Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course

Herman T. den Dekker[1,2,3,21], Kimberley Burrows[4,21], Janine F. Felix[1,3,5,21], Lucas A. Salas [6,7,8,9], Ivana Nedeljkovic [3], Jin Yao[10], Sheryl L. Rifas-Shiman[11], Carlos Ruiz-Arenas[6,8,9], N. Amin[3], Mariona Bustamante[6,8,9,12], Dawn L. DeMeo[13], A. John Henderson[4], Caitlin G. Howe[10], Marie-France Hivert[11], M. Arfan Ikram[3], Johan C. de Jongste[2], Lies Lahousse [3,14], Pooja R. Mandaviya[15,16], Joyce B. van Meurs[16], Mariona Pinart[6,8,9,17], Gemma C. Sharp[4], Lisette Stolk[16,18], André G. Uitterlinden[3,16,18], Josep M. Anto[6,8,9,17], Augusto A. Litonjua[13], Carrie V. Breton[10], Guy G. Brusselle[3,10,19], Jordi Sunyer[6,8,9,17], George Davey Smith[4], Caroline L. Relton[4,22], Vincent W.V. Jaddoe[1,3,5,22] and Liesbeth Duijts [2,20,22]

ABSTRACT

Rationale: We aimed to identify differentially methylated regions (DMRs) in cord blood DNA associated with childhood lung function, asthma and chronic obstructive pulmonary disease (COPD) across the life course.

Methods: We meta-analysed epigenome-wide data of 1688 children from five cohorts to identify cord blood DMRs and their annotated genes, in relation to forced expiratory volume in 1 s (FEV_1), FEV_1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC at ages 7–13 years. Identified DMRs were explored for associations with childhood asthma, adult lung function and COPD, gene expression and involvement in biological processes.

Results: We identified 59 DMRs associated with childhood lung function, of which 18 were associated with childhood asthma and nine with COPD in adulthood. Genes annotated to the top 10 identified DMRs were *HOXA5*, *PAOX*, *LINC00602*, *ABCA7*, *PER3*, *CLCA1*, *VENTX*, *NUDT12*, *PTPRN2* and *TCL1A*. Differential gene expression in blood was observed for 32 DMRs in childhood and 18 in adulthood. Genes related with 16 identified DMRs were associated with respiratory developmental or pathogenic pathways.

Interpretation: Our findings suggest that the epigenetic status of the newborn affects respiratory health and disease across the life course.

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

Introduction

Asthma and chronic obstructive pulmonary disease (COPD) have become major global health problems in recent decades [1]. Both diseases are characterised by airway obstruction, as indicated by a reduced forced expiratory volume in 1 s (FEV_1), FEV_1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC ($FEF_{75\%}$) [2]. Childhood lung function predicts lung function and risks of asthma and COPD in later life [3]. An accumulating body of evidence suggests that asthma and COPD have at least part of their origins in fetal life [4, 5]. Genetics alone fail to explain the quickly altering prevalence of allergies and chronic respiratory diseases in the past decades, because any mutation would require multiple generations to occur on a population level [6]. Furthermore, adverse fetal exposures, such as maternal smoking and suboptimal diet increase the risk of asthma and COPD [5]. The pathways linking genetic predisposition and environmental exposures in fetal life with life-course respiratory disease may include epigenetic changes, including DNA-methylation [5]. Epigenetic changes are influenced by environmental exposures and could exert population effects much more rapidly than genetic mutations [6]. DNA-methylation is currently the best understood epigenetic mechanism, and techniques have been developed to assess epigenome-wide DNA-methylation patterns in large population-based studies. Fetal development is characterised by high rates of DNA-methylation changes and rapid organ development [5]. DNA-methylation may affect fetal development through effects on gene transcription and expression [7]. Recent studies assessing the associations between DNA-methylation and childhood respiratory health are mainly limited to candidate genes and small sample sizes [3, 8]. We focused on identification of differential DNA-methylated regions (DMRs) because regional methylation of CpGs controls cell type specific transcription. Additionally, the use of DMRs increases statistical power and minimises the effects of genetic variants in the methylation analyses [9]. Identification of genomic regions with altered DNA-methylation levels related to lung function and respiratory diseases across the life course is important to understand mechanisms underlying associations of environmental and genetic factors with the development of lower lung function and risk of chronic respiratory diseases. We hypothesised that fetal differential DNA-methylation reflected in cord blood DNA of newborns affects gene expression and subsequent respiratory tract development, and predisposes individuals for obstructive airway diseases in later life [10, 11].

We meta-analysed five epigenome-wide association studies using data from 1688 children participating in prospective cohort studies to identify differential DMRs of newborns associated with childhood FEV_1, FEV_1/FVC ratio and $FEF_{75\%}$. Identified top DMRs were subsequently explored for their associations with childhood asthma, lung function in adolescence and adulthood, and COPD in adulthood, and explored for association with gene expression and involvement in biological processes.

Methods

Study design and data sources

We included population-based cohort studies participating in the Pregnancy and Childhood Epigenetics (PACE) consortium with data on epigenome-wide DNA-methylation at birth and lung function in childhood [12]. We used data from 1688 Caucasian children aged 7–13 years participating in the Avon

Affiliations: [1]The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [2]Dept of Pediatrics, Division of Respiratory Medicine and Allergology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [3]Dept of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [4]MRC Integrative Epidemiology Unit, University of Bristol, UK School of Social and Community Medicine, University of Bristol, Bristol, UK. [5]Dept of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [6]ISGLobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [7]Dept of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA. [8]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [9]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [10]Dept of Preventive Medicine, University of Southern California, Los Angeles, CA, USA. [11]Obesity Prevention Program, Dept of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA. [12]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. [13]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. [14]Dept of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [15]Dept of Clinical Chemistry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [16]Dept of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [17]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [18]Netherlands Consortium for Healthy Ageing (NCHA), Erasmus MC, University Medical Center Rotterdam, The Netherlands. [19]Dept of Respiratory Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [20]Dept of Pediatrics, Division of Neonatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [21]These authors contributed equally. [22]These authors contributed equally.

Correspondence: Liesbeth Duijts, Erasmus MC, University Medical Center Rotterdam, Sp-3435, PO Box 2060, 3000 CB, Rotterdam, The Netherlands. E-mail: l.duijts@erasmusmc.nl

Longitudinal Study of Parents and Children (ALSPAC) (United Kingdom), Generation R (the Netherlands), INfancia y Medio Ambiente Study (INMA) (Spain), the Children's Health Study and Project Viva (both USA). These data were used for the primary epigenome-wide meta-analysis to identify DMRs of newborns related to childhood lung function. We aimed to identify DMRs instead of single CpGs as differences at any individual CpG may be small, and the use of DMRs might minimise the effects of genetic variants in the methylation analysis [13, 14].

We used several resources for the secondary analyses. For clinical outcomes, we used childhood asthma data (Generation R, mean age 6 years), lung function data from adolescents (ALSPAC, mean age 15 years) and adults (Rotterdam Study, mean age 66 years, the Netherlands) and COPD data in adults (Rotterdam Study) (figure 1). For gene expression, we used blood samples from children (INMA, at birth and mean age 4 years) and adults (Rotterdam Study). Lastly, we used publicly available resources to relate identified DMRs with biological processes [15–17]. Parents, legal representatives or participants provided informed consent in accordance with local ethics policies. Detailed information about the study design and cohorts is provided in the supplementary appendix.

DNA-methylation

All cohorts extracted DNA from blood samples and used the EZ-96 DNA Methylation kit (Zymo Research Corporation, Irvine, CA, USA) for bisulfite conversion. Samples were processed with the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA, USA) followed by cohort-specific quality control, probe exclusion and data normalisation. Detailed information on cohort-specific data acquisition and quality control is provided in the supplementary appendix.

Respiratory outcome assessment

Lung function measures comprised pre-bronchodilator FEV_1, FEV_1/FVC ratio and $FEF_{75\%}$, which were converted into sex-, age-, height- and ethnicity-adjusted z-scores [18]. Physician-diagnosed asthma was obtained by questions adapted from the International Study of Asthma and Allergies in Childhood [19]. COPD was defined as pre-bronchodilator FEV_1/FVC <0.70 in the absence of asthma, or a doctor diagnosis [20].
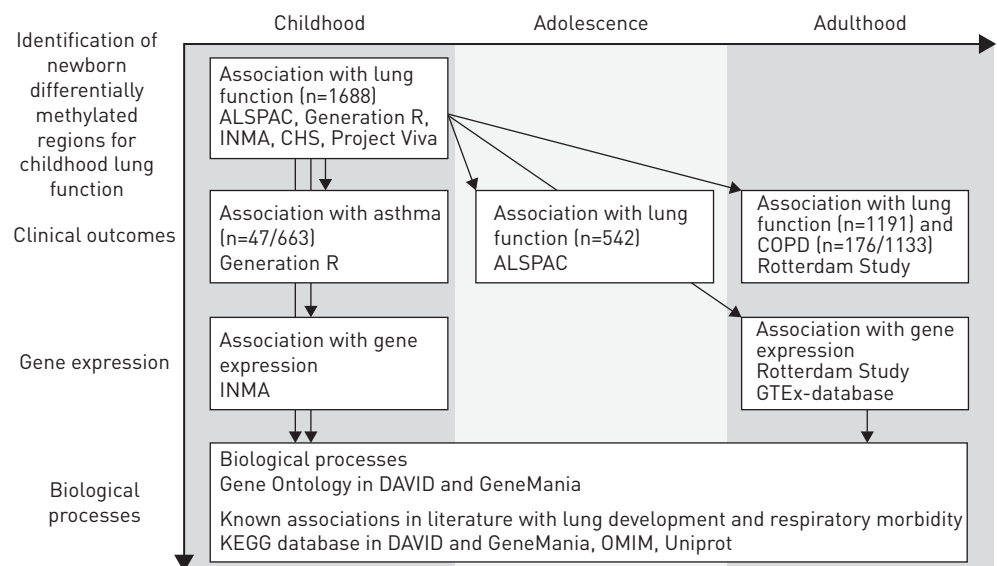

FIGURE 1 Overall study design. Epigenome-wide meta-analyses were performed to identify methylated CpGs associated with lung function in children using data from 1689 children participating in the Avon Longitudinal Study of Parents and Children (ALSPAC), Generation R, INfancia y Medio Ambiente Study (INMA), the Child Health Study (CHS) and Project Viva. Identified differentially methylated regions (DMRs) were annotated to their nearest gene using Peak Annotation and Visualization (PAVIS). Next, we examined if identified DMRs were associated with asthma in children participating in Generation R, lung function in adolescents and adults participating in the ALSPAC or the Rotterdam Study, or chronic obstructive pulmonary disease (COPD) in adults participating in the Rotterdam Study, and with gene expression levels in children participating in INMA, adults in the Rotterdam Study, and the Genotype-Tissue Expression (GTEx)-database. We further explored biological processes and associations with lung development and respiratory morbidity using publicly available resources (DAVID, GeneMania, Online Mendelian Inheritance in Man (OMIM) and Universal Protein Resource (UniProt)). KEGG: Kyoto Encyclopedia of Genes and Genomes.

Statistical analyses

Primary meta-analysis on childhood lung function

A detailed description of applied methods is presented in the supplementary appendix. Individual cohorts used robust linear regression models to examine the associations of DNA-methylation levels of CpGs with childhood FEV_1, FEV_1/FVC ratio and $FEF_{75\%}$. Analyses were adjusted for maternal age, socioeconomic status, smoking during pregnancy, parity, asthma or atopy, technical covariates and estimated cell counts [21]. Results were combined using inverse variance-weighted fixed-effect meta-analyses. Results from unadjusted models were similar to fully adjusted models (supplementary table S1). Using p-values obtained from the meta-analyses, we identified DMRs using the software-tool Comb-p, which is the most robust tool to identify DMRs with small effect sizes [9, 22]. Regions were defined as a minimum of two probes within a window size of 500 bases with a false discovery rate (FDR)-threshold <0.05 [9]. Comb-p uses unadjusted p-values for each probe as input, and calculates adjusted p-values for each probe that account for the correlation with nearby CpGs [23]. Next, the SLK p-values were adjusted for multiple testing and adjusted into q-values. Comb-p finds DMRs based on these q-values and calculates p-values for these DMRs based on the original p-values. Finally, the DMR p-values were adjusted for multiple testing using the Šidák-correction based on the size of the region and number of possible regions of that size. A sliding window identifies a DMR without any predefined regional borders, and therefore (theoretically) does not have a maximum number of windows and DMRs. A more extensive description of the identification of DMRS is provided in the supplementary appendix. Annotation of the genes located nearest to the DMRs was performed using Peak Annotation and Visualization (PAVIS) [24]. We limited annotation to a region of 500 kb (250 kb upstream, 250 kb downstream of the beginning and end of the region, respectively). All annotations were based on human GRCh37/hg19 assembly. Because genetic variants in Infinium probes could result in spurious methylation measurements, we performed a sensitivity analysis in a subset of high-quality probes (n=294 834) without single nucleotide polymorphisms (SNPs), insertions or deletions, repeats, polymorphic probes and bisulfite-induced reduced genomic complexity [25].

Secondary analyses on later-life lung function and respiratory diseases

We used linear and logistic regression models to examine the associations of CpGs within identified DMRs with asthma in childhood, FEV_1, FEV_1/FVC ratio and $FEF_{75\%}$ in adolescence and adulthood, and COPD in adulthood. Single CpG p-values were used to reconstruct the identified DMRs with Comb-p, applying identical parameter settings as in the discovery meta-analyses including FDR-correction [9, 26]. We did not apply Šidák-correction because analyses were limited to the identified DMRs.

Gene expression analyses

We assessed the associations of CpGs within identified DMRs with gene expression in a region of ±250 kb in blood samples from children and adults. p-values of CpGs associated with gene expression were combined for each DMR using a modified generalised Fisher method and FDR-correction [26, 27]. Additionally, we explored whether the annotated and differentially expressed genes were expressed in human lung specimens of the Genotype-Tissue Expression (GTEx) database [15].

Exploration biological processes

The gene ontology database implemented in DAVID and Genemania was used to examine gene function in biological processes [16, 17]. We examined pathways for all genes annotated to the DMRs and for genes with differential expression in association with the identified DMRs. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in DAVID and Genemania, the Online Mendelian Inheritance in Man (OMIM) database and the Universal Protein Resource (UniProt) to explore whether annotated or expressed genes have been related to respiratory development or diseases [28]. We used the Ensemble Genome browser to visualise the genomic structure of the identified DMRs.

Results

Meta-analysis of epigenome-wide association studies on childhood lung function

Characteristics of the participating cohorts are given in table 1 and supplementary table S2.

We identified 22, 15 and 22 DMRs associated with FEV_1, FEV_1/FVC ratio and $FEF_{75\%}$, respectively (figure 2, supplementary tables 3 and 4). A higher mean methylation of CpGs located within 37 (63%) of the identified DMRs was associated with higher lung function measures, and within 22 (37%) of identified DMRs with lower lung function measures. We observed a high homogeneity across the included studies (CpGs with I^2<50: FEV_1 140/163 (86%), FEV_1/FVC 82/89 (92%) and $FEF_{75\%}$ 139/148 (87%)) (supplementary tables S4a–c).

Of the top 10 significant DMRs associated with childhood lung function, the five DMRs and their annotated genes for FEV_1 were located at chr7:27 183 133–27 184 854 (*HOXA5*), chr10:135 202 522–135 203 201

TABLE 1 Characteristics of cohorts and their participants

	Participants	Type of blood sample for DNA-methylation	Available CpGs	Subjects with expression data	Age at lung function measurement years	Asthma		COPD	
						Cases	Controls	Cases	Controls
Primary analyses									
ALSPAC (UK)	654	Cord blood	471 193	n.a.	8.6±0.2	n.a.	n.a.	n.a.	n.a.
Generation R (the Netherlands)	643	Cord blood	436 013	n.a.	9.8±0.3	47	663	n.a.	n.a.
INMA (Spain)	140	Cord blood	439 306	107	6.9±0.3	n.a.	n.a.	n.a.	n.a.
CHS (USA)	75	Cord blood	383 857	n.a.	13.3±0.6	n.a.	n.a.	n.a.	n.a.
Project Viva (USA)	176	Cord blood	470 870	n.a.	7.9±0.7	n.a.	n.a.	n.a.	n.a.
Secondary analyses									
ALSPAC (UK)	542	Cord blood	n.a.	n.a.	15.4±0.2	n.a.	n.a.	n.a.	n.a.
Rotterdam Study – I (the Netherlands)	488	Peripheral blood	n.a.	488	64.0±6.3	n.a.	n.a.	63	425
Rotterdam Study – II (the Netherlands)	703	Peripheral blood	n.a.	703	67.5±5.9	n.a.	n.a.	92	611

Data are presented as n or mean±SD, unless otherwise stated. Lung function data were obtained using spirometry, and sex-, age-, height- and ethnicity-adjusted z-scores were calculated. COPD: chronic obstructive pulmonary disease; ALSPAC: Avon Longitudinal Study of Parents and Children; INMA: INfancia y Medio Ambiente Study; CHS: Child Health Study; n.a.: not applicable.

(PAOX), chr6:166 418 799–166 419 139 (LINC00602), chr19:1 063 624–1 064 219 (ABCA7) and chr1:7 887 199–7 887 561 (PER3). Three DMRs and their annotated genes for FEV1/FVC ratio were located at chr1:86 968 087–86 968 544 (CLCA1), chr10:135 051 149–135 051 582 (VENTX) and chr5:102 898 223–102 898 734 (NUDT12). Two DMRs for FEF75% and their annotated genes were located at chr7:158 045 980–158 046 359 (PTPRN2) and chr14:96 180 406–96 181 045 (TCL1A). After exclusion of potentially problematic probes containing genomic variants, 41 out of the 59 previously identified DMRs still contained two or more CpGs (supplementary table S4). Of these 41 DMRs, 54% (n=22) remained to be associated with childhood lung function (supplementary table S5).

Identified DMRs and lung function and respiratory diseases across the life course
Of all 59 identified DMRs related with childhood lung function, 18 (31%) were associated with childhood asthma (figure 3, supplementary table S6). Furthermore, 11 (19%) and nine (15%) DMRs were associated with lung function in adolescence and adulthood, respectively, and nine (15%) were associated with COPD. The DMRs annotated to HOXA5 and PAOX were associated with childhood and adolescence FEV1 and COPD, but not with childhood asthma or adult lung function. The DMRs annotated to PER3 and VENTX were associated with childhood and adolescence FEV1 and FEV1/FVC ratio, respectively. The DMR annotated to NUDT12 was associated with childhood FEV1/FVC and COPD. The DMRs annotated to PTPRN2 and TCL1A were associated with childhood FEF75% and asthma. The DMRs annotated to LINC00602, ABCA7 and CLCA1 were associated with childhood lung function, but not with other outcomes.

Identified DMRs and gene expression
Of the 59 identified DMRs, 32 (54%) DMRs at birth were associated with gene expression at age 4 years, and 18 (31%) DMRs with gene expression in adulthood (supplementary table S7). The DMR annotated to HOXA5 was associated with differential expression of several genes of the HOX-family (table 2). The DMRs annotated to PER3, VENTX, NUDT12 and TCL1A were associated with differential expression of their respective genes. The DMRs annotated to PAOX, LINC00602, ABCA7, CLCA1 and PTPRN2 were not associated with expression of their corresponding genes. Genes annotated to 28 (47%) of all identified DMRs were expressed in adult lung tissue, including the top significant DMRs annotated to PAOX, ABCA7, CLCA1, VENTX and NUDT12 (supplementary table S8).
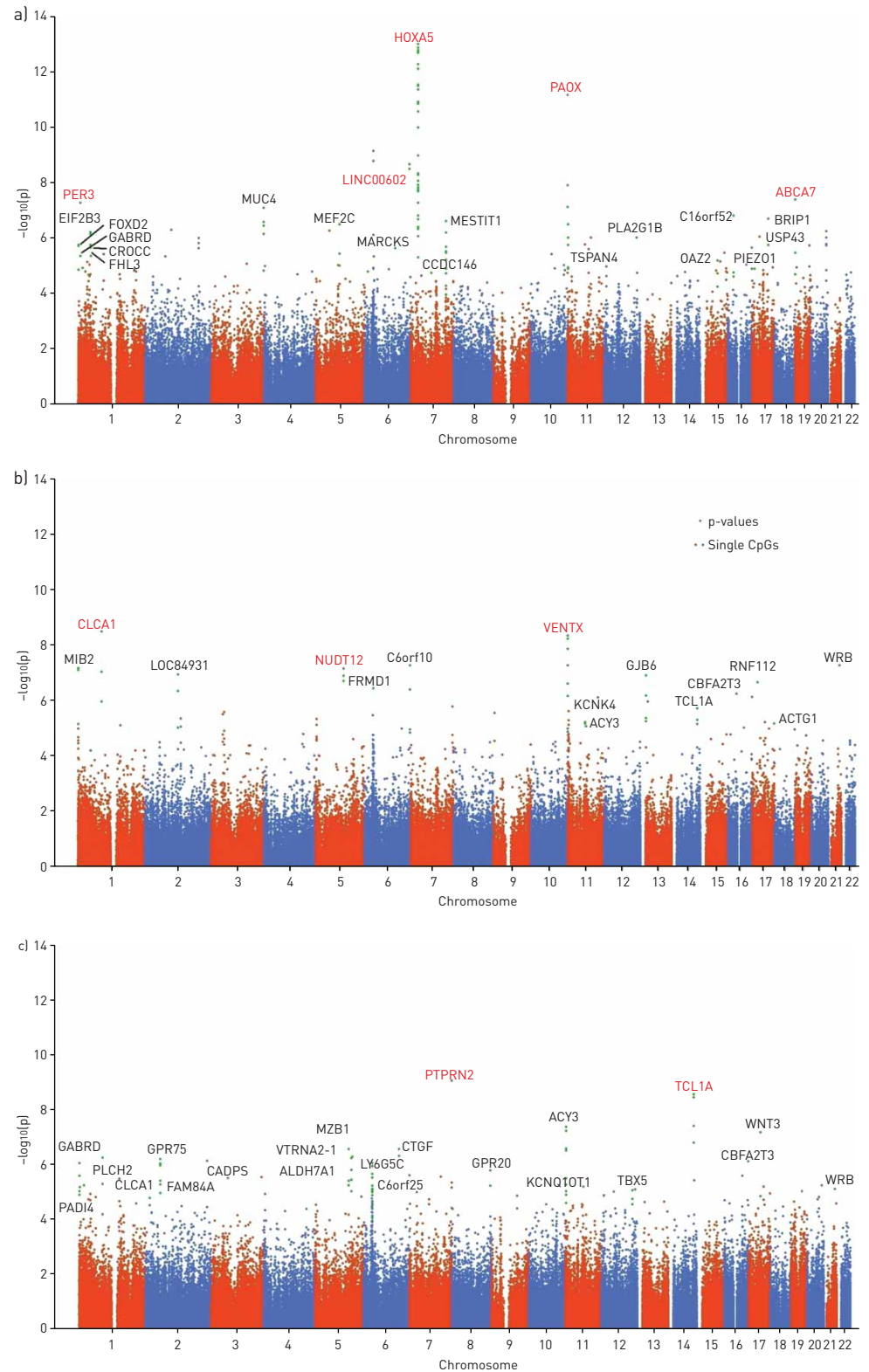
FIGURE 2 Manhattan plots of associations of CpGs located in differentially methylated regions (DMRs) with childhood lung function outcomes. Green dots represent p-values from associations of CpGs located in DMRs at birth with childhood forced expiratory volume in 1 s (FEV1), FEV1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC (FEF75%). p-values of DMRs ranged from 3.05×10^{-14} to 0.031, and details are provided in supplementary table S3. Nearest annotated genes of DMRs are provided. The genes annotated to the top 10 significant DMRs associated with childhood lung function are written in red. p-values of CpGs not located in DMRs are presented as red and blue dots, corrected for correlations with neighbouring CpGs.
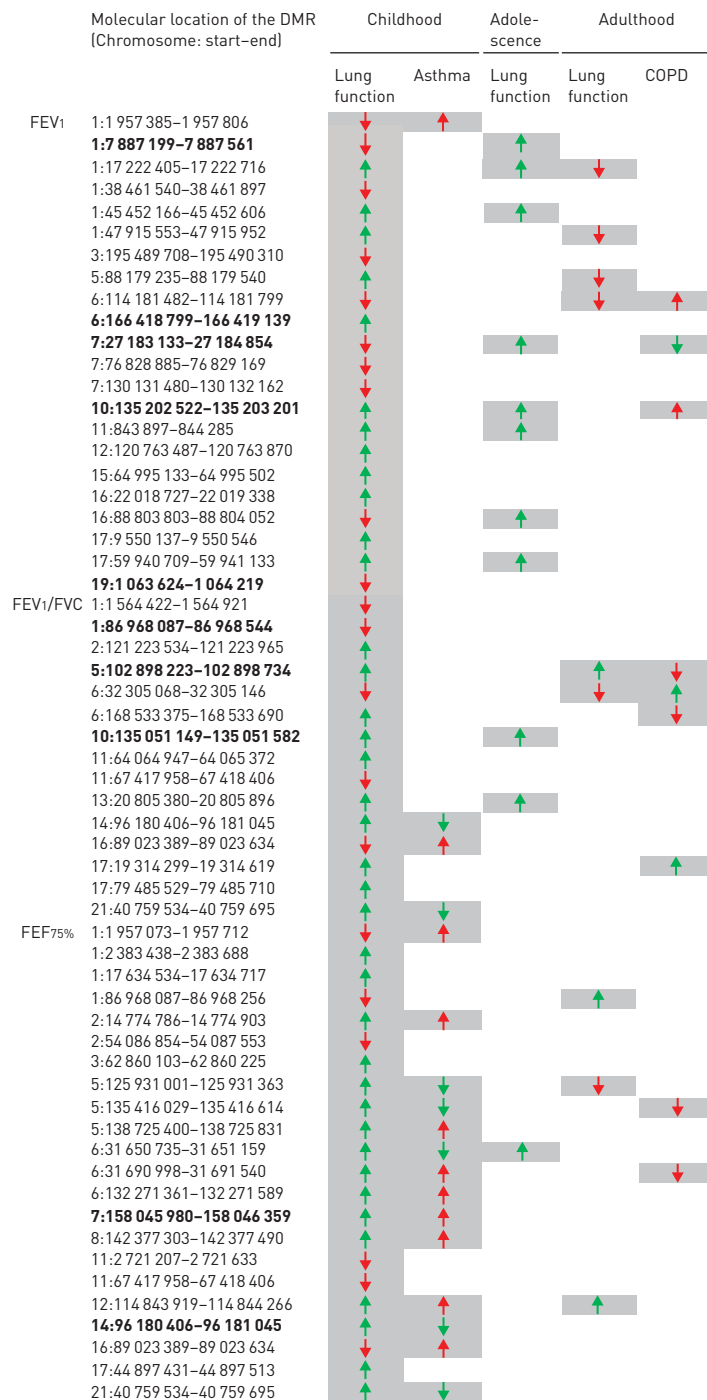
FIGURE 3 Identified differentially methylated regions (DMRs) and their location, and direction of associations with childhood lung function, childhood asthma, adolescent lung function and adult lung function and chronic obstructive pulmonary disease (COPD). Results present identified DMRs from association-analyses of DNA-methylation at birth with childhood forced expiratory volume in 1 s (FEV$_1$), FEV$_1$/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC (FEF$_{75\%}$), their location, and their direction of association with childhood lung function, childhood asthma, adolescent lung function and adult lung function and COPD. Molecular locations of the top 10 significant DMRs are presented in bold. Identified DMRs associated with childhood lung function and other respiratory outcomes are marked in grey, and if not associated with respiratory outcomes in white. Directions of associations are marked ↓ if a higher mean methylation of the DMRs was associated with a lower z-score for lung function or lower risk of asthma or COPD, and marked ↑ if a higher mean methylation of the DMRs was associated with a higher z-score of lung function or higher risk of asthma or COPD. Red arrows represent disadvantageous effect estimates (lower lung function, increased risk of asthma or COPD), and green arrows beneficial effect estimates (higher lung function, lower risk of asthma or COPD).

TABLE 2 Associations of the top 10 significant identified differentially methylated regions (DMRs) with gene expression and related respiratory outcomes

Molecular location of the DMR (chromosome: start – end)	Lung function	Annotated gene[#]	Expressed gene[¶]	Gene expression in children	Gene expression in adults	Previously associated with lung development or respiratory morbidity[+]
chr1: 7 887 199–7 887 561	FEV_1	PER3	PER3, RP3-467l1.4, RNA5SP23, RP4-726F1.1	↓	−	
			RP11-431K24.1	−	↑	
chr1: 86 968 087–86 968 544	FEV_1/FVC	CLCA1	No expression	−	−	Lung development, asthma, COPD
chr5: 102 898 223–102 898 734	FEV_1/FVC	NUDT12	NUDT12	↓	−	Smoking behaviour in COPD
			CMBL	↓	−	
chr6: 166 418 799–166 419 139	FEV_1	LINC00602	No expression	−	−	
chr7: 27 183 133–27 184 854	FEV_1	HOXA5	HOXA1, HOTTIP	↓	↓	Lung development, FEV_1, FEV_1/FVC
			EVX1, HOXA4, HOXA7	↓	−	Lung development, asthma, COPD
chr7: 158 045 980–158 046 359	$FEF_{75\%}$	PTPRN2	No expression	−	−	
chr10: 135 202 522–135 203 201	FEV_1	PAOX	No expression	−	−	
chr10: 135 051 149–135 051 582	FEV_1/FVC	VENTX	TUBGCP2, RP11-122K13.12	↓	−	
			VENTX, ECHS1	↑	−	
			SPRN	↑	↓	
			ZNF511	−	↑	
chr14: 96 180 406–96 181 045	$FEF_{75\%}$	TCL1A	TCL1A, CCDC85C	↓	−	Asthma
chr19: 1 063 624–1 064 219	FEV_1	ABCA7	No expression	−	−	

Results present identified DMRs from association-analyses of DNA-methylation at birth with childhood forced expiratory volume in 1 s (FEV_1), FEV_1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC ($FEF_{75\%}$). COPD: chronic obstructive pulmonary disease; ↓: a higher methylation of the DMR was associated with a decreased expression of the specific gene; ↑: a higher methylation of the DMR was associated with an increased expression of the specific gene; −: no direction of association observed. [#]: DMRs were annotated to their nearest gene; [¶]: identified DMRs at birth were associated with gene expression in childhood (INfancia y Medio Ambiente Study (INMA); mean age 4 years) and adulthood (the Rotterdam Study; mean age 66 years); gene expressions levels were assessed limited to 250 kb up- and downstream of the outer border of the DMR; [+]: associations of expressed genes with lung development and respiratory morbidity were explored in previous published studies.

Identified DMRs and related biological processes

Out of all 59 identified DMRs, 43 were annotated to genes not previously associated with lung function or respiratory morbidity (supplementary table S7). Of the genes annotated to the top 10 significant DMRs, HOXA5, CLCA1, TCL1A and NUDT12 have been previously associated with respiratory development including alveogenesis, respiratory diseases and cellular immunity (table 2). Integrative pathway analysis showed that the majority of genes annotated to DMRs associated with childhood lung function were in co-expression and are known for physical and genetic interactions (supplementary figures S1a–c). Genes related to the identified DMRs, including HOXA5, PER3, CLCA1, NUDT12 and PTPRN2 were located in pathways related to regionalisation, DNA- and RNA-regulation and embryonic development (supplementary table S9). The genes HLA-DRB4 and HLA-DRB5 were enriched in processes including asthma. These genes were associated with the DMR located at chr6:32 305 068–32 305 146, which was related with childhood and adulthood FEV_1/FVC ratio and COPD.

Of the top 10 significant DMRs, the DMRs annotated to HOXA5, CLCA1 and TCL1A contained CTCF-binding sites (supplementary figure S2). The DMRs annotated to HOXA5, PAOX, PER3 and NUDT12 were located in promotor regions of their respective genes. The DMR annotated to ABCA7 was located in a CpG-island.

Discussion

We identified 59 DMRs in neonatal cord blood associated with childhood lung function. 18 (31%) of all identified DMRs were also associated with childhood asthma, 11 (19%) and nine (15%) with adolescent and adult lung function, respectively, and nine (15%) with COPD. Differential gene expression was observed for 32 (54%) DMRs in childhood and 18 (31%) DMRs in adulthood. Multiple genes related to the identified DMRs have previously been associated with respiratory development and morbidity, and many identified DMRs were located within known regulatory elements for gene expression.

Reduced lung function in childhood is associated with reduced lung function and increased risks of asthma and COPD many decades later [3, 10]. Pathways of environmental exposures in early life that affect lung development might be modified by genetic susceptibility. Conversely, genetic susceptibility could partly explain the difference in adverse effects of early environmental exposures on the risk of chronic obstructive respiratory diseases. Identified genetic variants associated with childhood asthma in large-scale genome-wide association studies (GWAS) only account for up to 7.5% of the explained variance [29]. Epigenetic mechanisms could link environmental exposures with the unexplained heritability of childhood asthma [30, 31]. Studies that examine associations of DNA-methylation with lung function, asthma or COPD are scarce, limited to candidate genes or high-risk populations, and lack replication. An epigenome-wide study among 97 asthmatics and 97 healthy children aged 6–12 years identified 81 DMRs associated with asthma, of which 16 were also associated with FEV1 [8]. Out of these 81 DMRs, 19 were located within 500 kb of our identified DMRs and may affect the same genes. Another epigenome-wide study in 1454 adults identified 349 CpGs associated with COPD [32]. Four annotated genes in this adult study (*CBFA2T3*, *PADI4*, *LST1* and *KCNQ1*) were replicated in our study of children. Multiple genes associated with the identified DMRs have previously been related with asthma and COPD in GWAS. Nine (15%) of the 59 DMRs we identified were associated with adult lung function, and annotated to, or associated with differential expression of 11 genes. Nine of these genes were previously linked with pulmonary structures, immunity, asthma, COPD and smoking behaviour in COPD [28]. This suggests that genes associated with respiratory diseases could be influenced by differential DNA-methylation from early life onwards.

We explored the biological processes of the top significant DMRs for development of respiratory morbidity [28]. The DMR annotated to *HOXA5* was associated with childhood and adolescent FEV1, COPD and differential expression of *HOXA1*, *HOXA4* and *HOXA7*. One DMR associated with childhood FEV1/FVC ratio was annotated to *VENTX*, which is a member of the *HOX*-gene family. The DMR annotated to *LINC00602* (long intergenic non-protein coding RNA (lncRNA) 602) was linked to childhood FEV1. LncRNAs influence gene-specific epigenetic regulation and interact among others with the transcription of *HOX*-genes. *HOX*-genes are critical for segmental fetal development, and *HOXA5* especially is required for embryonic respiratory tract morphogenesis [33, 34]. The DMR annotated to *PAOX* was linked to childhood and adolescent FEV1 and COPD. *PAOX* is involved in the regulation of intracellular polyamine, which is essential for protein synthesis. The DMR linked to *ABCA7* was associated with FEV1 in childhood and adolescence. *ABCA7* is involved in the lipid homeostasis in the cellular immune system and is essential for phagocytosis of apoptotic cells by alveolar macrophages [35]. *PER3*, annotated to a DMR associated with FEV1 in children and adolescents, is a key element in the endogenous circadian rhythm. The DMR linked to *CLCA1* was associated with childhood FEV1/FVC ratio and FEF75%, and expressed in adult lung tissue. *CLCA1* affects interleukin-13-driven mucus production in human airway epithelial cells and is associated with asthma and COPD [36–38]. *NUDT12*, annotated to a DMR associated with childhood FEV1/FVC ratio and COPD, is involved in intracellular biochemical reactions. *NUDT12* is associated with smoking behaviour in COPD [39].

PTPRN2, annotated to a DMR associated with childhood FEF75% and asthma is member of a gene family regulating cell growth and differentiation, and is involved in vesicle-mediated secretory processes. DNA-methylation of *PTPRN2* differentiates between lung cancer, pulmonary fibrosis and COPD [40]. *TCL1A*, annotated to a DMR associated with childhood FEV1/FVC, FEF75% and asthma, is specific to developing lymphocytes when expressed and is associated with asthma [41]. Thus, many of the genes annotated to the top significant DMRs are involved in respiratory development, cellular immunity and respiratory morbidity. Furthermore, genes associated with 29 of the identified DMRs, including *HOXA5*, *PAOX*, *VENTX*, *PTPRN2* and *TCL1A*, have been reported to be differentially methylated in relation to maternal smoking during pregnancy [12]. Genes related with four identified DMRs associated with childhood lung function were differentially methylated in association with maternal folate levels during pregnancy [42].

This is the largest study to date evaluating the associations of newborn epigenome-wide DNA-methylation with lung function and respiratory disease in children and adults, and it provides new insights into the epigenetic changes in fetal life that increase the risk of lifetime respiratory morbidity. We aimed to strengthen our results using public databases on gene expression and biological pathways. Ideally, the presence of identified DMRs would be replicated in lung cells. However, in cohort studies, this is not done for reasons of ethics. It is unknown whether nasal cells, which are easier to acquire, have a high enough correlation of DNA-methylation with lung tissue. Therefore, further studies should aim to examine whether DNA-methylation in nasal cells is a good proxy for lung tissue and data on DNA-methylation and phenotypes should be shared in consortia to increase statistical power. These results cannot currently be used as predictors of disease in individuals, but are important from an aetiological perspective. Further

experimental or Mendelian randomisation studies on the identified DMRs and associated genes might inform strategies in early life to improve lung function and lower the lifetime risk of obstructive respiratory diseases.

Some limitations should be discussed. We measured DNA-methylation in blood because it is easily accessible in large cohort studies. Blood DNA-methylation does not necessarily reflect lung epithelial DNA-methylation. However, asthma and COPD have systemic manifestations, characterised by increased inflammatory blood markers [43, 44]. Although the analyses were adjusted for estimated cell counts, we cannot rule out residual confounding due to alterations in cell type distribution. Recently, two new reference sets for cell type adjustment in cord blood samples were published [45, 46]. These reference sets are currently being validated, and future studies will shed light on the differences between reference panels. In our secondary analyses, we assessed whether the identified DMRs were associated with lung function measured in adolescence and adulthood. DNA methylation patterns and expression of genes vary depending on the developmental stage, and these changes could be non-linear [47]. We were not able to assess the stability of DNA methylation in the identified DMRs in the same individuals from birth to adulthood. In a recent study addressing DNA methylation changes in early life, significantly reduced or increased methylation of single CpGs between ages 0–4 years and 4–8 years occurred in <4% of all CpGs, suggesting only a minor change in genome-wide DNA methylation in childhood [48]. Longitudinal changes in DNA methylation in the same individuals from early life until adulthood in relation to respiratory morbidity have not yet been studied.

We presented our primary results including all probes, and provided results of analyses excluding potentially problematic probes. In our stringent sensitivity analyses, we observed similar size and direction of the effect estimates in 54% of the identified DMRs associated with childhood lung function. However, the true exact impact of potentially problematic probes on the measurement of DNA methylation remains unknown [25, 49]. Discarding probes *a priori* may discard information. Therefore, we present all results of the main and sensitivity analysis.

Genetic variation might be influencing the DMRs associated with respiratory health. A recent study in two ethnic diverse adult cohorts in 557 subjects showed that DNA methylation of airway epithelium plays a central role in mediating the effects of SNPs and gene expression on asthma risk and its clinical course [50]. Another study in 115 adults reported a potential mediating effect of DNA methylation of single CpGs on the associations between SNPs located at chromosomal locus 17q21 and asthma [51]. The study identified six CpGs associated with gene expression of ORMDL3 and GSDMB. The authors did not assess the associations of DNA methylation with asthma or lung function. We did not identify any DMR located near the 17q21 locus. This could be explained by the young age of our study subjects or differences in main respiratory outcomes measurements.

Several identified DMRs were associated with gene expression other than the nearest and therefore annotated gene, which limits the potential biological effect of the annotated genes. The genomic inflation factor for the primary analyses ranged from 1.07 to 1.21 (supplementary figure S3). Recently, it was shown that the genomic inflation factor provides an invalid estimate of test-statistic inflation when the outcome of interest is associated with many small genetic effects [52]. Furthermore, estimating the inflation factor using the genomic inflation factor results both in an overestimation of the actual inflation and in imprecise estimates contributing to the previously unexplained high variability across studies. This might explain the genomic inflation in our analyses. The statistical steps in Comb-p limit the final number of DMRs identified, and genomic inflation in the identification of DMRs could not be tested. Further studies are needed to develop statistical tools dealing with genomic inflation in epigenome-wide studies.

There were no cohort studies available for replication or validation analyses. We included all available cohorts in the meta-analysis to obtain the largest possible power to detect new associations. A previously published study has shown that in (epi)genome-wide association analyses a meta-analysis of all participating cohorts rather than a split sample analysis with a properly selected level of (epi)genome-wide significance is the most powerful approach to identify new associations [53]. The high between-study homogeneity observed for the vast majority of CpGs in our meta-analysis (supplementary tables S4a–c) also provides support for the stability of the reported associations. Nevertheless, confirmatory studies are needed and presented results are limited by lack of independent cohorts for validation.

In conclusion, we identified 59 DMRs in cord blood that were associated with childhood lung function. Multiple DMRs were additionally related with childhood asthma, adolescent and adult lung function, or adult COPD. In addition, multiple DMRs were associated with differential gene expression of genes involved in embryonic and respiratory tract development, or were located in regulatory elements for gene expression. These findings suggest that epigenetic changes during fetal life might modify the risk of respiratory diseases across the life course.

References

1   World Health Organization. Chronic Obstructive Pulmonary Disease (COPD). 2015. www.who.int/mediacentre/factsheets/fs315/en/.

2   McDonough JE, Yuan R, Suzuki M, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. N Engl J Med 2011; 365: 1567–1575.

3   McGeachie MJ, Yates KP, Zhou X, et al. Patterns of growth and decline in lung function in persistent childhood asthma. N Engl J Med 2016; 374: 1842–1852.

4   Martinez FD. Early-life origins of chronic obstructive pulmonary disease. N Engl J Med 2016; 375: 871–878.

5   Duijts L, Reiss IK, Brusselle G, et al. Early origins of chronic obstructive lung diseases across the life course. Eur J Epidemiol 2014; 29: 871–885.

6   Begin P, Nadeau KC. Epigenetic regulation of asthma and allergic disease. Allergy Asthma Clin Immunol 2014; 10: 27.

7   Krauss-Etschmann S, Meyer KF, Dehmel S, et al. Inter- and transgenerational epigenetic inheritance: evidence in asthma and COPD? Clin Epigenetics 2015; 7: 53.

8   Yang IV, Pedersen BS, Liu A, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol 2015; 136: 69–80.

9   Pedersen BS, Schwartz DA, Yang IV, et al. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 2012; 28: 2986–2988.

10   Postma DS, Rabe KF. The asthma-COPD overlap syndrome. N Engl J Med 2015; 373: 1241–1249.

11   Fu JJ, McDonald VM, Baines KJ, et al. Airway IL-1β and systemic inflammation as predictors of future exacerbation risk in asthma and COPD. Chest 2015; 148: 618–629.

12   Joubert BR, Felix JF, Yousefi P, et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. Am J Hum Genet 2016; 98: 680–696.

13   Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462: 315–322.

14   Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet 2012; 13: 705–719.

15   Carithers LJ, Ardlie K, Barcus M, et al. A novel approach to high-quality postmortem tissue procurement: the GTEx Project. Biopreserv Biobank 2015; 13: 311–319.

16   Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1–13.

17   Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010; 38: W214–W220.

18   Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the Global Lung Function 2012 equations. Eur Respir J 2012; 40: 1324–1343.

19   Asher MI, Keil U, Anderson HR, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J 1995; 8: 483–491.

20   Terzikhan N, Verhamme KM, Hofman A, et al. Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. Eur J Epidemiol 2016; 31: 785–792.

21   Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012; 13: 86.

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

23   Eckhardt F, Lewin J, Cortese R, *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006; 38: 1378–1385.

24   Huang W, Loganantharaj R, Schroeder B, *et al.* PAVIS: a tool for Peak Annotation and Visualization. *Bioinformatics* 2013; 29: 3097–3099.

25   Naeem H, Wong NC, Chatterton Z, *et al.* Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* 2014; 15: 51.

26   Benjamini Y, Hochberg Y. Controlling for False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B* 1995; 57: 289–300.

27   Dai H, Leeder JS, Cui Y. A modified generalized Fisher method for combining probabilities from dependent tests. *Front Genet* 2014; 5: 32.

28   UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res* 2015; 43: D204–D212.

29   Soler Artigas M, Loth DW, Wain LV, *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082–1090.

30   Cookson W, Moffatt M, Strachan DP. Genetic risks and childhood-onset asthma. *J Allergy Clin Immunol* 2011; 128: 266–270.

31   English S, Pen I, Shea N, *et al.* The information value of non-genetic inheritance in plants and animals. *PLoS One* 2015; 10: e0116996.

32   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.

33   Golpon HA, Geraci MW, Moore MD, *et al.* HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema. *Am J Pathol* 2001; 158: 955–966.

34   Mandeville I, Aubin J, LeBlanc M, *et al.* Impact of the loss of Hoxa5 function on lung alveogenesis. *Am J Pathol* 2006; 169: 1312–1327.

35   Jehle AW, Gardai SJ, Li S, *et al.* ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J Cell Biol* 2006; 174: 547–556.

36   Poole A, Urbanek C, Eng C, *et al.* Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J Allergy Clin Immunol* 2014; 133: 670–678.

37   Hegab AE, Sakamoto T, Uchida Y, *et al.* CLCA1 gene polymorphisms in chronic obstructive pulmonary disease. *J Med Genet* 2004; 41: e27.

38   Alevy YG, Patel AC, Romero AG, *et al.* IL-13-induced airway mucus production is attenuated by MAPK13 inhibition. *J Clin Invest* 2012; 122: 4555–4568.

39   Siedlinski M, Cho MH, Bakke P, *et al.* Genome-wide association study of smoking behaviours in patients with COPD. *Thorax* 2011; 66: 894–902.

40   Wielscher M, Vierlinger K, Kegler U, *et al.* Diagnostic performance of plasma DNA methylation profiles in lung cancer, pulmonary fibrosis and COPD. *EBioMedicine* 2015; 2: 929–936.

41   George BJ, Reif DM, Gallagher JE, *et al.* Data-driven asthma endotypes defined from blood biomarker and gene expression data. *PLoS One* 2015; 10: e0117445.

42   Joubert BR, den Dekker HT, Felix JF, *et al.* Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nat Commun* 2016; 7: 10577.

43   Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 2016; 138: 16–27.

44   Nadif R, Siroux V, Boudier A, *et al.* Blood granulocyte patterns as predictors of asthma phenotypes in adults from the EGEA study. *Eur Respir J* 2016; 48: 1040–1051.

45   Bakulski KM, Feinberg JI, Andrews SV, *et al.* DNA methylation of cord blood cell types: applications for mixed cell birth studies. *Epigenetics* 2016; 11: 354–362.

46   Gervin K, Page CM, Aass HC, *et al.* Cell type specific DNA methylation in cord blood: a 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics* 2016; 11: 690–698.

47   Feng L, Wang J, Cao B, *et al.* Gene expression profiling in human lung development: an abundant resource for lung adenocarcinoma prognosis. *PLoS One* 2014; 9: e105639.

48   Xu CJ, Bonder MJ, Söderhäll C, *et al.* The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics* 2017; 18: 25.

49   Lehne B, Drong AW, Loh M, *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* 2015; 16: 37.

50   Kothari PH, Qiu W, Croteau-Chonka DC, *et al.* Role of local CpG DNA methylation in mediating the 17q21 asthma susceptibility gasdermin B (GSDMB)/ORMDL sphingolipid biosynthesis regulator 3 (ORMDL3) expression quantitative trait locus. *J Allergy Clin Immunol* 2018; 141: 2282–2286.

51   Nicodemus-Johnson J, Myers RA, Sakabe NJ, *et al.* DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* 2016; 1: e90151.

52   van Iterson M, van Zwet EW, BIOS Consortium, *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017; 18: 19.

53   Skol AD, Scott LJ, Abecasis GR, *et al.* Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; 38: 209–213.