

ORIGINAL RESEARCH

Hideyasu Yamada et al

## **Effects of lung function-related genes and *TSLP* on COPD phenotypes**

Hideyasu Yamada<sup>1,2</sup>, Norihito Hida<sup>2</sup>, Hironori Masuko<sup>1</sup>, Tohru Sakamoto<sup>1</sup>, Nobuyuki Hizawa<sup>1</sup>

<sup>1</sup>Department of Pulmonary Medicine, Faculty of Medicine, University of Tsukuba  
Tennodai 1-1-1, Tsukuba, Ibaraki 305-8577, Japan

<sup>2</sup>Department of Respiratory Medicine, Hitachi Ltd, Hitachinaka General Hospital, 20-1  
Ishikawa-cho, Hitachinaka, Ibaraki 312-0057, Japan

Correspondence:

Hideyasu Yamada

Department of Pulmonary Medicine, Faculty of Medicine, University of Tsukuba  
Tennodai 1-1-1, Tsukuba, Ibaraki 305-8577, Japan

E-mail address: h.yamada@md.tsukuba.ac.jp

**Tel: +81-29-354-5111 FAX: +81-029-354-6842**

## Abstract (243 words)

A weighted genetic risk score (GRS) based on 16 SNPs implicated in reduced lung function in both Japanese and non-Japanese populations was previously associated with the onset of COPD and asthma. We here examine the genetic impact of this lung function GRS on specific COPD phenotypes. A cohort of Japanese COPD patients (N=270) underwent lung function testing followed by genotyping with allele-specific arrays for 16 SNPs as well as expression quantitative trait loci at *TSLP* (rs2289276, rs3806933). Lung function GRS scoring and two-step cluster analyses grouped patients into different COPD phenotypes based on gender, age, smoking index, %FEV<sub>1</sub> and lung function GRS. The genetic effect of *TSLP* on COPD phenotypes was also examined for interactions with the lung function GRS. A total of 270 participants were grouped into 5 clusters. The cluster with the highest levels of lung function GRS was characterized by moderate to severe airflow obstruction and the highest blood eosinophil counts. Regarding *TSLP*, an increased number of T alleles at both SNPs was found in the cluster characterized by moderate to severe airflow obstruction and heavy smoking (rs2289276, *P* value = 0.035; rs3806933, *P* value = 0.047) independent of the lung function GRS. A genetic susceptibility to impaired lung function carries an increased risk of developing COPD characterized by increased eosinophil counts and severe airflow obstruction while individuals with increased *TSLP* responses to external stimuli have an independent risk of developing severe airflow obstruction in the presence of heavy smoking.

**Keywords:** COPD, Genetic risk score, TSLP, lung function, eosinophil

## Introduction

Chronic obstructive pulmonary disease (COPD) is a heterogeneous disease similar to asthma and previous studies have reported that various genes are associated with its onset and phenotypic expression [1,2]. Previously, we constructed a weighted, multi-single nucleotide polymorphism (SNP) genetic risk score (GRS) for decreased forced expiratory volume in 1 second over forced vital capacity ratio ( $FEV_1/FVC$ ) using genotypic information for 16 genes previously associated with  $FEV_1/FVC$  in a genome-wide association study (GWAS) of Japanese and non-Japanese populations. Some asthmatic children carry a risk for COPD as a result of poorly developed pulmonary growth and function [3] and our previous study showed that an increased lung function GRS was associated with COPD and asthma development. This stresses the importance and implications of genetic susceptibility to impaired lung development (as reflected in increased GRS) in the etiology of reduced lung function and later progression to asthma and COPD [4]. As the lung function GRS was significantly higher in patients with an atopic asthma characterized by early onset and severe airflow obstruction [4], we hypothesized that genetic susceptibility to impaired lung development plays an important role in the development of particular COPD phenotypes. Here, we identify specific COPD phenotypes associated with genetic susceptibility to impaired lung function as indicated by a multi-SNP GRS.

Exaggerated TSLP, expressed in the airways of patients with asthma and COPD [5] and correlating with respiratory deterioration in healthy and asthmatic individuals [6], plays a detrimental role in lung function and the pathogenesis of chronic inflammatory lung diseases [6,7]. COPD patients with higher lung function GRS may be more prone to

viral infections, lung barrier abnormalities, and increased sensitivity to smoking and/or inhaled allergens, indicating that the enhancement of TSLP production from airway tissues is a common pathobiology underlying specific asthma and COPD phenotypes. Thus, we also examined if the genetic impact of *TSLP* could modify the phenotypic expression of COPD interactions with lung function GRS.

## **Materials and methods**

### **Ethical Statement**

The Human Genome Analysis and Epidemiology Research Ethics Committee of the University of Tsukuba and the Human Genome/Gene Analysis Research Ethics Review Committees of the Tsukuba Medical Center and RIKEN approved the study. Written, informed consent was obtained from each participant in accordance with institutional requirements and the principles of the Declaration of Helsinki (latest IRB approval date: May 5th, 2019; IRB number: H29-294).

### **Study population**

Our COPD cohort consisted of 270 patients from the University of Tsukuba Hospital and its affiliated hospitals [4]. The healthy volunteers in this study were recruited from individuals who had an annual health exam at the Tsukuba Medical Center [6].

Participant characteristics are summarized in Table 1. The data are presented as mean  $\pm$  SD unless noted otherwise. All 270 patients were included in our previous report [4]. Inclusion criteria were physician diagnosis of COPD based on smoking history of at least 10 pack-years, clinical symptoms (including exertional dyspnea and/or chronic bronchitis), self-reported chronic cough and phlegm, and an abnormal spirometric value.

## **Lung function**

Spirometry was performed in accordance with the criteria established by the Japanese Respiratory Society (JRS) [8]. Briefly, experienced spirometry technicians continuously monitored the procedure and reviewed the flow-volume curves to ensure adherence to the standards. Participants performed up to 3 forced expiratory maneuvers so that acceptable readings could be obtained [9].

## **Genotyping**

We genotyped 270 COPD patients using the TaqMan allele-specific amplification method (Applied Biosystems, Foster City, CA, USA), as described previously, to identify 16 SNPs and functional SNPs at *TSLP* (rs2289276, rs3806933) [10].

Our previous studies [6,10] showed significant associations of *TSLP* SNPs, rs3806933 (−847C/T) and rs2289276 (−82C/T) with both asthma susceptibility and lower FEV<sub>1</sub> in non-COPD non-asthmatic subjects. Within the promoter region of the *TSLP* gene, the rs3806933 T allele creates a binding site for the transcription factor activating protein (AP)-1 and increases promoter-reporter activity of *TSLP* in response to poly(I:C)

stimulation in human bronchial epithelial cells [10]. Sequences containing the rs2289276 (–82C/T) SNP bind to the AP-2 $\alpha$  protein with a higher affinity than –82C (on the protective allele) [10].

## **Calculating GRS**

As previously described, the lung function GRS was calculated as a weighted sum of the number of high-risk alleles at 16 SNPs [4]. The 16 SNP GRS was significantly associated with reduced FEV<sub>1</sub>/FVC but not with annual decline of FEV<sub>1</sub> in healthy individuals and, thus, these 16 SNPs or their corresponding genetic pathways were selected as potential factors that influence maximum adult lung function involved in impaired lung growth or development rather than accelerating lung function decline [4]. Pathway analysis showed that epithelial-mesenchymal transition (EMT) pathway molecules, such as TGFB2, TNF, and NOTCH4, could be responsible for the biological relevance of the lung function GRS [4].

Briefly, the GRS, which combines weighted effects of multiple SNPs into a single variable, is calculated as a weighted sum of the number of high-risk alleles. The score was then divided by the highest risk score multiplied by 100 to rescale the scores to a range between 0 and 100 [4].

## **Statistical analysis**

Clustering gives GRS the specificity to serve as both a genetic marker to identify COPD phenotypes related to deregulated lung function and clarify the mechanisms

associated with COPD heterogeneity. Therefore, we conducted cluster analysis with respect to GRS. Two-step cluster analyses were performed with four factors: GRS, percent predicted FEV<sub>1</sub>, number of pack-years of cigarette smoking, and age using IBM SPSS Statistics, version 24. All factors were standardized, the similarity between two clusters was measured using log likelihood, and the best number of clusters was defined using Bayesian information criterion (BIC) values. We initially generated a dendrogram for estimation of the number of likely clusters within the population using Ward's method and fed this into a K-means cluster analysis, which was used as the principal clustering technique [11].

Differences in *TSLP* genotypes among clusters were analyzed both by a chi-square analysis and by a multinomial logistic regression analysis. For multinomial logistic analysis, the 1471 healthy adults were used as a reference group in R for Windows ver.3.4.3 [12]. Differences in blood eosinophil counts among clusters was analyzed by analysis of variance (ANOVA), correcting for age and dosage of inhaled corticosteroids.

As markers in the biological candidate genes might not yield significant p-values after correction for multiple testing but still replicate consistently over independent populations tested, we therefore set a liberal standard of “any nominal  $p < 0.05$ ”. The two *TSLP* SNPs are in strong linkage disequilibrium and were not considered multiple tests. For the multinomial regression analysis to predict the probabilities of the different possible phenotypic outcomes, we did not correct by “5” as this analysis was for hypothesis-generating purposes to see if any specific phenotype(s) could be predicted by the functional *TSLP* genotypes. In addition, examining a single multinomial regression

model using all individuals already adjusted the p-value inflation, which is different from running the multiple logistic regression analysis for each smaller group separately.

## **Results**

### **Subject characteristics**

All the patients in the current study primarily had a physician diagnosis of smoking-induced COPD (Table 1). The presence of bronchiectasis was ruled out in all these patients whereas eleven patients had a history of asthma. In this study, 13 patients whose FEV<sub>1</sub>/FVC was more than 0.7 after bronchodilator administration were also included in the study because they had a clinical diagnosis of chronic bronchitis based on the presence of chronic cough and sputum production for at least 3 months per year for two consecutive years. In addition, among the 270 participants, we found that 24 had an FEV<sub>1</sub> above the lower limit of normal (LLN) after bronchodilator administration.

### **Two-step cluster analysis**

Our two-step cluster analysis identified five clusters of COPD, the characteristics of which are shown in Table 2a. The characteristics of each cluster are as follows: Cluster 1: milder airflow obstruction with the lowest GRS; Cluster 2: severe airflow obstruction with higher blood eosinophil counts; Cluster 3: the most severe airflow obstruction with the lowest blood eosinophil count; Cluster 4: the highest GRS ( $P = 1.04 \times 10^{-24}$ , Figure 1) with moderate to severe airflow obstruction plus the highest blood



eosinophil counts (ANOVA adjusted for age and an ICS dose,  $P$  value = 0.011); and Cluster 5: moderate to severe airflow obstruction plus a history of heavy smoking.

### ***TSLP* functional SNP genotypes among clusters**

The functional *TSLP* genotypes among clusters were compared by chi-square testing and the differences were found to be statistically significant (rs2289276,  $P$  value = 0.035; rs3806933,  $P$  value = 0.047). In the multinomial logistic analysis, there was a statistically higher number of patients carrying T alleles in Cluster 5 compared to healthy participants (rs2289276,  $P$  value = 0.017; rs3806933,  $P$  value = 0.034, [Table 3a](#)). We found no genetic impact of the *TSLP* genotype on Cluster 4, indicating it has no influence in the effect of lung function GRS on the specific COPD phenotype.

- **Sensitivity analysis using COPD patients only whose FEV<sub>1</sub> was less than their LLN**

We also conducted the same analysis using patients whose diagnosis of COPD was based on an FEV<sub>1</sub> less than their LLN (N=246). In this analysis, similar genetic effects were found in the specific COPD phenotypes for both GRS and *TSLP* genotypes (Table 2b, Table 3b).

## **Discussion**

We conducted a genetic association analysis of candidate gene regions of lung function GRS and *TSLP* based upon the *a priori* understanding of the importance of these

genetic factors in COPD pathogenesis. Given that COPD is a heterogeneous condition involving several sub-phenotypes, any of which may differ in clinical presentation, disease mechanisms, underlying causes and genetic associations, we initially performed cluster analyses based on the lung function GRS, smoking index, respiratory function test values, and age in a cohort of Japanese COPD patients. Identifying at-risk patients based on these genotypes could allow better diagnosis of the disease at early stages with understanding of the involved molecular pathways that lead to disease manifestation.

We found that the cluster with the highest GRS was characterized by the highest blood eosinophil count and moderate to severe airflow obstruction. Population-based studies have shown significant heterogeneity in COPD patients with regard to both the attainment of maximal lung function and a subsequent decline over time [13,14] and, in addition, childhood asthma is a known risk factor for developing COPD [5]. Given that we previously identified the lung function-related GRS associated with early-onset atopic asthma patients with severe airflow obstruction, the results of the current study indicate that individuals with higher GRS are also likely to have an impaired attainment of maximal lung function and are more susceptible to the development of a specific phenotype of COPD (an eosinophilic COPD or asthma-COPD overlap) later in life [15,16]. COPD patients with higher GRS, therefore, may derive a greater benefit from treatments targeting type-2 inflammation, including ICS.

*TSLP* is another important gene implicated in the pathogenesis of asthma [10]. Our study found that the *TSLP* genotype functionally related to increased mRNA expression was increased in Cluster 5, characterized by heavy smoking and moderate to severe airflow obstruction. As *TSLP*, mRNA and protein are elevated in COPD [17,18],

and the number of TSLPR+ immunoreactive cells increased in the epithelium of patients with severe, stable COPD compared with control non-smokers [19], our findings suggest that, in individuals with an increased susceptibility to an exaggerated expression of *TSLP*, cigarette smoke-induced airway inflammation may be locally enhanced via reactive increases in TSLP signaling. This may thus represent a key regulator of disease pathogenesis in a specific COPD phenotype.

We note that the current study had a relatively small size given the complexity and heterogeneity of COPD that is associated with numerous genetic variants. However, a single study in a single population cannot cover all potential phenotypes, all possible gene-environment interactions, and all different genetic models. Our objective was not to identify a novel COPD gene; rather, the goal of this investigation was to explicate the role of genetic factors such as lung function GRS and *TSLP* genotypes in specific COPD phenotypes. Patients with COPD in the current study were mainly elderly men, reflecting the high prevalence of smoking among men in Japan. As women may be more susceptible to developing COPD in response to cigarette smoke exposure than men, our study failed to contemplate sex differences in COPD susceptibility and underlying mechanisms. Despite these study limitations, our findings indicate that these genetic factors continue to be compelling targets for functional studies and further genetic replication in independent populations. Therefore, we believe that such exploratory work is still useful to provide a basis for future studies.

## Conclusion

In conclusion, lung function GRS, in addition to childhood-onset asthma with advanced airflow obstruction, may serve as biomarkers for developing specific variants of COPD, especially eosinophilic COPD. Additionally, variable *TSLP* gene activation in the presence of heavy smoking, which is also related to asthma pathogenesis, may contribute to this progression into COPD phenotypes. The differential associations of the lung function GRS and *TSLP* genotypes with specific phenotypes may reflect the distinct pathogenesis underlying each COPD cluster. Such information will be valuable for disease treatment and primary prevention during childhood. Our analyses also provide further support for the overlap of the pathogenesis between asthma and COPD.

## **Acknowledgments**

We thank Dr. Bryan J. Mathis of the Medical English Communications Center, University of Tsukuba, for English revision.

## **Disclosure**

The authors have no conflicts of interest to declare.

## **Funding**

None

## References

1. Agustí A. The path to personalised medicine in COPD. *Thorax*. 2014; 69(9): 857-64.
2. Lange P, Celli B, Agustí A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2015; 373(2): 111-22.
3. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med*. 2016; 374: 1842-52
4. Yamada H, Masuko H, Yatagai Y, et al. Role of Lung Function Genes in the Development of Asthma. *PLoS One*. 2016; 11(1): e0145832.
5. Smelter DF, Sathish V, Thompson MA, et al. Thymic stromal lymphopoietin in cigarette smoke-exposed human airway smooth muscle. *J Immunol*. 2010; 185(5): 3035-40.
6. Masuko H, Sakamoto T, Kaneko Y, et al. Lower FEV1 in non-COPD, nonasthmatic subjects: association with smoking, annual decline in FEV1, total IgE levels, and TSLP genotypes. *Int J Chron Obstruct Pulmon Dis*. 2011; 6: 181-9.
7. Iijima H, Kaneko Y, Yamada H, et al. A distinct sensitization pattern associated with asthma and the thymic stromal lymphopoietin (TSLP) genotype. *Allergol Int*. 2013; 62(1): 123-30.
8. Kubota M, Kobayashi H, Quanjer PH, et al. Reference values for spirometry, including vital capacity, in Japanese adults calculated with the LMS method and compared with previous values. *Respir Investig*. 2014 ; 52(4): 242-50.
9. Yamada H, Yatagai Y, Masuko H, et al. Heritability of pulmonary function estimated from genome-wide SNPs in healthy Japanese adults. *Respir Investig*. 2015; 53(2): 60-7.

10. Harada M, Hirota T, Jodo AI, et al. Thymic stromal lymphopoietin gene promoter polymorphisms are associated with susceptibility to bronchial asthma. *Am J Respir Cell Mol Biol*. 2011;44(6):787-93.
11. Ball GH, Hall DJ. A clustering technique for summarizing multivariate data. *Behav Sci*. 1967; 12: 153–5
12. Sun B, VanderWeele T, Tchetgen EJ. A Multinomial Regression Approach to Model Outcome Heterogeneity. *Am J Epidemiol*. 2017;186(9):1097-103.
13. Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med*. 2013;1(9):728-42.
14. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019 ;51(3):481-93.
15. Belgrave DCM, Granell R, Turner SW, et al. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir Med*. 2018 ;6(7):526-34.
16. Woodruff PG, van den Berge M, Boucher RC, et al. American Thoracic Society/National Heart, Lung, and Blood Institute asthma-chronic obstructive pulmonary disease overlap workshop report. *Am J Respir Crit Care Med*. 2017;196:375–81.
17. Ying S, O'Connor B, Ratoff J, et al. Expression and cellular provenance of thymic stromal lymphopoietin and chemokines in patients with severe asthma and chronic obstructive pulmonary disease. *J Immunol*. 2008 ;181(4):2790-8.

18. Anzalone G, Albano GD, Montalbano AM, et al. IL-17A-associated IKK- $\alpha$  signaling induced TSLP production in epithelial cells of COPD patients. *Exp Mol Med*. 2018 ;50(10):131.
19. Di Stefano A, Caramori G, Barczyk A, et al. Innate immunity but not NLRP3 inflammasome activation correlates with severity of stable COPD. *Thorax*. 2014;69(6):516-24.



**Abbreviations**

COPD, chronic obstructive pulmonary disease; GRS, Genetic Risk Score; FEV<sub>1</sub>, Forced expiratory volume in 1 second; FVC, Forced vital capacity; SNP, Single Nucleotide Polymorphism; TSLP, thymic stromal lymphopoietin; ANOVA, analysis of variance.

	Tsukuba healthy volunteer (n=1471)	Tsukuba COPD Cohort (n=270)
Age, y	50.5 ± 9.6	74.6 ± 8.7
Female sex, n (%)	786 (53.4%)	24 (8.9%)
Pack years	9.00 ± 17.7	61.8 ± 33.6
GRS	56.1 ± 7.2	56.7 ± 6.39
Forced vital capacity (FVC), L	3.2 ± 0.78	2.65 ± 0.89
Predicted FVC, %	NA	85.3 ± 23.4
Forced vital capacity (VC), L	NA	2.97 ± 0.89
Predicted VC, %	93.8 ± 22.2	93.8 ± 22.2
Forced expiratory volume in 1 second (FEV <sub>1</sub> ), L	2.6 ± 0.66	1.28 ± 0.61
Predicted FEV <sub>1</sub> , %	91.2 ± 13.2	54.1 ± 22.5
FEV <sub>1</sub> /FVC, %	82.6 ± 6.1	47.8 ± 14.0
EOS, µl	NA	203 ± 178.5
Height, cm	163 ± 8.78	160 ± 7.32
Weight, kg	61.3 ± 11.3	56.4 ± 10.3
BMI	23.0 ± 3.11	21.8 ± 3.50

Table 1. Patient demographics and characteristics

Data are presented as means ± SDs unless otherwise indicated.

NA, Not available

Cluster	GRS	History of BA	Use of ICS	Age	VC	%VC	FVC	%FVC	FEV <sub>1</sub>	%FEV <sub>1</sub>	FEV <sub>1</sub> / FVC	Pack years	Log EOS
1	mean	54.2		68.7	3.7	108.4	3.5	106.1	2.1	82.3	59.6	46.6	2.0
	N	55	0	25	55	32	32	55	55	55	55	55	32
	SD	6.7		9.1	0.9	18.3	0.9	17.0	0.5	13.7	12.5	16.3	0.7
2	mean	57.1		65.7	2.9	86.5	2.6	78.4	1.1	40.2	42.7	57.3	2.1
	N	53	3	21	53	25	25	53	53	53	53	53	33
	SD	5.5		4.9	0.8	19.6	0.7	20.0	0.3	11.3	12.0	24.3	0.7
3	mean	54.3		80.5	2.6	87.1	2.2	75.0	0.9	41.0	41.4	44.2	1.8
	N	77	5	24	77	35	35	75	77	77	77	77	45
	SD	4.1		4.6	0.6	18.5	0.7	19.9	0.3	13.1	10.3	15.6	1.0
4	mean	66.2		78.5	2.8	94.9	2.7	90.5	1.3	59.7	48.5	53.0	2.2
	N	35	1	10	35	20	20	34	35	35	35	35	22
	SD	3.4		5.1	0.7	21.3	0.7	21.2	0.5	21.9	12.9	18.1	0.6
5	mean	56.4		78.6	2.8	91.1	2.5	82.1	1.2	54.3	49.8	116.4	2.0
	N	50	2	19	50	26	26	50	50	50	50	50	25
	SD	5.3		6.6	1.0	26.8	0.9	24.7	0.6	20.8	14.8	28.5	0.8
P value		1.0 ×10 <sup>-24</sup>		3.6 ×10 <sup>-38</sup>	8.3 ×10 <sup>-7</sup>	2.7 ×10 <sup>-4</sup>	7.2 ×10 <sup>-16</sup>	2.7 ×10 <sup>-15</sup>	1.2 ×10 <sup>-37</sup>	4.1 ×10 <sup>-39</sup>	1.5 ×10 <sup>-14</sup>	4.5 ×10 <sup>-55</sup>	0.011*

Table 2a. Characteristics of each cluster by cluster analysis of the COPD cohort

\* ANOVA adjusted for age and ICS dose

Cluster	GRS	History of BA	Use of ICS	Age	VC	%VC	FVC	%FVC	FEV <sub>1</sub>	%FEV <sub>1</sub>	FEV <sub>1</sub> / FVC	Pack years	Log EOS
1	mean	54.7		67.4	3.8	108.0	3.7	108.7	2.0	78.6	55.2	45.7	2.0
	N	40	0	19	40	25	25	40	40	40	40	40	28
	SD	7.5		8.6	0.9	19.3	1.0	17.0	0.5	12.6	11.7	15.7	0.8
2	mean	57.1		65.7	2.9	86.5	2.6	78.4	1.1	40.2	42.7	57.3	2.1
	N	53	2	21	53	25	25	53	53	53	53	53	33
	SD	5.5		4.9	0.8	19.6	0.7	20.0	0.3	11.3	12.0	24.3	0.7
3	mean	54.4		80.5	2.6	87.0	2.2	75.6	0.9	41.3	41.1	44.6	1.8
	N	76	5	24	76	35	35	74	76	76	76	76	45
	SD	4.0		4.6	0.6	18.5	0.6	19.1	0.3	13.0	9.8	15.3	1.0
4	mean	66.3		78.3	2.7	92.5	2.6	90.1	1.2	56.5	45.6	54.2	2.2
	N	31	1	8	31	18	18	30	31	31	31	31	21
	SD	3.4		4.8	0.6	21.1	0.6	20.8	0.4	20.4	10.5	18.2	0.6
5	mean	56.9		79.5	2.9	93.1	2.5	82.7	1.2	53.3	47.1	114.0	2.0
	N	46	2	18	46	25	25	46	46	46	46	46	25
	SD	5.3		6.1	1.0	25.4	0.9	24.1	0.5	20.0	12.0	26.3	0.8
P value		4.5 ×10 <sup>-21</sup>		4.0 ×10 <sup>-42</sup>	5.5 ×10 <sup>-7</sup>	0.002	1.9 ×10 <sup>-16</sup>	3.5 ×10 <sup>-14</sup>	1.2 ×10 <sup>-37</sup>	3.7 ×10 <sup>-30</sup>	9.0 ×10 <sup>-9</sup>	4.5 ×10 <sup>-55</sup>	0.011*

Table 2b. Characteristics of each cluster by cluster analysis of the COPD cohort (n=246, Exclude patients with FEV1 / FVC above lower limit of normal)

\* ANOVA adjusted for age and ICS dose

Cluster	rs2289276 genotype/rs3806933 genotype			P value rs2289276/rs3806933	Odds Ratio (95%CI) rs2289276/rs3806933
	CC	CT	TT		
Cluster 1	31/31	19/17	5/7	0.783/0.975	1.06(0.69-1.62)/1.01(0.66-1.53)
Cluster 2	35/33	16/17	2/3	0.146/0.161	0.70(0.43-1.13)/0.72(0.45-1.14)
Cluster 3	38/37	35/34	4/6	0.453/0.628	1.15(0.80-1.64)/1.09(0.77-1.54)
Cluster 4	19/19	9/8	7/8	0.152/0.261	1.44(0.87-2.38)/1.33(0.81-2.17)
Cluster 5	20/19	24/24	6/7	0.017/0.034	1.65(1.09-2.50)/1.55(1.03-2.32)
Healthy Volunteer	828/767	545/583	98/121		

Table 3a. Multinomial logistic regression between the *TSLP* SNP and the COPD clusters.

Results of chi-square test between COPD clusters. rs2289276 P=0.035, rs3806933 P=0.047.

To examine the association of a *TSLP* genotype with each of the 5 COPD clusters in comparison with healthy volunteers (n = 1471), multinomial logistic regression was modelled.

Cluster	rs2289276 genotype/rs3806933 genotype			P value rs2289276/rs3806933	Odds Ratio (95%CI) rs2289276/rs3806933
	CC	CT	TT		
Cluster 1	24/24	11/10	5/6	0.830/0.915	1.06(0.64-1.74)/0.97(0.60-1.59)
Cluster 2	35/33	16/17	2/3	0.147/0.162	0.70(0.43-1.14)/0.72(0.45-1.14)
Cluster 3	37/36	35/34	4/6	0.398/0.566	1.17(0.82-1.67)/1.11(0.78-1.57)
Cluster 4	17/17	8/7	6/7	0.213/0.321	1.40(0.82-2.39)/1.30(0.77-2.2)
Cluster 5	19/18	21/21	6/7	0.024/0.041	1.64(1.07-2.52)/1.55(1.02-2.36)
Healthy Volunteer	828/767	545/583	98/121		

Table 3b. Multinomial logistic regression between the *TSLP* SNP and the COPD clusters. (n=246, Exclude patients with FEV1 / FVC above lower limit of normal)

Results of chi-square test between COPD clusters. rs2289276 P=0.038, rs3806933 P=0.045

To examine the association of a *TSLP* genotype with each of the 5 COPD clusters in comparison with healthy volunteers (n = 1471), multinomial logistic regression was modelled.

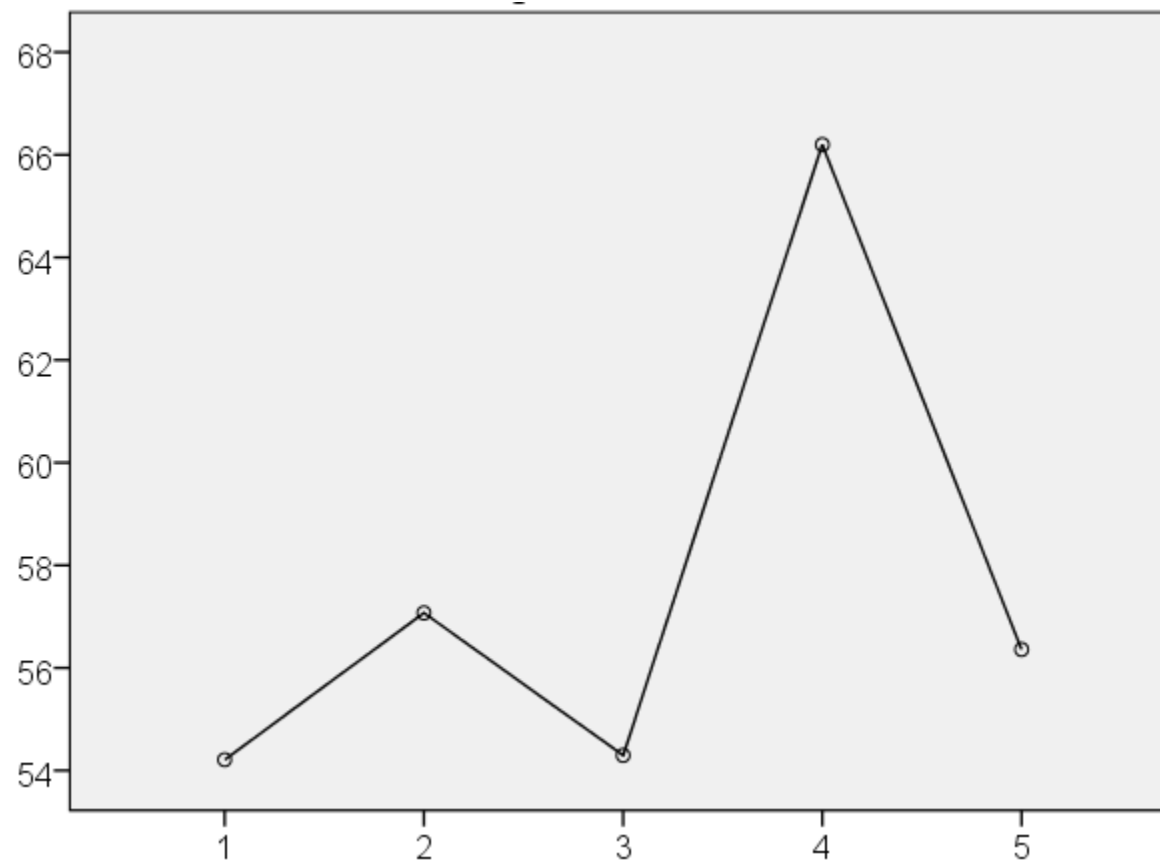


Figure 1. GRS of each cluster

Cluster 4 is characterized by the highest GRS.  $P = 1.04 \times 10^{-24}$