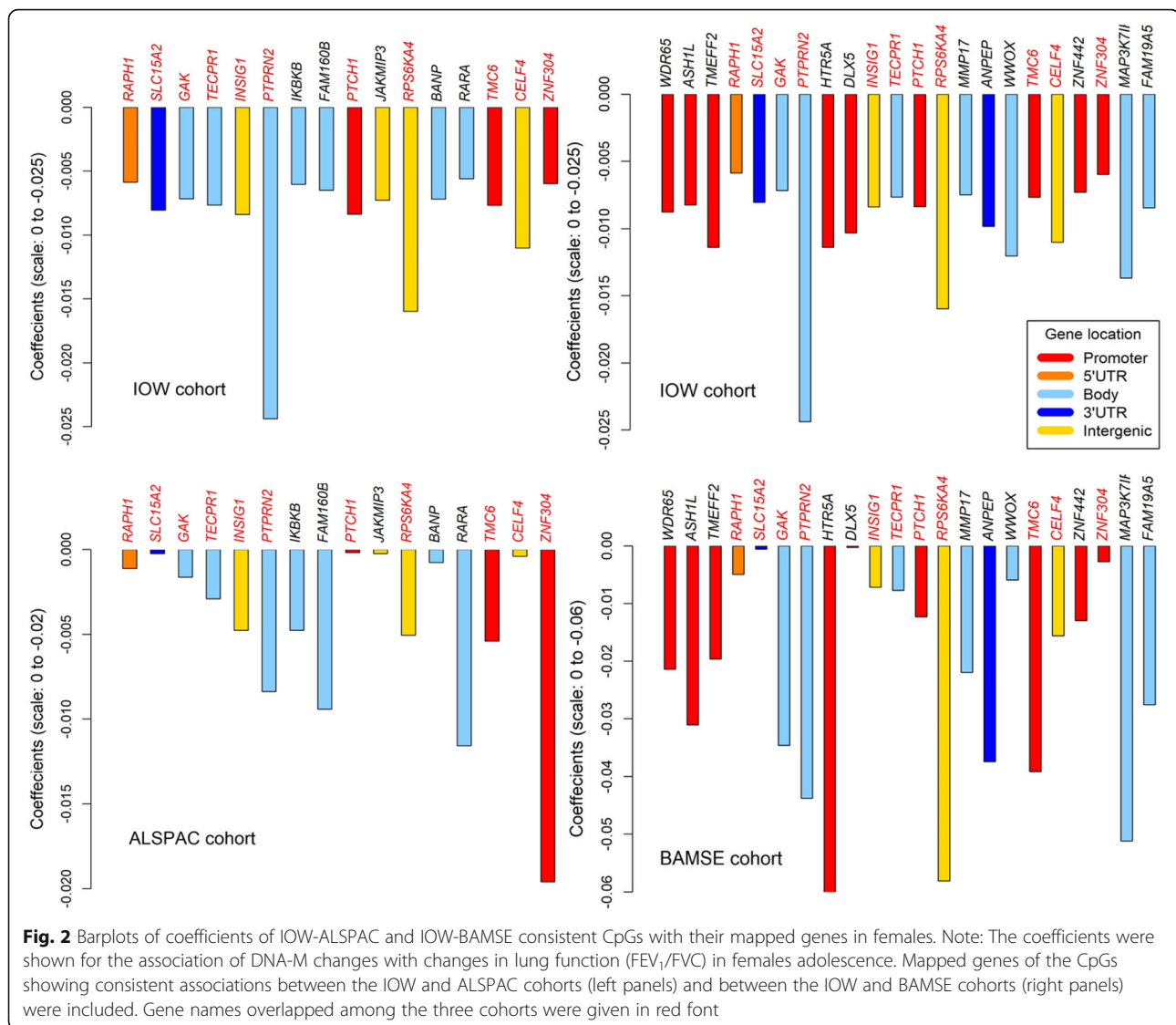
4) Chr. chromosome number, Coeff. coefficients

our study on these genes (*INSIG1*, *PTCH1*, and *PTPRN2*) further emphasizes their epigenetic contribution to the changes in lung function in adolescence.

CpGs cg11316510 and cg09573852 on genes *RARA* (retinoic acid receptor alpha) and *IKBKB* (Inhibitor of Nuclear Factor Kappa B Kinase), respectively, were among the IOW-ALSPAC consistent CpGs but not on the list of IOW-BAMSE consistent CpGs. Their significant involvement in lung function, as well as lung function development and pulmonary diseases such as asthma and COPD indicated the potential importance of these two CpGs and their mapped genes [11, 61–71]. *RARA* is the predominant isotype of the retinoic acid receptor (RAR) identified in alveolar type II epithelial cells and components of the retinoic acid

signaling pathway [63–68]. The retinoic acid signaling pathway plays important roles in lung development and alveolarization, and to regulate surfactant protein B gene expression in pulmonary epithelial cells. Adolescence is a period accompanied by significant lung function development and the functionality of this pathway supports the findings in our study. One of our recent studies also showed an epigenetic association of *RARA* with FEV<sub>1</sub>/FVC [11].

*IKBKB* is an enzyme complex that forms part of the nuclear factor-kappa B signaling pathway, which has been considered the master regulator of immune responses and demonstrated to play a cardinal role in allergic airways diseases [69–71]. In addition, gene *IKBKB* was required for the IL17-dependent signaling



that was associated with neutrophilia and pulmonary inflammation [72].

It is worth noting that the genes discussed above were based on the findings in females in our study. For CpGs located on those genes, no statistically significant associations were shown in males. The identified unique 11 CpGs in three population-based cohorts thus have the potential to serve as epigenetic markers related to lung function development during adolescence in females, but not in males. The absence of such epigenetic associations in males led us to postulate the possibility of either different underlying epigenetic mechanisms in each gender in the regulation of gene activity, or that these CpGs are biomarkers of female physiology and/or exposures that influence lung function growth in adolescence. Thus, our findings may help to explain the

various gender-associated health conditions related to lung function development in adolescence, such as gender reversal of asthma incidence in males and females.

There are some limitations of this study. Firstly, DNA-M measurements were made in peripheral blood leukocytes and provide no insight into epigenetic changes in structural cells of the airway. Secondly, concurrent instead of time-lagged modeling was applied to assess the association of DNA-M changes with lung function changes for each gender. In this context, we were not able to examine the potential of changes in DNA-M at the identified CpGs to predict lung function changes. In the IOW cohort, the analyses were based on data collected at ages 10 and 18 years representing pre- and post-adolescence. In the two replication cohorts, however, the corresponding ages were 7–8 years and 15 years

**Table 4** Biological processes were identified from the mapped genes based on the IOW-ALSPAC consistent CpGs

Name of the Biological process	$P_{Raw}$ -value	$P_{FDR}$ -value	Hit in Query List
Positive regulation of RNA metabolic process	3.39E-05	2.93E-02	<b>CELFA4, INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Positive regulation of nucleobase-containing compound metabolic process	9.30E-05	2.93E-02	<b>CELFA4, INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Positive regulation of gene expression	1.16E-04	2.93E-02	<b>CELFA4, INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Interleukin-1-mediated signaling pathway	1.33E-04	2.93E-02	<b>RPS6KA4, IKBKB</b>
Positive regulation of nitrogen compound metabolic process	1.42E-04	2.93E-02	<b>CELFA4, INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Positive regulation of nucleic acid-templated transcription	2.30E-04	3.19E-02	<b>INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Positive regulation of transcription, DNA-templated	2.30E-04	3.19E-02	<b>INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>
Positive regulation of RNA biosynthetic process	2.47E-04	3.19E-02	<b>INSIG1, PTCH1, RPS6KA4, ZNF304, RARA, IKBKB, BANP</b>

Note: The same biological processes were involved in the BAMSE cohort based on the CpGs identified BAMSE, although they did not survive multiple testing after controlling for FDR of 0.05

Genes which were formatted in **bold**, involved in the biological processes in both ALSPAC and BAMSE cohorts

for ALSPAC and 8 and 16 years for BAMSE. It is likely that many participants at age 15/16 years were still in the transition period or even just started puberty. This possibility accompanied by potentially significant changes in DNA-M during adolescence [23] might explain the non-replication of some CpGs identified in the IOW cohort. Other potential contributors to this non-replication may include some covariates being unavailable in the replication cohorts as well as variable characteristics unique to each cohort. On the other hand, the 11 CpGs showing consistent associations across all the three cohorts certainly deserve further assessment of their generalizability, as well as on the potential of predicting lung function changes.

## Conclusions

This epigenetic study represents an integrated strategy to understand lung function changes in males and females during adolescence. We identified 11 CpGs as potential markers for lung function development, which are applicable to females only. Findings from the study provide insight into the role of epigenetics in gender-dependent lung function development during this critical period of life and thus providing a strong foundation to evaluate gender reversal of asthma from male to female in adolescence period. In subsequent studies, the detected 11 CpGs could serve as candidate epigenetic markers to predict changes in lung function during adolescence.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12931-020-01342-y>.

**Additional file 1 Table S1.** List of CpGs ( $k = 42$ ) at which changes of DNA-M were significantly associated with changes of FEV<sub>1</sub>/FVC in females in IOW cohort and examined among the females in ALSPAC and BAMSE cohorts.

## Abbreviations

ATS: American Thoracic Society; ALSPAC: Avon Longitudinal Study of Children and Parents; BAMSE: Children, Allergy, Milieu, Stockholm, Epidemiology; BMI: Body mass index; CpGs: Cytosine-phosphate-guanine dinucleotide site or sites; COPD: Chronic obstructive pulmonary disease; DNA-M: DNA methylation; FVC: Forced vital capacity; FEV<sub>1</sub>: Forced expiratory volume in one second; FDR: False discovery rate; Hh signal: Hedgehog signal; IOW: Isle of Wight; PCs: Principal components; *tscreening*: Training and testing screening

## Acknowledgements

The authors gratefully acknowledge the cooperation of the children and parents who participated in this study and appreciate the hard work of 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 and MRC Hub grant G0900747 91070) for the generation of the methylation data. We acknowledge the contribution of Cory H. White in DNA-M data pre-processing of IOW cohort. The authors are thankful to the High-Performance Computing facility at the University of Memphis. For the ALSPAC cohort, 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.

## Authors' contributions

SKS tested hypotheses, conducted the analysis, and wrote the manuscript. HZ designed the study, guided the analysis, and involved in drafting and revision of the manuscript prior to submission. JWH and SE supervised the

DNA-methylation measurement and revised the manuscript. FIR was involved in the IOW cohort data management and pre-processing. SHA was involved sample collection, data acquisition, DNA-M arraying, study design, and critical revision in IOW cohort. AJH and CLR were involved in the ALSPAC study design and manuscript revision. EM, JH, and SKM were involved in statistical analyses in the BAMSE cohort. All authors read and approved the final manuscript.

#### Funding

The study conveyed in this publication was supported by the National Institute of Allergy and Infectious Diseases under Award Number R01 AI121226 (MPI: Hongmei Zhang and John Holloway). The 10-year follow-up of IOW cohort was funded by National Asthma Campaign, UK (Grant No 364) and the 18-year follow-up by a grant from the National Heart and Blood Institute (R01 HL082925, PI, SH Arshad). The UK Medical Research Council (MRC) and Wellcome (Grant ref.: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). Generation of methylation array data was specifically funded by NIH R01AI121226, R01AI091905, BBSRC BB1025751/1 and BB/1025263/1, MRC MC\_UU\_12013/1, MC\_UU\_12013/2, MC\_UU\_12013/8. Lung function measurements and were funded by grants from the MRC (G0401540/73080 and MR/M022501/1). BAMSE was supported by The Swedish Heart-Lung Foundation, The Swedish Research Council, Stockholm County Council (ALF), the Strategic Research Programme (SFO) in Epidemiology at Karolinska Institutet, the EU project MeDALL (Mechanisms of the Development of ALLergy; No. 261357). EM is supported by a grant from the European Research Council (ERC; No. 757919, TRIBAL).

#### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

#### Ethics approval and consent to participate

Ethics approvals for the Isle of Wight study were obtained from the Isle of Wight Local Research Ethics Committee (recruitment, 1, 2 and 4 years) and National Research Ethics Service, NRES Committee South Central – Southampton B (10 and 18 years) (06/Q1701/34). Written informed consent was obtained from parents to enroll newborns and at subsequent follow-up written informed consent was obtained from parents, participants, or both. At the University of Memphis, the internal review board first approved the project (FWA00006815) in 2015 (IRB ID: 3917).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no potential competing interests.

#### Author details

<sup>1</sup>Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN 38152, USA. <sup>2</sup>School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire MK43 0AL, England. <sup>3</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK. <sup>4</sup>Population Health Sciences, University of Bristol, Bristol BS8 2BN, UK. <sup>5</sup>Department of Clinical Sciences and Education Södersjukhuset, Karolinska Institutet, Stockholm, Sweden. <sup>6</sup>Sachs' Children's Hospital, Stockholm, Sweden. <sup>7</sup>Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK. <sup>8</sup>The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Parkhurst Road, Newport, Isle of Wight PO30 5TG, UK. <sup>9</sup>Large Animal Clinical Sciences, Michigan State University, East Lansing, MI, USA. <sup>10</sup>Human Development and Health, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK.

Received: 3 December 2019 Accepted: 25 March 2020

Published online: 07 April 2020

#### References

- Guthikonda K, Zhang H, Nolan VG, Soto-Ramirez N, Ziyab AH, Ewart S, et al. Oral contraceptives modify the effect of GATA3 polymorphisms on the risk

- of asthma at the age of 18 years via DNA methylation. *Clin Epigenetics*. 2014;6(1):17.
- Yousefi M, Karmaus W, Zhang H, Ewart S, Arshad H, Holloway J. The methylation of the LEPR/LEPROT genotype at the promoter and body regions influence concentrations of leptin in girls and BMI at age 18 years if their mother smoked during pregnancy. *Int J Mol Epidemiol Genet*. 2013;4:86–100.
- Piccioni P, Tassinari R, Carosso A, Carena C, Bugiani M, Bono R. Lung function changes from childhood to adolescence: a seven-year follow-up study. *BMC Pulm Med*. 2015;15:31.
- Mahmoud O, Granell R, Tilling K, Minelli C, Garcia-Aymerich J, Holloway JW, Custovic A, Jarvis D, Sterne J, Henderson J. Association of height growth in puberty with lung function: a longitudinal study. *Am J Respir Crit Care Med*. 2018;198(12):1539–48.
- Berry CE, Billheimer D, Jenkins IC, Lu ZJ, Stern DA, Gerald LB, Carr TF, Guerra S, Morgan WJ, Wright AL, et al. A distinct low lung function trajectory from childhood to the fourth decade of life. 2016;194(5):607–12.
- Becklake MR, Kauffmann F. Gender differences in airway behaviour over the human life span. *Thorax*. 1999;54(12):1119–38.
- Carey MA, Card JW, Voltz JW, Arbes SJ Jr, Germolec DR, Korach KS, et al. It's all about sex: gender, lung development and lung disease. *Trends Endocrinol Metab*. 2007;18(8):308–13.
- LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe (Sheff)*. 2018;14(2):131–40.
- Almqvist C, Worm M, Leynaert B. Working group of GALENWPG. Impact of gender on asthma in childhood and adolescence: a GA2LEN review. *Allergy*. 2008;63(1):47–57.
- Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013;38(1):23–38.
- Imboden M, Wielscher M, Rezwan FI, Amaral André FS, Schaffner E, Jeong A, Beckmeyer-Borowko A, Harris SE, Starr JM, Deary Ian J, et al. Deary Ian J et al. Epigenome-wide association study of lung function level and its change. *Eur Respir J*. 2019;54(1):1900457.
- Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med*. 2012; 185:373–81.
- Lepeule J, Baccarelli A, Motta V, Cantone L, Litonjua AA, Sparrow D, et al. Gene promoter methylation is associated with lung function in the elderly: the normative aging study. *Epigenetics*. 2012;7(3):261–9.
- Lange NE, Sordillo J, Tarantini L, Bollati V, Sparrow D, Vokonas P, Zanobetti A, Schwartz J, Baccarelli A, Litonjua AA, et al. Alu and LINE-1 methylation and lung function in the normative ageing study. *BMJ Open*. 2012;2(5):e001231.
- Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D. Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. *Respir Res*. 2016;17(1):143.
- Lee MK, Hong Y, Kim SY, Kim WJ, London SJ. Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. *Epigenomics*. 2017;9(7):971–84.
- Zhang H, Tong X, Holloway JW, Rezwan FI, Lockett GA, Patil V, et al. The interplay of DNA methylation over time with Th2 pathway genetic variants on asthma risk and temporal asthma transition. *Clin Epigenetics*. 2014;6(1):8.
- Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H. 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(5):1186–201.
- Wang D, Liu X, Zhou Y, Xie H, Hong X, Tsai HJ, et al. Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. *Epigenetics*. 2012;7(6):594–605.
- Madrigano J, Baccarelli AA, Mittleman MA, Sparrow D, Vokonas PS, Tarantini L, et al. Age and epigenetics: longitudinal changes in gene-specific DNA methylation. *Epigenetics*. 2012;7(1):63–70.
- Xu C-J, Bonder MJ, Söderhäll C, Bustamante M, Baiz N, Gehring U, et al. The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics*. 2017;18(1):25.
- Acevedo N, Reinius LE, Vitezic M, Fortino V, Söderhäll C, Honkanen H, et al. Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clin Epigenetics*. 2015;7(1):34.
- Han L, Zhang H, Kaushal A, Rezwan FI, Karmaus W, Henderson AJ, et al. Assessing DNA methylation changes pre- and post-adolescence and

- pubertal exposures via a longitudinal genome-scale study. *Clinical Epigenetics*. 2019;Minor revision and potential acceptable.
24. Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, et al. Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol*. 2018;47(4):1043–i.
  25. Crapo R. Guidelines for methacholine and exercise challenge testing-1999. This official statement of the American Thoracic Society was adopted by the ATS Board of directors, July 1999. *Am J Respir Crit Care Med*. 2000;161:309–29.
  26. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–38.
  27. McClelland M, Hanish J, Nelson M, Patel Y. KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res*. 1988;16(1):364.
  28. Bibikova M, Fan J-B. GoldenGate® assay for DNA methylation profiling. *DNA Methylation*: Springer; 2009. p. 149–63.
  29. Golden LC, Itoh Y, Itoh N, Iyengar S, Coit P, Salama Y, et al. Parent-of-origin differences in DNA methylation of X chromosome genes in T lymphocytes. *Proc Natl Acad Sci*. 2019;116(52):26779–87.
  30. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 2015;16:37.
  31. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9.
  32. Du P, Feng G, Huang S, Kibbe WA, Lin S. Analyze Illumina Infinium methylation microarray data. 2012.
  33. Hollams EM, de Klerk NH, Holt PG, Sly PD. Persistent effects of maternal smoking during pregnancy on lung function and asthma in adolescents. *Am J Respir Crit Care Med*. 2014;189(4):401–7.
  34. Patil VK, Holloway JW, Zhang H, Soto-Ramirez N, Ewart S, Arshad SH, et al. Interaction of prenatal maternal smoking, interleukin 13 genetic variants and DNA methylation influencing airflow and airway reactivity. *Clin Epigenetics*. 2013;5(1):22.
  35. Weiss ST. Lung function and airway diseases. *Nat Genet*. 2010;42(1):14.
  36. Sonnenschein-van der Voort AM, Howe LD, Granell R, Duijts L, Sterne JA, Tilling K, et al. Influence of childhood growth on asthma and lung function in adolescence. *J Allergy Clin Immunol*. 2015;135(6):1435–43 e7.
  37. Ogbuanu IU, Karmaus W, Arshad SH, Kurukulaaratchy RJ, Ewart S. Effect of breastfeeding duration on lung function at age 10 years: a prospective birth cohort study. *Thorax*. 2009;64(1):62–6.
  38. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson 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(1):111–27.
  39. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42(1):97–110.
  40. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data resource profile: accessible resource for integrated Epigenomic studies (ARIES). *Int J Epidemiol*. 2015;44(4):1181–90.
  41. Hallberg J, Ballardini N, Almqvist C, Westman M, van Hage M, Lijja G, et al. Impact of IgE sensitization and rhinitis on inflammatory biomarkers and lung function in adolescents with and without asthma. *Pediatr Allergy Immunol*. 2019;30(1):74–80.
  42. Schultz ES, Hallberg J, Andersson N, Thacher JD, Pershagen G, Bellander T, et al. Early life determinants of lung function change from childhood to adolescence. *Respir Med*. 2018;139:48–54.
  43. Schultz ES, Gruzieva O, Bellander T, Bottai M, Hallberg J, Kull I, 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(12):1286–91.
  44. Gref A, Merid SK, Gruzieva O, Ballereau S, Becker A, Bellander T, et al. Genome-wide interaction analysis of air pollution exposure and childhood asthma with functional follow-up. *Am J Respir Crit Care Med*. 2017;195(10):1373–83.
  45. Gruzieva O, Merid SK, Melén E. An update on epigenetics and childhood respiratory diseases. *Paediatr Respir Rev*. 2014;15:348–54.
  46. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
  47. Li X, Hawkins GA, Ampleford EJ, Moore WC, Li H, Hastie AT, et al. Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J Allergy Clin Immunol*. 2013;132(2):313–20 e15.
  48. Ray MA, Tong X, Lockett GA, Zhang H, Karmaus WJ. An efficient approach to screening Epigenome-wide data. *Biomed Res Int*. 2016;2016:2615348.
  49. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
  50. Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics*. 2009;10(1):73.
  51. Karmaus W, Mukherjee N, Janjanam VD, Chen S, Zhang H, Roberts G, 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):98.
  52. Altman DG, Bland JM. Statistics notes: absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.
  53. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *British Med J (Clin Res Ed)*. 1983;286(6376):1489.
  54. Insig1 Regulates SREBP Mediated Lipogenesis In Alveolar Type 2 Cells. C61 Gene regulation during development and in injury. p. A4953-A.
  55. Bridges JP, Schehr A, Wang Y, Huo L, Besnard V, Ikegami M, et al. Epithelial SCAP/INSIG/SREBP signaling regulates multiple biological processes during perinatal lung maturation. *PLoS One*. 2014;9(5):e91376–e.
  56. Li X, Howard TD, Moore WC, Ampleford EJ, Li H, Busse WW, et al. Importance of hedgehog interacting protein and other lung function genes in asthma. *J Allergy Clin Immunol*. 2011;127(6):1457–65.
  57. Kugler MC, Joyner AL, Loomis CA, Munger JS. Sonic hedgehog signaling in the lung. From development to disease. *Am J Respir Cell Mol Biol*. 2015;52(1):1–13.
  58. Tam A, Hughes M, McNagny KM, Obeidat M, Hackett TL, Leung JM, et al. Hedgehog signaling in the airway epithelium of patients with chronic obstructive pulmonary disease. *Sci Rep*. 2019;9(1):3353.
  59. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcicante KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2010;42(1):45–52.
  60. den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, 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):1801795.
  61. Na H, Lim H, Choi G, Kim BK, Kim SH, Chang YS, et al. Concomitant suppression of TH2 and TH17 cell responses in allergic asthma by targeting retinoic acid receptor-related orphan receptor gamma. *J Allergy Clin Immunol*. 2018;141(6):2061–73.e5.
  62. Xu L, Sun WJ, Jia AJ, Qiu LL, Xiao B, Mu L, et al. MBD2 regulates differentiation and function of Th17 cells in neutrophils- dominant asthma via HIF-1α. *J Inflamm (Lond)*. 2018;15:15.
  63. Yang L, Naltner A, Yan C. Overexpression of dominant negative retinoic acid receptor alpha causes alveolar abnormality in transgenic neonatal lungs. *Endocrinology*. 2003;144(7):3004–11.
  64. Desai TJ, Chen F, Lu J, Qian J, Niederreither K, Dolle P, et al. Distinct roles for retinoic acid receptors alpha and beta in early lung morphogenesis. *Dev Biol*. 2006;291(1):12–24.
  65. Wongtrakool C, Malpel S, Gorenstein J, Sedita J, Ramirez MI, Underhill TM, et al. Down-regulation of retinoic acid receptor alpha signaling is required for acculation and type I cell formation in the developing lung. *J Biol Chem*. 2003;278(47):46911–8.
  66. Manoli SE, Smith LA, Vyhlidal CA, An CH, Porrata Y, Cardoso WW, et al. Maternal smoking and the retinoid pathway in the developing lung. *Respir Res*. 2012;13(1):42.
  67. Yang L, Lian X, Cowen A, Xu H, Du H, Yan C. Synergy between signal transducer and activator of transcription 3 and retinoic acid receptor-alpha in regulation of the surfactant protein B gene in the lung. *Mol Endocrinol (Baltimore, Md)*. 2004;18(6):1520–32.
  68. Mendelsohn C, Lohnes D, Decimo D, Lufkin T, LeMeur M, Chambon P, et al. Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants. *Dev (Cambridge, England)*. 1994;120(10):2749–71.

69. Janssen-Heininger YM, Poynter ME, Aesif SW, Pantano C, Ather JL, Reynaert NL, et al. Nuclear factor  $\kappa$ B, airway epithelium, and asthma: avenues for redox control. *Proc Am Thorac Soc*. 2009;6(3):249–55.
70. Pannicke U, Baumann B, Fuchs S, Henneke P, Rensing-Ehl A, Rizzi M, et al. Deficiency of innate and acquired immunity caused by an IKK $\beta$  mutation. *N Engl J Med*. 2013;369(26):2504–14.
71. Edwards MR, Bartlett NW, Clarke D, Birrell M, Belvisi M, Johnston SL. Targeting the NF- $\kappa$ B pathway in asthma and chronic obstructive pulmonary disease. *Pharmacol Ther*. 2009;121(1):1–13.
72. Esposito S, Ierardi V, Daleno C, Scala A, Terranova L, Tagliabue C, et al. Genetic polymorphisms and risk of recurrent wheezing in pediatric age. *BMC Pulmonary Medicine*. 2014;14(1):162.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

