의학박사 학위논문

Analysis of biological pathways in elderly with asthma; sputum transcriptomic analysis

객담 전사체 분석을 통한 노인 천식의 생물학적 경로 분석

2019년 2월

서울대학교 대학원

의과대학 의학과 중개의학전공

김병근

Analysis of biological pathways in elderly with asthma; sputum transcriptomic analysis

지도교수 조상헌

이 논문을 의학박사 학위논문으로 제출함
2019년 1월

서울대학교 대학원
의과대학 의학과 중개의학전공
김병근

김병근의 의학박사 학위논문을 인준함
2019년 1월

위 원 장 이 은봉 (인)
부위원장 조상헌 (인)
위 원 박 홍석 (인)
위 원 강 혜련 (인)
위 원 이 병재 (인)

Abstract

Analysis of biological pathways in elderly with asthma; sputum transcriptomic analysis

Byung Keun Kim

Translational Medicine, Department of Medicine

The Graduate School

Seoul National University

Background: Elderly asthma (EA) shows characteristics different from those of conventional asthma. EA is increasing but its specific pathogenesis remains unclear. Currently, systems biology is widely used in biological research because of the rapid advancement of high-throughput technologies. This study aimed to identify EA-related biological pathways by analyzing genome-wide gene expression profiles in sputum cells using systems biology techniques.

Methods: We analyzed gene expression profiles in induced sputum of EA patients and healthy elderly controls. A total of 3,156 gene probes with significantly different expressions between the two groups were identified. We performed hierarchical clustering of genes to classify EA patients. Gene set enrichment analysis (GSEA) and weighted gene co-expression network analysis (WGCNA) were both performed to provide biological information. We also replicated our results using public gene expression data available from

Gene Expression Omnibus.

Results: Fifty-five EA patients and ten elderly control subjects were enrolled. Two distinct gene clusters were found. Cluster 1 (n=35) showed a lower eosinophil proportion in sputum and less severe airway obstruction compared to cluster 2 (n=20). The replication data set also identified two gene clusters (Cluster 1` and Cluster 2`). We found five gene sets enriched in cluster 1 and three gene sets enriched in cluster 2. Among these, we confirmed that two gene sets were significantly enriched in the replication data set (OXIDATIVE_PHOSPHORYLATION (OXPHOS) gene set in Cluster 1` and EPITHELIAL_MESENCHYMAL_TRANSITION (EMT) gene set in Cluster 2`). These were also enriched in a subgroup analysis that consisted of individuals who had never smoked. We also found four leading edge genes (*MRPS11*, *HSPA9*, *NUDF4*, and *ACTA1*) in the OXPHOS gene set and two (*SNTB1* and *FUCA1*) in the EMT gene set. WGCNA revealed four modules in cluster 1 and 18 modules in cluster 2. The brown module of cluster 1 and the magenta module of cluster 2were correlated with FEV1/FVC ratio in EA patients. These two modules were also replicated using the replication data set.

Conclusion: The findings of two distinct gene clusters in EA and different biological pathways within each gene cluster suggest two different pathogenic mechanisms underlying EA. We postulate that oxidative stress and cellular senescence-associated with aging may be important in the development or progress of EA, and these could be an important development for effectively treating EA.

Key words: Cluster analysis, Elderly asthma, Gene set enrichment analysis, Genetic pathway, Systems biology, Transcriptomics, Weighted gene co-expression network analysis

Student Number: 2015-30003

Contents

Introduction

The global human society is aging, and the population aged 65 and over is expected to expand by 150% in the next 35 years (1). Accordingly, elderly asthma (EA), encompassing asthma in people aged 65 years and over, is expected to increase dramatically. The estimated prevalence of EA in developed countries is 6-10% (2). Two-thirds of deaths attributed to asthma occur in people suffering from EA, and the mortality rate of EA is 7.3 deaths per 100,000 people (3).

Previous reports suggest that EA is phenotypically different from non-elderly asthma (NEA) (4-6). EA exhibits more heterogeneous characteristics and can be divided into long-standing asthma, wherein the patients suffer from the condition since childhood, and late-onset asthma, wherein the symptoms develop after the sixth decade of life (7). According to recent studies, the incidence of asthma is increasing and is the highest in elderly individuals, and more than two-thirds of these cases develop after the individuals reach 40 years of age (8). EA presents with characteristics different than those of NEA; its diagnosis is particularly challenging, and the symptoms of this disease often mimic alternative pathologies of elderly people, such as chronic obstructive pulmonary disease or congestive heart failure (7). The time required to achieve peak bronchodilator effect is 30 min, longer than 5-10 min in NEA (9), and the methacholine provocation test is less useful for the diagnosis of asthma in patients with a longer duration of asthma (10). Also, in EA patients, eosinophilic inflammation of airways is associated with airway hyper-responsiveness, and neutrophilic inflammation is determinant of airflow

limitation at rest and during bronchoconstriction (11).

These findings emphasize the observation that the pathogenesis of EA is different from that of typical allergic asthma, the latter being influenced by atopy, IgE, and Th2 responses. EA is an extraordinarily complex disease, and the pathogenesis underlying EA has not been clearly elucidated (12). There are several hypotheses concerning the development of EA. The first is based upon the relationship between the innate immune response and asthma. Environmental exposure to various proteases and to endotoxins such as those encountered from cigarette smoking can perpetuate this disease. Also chronic infection, colonization by bacteria or fungi, and viral infection can influence airway mucosa and accelerate submucosal hypertrophy and airway remodeling (12). Innate lymphoid cells (ILCs), a newly identified type of immune cell that belongs to the lymphoid lineage but does not respond in an antigen-specific manner, could also be a significant mediator of EA development. These cells are related to the development of asthma regardless of adaptive immune response, particularly in response to several stimuli such as several molecules, virus, mycobacterial infection that can cause repetitive epithelial damage (13). The second hypothesis concerns the role of microbiome in the development of EA. Evidence exists illustrating key roles for the microbiome in the development of asthma. Coupled with effects of immunosenescence and inflammation, the microbiome in EA is likely to influence the clinical phenotype observed in practice. If such associations are comparable in the context of late-onset disease, however, remains to be established (14). Recently, significant associations between Staphylococcal enterotoxin IgE (SE-IgE)-sensitization and late-onset asthma have been established (15), and this could

be an important clue for understanding the pathogenesis of EA.

As discussed above, a large amount of clinical and experimental evidence supports the idea that the mechanisms underlying EA are different from those of conventional NEA. Therefore, the mechanisms underlying EA should be studied from a new perspective. Given the current prevalence of genome-wide association studies (GWAS), many researchers are attempting to understand the pathophysiology and mechanisms regulating diseases by using both genomic and clinical data. Currently, the study of genetics in human diseases has become "big data science." The rapid advancement of high-throughput technologies has allowed the affordable profiling of genomes, transcripts, proteins, and metabolites. The systems approach has become much more attractive in the post-GWAS era (16). Systems biology approaches that integrate various indicators to facilitate analysis within one system have been widely used in biological research. As human biology and specific disease conditions cannot be fully understood in terms of cross-section, systems biology seeks to determine meaningful signals through objective mathematical analysis allowing for understanding of the intricate connections among the varied and large-scale biological data. Recently, methods such as computational unsupervised cluster analysis techniques, gene set enrichment analysis (GSEA), and weighted gene co-expression network analysis (WGCNA) have been widely used for genome expression studies.

Computational unsupervised cluster analysis techniques using co-relation networks can be performed using a three step process that includes pre-processing, cluster analysis, and cluster validation (17). This approach has been

widely used in many applications such as mRNA expression studies. Many tools for validation of cluster analysis have been used. Pvclust is an add-on package to assess the uncertainty in hierarchical cluster analysis for the popular statistical software R. Pvclust can easily perform bootstrap analysis of clustering for general statistical problems (18). GSEA is an algorithm that evaluates if a number of genes sharing biological traits are expressed with statistically significant differences. This technique verifies related genes in a given set are distributed more toward the top or bottom of a ranked list of differential expression results between two biologic states than expected by chance (19). The enrichment score of GSEA is determined by repeatedly permuting the observed expression-phenotype relationships. The leading edge of an enriched gene set represents the core biology of interest. Gene co-expression networks are used to associate genes of unknown function with biological processes, to prioritize candidate disease genes, or to discern transcriptional regulatory programs (20). Although they do not provide information concerning causality, new methods for differential co-expression analysis can provide clues to identify regulatory genes underlying various phenotypes. WGCNA is the first co-expression tool and the most widely used clustering package for co-expression analysis (20, 21). This technique performs hierarchical clustering and divides each cluster into sub-clusters representing co-expression modules. The hub gene is a highly connected gene within each co-expression network, and identification of this gene could provide novel functional insight into specific disease. NetRep is a rapid and computationally efficient method that can be used to verify if specific co-expression modules are preserved in other gene expression datasets without assuming data are

normally distributed (22).

Gene expression data obtained from specific tissue are advantageous in terms of exploring the mechanisms underlying specific diseases due to the tissue-specificity of these data. Gene expression data of asthma can be obtained from various specimens such as induced sputum, bronchial epithelial cells, bronchoalveolar lavage fluid, and biopsy specimens. Among these, induced sputum is a reliable non-invasive specimen of bronchial inflammation in asthma. It is recognized as a very useful specimen for both research and clinical applications, and induced sputum aids in both the diagnosis and monitoring of asthma (23). The evaluation of gene expression profiles of sputum cells has been applied successfully to understand asthma pathogenesis (24, 25).

The present study examined biological pathways related to EA using genome-wide gene expression profiles of sputum cells from EA patients. We selected genes exhibiting significantly different expression profiles between EA and healthy elderly controls, and we identified two distinct clusters by hierarchical clustering of these genes. We performed gene set enrichment analysis (GSEA) and weighted gene co-expression network analysis (WGCNA) to gain further biological insights into each cluster. Finally, we confirmed that our findings were replicated in a dependent gene expression profile of sputum cells obtained from Gene Expression Omnibus (GEO), a publicly available database of gene expression profiles. To the best of our knowledge, this is the first study to identify specific biological pathways contributing to EA pathogenesis using sputum gene expression profiles.

Methods

This study was approved by the Seoul National University Hospital Review Board (1608-101-786), and informed consent was obtained from all study participants.

1. Discovery data set

Asthma was diagnosed according to the Global Initiative for Asthma guidelines. Elderly patients with asthma, who were older than 65 years and exhibited symptoms and signs suggesting chronic airway diseases and reversible airflow limitation or airway hyper-responsiveness were recruited from the Seoul National University Hospital, Seoul, Korea. Chronic signs and symptoms were defined based on current (past 12 months) episodic respiratory symptoms including dyspnea, coughing, wheezing, and sputum. Reversible airflow limitation was defined as the ratio of forced expiratory volume in the first second (FEV1) to forced vital capacity (FVC) being below 0.7 with FEV1 increased more than 12% and 150ml after bronchodilator use, or FEV1 being increased more than 15% spontaneously or after treatment. Airway hyper-responsiveness was assessed by the methacholine provocation test when the provocative concentration 20% (PC20) was below 16mg / ml. (26). As acute exacerbation by various causes is also likely to alter gene expression in induced sputum, patients with severe comorbidities who were unable to perform the study or who could not obtain induction sputum in a stable state were excluded.

Healthy elderly controls without asthma (n = 10) were recruited by advertisement. None of the healthy controls exhibited respiratory symptoms, sputum eosinophilia, abnormality on chest radiography, or obstructive pulmonary function test results. Three of the control patients were smokers (one current smoker and two ex-smokers). Baseline assessment at recruitment included sex, age, height, weight, smoking status, and previous medical history. Exclusion criteria included recent (past month) respiratory tract infection, recent change in maintenance therapy, and recent asthma exacerbation; however, to reflect real-life situations, smoking status was not considered as an exclusion criterion. Sputum induction and processing were performed as previously described (27).

2. Gene expression arrays

We collected induced sputum prior to drug administration when patient disease condition was stable to minimize drug effect. Collection of induced sputum was performed in accordance with a previously described method (28). Cells were stored at -80°C until use. RNA was extracted from induced sputum samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany). Gene expression levels were measured using the GeneChip Human Gene 2.0 ST (Affymetrix, Santa Clara, CA, US). We removed probes exhibiting poor chromosome annotation and probes within the X or Y chromosome. We then performed VST transformation and quantile normalization, respectively, to reduce the effects of technical noise and to ensure that the distribution of

expression levels for each array was closer to a normal distribution.

3. Statistical analysis

Figure 1 shows an overview of the statistical analysis. A total of 3,156 gene probes showing significantly ($P < 0.05$) different expressions between EA and healthy elderly controls were used for further analysis. To search for meaningful information patterns and dependencies in gene expression data, we performed hierarchical clustering using the Pvclust package in R version 3.4.3 (www.r-project.org). P-values indicate how strongly clusters are supported by the data (18). An approximate and unbiased P-value greater than 95% was used to define a cluster. Comparison of the phenotype between two clusters was also performed. Pearson's chi square test and Fisher's exact test were to determine categorical variable, and *t*-test and Mann-Whitney U test were used for the continuous variable.

We next performed GSEA using the GSEA software (version 3.0) provided by the Broad Institute (Boston, MA, USA) to compare the leading edge subsets of each cluster as previously described (19). We used the hallmark gene sets (H collection) obtained from the Molecular Signatures Database (MSigDB, version 6.0), and we defined significantly enriched gene sets as those with a false discovery rate threshold of less than 0.05. We also performed WGCNA on gene expression datasets of two clusters (21). Using 3156 gene probes, we characterized unsigned correlation networks and determined their relationship to each other. We also found gene modules within each cluster. Modules were

defined as groups of highly interconnected genes (29). We computed eigengene values for the modules identified, and we performed linear regression analysis to search phenotype-associated gene modules that were significantly related to the target phenotype (29). GSEA based on Reactome pathway database was also done with individual genes of each gene modules we found.

4. Replication data set

A dependent gene expression profile of sputum cells (GSE41863) obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41863) was used to replicate our results. The dataset contains gene expression data from 118 patients diagnosed with moderate-to-severe asthma and 21 control subjects. To identify markers associated with various asthma subtypes, sputum samples were collected from asthmatics and healthy controls and subjected to expression profiling using Affymetrix HG-U133Plus2.0. From this profile, we selected 20 subjects aged 65 years or older (15 asthmatics and 5 healthy controls).

We also selected 3,264 gene probes ($P < 0.05$) that are differently expressed form normal elderly subjects and performed cluster analysis. From the gene probes, two clusters were identified. Using these gene probes, GSEA based on hallmark gene sets of MSigDB was also performed in the same manner. We compared GSEA results obtained from the discovery and replication datasets, and we searched pathways and individual genes that may participate in the pathogenesis of EA. We also tested if gene modules from our two clusters were

present in the public database. R package “NetRep” was used for this analysis, and this approach can quantify the preservation of gene co-expression modules across different datasets and can produce unbiased p-values based on a permutation approach to score module preservation without assuming normal distribution of data (22).

Results

1. Baseline characteristics

Fifty-five patients with EA were enrolled and their phenotypic data and gene expression profile in induced sputum were collected. Based on the differential gene expression patterns of sputum cells, two distinct clusters were identified excluding three outliers. Cluster 1 consisted of 35 patients and cluster 2 consisted of 20 patients. (Figure 2 and Figure 3). Cluster 1 featured a significantly lower proportion of eosinophils in the sputum and less severe airway obstruction as measured by the post-bronchodilator (BD) ratio of the forced expiratory volume in one second and forced vital capacity (FEV1/FVC) compared to cluster 2. Although it was not statistically significant, patients in cluster 2 were older and have longer duration of diseases than patients in cluster 1. Detailed characteristics of the two clusters are provided in Table 1.

2. Gene set enrichment analysis (GSEA)

We performed a GSEA of our gene probes in each cluster to identify the interactions of functionally related biologic processes. With hallmark gene sets in MSigDB, we found several significant pathways (Table 2 and Figure 4). Five gene sets were significantly enriched in cluster 1 [OXIDATIVE_PHOSPHORYLATION (OXPHOS), UNFOLDED_PROTEIN_RESPONSE (UPR), MYC_TARGETS_V1, DNA_REPAIR, and ADIPOGENESIS] and three gene sets were significantly

enriched in cluster 2 [EPITHELIAL_MESENCHYMAL_TRANSITION (EMT), MYOGENESIS, and KRAS_SIGNALING_DN]. This result shows biologic support that gene expression profile of our two cluster is statistically different and clues that pathologic process of asthma phenotype might be different.

3. Subgroup analysis with patients who have never smoked

We also analyzed with patients who have never smoked. Based on the differential gene expression patterns of sputum cells, a total of 29 patents was separated with two clusters (nineteen patients in cluster 1 and ten patients in cluster 2) (Figure 5). We performed GSEA with the result and we found two significant pathway in each of the two clusters. UNFOLDED_PROTEIN_RESPONSE (UPR) and OXIDATIVE_PHOSPHORYLATION (OXPHOS) gene sets were enriched in cluster 1 and MYOGENESIS and EPITHELIAL_MESENCHYMAL_TRANSITION (EMT) gene sets were enriched in cluster 2 (Figure 6). With analysis only with never smokers, two gene sets in cluster 1 showed statistical significance but they did not after correction of multiple comparisons (Table 3)

4. Validation of gene sets from GSEA analysis

In the replication analysis, hierarchical clustering of gene expression patterns also revealed two clusters (cluster 1` with 8 EA patients and cluster 2` with 7 EA patients, Figure 7). Clinical characteristics of each cluster are shown in Table 4. GSEA identified eight gene sets (OXPHOS, ESTROGEN_RESPONSE_EARLY, E2F_TARGETS, MYC_TARGETS_V1, FATTY_ACID_METABOLISM, MYC_TARGETS_V2, ESTROGEN_RESPONSE_LATE, and DNA_REPAIR) were significantly enriched in cluster 1 and three gene sets (TNFA_SIGNALING_VIA_NFKB, INFLAMMATORY_RESPONSE, and EMT) in cluster 2` (Table 5 and Figure 8). The OXPHOS gene set was significantly enriched in both cluster 1 and cluster 1` and the EMT gene set in both cluster 2 and cluster 2`.

5. Relationship between phenotype and replicated gene sets

To assess the potential functional relevance of the replicated gene sets, the leading edge genes of gene sets enriched in both discovery and replication dataset (Table 6) were summarized into a single metagene metric using a principal component analysis. Then, the correlations between the first principal component (PC1) and serum uric acid levels and post-BD FEV1/FVC values were measured in each cluster. Circulating uric acid is a major antioxidant that might help protect against oxidative stress (30). The post-BD FEV1/FVC value is an indirect indicator of airway remodeling (31). Levels of serum uric acid were lower in cluster 1 with a borderline significance ($P = 0.083$) compared to

cluster 2, whereas post-BD FEV1/FVC values were significantly higher in cluster 1 (P = 0.008) (Figure 9). PC1 of leading edge genes in the OXPHOS gene set showed a negative correlation with serum uric acid levels (P = 0.075) only in cluster 1 (Figure 10). Meanwhile, PC1 of leading edge genes in the EMT gene set showed a significantly negative correlation with post-BD FEV1/FVC values (P = 0.005) only in cluster 2 (Figure 10). For individual gene, we found 4 genes (MRPS11, HSPA9, NDUFB4, and ACAT1) in the leading edge genes of the OXPHOS gene set and 2 genes (SNTB1 and FUCA1) in leading edge genes of the EMT gene set belonged to genes which showed more than 1.5 log2-fold expression difference between cluster 1 and cluster 2 with P values less than 0.01 (Figure 11).

6. Weighted gene co-expression network analysis (WGCNA) and modules correlates with phenotype of cluster

Applying WGCNA to the 3,156 genes of two clusters, we identified four modules in cluster 1 and 18 modules in cluster 2. (Figure 12) The size of modules were various ranging from 62 (lightgreen module in cluster 2) to 1728 genes (grey module in cluster 1). Grey module is group of genes which could not be assigned to a module and they were not considered for further analysis. Among modules identified, FEV1/FVC ratio correlates significantly with brown module of cluster 1 (P value = 0.003632) and magenta module of cluster 2. (P value = 0.03697). After multiple regression with other phenotypic variable

including BMI, symptom duration, FVC, FEV1 and cellular profile of induced sputum, the association between FEV1/FVC and brown module of cluster 1 (P=0.0347) and magenta module of cluster 2 (P value=0.003117) remained statistically significant (Figure 13). GSEA based on Reactome database with individual genes of two modules identified two gene sets which were significantly enriched (CELL_CYCLE gene set in brown module of cluster 1 and SIGNALING_BY_GPCR gene set in magenta module of cluster 2) (Figure 14).

7. Validation of gene sets from WGCNA and enriched pathways

Among four modules identified in cluster 1 of discovery dataset, two modules (module 1 and 3, P value = 0.00319968 and 0.00209979, respectively) were significantly preserved in cluster 2` of replication dataset with NetRep analysis (P value = 0.00319968 and 0.00209979, respectively) (Figure 15). Also, four modules in cluster 2 of discovery dataset were preserved in cluster 1` (Module 1, 6, 9, and 13, P value = 0.00039996, 0.00009999, 0.00079992 and 0.02329767, respectively). Brown module of cluster 1 and magenta module of cluster in cluster 2 also preserved in the replication cohort. Thus, we denoted them as the significant eigengene sets that shows phenotypic difference in elderly asthma. Table 7 shows genes consists brown module of cluster 1 and magenta module in cluster 2.

Table 1. Baseline characteristics (discovery dataset)

Characteristics	**Cluster 1 (n=35)**	**Cluster 2 (n=20)**	**P value**
Male, N (%)	16 (45.7)	9 (45)	0.959
Age, year	72.6 (5.4)	75.1 (5.4)	0.103
BMI, kg/m2	24.9 (3.7)	26.4 (4.1)	0.205
Smoking status			
Current smoker, N (%)	4 (11.4)	2 (10)	0.841
Ex-smoker, N (%)	12 (34.3)	8 (40)	
Never-smoker, N (%)	19 (54.3)	10 (50)	
Age of symptom onset, year	67.6 (6.8)	69.1 (6.9)	0.430
Age of symptom onset < 65 year, N (%)	11 (31.4)	6 (30)	0.912
Symptom duration, year	4.9 (6.0)	5.9 (4.7)	0.524
Chronic rhinosinusitis, N (%)	18 (51.7)	8 (40)	0.514
Atopy, N (%)	4 (44.4)	2 (22.2%)	0.620

Sputum eosinophil, %	2.8 (4.3)	5.9 (8.2)	0.033
Sputum neutrophil, %	35.9 (14.3)	37.9 (14.8)	0.580
Serum uric acid, mg/L	5.2 (1.2)	6.1 (2.3)	0.083
Peripheral eosinophil count	209.6(192.3)	197.2(193.7)	0.819
Serum total IgE	91.0(143.1)	130.0(226.7)	0.616
Post-BD FVC, L	2.75 (0.80)	2.74 (0.93)	0.974
Post-BD FVC predicted, %	96.4 (16.0)	104.8 (15.8)	0.068
Post-BD FEV1, L	1.80 (0.44)	1.51 (0.37)	0.011
Post-BD FEV1 predicted, %	91.0 (18.7)	82.0 (15.9)	0.065
Post-BD FEV1/FVC, %	66.3(11.8)	56.3 (13.0)	0.007

Data were presented as mean (standard deviation) except those indicated otherwise

BD, Bronchodilator; BMI, Body Mass Index, FEV1, Forced expiratory volume in 1 second; FVC, Forced vital capacity; n, Number

Table 2. Gene sets enriched significantly in discovery dataset

Gene set name	Number of overlapped genes	Normalized enrichment score	FDR P-value
Cluster 1			
OXIDATIVE_PHOSPHORYLATION	28	2.99	<0.001
UNFOLDED_PROTEIN_RESPONSE	16	2.55	<0.001
MYC_TARGETS_V1	22	1.83	0.027
DNA_REPAIR	15	1.81	0.025
ADIPOGENESIS	31	1.80	0.021
Cluster 2			
EPITHELIAL_MESENCHYMAL_TRANSITION	31	-3.29	<0.001
MYOGENESIS	27	-2.67	<0.001
KRAS_SIGNALING_DN	29	-1.81	0.024

Table 3. Gene sets enriched in analysis with never smokers of discovery dataset

Gene set name	Number of overlapped genes	Normalized enrichment score	Nominal P-value	FDR P-value
Cluster 1				
UNFOLDED_PROTEIN_RESPONSE	16	1.3913496	< 0.001	0.125
OXIDATIVE_PHOSPHORYLATION	28	1.3671378	0.012	0.324
MYC_TARGETS_V1	22	1.3158339	0.156	0.671
Cluster 2				
EPITHELIAL_MESENCHYMAL_TRANSITION	31	-1.3612622	0.083	0.573
MYOGENESIS	27	-1.3715386	0.085	1
HYPOXIA	31	-1.3222681	0.107	0.573

Table 4. Clinical characteristics of replication data set

Patient	GSE41863 ID	Age	Gender	Sputum inflammatory profile
Cluster 1'				
Pt7	GSM1026689	66	Male	Eosinophilic
Pt8	GSM1026702	68	Female	Neutrophilic
Pt9	GSM1026712	71	Male	Neutrophilic
Pt10	GSM1026714	77	Male	Eosinophilic
Pt12	GSM1026716	65	Male	Neutrophilic
Pt13	GSM1026719	67	Male	Eosinophilic
Pt14	GSM1026723	78	Female	Pauci-granulocytic
Pt15	GSM1026740	72	Male	Pauci-granulocytic
Cluster 2'				
Pt1	GSM1026697	76	Male	Neutrophilic

Pt2	GSM1026701	78	Male	Neutrophilic
Pt3	GSM1026703	82	Female	Neutrophilic
Pt4	GSM1026711	68	Female	Mixed
Pt5	GSM1026727	68	Male	Neutrophilic
Pt6	GSM1026730	67	Male	Neutrophilic
Pt11	GSM1026735	76	Male	Mixed

Eosinophilic: sputum eosinophil proportion > 3%; Neutrophilic: sputum neutrophil proportion > 61%; Mixed: both increased neutrophils and eosinophils; Pauci-granulocytic: normal levels of both neutrophils and eosinophils

Table 5. Gene sets enriched significantly in replication dataset

Gene set name	Number of overlapped genes	Normalized enrichment score	FDR P-value
Cluster 1`			
OXIDATIVE_PHOSPHORYLATION	85	1.96	0.005
ESTROGEN_RESPONSE_EARLY	24	1.89	0.005
E2F_TARGETS	37	1.75	0.019
MYC_TARGETS_V1	67	1.65	0.048
FATTY_ACID_METABOLISM	38	1.64	0.041
MYC_TARGETS_V2	28	1.64	0.034
ESTROGEN_RESPONSE_LATE	26	1.64	0.031
DNA_REPAIR	39	1.61	0.038
Cluster 2`			
TNFA_SIGNALING_VIA_NFKB	27	-2.59	<0.001
INFLAMMATORY_RESPONSE	36	-2.05	0.002

EPITHELIAL_MESENCHYMAL_TRANSITION	28	-1.74	0.019

Table 6. Leading edge genes of gene sets enriched in both discovery and replication datasets

OXIDATIVE_PHOSPHORYLATION gene set				EPITHELIAL_MESENCHYMAL_TRANSITION gene set			
Gene	**Rank metric score**	**Running ES**	**Core enrichment**	**Gene**	**Rank metric score**	**Running ES**	**Core enrichment**
NDUFB4	1	0.376	0.078	EMP3	2577	-0.099	-0.743
MRPS11	3	0.361	0.152	CDH11	2591	-0.100	-0.723
ACAT1	6	0.290	0.212	FAP	2605	-0.101	-0.702
HSPA9	7	0.286	0.271	SDC1	2617	-0.101	-0.680
SDHC	8	0.278	0.329	CRLF1	2649	-0.105	-0.664
OXA1L	24	0.244	0.374	LOXL2	2650	-0.105	-0.638
GRPEL1	30	0.238	0.422	COMP	2691	-0.109	-0.624
GPX4	77	0.201	0.449	TNC	2712	-0.110	-0.603
VDAC2	106	0.188	0.479	CAP2	2761	-0.116	-0.589
ATP6V0C	115	0.185	0.514	FOXC2	2799	-0.120	-0.572
ETFA	123	0.180	0.549	COPA	2812	-0.122	-0.545
UQCRC2	134	0.177	0.583	FBN1	2837	-0.126	-0.522

FXN	167	0.167	0.607	DPYSL3	2851	-0.128	-0.494
BAX	169	0.167	0.641	SFRP4	2871	-0.131	-0.467
HADHB	178	0.165	0.673	WIPF1	2900	-0.137	-0.443
ALDH6A1	320	0.138	0.656	COL5A1	2923	-0.141	-0.415
GLUD1	352	0.134	0.673	MMP2	2964	-0.151	-0.390
COX10	393	0.128	0.687	GREM1	2978	-0.155	-0.356
AFG3L2	396	0.128	0.713	NOTCH2	3015	-0.172	-0.324
NDUFA4	406	0.126	0.736	VIM	3028	-0.179	-0.284
COX8A	414	0.125	0.760	PFN2	3038	-0.186	-0.240
NDUFA5	416	0.124	0.785	DAB2	3055	-0.199	-0.196
				PRRX1	3058	-0.203	-0.146
				MGP	3061	-0.205	-0.095
				SNTB1	3078	-0.219	-0.046
				FUCA1	3079	-0.220	0.009

Table 7. Genes consists (A) brown module of cluster 1 and (B) magenta module in cluster 2.

(A)

kc1 module 3 gene			
DUOXA1	EYA2	RIMS1	MUC5B
TMC5	CLIC6	GSTA1	PTPRF
CX3CL1	RIBC2	UBXN10	COLCA2
CDH3	CHL1	PERP	LAYN
CDH1	UPK1B	CLDN4	TSPAN1
ATP2C2	EFHB	STEAP2	CYP4B1
FA2H	PLXNB1	CFTR	MUC15
ST6GALNAC1	LRIG1	PTPRZ1	DCDC1
DSG2	MUC13	AGR2	CACHD1
SERPINB5	CP	COBL	KRT18
SERPINB11	PFN2	IQUB	C12orf74
CAPN9	PLCH1	GRHL2	KRT5
CHST9	MECOM	EYA1	LRIG3
SERPINB4	C1orf194	DNAI1	ATP12A
CYP2F1	CCDC39	PIGR	SOX21-AS1
MED25	SLC34A2	ASS1	NEK5
CXCL17	SPARCL1	NFIB	SLITRK6
DRC1	ADH1C	IRF6	DZIP1
EPCAM	ADH7	TNC	AK7
DAW1	MARVELD2	ENAH	SIX1
EFEMP1	SELENBP1	PLEKHS1	
WFDC2	CAP2	ENKUR	

(B)

kc2 module 9 gene			
CHTOP	HOXD9	PRICKLE4	KLF6
ARID3B	PRDM2	GATAD2B	DNAJB12
MAN2A2	SLC11A1	C6orf106	LINC00294
INO80	TMEM198	USP42	ATG13
PARP6	TLDC2	CREB5	SLC22A9
DDR2	SLC23A2	KCTD7	RBM14
TCF25	EWSR1	BCL7B	CARS
CES5A	HMGXB4	GIGYF1	ZNF195
SMCR8	EP300	TNFSF18	CPSF7
ZNF830	LOC388882	MKRN1	LOC100133315
NPL	SF3A1	ZNF746	LOC100049716
RPS6KB1	DUSP18	PHYHIP	PFDN5
AA06	NR2C2	FAM214B	SP1
MIR205HG	LOC100129550	KLF9	PCBP2
LINC00511	CC2D1B	FAM120AOS	PSMD9
TGIF1	LPP	ZFP37	LOC414300
ARF1	DCP1A	SH2D3C	PRPF3
GATAD2A	NPY2R	WDR37	SMAD9
WTIP	PCBD2	DHTKD1	RCBTB2
MED25	ANKHD1	LOC439994	IRS2
DDX39A	CXCL14	WBP1L	TOX4
ZNF329	PDLIM7	TAF5	MAPK1IP1L
LINC00486	HMGN4	MXI1	RGS6

Figure 1. Overview of analysis

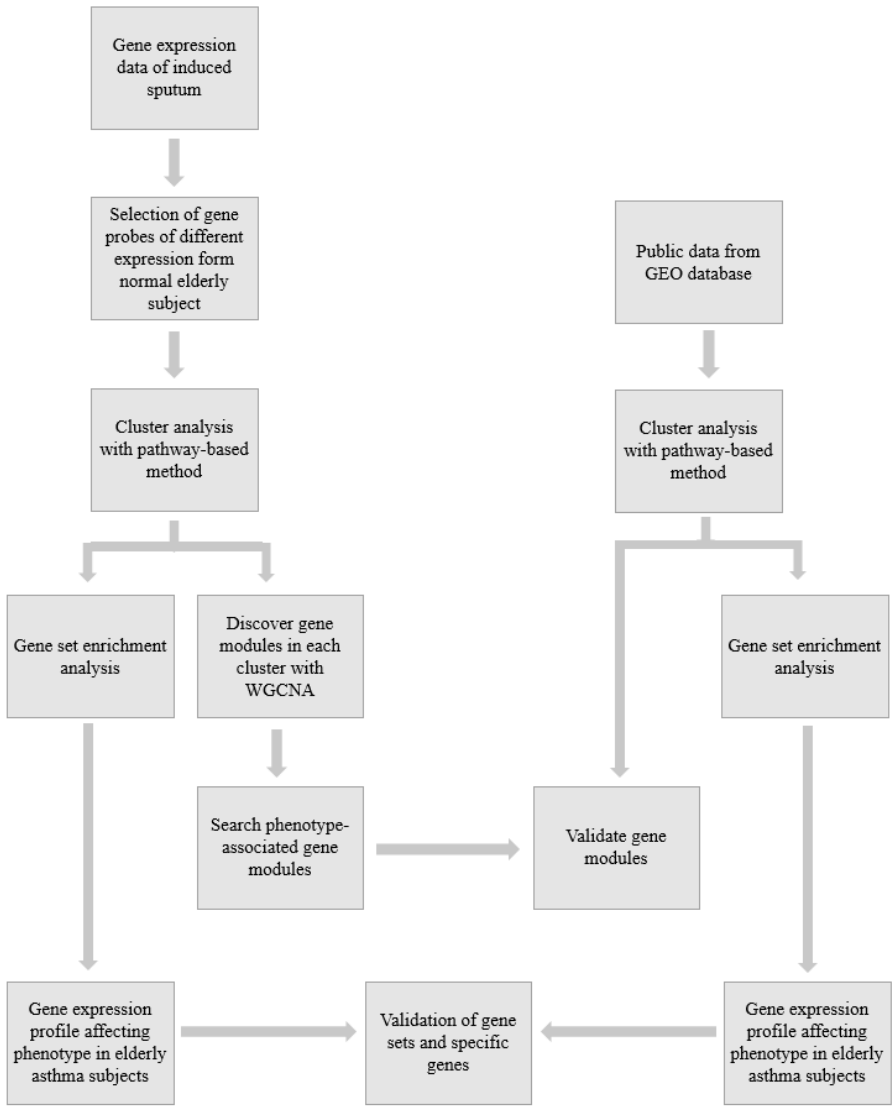


Figure 2. Two gene clusters identified in the discovery dataset. Three outliers (Pt5, Pt21, and Pt28) were excluded from analysis.

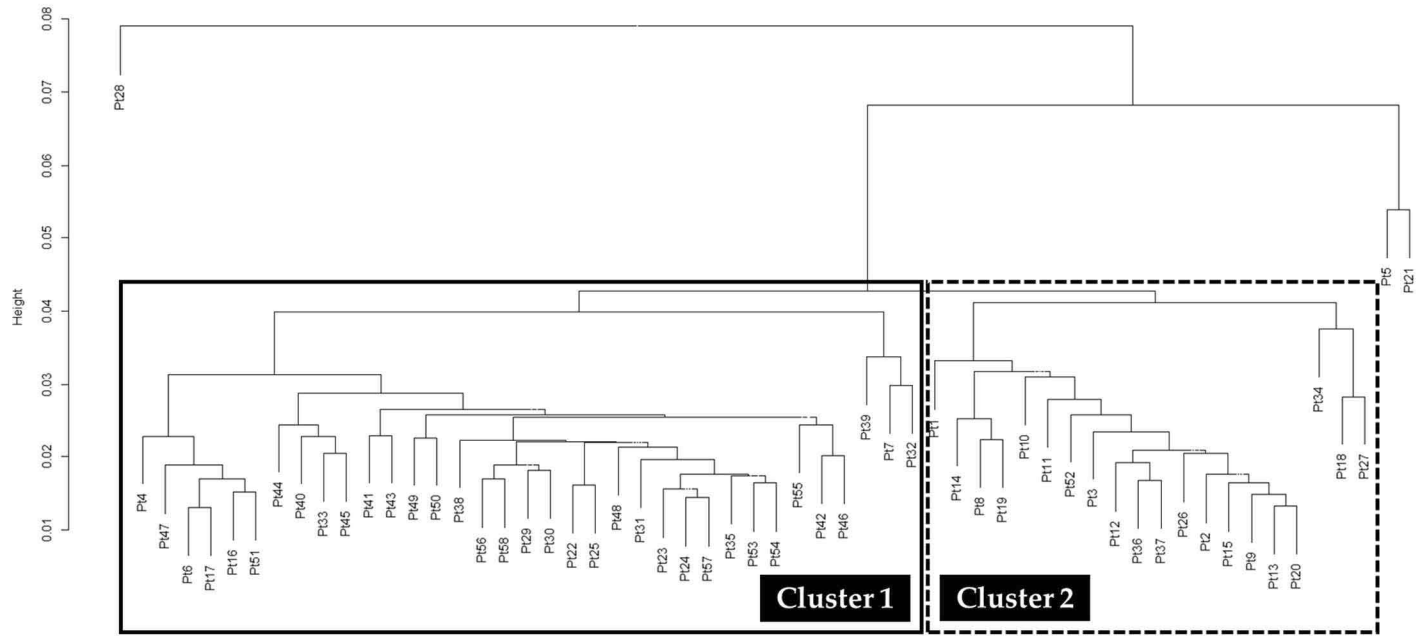


Figure 3. Heat map of two gene clusters identified in the discovery dataset

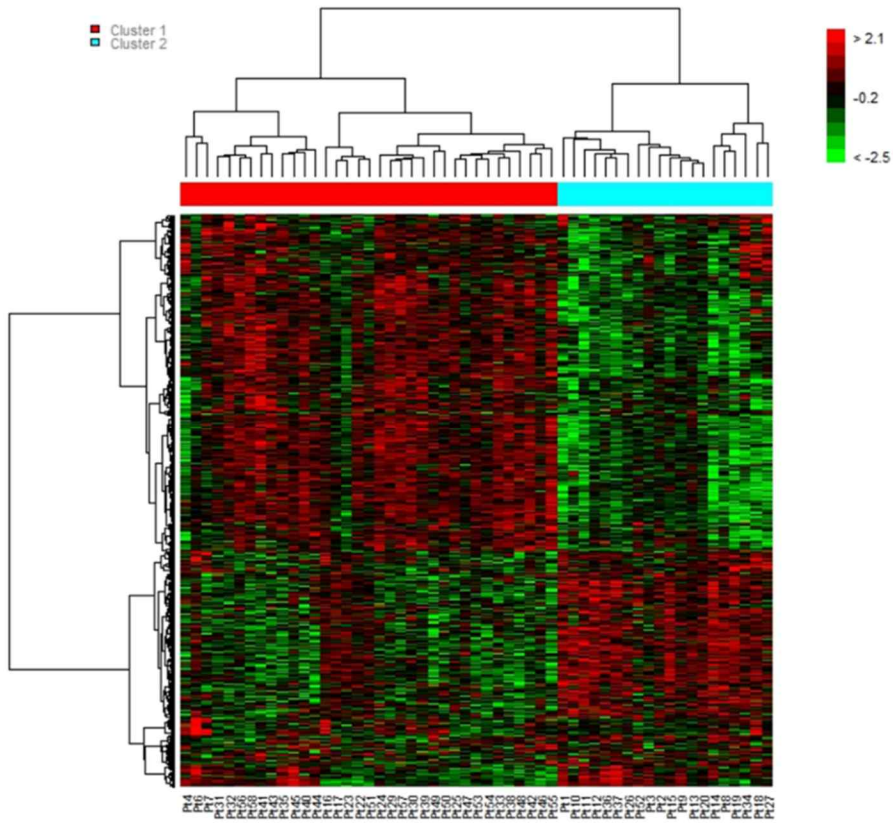
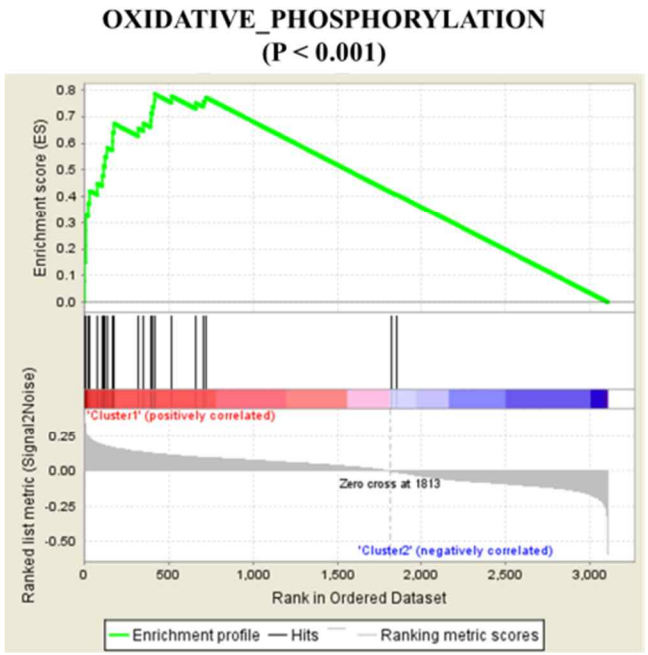
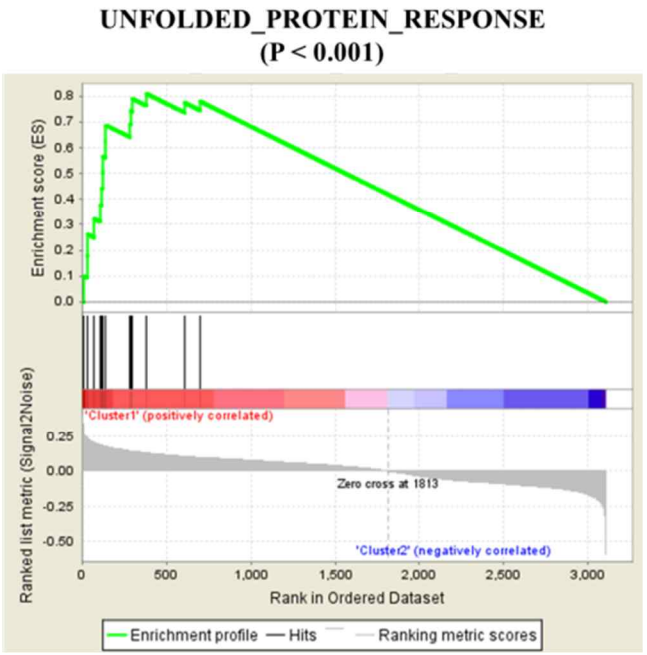


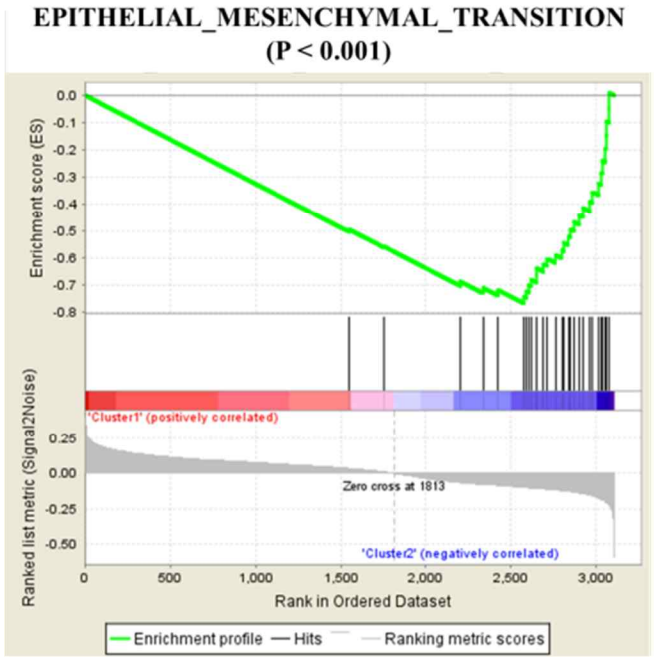
Figure 4. Gene sets enriched in each cluster identified in the discovery dataset with FDR P-values less than 0.001

A. Cluster 1





B. Cluster 2


EPITHELIAL_MESENCHYMAL_TRANSITION
(P < 0.001)
Enrichment score (ES)
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
'Cluster1' (positively correlated)
Ranked list metric (Signal2Noise)
0.25
0.00
-0.25
-0.50
Zero cross at 1813
'Cluster2' (negatively correlated)
0
500
1,000
1,500
2,000
2,500
3,000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

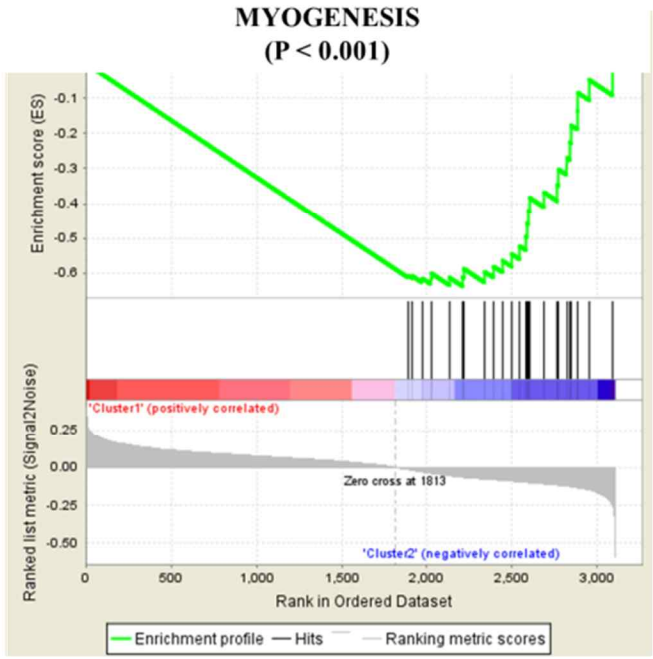

MYOGENESIS
(P < 0.001)
Enrichment score (ES)
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
'Cluster1' (positively correlated)
Ranked list metric (Signal2Noise)
0.25
0.00
-0.25
-0.50
Zero cross at 1813
'Cluster2' (negatively correlated)
0
500
1,000
1,500
2,000
2,500
3,000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

Figure 5. Two gene clusters with patients who have never smoked in the discovery cohort

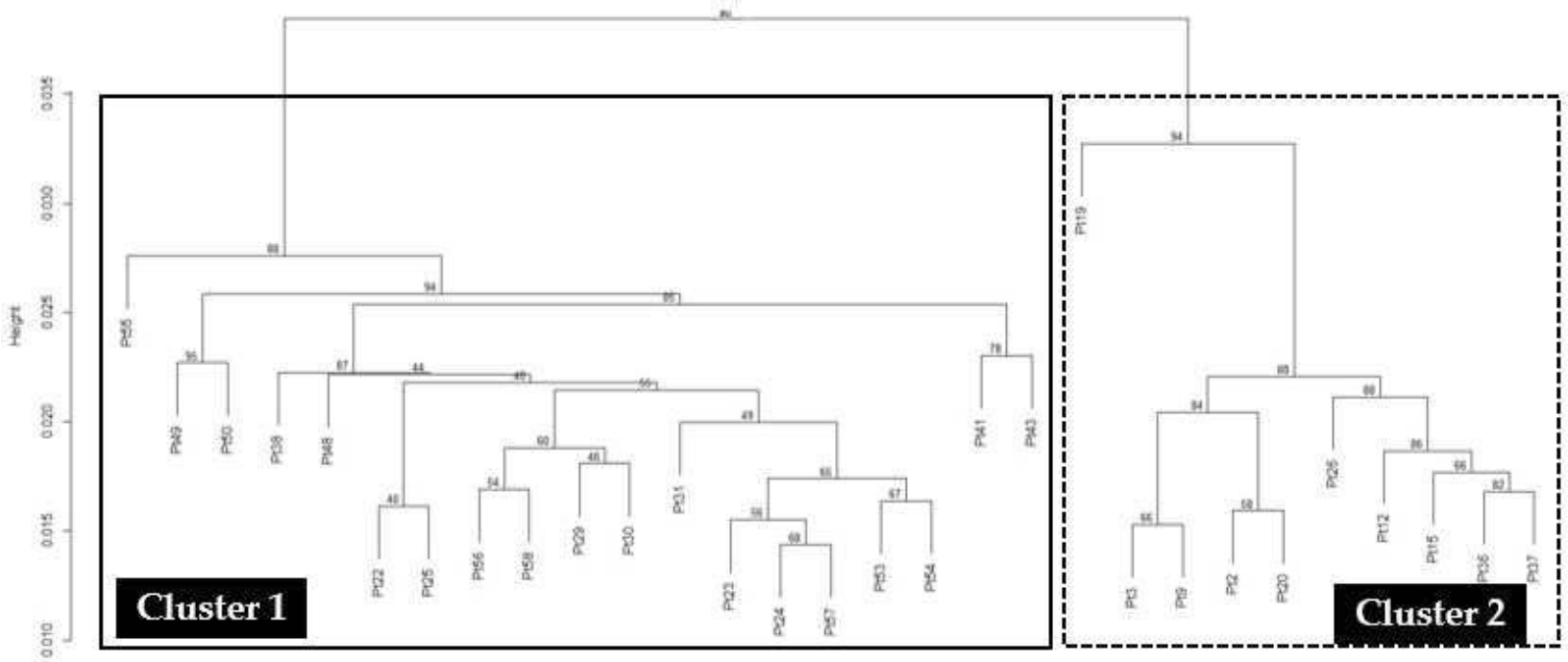
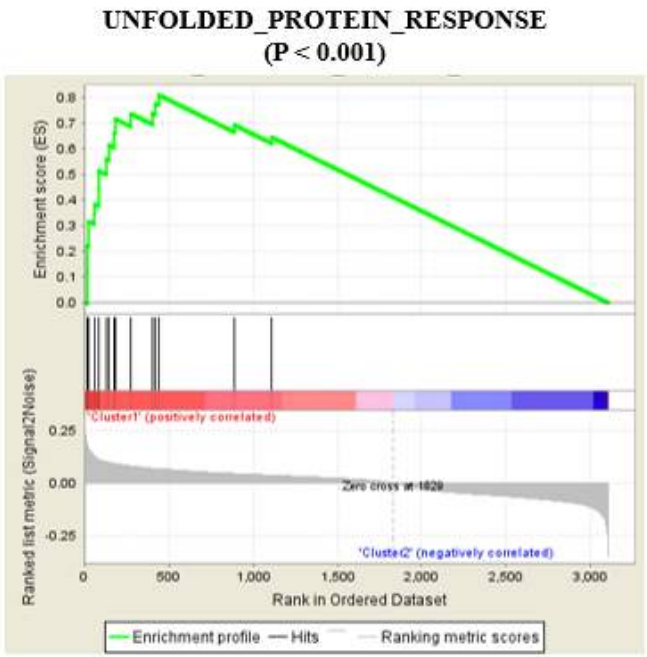
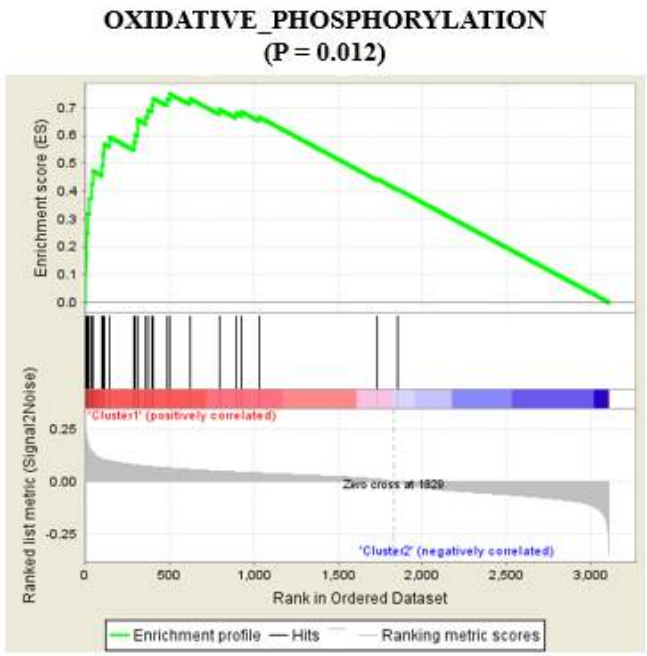


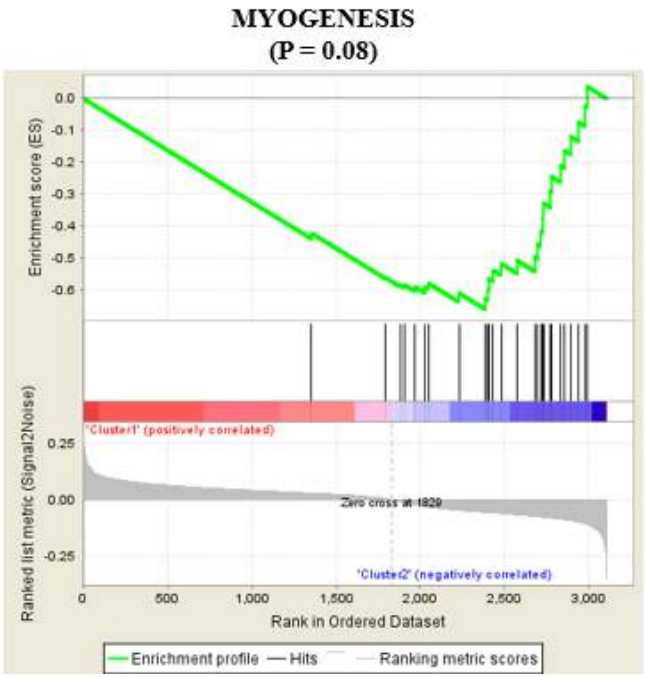
Figure 6. Gene sets enriched in patients who have never smoked in discovery dataset

A. Cluster 1





B. Cluster 2


MYOGENESIS
(P = 0.08)
Enrichment score (ES)
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
'Cluster1' (positively correlated)
Ranked list metric (Signal2Noise)
0.25
0.00
-0.25
Zero cross at 1829
'Cluster2' (negatively correlated)
0
500
1,000
1,500
2,000
2,500
3,000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

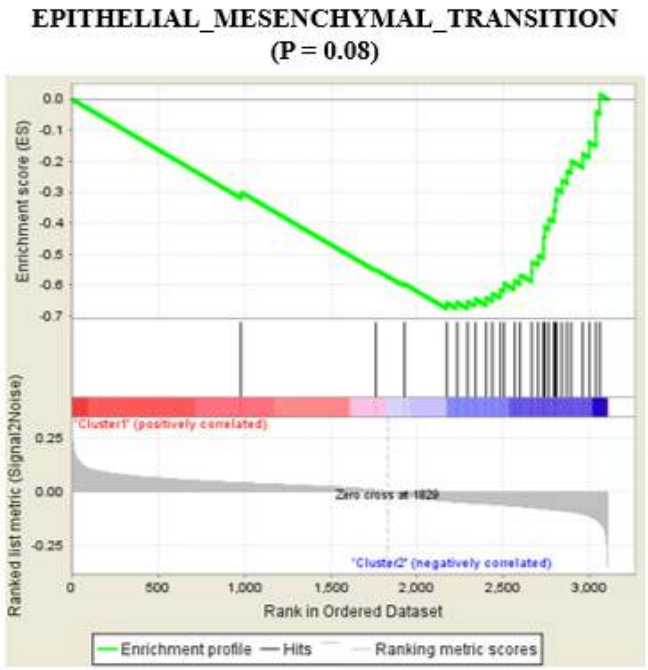

EPITHELIAL_MESENCHYMAL_TRANSITION
(P = 0.08)
Enrichment score (ES)
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
'Cluster1' (positively correlated)
Ranked list metric (Signal2Noise)
0.25
0.00
-0.25
Zero cross at 1829
'Cluster2' (negatively correlated)
0
500
1,000
1,500
2,000
2,500
3,000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

Figure 7. Two gene clusters identified in the replication dataset

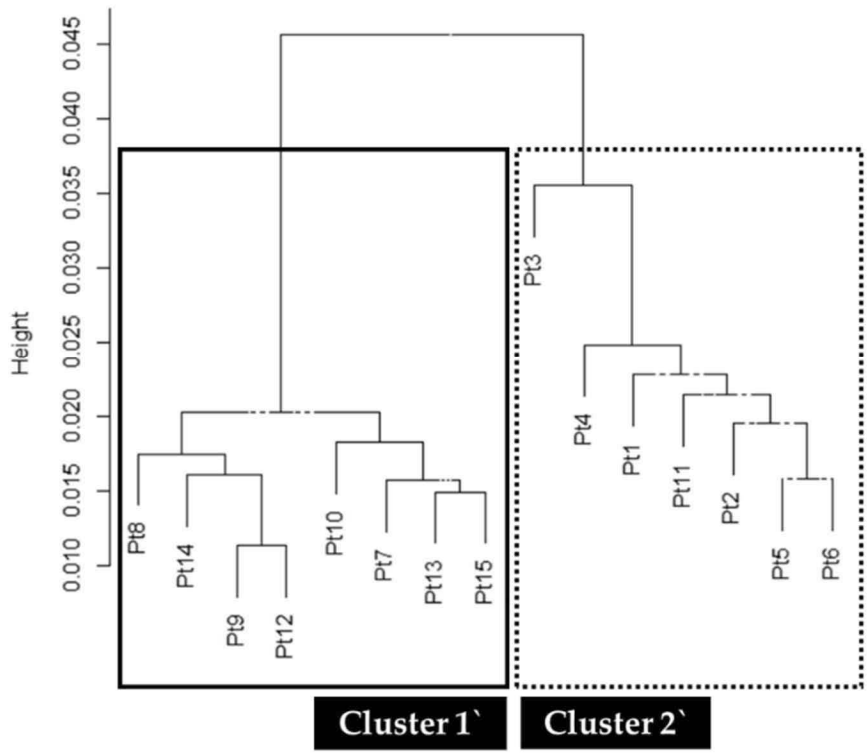


Figure 8. Gene sets enriched in each cluster identified in the replication dataset with FDR P-values less than 0.001. Only two gene sets were presented which were also significantly enriched in the discovery dataset.

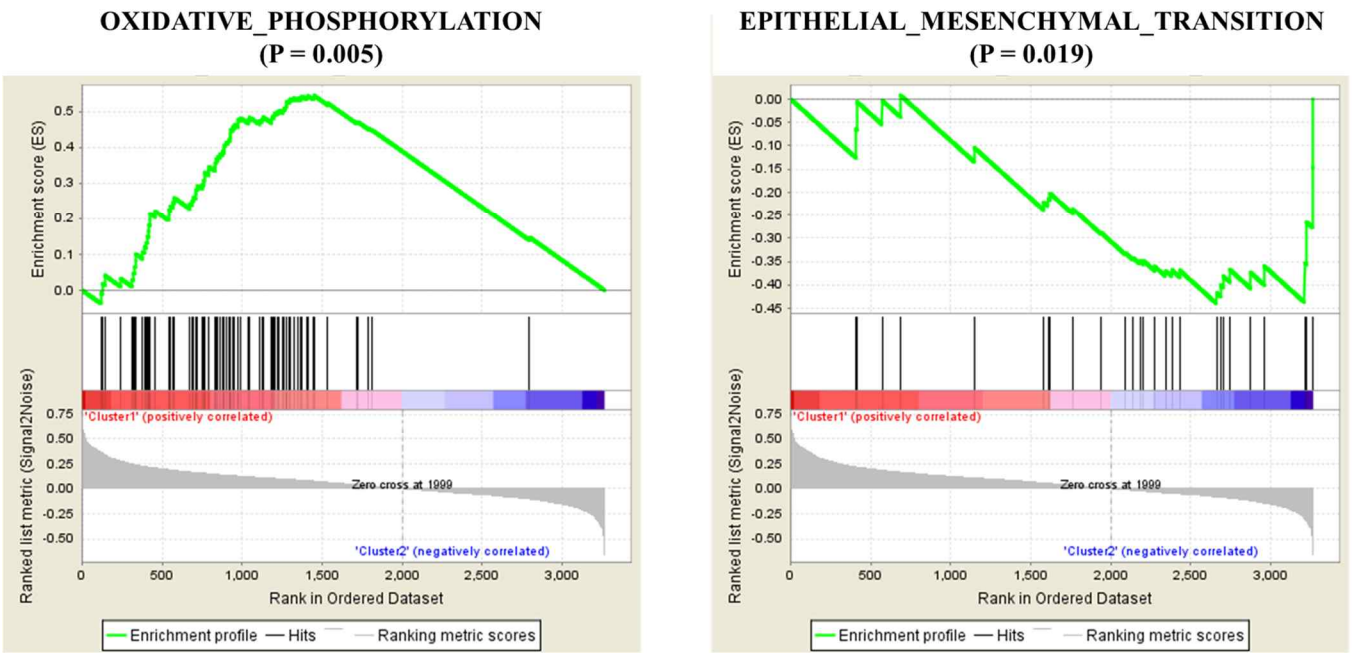


Figure 9. Differences of serum uric acid levels and post-BD FEV1/FVC values between cluster 1 and cluster 2 in the discovery dataset.

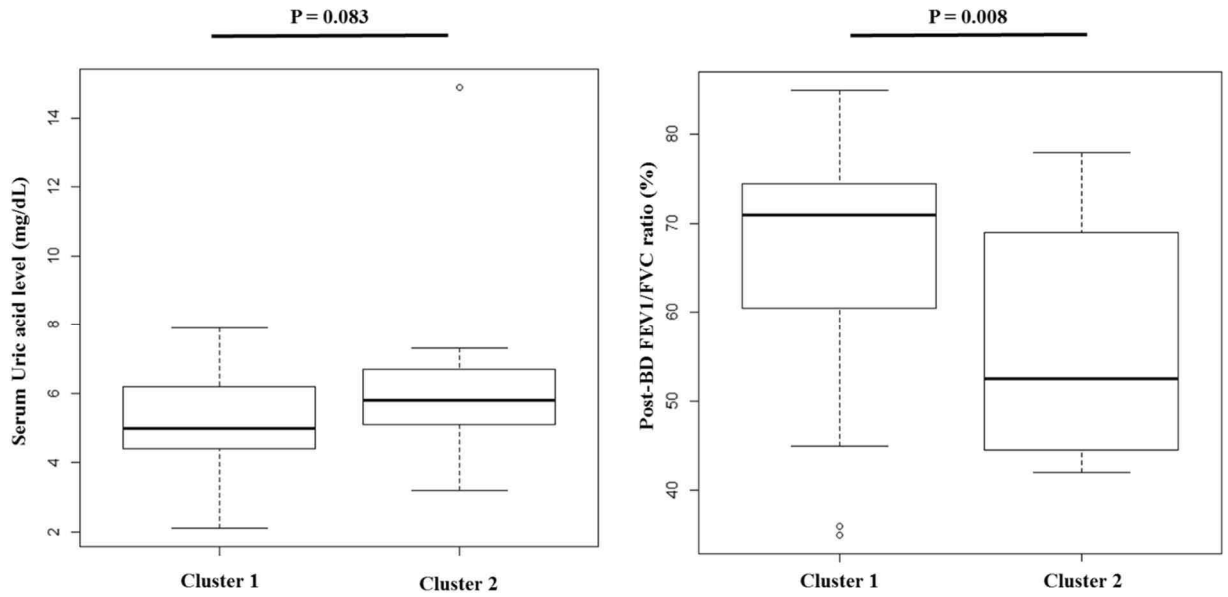
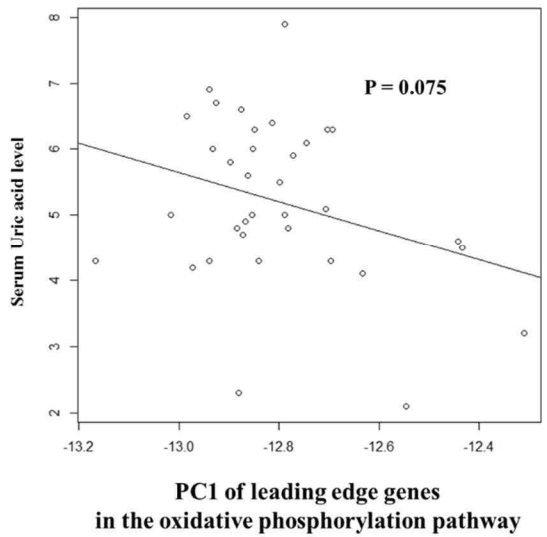
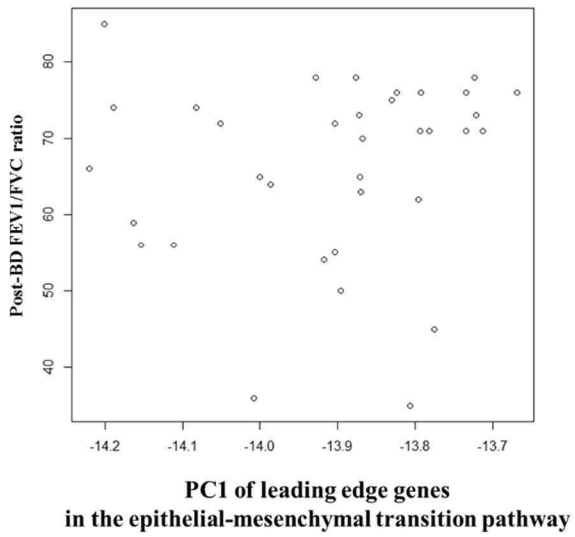


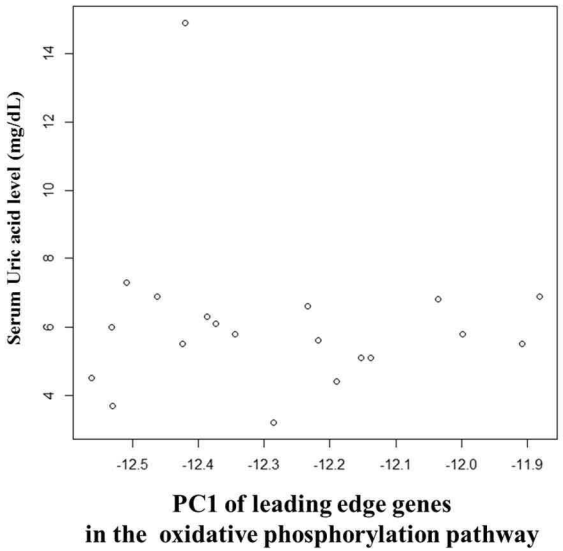
Figure 10. Association between clinical variables and PC1 of the leading edge genes from gene sets enriched in both discovery and replication datasets. Figures without P values denote statistically insignificant associations.

A. Cluster 1





B. Cluster 2


Serum Uric acid level (mg/dL)
4
6
8
10
12
14
-12.5
-12.4
-12.3
-12.2
-12.1
-12.0
-11.9
PC1 of leading edge genes
in the oxidative phosphorylation pathway

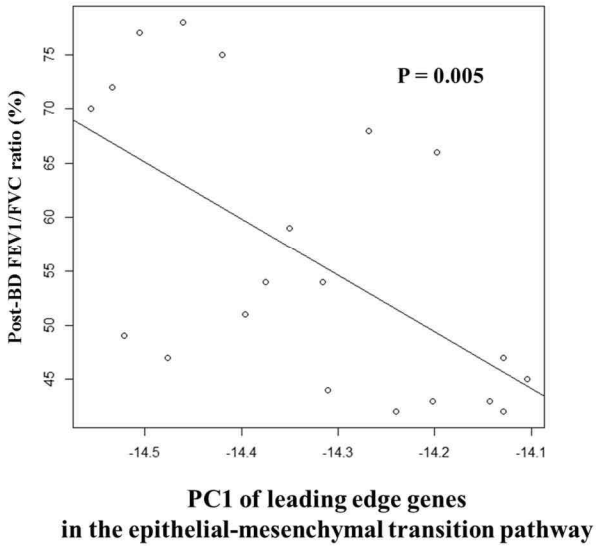

Post-BD FEV1/FVC ratio (%)
45
50
55
60
65
70
75
P = 0.005
-14.5
-14.4
-14.3
-14.2
-14.1
PC1 of leading edge genes
in the epithelial-mesenchymal transition pathway

Figure 11. Volcano plot displaying differential expressed genes. Four genes (MRPS11, HSPA9, NUDF4, and ACTA1) belong to the leading edge genes of the OXPHOS gene set and two genes (SNTB1 and FUCA1) belong to the leading edge genes of the EMT gene set.

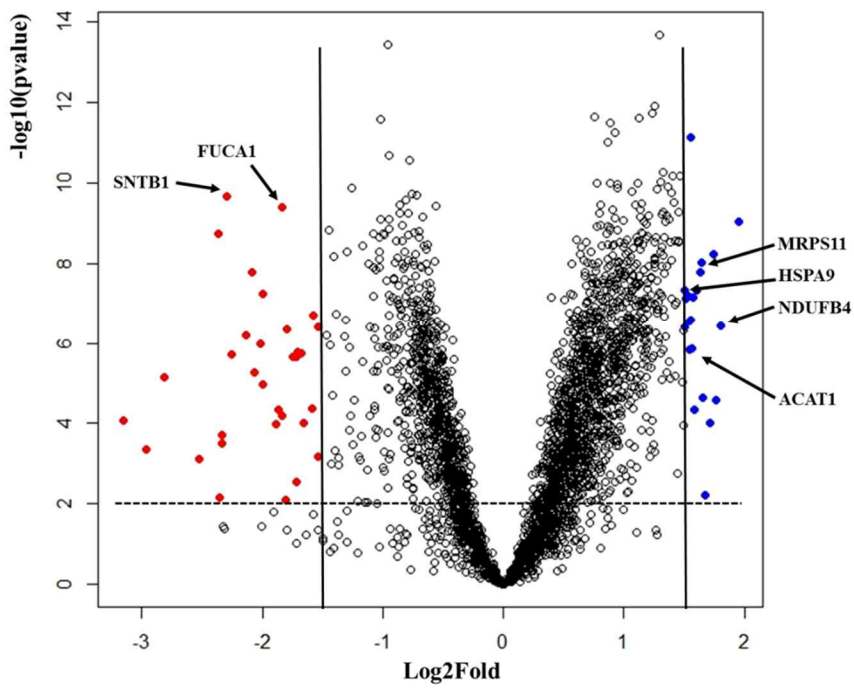
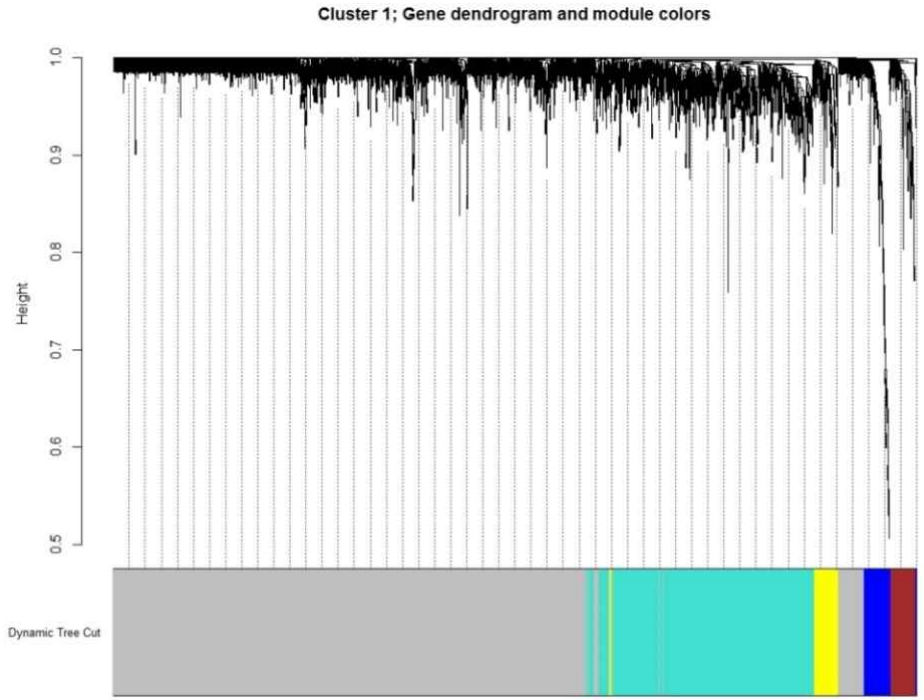


Figure 12. Result of weighted gene co-expression network analysis of discovery dataset

(A)



(B)

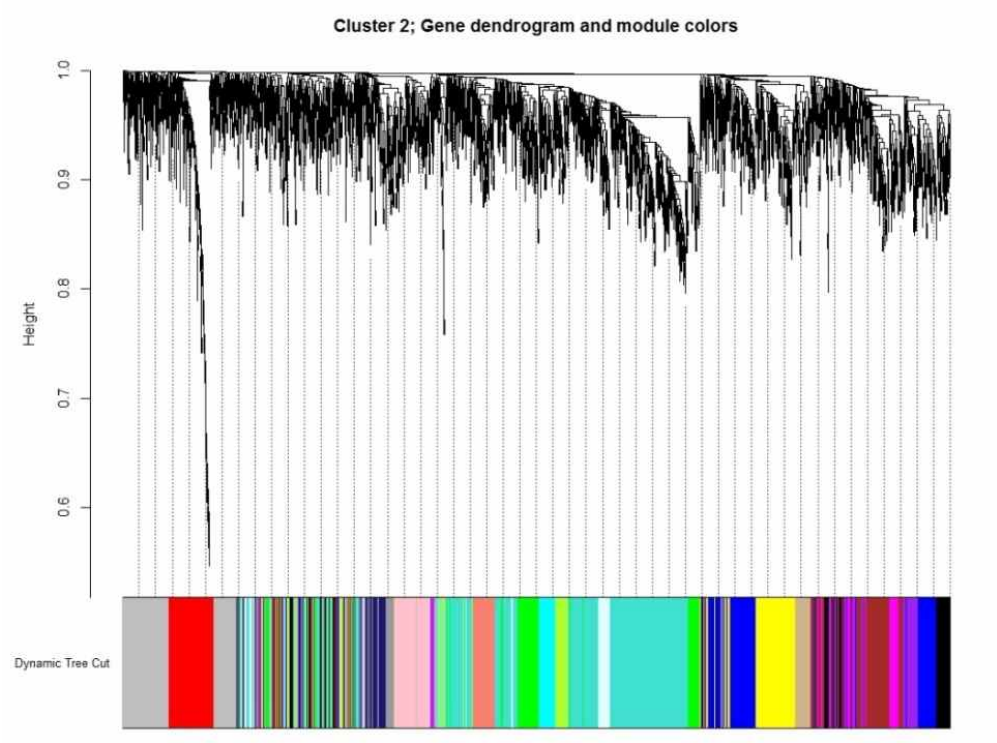
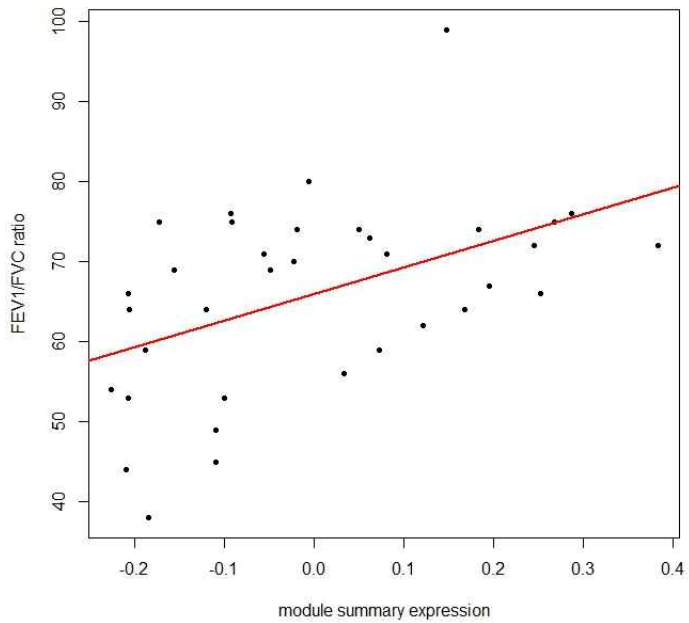

Cluster 2; Gene dendrogram and module colors
Height
1.0
0.9
0.8
0.7
0.6
Dynamic Tree Cut

Figure 13. Correlation between FEV1/FVC and gene modules (A) brown module of cluster 1. (B) Magenta module of cluster 2

(A)



(B)

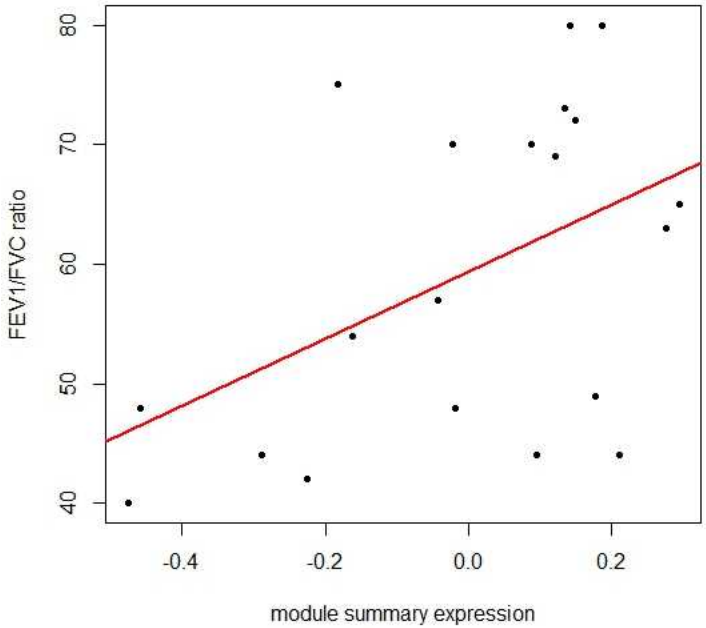
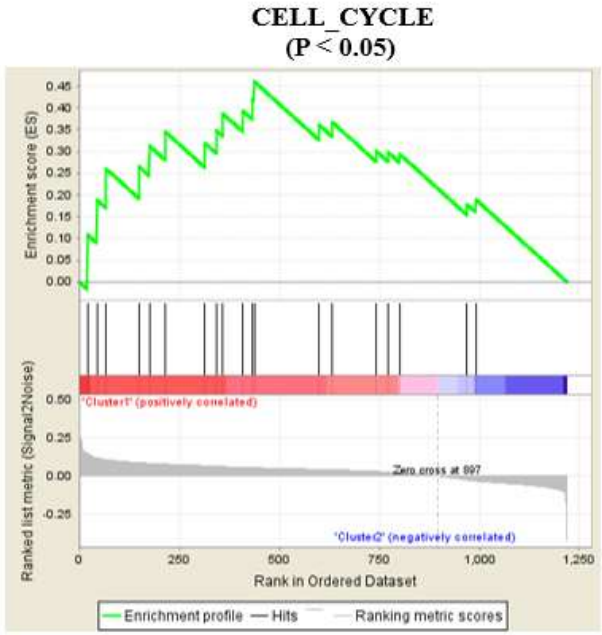


Figure 14. Gene sets enriched in (A) brown module of cluster 1 and (B) magenta module of cluster 2

(A)



(B)

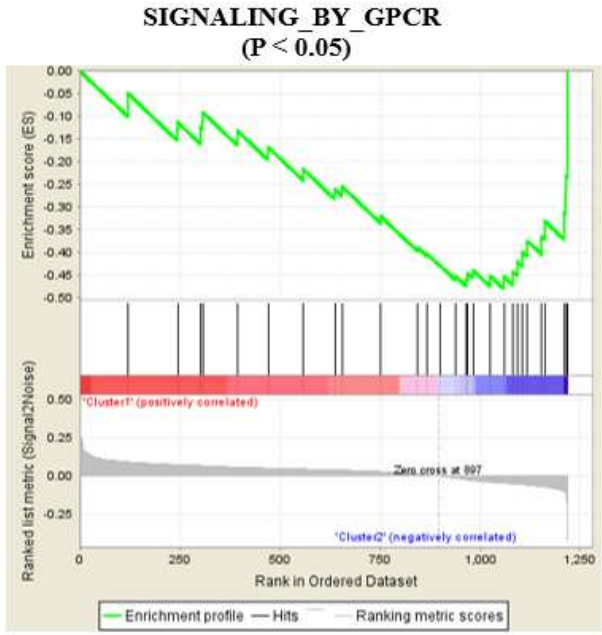
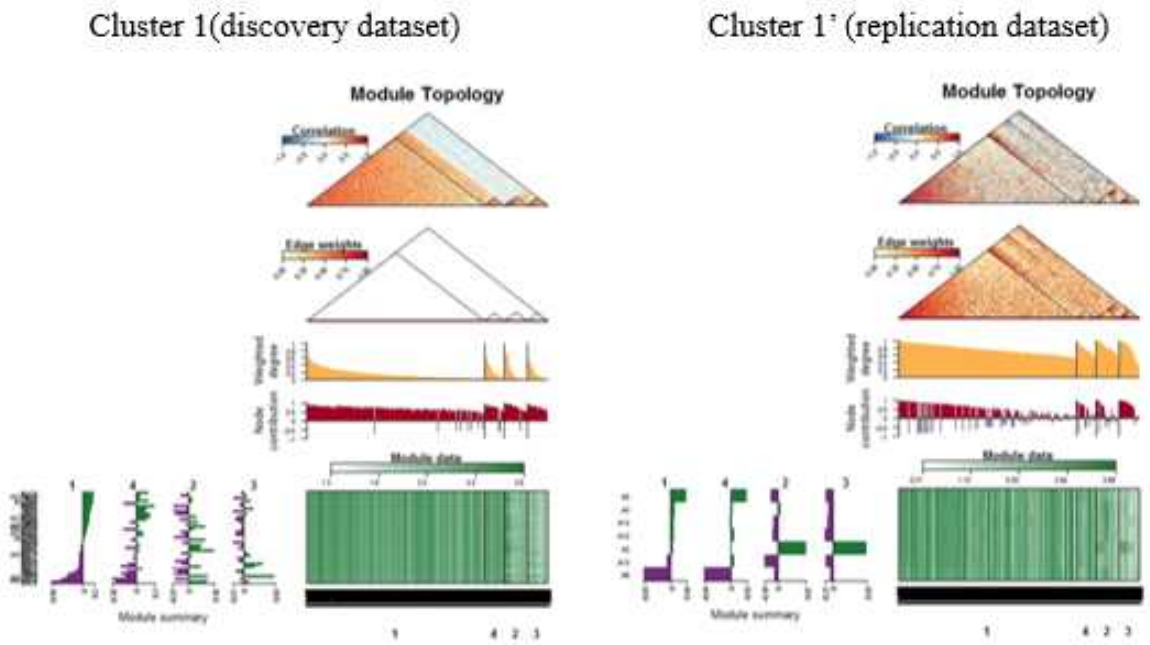
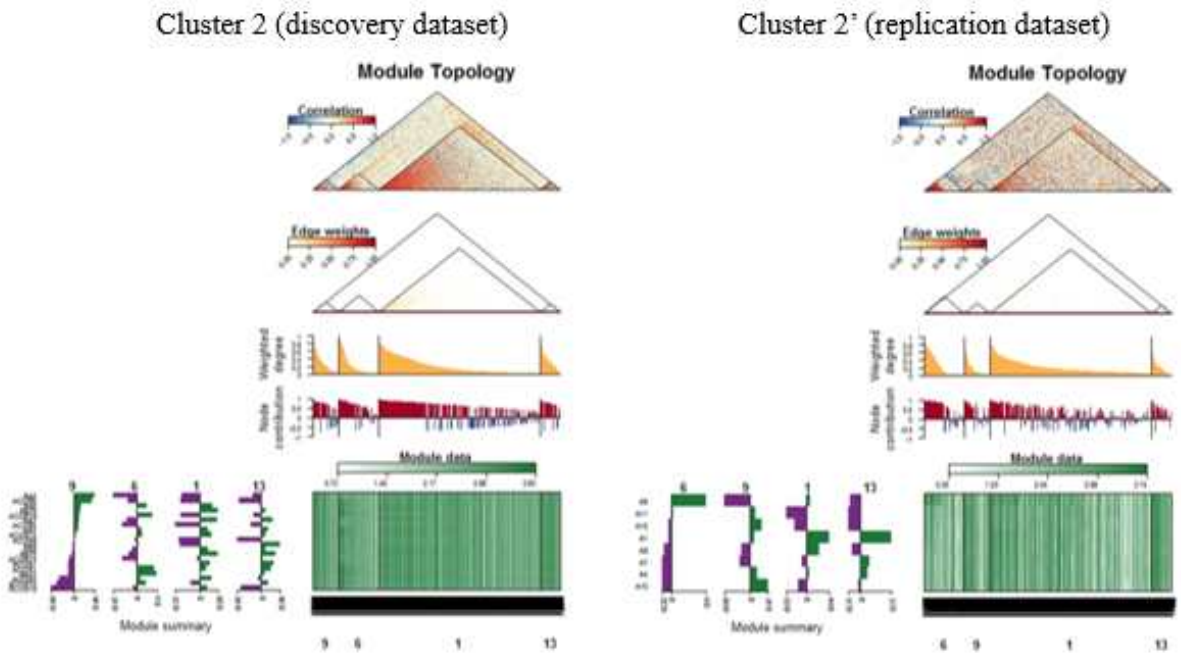


Figure 15. Preservation of gene modules identified in each cluster of discovery and replication dataset.

(A) Cluster 1 of discovery dataset and cluster 1` of replication dataset



(B) Cluster 2 of discovery dataset and cluster 2` of replication dataset



Discussion

Using gene expression profiles derived from sputum cells, we identified 3,156 gene probes showing different expression profiles between EA patients and healthy elderly controls. Using those gene probes, we could identify two molecular clusters within the EA patients. Cluster 1 showed a significantly lower proportion of eosinophils in the sputum and less severe airway obstruction compared to cluster 2. GSEA revealed that different biologic pathways were enriched in each cluster, implying that distinct and discriminative pathogenic mechanisms may exist. Although it was not statistically significant after correction using multiple comparisons, two pathways in cluster 1 of the discovery dataset (UPR, and OXPHOS) were also enriched according to subgroup analysis of individuals who had never smoked. Additionally, similar to the discovery dataset, the OXPHOS gene set was significantly enriched in one cluster and the EMT gene set was significantly enriched in the other cluster. WGCNA revealed several modules that differed between cluster 1 and 2. The brown module of cluster 1 and the magenta module of cluster 2 showed significant correlation with the FEV1/FVC ratio, and there were significantly different phenotypes between two modules. Those two modules obtained from the discovery dataset were preserved in the replication dataset after NetRep analysis, and we observed genes consistent with each module. With individual genes with two modules by WGCNA analysis of each cluster, we found two enriched gene sets (CELL_CYCLE gene set and SIGNALING_BY_GPCR gene set).

As mentioned above, induced sputum analysis is a well-known, non-invasive, and

reliable method used to evaluate inflammation, gene expression, and proteomics in airway disease including asthma (23). There are several studies that analyzed the gene expression profile of induced sputum using systems biology methodology. Yan *et al.* performed the first study in 2015 using a systems biological method to study gene expression data of induced sputum in asthma patients (32). They performed unsupervised clustering analysis and found three distinct clusters. Each of these clusters showed statistically significant phenotypic differences, and several genes were differentially expressed between the three clusters and control subjects. Kuo and his colleagues also evaluated induced sputum of moderate-to-severe asthma patients from the European Unbiased Biomarkers for the Prediction of Respiratory Diseases Outcomes (U-BIOPRED) cohort (33). They found three transcriptome-associated clusters, and co-expression network analysis revealed several hub genes that appeared to be important in asthma pathogenesis. They also found several gene pathways that were differently enriched according to the enrichment score of gene set variation analysis (GSVA). The enriched gene signatures were specific to pathways involving IL-13/Th2, ILC1, ILC2, ILC3, Th17, neutrophil activation, inflammasome, ageing, and OXPHOS signatures. Lefaudeux *et al.* also evaluated gene expression in induced sputum from asthmatic patients of U-BIOPRED (34). They found four clusters where transcriptomics and proteomic data of induced sputum showed phenotypic differences. Enrichment analysis was performed using the g:Profiler tool, and several enrichment pathways were found through the Kyoto Encyclopedia of Genes and Genomes and Reactome databases.

This is the first study analyzing gene expression profiles in induced sputum from

EA patients to evaluate candidate gene pathways affecting the pathogenic mechanism of EA. Approximately 70% of EA patients enrolled in the present study displayed symptom onset after age 65, and the mean symptom duration was 5 years. These findings suggest that EA is not merely a prolongation of NEA, and that aging itself may cause susceptibility to the development of asthma, as discussed elsewhere (35). In addition to anatomical changes, repetitive and long-standing exposures to environmental noxious stimuli cause a pro-inflammatory state in elderly subjects that results in aging-associated diseases in combination with oxidative stress (36, 37). The OXPHOS system embedded in mitochondria is the final biochemical pathway to produce ATP. A defect in the coupling between oxidation and phosphorylation causes various pathologic conditions including airway disorders (38-40). At least 70 of the OXPHOS subunits are encoded by nuclear genes whose expressions were measured in the present study, and certain defects in these genes were linked to several diseases (41). The OXPHOS system is an important source of reactive oxygen species (ROS) within cells. ROS production contributes to mitochondrial damage in a range of pathologic conditions, and this process is also important in redox reactions between the organelle to the rest of the cell (42, 43). Oxidative stress has also been linked to endoplasmic reticulum (ER) stress and to the activation of the unfolded protein response (UPR), leading to the activation of various inflammatory responses and dysregulation of the innate immune functions in the airways (44, 45). These prior observations support our findings that both the OXPHOS and UPR gene sets were significantly enriched in cluster 1. Given the previous suggestions that the OXPHOS and UPR systems play specific roles in the pathogenesis of asthma (45, 46), age-related changes in the OXPHOS and UPR systems may influence the development of

asthma in the elderly people. The observed decrease in serum uric acid levels in cluster 1 was also interesting. As anti-oxidants may scavenge ROS, an imbalance between ROS and anti-oxidant capacity is an important factor determining the effects of oxidative stress. Presently, it is not clear if EA patients in cluster 1 possess an intrinsic defect in anti-oxidant capacity. Combined with changes in the OXPHOS and UPR systems, however, reduced anti-oxidant capacity may contribute to the pathogenesis of asthma in the elderly people. Also, genes consisting brown module in cluster 1 which is correlated with FEV1/FVC ratio showed enrichment of CELL_CYCLE gene set. This also should be important in airway obstruction of EA because both OXPHOS and UPR pathway promote apoptosis (47)

Cluster 2 is characterized by the enrichment of EMT gene sets. EMT describes a situation where epithelial and mesenchymal cells can, under certain conditions, alter their phenotypes (48), and emerging evidence suggests that EMT is an important mechanism contributing to airway remodeling in asthma (49), providing an explanation for the lower post-BD FEV1 and post-BD FEV1/FVC ratio in cluster 2 compared to those in cluster 1. Senescent cells secrete a variety of proteins collectively known as the senescence-associated secretory phenotype (SASP), and these proteins can induce cellular plasticity and tissue change in a paracrine manner (50). Recently, it was reported that SASP might play a causal role in the pathogenesis of chronic obstructive pulmonary disease (51, 52). The role of SASP in the pathogenesis of asthma in the elderly people has remained completely unknown; however, cellular senescence is believed to contribute to cancer progression via EMT (53, 54). Cellular senescence observed in the airway may induce EMT and increase

the severity of asthma in the elderly people. Genes consisting magenta module in cluster 2 which is correlated with FEV1/FVC ratio showed enrichment of SIGNALING_BY_GPCR gene set. Downstream signals of G protein-coupled receptors are also well-known biologic pathway concerning airway remodeling (55). Therefore, it should be important in development of airway obstruction of EA patients of cluster 2.

Four leading edge genes (*NDUFB4*, *HSPA9*, *MRPS11*, and *ACAT1*) (Figure 11) of the OXPHOS gene set exhibited significantly dominant enrichment in cluster 1. All these genes were involved in the expression of mitochondrial enzymes or proteins. *NUDFB4* is a subunit of NADH dehydrogenase, and its expression was increased in smoking associated severe asthmatics (56, 57). Similar to the present study, it has been reported that a considerable portion of elderly patients with asthma were former or current smokers (5, 6). *HSPA9* is a gene involved in the expression of the intracellular hsp-70 group, and this gene showed significant association with stimulation of PM 2.5 to bronchial epithelial cell of the small airway of smokers (58). *MRPS11* encodes a mitochondrial ribosomal protein that is overexpressed in IL-6 stimulated bronchial epithelial cell (59). No studies exist regarding the relationship between the *ACAT1* gene and asthma or airways. Thus, these genes appear to be associated with airway responses to noxious substances, and these genes are likely to be related to the mechanism of elderly asthma, especially in cluster 1 of our study when compared to the phenotype of cluster 2. Two genes of the EMT gene set (*FUCA1* and *SNTB1*) (Figure 11) were identified as leading edge genes, with significantly dominant enrichment in cluster 2. IFN-γ increases the expression of the

FUCA1 gene and α-L-fucosidase, which is the final product of *FUCA1* gene (60). Increased α-L-fucosidase blocks wound repair in primary airway epithelium (61). It is well established that a delay in airway epithelial damage is associated with airway remodeling (62). In general, IFN-γ is known to block Th2 inflammation in the airway of asthmatics; however, mouse models of chronic airway inflammation induced by prolonged allergen challenge indicate that IFN-γ can be increased to promote airway remodeling (63). Therefore, given the phenotype, *FUCA1* may act as a key gene in elderly asthma, especially in cluster 2 of our study. The *SNTB1* gene encodes a subunit of the dystrophin-associated protein complex. This complex is known to affect the maturation of airway smooth muscle (64), but other than that, little is known concerning the effects of the *SNTB1* gene in the context of the asthmatic airway.

As mentioned above, innate immunity induced by repetitive insult to airway epithelium is presented as a possible hypothesis for EA pathogenesis, and UPR can function to integrate cell survival with the need to respond to and clear pathogens and infected cells from airways (65). The UPR integrates the metabolic challenge of producing large amounts of complex proteins of the innate immune system (66). Also, the OXPHOS system is involved in the activation of group 2 ILCs in airway inflammation that results in increased fatty acid oxidation (FAO)-dependent OXPHOS (67). These finding provide evidence that enriched gene pathways in cluster 1 may be related to EA pathogenesis via innate immunity. The other possible explanation for EA pathogenesis is to the observation of SE-IgE sensitization in late onset asthma. That phenotype of asthma is characterized by less atopy, increased sinus disease, higher eosinophilia, and more severe asthma. This phenotype also

shows a higher prevalence of chronic rhinosinusitis and nasal polyposis (15). The genetic pathway and phenotype exhibited by cluster 2 may be related to this phenotype. Although we did not find any specific pathway directly related to eosinophil biology, a previous report indicated that eosinophils promoted EMT of bronchial epithelial cells (68), and a significantly elevated sputum eosinophil proportion was observed in cluster 2. Therefore, airway eosinophilia may play an important role in the pathogenesis underlying cluster 2. Additionally, changes in EMT markers were observed in epithelial cells of nasal polyp and chronic rhinosinusitis (69, 70) .

We aimed to identify biological pathways encompassing only EA, and thus, we did not consider asthma-COPD overlap (ACO) as a subtype of EA in this study. If we define ACO as exhibiting a post-BD FEV1/FVC < 0.7 along with classical symptoms of chronic bronchitis or signs of emphysema with chest radiography or pulmonary function test (71), the proportion of ACO is higher in cluster 2 (28.6 % vs 55.5%). Although the difference exhibited a borderline significance (P = 0.083), the present study raised new insight into understanding ACO in the elderly participants. It is established that ACO is a common clinical problem in the elderly people (72). Therefore, large scale studies to examine possible associations between biological pathways enriched in cluster 2 and ACO are necessary.

There are a few general limitations to our findings. One is the small number of participants. Our results, however, were replicated using an independent data set. This was a cross-sectional study, and thus, we cannot know if the pathways identified were

the primary cause or the secondary effect of asthma. Second, we used uric acid levels as an indirect biomarker to reflect oxidative stress status, as this was easily available from routine laboratory examinations. The measurement of additional biomarkers is required to confirm our observations. Third, we analyzed gene expression of EA patients and found two different biologic pathways. As mentioned above, several studies suggest that the pathogenic mechanisms of EA and NEA are different, and there is no study which directly compares the biological pathways involved with these two subsets of asthma. Therefore, studies to elucidate different genetic mechanisms between EA and NEA are required.

In this study, we identified two distinct molecular clusters using gene expression profiles in sputum cells from patients diagnosed with EA. The OXOPHOS and UPR gene sets were significantly enriched in one cluster with lower serum levels of uric acid. The EMT gene sets were significantly enriched in one cluster with airway remodeling defined by lower post-BD FEV/FVC ratios. This is the first study to show biologic pathways that are possibly related to EA pathogenesis. EA is an extraordinary phenotype of asthma that is characterized as difficult to diagnose and manage. It likely involves mechanisms other than those observed in conventional asthma. Further studies on these mechanisms and new approaches to examine EA are needed. We postulate that genes associated with oxidative-phosphorylation and epithelial mesenchymal transition may be important in the development or progression of EA, and positive findings from future studies may result in important developments in the treatment of EA.

References

1. He W, Goodkind D, Kowal P. An Aging Wolrd: 2015 2015 [Available from: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf.

2. Gibson PG, McDonald VM, Marks GB. Asthma in older adults. The Lancet. 2010;376(9743):803-13.

3. Australian Centre for Asthma Monitoring. Asthma in Australia. Canberra: Australian Institute of Health and Welfare. 2008.

4. Park HW, Kwon HS, Kim TB, Kim SH, Chang YS, Jang AS, et al. Differences between asthma in young and elderly: results from the COREA study. Respir Med. 2013;107(10):1509-14.

5. Sano H, Iwanaga T, Nishiyama O, Sano A, Higashimoto Y, Tomita K, et al. Characteristics of phenotypes of elderly patients with asthma. Allergol Int. 2016;65(2):204-9.

6. Inoue H, Niimi A, Takeda T, Matsumoto H, Ito I, Matsuoka H, et al. Pathophysiological characteristics of asthma in the elderly: a comprehensive study. Ann Allergy Asthma Immunol. 2014;113(5):527-33.

7. Al-Alawi M, Hassan T, Chotirmall SH. Advances in the diagnosis and management of asthma in older adults. Am J Med. 2014;127(5):370-8.

8. Shin JY, Sohn KH, Shin JE, Park M, Lim J, Lee JY, et al. Changing patterns of adult asthma incidence: results from the National Health Insurance Service-

National Sample Cohort (NHIS-NSC) database in Korea. Sci Rep. 2018;8(1):15052.

9. Kradjan WA, Driesner NK, Abuan TH, Emmick G, Schoene RB. Effect of Age on Bronchodilator Response. Chest. 1992;101(6):1545-51.

10. Grönke L, Kanniess F, Holz O, Jörres RA, Magnussen H. The relationship between airway hyper-responsiveness, markers of inflammation and lung function depends on the duration of the asthmatic disease. Clinical & Experimental Allergy. 2002;32(1):57-63.

11. Porsbjerg CM, Gibson PG, Pretto JJ, Salome CM, Brown NJ, Berend N, et al. Relationship between airway pathophysiology and airway inflammation in older asthmatics. Respirology. 2013;18(7):1128-34.

12. Reed CE. Asthma in the elderly: What we do not know yet but should find out. J Allergy Clin Immunol. 2011;128(3 Suppl):S1-3.

13. Kim HY. Innate immunity in asthma. Allergy Asthma Respir Dis. 2014;2(5):317-25.

14. Chotirmall SH, Burke CM. Aging and the microbiome: implications for asthma in the elderly? Expert Rev Respir Med. 2015;9(2):125-8.

15. Song WJ, Sintobin I, Sohn KH, Kang MG, Park HK, Jo EJ, et al. Staphylococcal enterotoxin IgE sensitization in late-onset severe eosinophilic asthma in the elderly. Clin Exp Allergy. 2016;46(3):411-21.

16. Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. Biochim Biophys Acta. 2014;1842(10):1889-95.

17. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. Bioinformatics. 2005;21(15):3201-12.

18. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22(12):1540-2.

19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.

20. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2017.

21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.

22. Ritchie SC, Watts S, Fearnley LG, Holt KE, Abraham G, Inouye M. A Scalable Permutation Approach Reveals Replication and Preservation Patterns of Network Modules in Large Datasets. Cell Systems.3(1):71-82.

23. Nicholas B, Djukanovic R. Induced sputum: a window to lung pathology. Biochem Soc Trans. 2009;37(Pt 4):868-72.

24. 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(1):153-60, 60 e1-9.

25. Hekking PP, Loza MJ, Pavlidis S, De Meulder B, Lefaudeux D, Baribaud F, et al. Transcriptomic gene signatures associated with persistent airflow limitation in patients with severe asthma. Eur Respir J. 2017;50(3).

26. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention 2018 [Available from: www.ginasthma.org.

27. Sohn SW, Lee HS, Park HW, Chang YS, Kim YK, Cho SH, et al. Evaluation of cytokine mRNA in induced sputum from patients with allergic rhinitis: relationship to airway hyperresponsiveness. Allergy. 2008;63(3):268-73.

28. Kim MY, Jo EJ, Lee SE, Lee SY, Song WJ, Kim TW, et al. Reference ranges for induced sputum eosinophil counts in Korean adult population. Asia Pac Allergy. 2014;4(3):149-55.

29. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Systems Biology. 2007;1(1):54.

30. Fabbrini E, Serafini M, Colic Baric I, Hazen SL, Klein S. Effect of plasma uric acid on antioxidant capacity, oxidative stress, and insulin sensitivity in obese subjects. Diabetes. 2014;63(3):976-81.

31. Rasmussen F, Taylor DR, Flannery EM, Cowan JO, Greene JM, Herbison GP, et al. Risk Factors for Airway Remodeling in Asthma Manifested by a Low Postbronchodilator FEV1/Vital Capacity Ratio. American Journal of Respiratory and Critical Care Medicine. 2002;165(11):1480-8.

32. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma.

Am J Respir Crit Care Med. 2015;191(10):1116-25.

33. Kuo C-HS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. European Respiratory Journal. 2017;49(2).

34. Lefaudeux D, De Meulder B, Loza MJ, Peffer N, Rowe A, Baribaud F, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. J Allergy Clin Immunol. 2017;139(6):1797-807.

35. Dunn RM, Busse PJ, Wechsler ME. Asthma in the elderly and late-onset adult asthma. Allergy. 2018;73(2):284-94.

36. Monica De la F, Jaime M. An Update of the Oxidation-Inflammation Theory of Aging: The Involvement of the Immune System in Oxi-Inflamm-Aging. Current Pharmaceutical Design. 2009;15(26):3003-26.

37. Bullone M, Lavoie JP. The Contribution of Oxidative Stress and Inflamm-Aging in Human and Equine Asthma. Int J Mol Sci. 2017;18(12).

38. DiMauro S, Bonilla E, Davidson M, Hirano M, Schon EA. Mitochondria in neuromuscular disorders. Biochimica et Biophysica Acta (BBA) - Bioenergetics. 1998;1366(1):199-210.

39. Shoubridge EA. Mitochondrial encephalomyopathies. Current opinion in neurology. 1998;11(5):491-6.

40. Prakash YS, Pabelick CM, Sieck GC. Mitochondrial Dysfunction in Airway Disease. Chest. 2017;152(3):618-26.

41. Smeitink J, van den Heuvel L, DiMauro S. The genetics and pathology of oxidative phosphorylation. Nature Reviews Genetics. 2001;2:342.

42. Balaban RS, Nemoto S, Finkel T. Mitochondria, oxidants, and aging. Cell. 2005;120(4):483-95.

43. Dröge W. Free Radicals in the Physiological Control of Cell Function. Physiological Reviews. 2002;82(1):47-95.

44. Malhotra JD, Kaufman RJ. Endoplasmic reticulum stress and oxidative stress: a vicious cycle or a double-edged sword? Antioxid Redox Signal. 2007;9(12):2277-93.

45. Pathinayake PS, Hsu AC, Waters DW, Hansbro PM, Wood LG, Wark PAB. Understanding the Unfolded Protein Response in the Pathogenesis of Asthma. Front Immunol. 2018;9:175.

46. Reddy PH. Mitochondrial Dysfunction and Oxidative Stress in Asthma: Implications for Mitochondria-Targeted Antioxidant Therapeutics. Pharmaceuticals (Basel). 2011;4(3):429-56.

47. Yadav N, Kumar S, Marlowe T, Chaudhary AK, Kumar R, Wang J, et al. Oxidative phosphorylation-dependent regulation of cancer cell apoptosis in response to anticancer agents. Cell Death Dis. 2015;6:e1969.

48. Micalizzi DS, Farabaugh SM, Ford HL. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. J Mammary Gland Biol Neoplasia. 2010;15(2):117-34.

49. Hackett TL. Epithelial-mesenchymal transition in the pathophysiology of airway remodelling in asthma. Curr Opin Allergy Clin Immunol. 2012;12(1):53-9.

50. Ritschka B, Storer M, Mas A, Heinzmann F, Ortells MC, Morton JP, et al. The senescence-associated secretory phenotype induces cellular plasticity and tissue regeneration. Genes Dev. 2017;31(2):172-83.

51. Dagouassat M, Gagliolo JM, Chrusciel S, Bourin MC, Duprez C, Caramelle P, et al. The cyclooxygenase-2-prostaglandin E2 pathway maintains senescence of chronic obstructive pulmonary disease fibroblasts. Am J Respir Crit Care Med. 2013;187(7):703-14.

52. Houssaini A, Breau M, Kebe K, Abid S, Marcos E, Lipskaia L, et al. mTOR pathway activation drives lung cell senescence and emphysema. JCI Insight. 2018;3(3).

53. Smit MA, Peeper DS. Epithelial-mesenchymal transition and senescence: two cancer-related processes are crossing paths. Aging (Albany NY). 2010;2(10):735-41.

54. Laberge RM, Awad P, Campisi J, Desprez PY. Epithelial-mesenchymal transition induced by senescent fibroblasts. Cancer Microenviron. 2012;5(1):39-44.

55. Billington CK, Penn RB. Signaling and regulation of G protein-coupled receptors in airway smooth muscle. Respiratory Research. 2003;4(1):4.

56. Bigler J, Boedigheimer M, Schofield JPR, Skipp PJ, Corfield J, Rowe A, et al. A Severe Asthma Disease Signature from Gene Expression Profiling of Peripheral Blood from U-BIOPRED Cohorts. American Journal of Respiratory and Critical Care

Medicine. 2016;195(10):1311-20.

57. Takahashi K, Pavlidis S, Ng Kee Kwong F, Hoda U, Rossios C, Sun K, et al. Sputum proteomics and airway cell transcripts of current and ex-smokers with severe asthma in U-BIOPRED: an exploratory analysis. Eur Respir J. 2018;51(5).

58. O'Beirne SL, Shenoy SA, Salit J, Strulovici-Barel Y, Kaner RJ, Visvanathan S, et al. Ambient Pollution Related Reprogramming of the Human Small Airway Epithelial Transcriptome. American Journal of Respiratory and Critical Care Medicine. 2018.

59. Jevnikar Z, Ostling J, Ax E, Calven J, Thorn K, Israelsson E, et al. Epithelial IL-6 trans-signaling defines a new asthma phenotype with increased airway inflammation. J Allergy Clin Immunol. 2018.

60. Sobkowicz AD, Gallagher ME, Reid CJ, Crean D, Carrington SD, Irwin JA. Modulation of expression in BEAS-2B airway epithelial cells of alpha-L-fucosidase A1 and A2 by Th1 and Th2 cytokines, and overexpression of alpha-L-fucosidase 2. Mol Cell Biochem. 2014;390(1-2):101-13.

61. Dorscheid DR, Wojcik KR, Yule K, White SR. Role of cell surface glycosylation in mediating repair of human airway epithelial cell monolayers. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2001;281(4):L982-L92.

62. Holgate ST, Davies DE, Puddicombe S, Richter A, Lackie P, Lordan J, et al. Mechanisms of airway epithelial damage: epithelial-mesenchymal interactions in the pathogenesis of asthma. European Respiratory Journal. 2003;22(Supplement 44):24s-

9s.

63. Bergeron C, Tulic MK, Hamid Q. Airway remodelling in asthma: From benchside to clinical practice. Canadian Respiratory Journal : Journal of the Canadian Thoracic Society. 2010;17(4):e85-e93.

64. Sharma P, Tran T, Stelmack GL, McNeill K, Gosens R, Mutawe MM, et al. Expression of the dystrophin-glycoprotein complex is a marker for human airway smooth muscle phenotype maturation. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2008;294(1):L57-L68.

65. Osorio F, Lambrecht B, Janssens S. The UPR and lung disease. Semin Immunopathol. 2013;35(3):293-306.

66. Whitsett JA, Alenghat T. Respiratory epithelial cells orchestrate pulmonary innate immunity. Nat Immunol. 2015;16(1):27-35.

67. Wilhelm C, Kharabi Masouleh S, Kazakov A. Metabolic Regulation of Innate Lymphoid Cell-Mediated Tissue Protection-Linking the Nutritional State to Barrier Immunity. Front Immunol. 2017;8:1742.

68. Yasukawa A, Hosoki K, Toda M, Miyake Y, Matsushima Y, Matsumoto T, et al. Eosinophils promote epithelial to mesenchymal transition of bronchial epithelial cells. PLoS One. 2013;8(5):e64281.

69. Shin HW, Cho K, Kim DW, Han DH, Khalmuratova R, Kim SW, et al. Hypoxia-inducible factor 1 mediates nasal polypogenesis by inducing epithelial-to-mesenchymal transition. Am J Respir Crit Care Med. 2012;185(9):944-54.

70. Hupin C, Gohy S, Bouzin C, Lecocq M, Polette M, Pilette C. Features of mesenchymal transition in the airway epithelium from chronic rhinosinusitis. Allergy. 2014;69(11):1540-9.

71. Lee H, Tho NV, Nakano Y, Lee BJ, Park HY. A diagnostic approach and natural course of a patient with asthma-COPD overlap syndrome. Respirol Case Rep. 2015;3(4):119-21.

72. Zeki AA, Schivo M, Chan A, Albertson TE, Louie S. The Asthma-COPD Overlap Syndrome: A Common Clinical Problem in the Elderly. J Allergy (Cairo). 2011;2011:861926.

국 문 초 록

객담 전사체 분석을 통한 노인 천식의 생물학적 경로 분석

김 병 근

서울대학교 대학원

의학과 중개의학전공

서론: 노인 천식은 전형적인 천식과는 다른 특징을 가지고 있다. 노인 천식의 유병율은 지속적으로 증가하고 있으나 그 병리 기전은 불분명하다. 본 연구는 객담에서 추출한 유전자 발현을 최근의 새로운 방법론을 이용하여 분석하고 이를 통하여 노인 천식과 관련된 생물학적 경로를 규명하고자 한다.

방법: 노인 천식 환자와 정상 대조군의 유도 객담에서 유전자 발현을 분석하였다. 두 군 간 차별 발현되는 3156개의 유전자 프로브를 확인하고 이를 이용하여 계층 클러스터링을 시행하여 노인 천식 환자를 두 클러스터 (Cluster) 로 구분하였다. Gene set enrichment analysis (GSEA)과

Weighted gene co-expression network analysis (WGCNA)를 통하여 유전자 발현을 분석하였고 노인 천식의 새로운 생물학적 경로를 확인하였다. 또한 Gene Expression Omnibus (GEO) 에 공개된 유전자 정보를 이용하여 이러한 결과를 재확인하였다.

결과: 55명의 노인 천식 환자와 10명의 정상 노인 대조군이 연구에 참여하였다. 클러스터 분석을 통하여 뚜렷한 두개의 군집이 확인되었다. 35명의 환자로 구성된 첫번째 군집 (Cluster 1)은 20명의 환자로 구성된 두번째 군집 (Cluster 2)에 비하여 객담 호산구 분율이 낮고 기도 폐색이 경미한 특징을 보였다. GEO에 공개된 유전자 정보를 이용하여 클러스터 분석을 시행하였을 때 역시 두개의 군집이 확인되었다 (Cluster 1‘와 Cluster 2‘). GSEA 를 시행하여 Cluster 1에서 5개의 유전자 세트와 Cluster 2에서 3개의 유전자 세트의 발현이 증가하여 있음을 확인하였다. GEO에서 공개된 유전자 정보를 이용하여 같은 방법으로 분석하여 발현이 증가되어 있는 유전자 군집을 확인하였고 이 중 2개의 유전자 세트 [Cluster 1‘ 에서 OXIDATIVE_PHOSPHORYLATION (OXPHOS), Cluster 2‘ 에서 EPITHELIAL_MESENCHYMAL_TRANSITION (EMT))가 공통적으로 발현이 증가되어 있음을 확인하였다. 또한 이 두 유전자 세트는 비흡연자들만을 대상으로 분석하였을 때도 발현이 증가되어 있었다. 두개의 유전자 군집을 추가로 분석하였을 때 첫번째 군집의 OXPHOS

유전자 세트에서 4개의 leading edge 유전자 (*MRPS11*, *HSPA9*, *NUDF4*, *ACTA1*) 와 두번째 군집의 EMT 유전자 세트에서 2개의 leading edge 유전자 (*SNTB1*, *FUCA1*)를 확인하였다. WGCNA를 통한 분석에서 첫번째 군집에서 4개, 두번째 군집에서 18개의 모듈을 확인하였고, 첫번째 군집의 갈색 모듈과 두번째 군집의 자홍색 모듈이 노인 천식 환자의 FEV1/FVC 비율과 연관이 있음을 확인하였다. 이 두 모듈은 GEO 자료를 이용한 분석에서도 보존됨을 확인하였다.

결론: 노인 천식 환자의 객담에서 두개의 뚜렷한 유전자 세트가 각 군집에서 발현이 증가되어 있음을 확인하였다. 이는 노인 천식의 병리 기전에서 두개의 다른 생물학적 경로가 작용함을 시사하는 소견이라 할 수 있겠다. 이는 노인 천식의 발생과 진행에서 중요한 경로로 보이며 향후 노인 천식의 치료에도 중요할 수 있을 것으로 보인다.

주요어: 군집 분석, 노인 천식, 유전자 경로, Gene set enrichment analysis, 시스템 생물학, 유전자 전사체 분석, Weighted gene co-expression network analysis

학번: 2015 -30003