Supplementary material

Title: Sex-specific longitudinal association of DNA methylation with lung function

Authors

Shadia Khan Sunny[1], MBBS, MPH, email: ssunny@memphis.edu

Hongmei Zhang[1], PhD, email: hzhang6@memphis.edu

Caroline L. Relton[2], PhD, email: Caroline.Relton@bristol.ac.uk

Susan Ring[2,3], PhD, email: email: S.M.Ring@bristol.ac.uk

Latha Kadalayil[4], PhD, email: lpk1r12@soton.ac.uk

Fawaz Mzayek[1], MD, MPH, PhD, email: fmzayek@memphis.edu

Susan Ewart[5], PhD, email: ewarts@cvm.msu.edu

John W. Holloway[4,6], PhD, email: J.W.Holloway@soton.ac.uk

S. Hasan Arshad[6,7,8], MBBS, MRCP, email: S.H.Arshad@soton.ac.uk

Affiliations:

[1] Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN, USA.

[2] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK.

[3] Population Health Sciences, University of Bristol, Bristol, BS8 2BN, UK.

[4] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK.

[5] Large Animal Clinical Sciences, Michigan State University, East Lansing, MI.

[6] NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Southampton, SO16 6YD, UK.

[7] Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK.

[8] The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Parkhurst Road, Newport, Isle of Wight PO30 5TG, UK.

Corresponding author:

Hongmei Zhang, PhD

Division of Epidemiology, Biostatistics, and Environmental Health Sciences,

School of Public Health, University of Memphis, Memphis, TN 38152, U.S.A.

Email: hzhang6@memphis.edu

Keywords:

Epigenome-wide, DNA methylation, longitudinal association, sex-specific effects, Population-based cohorts (IOWBC, ALSPAC), gene expression

MATERIAL AND METHODS

The IOWBC – Discovery cohort

The IOWBC is a prospective population-based birth cohort established in 1989, UK to investigate the natural history of asthma, lung function, and allergic diseases and identify genetic and environmental factors. The study was approved by the IOW Local Research Ethics Committee at recruitment and further assessments were approved by the Local/National Research Ethics Service, Committee South Central – Southampton B (06/Q1701/34). The population is largely Caucasian (~ 99%). Informed consent was obtained from parents of 1456 out of 1536 (~95%) newborns (after exclusion of adoptions, infant deaths, and non-consent) and details are described in Arshad et al. [1]. Longitudinal monitoring of diseases and assessments of environmental exposures was conducted at birth, and age 1, 2, 4, 10, 18, and 26-years with excellent retentions (~70.9% to 94.4%). This study focused on DNA-M data collected at ages 10- and 18-years, and spirometric measurements performed at ages 18- and 26-years.

Lung function

Forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1) at ages 18-years (n=838) and 26-years (n=547) were measured using a Koko spirometer and software with a portable desktop device (both PDS Instrumentation, Louisville, KY, USA) and the ratio of FEV_1 over FVC (FEV_1/FVC) was calculated. Spirometry was assessed according to the American Thoracic Society (ATS) guidelines [2, 3]. Before spirometry test, participants had to be free of respiratory infection and had not taken oral steroids for two weeks, not taken β-adrenergic

agonist medication for 6 hours, and abstained from caffeine intake for at least 4 hours

Measuring DNA Methylation (DNA-M)

Peripheral blood samples collected at ages 10 (n=330), 18 (n=476), 26 (n=303)-years from randomly selected subjects were used for DNA extraction via a standard salting out procedure [4]. 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., SanDiego, CA, USA). Arrays were processed using a standard protocol as described elsewhere [5], 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 with a detection *p*-value of less than 10^{-16} in at least 95% of samples were excluded. CpGs on sex chromosomes were also omitted to avoid potential bias in DNA-M as there are parent of origin diffrences in methylation of paternally and maternally inherited X chromosomes [6]. DNA-M data were pre-processed using the "CPACOR" pipeline for data from both platforms [7]. DNA-M intensities were quantile normalized using the R computing package, *minfi* [8]. DNA-M β values for each CpG was calculated as a ratio of methylated (M) over the sum of methylated and unmethylated (U) probes (β=M/[c+M+U]) interpreted as the

percentage of methylation [9], 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) variations. Since DNA-M data were from two different platforms (450K and EPIC), we determined the PCs based on DNA-M at shared control probes between the two platforms. The 450K 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, the top 15 of which were used to represent latent batch factors [7].

After pre-processing, a total of 473,864 and 847,155 CpGs were available in the 450K and EPIC methylation array data, respectively, with 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 ≥ 0.7% in Caucasians (corresponding to at least 10 subjects in the IOW cohort with n = 1,456) 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.

Potential Confounders

Gestational age, birth weight, sex, duration of breast feeding, maternal smoking exposure during pregnancy, recurrent chest infection at ages 1, 2 and 4-years, socioeconomic status (SES), repeated measures of height, body mass index (BMI), smoking status, and paracetamol (acetaminophen) use at ages 18 and 26-years were selected and adjusted in the model based on prior knowledge in the published literature of lung function and DNA-M [10-

13].

Information on gestational age was recorded during delivery. Birth weight was measured immediately after birth. Heights and weights at age 18 and 26-years were measured before spirometry tests and BMI was calculated accordingly. Smoking status was defined by the current, ever and never personal smoking status at age 18 and 26-years. SES was categorized using a composite "SES-cluster" based on the following three variables: (a) British socioeconomic classes (1–6) derived from parental occupation reported at birth; (b) the number of children in the index child's bedroom (collected at age 4 years); and (c) family income at age 10-years [14]. This composite variable captures the family social class across the total study period. Information on paracetamol use (frequency of taking paracetamol in a month) was collected by questionnaire at age 18-years

Gene expression (GE) data

RNA-seq GE data for subjects at age 26-years was available in IOWBC, which was used to evaluate biological relevance of identified CpGs showing longitudinal association with lung function. We used paired-end (2 × 75 bp) RNA sequencing with the Illumina Tru-Seq Stranded mRNA Library Preparation Kit with IDT for Illumina Unique Dual Index (UDI) barcode primers following manufacturer's recommendations. RNA samples were extracted from whole blood of IOWBC participants at age 26-years. All samples were sequenced a second time using the identical protocol and for each sample the output from both runs were combined. FASTQC were run to assess the quality of the FASTQ files [15]. Reads were mapped against Human Genome (GRch37 version 75) using HISAT2 (v2.1.0) aligner [16]. The alignment files, produced

in the Sequence Alignment Map (SAM) format, were converted into the Binary Alignment Map (BAM) format using SAMtools (v1.3.1) [17]. HTseq (v0.11.1) was used to count the number of reads mapped to each gene in the same reference genome used for alignment [18]. Normalized read count FPKM (Fragments Per Kilobase of transcript per Million mapped reads) were calculated using the countToFPKM package (https://github.com/AAlhendi1707/countToFPKM), and were included for subsequent data for analysis.

Replication cohort –ALSPAC

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a population-based birth cohort study established in 1991 in Avon, UK. Details of the cohort were described elsewhere [19, 20]. All pregnant women residing in the Avon region of the South West of UK during 1990–1992 were eligible to enroll in the cohort, and 14062 live newborns were recruited. All participants provided written informed consent. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Spirometry (Vitalograph 2120; Vitalograph, Maids Moreton, United Kingdom) was performed at 15 and 24-years of ages according to ATS standards [3, 12]. Information on environmental exposures, lifestyle, health of the child and family, and demographic data were collected. 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 levels in peripheral blood of children at ages 7-years (n=968) and 15-years (n=968) were assessed using the Infinium HumanMethylation450K BeadChip. The pre-

processing of DNA-M was performed by adjusting batch effect, excluding CpGs with detection *p*-value ≥0.01, and excluding samples that were flagged as sex-mismatch based on X-chromosome methylation [21]. CpGs on sex chromosomes were not included in the analyses. The participants with DNA-M at age 7-years with lung function at age 15-years and DNA-M at age 15-years with lung function at age 24-years were represented as time-lagged periods 1 and 2, respectively and included in the replication study. Details of pre-processing, quality control, and quantile normalization of DNA-M data have been described elsewhere [21, 22]. The procedure for DNA sample preparation, spirometry tests conduction, and other confounders collection were comparable to that applied in the IOWBC.

Statistical analyses

Adjustment of DNA methylation (DNA-M)

DNA-M level β values were logit-transformed to M values using log2(β value/(1- β value)) [23] due to their heteroscedasticity [9]. To adjust the impact of cellular heterogeneity of whole blood on DNA-M, different batches effects, and technical variation in the process of analyzing DNA samples, linear regression was applied with DNA-M as the outcome variable, and cell type proportions, batch information, and top 15 principal components (PCs) of the control probes were included as independent variables for each age (ages 10- and 18-years in IOWBC). Cell-type proportions (CD4+ T, CD8+ T, natural killer, B cells, monocytes, neutrophils, and eosinophils) were inferred from methylation data for each sample using the R computing package *minfi* [8, 24]. The adjusted DNA-M (or residuals from the linear regression analyses) at each CpGs were included in subsequent analyses.

training and testing (ttScreening)-based method:

A screening package, "*ttScreening"* in R 3.3.2 version [25, 26], was applied to filter out CpGs not potentially associated with lung function in either of the two periods. This method utilizes training and testing data in robust linear regressions with surrogate variables included in the regressions to adjust for unknown factor effects. The training and testing steps were repeated 100 times. The CpGs that were statistically significant in both training and testing steps at least 60 times for the longitudinal associations with lung function were included in subsequent analyses.

Analyses of differentially methylated regions (DMRs)

Regional differential methylation signals among the CpGs that passed screening for their potential association with each lung function parameter using *ttScreening*, were examined using an R package DMRcate [27]. In DMR enrichment analysis, a frequency of 20 or above was used in screening as a cutoff point to secure enough numbers of CpGs was used. The default settings in DMRcate include having ≥2 significant CpGs that passed screening in a region and a minimum length of 1000 nucleotides. A DMR was considered to be statistically significant if the FDR-adjusted $P < 0.05$ [27]. A significant DMR can be detected even if there is no genome-wide significant individual CpGs in the region.

REFERENCES

1. Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, Matthews S, Hodgekiss C, Roberts G, Kurukulaaratchy R: **Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC)**. *Int J Epidemiol* 2018, **47**(4):1043-1044i.

2. 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**:309-329.

3. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P *et al*: **Standardisation of spirometry**. *Eur Respir J* 2005, **26**(2):319-338.

4. McClelland M, Hanish J, Nelson M, Patel Y: **KGB: a single buffer for all restriction endonucleases**. *Nucleic Acids Research* 1988, **16**(1):364.

5. Bibikova M, Fan J-B: **GoldenGate® assay for DNA methylation profiling**. In: *DNA Methylation.* edn.: Springer; 2009: 149-163.

6. Golden LC, Itoh Y, Itoh N, Iyengar S, Coit P, Salama Y, Arnold AP, Sawalha AH, Voskuhl RR: **Parent-of-origin differences in DNA methylation of X chromosome genes in T lymphocytes**. *Proceedings of the National Academy of Sciences* 2019, **116**(52):26779-26787.

7. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P *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.

8. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays**. *Bioinformatics* 2014, **30**(10):1363-1369.

9. Du P, Feng G, Huang S, Kibbe WA, Lin S: **Analyze Illumina Infinium methylation microarray data**. In.; 2012.

10. Patil VK, Holloway JW, Zhang H, Soto-Ramirez N, Ewart S, Arshad SH, Karmaus W: **Interaction of prenatal maternal smoking, interleukin 13 genetic variants and DNA methylation influencing airflow and airway reactivity**. *Clinical epigenetics* 2013, **5**(1):22.

11. Weiss ST: **Lung function and airway diseases**. *Nature genetics* 2010, **42**(1):14.

12. Sonnenschein-van der Voort AM, Howe LD, Granell R, Duijts L, Sterne JA, Tilling K, Henderson AJ: **Influence of childhood growth on asthma and lung function in adolescence**. *J Allergy Clin Immunol* 2015, **135**(6):1435-1443 e1437.

13. Sunny SK, Zhang H, Rezwan FI, Relton CL, Henderson AJ, Merid SK, Melén E, Hallberg J, Arshad SH, Ewart S *et al*: **Changes of DNA methylation are associated with changes in lung function during adolescence**. *Respiratory Research* 2020, **21**(1):80.

14. 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):62-66.

15. **A quality control tool for high throughput sequence data** [https://www.bioinformatics.babraham.ac.uk/projects/fastqc/]

16. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nature methods* 2015, **12**(4):357-360.

17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

18. Anders S, Pyl PT, Huber W: **HTSeq--a Python framework to work with high-throughput sequencing data**. *Bioinformatics* 2015, **31**(2):166-169.

19. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G: **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):111-127.

20. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A *et al*: **Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort**. *Int J Epidemiol* 2013, **42**(1):97-110.

21. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, Woodward G, Lyttleton O, Evans DM, Reik W *et al*: **Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)**. *International journal of epidemiology* 2015, **44**(4):1181-1190.

22. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M: **Meffil: efficient normalization and analysis of very large DNA methylation datasets**. *Bioinformatics* 2018, **34**(23):3983-3989.

23. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis**. *BMC Bioinformatics* 2010, **11**:587.

24. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT: **DNA methylation arrays as surrogate measures of cell mixture distribution**. *BMC Bioinformatics* 2012, **13**:86.

25. Li X, Hawkins GA, Ampleford EJ, Moore WC, Li H, Hastie AT, Howard TD, Boushey HA, Busse WW, Calhoun WJ *et al*: **Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients**. *J Allergy Clin Immunol* 2013, **132**(2):313-320 e315.

26. Ray MA, Tong X, Lockett GA, Zhang H, Karmaus WJ: **An Efficient Approach to Screening Epigenome-Wide Data**. *Biomed Res Int* 2016, **2016**:2615348.

27. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RV, Clark SJ, Molloy PL: **De novo identification of differentially methylated regions in the human genome**. *Epigenetics & chromatin* 2015, **8**(1):6.

Supplementary Table:

Table S1: DNA-M at CpGs (k=14) at an earlier age associated with lung function at a later age in IOWBC.

Lung function	CpGs Name	Chr. No.	Gene	Location	IOWBC			ALSPAC	
					Coeff.	P_{RAW}	P_{FDR}	Coeff	*P*
FVC	cg07991621	4	*SH3BP2*	Body	0.17	0.0002	0.010	-0.097	0.73
	cg13394305	2	*SLC40A1*	TSS200	-0.058	0.0001	0.010	15.23	0.056
	cg21492378	9	*CEP110*	TSS1500	0.53	<.0001	0.001	-0.016	0.91
FEV_1	cg10729557	14	*ANKRD9*	Intergenic	-0.18	0.0008	0.035	-2.02	0.22
	cg16582803	2	*WNT10A*	Intergenic	-0.16	0.0002	0.028	-0.59	0.034
	cg17315331	11	*HINFP*	TSS200	0.06	0.0003	0.028	1.60	0.84
	cg21556039	21	*C21orf58*	Body	-0.19	0.0008	0.035	0.58	0.17
	cg26174454	19	*EFNA2*	Intergenic	0.23	0.0007	0.035	0.44	0.15
	cg27599129	7	*ZNF727*	Intergenic	-0.23	0.001	0.036	-0.061	0.86
FEV_1/FVC	cg11401293	21	*COL6A1*	Body	-0.02	0.0007	0.041	0.022	0.74
	cg12614529	4	*MND1*	Body	0.060	0.0012	0.041	-0.031	0.35
	cg18760835	6	*NRN1*	Intergenic	-0.01	0.0004	0.041	-0.22	0.33
	cg21240861	7	*DNAJB6*	TSS200	-0.004	0.0014	0.041	-0.045	0.95
	cg27601198	16	*C16orf87*	Intergenic	0.01	0.0011	0.041	0.083	0.28

Note to table 2: 1) Coefficients of CpGs for the association of DNA-M at earlier age with lung function at a later age in IOWBC.

2) *Genes located at intergenic location were not found in Illumina annotation file and were identified using online tool SNIPPER.

3) Chr. no. = Chromosome number; Coeff. = Coefficients.

Table S2: DNA-M at CpGs (k=23) at an earlier age associated with lung function at a later age in IOWBC.

CpGs	Sex	Chr. no.	Gene name	Location	IOWBC Coeff.	P_{RAW}	P_{FDR}	ALSPAC Coeff.	*P*
FVC									
cg01376079		11	*SSH3*	TSS1500	0.085	0.38		0.22	0.49
cg01376079*sex	Male				-0.41	0.0019	0.042	-0.71	0.12
cg07230380		10	*SCD*	TSS1500	-0.015	0.25		-4.77	0.62
cg07230380*sex	Male				0.052	0.0012	0.042	16.21	0.23
cg07557690		1	*TGFBR3*	TSS1500	0.31	0.0014		0.21	0.56
cg07557690*sex	Male				-0.41	0.0017	0.042	-0.40	0.46
cg10123952		3	*ALCAM*	Intergenic	-0.36	0.001		0.38	0.29
cg10123952*sex	Male				0.47	0.0019	0.042	-0.68	0.12
cg12040830		11	*NCAM1*	Body	-0.075	0.0322		1.54	0.42
cg12040830*sex	Male				0.14	0.0011	0.042	-3.05	0.23
cg14083603		20	*ZGPAT*	Body	-0.49	<.0001		-0.55	0.76
cg14083603*sex	Male				0.69	0.0007	0.042	6.22	0.018
cg15757271		3	*WNT5A*	TSS1500	-0.31	0.0024		-0.016	0.99
cg15757271*sex	Male				0.49	0.0006	0.042	-5.76	0.039
cg19476368		11	*MIR100HG*	Intergenic	-0.21	0.0486		-0.30	0.82
cg19476368*sex	Male				0.59	0.0006	0.042	-0.92	0.60
cg23026420		11	*PPP2R1B*	TSS200	0.17	0.0059		2.96	0.66
cg23026420*sex	Male				-0.23	0.0019	0.042	-11.86	0.19
FEV_1									
cg05849324		6	*NHLRC1*	1stExon	-0.12	0.19		0.18	0.62
cg05849324*sex	Male				0.47	0.0003	0.016	0.10	0.86
cg09205595		7	*AGAP3*	TSS1500	-0.40	0.0004		0.22	0.59
cg09205595*sex	Male				0.63	0.0002	0.016	0.38	0.51
cg13468252		1	*C1orf128*	TSS1500	-0.013	0.47		1.38	0.76
cg13468252*sex	Male				0.083	0.001	0.032	-10.55	0.085
cg15981851		2	*AGAP1*	Body	0.071	0.27		0.10	0.77
cg15981851*sex	Male				-0.361	0.0002	0.0156	-0.057	0.91
cg16582803		2	*WNT10A*	Intergenic	-0.018	0.7701		-0.48	0.21
cg16582803*sex	Male				-0.287	0.0009	0.0320	-0.23	0.67
cg19736286		2	*MSH6*	TSS200	0.093	0.4143		0.23	0.90
cg19736286*sex	Male				-0.562	0.0011	0.0320	-1.36	0.59
cg20804831		3	*NUDT16P*	TSS200	-0.211	0.0612		0.23	0.69
cg20804831*sex	Male				0.725	<.0001	0.0060	-0.54	0.54
FEV_1/FVC									
cg02397934		6	*H2BC13*	Intergenic	-0.010	0.0354		-0.094	0.46
cg02397934*sex	Male				0.033	0.0004	0.0236	0.21	0.24
cg02466892		3	*ABI3BP*	Intergenic	0.049	0.0015		-0.017	0.57
cg02466892*sex	Male				-0.069	0.002	0.0327	0.014	0.76

cg04199473		14	*STRN3*	Body	0.009	0.2186		-0.060	0.34
cg04199473*sex	Male				-0.041	0.0013	0.0270	0.061	0.52
cg08650125		8	*LY6H*	Intergenic	0.039	0.0014		0.010	0.73
cg08650125*sex	Male				-0.058	0.0005	0.0236	-0.029	0.50
cg09010372		4	*MAEA*	Body	-0.043	0.0232		-0.012	0.80
cg09010372*sex	Male				0.089	0.0016	0.0295	-0.037	0.63
cg09059988		1	*HORMAD1*	1stExon; 5'UTR	0.011	0.3599		-0.012	0.89
cg09059988*sex	Male				-0.059	0.0008	0.0247	-0.030	0.80
cg11579646		6	*LOC154449*	Intergenic	-0.023	0.0042		0.12	0.49
cg11579646*sex	Male				0.036	0.0023	0.0343	-0.17	0.50
cg18499321		12	*RIMBP2*	Body	0.033	0.0296		0.014	0.77
cg18499321*sex	Male				-0.070	0.001	0.0260	0.038	0.58
cg20038169		19	*GMIP*	1stExon; 5'UTR	-0.017	<.0001		-0.59	0.66
cg20038169*sex	Male				0.020	0.0007	0.0247	1.82	0.34
cg23370466		5	*TRIM41*	Intergenic	-0.021	0.0045		-0.032	0.43
cg23370466*sex	Male				0.039	0.0002	0.0236	0.040	0.49

Note to table S2: 1) Coefficients of CpGs for the sex-specific association of DNA-M at earlier age with lung function at a later age in IOWBC.

2) *Genes located at intergenic location were not found in Illumina annotation file and were identified using online tool SNIPPER.

3) Chr. no. = Chromosome number; Coeff. = Coefficients

Table S3: DMRs of lung function at later age in relation to DNA-M at earlier age identified by DMRcate method

Lung function	Molecular location of DMR (chromosome: start – end)	No. CpGs	Stouffer	Annotated Gene
Male				
FEV_1				
	chr4: 1004525-1004678	3	0	*FGFRL1*
	chr6: 33084825-33085031	3	0	*HLA-DPB2*
	chr10: 100993826-100994478	2	0	*HPSE2*
	chr11: 105948706-105949099	2	0	*AASDHPPT, KBTBD3*
	chr12: 49259786-49259997	2	0	*RND1*
	chr15: 52970418-52971181	2	0	*FAM214A*
	chr17: 27400787-27401144	2	0	*MYO18A*
	chr17: 61904053-61905004	2	0	*PSMC5, FTSJ3*
	chr19: 11669574-11669730	2	0	*ELOF1, ZNF627*
	chr20: 62710905-62711729	2	0	*RGS19, OPRL1*
	chr7: 93204985-93205240	2	0	*CALCR*
	chr10: 94826314-94826319	2	1.6933E-307	*CYP26C1*
	chr6: 30029232-30029760	2	5.5778E-303	*ZNRD1*
	chr1: 43920090-43920103	2	9.3906E-294	*HYI*
	chr13: 112870385-112870414	2	1.7669E-272	*SOX1*
	chr11: 65601265-65601301	2	6.0677E-270	*SNX32*
	chr2: 219157103-219157119	2	3.4719E-264	*TMBIM1*
	chr1: 27960788-27961680	2	4.4378E-242	*FGR*
	chr19: 523300-523360	2	1.1118E-224	*TPGS1*
	chr15: 30163660-30163825	3	1.9556E-203	*TJP1*
	chr4: 7033722-7033761	2	8.2315E-183	*LOC100129931*
	chr15: 40364524-40364740	2	5.29417E-89	*BMF*
	chr6: 2891973-2892150	3	9.07093E-66	*SERPINB9*
	chr1: 17023008-17023283	2	1.28724E-45	*ESPNP*
FVC				
	chr4: 2819770-2820479	4	0	*SH3BP2*
	chr8: 11659832-11660733	3	0	*FDFT1, RP11-297N6.4*
	chr2: 69614945-69615105	2	0	*GFPT1*
	chr3: 51975220-51976003	2	0	*RRP9, PARP3*
	chr5: 1949480-1950271	2	0	*IRX4*
	chr5: 176943423-176943966	2	0	*DDX41*
	chr6: 89827135-89827915	2	0	*SRSF12*
	chr7: 29605897-29606082	2	0	*PRR15-002*
	chr8: 120220410-120221268	2	0	*MAL2*
	chr2: 198649783-198650123	2	4.027E-205	*BOLL*
	chr3: 194014481-194014745	3	1.2594E-198	*CPN2*
	chr1: 102312608-102312610	2	1.129E-190	*OLFM3*

	chr6: 144386416-144386457	2	8.3492E-174	*PLAGL1*
	chr19: 519609-519611	2	1.7209E-149	*IRX4*
	chr19: 19626525-19626576	2	5.56579E-84	*NDUFA13, TSSK6, NDUFA13, YJEFN3*
	chr17: 1395864-1395880	2	2.30885E-80	*MYO1C*
	chr6: 110720918-110721138	2	3.51751E-44	*DDO*
FEV_1/FVC				
	chr6: 30038929-30039435	10	0	*RNF39*
	chr19: 1467008-1467032	3	0	*APC2*
	chr1: 35544839-35545196	2	0	*ZMYM1*
	chr10: 132239568-132239652	2	0	*NA*
	chr14: 50319271-50319614	2	0	*NEMF, RN7SL3*
	chr16: 24740859-24740939	2	0	*TNRC6A*
	chr2: 128458399-128458845	2	0	*SFT2D3*
	chr3: 129147541-129147553	2	0	*EFCAB12*
	chr4: 95128817-95128914	2	0	*SMARCAD1, RP11-363G15.2*
	chr6: 146285012-146285424	2	0	*SHPRH*
	chr8: 124252970-124253478	2	0	*C8orf76*
	chr19: 292167-292245	2	5.1994E-291	*PPAP2C*
	chr6: 33871907-33872861	3	4.7088E-188	*MIR1275*
	chr7: 1003645-1003750	2	2.6165E-181	*COX19*
	chr15: 23115232-23115432	2	1.9713E-112	*RP11-566K19.6*
	chr19: 613433-613505	2	1.44724E-67	*HCN2*
Female				
FVC				
	chr1: 92352293-92352481	3	0	*TGFBR3*
	chr10: 135191624-135192230	3	0	*PAOX, AL360181.1-201*
	chr11: 2322500-2322808	3	0	*TSPAN32*
	chr20: 25677290-25677582	3	0	*ZNF337*
	chr1: 36948570-36949518	2	0	*CSF3R*
	chr1: 206808936-206809102	2	0	*DYRK3*
	chr1: 213223461-213224450	2	0	*RPS6KC1*
	chr11: 2444485-2445216	2	0	*TRPM5*
	chr11: 124767720-124768554	2	0	*ROBO4*
	chr12: 99038290-99038766	2	0	*IKBIP, APAF1*
	chr13: 36050158-36050993	2	0	*MAB21L1, NBEA*
	chr13: 111301379-111301576	2	0	*CARS2*
	chr14: 90421085-90422082	2	0	*EFCAB11, TDP1*
	chr15: 43809689-43809865	2	0	*MAP1A*
	chr16: 30615018-30615808	2	0	*ZNF689*
	chr16: 51185346-51185772	2	0	*SALL1*
	chr16: 57278759-57279645	2	0	*ARL2BP, RP11-*
	chr19: 17666205-17666514	2	0	*COLGALT1*
	chr19: 19754386-19755321	2	0	*GMIP*

	chr21: 45926167-45926719	2	0	*TSPEAR-AS1*
	chr22: 31477112-31477330	2	0	*SMTN*
	chr3: 50311211-50311213	2	0	*SEMA3B*
	chr3: 55521351-55521789	2	0	*WNT5A*
	chr5: 68665393-68665965	2	0	*TAF9, RAD17*
	chr5: 139554569-139555269	2	0	*CYSTM1*
	chr6: 31595653-31595725	2	0	*PRRC2A*
	chr6: 31646077-31646262	2	0	*LY6G5C*
	chr6: 31831489-31831599	2	0	*NEU1*
	chr7: 22539741-22539822	2	0	*STEAP1B*
	chr7: 95951432-95951712	2	0	*SLC25A13*
	chr15: 40401038-40401272	2	2.2609E-296	*BMF*
	chr4: 6729081-6729744	2	3.1708E-288	*ZNF689*
	chr21: 27944586-27944779	2	4.6994E-284	*CYYR1*
	chr13: 110961330-110961606	2	1.2844E-281	*COL4A2*
	chr22: 30476089-30476525	5	7.7543E-280	*HORMAD2, CTA*
	chr14: 36278529-36278684	2	8.8314E-270	*RALGAPA1, AL162311.1*
	chr7: 20818725-20818928	2	1.808E-265	*SP8*
	chr5: 112824497-112824765	3	7.9303E-254	*MCC*
	chr4: 184961220-184961374	2	7.4488E-235	*STOX2*
	chr19: 1009048-1009949	2	2.5482E-222	*TMEM259*
	chr17: 56744332-56744490	3	9.774E-216	*RNU1-108P*
	chr11: 128693961-128694915	2	2.5621E-185	*FLI1*
	chr6: 32729563-32729647	2	6.3305E-143	*HLA-DQB2*
	chr10: 34408530-34408654	3	3.5373E-139	*PARD3*
	chr4: 110625010-110625080	2	2.272E-137	*CASP6*
	chr7: 150038598-150038898	2	1.5089E-132	*RARRES2, RP4-584D14.7*
	chr11: 1785618-1785631	2	5.2265E-125	*CTSD, RP4-584D14.7*
	chr9: 123605570-123605666	2	2.8072E-123	*PSMD5*
	chr3: 142666320-142666476	2	2.862E-119	*PAQR9*
	chr6: 32294470-32294503	2	6.2162E-114	*HNRNPA1P2*
	chr6: 32202748-32202844	2	2.1001E-109	*NOTCH4*
	chr15: 99975310-99975470	3	1.2901E-107	*LRRC28*
	chr13: 106063138-106063150	2	1.5181E-96	*DAOA*
	chr6: 88757302-88757392	5	7.45651E-74	*SPACA1*
	chr1: 227746882-227747268	2	2.53884E-63	*RNA5SP77*
	chr10: 90985055-90985062	2	7.80545E-23	*LIPA*
	chr6: 110721138-110721349	2	2.9673E-19	*DDO*
FEV_1				
	chr17: 45949743-45949878	4	0	*SP6*
	chr6: 32016257-32017229	4	0	*TNXB*
	chr7: 157512397-157513707	3	0	*PTPRN2*
	chr1: 51434014-51434666	2	0	*CDKN2C*
	chr1: 92352293-92352407	2	0	*TGFBR3*

chr10: 135049999-135050355	2	0	*VENTX*
chr11: 3862089-3862297	2	0	*RHOG*
chr12: 8995591-8995660	2	0	*A2ML1*
chr12: 99038639-99038766	2	0	*IKBIP, APAF1*
chr12: 121947315-121947522	2	0	*KDM2B*
chr14: 61116382-61117162	2	0	*SIX1*
chr14: 103058807-103058815	2	0	*RCOR1*
chr16: 23568656-23569246	2	0	*UBFD1, EARS2*
chr16: 51185346-51185772	2	0	*SALL1*
chr17: 26988607-26989222	2	0	*SDF2, UPT6H*
chr18: 579237-580188	2	0	*CETN1*
chr2: 26915349-26915355	2	0	*KCNK3*
chr2: 217498574-217499384	2	0	*IGFBP2*
chr20: 62179030-62179752	2	0	*SRMS*
chr21: 45926167-45926719	2	0	*TSPEAR-AS1*
chr3: 42543161-42544067	2	0	*VIPR1*
chr4: 178528415-178528594	2	0	*AGA*
chr6: 10886999-10887023	2	0	*SYCP2L*
chr6: 26537980-26538671	2	0	*HMGN4*
chr6: 31515296-31515404	2	0	*ATP6V1G2, NFKBIL1*
chr6: 31595653-31595725	2	0	*PRRC2A*
chr6: 31633420-31634141	2	0	*CSNK2B, GPANK1*
chr6: 32185954-32185995	2	0	*NOTCH4*
chr9: 94486741-94487105	2	0	*ROR2*
chr18: 21017905-21018217	2	2.40E-294	*TMEM241*
chr19: 5478473-5478484	2	2.03E-281	*ZNRF4*
chr11: 67286645-67287418	2	5.30E-273	*CABP2*
chr7: 26676482-26677374	2	1.23E-267	*C7orf71*
chr1: 154298543-154298956	2	5.69E-260	*ATP8B2*
chr17: 48912543-48912545	2	9.83E-260	*WFIKKN2*
chr9: 124982413-124982834	2	3.04E-247	*LHX6*
chr18: 77289084-77289104	2	5.98E-242	*NFATC1*
chr17: 44343683-44343776	2	8.17E-231	*RP11*
chr2: 190044636-190044638	2	6.12E-228	*COL5A2*
chr7: 150037890-150038898	3	3.57E-219	*RARRES2, RP4*
chr6: 30509642-30510300	2	3.40E-197	*GNL1*
chr6: 32294470-32294577	3	1.28E-190	*HNRNPA1P2*
chr16: 54967714-54967786	2	3.34E-184	*IRX5, CTD*
chr22: 30476089-30476525	4	6.51E-182	*HORMAD2, CTA*
chr20: 62328084-62328427	3	7.15E-181	*TNFRSF6B*
chr18: 72916776-72917101	2	4.88E-168	*ZADH2*
chr10: 34408530-34408654	3	2.31E-139	*PARD3*
chr9: 123605570-123605666	2	5.93E-118	*PSMD5*
chr16: 86795398-86795490	2	8.52E-115	*FOXL1*

	chr20: 61905223-61905353	2	1.50E-110	*ARFGAP1, NKAIN4*
	chr16: 1133168-1133172	2	1.83E-110	*SSTR5*
	chr6: 32202748-32202844	2	2.18E-109	*NOTCH4*
	chr15: 99975310-99975470	3	4.27E-108	*LRRC28*
	chr8: 37605517-37605783	3	6.66E-95	*RP11*
	chr6: 160023626-160024144	2	3.74E-92	*SOD2*
	chr17: 154420-154671	3	1.37E-88	*RPH3AL*
	chr5: 1867978-1868693	5	2.75E-84	*IRX4*
	chr11: 18477303-18477379	2	1.34E-68	*LDHAL6A*
	chr21: 34405681-34405997	2	3.71E-57	*OLIG2*
	chr16: 90016004-90016020	2	2.87E-56	*DEF8*
	chr9: 130955380-130955436	2	7.59E-52	*CIZ1*
	chr11: 109785847-109786133	2	1.45E-41	*ZC3H12C*
	chr11: 18433554-18433745	4	5.93E-39	*LDHC*
	chr6: 110720918-110721349	4	8.70E-26	*DDO*
	chr8: 144120106-144120335	2	2.07E-21	*C8orf31*
	chr6: 88757358-88757392	2	5.77E-18	*SPACA1*
FEV_1/FVC				
	chr1: 2058230-2059086	3	0	*PRKCZ*
	chr17: 40936570-40937362	3	0	*WNK4*
	chr20: 44829602-44829821	3	0	*CDH22*
	chr1: 64014340-64014796	2	0	*DLEU2L, EFCAB7*
	chr1: 147245485-147245494	2	0	*GJA5*
	chr1: 155265026-155265033	2	0	*PKLR*
	chr10: 102045959-102046263	2	0	*BLOC1S2*
	chr10: 102988831-102989311	2	0	*LBX1*
	chr11: 57157508-57157632	2	0	*PRG2*
	chr11: 89956573-89956708	2	0	*CHORDC1*
	chr13: 34392073-34392492	2	0	*RFC3*
	chr15: 25683909-25684085	2	0	*UBE3A*
	chr15: 49170643-49170751	2	0	*SHC4, AC012379.1, EID1*
	chr16: 21289268-21289812	2	0	*CRYM*
	chr17: 1953268-1953382	2	0	*MIR212*
	chr17: 57784674-57784779	2	0	*VMP1, PTRH2*
	chr19: 46521569-46522090	2	0	*MIR769, CCDC61*
	chr2: 397730-398382	2	0	*ALKAL2*
	chr3: 14166507-14167245	2	0	*CHCHD4, TMEM43*
	chr3: 172165696-172166517	2	0	*GHSR*
	chr5: 1794232-1794420	2	0	*MRPL36*
	chr6: 33267268-33267505	2	0	*RGL2*
	chr7: 150756561-150757210	2	0	*CDK5, SLC4A2*
	chr8: 56015399-56015750	2	0	*XKR4*
	chr8: 125740451-125740636	2	0	*MTSS1*
	chr8: 144599027-144599347	2	2.0952E-300	*ZC3H3*

	chr22: 39713018-39713086	2	1.5662E-295	*SNORD83A, RPL3*
	chr20: 35504198-35504371	2	2.8022E-280	*TLDC2*
	chr5: 143191565-143191663	2	8.8312E-273	*HMHB1*
	chr10: 125033770-125034002	2	8.806E-248	*BUB3*
	chr6: 26195488-26195995	2	7.1254E-240	*HIST1H3D*
	chr10: 105344798-105344807	2	2.1525E-239	*NEURL1*
	chr7: 2143886-2143942	2	2.2096E-221	*MAD1L1*
	chr1: 153599479-153600064	2	7.8863E-216	*S100A1*
	chr10: 131567828-131568735	2	1.2979E-205	*RP11*
	chr20: 17943403-17943694	2	6.9542E-198	*SNORD17, SNX5*
	chr2: 3486749-3487164	2	6.6271E-175	*TRAPPC12-AS1*
	chr5: 66462471-66462662	2	1.4278E-166	*MAST4*
	chr17: 78863570-78863674	2	2.8514E-163	*RPTOR*
	chr2: 113993052-113994035	3	3.0569E-156	*PAX8-AS1*
	chr8: 144260671-144260730	2	2.1786E-144	*LY6H*
	chr7: 57471759-57472367	3	1.7943E-131	*MIR3147*
	chr10: 369977-370009	2	7.4426E-103	*DIP2C*
	chr12: 58012601-58013109	2	1.0159E-95	*SLC26A10, AC025165.8*
	chr3: 113234510-113235015	2	1.83952E-49	*SPICE1*
	chr1: 43814764-43815035	3	2.22175E-21	*MPL*
	chr1: 205819463-205819492	2	0.000863876	*PM20D1*

Note to table S3:

1) DMRcate annotates to UCSC RefGene from the Illumina annotation file.