An et al. World Allergy Organization Journal (2024) 17:100871
http://doi.org/10.1016/j.waojou.2024.100871

WORLD ALLERGY ORGANIZATION JOURNAL

Open Access

Blood transcriptome differentiates clinical clusters for asthma

Jin An, MD, PhD[a,1], Seungpil Jeong, PhD[b,1], Kyungtaek Park, PhD[c], Heejin Jin, PhD[c], Jaehyun Park, PhD[d], Eunsoon Shin, PhD[e], Ji-Hyang Lee, MD, PhD[f], Woo-Jung Song, MD, PhD[f], Hyouk-Soo Kwon, MD, PhD[f], You Sook Cho, MD, PhD[f], Jong Eun Lee, PhD[e], Sungho Won, PhD[b,c,g**,1] and Tae-Bum Kim, MD, PhD[f*,1]

ABSTRACT

Background: In previous studies, several asthma phenotypes were identified using clinical and demographic parameters. Transcriptional phenotypes were mainly identified using sputum and bronchial cells.

Objective: We aimed to investigate asthma phenotypes via clustering analysis using clinical variables and compare the transcription levels among clusters using gene expression profiling of the blood.

Methods: Clustering analysis was performed using 6 parameters: age of asthma onset, body mass index, pack-years of smoking, forced expiratory volume in 1 s (FEV1), FEV1/forced vital capacity, and blood eosinophil counts. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples and RNA was extracted from selected PBMCs. Transcriptional profiles were generated (Illumina NovaSeq 6000) and analyzed using the reference genome and gene annotation files (hg19.refGene.gft). Pathway enrichment analysis was conducted using GO, KEGG, and REACTOME databases.

Results: In total, 355 patients with asthma were included in the analysis, of whom 72 (20.3%) had severe asthma. Clustering of the 6 parameters revealed 4 distinct subtypes. Cluster 1 (n = 63) had lower predicted FEV1 % and higher pack-years of smoking and neutrophils in sputum. Cluster 2 (n = 43) had a higher proportion and number of eosinophils in sputum and blood, and severe airflow limitation. Cluster 3 (n = 110) consisted of younger subjects with atopic features. Cluster 4 (n = 139) included features of late-onset mild asthma. Differentially expressed genes between clusters 1 and 2 were related to inflammatory responses and cell activation. Th17 cell differentiation and interferon gamma-mediated signaling pathways were related to neutrophilic inflammation in asthma.

Conclusion: Four clinical clusters were differentiated based on clinical parameters and blood eosinophils in adult patients with asthma form the Cohort for Reality and Evolution of Adult Asthma in Korea (COREA) cohort. Gene expression profiling and molecular pathways are novel means of classifying asthma phenotypes.

Keywords: Asthma, Cluster, Transcriptome, Pathway

[a]Department of Pulmonary, Allergy and Critical Care Medicine, College of Medicine, Kyung Hee University Hospital at Gangdong, Kyung Hee University, Seoul, South Korea
*Corresponding author. Department of Allergy and Clinical Immunology, Asan Medical Center, University of Ulsan College of Medicine, 86 Asanbyeongwon-gil, Songpa-gu, Seoul 138-736, South Korea. E-mail: tbkim@amc.seoul.kr
**Corresponding author. Department of Public Health Sciences, Seoul National University, Kwanak-ro 1, Kwanak-gu, Seoul 151-742, South Korea. Email: sunghow@gmail.com

[1] These authors contributed equally to this work.
Full list of author information is available at the end of the article

http://doi.org/10.1016/j.waojou.2024.100871
Received 2 May 2023; Received in revised from 4 January 2024; Accepted 11 January 2024
Online publication date xxx

INTRODUCTION

Asthma is a common chronic inflammatory disease of the airways that involves airway inflammation, variable airway obstruction, and hyper-responsiveness.[1] Various factors, including genetic susceptibility, environmental exposure, and medication compliance, influence the heterogeneous clinical manifestations of asthma, highlighting the growing emphasis on phenotypes.[1-3] Asthma phenotypes have been defined via cluster analysis based on clinical variables.[4,5] Previously, cluster analysis of the Korean asthmatic cohort revealed 4 distinct clusters: smoking asthma, severe obstructive asthma, early onset atopic asthma, and late-onset mild asthma.[6]

Pathobiologic alterations in asthma are increasingly identified as heterogeneous and differences in the expression of many biological pathways underlie differences in the phenotypic expressions of the disease.[7] Asthma phenotypes can be considered as different biological processes with distinctive but overlapping genomic, transcriptomic, and physiologic features.[8] Accordingly, the identification of differentially expressed genes (DEGs) associated with asthma phenotypes may be a better approach to capture the various pathways of asthma pathophysiology.

Recent transcriptomic profiling of adult asthmatics mainly involved the use of airway cells, such as induced sputum, bronchial biopsies, and brushing.[9-12] According to a study that employed transcriptomics and peripheral blood mononuclear cell (PBMC), PBMC could recapitulate systemic changes accompanying asthma exacerbation.[13] We aimed to identify asthma phenotypes via clustering analysis with clinical parameters and compare the levels of transcription among clusters via gene expression profiling of PBMCs.

METHODS

Study subjects

A total of 456 adult asthmatics with available PBMC samples and medical records from the Cohort for Reality and Evolution of Adult Asthma in Korea (COREA) between 2005 and December 2020 were enrolled in the study. The diagnosis of asthma was confirmed via airway hyperresponsiveness, as indicated by a 20% reduction in forced expiratory volume in 1 s (FEV1) with a methacholine dose of 16 mg/mL (PC20) through a provocation test,[14] or airway reversibility in FEV1 $>$12% (and at least 200 mL) after inhalation of a short-acting β-agonist.[15] Patients with active infection, such as pneumonia, severe lung damage, bronchiectasis, or a history of lung resection were excluded. Induced sputum samples were obtained by inhalation of nebulized sterile saline solution followed by coughing and expectoration of airway secretions. Eosinophilic inflammation was defined as positive if it induced sputum eosinophils $\geq$2%, blood eosinophil count $\geq$150 cells/μL, or fractional excretion of nitric oxide (FeNO) level $\geq$20 ppb.[16] Neutrophilic asthma was defined as having neutrophil percentages in the sputum of 61% or higher.[17] Eosinophilic inflammation was confirmed, followed by checking for neutrophilic inflammation in patients without eosinophilic inflammation. Severe asthma was determined according to the international European Respiratory Society/American Thoracic Society (ERS/ATS) guidelines.[18] This study was approved by the Institutional Review Board of the Asan Medical Center (2019-0376). All patients provided written informed consent.

Cluster analysis

We conducted a two-phase cluster analysis to categorize adult patients with asthma into the COREA cohort. This approach employs unsupervised learning techniques to ensure accurate and reliable clustering.

In the initial phase, we used the Gaussian Mixture Model (GMM) with the assistance of the MCLUST package[19] to determine the optimal number of clusters and their centroids. This method was specifically chosen to uncover the underlying patterns within complex datasets. To determine this, we relied on both the Bayesian Information Criterion (BIC) and Integrated Complete-data

Likelihood (ICL)[20] to select the most suitable combination of components and the covariance architecture. The GMM was parameterized and initiated using a model-based hierarchical clustering approach. Remarkably, both methods consistently identified the division into 4 distinct groups, as depicted in Supplementary Fig. 1A.

In the subsequent phase, we employ K-means clustering using ECLUST,[21] considering the cluster counts and centroids established in the initial phase. The resulting cluster are shown in Fig. 1B. To validate these findings, we assessed the silhouette score (Supplementary Fig. 1C) and performed principal component analysis (PCA) to examine the cluster distributions (Supplementary Fig. 1D).

Six variables were selected for cluster analysis based on their contribution to the characterization of asthma phenotypes, namely age of asthma onset, body mass index (BMI), pack-years of smoking, FEV1 as a percentage of the predicted value, FEV1/forced vital capacity (FVC), and blood eosinophil counts. One notable aspect of the cluster analysis in this study was the inclusion of eosinophilic inflammation, differentiating it from previous cluster analyses conducted on the COREA cohort.[6] Of the 456 asthmatics, 355 with complete data for the 6 variables were employed for the cluster analysis. The study protocol for cluster analysis is presented in Fig. 1.

RNA extraction from PBMC

RNA extraction from PBMC was performed using TRI Reagent solution (Invitrogen), following the manufacturer's instructions. Initially, the PBMC sample was homogenized in 1 mL of TRI Reagent solution. Subsequently, 200 µL of chloroform was added, thoroughly mixed, and incubated at room temperature (RT) for 10 min. The mixture was then

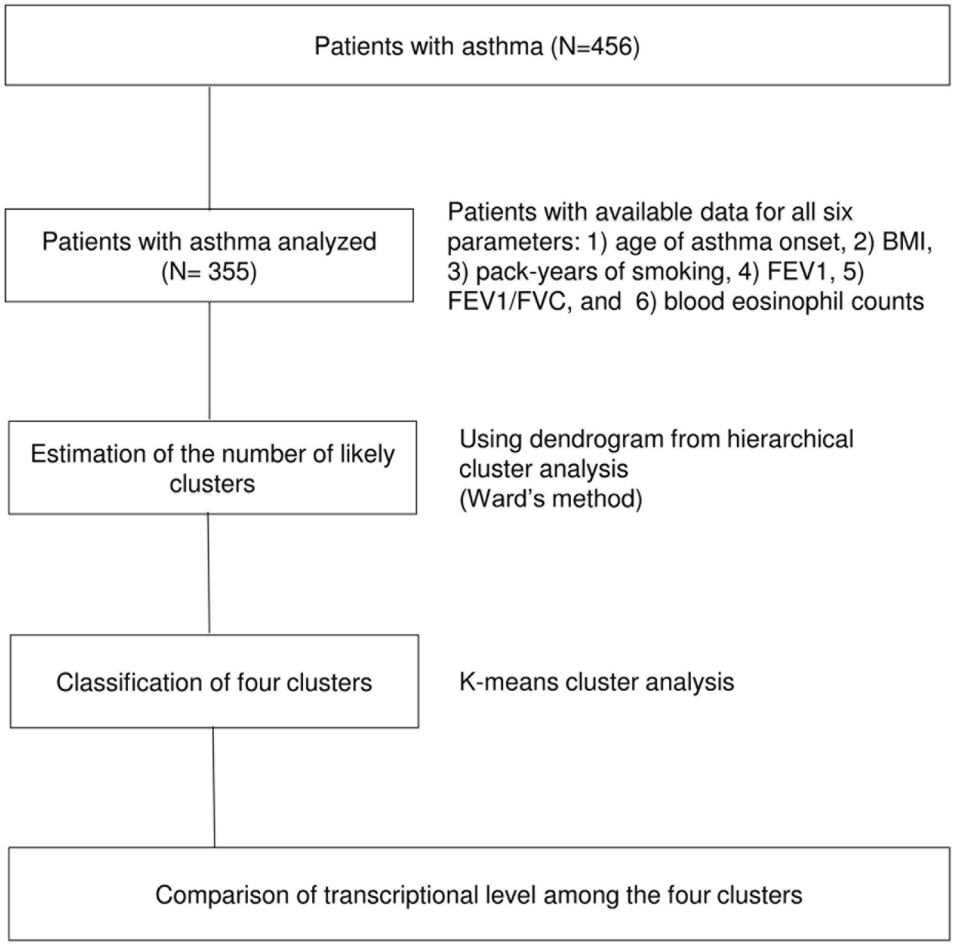

Fig. 1 Summary of the statistical methods. BMI, body index mass; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity

centrifuged at 12 000×*g* for 15 min at 4°C, and the aqueous phase was transferred to a 1.5 mL tube. Following that, 500 µL of isopropanol was added, vortexed for 5 s, and incubated at RT for 7 min. The mixture was centrifuged at 12 000×*g* for 8 min at 4°C, and the supernatant was discarded. Subsequently, 1 mL of 75% ethanol was added, and the mixture was centrifuged at 7500×*g* for 5 min at 4°C. The ethanol was removed, and the RNA pellet was briefly air-dried. Finally, the RNA pellet was dissolved in nuclease-free water.[22]

RNA exome sequencing

The assessment of RNA purity was conducted by analyzing 1 µL of the total RNA extract using a NanoDrop 8000 spectrophotometer. Additionally, the integrity of the total RNA was evaluated using an Agilent Technologies 2100 Bioanalyzer, which provided 2 key metrics: the RNA Integrity Number (RIN) and the percentage of RNA fragments >200 nucleotides in the fragment distribution value (DV200).

RNA exome sequencing libraries were prepared according to the manufacturer's instructions (Swift RNA Library Kit and IDT Exome v2.0 kit).[23] The process commenced with the fragmentation of 25 ng of total RNA, resulting in the generation of small RNA fragments typically ranging from 200 to 250 bp in length. Subsequently, the fragmented RNA underwent random priming and reverse transcription to produce the first-strand cDNA. Simultaneously, tailing and ligation were performed to incorporate the cleaved i7 adapter into the 3′ end of the cDNA molecule. The extension step led to the formation of a dsDNA duplex for adapter ligation, thereby appending truncated i5 adapter to the 3′ ends of the primer-extended cDNA molecules.

The constructed libraries were quantified based on the absorbance using a Qubit® 2.0 fluorometer with a Quant iT dsDNA HS Assay Kit (Thermo Fisher). As per the protocol for the IDT Exome v2.0, hybridization required a minimum of 500 ng of each library. The pre-hybridization solution, including library and blocker components, was dried in the tubes designated for the hybridization reaction employing a SpeedVac system. Subsequently, the hybridization mix was then added to the desiccated capture library. The resulting mixture was then incubated for 16 h at 65°C within a thermal cycler.

Captured libraries were subsequently subjected to PCR-amplification to enrich the fragments, using the KAPA HiFi HotStart Ready Mix and an eleven-cycle PCR program. The quality of the amplified libraries was verified using automated electrophoresis (Tapestation; Agilent Technologies). After the performance of qPCR using KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems), we combined libraries that were index tagged in equimolar amounts in the pool. RNA sequencing was performed using an Illumina NovaSeq 6000 (DNA-LINK, Inc., Seoul, Korea).

Bioinformatics analysis

A total of 456 samples were sequenced using Illumina NovaSeq 6000. The sequenced data were converted to count data using the following steps. Adapter sequences were trimmed by Trimmomatic (v0.39) with a phred score of 33 and the paired end mode setting.[24] The trimmed reads were aligned using HISAT2 (v2.2.0) with reference to the hg19 genome.[25] Finally, the aligned data were sorted using samtools (v1.9)[26] and counted using HTSeq (v0.12.4)[27]. The reference genome and gene annotation file (hg19.refGene.gft) were downloaded from the UCSC Genome Browser (https://genome.ucsc.edu/). Pathway enrichment analysis was performed using DAVID Bioinformatics Resources (https://david.ncifcrf.gov) and Enrichr guideline.[28] DEGs were mapped to the GO database, and the number of genes corresponding to each entry was calculated. Enrichment analysis was conducted using the KEGG[29] and REACTOME pathway databases (https://reactome.org) for DEGs.

Statistical analysis

Analysis of variance (ANOVA) for parametric data and Kruskal-Wallis tests for nonparametric data were used to determine the significance of differences among the 4 clusters. Bonferroni post hoc analysis was conducted to identify between-group differences. LIMMA[30] was performed on data adjusted for age, sex, and BMI, and transcript expression was compared between clusters. Batch effects in the high-throughput

experiments were corrected. *P* values were determined using Fisher's exact and binomial tests. To further enhance the accuracy and reliability of identifying differentially expressed genes (DEGs) as markers distinct from each cluster, the Benjamini-Hochberg method was applied for false discovery rate correction. All calculations were performed using R software (version 4.1.1). $P < .05$ was considered significant.

RESULTS

Study population

The initial dataset included 456 subjects; however, the final analysis included 355 individuals who had complete data for the 6 clinical variables used in the cluster analysis; these variables include age of asthma onset, BMI, pack-years of smoking, FEV1, FEV1/FVC, and blood eosinophil counts (Fig. 1). A four-cluster model that best fit the study subjects was determined via hierarchical cluster analysis using Ward's method and K-means analysis (Fig. 2A and B). The clinical characteristics of the study population are presented in Table 1. Among the patients with asthma, the mean age (SD) of 49.6 (15.6) years and the mean BMI (SD) of 25 (4.4). The study group included 204 (57.5%) females, with 164 (46.2%) individuals being atopic. Additionally, there were 111 (31.3%) ex-smokers and 44 (12.4%) current smokers. Notably, 72 (20.3%) patients had severe asthma. The mean predicted FEV1 % predicted (SD) stood at 75.1 (18.3) (Table 1).

Phenotypic characteristics of the clinical clusters

Four clusters were identified using the cluster approach outlined in the methods section. The detailed characteristics of the 4 clusters are presented in Table 1.

Cluster 1

Eighteen percent of subjects (n = 63) were grouped into Cluster 1, which featured "smoking asthma with neutrophilic inflammation." This cluster was characterized by older, predominantly male (74.6%) subjects with a heavy smoking history, and lower potential of being atopic (33.3%). The mean age (SD) of asthma onset was 51.3 (14.9) and the BMI for the subjects in this cluster was the highest (27.0 ± 6.0) among the clusters. This result differs from that of other clusters with severe airflow limitation and neutrophil-dominant airway inflammation. The baseline predicted FEV1 was 56.3 ± 13.0 and FEV1/FVC for airway obstruction was 58.8 ± 9.9. Subjects in this cluster had

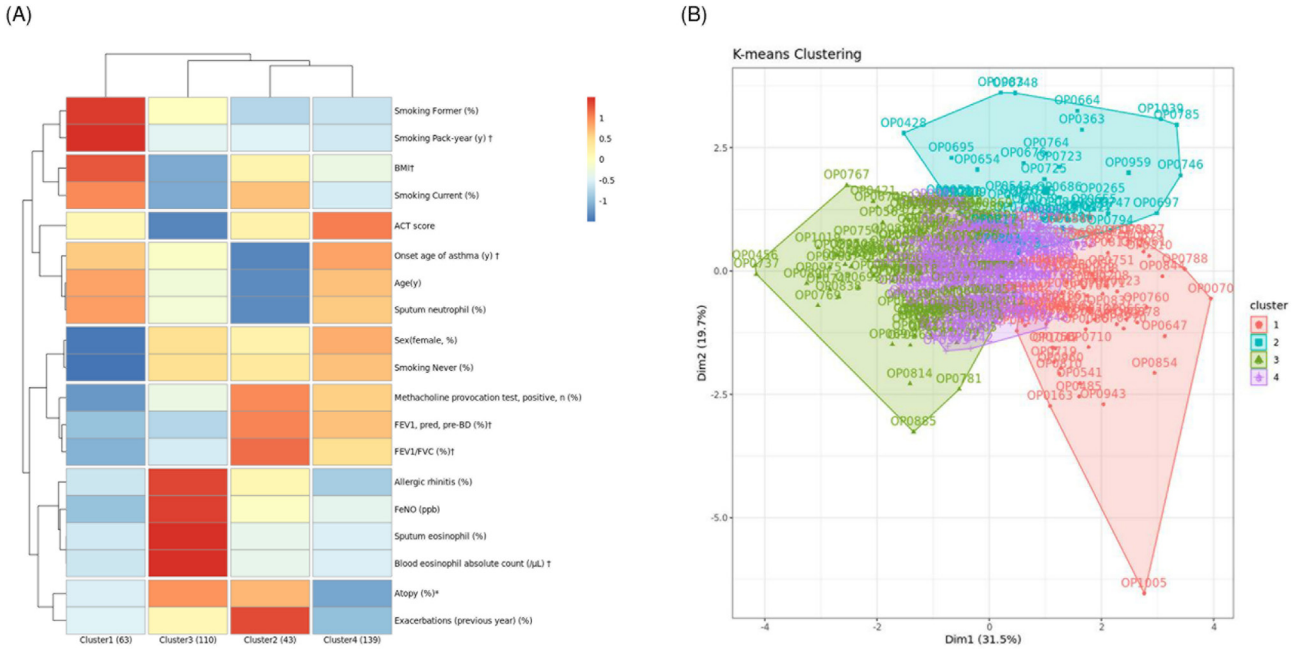

Fig. 2 Cluster analysis. (A) Heatmap showing the clustering results and (B) K-means clustering

	Total (n = 355)	Cluster 1 (n = 63)	Cluster 2 (n = 43)	Cluster 3 (n = 110)	Cluster 4 (n = 139)	*P*
Age (y)	49.6 ± 15.6	60.1 ± 13.5	47.1 ± 10.8	33.8 ± 11.0	58.1 ± 9.3	<.001
Sex (female, %)	204 (57.5)	16 (25.4)	27 (62.3)	64 (58.2)	97 (69.8)	<.001
BMI[b]	25.0 ± 4.4	27.1 ± 6.0	23.1 ± 4.1	25.3 ± 4.6	24.5 ± 3.0	<.001
Age of asthma onset (y)[b]	43.1 ± 17.1	51.3 ± 14.9	41.9 ± 11.0	24.5 ± 10.1	54.6 ± 9.7	<.001
Smoking						<.001
Never (%)	200 (56.3)	12 (19.0)	27 (62.8)	66 (60.0)	95 (68.4)	
Former (%)	111 (31.3)	38 (60.3)	15 (34.9)	24 (21.8)	34 (24.5)	
Current (%)	44 (12.4)	13 (20.6)	1 (2.3)	20 (18.2)	10 (7.2)	
Pack-year (y)[b]	16.8 ± 12.6	24.0 ± 16.3	14.8 ± 7.4	12.4 ± 8.9	13.7 ± 8.0	<.001
Atopy (%)[a]	164 (46.2)	21 (33.3)	25 (58.1)	71 (64.5)	47 (33.8)	<.001
Allergic rhinitis (%)	174 (49.0)	24 (38.1)	22 (51.2)	57 (51.8)	71 (51.1)	0.254
Exacerbations (previous year) (%)	75 (21.1)	9 (14.3)	10 (23.3)	27 (24.5)	29 (20.9)	0.369
ICS dose (%)						0.773
Low	164 (48.0)	27 (42.9)	19 (45.2)	53 (51.5)	65 (48.1)	
Moderate to high	178 (52.0)	35 (56.5)	23 (54.8)	50 (48.5)	70 (51.9)	
ACT score	16.8 ± 4.5	16.5 ± 4.7	14.7 ± 3.6	16.6 ± 4.6	17.6 ± 4.3	0.006
Severity						<.001
Mild (%)	160 (45.1)	3 (4.8)	8 (18.6)	72 (65.5)	77 (55.4)	
Moderate (%)	123 (34.6)	24 (38.1)	10 (23.3)	36 (32.7)	53 (38.1)	
Severe (%)	72 (20.3)	36 (57.1)	25 (58.1)	2 (1.8)	9 (6.5)	
Total IgE (kU/L)	409.4 ± 585.7	444.5 ± 656.4	450.3 ± 492.5	454.6 ± 678.0	344.8 ± 496.5	0.095
FEV1, pred, pre-BD (%)[b]	75.1 ± 18.3	56.3 ± 13.0	58.7 ± 17.8	85.6 ± 13.2	80.2 ± 12.7	<.001
FEV1/FVC (%)[b]	72.1 ± 12.6	58.8 ± 9.9	63.6 ± 10.7	80.6 ± 11.5	73.7 ± 7.0	<.001
FeNO (ppb)	51.6 ± 43.2	41.0 ± 42.0	71.6 ± 54.2	52.8 ± 40.0	47.8 ± 39.8	0.002
Methacholine provocation test, positive, n (%)[c]	187 (53.9)	11 (17.7)	17 (40.5)	75 (71.4)	84 (60.9)	<.001
Sputum neutrophil (%)[d]	55.7 ± 32.1	63.1 ± 32.4	54.4 ± 28.2	45.3 ± 33.2	61.0 ± 30.4	0.001

Sputum eosinophil (%)[d]	11.3 ± 18.2	8.6 ± 16.6	24.6 ± 22.2	10.3 ± 18.9	9.2 ± 15.1	<.001
Blood neutrophil absolute count (/μL)	4216.2 ± 2223.4	5404.8 ± 2800.6	3684.3 ± 1822.5	4112.0 ± 2131.0	3932.8 ± 1945.8	<.001
Blood eosinophil absolute count (/μL)[b]	434.4 ± 488.6	270.4 ± 221.8	1237.8 ± 882.8	374.5 ± 285.3	307.6 ± 237.2	<.001

Table 1. Phenotypic characteristics of the clinical clusters. BMI, *body mass index*; ICS, *inhaled corticosteroid*; ACT, *Asthma control test*; FEV1, *forced expiratory volume in 1 s*; BD, *bronchodilator*; FVC, *forced vital capacity*; FeNO, *fraction of exhaled nitric oxide*.Values reported as means ± SD. [a]Atopy is defined as positive skin prick test reactivity to at least 1 common aeroallergen or a history of atopic dermatitis or allergic rhinitis. [b]Variables used for clustering. [c]347 patients underwent a methacholine provocation test. [d]332 suitable sputum samples were analyzed

increased sputum neutrophil (63.1 ± 32.4%) with low level of sputum eosinophil percentages and absolute eosinophil counts compared with those in other clusters. Approximately 55.6% of these received regular moderate to high dose of inhaled corticosteroids (ICS) as controllers.

Cluster 2

Cluster 2 was the smallest cluster (n = 43, 12% of subjects) with "severe eosinophilic asthma" and comprised middle-aged subjects (mean age, 47 years), two-thirds female, with severe airflow limitation at baseline (predicted FEV1 58.7 ± 17.8). This group was distinguished by T2 inflammation with higher sputum eosinophil and blood eosinophil counts than the other clusters (sputum eosinophil, 24.6 ± 22.2%; blood eosinophil count, 1237.8 ± 882.8). Subjects in this cluster had high levels of exhaled nitric oxide (FeNO, 71.6 ± 54.2 ppb). Although approximately 53.5% of these subjects received regular moderate to high doses of ICS, this cluster had a low asthma control test (ACT) score (14.7 ± 3.6).

Cluster 3

This cluster (n = 110, 31% of subjects) consisted of younger subjects with an atopic feature, named "early-onset atopic asthma with normal lung function." This cluster had the youngest age of asthma onset (24.5 ± 10.1) among the clusters. The number of subjects with atopy was 71 (64.5%). The baseline predicted FEV1 was 85.6 ± 13.2 and FEV1/FVC was 80.6 ± 11.5.

Cluster 4

Cluster 4 had the largest number of subjects (n = 139, 39% of subjects) and featured "late-onset mild asthma." The cluster consisted of more women (69.8%) with mainly late-onset asthma (mean age of asthma onset, 54.6 ± 9.7) than the other clusters. In addition, subjects in this group had nearly normal lung function (predicted FEV1, 80.2 ± 12.7; FEV1/FVC, 73.7 ± 7.0).

Transcriptional differences between clinical clusters

Each cluster was considered a unique pathobiologic process and the transcriptome in PBMC was compared to determine the differences in

Gene Symbol	Gene name	Biologic Processes	Function	Log Fold Change	*P* value
B4GALNT1	Beta-1,4-N-Acetyl-Galactosaminyltransferase 1	Lipid metabolic process	Transferase activity	1.0924	0.0084
CFD	Complement Factor D	Complement activation	Protein binding	−1.0241	0.0084
GJC2	Gap Junction Protein Gamma 2	Cell communication, Response to toxic substance	Gap junction channel activity	1.0537	0.0112
GNG11	G Protein Subunit Gamma 11	Signal transduction	G-protein beta-subunit binding	1.0549	0.0112
HGH1	HGH1 Homolog	Biological process	NA	1.5674	0.0112
HLA-DRA	Major Histocompatibility Complex, Class II, DR Alpha	Antigen processing and presentation, Regulation of T-helper cell differentiation	Peptide antigen binding, T cell receptor binding	−1.2499	0.0146
HLA-DRB3	Major Histocompatibility Complex, Class II, DR Beta 3	Antigen processing and presentation, Immunoglobulin production	Peptide antigen binding	−1.1875	0.0176
KCTD8	Potassium Channel Tetramerization Domain Containing 8	Protein homooligomerization	Protein binding	1.7379	0.0208
MARVELD1	MARVEL Domain Containing 1	Cell cycle, myelination	Structural constituent of myelin sheath	−1.1181	0.0223
ONECUT2	One Cut Homeobox 2	Regulation of cell migration	DNA binding	−1.0613	0.0251
PPBP	Pro-Platelet Basic Protein	Chemotaxis	Transmembrane transporter activity, chemokine activity	1.6594	0.0239
PTF1A	Pancreas Associated Transcription Factor 1a	DNA-dependent regulation of transcription, Neuron generation	NA	−1.0614	0.033
SERPINB12	Serpin Family B Member 12	Regulation of peptidase activity	Peptidase inhibitor activity, enzyme binding	−1.6900	0.0294

SNORD35B	Small Nucleolar RNA, C/D Box 35B	RNA processing	NA	1.4851	0.0308
SOWAHD	Sosondowah Ankyrin Repeat Domain Family Member D	NA	Protein binding	1.3325	0.0351
SOX17	SRY-Box Transcription Factor 17	Angiogenesis, Vasculogenesis, endoderm formation	DNA binding	1.5652	0.0351
TPTEP1	TPTE Pseudogene 1	NA	NA	−1.1109	0.0351

Table 2. Differentially expressed genes in cluster 1 compared with cluster 2. *NA, not applicable. P values indicate that the difference is still significant, even when correcting for the false discovery rate using the Benjamini-Hochberg method.The fold change (FC) is presented in the base 2 logarithm. If it is positive in X versus Y, the analyte is more abundant in X in Y; negative values indicate less abundance*

gene expression between the clusters. A total of 309 transcripts were found to be significantly differentially expressed in at least 1 of the pairwise comparisons between clusters. Significantly expressed genes were only found in 2 comparisons: cluster 1 vs 2 and cluster 2 vs 3. There were no significant transcriptional differences across clusters or between the other 2 clusters. The results are presented in Tables 2 and 3, respectively. Many genes were upregulated between clusters 1 and 2. In particular, 17 DEGs were identified (9 upregulated and 8 downregulated). Smaller differences were found between clusters 2 and 3 as only 2 DEGs were identified (1 upregulated and 1 downregulated) (Fig. 3A and B).

Pathway analysis of the DEGs

GO analysis was performed using the DEGs identified through sequencing. A total of 24 GO terms for biological processes were found between clusters 1 and 2, mainly cellular processes, such as "regulation of T-helper cell differentiation," "interferon gamma (IFN-γ)- mediated signaling pathway," and "cellular response to IFN-" (Fig. 4A). KEGG pathway enrichment analysis revealed that DEGs between cluster 1 and cluster 2 were significantly enriched in biological pathways, including "Th17 cell differentiation," "Th1 and Th2 cell differentiation," "Chemokine signaling pathway," and "asthma" (Table 4 and Fig. 4B). The DEGs between cluster 2 and cluste 3 were significantly enriched in biological pathways, such as "NOD-like receptor signaling pathway" and "Calcium signaling pathway" (Table 4).

DISCUSSION

In this study, 4 clinical phenotypes were identified using cluster analysis based on age of asthma onset, BMI, pack-years of smoking, FEV1 as a percentage of predicted value, FEV1/FVC, and blood eosinophil counts. The following names were assigned to the 4 clusters: smoking asthma with neutrophilic inflammation (cluster 1), severe eosinophilic asthma (cluster 2), early onset atopic asthma with normal lung function (cluster 3), and late-onset mild asthma (cluster 4). Transcriptomic profiling of the PBMC revealed genes and pathways that differentiate the clusters, especially between neutrophilic and eosinophilic asthma.

Gene Symbol	Gene name	Biologic Processes	Function	Log Fold Change	*P* value
LINC02652	Long Intergenic Non-Protein Coding RNA 2652	NA	NA	0.8757	0.0248
P2RX7	Purinergic Receptor P2X 7	Positive regulation of cytokine production, positive regulation of protein phosphorylation, Calcium ion transport, positive regulation of the apoptotic process	Protein binding	−0.7822	0.05

Table 3. Differentially expressed genes in cluster 2 compared with cluster 3. *NA, not applicable. P values indicate that the difference is still significant, even when correcting for the false discovery rate using the Benjamini-Hochberg method.The fold change (FC) is presented in the base 2 logarithm. If it is positive in X versus Y, the analyte is more abundant in X in Y; negative values indicate less abundance*

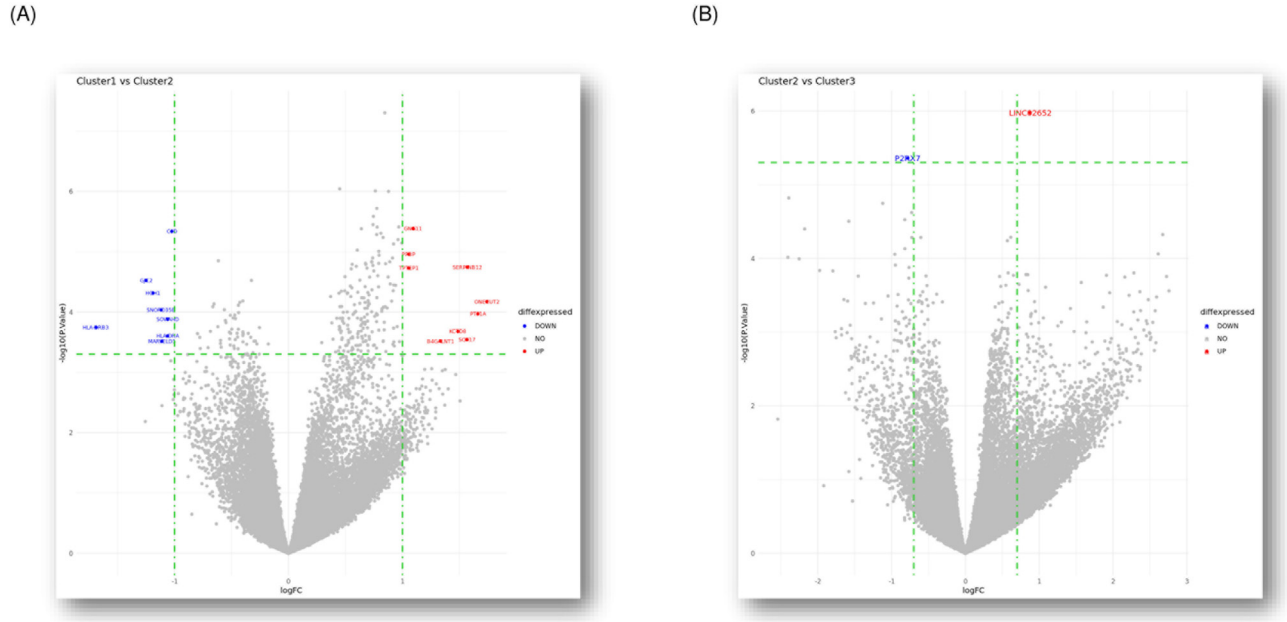

Fig. 3 Volcano plots based on the transcriptomics data between clusters. Volcano plots represent -log10 (*P*-value) for genome-wide genes (Y-axis) plotted for each log2 (fold change) (X-axis). Red and blue dots indicate genes that are significantly up (red)- and down (blue)-regulated between (A) cluster 1 vs. cluster 2 and (B) cluster 2 vs. cluster 3, respectively. Green dot lines indicate the cut-off criteria of 5% false discovery rate (horizontal) and a log2 fold change (Log2FC) of 1 (vertical).

Recent studies revealed different asthma phenotypes through cluster analysis using clinical and demographic characteristics, owing to the heterogeneity and diverse nature of the disease.[4,6,31] Previously, we reported 4 asthma clusters based on clinical variables such as FEV1 predicted, BMI, onset age, skin test, smoking history, and history of exacerbation, in 2 independent Korean adult asthma cohorts: smoking asthma, early onset atopic asthma, severe obstructive asthma, and late-onset mild asthma.[6] However, the previous study lacked specific information on airway inflammation. This study is a valuable addition to the previous study as specific data on the patterns of eosinophilic inflammation were employed in the cluster analysis.

The present study aimed to investigate the transcriptional differences among the 4 clusters,

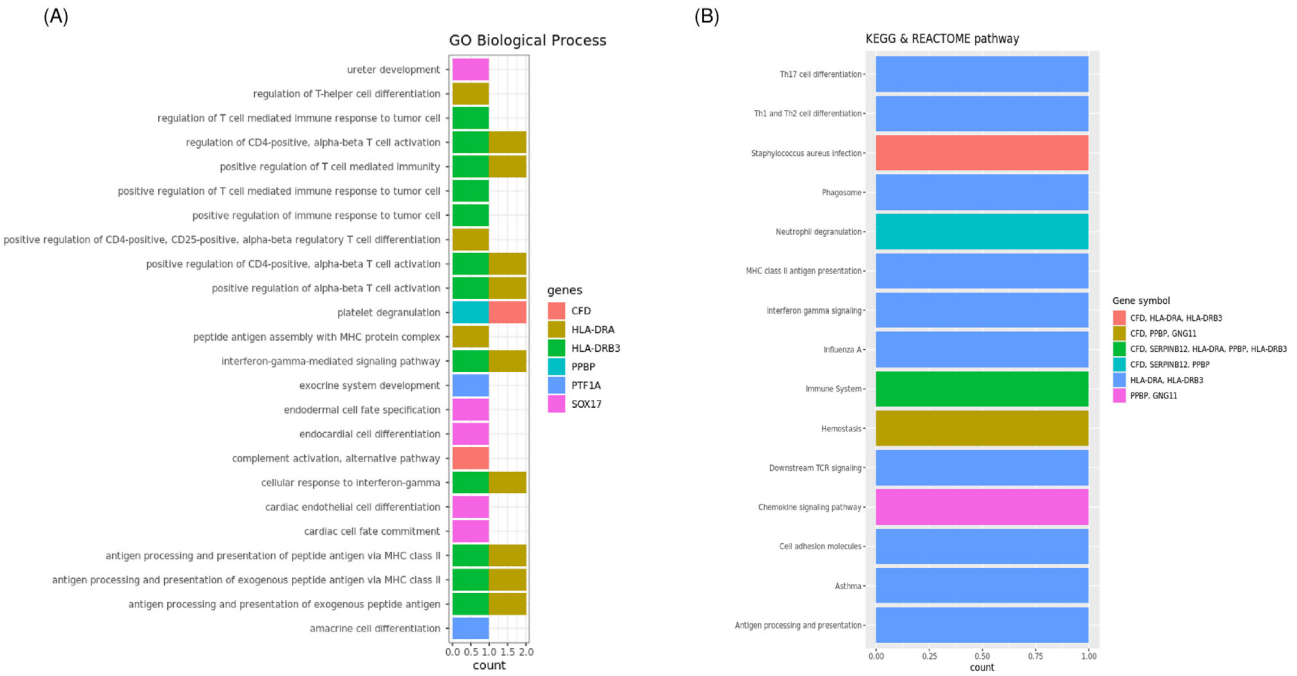

Fig. 4 Bar charts for the top 17 (A) Gene Ontology and (B) KEGG and REACTOME pathway enrichment analysis terms based on the differentially expressed genes between cluster 1 and cluster 2

particularly between Cluster 1 (neutrophil-enriched type) and Cluster 2 (eosinophil-enriched type). Cluster 1 was mainly composed of male smokers and may be proposed to have a chronic obstructive pulmonary disease (COPD) overlap (ACO). The subjects in this cluster exhibited more severe airflow obstruction with neutrophilic asthma than those in other clusters. Meanwhile, cluster 2, with T2 inflammation patterns and worsened airflow limitation, could be considered as representing severe eosinophilic asthma. Differential gene expression between clusters 1 and 2 revealed upregulated genes in cluster 1 that were related to cell chemotaxis and response to toxic substances, indicating active cellular recruitment. Especially, HLA-DRA and HLA-DRB3 are known to play important roles in Th17, Th1, and Th2 inflammatory pathways. Other genes (eg, CFD, SERPINB12, PPBP, and GNG11) can also influence the pathogenesis of asthma by regulating various inflammatory responses.[32] Recent studies have demonstrated that genes related to cell chemotaxis are upregulated in the neutrophilic inflammation group based on transcriptional clustering,[33,34] which is consistent with our results.

Recent studies focused on the increased Th17/IL-17 and Th1/IFN-γ responses in neutrophilic asthma, which is associated with corticosteroid-resistant asthma.[35] In this study, differences in pathways related to Th17 cell differentiation and IFN-γ-mediated signaling were found between cluster 1 (neutrophilic) and cluster 2 (eosinophilic) DEGs. Th17 cells produce IL-17 A, IL-17F, and IL-22 and are primarily involved in forms of asthma in which neutrophils contribute more to inflammation than eosinophils.[36,37] IFN-γ was demonstrated to be most commonly associated with neutrophilic airway inflammation.[38] Some studies using transcriptome-associated clusters with induced sputum reported upregulation of IFN- and TNF-associated genes and IL-1 and TNF-α/NF-κB pathways in clusters related to neutrophilic asthma.[10,12] The present data supported these findings as relevant mechanisms in neutrophilic asthma and further suggested specific molecular pathways involving the IFN-γ-mediated signaling.

Interestingly, although the clinical characteristics allowed discrimination between cluster 3 (named "atopic asthma with normal lung function") and cluster 2 (named "severe eosinophilic asthma"), few significant transcriptional differences were identified in this comparison. These 2 phenotypes have been considered to markedly overlap as subtypes

Pathway	No. of overlapping/ total genes	*P* value		Gene symbol
		Cluster 1 vs. 2	Cluster 2 vs. 3	
Staphylococcus aureus infection[a]	3/96	0.0997		CFD, HLA-DRA, HLA-DRB3
Asthma[a]	2/31	0.2489		HLA-DRA, HLA-DRB3
Antigen processing and presentation[a]	2/78	0.2489		HLA-DRA, HLA-DRB3
Th1 and Th2 cell differentiation[a]	2/92	0.2489		HLA-DRA, HLA-DRB3
Th17 cell differentiation[a]	2/108	0.2489		HLA-DRA, HLA-DRB3
Cell adhesion molecules[a]	2/148	0.0183		HLA-DRA, HLA-DRB3
Phagosome[a]	2/152	0.0183		HLA-DRA, HLA-DRB3
Influenza A[a]	2/172	0.0220		HLA-DRA, HLA-DRB3
Chemokine signaling pathway[a]	2/192	0.0247		PPBP, GNG11
Neutrophil degranulation[b]	3/482	0.7178		CFD, SERPINB12, PPBP
Immune System[b]	5/2016	0.7178		CFD, SERPINB12, HLA-DRA, PPBP, HLA-DRB3
Interferon gamma signaling[b]	2/91	0.7178		HLA-DRA, HLA-DRB3
Downstream TCR signaling[b]	2/98	0.7178		HLA-DRA, HLA-DRB3
Hemostasis[b]	3/621	0.0896		CFD, PPBP, GNG11
MHC class II antigen presentation[b]	2/123	0.0972		HLA-DRA, HLA-DRB3
NOD-like receptor signaling pathway[a]	1/181		0.0338	*P2RX7*
Calcium signaling pathway[a]	1/240		0.0338	*P2RX7*
Neuroactive ligand-receptor interaction[a]	1/341		0.0338	*P2RX7*

Table 4. Top pathways enriched by the top differentially expressed genes between clusters *No, number. P-values were corrected using the Benjamini-Hochberg method. [a]KEGG pathway. [b]REACTOME pathway*

of allergic asthma.[39] Cluster 1, 2, and 3 showed no significant transcriptional differences as compared to cluster 4, which represents late-onset mild asthma. As this study was a cross-sectional study performed with a small number of asthmatics, a large-scale prospective asthma transcriptomics study is needed to obtain distinct results.

To our knowledge, this study is the first to confirm the usefulness of PBMC for transcriptomic analysis of Korean asthmatics. Blood expresses approximately 80% of the genes encoded by the human genome and contains many cells involved in immune responses.[40] In addition, blood is easily accessible. Currently, available transcriptome

studies in asthmatics were mainly performed with airway epithelium and induced sputum.[10,41] However, a study comparing the transcriptome profiles in PBMC from 118 adult asthmatics between stable and exacerbation status revealed that PBMC could reflect systemic changes according to asthma exacerbation at the gene expression level.[13] This study supports the hypothesis that PBMCs cab recapitulate airway inflammation and biological processes, and that transcriptomic analysis using of PBMCs can yield similar results to those obtained from airway samples.

The present study had several limitations. First, cluster analysis with several clinical variables may not be appropriate for the objective classification of asthma with heterogenous features. However, the clinical variables used for this cluster analysis were considered to have significant influences on asthma. The appropriateness of the clinical variables used must be verified. Second, airway samples such as sputum, bronchial epithelium, and bronchoalveolar lavage fluid, were not utilized for the transcriptomic analysis. Future studies are necessary to establish a direct link between blood and airway transcriptome profiles. Third, this study exclusively focused on patients diagnosed with asthma and did not include a healthy control group. Furthermore, there were no comparative results for clusters 3 and 4. Fourth, the identified clinical and molecular phenotypes of asthma require external validation in an independent asthma cohort; however, this was not conducted in this study. Future external validation is necessary to expand the scope of our research. Fifth, in the comparison between clusters 2 and 3, 3 pathways associated with a single gene were identified, and the inability to identify impact of this gene might limit interpretation. Lastly, the contribution of environmental exposure to asthma development has not been fully assessed. As asthma is a heterogeneous disease with gene-environment interactions, further studies are required to evaluate gene-environment associations via epigenetic analysis in a larger sample of Korean asthmatics.

In conclusion, we have identified 4 clinical clusters and confirmed transcriptional differences in PBMC between these clusters in Korean asthmatics. Furthermore, for asthmatics with severe airflow limitation, differences in DEGs between eosinophil-enriched and neutrophil-enriched clusters were found to be related to chemotaxis and immune and inflammatory responses, including IFN-γ-mediated signaling pathways. Overall, analyzing transcriptional levels in blood may provide valuable insights into the molecular mechanisms of asthma phenotypes.

Acknowledgements
This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF), funded by the Korean government (MSIT) (2019M3E5D3073365).

Funding
This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF), funded by the Korean government (MSIT) (2019M3E5D3073365).

Availability of data and materials
The data used or analyzed during the study is available upon reasonable request from the corresponding author.

Authors' contributions
All authors have contributed to the content of the manuscript, and the respective roles of each author are as follows:

Study conception, design, image assessment, and supervision: Tae-Bum Kim, Sungho Won.

Study design, review of experiments, data analysis and interpretation, image assessment, and writing of the draft: Jin An, Seungpil Jeong.

Review of image assessment, and data analysis and interpretation: Jin An, Seungpil Jeong, Kyungtaek Park, Heejin Jin, Jaehyun Park, Eunsoon Shin, Ji-Hyang Lee, Woo-Jung Song, Hyouk-Soo Kwon, You Sook Cho, Jong Eun Lee.

All authors revised the manuscript.

Ethics approval
All patients provided written informed consent at the time of cohort enrollment, and this study was approved by the institutional review boards of Asan Medical Center (2019-0376).

Consent for publication
The manuscript's publishing is approved by all of the authors.

Declaration of competing interest
The authors declare that they have no competing interests.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.waojou.2024.100871.

Author details
[a]Department of Pulmonary, Allergy and Critical Care Medicine, College of Medicine, Kyung Hee University Hospital at Gangdong, Kyung Hee University, Seoul, South Korea. [b]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, South Korea. [c]Institute of Health and Environment, Seoul National University, Seoul, South Korea. [d]Interdisciplinary Program of Bioinformatics, Seoul National University, Seoul, South Korea. [e]DNA Link, Seoul, South Korea. [f]Department of Allergy and Clinical Immunology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea. [g]Interdisciplinary Program of Bioinformatics, College of Natural Science, Seoul National University, Seoul, South Korea.

REFERENCES

1. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med*. 2012;18:716–725.
2. Proceedings of the ATS workshop on refractory asthma: current understanding, recommendations, and unanswered questions. American Thoracic Society. *Am J Respir Crit Care Med*. 2000;162:2341–2351.
3. Montuschi P, Barnes PJ. New perspectives in pharmacological treatment of mild persistent asthma. *Drug Discov Today*. 2011;16:1084–1091.
4. Haldar P, Pavord ID, Shaw DE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008;178: 218–224.
5. Bradding P, Green RH. Subclinical phenotypes of asthma. *Curr Opin Allergy Clin Immunol*. 2010;10:54–59.
6. Kim T-B, Jang A-S, Kwon H-S, et al. Identification of asthma clusters in two independent Korean adult asthma cohorts. *Eur Respir J*. 2013;41:1308–1314.
7. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363:1211–1221.
8. Anenberg SC, Henze DK, Tinney V, et al. Estimates of the global burden of ambient [formula: see text], ozone, and [formula: see text] on asthma incidence and emergency room visits. *Environ Health Perspect*. 2018;126, 107004.
9. Yan X, Chu JH, Gomez J, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. *Ann Am Thorac Soc*. 2016;13(Suppl 1): S104–S105.
10. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol*. 2011;127:153–160, 160 e151-159.
11. Kuo C-HS, Pavlidis S, Loza M, et al. A transcriptome-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in U-BIOPRED. *Am J Respir Crit Care Med*. 2017;195:443–455.
12. Kuo C-HS, Pavlidis S, Loza M, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J*. 2017;49.
13. Bjornsdottir US, Holgate ST, Reddy PS, et al. Pathways activated during human asthma exacerbation as revealed by gene expression patterns in blood. *PLoS One*. 2011;6, e21902.
14. Coates AL, Wanger J, Cockcroft DW, et al. ERS technical standard on bronchial challenge testing: general considerations and performance of methacholine challenge tests. *Eur Respir J*. 2017;49.
15. An J, Lee J-H, Sim JH, et al. Serum eosinophil-derived neurotoxin better reflect asthma control status than blood eosinophil counts. *J Allergy Clin Immunol Pract*. 2020;8:2681–2688. e2681.
16. Carr TF, Zeki AA, Kraft M. Eosinophilic and noneosinophilic asthma. *Am J Respir Crit Care Med*. 2018;197:22–37.
17. Seys SF, Lokwani R, Simpson JL, Bullens DMA. New insights in neutrophilic asthma. *Curr Opin Pulm Med*. 2019;25:113–120.
18. Chung KF, Wenzel SE, Brozek JL, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2014;43:343–373.
19. Scrucca L, Fop M, Murphy TB, Raftery AE. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*. 2016;8:289.
20. Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans Pattern Anal Mach Intell*. 2000;22:719–725.
21. Bhatnagar SR, Yang Y, Khundrakpam B, et al. An analytic approach for interpretable predictive models in high-dimensional data in the presence of interactions with exposures. *Genet Epidemiol*. 2018;42:233–249.
22. Nojima T, Gomes T, Carmo-Fonseca M, Proudfoot NJ. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat Protoc*. 2016;11:413–428.
23. Wan Y, Qu K, Ouyang Z, Chang HY. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat Protoc*. 2013;8:849–869.
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30: 2114–2120.
25. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–915.
26. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
27. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–169.
28. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–W97.

29. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
30. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47. e47.
31. Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med*. 2010;181:315–323.
32. Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy Asthma Immunol Res*. 2019;11:170–187.
33. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol*. 2011;127:153–160. e159.
34. Baines KJ, Simpson JL, Bowden NA, Scott RJ, Gibson PG. Differential gene expression and cytokine production from neutrophils in asthma phenotypes. *Eur Respir J*. 2010;35:522–531.
35. Ray A, Kolls JK. Neutrophilic inflammation in asthma and association with disease severity. *Trends Immunol*. 2017;38:942–954.
36. Cosmi L, Liotta F, Maggi E, Romagnani S, Annunziato F. Th17 cells: new players in asthma pathogenesis. *Allergy*. 2011;66:989–998.
37. Ciepiela O, Ostafin M, Demkow U. Neutrophils in asthma–a review. *Respir Physiol Neurobiol*. 2015;209:13–16.
38. Carr TF, Zeki AA, Kraft M. Eosinophilic and noneosinophilic asthma. *Am J Respir Crit Care Med*. 2018;197:22–37.
39. Tran TN, Zeiger RS, Peters SP, et al. Overlap of atopic, eosinophilic, and TH2-high asthma phenotypes in a general population with current asthma. *Ann Allergy Asthma Immunol*. 2016;116:37–42.
40. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med*. 2006;147:126–132.
41. Tsai YH, Parker JS, Yang IV, Kelada SNP. Meta-analysis of airway epithelium gene expression in asthma. *Eur Respir J*. 2018;51.