Genetic and epigenetic associations with pre-COPD lung function trajectories.

David J Martino[1], Dinh S Bui[2], Shuai Li[2,3,4,5], Sabrina Idrose[2], Jennifer Perret[2], Adrian Lowe[2], Caroline Lodge[2], Gayan Bowatte[2], Yuben Moodley[6], Paul S Thomas[7], Graeme Zosky[8], Phillip Hansbro[9], John W Holloway[10], Cecile Svanes [11,12], Rosa Faner[13,14,15], Eugene H Walters [2,16*] and Shaymali C Dharmage [2] *.

** Joint Senior authors*

1 Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia.

2 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne

3 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge

4 Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

5 Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria 3051, Australia

6 Department of Respiratory Medicine, Fiona Stanley Hospital, Murdoch, Australia

7 Respiratory Medicine Prince of Wales' Hospital & Prince of Wales' Clinical School, UNSW, Randwick, NSW 2031, Australia

8 Menzies Institute for Medical Research, University of Tasmania

9 Centre for Inflammation, Centenary Institute and University of Technology Sydney, School of Life Sciences, Faculty of Science, Sydney NSW 2007, Australia

10 Human Development and Health, Faculty of Medicine, University of Southampton, Southampton UK.

11 Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway.

12 Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway.

13 University of Barcelona, Immunology Unit, Biomedicine Department. Spain.

14. IDIBAPS-FCRB Institut d'Investigacions Biomediques August Pi i Sunyer-Fundació Clinic Recerca, Barcelona, Spain.

15. Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBER), Spain.

16 School of Medicine, University of Tasmania and School of Population and Global Health, University of Melbourne

Running title: Genomics of pre-COPD lung function trajectories.

Corresponding author: David Martino

Telethon Kids Institute

PO Box 855, West Perth, Western Australia, 6872

Abbreviations

COPD – Chronic Obstructive Pulmonary Disease

CpG – Cytosine phosphate Guanine

DMR – Differentially Methylated Region

GWAS – Genome Wide Association Study

mQTL – Methylation Quantitative Trait Loci

SNP – Single Nucleotide Polymorphism

TAHS – Tasmanian Longitudinal Health Study

To the Editor,

Understanding the molecular mechanisms of lung function trajectories that progress to chronic obstructive pulmonary disease (COPD) (pre-COPD trajectories), especially those with a rapidly declining phenotype, should inform preventive interventions. The Tasmanian Longitudinal Health Study (TAHS) previously defined life-course lung function trajectories by serial spirometry in a cohort of all seven-year-old school children in the state of Tasmania recruited in 1968 and followed up to age 53 years (1). Of the six pre-bronchodilator FEV_1 lifetime trajectories identified, three collectively accounted for 75% of chronic obstructive pulmonary disease (COPD) prevalence at age 53 years (2). These high-risk trajectories were: 1) early below average lung function (with usual rate of subsequent decline), 2) persistently low, and 3) early below average lung function with accelerated decline. The TAHS cohort provides a unique opportunity to investigate molecular factors associated with disadvantaged trajectories, and we conducted a pilot study in this cohort to characterize associations with COPD high-risk trajectories to inform more extensive longitudinal studies in the future.

The rationale for our approach was based on previous studies that demonstrate declining lung function (3-6) and lower lung function associated with COPD (7, 8) are complex phenotypes involving the interplay between genomic and environmental factors. Genetic (9) and epigenetic associations (3, 7, 10) have been previously been described for disadvantaged lung function trajectories, including epigenetic aging related phenotypes (epigenetic age acceleration) (5, 10). We therefore conducted a screen of both epigenetic (DNA methylation) and genetic (single nucleotide polymorphisms; SNPs) markers on available whole blood samples collected at the 45-year follow-up. We used an extremes of phenotype design to maximise power randomly selecting 80 subjects from across the three high-risk trajectories

and matching on age and smoking status to 80 subjects from the persistently high trajectory. By design, individuals belonging to the different lung function trajectories varied in post bronchodilator FEV_1, but also steroid medication use, and sex so these differences were examined in adjusted models. We quantified 787,111 DNA methylation markers (CpGs) and 4,456,571 SNPs using the InfiniumMethylationEPIC (v1) and Infinium Global Screening Array (v3) genotyping microarrays. Some of the results of this study have been previously reported in the form of an abstract (11).

We detected DNA methylation differences at 55 differentially methylated regions (DMRs) containing 73 unique genes and 6 non-coding regions (FDR adjusted $P < 0.05$; Figure 1A). Notable genes in DMRs included *LY6G5C and HLA-DQB1* within the major histocompatibility complex, HOX cluster transcription factors (*HOXB-AS3, HOXB3, HOXB6)* which have been implicated in the pathogenesis of pulmonary diseases (12, 13), and transmembrane glycoproteins (*LGALS3BP, OCA2, KCNE1, PTPRN2, TNXB, PCDHGA5, CDSN, PCDHGA4, PCDHGA3, PCDHGB3, PCDHGA2, PCDHGB2, PCDHGA1, EGFR, DPP6, FOLH1, SGCD, CRTAC1, PCDHGB1, FIBIN, CHST1, MUC4, DPEP3)* that play a role in epithelial biology and when disrupted may lead to EMT (14).

This DMR signature was only partially consistent across the high-risk trajectory sub-groups (24% of DMRs shared across all sub-groups), whereas sub-group specific regions predominated (Figure 1B) consistent with the notion of COPD risk factor etiotypes that likely exhibit different molecular drivers (15). Current COPD or current asthma explained 17-30% of methylation differences across the DMRs respectively, but sex and blood cell counts were not mediators or confounders of these associations. Integrating the genetic and epigenetic data sets we performed methylation quantitative trait (mQTL) mapping and found that genetic variation at 381 nearby SNPs (+/- 500kb of DMRs) in 17 genomic loci were associated with

23% of CpGs within DMR regions (Figure 1C). The strongest mQTL region was on chromosome 6 at the major histocompatibility locus. Using publicly available tissue specific gene expression signatures (GTEx catalogue v8) we determined these mQTL SNPs were statistically enriched among transcripts primarily expressed in the lung (ATP13A4, MUC4, PSORS1C1) (Figure 1D). Several mQTLs have previously been associated with Lung function phenotypes (HAPLN1, HLA-DRB1, HLA-DQA1, HLA-DQB1), COPD (HLA-DQB1, HLA-DQA1) and Asthma risk (HLA-DQB1, HLA-DQA1, HLA-DRB1, HLA-DRB6, PSORS1C1) in the genome-wide association study (GWAS) catalogue. Consistent with previous studies (5, 10) we also found that epigenetic age predictions were significantly higher in the high-risk group when measured using the phenoAge algorithm (16) (Table 1). On average, individuals in the high-risk category had increased mean predicted chronological age of 1.5 years relative to controls (40.2 v 38.7, P=0.03, t-test). Stratified analysis suggested age-acceleration was strongest in the early below average, accelerated decline group (Beta = 2.1, P=0.06 v 1.4, P=0.11, *below average*; v 1.1, P=0.19, *persistently low*) although sample size was a limiting factor.

To our knowledge this was the first epigenome-wide association analysis in individuals from COPD-risk lung function trajectories, providing a strong foundation for further delineation of phenotypes and risk factors to enable precision molecular profiling. We determined blood to be a phenotypically relevant tissue to explore molecular associations with life-time lung function trajectories in this cohort. Although causality of the epigenetic associations cannot be established in this pilot, a subset of epigenetic changes in the high-risk trajectory were mQTLs whereby genetic variants affected the methylation patterns at these genes. Since the causality of genetic variation on DNA methylation levels is uni-directional, these analyses aid in prioritization of methylation-trait associations from epigenome-wide scans. The mQTL associations are compelling candidates for gene-environment interactions, and might be

linked to early life events, as well as processes related to disease progression. Confirmatory longitudinal are now planned to dissect these environmental and host genomic risk factors that are reflected in the epigenome. Our analysis of DMR sharing across sub-groups suggests molecular risk factors will be unique across different life-time lung function trajectories warranting follow-up studies at cohort-wide scale. We also determined that comorbidities including current asthma and COPD explained a proportion of variation in the blood epigenetic markers. Consistent with previous reports for declining lung function and epigenetic ageing phenotypes (5, 10), we also found that epigenetic age acceleration was detectable in the high-risk trajectory group, the biology of which is still poorly understood. In summary this pilot study confirms the utility of our approach and paves the way for future profiling studies in this unique cohort. An enhanced understanding of molecular risk factors associated with disadvantaged trajectories will enable more precise biomarker-driven interventions in the future with potential to redirect the course of respiratory health in vulnerable individuals.

References

1. Matheson MC, Abramson MJ, Allen K, Benke G, Burgess JA, Dowty JG, Erbas B, Feather IH, Frith PA, Giles GG, Gurrin LC, Hamilton GS, Hopper JL, James AL, Jenkins MA, Johns DP, Lodge CJ, Lowe AJ, Markos J, Morrison SC, Perret JL, Southey MC, Thomas PS, Thompson BR, Wood-Baker R, Haydn Walters E, Dharmage SC, group Ti. Cohort Profile: The Tasmanian Longitudinal Health STUDY (TAHS). *Int J Epidemiol* 2017; 46: 407-408i.
2. Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, Bowatte G, Gurrin L, Johns DP, Thompson BR, Hamilton GS, Frith PA, James AL, Thomas PS, Jarvis D, Svanes C, Russell M, Morrison SC, Feather I, Allen KJ, Wood-Baker R, Hopper J, Giles GG, Abramson MJ, Walters EH, Matheson MC, Dharmage SC. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Medicine* 2018; 6: 535-544.
3. Ngo D, Pratte KA, Flexeder C, Petersen H, Dang H, Ma Y, Keyes MJ, Gao Y, Deng S, Peterson BD, Farrell LA, Bhambhani VM, Palacios C, Quadir J, Gillenwater L, Xu H, Emson C, Gieger C, Suhre K, Graumann J, Jain D, Conomos MP, Tracy RP, Guo X, Liu Y, Johnson WC, Cornell E, Durda P, Taylor KD, Papanicolaou GJ, Rich SS, Rotter JI, Rennard SI, Curtis JL, Woodruff P, Comellas AP, Silverman EK, Crapo JD, Larson MG, Vasan RS, Wang TJ, Correa A, Sims M, Wilson JG, Gerszten RE, O'Connor GT, Barr RG, Couper D, Dupuis J, Manichaikul A, O'Neal W, Tesfaigzi Y, Schulz H, Bowler R. Systemic Markers of Lung Function and FEV(1) Decline across Diverse Cohorts. *Ann Am Thorac Soc* 2023.
4. Lee M, Huan T, McCartney DL, Chittoor G, de Vries M, Lahousse L, Nguyen JN, Brody JA, Castillo-Fernandez J, Terzikhan N, Qi C, Joehanes R, Min JL, Smilnak GJ, Shaw JR, Yang CX, Colicino E, Hoang TT, Bermingham ML, Xu H, Justice AE, Xu CJ, Rich SS, Cox SR, Vonk JM, Prokic I, Sotoodehnia N, Tsai PC, Schwartz JD, Leung JM, Sikdar S, Walker RM, Harris SE, van der Plaat DA, Van Den Berg DJ, Bartz TM, Spector TD, Vokonas PS, Marioni RE, Taylor AM, Liu Y, Barr RG, Lange LA, Baccarelli AA, Obeidat M, Fornage M, Wang T, Ward JM, Motsinger-Reif AA, Hemani G, Koppelman GH, Bell JT, Gharib SA, Brusselle G, Boezen HM, North KE, Levy D, Evans KL, Dupuis J, Breeze CE, Manichaikul A, London SJ. Pulmonary Function and Blood DNA Methylation: A Multiancestry Epigenome-Wide Association Meta-analysis. *Am J Respir Crit Care Med* 2022; 206: 321-336.
5. Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, Colicino E, Karrasch S, Wahl S, Kunze S, Jafari N, Zheng Y, Hou L, DeMeo DL, Litonjua AA, Vokonas PS, Peters A, Lin X, Schwartz J, Schulz H, Baccarelli AA. Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. *Epigenetics* 2018; 13: 1039-1055.
6. Fang J, Gao Y, Zhang M, Jiang Q, Chen C, Gao X, Liu Y, Dong H, Tang S, Li T, Shi X. Personal PM(2.5) Elemental Components, Decline of Lung Function, and the Role of DNA Methylation on Inflammation-Related Genes in Older Adults: Results and Implications of the BAPE Study. *Environ Sci Technol* 2022; 56: 15990-16000.

7. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W, Lomas DA, DeMeo DL. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med* 2012; 185: 373-381.
8. Casas-Recasens S, Noell G, Mendoza N, Lopez-Giraldo A, Garcia T, Guirao A, Agusti A, Faner R. Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity. *Am J Respir Crit Care Med* 2021; 203: 129-134.
9. Busch R, Cho MH, Silverman EK. Progress in disease progression genetics: dissecting the genetic origins of lung function decline in COPD. *Thorax* 2017; 72: 389.
10. Rezwan FI, Imboden M, Amaral AFS, Wielscher M, Jeong A, Triebner K, Real FG, Jarvelin M-R, Jarvis D, Probst-Hensch NM, Holloway JW. Association of adult lung function with accelerated biological aging. *Aging Albany Ny* 2020; 12: 518-542.
11. Martino D, Bui D, Lodge C, Perret J, Boweatte G, Lowe A, Faner R, Walters E, Thomas P, Moodley Y, Zosky G, Hamilton G, Holloway J, Svanes C, Abramson M, Dharmage S. Epigenetic study of COPD-risk lung function trajectories. *Respirology* 2022; 27: 3-4.
12. Golpon HA, Geraci MW, Moore MD, Miller HL, Miller GJ, Tuder RM, Voelkel NF. HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema. *Am J Pathol* 2001; 158: 955-966.
13. Portas L, Pereira M, Shaheen SO, Wyss AB, London SJ, Burney PGJ, Hind M, Dean CH, Minelli C. Lung Development Genes and Adult Lung Function. *Am J Respir Crit Care Med* 2020; 202: 853-865.
14. Sohal SS, Mahmood MQ, Walters EH. Clinical significance of epithelial mesenchymal transition (EMT) in chronic obstructive pulmonary disease (COPD): potential target for prevention of airway fibrosis and lung cancer. *Clin Transl Med* 2014; 3: 33.
15. Stolz D, Mkorombindo T, Schumann DM, Agusti A, Ash SY, Bafadhel M, Bai C, Chalmers JD, Criner GJ, Dharmage SC, Franssen FME, Frey U, Han M, Hansel NN, Hawkins NM, Kalhan R, Konigshoff M, Ko FW, Parekh TM, Powell P, Mölken MR-v, Simpson J, Sin DD, Song Y, Suki B, Troosters T, Washko GR, Welte T, Dransfield MT. Towards the elimination of chronic obstructive pulmonary disease: a Lancet Commission. *Lancet* 2022; 400: 921-972.
16. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, Aviv A, Lohman K, Liu Y, Ferrucci L, Horvath S. An epigenetic biomarker of aging for lifespan and healthspan. *Aging Albany Ny* 2018; 10: 573-591.

Figure Legends

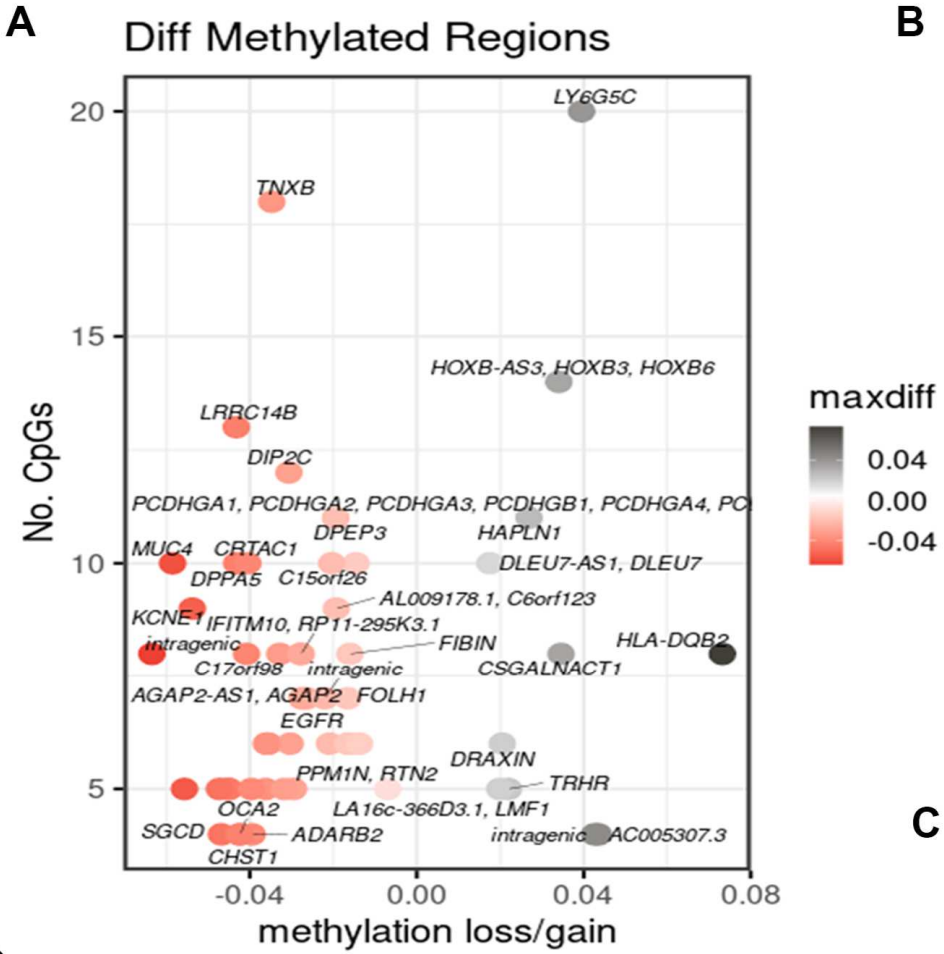
Figure 1 – (A) Volcanoplot of differentially methylated regions**.** Each point represents a genomic region, and the number of individual CpGs in the region is shown on the y-axis as a function of the effect size (x-axis) interpreted as the percent change in methylation ratios (10^{-2}). (**B**) Upset plot showing the number of overlapping DMRs per trajectory sub-group. *ph= persistently high*; *acc.dec = accelerated decline; bl.ave = below average.* **(C)** Boxplot of the Mucin 4 mQTL showing methylation ratios expressed as a percentage (10^{-2}) stratified by genotype. Means comparisons by t-test, exact P-values shown. **(D)** Summary statistics of tissue-specific enrichment testing for mQTLs and sets of differentially expressed genes for 30 general tissue types in the GTExv8 catalogue.

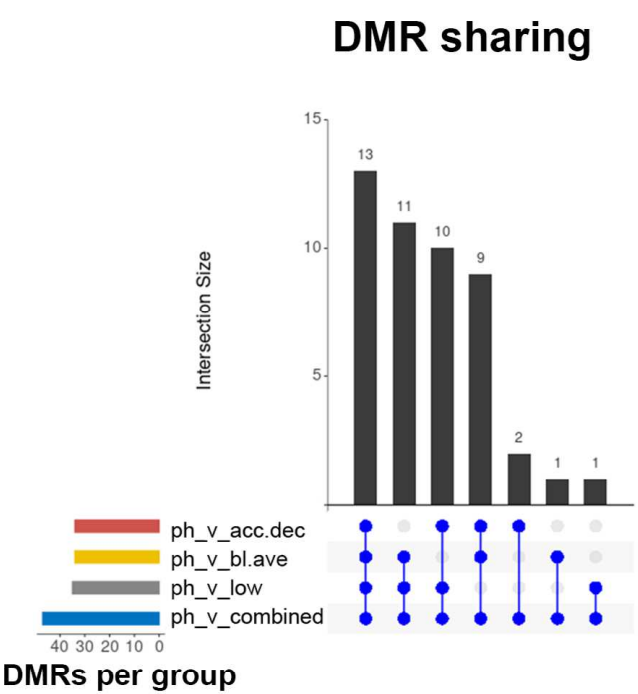
Tables

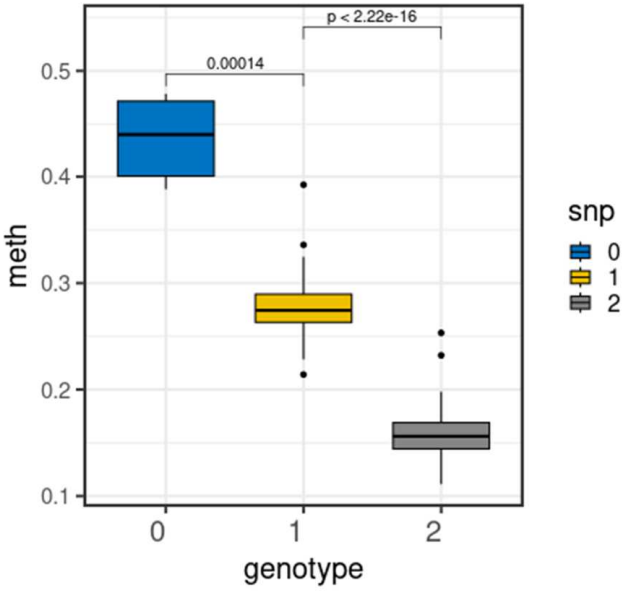
Table 1 - Logistic regression of PhenoAge clock with case - control group

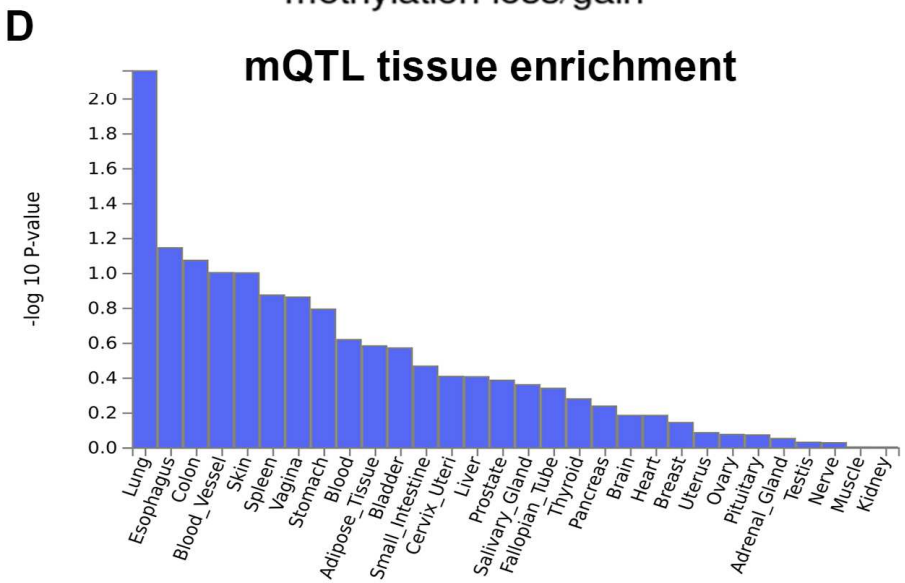
	Coefficient	Std. Error	z value	P value
AA	0.09	0.05	1.85	0.06
Male sex	1.11	0.38	2.90	0.06
Asthma	2.23	0.48	4.67	<0.01*
EAA	0.09	0.05	1.85	0.06
Male sex	1.11	0.38	2.90	<0.01*
Asthma	2.23	0.48	4.66	<0.01*
IEAA	0.10	0.05	1.94	0.05*
Male sex	1.09	0.38	2.87	<0.01*
Asthma	2.26	0.48	4.70	<0.01*

*Outcome variable = high-risk/persistently low, predictors: AA = Age acceleration residual, EAA = Extrinsic age acceleration residual, IEAA = Intrinsic age acceleration residual. * = P <0.05.*


A
Diff Methylated Regions
No. CpGs
methylation loss/gain
maxdiff
0.04
0.00
-0.04
LY6G5C
TNXB
HOXB-AS3, HOXB3, HOXB6
LRRC14B
DIP2C
PCDHGA1, PCDHGA2, PCDHGA3, PCDHGB1, PCDHGA4, PC
DPEP3
HAPLN1
MUC4
CRTAC1
DLEU7-AS1, DLEU7
DPPA5
C15orf26
AL009178.1, C6orf123
KCNE1
IFITM10, RP11-295K3.1
intragenic
FIBIN
HLA-DQB2
C17orf98
intragenic
CSGALNACT1
AGAP2-AS1, AGAP2
FOLH1
EGFR
DRAXIN
PPM1N, RTN2
TRHR
OCA2
LA16c-366D3.1, LMF1
SGCD
ADARB2
intragenic
AC005307.3
CHST1


B
DMR sharing
Intersection Size
ph_v_acc.dec
ph_v_bl.ave
ph_v_low
ph_v_combined
DMRs per group


C
MUC4 mQTL
chr5:82973396;cg16102102
p < 2.22e-16
0.00014
meth
genotype
snp
0
1
2


D
mQTL tissue enrichment
-log 10 P-value
Lung
Esophagus
Colon
Blood_Vessel
Skin
Spleen
Vagina
Stomach
Blood
Adipose_Tissue
Bladder
Small_Intestine
Cervix_Uteri
Liver
Prostate
Salivary_Gland
Fallopian_Tube
Thyroid
Pancreas
Brain
Heart
Breast
Uterus
Ovary
Pituitary
Adrenal_Gland
Testis
Nerve
Muscle
Kidney