

RESEARCH

Open Access



# Integrative analysis identifies gene signatures mediating the effect of DNA methylation on asthma severity and lung function

Eskezeia Y. Dessie<sup>1</sup>, Lili Ding<sup>2</sup> and Tesfaye B. Mersha<sup>1\*</sup>

## Abstract

DNA methylation (DNAm) changes play a key role in regulating gene expression in asthma. To investigate the role of epigenetics and transcriptomics change in asthma, we used publicly available DNAm (asthmatics,  $n=96$  and controls,  $n=46$ ) and gene expression (asthmatics,  $n=79$  and controls,  $n=39$ ) data derived from bronchial epithelial cells (BECs). We performed differential methylation/expression and weighted co-methylation/co-expression network analyses to identify co-methylated and co-expressed modules associated with asthma severity and lung function. For subjects with both DNAm and gene expression data (asthmatics,  $n=79$  and controls,  $n=39$ ), machine-learning technique was used to prioritize CpGs and differentially expressed genes (DEGs) for asthma risk prediction, and mediation analysis was used to uncover DEGs that mediate the effect of DNAm on asthma severity and lung function in BECs. Finally, we validated CpGs and their associated DEGs and the asthma risk prediction model in airway epithelial cells (AECs) dataset. The asthma risk prediction model based on 18 CpGs and 28 DEGs showed high accuracy in both the discovery BEC dataset with area under the receiver operating characteristic curve (AUC) = 0.99 and the validation AEC dataset (AUC = 0.82). Genes in the three co-methylated and six co-expressed modules were enriched in multiple pathways including WNT/beta-catenin signaling and notch signaling. Moreover, we identified 35 CpGs correlated with DEGs in BECs, of which 17 CpGs including cg01975495 (SERPINE1), cg10528482 (SLC9A3), cg25477769 (HNF1A) and cg26639146 (CD9), cg17945560 (TINAGL1) and cg10290200 (FLNC) were replicated in AECs. These DEGs mediate the association between DNAm and asthma severity and lung function. Overall, our study investigated the role of DNAm and gene expression change in asthma and provided an insight into the mechanisms underlying the effects of DNA methylation on asthma, asthma severity and lung function.

**Keywords** Asthma, Asthma severity, Lung function, Multi-omics analysis, Weighted correlation network analysis, Mediation analysis, Biomarkers

\*Correspondence:

Tesfaye B. Mersha

tesfaye.mersha@cchmc.org

<sup>1</sup> Division of Asthma Research, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA

<sup>2</sup> Division of Biostatistics and Epidemiology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA

## Introduction

Recent advances in high-throughput multi-omics approaches have enabled the collection of molecular assessments at different layers, providing complementary perspectives of complex diseases including asthma. While having a family history is a known risk factor for developing asthma [1], the low concordance rate of asthma between monozygotic twins suggests that



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

epigenomics and transcriptomics profiling and/or environmental factors may play substantial role in asthma pathogenesis [2].

DNA methylation (DNAm) may serve as surrogate endpoint for environmental exposures and could serve as a potential biomarker in the association between environmental factors and asthma endotypes [3, 4]. Reduced lung function is a hallmark of asthma disease. Lung function measures such as forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) are strongly associated with decline in airflow in asthma [5–7]. Many genomics studies of asthma have been performed either on gene expression data or DNAm data alone. For example, there were studies that showed the association of DNAm with lung function and asthma [8, 9], as well as a study that showed global changes of DNAm affect gene expression derived from BECs in asthma [10]. Furthermore, Zhang et al. revealed differentially expressed genes as well as enrichment of neutrophil activation and cytokine receptor pathways for severe asthma [11]. Thürmann et al. showed that DNAm plays key roles in various biological processes such as regulating gene expression in childhood asthma [12]. Perry et al. identified differentially methylated DNAm CpG sites that might play a role in regulating several key asthma-related genes such as DBX2, ACP6 and KCNJ11 [13]. However, the results of their study were primarily using either DNAm or gene expression alone, and the role of DNAm on gene expression associated with asthma status, severity and lung function was not thoroughly investigated through integrated analysis of DNAm and gene expression. Epigenetic and transcriptomic profiling are well suited to identify novel genes and pathways involved in asthma pathogenesis [14, 15].

In this study, we address the limitations of the previous asthma studies with the aim of identifying novel epigenetic regulation of gene expression associated with asthma severity and lung function through integrated analysis of DNAm and gene expression. The aims of the integrated analysis were to: (1) identify differentially co-methylated or co-expressed network modules associated with asthma status, severity and lung function; (2) identify dysregulated biological function within these identified modules and unveil shared pathways underlying asthma severity and lung function; (3) identify significant CpGs and DEGs using machine learning and construct multi-omics asthma risk predictive models; and (4) infer the mediational roles of gene expression on the relationships between DNA methylation changes and asthma severity as well as lung function in BECs. A subset of CpGs–DEGs pairs showed significant association in asthma and were used to validate the prediction model in AECs dataset. Overall, our results provide an insight into

the role of DNAm in asthma risk, severity and identify important genes that may mediate the effect of DNAm changes on asthma severity and lung function.

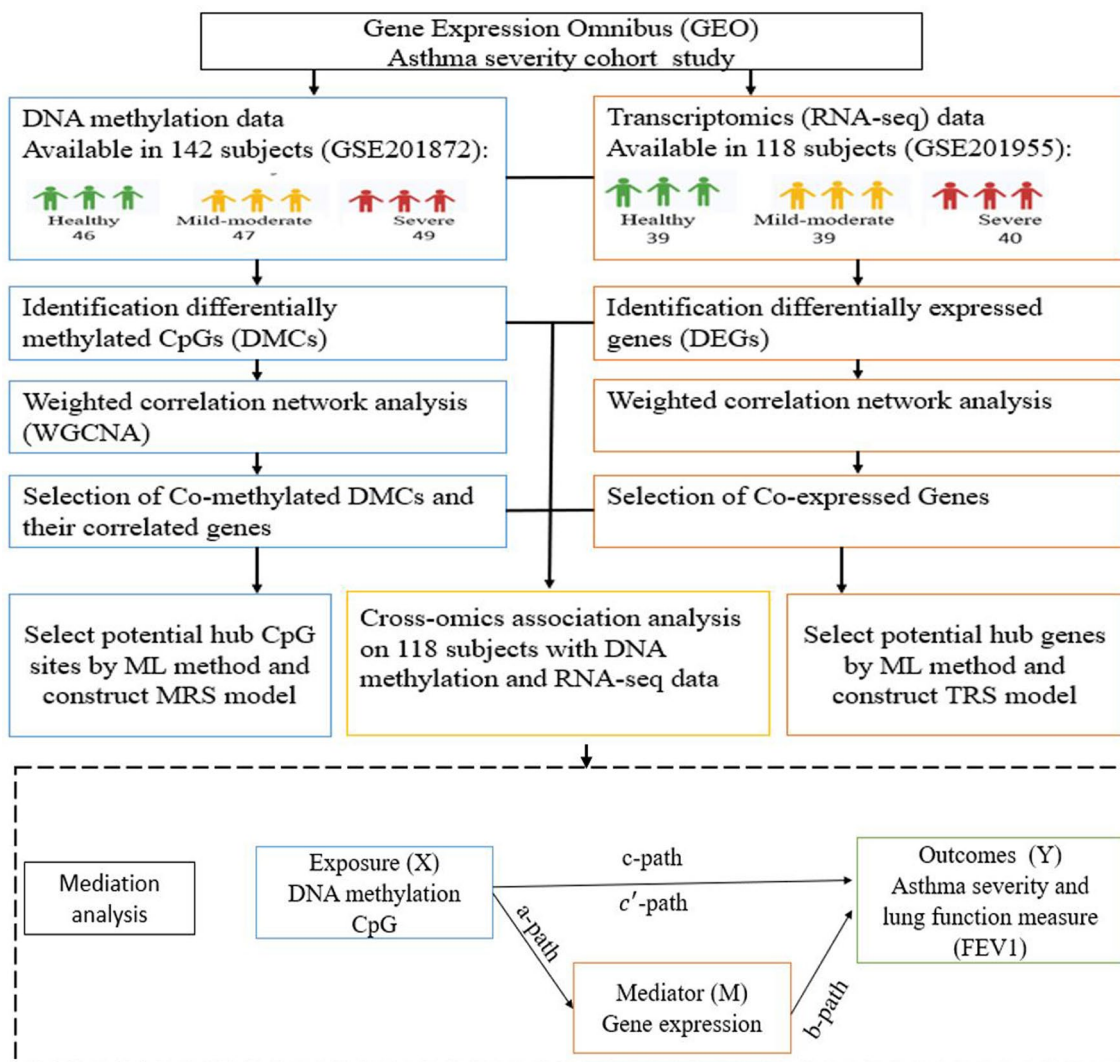
## Results

### Data description

The overall workflow of the study is described in Fig. 1. The GSE201872 dataset contained DNAm data profiles of 142 subjects including 46 controls, 47 mild-moderate asthmatic and 49 severe asthmatic samples derived from BECs. The GSE201955 dataset contained RNA-seq profiles of 118 subjects, a subset of the 142 subjects in the GSE201872 DNAm dataset, including 39 controls, 39 mild-moderate asthmatic and 40 severe asthmatic samples derived from BECs. Baseline characteristics of the 142 subjects are summarized in Table 1. Age, IgE, BMI, blood eosinophilia, and BAL eosinophilia are higher in mild-moderate and severe asthmatic subjects compared with controls ( $P < 0.01$ ). The mean ( $\pm$  standard deviation [SD]) of FEV1 was 3.2 ( $\pm 0.68$ ) for controls, 2.75 ( $\pm 0.99$ ) for mild-moderate asthmatics, 1.99 ( $\pm 0.59$ ) for severe asthmatics ( $P < 0.001$ ), and the mean ( $\pm$  SD) of FEV1/FVC was 0.95 ( $\pm 0.11$ ) for controls, 0.82 ( $\pm 0.18$ ) for mild-moderate asthmatic subjects and 0.66 ( $\pm 0.18$ ) for severe asthmatic subjects ( $P < 0.001$ ).

### Construction of co-methylation networks and identification of modules associated with asthma severity and lung function

Initially, we conducted differential methylation analysis between asthmatic subjects ( $n=96$ ) and controls ( $n=46$ ) based on DNAm data ( $M$ -values) adjusting for covariates including age, sex, smoking status and the first three ancestry PCs in the discovery DNA methylation dataset. The differential analysis identified 1845 differentially methylated CpGs (DMCs, including 1500 hyper-methylated and 345 hypo-methylated) in asthmatic subjects compared to controls (adjusted  $P$  value  $< 0.05$  and  $\Delta_{\text{meth}} > 1\%$ ; Fig. 2A). We reported the methylation difference (effect size,  $\Delta_{\text{meth}}$ ) between asthmatic subjects and controls as a percentage (e.g., effect size of 0.01 = 1%). The 1845 DMCs whose effect size ranged from  $-14.1$  to 21.0% were further analyzed to identify co-methylation networks/modules using WGCNA [16] package and the identified modules were tested for association with asthma severity, lung function and asthma-relevant clinical measures. In co-methylation network analysis, the suitable soft threshold power ( $\beta$ ) = 7 was used as the correlation coefficient threshold to ensure relatively balanced mean connectivity and scale free network (Additional file 1: Fig. S1A). WGCNA analysis clustered the 1845 DMCs into four modules with 95–1382 methylated CpGs per module (i.e., the largest is turquoise with 1382 CpGs, 723 genes) (Fig. 2B and Additional file 2: Table S1).



**Fig. 1** The overall workflow of the study. Initially, preprocessed, and normalized gene expression and DNA methylation data were downloaded and analyzed for differential methylation and expression analyses and weighted correlation network analysis (WGCNA) to generate DEG and DMC modules associated with asthma severity and lung function. Then, both DEGs and DMCs (adjusted *P* value < 0.05) in modules significantly associated with asthma severity and lung function (*P* value < 0.05) were selected. Finally, key DMCs and DEGs were selected to develop asthma-risk prediction models (methylation-based risk score (MRS) and transcriptomic-based risk score (TRS). Mediation analysis was conducted to select DMCs regulated DEGs in BECs

The gray module contained uncorrelated CpG sites and hence was not considered in downstream analyses. Notably, module eigenCpGs from three co-methylation modules including blue, brown, and turquoise modules were significantly associated with asthma relevant clinical measures including asthma severity, and lung function measures (Fig. 2C; *P* value < 0.05). Turquoise and brown modules were found to be positively associated with asthma severity, while blue module was negatively associated with asthma severity. Age was associated with blue and turquoise modules. Clinical phenotypes such as IgE, FEV1, FEV1/FVC

and BAL eosinophilia were significantly associated with all three modules. BMI was associated with blue, brown, and turquoise modules, whereas blood eosinophilia was significantly associated with brown module.

**Construction of co-expression networks and identification of modules associated with asthma severity and lung function**

For co-expression network analysis, we also first performed differential analysis between asthmatic subjects (*n* = 79) and controls (*n* = 39) using normalized

**Table 1** Clinical characteristics of 142 asthmatic and control subjects

Characteristic	Asthma-risk group			P value
	Control (n = 46)	Mild-moderate (n = 47)	Severe (n = 49)	
Age (years), mean ± SD	36.78 ± 11.58	37.04 ± 14.28	43.55* ± 11.05	0.010
Gender (% female)	65	64	82	
Race (% AA/EA/Other)	63/28/9	70/30/0	45/55/0	
Current smoking status at bronchoscopy (% yes)	15	4	4	
Maternal asthma (% yes)	9	40	33	
BMI (kg/m <sup>2</sup> ), mean ± SD	28.64 ± 5.80	29.58 ± 6.46	38.09*** ± 9.65	< 0.001
Mean abs FEV <sub>1</sub> , mean ± SD	3.2 ± 0.68	2.75* ± 0.99	1.99*** ± 0.59	< 0.001
Mean FEV1/FVC, mean ± SD	0.95 ± 0.11	0.82*** ± 0.18	0.66*** ± 0.18	< 0.001
Total serum IgE (IU/mL), Median	50.50	115.50*	194.00	0.049
(Lower, upper quartile)	21.25–167.75	47.50–349.25	22.00–312.00	
BAL eosinophilia (%), Median	0.00	2.90***	3.05***	< 0.001
(Lower, upper quartile)	0.00–0.50	1.55–5.10	1.28–6.33	
BAL neutrophilia (%), Median	4.90	5.40	4.60	0.520
(Lower, upper quartile)	3.20–6.40	3.55–7.00	2.65–6.50	
Blood eosinophilia (cells/μL), Median	10.00	17.00*	17.50	0.040
(Lower, upper quartile)	7.00–17.00	9.50–28.50	7.00–30.25	

Asthma severity is categorized based on STEP classification as mild, moderate, and severe asthma

AA African American, EA European American, BMI body mass index, FEV<sub>1</sub> forced expiratory volume at 1 s, FVC forced vital capacity, FEV<sub>1</sub>/FVC the ratio of FEV<sub>1</sub> to FVC, IgE immunoglobulin E, BAL Bronchoalveolar lavage

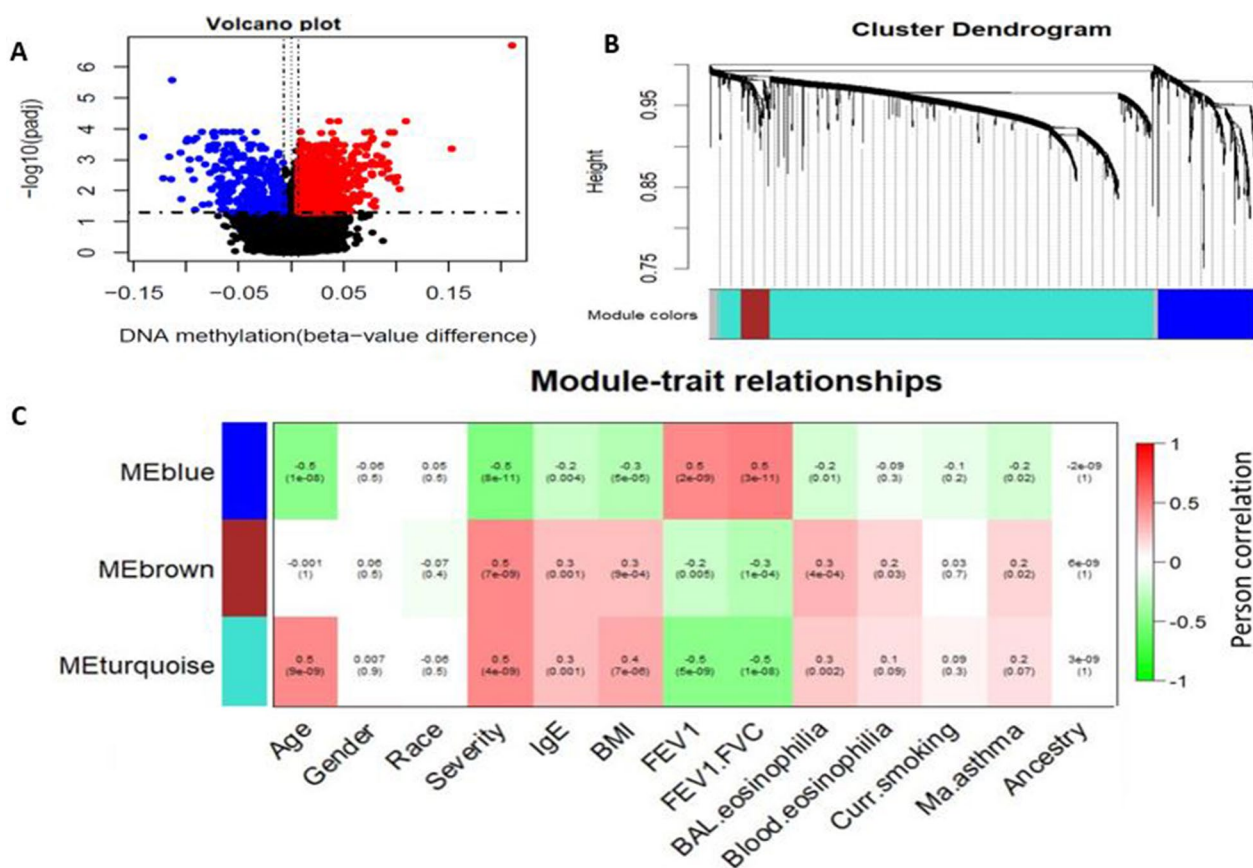
The significance levels of comparison (mean/median of mild-moderate asthmatics vs controls and severe asthmatics vs controls) were indicated by the stars (\*\*\**P* < 0.001, \*\**P* < 0.01, \**P* < 0.05). The *P*-values are for three-group (control, mild-moderate and severe) comparison

RNA-seq data (adjusted for covariates including age, gender, smoking status and the first three ancestry PCs). The differential analysis identified a total of 1891 DEGs including 1058 downregulated and 833 upregulated genes in asthmatic subjects compared with controls (absolute value log<sub>2</sub> (Fold change) > 0.1 and adjusted *P* value < 0.05) (Fig. 3A). Next, we performed WGCNA analysis to cluster and characterize correlation structure of the 1891 DEGs. WGCNA analysis using suitable  $\beta = 10$  (Additional file 1: Fig. S1B) was conducted, and a total of seven co-expression modules with 32–785 genes per module (Fig. 3B) were identified excluding the gray module. The module eigengenes of the rest seven co-expression modules were significantly associated asthma severity and asthma relevant clinical features including lung function measures (Fig. 3C; *P* value < 0.05). The modules turquoise, brown, and green were negatively associated with asthma severity. The modules red, black, blue, and yellow were positively associated with asthma severity. Among clinical phenotypes, IgE was significantly associated with yellow and black modules. BMI was associated with red, brown, and turquoise modules. FEV<sub>1</sub>/FVC and BAL eosinophilia were significantly correlated with all modules. FEV<sub>1</sub> and blood eosinophilia were correlated with all modules except red module.

### Functional enrichment analysis of co-methylated and co-expressed modules

#### Pathway analysis of co-methylated modules

Methylation CpGs in each module were mapped to genes based on IlluminaHumanMethylation450kanno.ilmn12.hg19 in the minfi package, and these co-methylated module genes were used for IPA analysis. Significantly enriched IPA results of three asthma severity and lung function associated co-methylation modules (turquoise, brown and blue) are shown in Fig. 4A–C and Additional file 2: Table S2. The 723 unique genes in turquoise co-methylation module were enriched in several pathways including pulmonary healing signaling, WNT/beta-catenin signaling, HOTAIR regulatory, regulation of epithelial development, and axonal guidance signaling pathways (Fig. 4A). The 70 genes associated with the brown co-methylation module were significantly enriched for pathways such as notch signaling, role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis, epithelial adherens junction signaling and WNT/beta-catenin signaling (Fig. 4B). The 205 mapped genes associated with blue co-methylated module were enriched in G-protein coupled receptor signaling and chronic myeloid leukemia signaling (Fig. 4C).



**Fig. 2** Identification of CpGs associated with asthma severity in the discovery BECs dataset. **A** Volcano plot showing CpGs associated with asthma. DMCs are identified as those with adjusted *P* value < 0.05 and absolute effect size ( $\Delta_{\text{meth}}$ ) > 1%. **B** Cluster dendrogram of four modules (blue, brown, turquoise, and gray) containing 1891 DMCs. **C** Correlation between co-methylation modules and various clinical measures

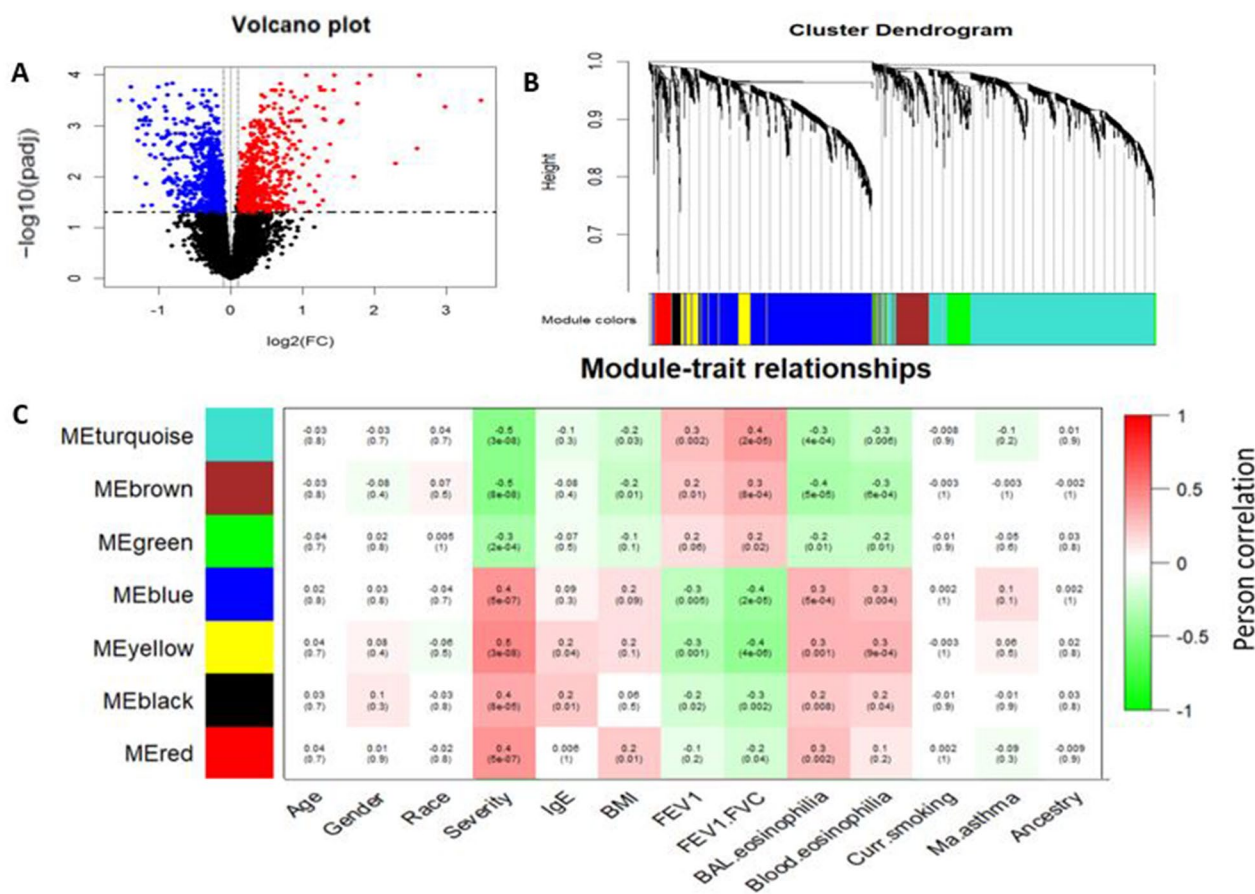
**Overrepresentation of epigenetic regulated genes in the co-expression modules**

We conducted hypergeometric tests on overlapping genes between the co-expression modules and the co-methylation modules that are associated with asthma severity and lung function to identify epigenetic regulated genes that overrepresented in the co-expression modules. The analysis showed that genes in three co-methylated modules of interest were overrepresented in the six co-expressed modules (turquoise, brown, green, red, blue, and yellow) (Additional file 2: Tables S1 and S3).

**Pathway analysis of co-expressed modules**

The results of significantly enriched pathways for genes in the co-expression modules of interest are shown in Fig. 5A–D and Additional file 2: Table S4. The 785 unique genes associated with turquoise module were significantly enriched in several biological pathways

such as RHO GTPase cycle, generic transcription pathway and WNT/beta-catenin signaling (Fig. 5A). The 61 unique genes associated with red module were significantly enriched in several biological pathways such as pathogen induced cytokine storm signaling, activin inhibin signaling and STAT3 Pathway (Fig. 5B). The 103 genes associated with the green module were mainly involved in protein kinase signaling and notch signaling (Fig. 5C). The 140 genes associated with the brown module were involved in axonal guidance signaling (Additional file 2: Table S4). Notably, uniquely annotated genes shared between the co-methylation modules and genes co-expression modules were enriched in several pathways including axonal guidance signaling, notch signaling, WNT/beta-catenin signaling, pulmonary fibrosis idiopathic signaling and myelination signaling (Additional file 1: Fig. S2), suggesting DNA methylation changes in these pathways may potentially regulate genes in asthma pathogenesis.

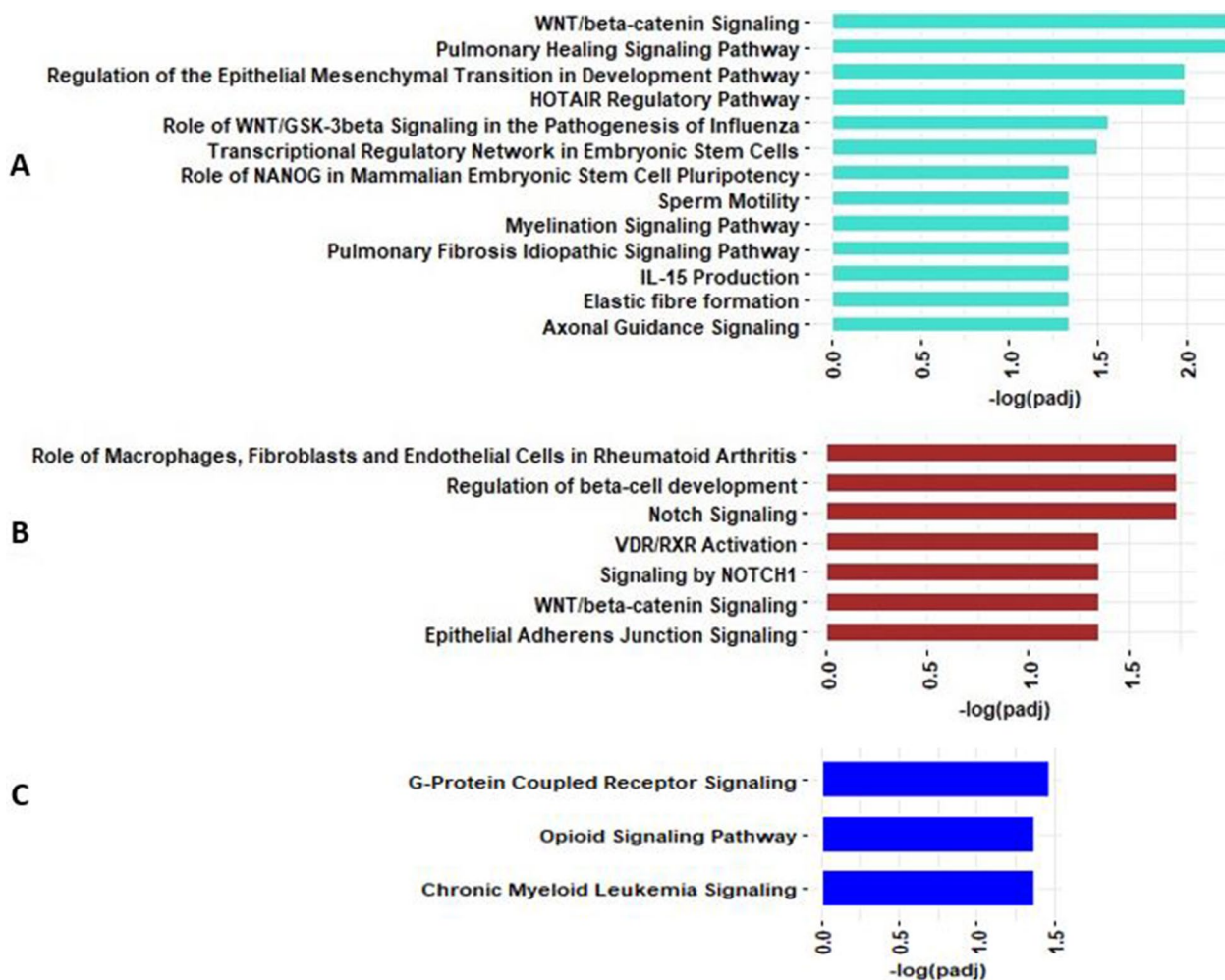


**Fig. 3** Identification of DEGs associated with asthma severity in the discovery BECs dataset. **A** Volcano plot showing DEGs related to asthma. Differentially expressed genes (DEGs) are those with adjusted *P* value < 0.05 and absolute value of fold change > 0.1. **B** Cluster dendrogram of 1891 DEGs revealed eight modules including turquoise, brown, green, blue, yellow, black, red, and gray. **C** Correlation between co-expression modules and various clinical measures

### Construction and validation of asthma multi-omics risk model

To construct a multi-omics asthma prediction model in the discovery BECs datasets, we used Boruta method to identify important CpGs and genes whose DNAm or expression that showed high asthma-risk prediction performance. The results showed that 18 DMCs and 28 DEGs yielded the “confirm” status in Boruta iterations in methylation and expression data, respectively, in the discovery BECs datasets (Additional file 2: Tables S5 and S6) with high asthma risk prediction performance. Next, based on 18 important CpGs and 28 genes, we constructed Methylation risk score (MRS) and transcriptome risk score (TRS), respectively, and their asthma-risk prediction performances were assessed. The asthma-risk prediction MRS model had an AUC of 0.92, and the TRS model had an AUC of 0.98 in the discovery datasets. We also constructed multi-omics risk model by integrating MRS and TRS models based on 118 samples having

both DNA methylation and gene expression data and this model revealed high asthma diagnostic performance with AUC=0.99 (Fig. 6A). The MRS is significantly related to asthma severity (Additional file 1: Fig. S3A), FEV1 ( $r = -0.43$ ,  $P$  value = < 0.001; Additional file 1: Fig. S3B) and FEV1/ FVC ( $r = -0.5$ ,  $P$  value = < 0.001; Additional file 1: Fig. S3C). The TRS also associated with asthma severity (Additional file 1: Fig. S3D), FEV1 ( $r = -0.23$ ,  $P$  value < 0.05; Additional file 1: Fig. S3E) and FEV1 ( $r = -0.33$ ,  $P$  value < 0.001; Additional file 1: Fig. S3F). Notably, the risk scores significantly associated with several asthma clinical features such as BAL eosinophilia, BAL neutrophils, IGE, and blood eosinophilia (Additional file 1: Figs. S4 and S5). To validate the robustness of our developed asthma multi-omics risk model, we constructed 18 CpG sites-based MRS model, and 28 genes-based TRS model using 81 samples having both DNA methylation and gene expression data in validation AECs dataset. The diagnostic ability of multi-omics asthma risk



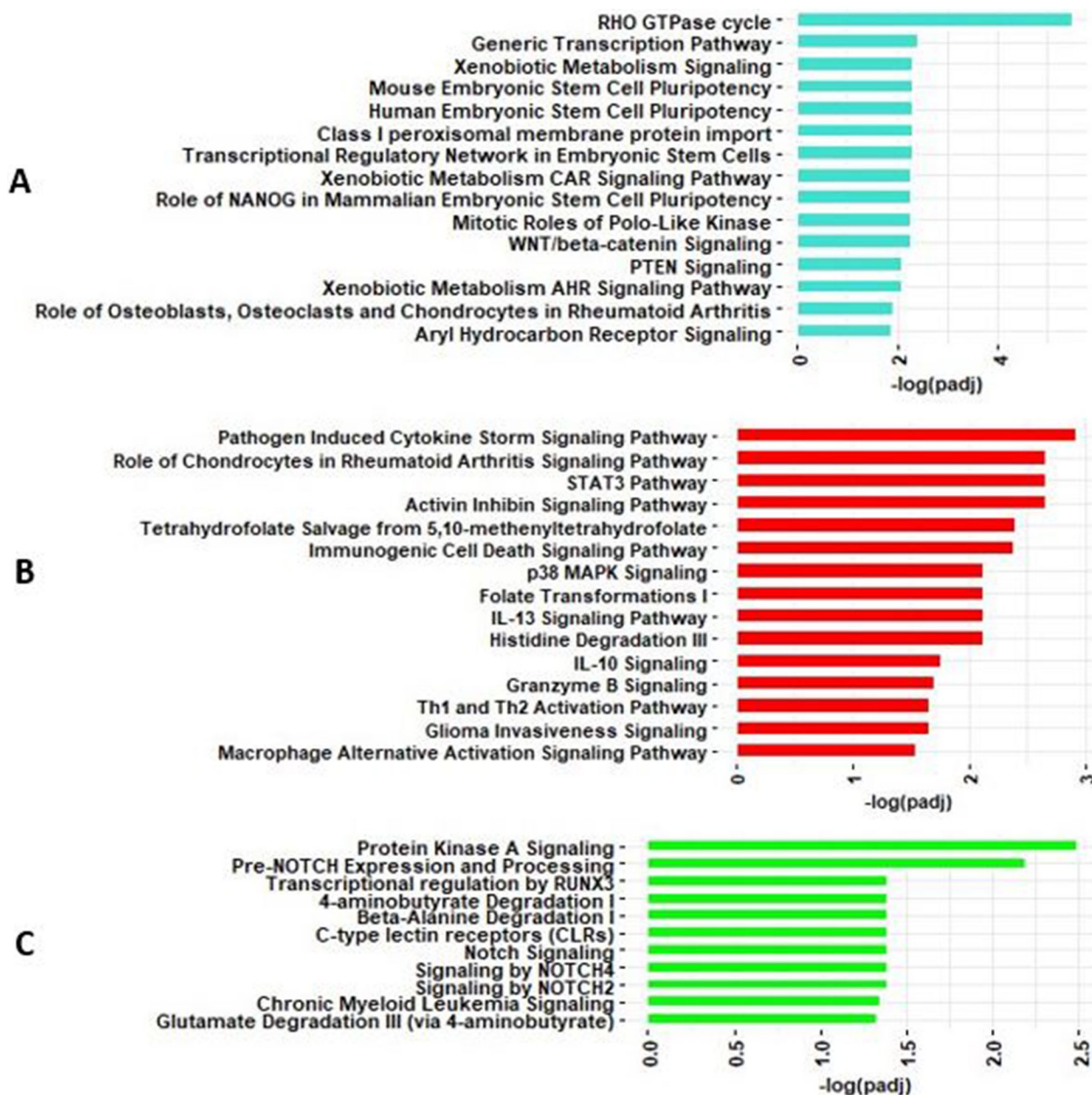
**Fig. 4** Pathways significantly enriched by genes in the **A.** turquoise, **B.** brown, and **C.** blue co-methylation modules that are significantly ( $P$  value < 0.01) associated with asthma severity and/or lung function

model in validation AECs dataset revealed high performance in asthma-risk prediction (AUC=0.82) (Fig. 6B).

**Integrating DNA methylation and gene expression data**

We performed correlation analysis between methylation levels of DMCs and the expression level of DEGs in the modules associated with asthma severity and lung function to explore the biological relevance of the DMCs in gene expression. Initially, the 1845 DMCs were mapped to 1035 genes. Then, correlation analysis between the methylation levels of DMCs and gene expression levels identified a total of 35 significantly correlated DMC-DEG pairs (Fig. 7A–F and Additional file 2: Table S7) including cg14015211 (*TLR5*), cg21385480 (*LRIG1*), cg01975495 (*SERPINE1*), cg10528482 (*SLC9A3*), cg25477769 (*HNFI1A*) and

cg26639146 (*CD9*). Using these results, we conducted a mediation analysis for each of the 35 DMC-DEG pairs to examine direct effect (CpG methylation → asthma severity or FEV1) and mediational effect (CpG methylation → gene expression → asthma severity or FEV1) pathways adjusting for age, gender, and ancestry PCs. We identified several significant mediation genes (Table 2 and Additional file 2: Table S8) including *INAGL1*, *SERPINE1*, *TLR5*, *SLC9A3* and *CD9* for asthma severity and *INAGL1*, *TLR5*, *CD9* for FEV1 (Table 2 and Additional file 2: Table S9), suggesting potential regulation effect of DNAm on gene expression in asthma. Moreover, we found a significant direct effect of DNAm CpG on asthma severity and FEV1 after adjusting for age, gender, ancestry, and mediator gene (Table 2, Additional file 2: Tables S7 and S8).



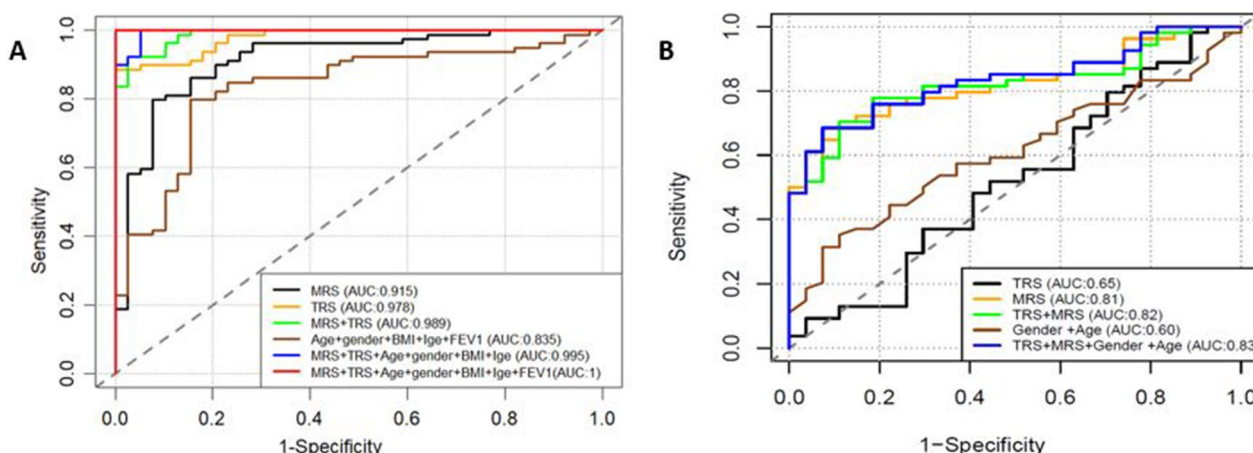
**Fig. 5** Pathways significantly enriched by genes in the **A.** turquoise, **B.** red, and **C.** green co-expression modules that are significantly ( $P$  value  $< 0.01$ ) associated with asthma severity and/or lung function

**Validation of differential methylation CpG sites associated with DEGs in an independent AEC dataset**

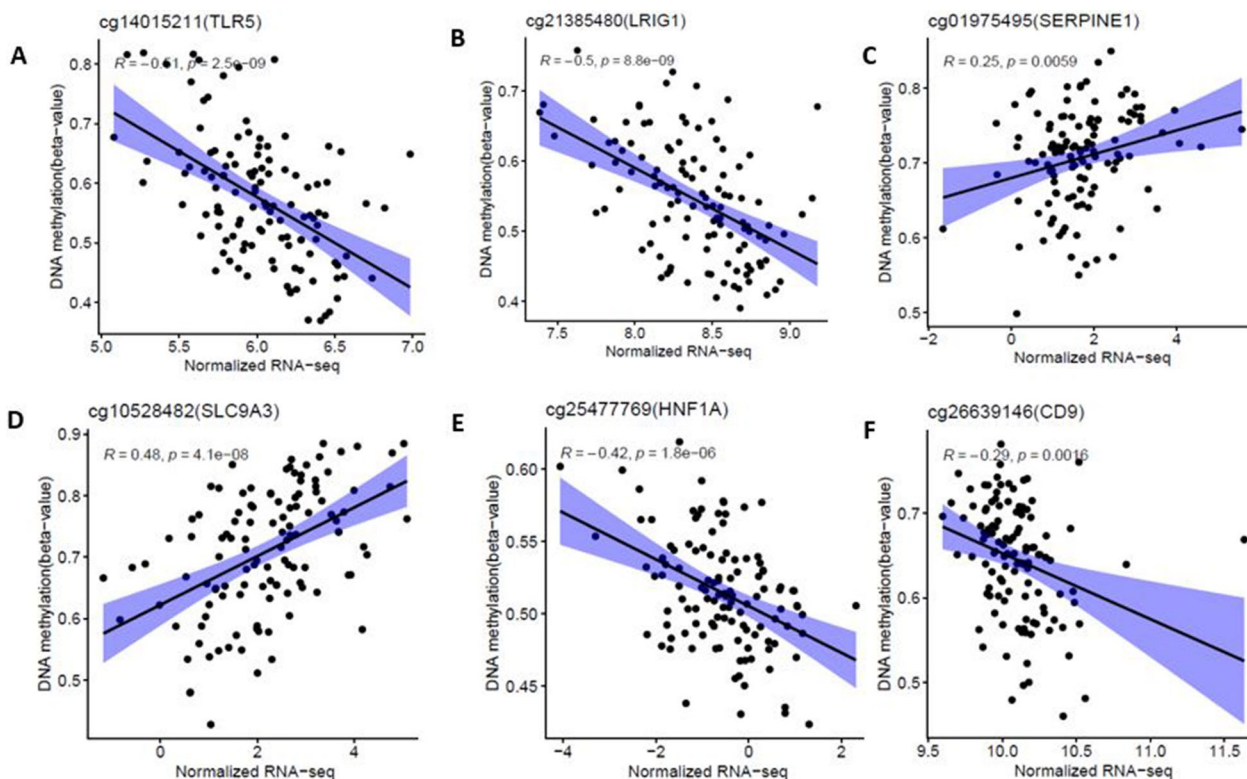
To validate the 35 DMC-DEG pairs associated with asthma severity and lung function identified in BECs, we further used an independent AECs validation dataset. First, we conducted differential methylation analysis between asthmatic subjects ( $n=74$ ) and Controls ( $n=41$ ) in AECs adjusting age and gender and showed that 29 CpG sites out of 35 DMCs identified in BECs were

found to be differentially methylated in AECs (Table 3) with the same directional effect (adjusted  $P$  value  $< 0.05$  and  $\Delta_{meth} > 1\%$ ). Next, we examined correlation between 29 DMCs and their mapped DEGs based on 81 samples having both DNA methylation and gene expression data in validation AECs dataset. We validated 17 DMCs significantly correlated with DEGs (Fig. 8A–F and Additional file 1: Fig. S6) including cg01975495 (*SERPINE1*), cg10528482 (*SLC9A3*), cg25477769 (*HNFL1A*)





**Fig. 6** The risk prediction performance of different risk models including MRS, TRS and clinical biomarkers. **A** in the discovery BECs dataset **B** in the validation AECs dataset. MRS-methylation-based risk Score, TRS-transcriptomic-based risk score



**Fig. 7** Scatter plot showing the correlation between DNA methylation levels and gene expression levels for 118 samples with both DNA methylation and gene expression data in the discovery-BECs dataset

and cg26639146 (*CD9*), cg17945560 (*TINAGL1*) and cg10290200 (*FLNC*). More importantly, several of these identified DEGs that are correlated with DMCs have been implicated in asthma [17–19].

### Discussion

In this study, DNAm and gene expression change in BECs of asthmatics and controls were analyzed at a genome-wide scale. Based on differential methylation, expression, and weighted correlation network analysis, we identified co-methylation modules and co-expressed modules that

**Table 2** Mediation analysis of gene expression in the association between DNA methylation and asthma severity and lung function measure (FEV1) in the discovery BECs dataset

CpG	Gene	Total effect		Direct effect (DE)		Mediational effect (ME)		Mediation proportion
		$\beta$ (95% CI)	P value	$\beta$ (95% CI)	P value	$\beta$ (95% CI)	P value	
<i>Asthma severity association</i>								
cg17945560	TINAGL1	0.34 (0.06, 0.59)	0.016	0.076 (-0.15, 0.28)	0.52	0.26 (0.09, 0.44)	<b>2E-16</b>	<b>77.5</b>
cg01975495	SERPINE1	0.69 (0.36, 0.99)	2E-16	0.57 (0.24, 0.88)	2E-16	0.12 (0.03, 0.25)	<b>0.004</b>	<b>17.3</b>
cg14015211	TLR5	0.314 (0.21, 0.43)	2E-16	0.21 (0.07, 0.35)	0.008	0.106 (0.04, 0.191)	<b>0.004</b>	<b>33.77</b>
cg26639146	CD9	-0.29 (-0.5, -0.06)	0.016	-0.18 (-0.38, 0.09)	0.17	-0.10 (-0.25, -0.031)	<b>0.002</b>	<b>36.24</b>
cg10290200	FLNC	-0.20 (-0.3, -0.1)	2E-16	-0.15 (-0.26, -0.03)	0.006	-0.05 (-0.11, 0.1)	0.06	<b>26.69</b>
cg25477769	HNF1A	0.57 (0.203, 0.93)	0.002	0.39 (-0.02, 0.82)	0.068	0.18 (-0.02, 0.391)	0.08	31.18
cg21385480	LRIG1	0.41 (0.281, 0.55)	2E-16	0.3 (0.13, 0.45)	2E-16	0.12 (0.03, 0.221)	<b>0.002</b>	<b>27.99</b>
cg10528482	SLC9A3	0.28 (0.183, 0.38)	2E-16	0.2 (0.08, 0.31)	2E-16	0.08 (0.02, 0.141)	<b>0.002</b>	<b>27.97</b>
<i>Lung function association</i>								
g17945560	TINAGL1	-0.27 (-0.53, 0.01)	0.06	-0.11 (-0.36, 0.13)	0.07	-0.17 (-0.27, -0.06)	<b>2E-16</b>	<b>61.2</b>
cg01975495	SERPINE1	-0.61 (-0.84, -0.38)	2E-16	-0.55 (-0.76, -0.32)	2E-16	-0.06 (-0.15, 0.01)	0.13	9.08
cg14015211	TLR5	-0.41 (-0.56, -0.27)	2E-16	-0.27 (-0.43, -0.12)	2E-16	-0.14 (-0.26, -0.03)	<b>0.022</b>	<b>34.78</b>
cg26639146	CD9	0.34 (0.018-0.63)	0.046	0.25 (-0.18, 0.55)	0.218	0.09 (0.003, 0.33)	<b>0.034</b>	<b>25.44</b>
cg10290200	FLNC	0.22 (0.08-0.36)	0.006	0.13 (-0.04, 0.29)	0.154	0.09 (0.004, 0.17)	<b>0.036</b>	42.72
cg25477769	HNF1A	-0.65 (-1.18, -0.11)	0.028	-0.47 (-1.07, 0.16)	0.162	-0.18 (-0.43, 0.05)	0.124	27.8
cg21385480	LRIG1	-0.37 (-0.58, -0.13)	0.002	-0.41 (-0.66, -0.09)	0.006	0.04 (-0.12, 0.16)	0.614	10.11
cg10528482	SLC9A3	-0.25 (-0.41, -0.09)	0.002	-0.18 (-0.35, -0.01)	0.044	-0.07 (-0.16, 0.01)	0.076	28.5

ME denotes causal mediational effect (indirect effect of each CpG on asthma severity/FEV1 that goes through mediating gene). The values in bold show genes that are significantly mediating the DNA methylation effect on asthma severity/FEV1 association

were correlated with asthma severity, and lung function. The functional enrichment analysis of co-methylation module genes and co-expression module genes showed enrichment in several shared pathways including axonal guidance signaling, notch signaling, WNT/beta-catenin signaling, pulmonary fibrosis idiopathic signaling and myelination signaling. Subsequently, machine learning methods identified the most informative subset of DMCs and DEGs that leads to better asthma diagnostic performance. Moreover, correlation analysis followed by mediation analysis identified potential DNAm regulated DEGs that were associated with asthma severity and lung function and the findings were validated using an independent AECs dataset.

In this study, we first identified DMCs and DEGs in asthmatic samples compared to controls and further conducted weighted correlation network analysis to identify network modules associated with asthma severity and lung function and identified several shared pathways including WNT/beta-catenin and notch signaling. The Wnt/beta-catenin signaling pathway regulates airway remodeling, which is involved in chronic asthma [20]. Indeed, several genes identified in this study are enriched in WNT/beta-catenin signaling, including *FZD7* and *HNF1A*, suggesting methylation signals associated with

the activation WNT/beta-catenin signaling in the development of asthma pathology. Moreover, differentially methylated CpGs annotated genes and differentially expressed genes (*DTX1*, *MAML2*, and *NOTCH1*) are involved in notch signaling and this result is consistent with previous finding, showing DNAm associated with notch signaling pathway activation in severe asthmatic subjects [13]. The notch signaling regulates the immune response and a therapeutic target for asthma [21].

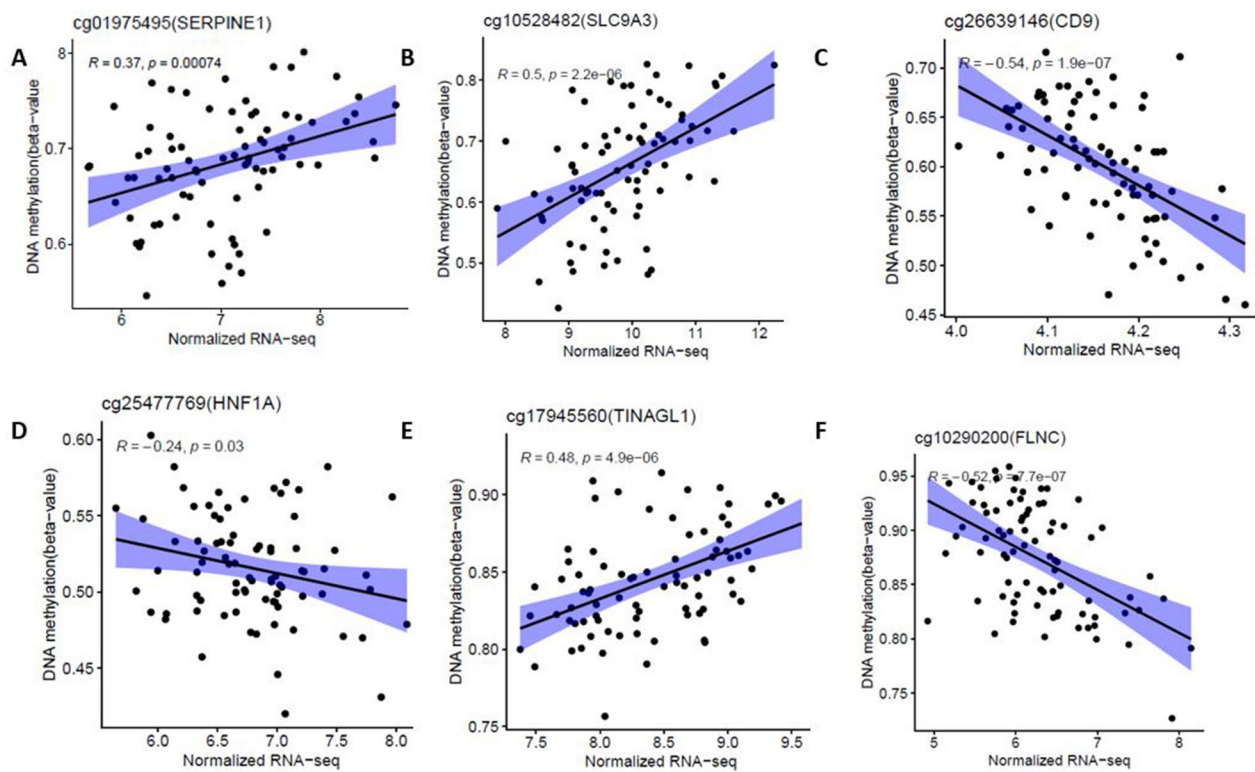
To evaluate predictive performance of DMCs and DEGs, we used machine learning methods to identify important DMCs and DEGs and constructed asthma risk prediction models. Most previous studies that constructed asthma risk prediction model were limited to either using CpGs or DEGs [22], and there were limited studies developed risk model based on integrated transcriptomic and DNAm data. A recent study by Liu et al. [23] used co-expression and co-methylation network analyses to identify genetic signatures and methylation pathways associated with Opioid use disorder. Zhang et al. [11] also showed that the roles of co-expression network analysis to identify hub DEGs associated with asthma severity. In this study, we constructed multi-omics-based risk models in predicting asthma, as multi-omics data potentially predict outcome more accurately

**Table 3** Validation of the methylation differences of CpGs between asthmatic subjects and controls in independent validation ACEs dataset

CpG	Gene	Effect size in discovery set (n = 96 asthmatics, n = 46 controls)		Effect size in validation set (n = 74 asthmatics, n = 41 controls)		Direction of association
		$\Delta_{\text{meth}}$ (100%)	Adjusted P value	$\Delta_{\text{meth}}$ (100%)	Adjusted P value	
cg00406211	GRK5	-11.335	2.63E-06	-12.978	1.39E-09	Hypo
cg21385480	LRIG1	7.7035	0.0002	9.3715	4.45E-09	Hyper
cg10528482	SLC9A3	8.1986	0.0007	8.968	2.76E-05	Hyper
cg16393012	ARHGDI3	-5.627	0.010	-8.863	1.46E-05	Hypo
cg25267808	MAML2	6.403	0.006	6.661	6.16E-05	Hyper
cg18181229	PBX1	5.285	0.009	5.417	0.001	Hyper
cg14185918	KLF4	-4.071	0.006	-5.009	0.0001	Hypo
cg10290200	FLNC	-2.633	0.001	-4.843	3.08E-06	Hypo
cg02766259	AACS	-4.119	0.011	-4.623	0.0004	Hypo
cg06128142	GPT2	3.595	0.001	4.277	5.91E-06	Hyper
cg23817893	CCDC81	4.435	0.003	4.1574	0.003	Hyper
cg01975495	SERPINE1	4.669	0.003	3.989	0.002	Hyper
cg26639146	CD9	-3.696	0.039	-3.709	0.009	Hypo
cg03441945	ABAT	3.402	0.014	3.573	0.001	Hyper
cg17602126	HEYL	3.188	0.002	3.307	0.001	Hyper
cg25477769	HNF1A	2.497	0.017	3.304	7.98E-05	Hyper
cg19212949	PEG3	-1.943	0.027	-3.243	0.0001	Hypo
cg08801887	TCIRG1	2.366	0.017	3.127	4.58E-05	Hyper
cg11419403	CLPTM1L	-2.66	0.009	-2.749	0.016	Hypo
cg15012607	ETHE1	2.046	0.043	2.554	0.0005	Hyper
cg10672136	TPO	1.551	0.001	2.512	6.98E-08	Hyper
cg17945560	TINAGL1	2.037	0.011	2.383	0.0032	Hyper
cg15562220	SCGN	1.964	0.018	2.329	0.006	Hyper
cg23556108	BCL11A	1.652	0.001	2.1769	0.0001	Hyper
cg19651003	BSG	-1.589	0.004	-2.164	1.44E-05	Hypo
cg10336131	CNIH2	1.941	0.007	1.963	0.008	Hyper
cg20491914	KCNK3	1.281	0.045	1.632	0.017	Hyper
cg09048665	WDR90	-1.233	0.033	-1.412	0.012	Hypo
cg17012160	FMN2	1.200	0.019	1.290	0.002	Hyper

than single omics data [24]. Our multi-omics-based risk model prediction accuracy (AUC) was 0.99 in BECs and 0.82 in AECs. Moreover, mediational analysis was performed to examine whether the association between methylated CpG site and asthma severity and lung function is explained at least partially by differentially expressed gene with adjustment for covariates. We found 35 correlated DMC-DEG pairs and revealed several significant mediation relationships among them including *TLR5*, *SLC9A3*, *SERPINE1* and *GRK5*. Other genes such as *RMD4A*, *HNF1A* and *GAL3ST2* showed insignificant mediation, implying that the association between DNAm and asthma severity and lung function occur via other mechanisms in BECs. Notably, our validation analysis in independent AECs datasets replicated 29 out of the 35

differentially methylated CpGs associated with asthma in BECs and 17 of the differentially methylated CpGs were significantly correlated with expression of annotated DEGs including *SERPINE1*, *SLC9A3*, *CD9*, *HNF1A*, *TINAGL1*, and *FLNC*. Two CpG sites that were methylated across two studies are annotated to hypo-methylated *SERPINE1* and *SLC9A3*. Pampuch et al. showed that *SERPINE1* was associated with IgE response and bronchial activity [25]. In addition, Zhou et al. suggested that DNA methylation changes of *SERPINE1* might be responsible for mediating the effect of genetic variate on the development of food allergy [17]. Importantly, our study revealed that *SERPINE1* was hypo-methylated in BECs and AECs and its expression mediates the methylation effect on asthma severity, and lung function.



**Fig. 8** Scatter plot showing the correlation between DNA methylation levels and gene expression levels for 81 samples with both DNA methylation and gene expression data in validation AECs dataset

The validated hyper-methylated co-methylated CpG site annotated to *SLC9A3* is part of our asthma-risk predictive model. This gene is related to airway epithelium and immune regulation. Notably, our findings showed that expression level of *SLC9A3* was associated with DNA methylation changes and mediated methylation effect on asthma severity and lung function association. Our finding aligned with previous genome-wide study in atopy [18], indicating expression changes of *SLC9A3* mediate the methylation and atopy association. Other hyper-methylated validated genes were mapped to genes such as *CD9*, *SCGN*, *KLF4*, *KCNK3* and *HNF1A*. Previous study reported deficiency of *KLF4* compromises the lung function in an acute mouse model of allergic asthma [26]. Besides, *KLF4* regulates immune response-related genes including *IL1RL1*, *CD274*, and *CD44* in human nasal epithelial cells of allergic rhinitis [27]. In line with previous study, our study also found that *KLF4* is significantly associated with asthma severity and lung function. In addition, we found hypo-methylated *KLF4* significantly associated with asthma severity and lung function, which suggests that methylation effect may regulate *KLF4* in asthma and lung function disease development. We also found that hyper-methylated CpG annotated to *HNF1A* was associated with asthma severity and reduced

lung function. Our finding was supported by the previous study showing hyper-methylated *HNF1A* linked with asthma pathogenesis [28].

Our study has some limitations. Sample size of the DNA methylation and transcription expression data is relatively small. Hence, this study may not detect all association between DNA methylation CpG and gene expression changes and with clinical traits such as asthma status, severity and lung function measures. Since CpG can be influenced by lifestyle factors and environmental risk factors, in future, considering lifestyle and environmental risk factors and their association with methylation changes in asthma disease is recommended. The study was limited by datasets deposited in public domains. Our analysis is also limited to DNAm and DEGs, and future studies including genomics, proteomics and metabolomics need to be conducted.

Our study also has several strengths. First, we constructed co-methylation and co-expression modules associated with asthma severity and lung function measures and found that differentially co-methylation module CpGs and co-expression module DEGs enriched common biological pathways. Second, we developed MRS and TRS in asthma-specific tissue and further validated using asthma-relevant tissues [29]. Third, to our

knowledge, this is the first study that constructed multi-omics asthma risk prediction model by integrating TRS and MRS in BECs and conducted mediational analysis to assess whether the methylation-asthma severity or lung function association is explained by transcriptomics changes.

In conclusion, DNAm is an epigenetic mechanism that plays a key role in regulating gene expression. An integrated analysis of DNAm and gene expression data identified several DNA methylation-regulated genes associated with asthma severity and lung function in BECs dataset and subset of which including cg01975495 (*SERPINE1*), cg10528482 (*SLC9A3*), cg25477769 (*HNF1A*) and cg26639146 (*CD9*), cg17945560 (*TINAGLI*) and cg10290200 (*FLNC*) were further confirmed in validation AECs dataset. The validation results showed that these methylation CpG sites/genes may serve for asthma diagnosis and the development of new epigenetic therapies for asthma. In addition, the integration of multi-omics data has expanded our understanding of how various omics data correlated with one another and inform better prediction of disease risk.

## Materials and methods

### Datasets and filtering criteria

The overall pipeline of our study is shown in Fig. 1. Eligible and publicly available DNAm and gene expression datasets were selected based on the following criteria. (1) Homo sapiens, (2) sample size  $\geq 80$ , and (3) consists of DNAm and gene expression profiles of asthmatic and control subjects. DNAm data (accession: GSE201872) and RNA-seq data (accession: GSE201955) obtained from BECs that satisfied the inclusion criteria were selected and used as discovery datasets. Other datasets with DNAm (accession: GSE85568) and RNA-seq data (accession: GSE85567) obtained from AECs that satisfied the inclusion criteria were selected and used for validation datasets. Initially, DNAm and gene expression profiles and corresponding clinical information were downloaded from the identified Gene Expression Omnibus (GEO) database.

The quality controls and data normalization procedures for DNAm data (GSE201872) and RNA-seq data (GSE201955) were already conducted and described in the original study [30].

For GSE201872, a matrix of DNAm data with  $M$ -values of a total of 61,162 methylation CpGs in 142 samples (including 46 controls, 47 mild-moderate asthmatic and 49 severe asthmatic) based on both Illumina Infinium MethylationEPIC and Infinium HumanMethylation450K assays were obtained. We further normalized the between platform differences using surrogate variable analysis (SVA) [31] package. The

GSE201955-RNA-seq data contain a total of 118 samples, which are subset of the 142 samples in the DNAm GSE201872 dataset, including 39 control, 39 mild-moderate asthmatic and 40 severe asthmatics and a total of 13,757 genes with normalized RNA-seq expression profiles. Therefore, there are a total of 118 samples with both DNAm and gene expression data available between the GSE201872 and GSE201955 datasets. Three outcome variables: asthma status only (cases vs. controls) for identification of DMCs and DEGs and the development of risk prediction models, asthma severity (no, step = 0, mild-moderate, step = 1, 2, 3, 4 and severe, step = 5, 6) and lung function measures (including FEV1 and FEV1/FVC) for associations with co-methylation and co-expression modules, were considered in this study. Other demographic and clinical variables such as age, gender, race, current smoking status, maternal asthma, BMI, IgE, BAL eosinophilia, BAL neutrophilia, and blood eosinophilia were included in our study. In addition to clinical and demographic variables, we included the three ancestry principal components (PCs) information, which were estimated in the original study [10]. To replicate and validate our findings, DNAm data from GSE85568 (contained 71 asthmatic samples and 42 control samples) and gene expression data from GSE85567 (contained 57 asthmatic samples and 28 control samples) derived from AECs were used. Between DNAm data from GSE85568 and gene expression data from GSE85567 datasets, a total of 81 same samples have both methylation and gene expression data available. The summary of the datasets used in this study is shown in Additional file 2: Table S10.

### Identification of differentially methylated and expressed genes

We conducted differential methylated and expressed analysis between asthmatic and control samples to identify DMCs and DEGs using Empirical Bayes linear model in the limma package [32] adjusting for age, gender, current smoking status and ancestry information. Differential methylation analysis on  $M$ -values, the Benjamini and Hochberg (B-H) [33]-based adjusted  $P$  value  $< 0.05$  and an absolute effect size ( $|\Delta_{\text{meth}}| > 1\%$ ) were used to define DMCs. The effect size was defined based on beta-values as follows:  $(\Delta_{\text{meth}}) = \text{mean methy}_{\text{asthmatics}} - \text{mean methy}_{\text{Control}}$ . Similarly, differential expression analysis was conducted to identify DEGs adjusting age, gender, and smoking status in the discovery dataset. We used an absolute value of  $\log_2$  (fold change =  $\text{gene expression}_{\text{asthmatics}} / \text{gene expression}_{\text{controls}}$ )  $> 0.1$  and a significance threshold of adjusted  $P$  value  $< 0.05$  based on B-H procedure to identify DEGs [33].

### Weighted correlation network analysis

We conducted weighted correlation network analyses to identify co-methylation and co-expression networks based on the connectivity of methylation of CpGs and gene expression profiles using WGCNA package [34]. Initially, similarity matrix of DMCs or DEGs were constructed using pairwise correlation  $S_{ij} = \text{cor}(x_i, x_j)$ , where  $x_i$  and  $x_j$  represent the  $i$ th row and the  $j$ th row of DNAm/gene expression data matrix  $X$ , respectively. The similarity matrix was transformed into an adjacency matrix, represented by  $A_{ij} = |\text{cor}(x_i, x_j)|^\beta$ , where the suitable soft-thresholding power  $\beta$  was set ranging from 1 to 20 utilizing the pick Soft Threshold function. Dynamic tree cutting algorithm was used to divide DMCs or DEGs into different groups of connected CpGs (co-methylation modules) or connected DEGs (co-expressed modules) based WGCNA signed network, respectively. Methylated CpGs/DEGs with similar methylation/expression patterns are clustered into the same module and may potentially share common biological roles [35]. The eigenvectors of the co-methylation and co-expression modules were derived and examined for associations with clinical features of asthma including the asthma severity outcome (no, mild-moderate or severe asthma), lung function measures (FEV1 and FEV1/FVC), IgE, BMI, BAL eosinophilia, blood eosinophilia, maternal asthma status, age, gender, race, current smoking status and ancestry PCs. Lastly, modules significantly correlated with asthma severity and lung function measures were selected for further analysis ( $P$  value  $< 0.05$ ). The DMCs within significant co-methylated modules and DEGs within co-expressed modules are hypothesized to have an influential role in regulating diseases [36].

### Pathway analysis of module genes

The network module level analysis between co-methylated and co-expressed module genes was conducted to integrate DNAm and gene expression analysis. At the network module level, we used a hypergeometric test in the GeneOverlap [37] package to measure overlapping genes between CpG mapped genes in the co-methylation modules associated with asthma severity and lung function and DEGs in the co-expression modules associated with asthma severity and lung function. A hypergeometric test with  $P$  value  $< 0.05$  was used to assess the significance of overlapping genes between co-methylation modules and co-expression modules. Then, pathway enrichment analyses by Ingenuity Pathway Analysis (IPA) software [38] were used to understand the biological function of genes in the co-methylated and co-expression modules that were associated with asthma severity and lung function. The IPA method evaluates proportional

representation of module genes from a defined set in a canonical pathway in all set of known genes. Canonical pathways related to input module genes were elucidated with a ratio to evaluate significant pathway enrichment adjusted for multiple testing. The adjusted  $P$  value was calculated based on B-H procedure and biological functions with adjusted  $P$  value  $< 0.05$  were defined as significant canonical pathways.

### Construction and validation of multi-omics risk score models

We constructed three asthma risk prediction models: methylation-based risk score (MRS), transcriptomic-based risk score (TRS) and multi-omics risk model (MRS + TRS). Differentially co-methylated CpG sites and differentially co-expressed genes obtained from differential and WGCNA analysis with adjusted  $P$  value  $< 0.05$  were included in supervised Boruta algorithm. The Boruta algorithm is type of machine learning technique, which was used to select important CpG sites and genes whose DNAm levels and gene expression levels are relevant for asthma risk prediction [39] using Boruta R package. Important CpG sites and genes yielded the 'confirmed' status in Boruta iterations were selected for subsequent analysis. Next, multiple logistic regression model was utilized to generate coefficients of the CpGs and genes, separately for asthma risk prediction. Then,  $MRS = \sum \beta_i \times \text{methyl-CpG}_i$ , where methyl-CpG is the  $M$  value and  $\beta$  is the regression coefficient from logistic regression analysis [40] and  $TRS = \sum \beta_i \times \text{exp-gene}_i$ , where exp-gene is the normalized gene expression value and  $\beta$  is the coefficient from logistic regression analysis were constructed. Finally, multi-omics risk score model by integrating TRS and MRS was constructed using 118 samples having both DNAm and gene expression data in the discovery BCEs dataset. The prediction performances were assessed in the discovery BEC datasets using area under the ROC curve (AU-ROC) as implemented in pROC [41] R package. The risk prediction performances were validated using 81 samples having both DNAm and gene expression data in the validation AECs dataset.

### Correlation analysis between CpG methylation level and expression level of DEGs

To evaluate whether identified differentially methylated CpGs are correlated with corresponding mapped DEGs, we performed correlation analysis using Pearson correlation coefficient ( $R_i$ ) as calculated below based on  $M$ -values of methylated CpGs and normalized expression values of DEGs [42] of 118 subjects in discovery BECs dataset.

$$R_i = \frac{\sum_{i=0}^n (p_{ij} - \bar{p}_i)(g_{ij} - \bar{g}_i)}{\sqrt{\sum_{i=0}^n (p_{ij} - \bar{p}_i)^2 \sum_{i=0}^n (g_{ij} - \bar{g}_i)^2}}$$

Here,  $p_{ij}$  represents the  $M$ -values of  $i$ -th CpG site in the  $j$ -th subject,  $\bar{p}_i$  is the mean  $M$ -value of  $i$ -th CpG site over subjects;  $g_{ij}$  represent the expression value of  $i$ -th gene in the  $j$ -th subject,  $\bar{g}_i$  is the mean expression value of  $i$ -th gene over subjects and  $R_i$  is correlation coefficient between  $i$ -th CpG site  $M$ -values and  $i$ -th gene expression values, where the  $i$ -th CpG site is mapped to the  $i$ -th gene. Significant correlations between the differentially methylated CpGs and mapped DEGs in discovery BECs dataset were validated based on 81 samples having both DNA methylation and gene expression data in the independent validation AECs dataset.

### Mediation analysis

After evaluating the correlation between methylated CpGs and the corresponding mapped DEGs in the discovery dataset, we conducted mediation analysis to explore whether any proportion of the association between methylated CpGs and asthma severity/FEV1 is mediated by DEG [43–45] among the subjects with both DNAm and gene expression data. In our mediation analysis, we constructed three models: including (1) mediator model (a-path) to evaluate the association between the gene expression and DNAm of CpG using linear regression model. (2) Outcome models (b-path and  $c'$ -path) to evaluate the effect of DNAm and gene expression on asthma severity using ordinal logistic regression or FEV1 using linear regression. (3) Total effect model (c-path) to evaluate the effect of DNAm on the asthma severity using ordinal logistic regression or FEV1 using linear regression. All models were adjusted for covariates: age, gender, and ancestry. The regression models used in our study are described in Fig. 1 and as follows.

- *a-path*: gene expression ~ DNA methylation + covariates,
- b-path and  $c'$ -path: FEV1 or asthma severity ~ DNA methylation + Gene expression + covariates,
- c-path: FEV1 or asthma severity ~ DNA methylation + covariates (total effect).

The mediation effect of each CpGs was performed separately. The standardized coefficients were estimated based on 1000 nonparametric bootstrap resampling quasi-Bayesian approximation method using mediation package [46]. The total effect from DNAm to asthma severity/FEV1 was decomposed into the indirect effect or mediated effect (ME) through gene expression (methylation CpGs → gene expression → asthma severity or FEV1)

and the direct effect (methylation CpGs → asthma severity or FEV1) not mediated by gene expression but could be mediated by other factors or a direct link between DNA methylation changes and asthma severity or FEV1. The ME is equal to  $a \times b$ , which is equivalent to  $c - c'$  and denotes the part of the DNAm effect of each CpG on asthma severity or FEV1 mediated by gene expression. Significance of association was defined based a  $P$  value  $< 0.05$  as cut-off.

### Statistical analysis

All preprocessing and statistical analysis were performed using R software version 4.2.2 and associated Bioconductor packages. Comparisons of median among three groups were tested using pairwise median test. The ANOVA test was used to compare means of three groups using multcomp package. The Tukey Honest significant differences test was utilized to compare mean between two groups. Associations between two continuous variables were assessed by either Spearman or Pearson correlation coefficient. Statistical significance based on B-H adjusted  $P$  value  $< 0.05$  was used, unless otherwise stated.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-023-01611-9>.

**Additional file 1.** Supplementary Figures S1–S6.

**Additional file 2.** Supplementary Tables S1–S10.

### Acknowledgements

The authors are thankful to Joseph Beyene and Mekibib Altaye for their valuable input.

### Author contributions

TBM conceived the study, EYD and TBM designed the study, EYD performed the analysis and supervised by TBM, and LD and EYD drafted and wrote the manuscript with help from TBM and LD. All authors read and approved the paper.

### Funding

This work was supported by the National Institutes of Health (NIH) NHGRI (R01 HG011411) grant support.

### Availability of data and materials

The DNA methylation data (Accession Numbers: GSE201872) and RNA sequencing (Accession Numbers: GSE201872) as well as the DNA methylation (GSE85568) and RNA sequencing data (GSE85567) that support the findings of this study openly available in the GEO database.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

**Competing interests**

The authors declare no competing interests.

Received: 22 March 2023 Accepted: 2 December 2023

Published online: 20 January 2024

**References**

- London SJ, et al. Family history and the risk of early-onset persistent, early-onset transient, and late-onset asthma. *Epidemiology*. 2001;12(5):577–83.
- Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med*. 2014;6(7):60.
- Yang IV, Schwartz DA. Epigenetic mechanisms and the development of asthma. *J Allergy Clin Immunol*. 2012;130(6):1243–55.
- Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33(Suppl):245–54.
- Moeller A, et al. Monitoring asthma in childhood: lung function, bronchial responsiveness and inflammation. *Eur Respir Rev*. 2015;24(136):204–15.
- Agusti A, et al. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med*. 2017;5(12):935–45.
- Hole DJ, et al. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ*. 1996;313(7059):711–5 (**discussion 715–6**).
- Herrera-Luis E, et al. Epigenome-wide association study of lung function in Latino children and youth with asthma. *Clin Epigenet*. 2022;14(1):9.
- DeVries A, Vercelli D. Epigenetic mechanisms in asthma. *Ann Am Thorac Soc*. 2016;13(Suppl 1):S48–50.
- Magnaye KM, et al. DNA methylation signatures in airway cells from adult children of asthmatic mothers reflect subtypes of severe asthma. *Proc Natl Acad Sci USA*. 2022;119(24):e2116467119.
- Zhang Z, Wang J, Chen O. Identification of biomarkers and pathogenesis in severe asthma by coexpression network analysis. *BMC Med Genom*. 2021;14(1):51.
- Thürmann L, et al. Global hypomethylation in childhood asthma identified by genome-wide DNA-methylation sequencing preferentially affects enhancer regions. *Allergy*. 2023;78(6):1489–506.
- Perry MM, et al. DNA methylation modules in airway smooth muscle are associated with asthma severity. *Eur Respir J*. 2018;51(4):1701068.
- Perez-Garcia J, et al. Epigenomic response to albuterol treatment in asthma-relevant airway epithelial cells. *Clin Epigenet*. 2023;15(1):156.
- Singh P, et al. Transcriptomic analysis delineates potential signature genes and miRNAs associated with the pathogenesis of asthma. *Sci Rep*. 2020;10(1):13354.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9(1):559.
- Zhou X, et al. Targeted DNA methylation profiling reveals epigenetic signatures in peanut allergy. *JCI Insight*. 2021;6(6):e143058.
- Forno E, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. *Lancet Respir Med*. 2019;7(4):336–46.
- Hoang TT, et al. Epigenome-wide association study of DNA methylation and adult asthma in the agricultural lung health study. *Eur Respir J*. 2020;56(3):2000217.
- Kwak HJ, et al. The Wnt/ $\beta$ -catenin signaling pathway regulates the development of airway remodeling in patients with asthma. *Exp Mol Med*. 2015;47(12):e198.
- Qu SY, et al. Disruption of the Notch pathway aggravates airway inflammation by inhibiting regulatory T cell differentiation via regulation of plasmacytoid dendritic cells. *Scand J Immunol*. 2020;91(5):e12865.
- Kothalawala DM, et al. Integration of genomic risk scores to improve the prediction of childhood asthma diagnosis. *J Pers Med*. 2022;12(1):75.
- Liu A, et al. Genome-wide correlation of DNA methylation and gene expression in postmortem brain tissues of opioid use disorder patients. *Int J Neuropsychopharmacol*. 2021;24(11):879–91.
- Radzikowska U, et al. Omics technologies in allergy and asthma research: an EAACI position paper. *Allergy*. 2022;77:2888–908.
- Pampuch A, et al. The -675 4G/5G plasminogen activator inhibitor-1 promoter polymorphism in house dust mite-sensitive allergic asthma patients. *Allergy*. 2006;61(2):234–8.
- Nimpong JA, et al. Deficiency of KLF4 compromises the lung function in an acute mouse model of allergic asthma. *Biochem Biophys Res Commun*. 2017;493(1):598–603.
- Jeon Y, et al. Gene signatures and associated transcription factors of allergic rhinitis: KLF4 expression is associated with immune response. *Biomed Res Int*. 2023;2023:1317998.
- Clifford RL, et al. TWIST1 DNA methylation is a cell marker of airway and parenchymal lung fibroblasts that are differentially methylated in asthma. *Clin Epigenet*. 2020;12(1):145.
- Sajuthi SP, et al. Nasal airway transcriptome-wide association study of asthma reveals genetically driven mucus pathobiology. *Nat Commun*. 2022;13(1):1632.
- Magnaye KM, et al. DNA methylation signatures in airway cells from adult children of asthmatic mothers reflect subtypes of severe asthma. *Proc Natl Acad Sci*. 2022;119(24):e2116467119.
- Leek JT, et al. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
- Ritchie ME, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47.
- Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990;9(7):811–8.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
- van Dam S, et al. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 2018;19(4):575–92.
- Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat Commun*. 2018;9(1):1090.
- Shen L, Sinai M. GeneOverlap: test and visualize gene overlaps. R package version. 2022.
- Shao Z, et al. Ingenuity pathway analysis of differentially expressed genes involved in signaling pathways and molecular networks in RhoE gene-edited cardiomyocytes. *Int J Mol Med*. 2020;46(3):1225–38.
- Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010;36(11):1–13.
- Dessie EY, et al. A novel miRNA-based classification model of risks and stages for clear cell renal cell carcinoma patients. *BMC Bioinform*. 2021;22(Suppl 10):270.
- Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;12(1):77.
- Cao YN, Li QZ, Liu YX. Discovered key CpG sites by analyzing DNA methylation and gene expression in breast cancer samples. *Front Cell Dev Biol*. 2022;10:815843.
- Alfons A, Ateş NY, Groenen PJF. A robust bootstrap test for mediation analysis. *Organ Res Methods*. 2022;25(3):591–617.
- Wielscher M, et al. DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. *Nat Commun*. 2022;13(1):2408.
- Meng H, et al. Epigenome-wide DNA methylation signature of plasma zinc and their mediation roles in the association of zinc with lung cancer risk. *Environ Pollut*. 2022;307:119563.
- Tingley D, et al. Mediation: R package for causal mediation analysis. *J Stat Softw*. 2014;59(5):1–38.

**Publisher's Note**

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