Chronic Obstructive Pulmonary Diseases:
Journal of the COPD Foundation

Original Research

Subtyping Chronic Obstructive Pulmonary Disease Using Peripheral Blood Proteomics

Sara Zarei, PhD[1,2] Ali Mirtar, PhD[3] Jarrett D. Morrow, PhD[1] Peter J. Castaldi, MD[1] Paula Belloni, PhD[4] Craig P. Hersh, MD, MPH[1]

Abstract

Chronic obstructive pulmonary disease (COPD) is a heterogeneous disorder. COPD patients may have different clinical features, imaging characteristics and natural history. Multiple studies have investigated heterogeneity using statistical methods such as unsupervised clustering to define different subgroups of COPD based largely on clinical phenotypes. Some studies have performed clustering using genetic data or limited numbers of blood biomarkers. Our primary goal was to use proteomic data to find subtypes of COPD within clinically similar individuals. In the Treatment of Emphysema with a gamma-Selective Retinoid Agonist (TESRA) study, multiplex biomarker panels were run in serum samples collected prior to randomization. After implementing an algorithm to minimize missing values, the dataset included 396 COPD individuals and 87 biomarkers. Using hierarchical clustering, we identified 3 COPD subgroups, containing 267 (67.4%), 104 (26.3%), and 25 (6.3%) individuals, respectively. The third cluster had less emphysema on quantitative analysis of chest computed tomography scans (p=0.03) and worse disease-related quality of life based on the St. George's Respiratory Questionnaire (total score cluster 1: 45.6, cluster 2: 45.4, cluster 3: 56.6; p=0.01), despite similar levels of lung function impairment (forced expiratory volume in 1 second (49.2%, 49.2%, 48.2% predicted, respectively). Enrichment analysis showed the biomarkers distinguishing cluster 3 mapped to platelet alpha granule and cell chemotaxis pathways. Thus, we identified a subgroup which has less emphysema but may have greater inflammation, which could be potentially targeted with anti-inflammatory therapies.

Abbreviations: chronic obstructive pulmonary disease, **COPD**; Treatment of Emphysema with a gamma-Selective Retinoid Agonist trial, **TESRA**; computed tomography, **CT**; Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints, **ECLIPSE**; forced volume in 1 second, **FEV_1**; forced vital capacity, **FVC**; diffusing capacity of the lungs for carbon monoxide, **DL_{CO}**; St. George's Respiratory Questionnaire, **SGRQ**; 15th percentile of lung density, **Perc 15**; Hounsfield units, **HU**; ethylenediamine-tetraacetic acid, **EDTA**; analysis of variance, **ANOVA**; Database for Annotation, Visualization and Integrated Discovery, **DAVID**; matrix metalloproteinase 9, **MMP9**; transforming growth factor beta, **TGF-β**; C-reactive protein, **CRP**; pulmonary and activation-regulated chemokine, **PARC**; chemokine ligand 18, **CCL18**; interleukin-18, **IL-18**; brain-derived neurotrophic factor, **BDNF**; platelet-derived growth factor, **PDGF**; epidermal growth factor, **EGF**; vascular endothelial growth factor, **VEGF**; CC chemokine ligand 16, **CCL16**; interleukin-1, **IL-1**; interleukin-8, **IL-8**; interleukin-10, **IL-10**; Body mass index-airflow Obstruction-Dyspnea-Exercise capacity index, **BODE**; modified Medical Research Council dyspnea index, **MMRC**

Funding Support: Supported by National Institutes of Health grants R01HL094635, P01HL105339, R01HL125583, R01HL130512. The TESRA trial was supported by Roche.
Date of Acceptance: November 1, 2016
Citation: Zarei S, Mirtar A, Morrow JD, Castaldi PJ, Belloni P, Hersh CP. Subtyping chronic obstructive pulmonary disease using peripheral blood proteomics. *Chronic Obstr Pulm Dis.* 2017;4(2):97-108. doi: https://doi.org/10.15326/jcopdf.4.2.2016.0147

This article has an online data supplement.

1 Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

2 San Juan Bautista School of Medicine, Caguas, Puerto Rico

3 University of California, San Diego

4 Genentech, San Francisco, California

Address correspondence to:

Craig P. Hersh, MD, MPH
Channing Division of Network Medicine
Brigham and Women's Hospital
181 Longwood Ave
Boston, MA 02115
Phone: 617-525-0729
Email: craig.hersh@channing.harvard.edu

Keywords:

biomarkers, cluster analysis, emphysema, inflammation

Introduction

Chronic obstructive pulmonary disease (COPD) is a common, progressive disease defined by airflow limitation on lung function tests. COPD is a heterogeneous condition, characterized by varying symptoms, natural history, and anatomic processes, which can be visualized on chest computed tomography (CT) scans.[1] These differences may be attributable to the molecular heterogeneity of the disease.[2] Starting almost 40 years ago, researchers have attempted to define the spectrum of heterogeneity in COPD.[3] However the underlying pathogenesis for the heterogeneity of the disease is still under investigation. Hence, identifying COPD phenotypes with the goal of achieving individualized treatment remains an important goal.[4] To date, studies have identified several clinical subtypes of COPD as well as a genetic subtype, alpha-1antitrypsin deficiency, that can be targeted with specific treatments.[5-7]

Statistical techniques such as cluster analysis have been used to assign COPD individuals to different groups where individuals within each cluster have more common characteristics compared to individuals from other groups or clusters.[4] These techniques have relied upon clinical and physiologic variables, chest CT scans, and gene expression data.[4,8,9] A recent COPD subtyping analysis in the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study used a limited set of blood biomarkers, including club cell secretory protein-16, surfactant protein-D, interleukin-8, tumor necrosis factor, fibrinogen and white blood cell counts.[10] The latter 2 biomarkers may be the most clinically applicable.[10] However, none of the previous studies have used large-scale proteomics to find different subgroups of COPD. In this study, we analyzed a large set of peripheral blood biomarkers in former smokers with emphysema from a clinical trial. We hypothesized that unbiased mathematical tools such as hierarchical clustering will be able to define sets of biomarkers that delineate the clinical variation among COPD patients and provide insight into the potential pathogenic mechanisms behind the subgroups.

Methods

Study Participants

The Treatment of Emphysema with a Selective Retinoid Agonist (TESRA) trial was a multicenter randomized, placebo-controlled clinical trial of palovarotene in COPD (clinicaltrials.gov identifier NCT00413205).[11] Details on the design, methods and the collection of clinical data of the TESRA study have been previously described.[11,12] Study participants provided written informed consent, and the study was approved by the institutional review boards at all participating centers. The study enrolled former smokers (abstinent at least 12 months) with a minimum of 10 pack years smoking history. Study participants had moderate-to-severe COPD, defined by post-bronchodilator forced expiratory volume in 1 second (FEV_1) to forced vital capacity (FVC) ratio <0.7 and FEV_1 <70% predicted, with diffusing capacity of the lung for carbon dioxide (DLCO) <70% predicted. All participants had emphysema based on visual review of chest CT scans. Baseline measurements also included assessments of symptoms and exacerbations, a 6-minute walk test, and body plethysmography to measure lung volumes. Disease-related quality of life was assessed using the St. George's Respiratory Questionnaire (SGRQ).[13] Chronic bronchitis was defined using the SGRQ per Kim et al.[14] Exacerbations were identified by hospitalization or treatment with oral steroids or antibiotics. Emphysema was quantified by the 15th percentile (Perc15) of lung density histogram.[15] Since Perc15 values are negative, the variable was converted to a positive value before log-transformation i.e., Perc15-transformed = $\log_{10}$ (1000 + Perc15). Therefore, higher values of the transformed variable correspond to lower quantitative emphysema. The percent of voxels with attenuation < -910 Hounsfield units (HU) was also used to assess emphysema.

Biomarker Measurements

Prior to randomization, blood samples were collected from 458 participants for biomarker analysis. These biomarkers were selected based on presumed biologic

mechanisms in COPD.[12] To measure the biomarker levels, a custom 15-panel assay was used. Concentrations of 140 protein biomarkers were measured in ethylenediamine-tetraacetic acid (EDTA) plasma in duplicate at Rules Based Medicine (Austin, Texas) and Quest Diagnostics (Valencia, California). Full details of the biomarker testing in TESRA, including the list of biomarkers, have been previously published.[12,16] For this analysis, biomarker measurements below the lower limit of quantification were set to missing. Otherwise, untransformed biomarker values were used.

Cluster Analysis

We implemented a simple optimization (maximization) algorithm to include the maximum number of participants and biomarkers, while minimizing missing data. Details of the algorithm can be found in the online supplement, Supplemental Figure 1. The maximum amount of data without missing values is found with 87 biomarkers and 396 participants. We performed an agglomerative McQuitty hierarchical clustering using only the biomarker dataset with a Canberra distance method to identify participant clusters on the basis of their individual biomarker profiles.[17] Canberra distance is suitable for biomarker data which contains large values, outliers and non-normally distributed variables.[18] To evaluate the quality of clusters and to find the optimal number of clusters, the R package NbClust was used.[19] NbClust provides multiple indices which determine the optimal number of clusters in a dataset. We used analysis of variance (ANOVA) to compare the mean values of phenotypes and biomarkers across different clusters. When the ANOVA *p*-value was significant, Tukey's test was used for pairwise comparisons between the groups.

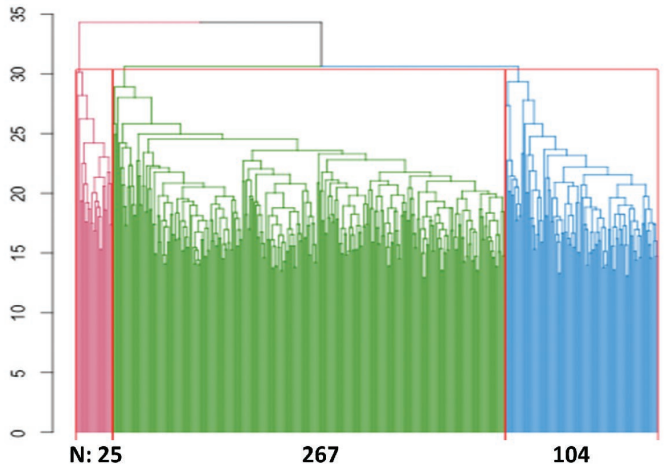
Enrichment Analysis

For the biomarkers that had different mean values among subgroups of COPD, we utilized enrichment analysis to find the corresponding molecular pathways for these biomarkers using Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7[20] and Reactome gene enrichment tools.[21] The biomarker proteins were mapped to their corresponding genes. Gene lists were input into GeneMANIA to create graphical networks.[22] GeneMANIA draws edges between the query genes and extends the set of genes based on databases of genetic interactions, molecular pathways, co-expression, co-localization, physical interactions and shared protein domains.

Results

Based on the majority (12 of 23) of the metrics implemented in the NbClust package,[19] we determined the optimal number of clusters to be 3 (online supplement, Supplemental Table 1). Therefore, we used hierarchical clustering of the biomarker measurements to divide the COPD participants into 3 subgroups, containing 267 (67.4%), 104 (26.3%) and 25 (6.3%) participants (Figure 1, Supplemental Figure 2). The characteristics of participants in each cluster are listed in Table 1. According to the ANOVA, the mean values for emphysema, measured by the 15th percentile of lung density histogram (transformed, see Methods), as well as the total score on the SGRQ were different between the 3 groups. Participants in the third cluster had higher SGRQ scores, consistent with a lower quality of life. However, they had less emphysema on chest CT scans (higher transformed Perc15 values). Cluster 3 showed a trend for the lowest emphysema, based on untransformed Perc15 and -910HU threshold, but the difference between clusters was not statistically significant. SGRQ scores and emphysema were not significantly different between clusters 1 and 2. Table 2 lists the 18 biomarkers that are significantly different between the 3 clusters, based on ANOVA *p*-value<0.05 and *p*-values<0.05 for all pairwise Tukey tests. Several of these proteins are known to be important in COPD

Figure 1. Hierarchical Clustering of COPD Participants



Hierarchical clustering of COPD participants based on serum biomarkers identified 3 participant clusters. Number of participants in each cluster is shown.

Table 1. Characteristics of Participants in Each Cluster Within the TESRA Study

Characteristic	Cluster 1		Cluster 2		Cluster 3		
	Mean or N	SD or %	Mean or N	SD or %	Mean or N	SD or %	*p*-value[a]
Number of Participants	267	67.4%	104	26.3%	25	6.3%	
Female Gender	90	33.7%	34	32.7%	7	28.0%	0.84
Age, years	66.63	8.51	67.77	7.06	65.40	8.66	0.32
Body Mass Index, kg/m^2	26.20	4.70	26.07	4.84	27.36	4.83	0.46
Pack Years of Smoking	45.79	25.12	48.76	25.19	57.19	29.26	0.08
Number of Exacerbations in the Past 12 Months	0.60	0.68	0.52	0.62	0.60	0.65	0.57
Diffusing Capacity (DL_{CO}) % predicted	48.84	13.24	48.24	12.25	49.60	14.44	0.87
FEV_1/FVC ratio	0.43	0.08	0.43	0.09	0.43	0.09	0.79
FEV_1 % predicted	49.21	9.22	49.18	8.76	48.22	9.61	0.87
Emphysema, 15th Percentile of Lung Density Histogram (transformed[b])	1.70	0.21	1.64	0.28	1.76	0.19	**0.03**
Emphysema, 15th percentile of Lung Density Histogram (untransformed)	-945	24	-949	24	-937	26	0.10
Emphysema % at -910HU	40.3	16.1	43.0	16.6	35.8	18.4	0.15
6-minute Walk Distance, m	320.80	106.95	327.60	99.81	281.80	104.54	0.15
BODE Index	3.56	1.53	3.65	1.35	4.08	1.82	0.25
mMRC Dyspnea Score	1.99	0.67	2.04	0.78	2.24	0.78	0.21
Total Lung Capacity % predicted	101.00	16.61	101.40	14.71	102.30	14.22	0.90
SGRQ Total Score	45.55	16.43	45.44	17.90	56.58	16.80	**0.007**
Chronic Bronchitis	62	23.2%	19	18.3%	6	24.0%	0.57

[a]Chi-square test for gender; ANOVA for quantitative outcomes.
[b]See Methods. Higher values indicate lower emphysema.
Mean and standard deviation or N and percentage are shown.
P-values <0.05 are in bold.
SD=standard deviation; FEV_1=forced expiratory volume in 1 second; FVC=forced vital capacity; HU=Hounsfield units;
BODE= Body mass index- airflow Obstruction-Dyspnea-Exercise capacity index; mMRC=modified Medical Research Council;
SGRQ=St. George's Respiratory Questionaire

pathogenesis, such as matrix metalloproteinase 9 (MMP9) and transforming growth factor beta (TGF-β).[23,24] Many of the biomarkers remained significant after Bonferroni correction for multiple testing (*p*-value <0.05/87=5.7e-4). The biomarkers in Table 2 were only weakly correlated with pack years of smoking (absolute value of Pearson r ≤0.2).

Enrichment analysis using DAVID, GeneMANIA and Reactome was performed using the list of genes coding for the biomarkers in Table 2. Figure 2 illustrates the gene and pathway enrichment analysis methods used, and the resulting top enriched pathways and the relevant genes. The Table 2 biomarkers were enriched for platelet-related pathways.

Supplemental Table 2 displays the result of DAVID enrichment analysis for the genes coding for the biomarkers listed in Table 2, limited to results with false discovery rate <0.05. Pathways are presented in annotation clusters with similar mechanisms and similar genes. The second annotation cluster included platelet granule genes, which are listed in Figure 2. Similar results were seen with GeneMANIA and Reactome. The GeneMANIA network of genes involved in platelet alpha granule pathway is displayed in Figure 3. The identified biomarker genes have multiple network connections in this pathway.

Table 2. Biomarkers Different Between All Three Clusters

Biomarker	ANOVA *P*-value	C1-C2 Tukey's *P*-value	C2-C3 Tukey's *P*-value	C3-C1 Tukey's *P*-value	Gene(s)
Brain-derived Neurotropic Factor[a]	1.8E-45	<2.2E-16	<2.2E-16	4.2E-07	BDNF
Cluster of Differentiation 40[a]	7.3E-19	9.7E-05	<2.2E-16	7.7E-10	CD40
CD40 Ligand[a]	6.6E-51	<2.2E-16	<2.2E-16	<2.2E-16	CD154
Epidermal Growth Factor[a]	2.9E-50	<2.2E-16	<2.2E-16	1.1E-09	EGF
Eotaxin 1	1.6E-10	1.9E-04	5.3E-09	7.5E-04	CCL11
Fibrinogen	7.0E-13	4.1E-02	4.6E-11	<2.2E-16	FGA, FGB, FGG
Growth-related Oncogene Alpha[a]	1.6E-17	2.7E-07	<2.2E-16	4.4E-06	CXCL1
Interleukin-18	1.0E-06	4.5E-02	1.7E-06	1.3E-03	IL18
Matrix Metalloproteinase 10	4.8E-11	2.5E-03	1.9E-10	2.0E-05	MMP10
Matrix Metalloproteinase 9	1.7E-12	1.7E-03	<2.2E-16	2.5E-06	MMP9
Matrix Metalloproteinase 9, total	1.6E-08	1.7E-02	3.3E-08	1.8E-04	MMP9
Myeloperoxidase	1.8E-15	2.5E-02	<2.2E-16	3.1E-10	MPO
Osteoprotegerin	5.8E-10	4.0E-05	8.2E-08	7.1E-03	TNFRSF11B
Platelet-derived Growth Factor[a]	4.3E-48	<2.2E-16	<2.2E-16	3.9E-08	PDGFB
Stem Cell Factor	1.1E-23	<2.2E-16	<2.2E-16	1.4E-02	KITLG
Sortilin[a]	1.5E-43	<2.2E-16	<2.2E-16	2.8E-10	SORT1
Transforming Growth factor Beta[a]	9.1E-57	<2.2E-16	<2.2E-16	6.2E-09	TGFB1
Vascular Endothelial Growth Factor[a]	1.8E-30	1.5E-11	<2.2E-16	<2.2E-16	VEGFA

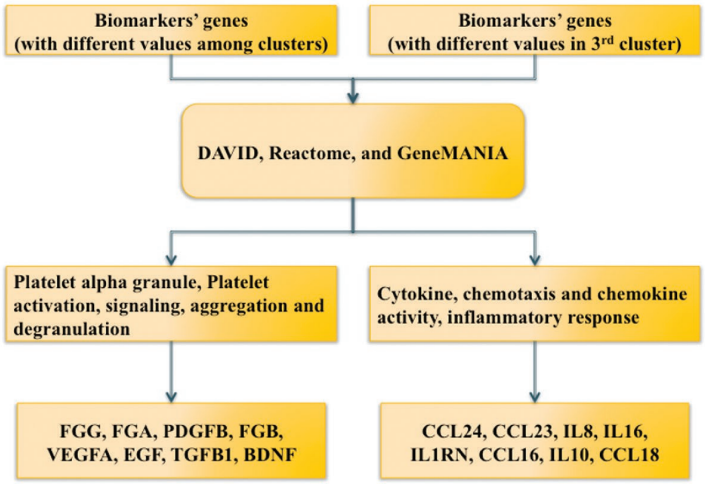
[a]Biomarkers significant based on Bonferroni correction for multiple testing, with $p<5.7E-4$ for ANOVA and for Tukey test for all 3 comparisons between clusters.
Shown are biomarkers with ANOVA $p<0.05$ along with p-values<0.05 for Tukey test for all 3 comparisons between clusters.

Based on the phenotype differences in cluster 3 seen in Table 1, we focused on biomarkers that showed different values in the participants from the third cluster compared with each of the first 2 clusters; these 21 biomarkers are listed in Table 3. Several of these biomarkers have been linked to COPD in previous studies. C-reactive protein (CRP), pulmonary and activation-regulated chemokine (PARC)/chemokine ligand 18 (CCL18), and interleukin (IL)-18 are known inflammatory biomarkers relevant to COPD[25-27]; all had higher levels in cluster 3. Alpha-1 antitrypsin levels were higher in cluster 3, consistent with the lower quantitative emphysema values in these participants. Similar to Table 2, the biomarkers in Table 3 were weakly correlated with pack years (absolute value of Pearson r ≤0.2).

Supplemental Table 3 shows the DAVID enrichment analysis for the genes encoding the biomarkers in Table 3. The top 2 annotation clusters relate to cytokines and chemotaxis, which was confirmed in GeneMANIA and Reactome. The GeneMANIA network analysis of these corresponding genes involved in cytokine/chemotaxis activities is displayed in Figure 4.

Supplemental Tables 4 and 5 show the mean levels by cluster for those biomarkers in the platelet alpha granule and cytokine/chemotaxis annotations, respectively. Participants from the first cluster have the lowest values of alpha platelet granule biomarkers while participants from the third cluster have the highest values. Participants from the third cluster have the highest values for the chemotaxis-related biomarkers.

Figure 2. Gene and Pathway Enrichment Analysis



Genes from Tables 2 and 3 were analyzed using DAVID, GeneMANIA and Reactome. The top enriched pathways along with the genes from those pathways are shown.

Discussion

In a study of former smokers with moderate-to-severe COPD with emphysema on chest CT scans, we were able to identify 3 subgroups based on cluster analysis of blood proteomics data. We specifically defined a small subgroup of participants with increased inflammatory biomarkers, yet less emphysema on quantitative analysis of chest CT scans, based on the 15th percentile of lung density histogram, which may be a better measure of emphysema in COPD than the -910HU threshold.[28] These participants had similar reductions in lung function, suggesting that airway disease may be present as well. This subgroup had reduced disease-related quality of life, exceeding the minimum clinically important difference of 4 points in total SGRQ score.[29] A previous study had used a panel of blood and sputum biomarkers to define 4 biologic clusters of acute exacerbations,[30] but our study is the first to use proteomics to subtype stable COPD patients.

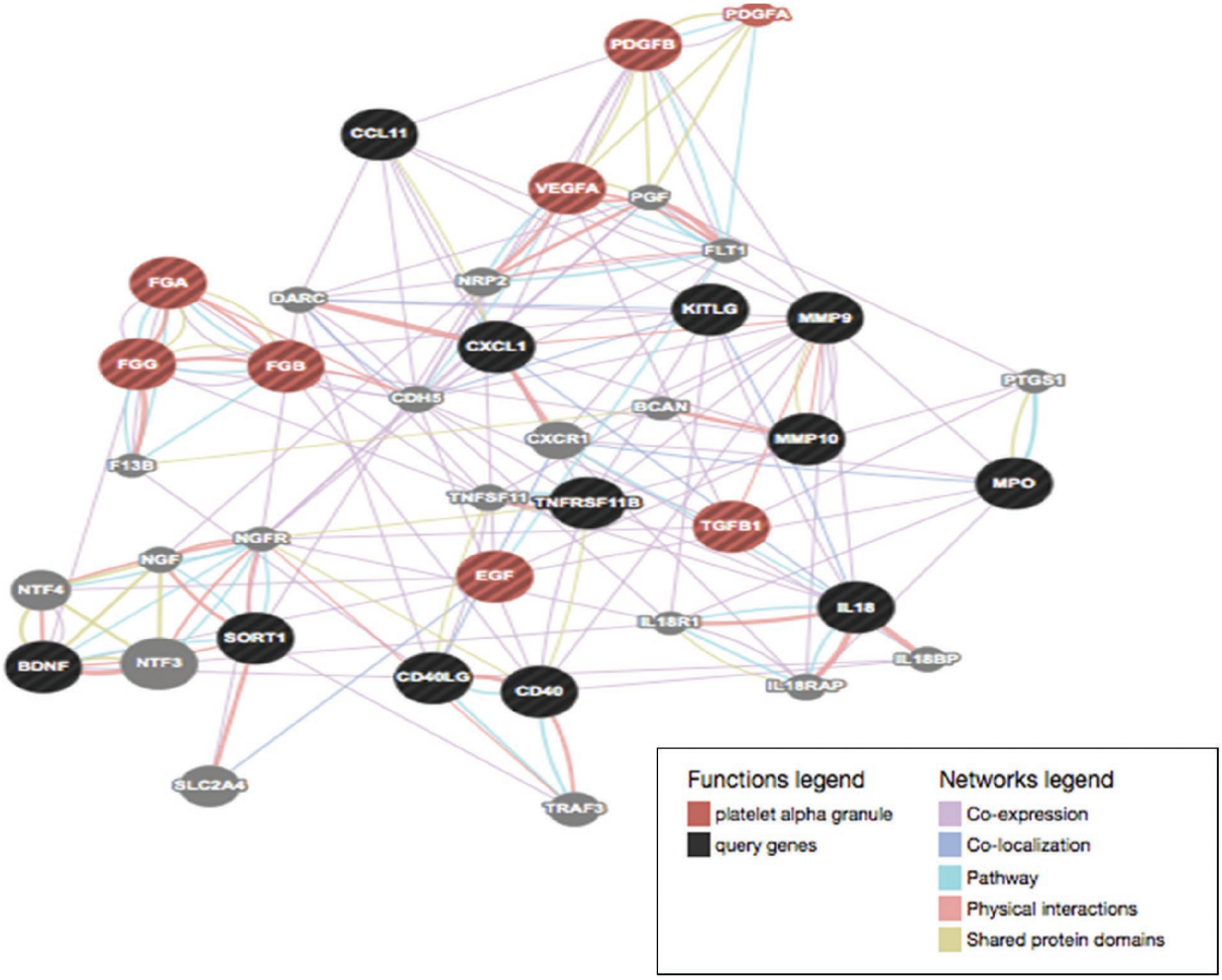
Gene enrichment analysis showed that biomarkers annotated to the platelet alpha granule pathway were different between the 3 subgroups of COPD. Platelets contain different storage granules including alpha granules, dense granules and lysosomes. Alpha granules, the main storage granules, contain fibrinogen, von Willebrand factor, growth factors and protease inhibitors that enhance thrombin formation at the site of injury.[31] Several of the platelet alpha granule biomarkers have previously been shown to be associated with COPD and related phenotypes. Systemic inflammatory markers such as fibrinogen are associated with COPD risk, mortality and exacerbations.[32,33] Brain-derived neurotrophic factor (BDNF) is stored in platelets and is released during an inflammatory response.[34] BDNF has been shown to be elevated in COPD individuals when compared to controls.[34,35] Growth factors, including platelet-derived growth factor (PDGF), epidermal growth factor (EGF), and TGF-β, are mitogens for smooth muscle cells and fibroblasts. PDGF may be involved in vascular and small airway remodeling in COPD.[36,37] EGF and TGF-β are both upregulated in airway epithelium and submucosal cells of patients with COPD.[38,39] The role of these growth factors, especially TGF-β, in airway remodeling in COPD has been well-documented.[23,40-42]

The important pro-angiogenic, regulatory protein, vascular endothelial growth factor (VEGF), is also housed in alpha granules. VEGF plays a role in maintaining the homeostasis of alveoli; decrease in VEGF expression is associated with pulmonary endothelial cell apoptosis.[43,44] In our study, participants from the third cluster have the highest mean values for the platelet alpha granule biomarkers, including VEGF, and the lowest amount of emphysema on chest CT scans.

In the analysis focusing on differences between the third cluster and the other 2 clusters, gene enrichment pointed to genes related to cell chemotaxis and inflammatory response. The corresponding biomarkers included PARC, CC chemokine ligand 16 (CCL16), interleukin-1 (IL-1) receptor antagonist, interleukin-8 (IL-8) and interleukin-10 (IL-10). PARC/CCL-18 is primarily synthesized in dendritic cells and monocytes and is highly expressed in the lungs.[45,46] Serum PARC/CCL-18 is elevated in COPD and is associated with mortality.[25]

Balanced secretion of pro- and anti-inflammatory cytokines is essential in limiting pulmonary inflammation in the stable state and during respiratory infections. IL-10, an anti-inflammatory cytokine, was shown to be decreased in serum and sputum of COPD participants and healthy smokers compared to non-smokers.[47] On the other hand, pro-inflammatory cytokines such as IL-8, secreted by alveolar macrophages, were found to be elevated in COPD patients and were further increased during exacerbations.[40] On average, participants from the third cluster had the highest mean

Figure 3. Platelet Alpha Granule Pathway Gene Network Resulting From GeneMania



The black nodes (query genes) represent the genes for the biomarkers that were significantly different across all 3 clusters. The red nodes are query genes involved in the platelet alpha granule pathway. The edges are color coded as described in the figure.

values for both pro- and anti-inflammatory biomarkers.

Our study has several limitations. The biomarker panel used was selected from available assays, based on possible mechanisms in COPD. The panel was heavily weighted towards inflammatory markers, which may explain why these pathways were prominent in our results. Unbiased proteomic assays would be required to identify novel pathways in COPD subgroups. Based on enrollment criteria in the TESRA trial, which included moderate-to-severe COPD with emphysema, study participants tended to be more homogeneous than the general COPD patient population, which may limit the ability to identify subgroups (as demonstrated in Supplemental Figure 2) and may also limit generalizability. Despite these limitations, we were able to identify 3 subgroups of COPD participants using clustering and network analysis of a large panel of serum biomarkers. We found participants in the smallest subgroup to have the highest levels of platelet alpha granule biomarkers and inflammatory cytokines. These participants had less emphysema and a lower quality of life, despite similar levels of lung function

Table 3. Biomarkers Specific to the Third Cluster

Biomarker	ANOVA *P*-value	C1-C2 Tukey's *P*-value	C2-C3 Tukey's *P*-value	C3-C1 Tukey's *P*-value	Gene(s)
Alpha-1 Antitrypsin[a]	8.8E-05	6.0E-01	1.7E-04	5.9E-05	SERPINA1
Apolipoprotein C-III	1.2E-03	5.9E-01	2.4E-03	8.2E-04	APOC3
Chromogranin A	3.9E-02	9.1E-01	4.2E-02	3.5E-02	CGA
Creatine Kinase-MB	1.7E-02	8.4E-01	2.0E-02	1.4E-02	CKM
C-reactive Protein[a]	1.7E-12	6.5E-01	<2.2E-16	<2.2E-16	CRP
Eotaxin 2	1.4E-03	9.3E-01	1.4E-03	1.5E-03	CCL24
Fatty-acid-binding Protein 1	2.7E-02	9.8E-01	2.5E-02	2.8E-02	FABP1
Hemofiltrate CC-Chemokine-4[a]	5.8E-07	2.2E-01	4.6E-06	2.7E-07	CCL16
Hepatocyte Growth Factor[a]	7.4E-06	7.3E-01	3.5E-06	8.2E-05	HGF
Haptoglobin[a]	7.5E-07	8.2E-01	8.9E-07	9.2E-07	HP
Interleukin-10	3.8E-03	6.6E-01	6.4E-03	2.7E-03	IL10
Interleukin-16	1.3E-10	7.8E-01	3.2E-11	5.3E-09	IL16
Interleukin-1 Receptor Antagonist[a]	2.1E-07	9.9E-01	1.2E-07	9.3E-07	IL1RN
Interleukin-2 Receptor Antagonist[a]	4.8E-10	5.2E-02	3.8E-10	2.8E-06	IL2RA
Interleukin-8[a]	3.9E-12	1.8E-01	<2.2E-16	7.6E-09	IL8
Myeloid Progenitor Inhibitory Factor-1	7.0E-04	2.9E-01	3.4E-03	4.1E-04	CCL23
Prostatic Acid Phosphatase	2.0E-04	9.0E-01	1.1E-04	7.7E-04	ACPP
Pulmonary and Activation-Regulated Chemokine[a]	7.0E-10	6.9E-01	1.2E-09	9.8E-10	CCL18
Tenascin C[a]	1.1E-04	1.0E+00	7.0E-05	2.3E-04	TNC
Tumor Necrosis Factor Receptor 1[a]	6.6E-08	9.8E-01	3.4E-08	4.1E-07	TNFRSF1A, TNFRSF1B
Tumor Necrosis Factor-related Apoptosis-inducing Ligand Receptor 3	3.3E-02	9.5E-01	3.1E-02	3.1E-02	TNFRSF10C

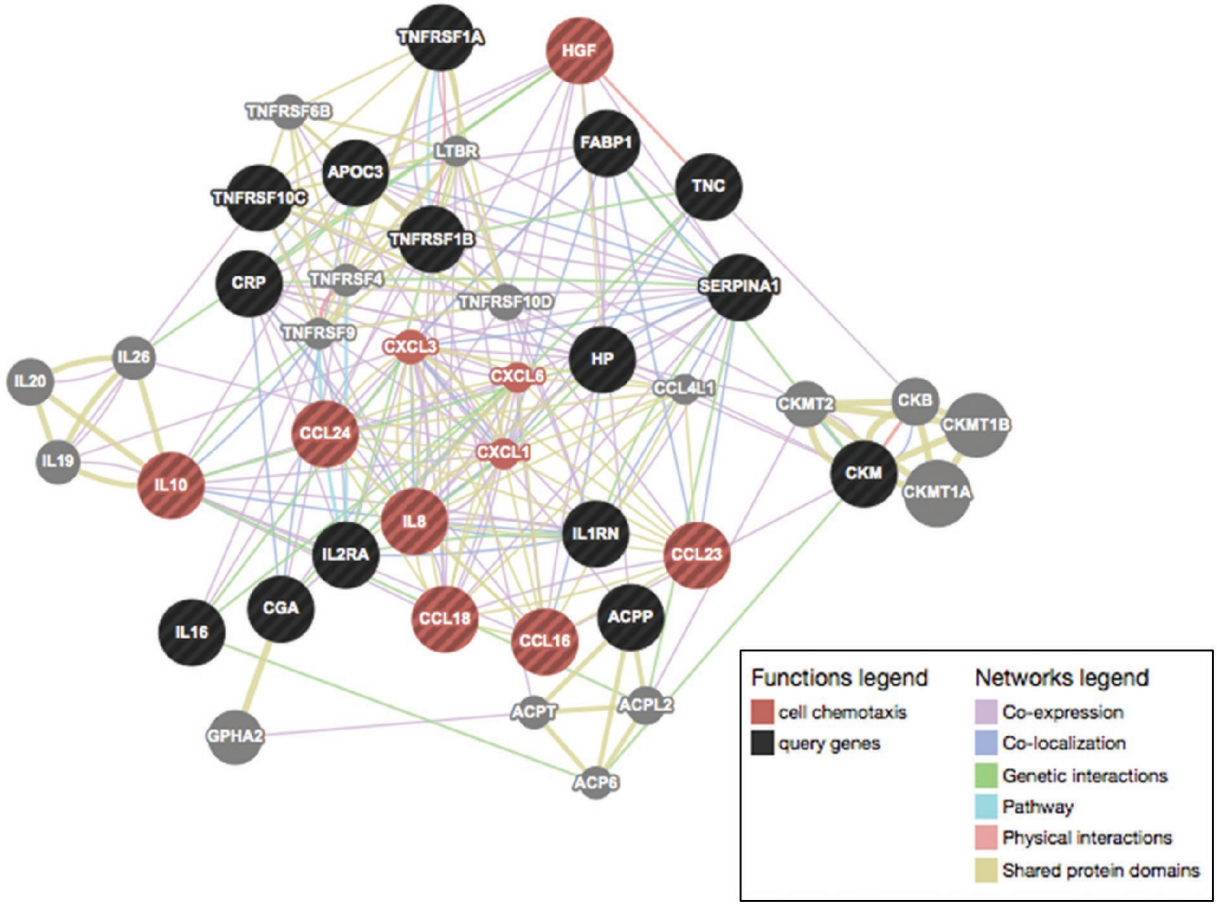
Listed are the biomarkers with ANOVA $p<0.05$ and Tukey test $p<0.05$ for comparisons between clusters 1 and 3 and clusters 2 and 3.
[a]Biomarkers significant based on Bonferroni correction for multiple testing, with $p<5.7E-4$ for ANOVA and for Tukey test for comparisons between clusters 1 and 3 and clusters 2 and 3.

impairment. Thus, these individuals may have more airway inflammation compared to the other groups. Future studies measuring similar biomarkers in a broader range of COPD participants will be required to validate and to expand upon these results. The ultimate goal is to use serum biomarkers to define clinically-relevant COPD subgroups which may have different outcomes and could potentially be treated with different therapies.

Acknowledgements

TESRA Investigators: Ognian Georgiev, Dimitar Popov, and Vasil Dimitrov, Sofia, Bulgaria; Hristo Metev, Ruse, Bulgaria; Yavor Ivanov, St. Pleven, Bulgaria; Libor Fila and Jiri Votruba, Praha, Czech Republic; Vladimir Zindr Vitezna, Karlovy Vary, Czech Republic; Kamil Klenha, Tabor, Czech Republic; Jaromir Roubec, Ostrava, Czech Republic; Barna Szima, Szombathely, Hungary; Zsuzsanna Mark, Torokbalint, Hungary; Zoltan Baliko, Pecs Hungary; Zoltan Bartfai, Budapest, Hungary; Katalin Gomori, Balassagyarmat, Hungary; Andres Sigvaldason, Reykjavik, Iceland; Mordechai Kramer, Petach Tikva, Israel; Gershon Ya Fink, Rehovot, Israel; Zeev Weiler, Ashkelon, Israel; Joel Greif, Tel Aviv, Israel; Issahar Ben-Dov, Ramat Gan, Israel; Mordechai Yigla, Haifa, Israel; Leonardo Fabbri, Modena, Italy; Pierluigi Paggiaro, Pisa, Italy; Giorgio Canonica, Genova, Italy; Isa Cerveri, Pavia, Italy; Antra Bekere and Aurika Babjoniseva, Riga, Latvia; Alvil Krams and Stopinu Pagasts, Rigas Rajons, Latvia; Wladyslaw Pierzchala, Katowice, Poland; Dariusz Nowak, Lodz, Poland; Robert Mroz, Bialystok, Poland; Hanna Szelerska-Twardosz

Figure 4. Cell Chemotaxis Pathway Gene Network Resulting From GeneMania



The black nodes (query genes) represent the genes for the biomarkers that were significantly different in the third cluster. The red nodes are query genes involved in cell chemotaxis pathways. The edges are color coded as described in the figure.

and Malgorzata Rzymkowska, Poznan, Poland; Ismail Abdullah, Durban, South Africa; Christo Van Dyk, Western Cape, Worcester, South Africa; Nyda Fourie, Bloemfontein, South Africa; John O'Brien and Mary Batema, Cape Town, South Africa; J. Joubert, Bellville-Cape Province, South Africa; Abdool Gafar, Kwa-Zulu Natal, Amanzimtoti, South Africa; Hannes Van Rensburg, Centurion, South Africa; Lyudmila Yashina, Oleksandr Dzyublik, Volodymyr Gavrysyuk and Yuriy Feshchenko, Kiev, Ukraine; Nadezda Monogarova, Donetsk, Ukraine; David Parr, Coventry, United Kingdom; Stephen Rennard, Omaha, Nebraska; Richard Casaburi, Torrance, California; Gerard Criner, Philadelphia, Pennsylvania; Mark Dransfield, Birmingham, Alabama; Charles Fogarty, Spartanburg, South Carolina; Nicola Hanania and Amir Harafkhaneh Houston, Texas; Carl Griffin and Kathi Mcdavid, Oklahoma City, Oklahoma;, Paul Kvale, Detroit, Michigan; Barry Make, Denver, Colorado; Joe Ramsdell, San Diego, California; Michael D. Roth, Los Angeles, California; Peter Sporn, Chicago, Illionois.

Steering Committee: Alvar Agustı´ (Spain), Peter Calverley (United Kingdom), Leonardo Fabbri (Italy), Klaus F. Rabe (Netherlands), Nicolas Roche (France), Michael Roth (United States), Jorgen Vestbo (Denmark), Stephen Rennard (United States).

Declaration of of Interest

Dr. Hersh reports consulting fees from AstraZeneca, Concert Pharmaceuticals and Mylan. Dr. Belloni is an employee of Genentech. Drs. Zarei, Mirtar, Morrow, and Castaldi report no competing interests. Roche designed the TESRA trial and collected the data. The funders

had no role in the data analysis in this manuscript, the writing of the manuscript or the decision to submit the manuscript for publication.

References

1. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598-604. doi: https://doi.org/10.1164/rccm.200912-1843CC.

2. Agusti A, Calverley PM, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11:122. doi: https://doi.org/10.1186/1465-9921-11-122.

3. Burrows B, Fletcher CM, Heard BE, Jones NL, Wootliff JS. The emphysematous and bronchial types of chronic airways obstruction. *Lancet*. 1966;287:830-5. doi: https://doi.org/10.1016/S0140-6736(66)90181-184.

4. Pinto LM, Alghamdi M, Benedetti A, Zaihra T, Landry T, Bourbeau J. Derivation and validation of clinical phenotypes for COPD: a systematic review. *Respir Res*. 2015;16:50. doi: https://doi.org/10.1186/s12931-015-0208-4.

5. Chapman KR, Burdon JGW, Piitulainen E, et al. Intravenous augmentation treatment and lung density in severe α1 antitrypsin deficiency (RAPID): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;386:360-8. doi: https://doi.org/ 10.1016/S0140-6736(15)60860-1.

6. Calverley PM, Rabe KF, Goehring UM, Kristiansen S, Fabbri LM, Martinez FJ. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. *Lancet*. 2009;374(9691):685-694. doi: https://doi.org/10.1016/S0140-6736(09)61255-1.

7. Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med*. 2003;348(21):2059-2073. doi: https://doi.org/10.1056/NEJMoa030287.

8. Castaldi PJ, Dy J, Ross J, et al. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax*. 2014;69(5):415-422. doi: https://doi.org/10.1136/thoraxjnl-2013-203601.

9. Menche J, Sharma A, Cho MH, et al. A diVIsive Shuffling Approach (VIStA) for gene expression analysis to identify subtypes in chronic obstructive pulmonary disease. *BMC Syst Biol*. 2014;8 Suppl 2:S8. doi: https://doi.org/10.1186/1752-0509-8-S2-S8.

10. Rennard SI, Locantore N, Delafont B, et al. Identification of five chronic obstructive pulmonary disease subgroups with different prognoses in the ECLIPSE cohort using cluster analysis. *Ann Am Thorac Soc*. 2015;12(3):303-312. doi: https://doi.org/10.1513/AnnalsATS.201403-125OC.

11. Jones PW, Rames AD. TESRA (Treatment Of Emphysema With A Selective Retinoid Agonist) Study Results[abstract]. *Am J Respir Crit Care Med*. 2011;183:A6418.

12. Cheng DT, Kim DK, Cockayne DA, et al. Systemic soluble receptor for advanced glycation endproducts is a biomarker of emphysema and associated with AGER genetic variants in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;188(8):948-957. doi: https://doi.org/10.1164/rccm.201302-0247OC.

13. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis*. 1992;145(6):1321-1327.

14. Kim V, Crapo J, Zhao H, et al. Comparison between an alternative and the classic definition of chronic bronchitis in COPDGene. *Ann Am Thorac Soc*. 2015;12(3):332-339. doi: https://doi.org/10.1513/AnnalsATS.201411-518OC.

15. Dirksen A, Piitulainen E, Parr DG, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in alpha1-antitrypsin deficiency. *Eur Respir J*. 2009;33(6):1345-1353. doi: https://doi.org/10.1183/09031936.00159408.

16. Carolan BJ, Hughes G, Morrow J, et al. The association of plasma biomarkers with computed tomography-assessed emphysema phenotypes. *Respir Res*. 2014;15:127. doi: https://doi.org/10.1186/s12931-014-0127-9.

17. Emmert-Streib F, Abogunrin F, de Matos Simoes R, et al. Collectives of diagnostic biomarkers identify high-risk subpopulations of hematuria patients: exploiting heterogeneity in large-scale biomarker data. *BMC Med*. 2013;11:12. doi: https://doi.org/10.1186/1741-7015-11-12.

18. Quinn GP, Keogh MJ. Experimental design and data analysis for biologists. New York: Cambridge University Press; 2002.

19. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust : an R package for determining the relevant number of clusters in a data set. *J Stat Softw*. 2014;61(6):1-36. doi: https://doi.org/10.18637/jss.v061.i06.

20. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi: https://doi.org/10.1038/nprot.2008.211.

21. Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;42 (Database issue): D472-477. doi: https://doi.org/10.1093/nar/gkt1102.

22. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(Web Server issue):W214-220. doi: https://doi.org/10.1093/nar/gkq537.

23. Bartram U, Speer CP. The role of transforming growth factor beta in lung development and disease. *Chest*. 2004;125(2):754-765.

24. Houghton AM. Matrix metalloproteinases in destructive lung disease. *Matrix Biol*. 2015;44-46:167-174. doi: https://doi.org/10.1016/j.matbio.2015.02.002.

25. Sin DD, Miller BE, Duvoix A, et al. Serum PARC/CCL-18 concentrations and health outcomes in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2011;183(9):1187-1192. doi: https://doi.org/10.1164/rccm.201008-1220OC.

26. Dahl M, Vestbo J, Zacho J, Lange P, Tybjaerg-Hansen A, Nordestgaard BG. C reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach. *Thorax*. 2011;66(3):197-204. doi: https://doi.org/10.1136/thx.2009.131193.

27. Kang MJ, Choi JM, Kim BH, et al. IL-18 induces emphysema and airway and vascular remodeling via IFN-gamma, IL-17A, and IL-13. *Am J Respir Crit Care Med*. 2012;185(11):1205-1217. doi: https://doi.org/10.1164/rccm.201108-1545OC.

28. Lynch DA. Progress in imaging COPD, 2004-2014. *Chronic Obstr Pulm Dis (Miami)*. 2014;1(1):73-82. doi: http://dx.doi.org/10.15326/jcopdf.1.1.2014.0125.

29. Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD*. 2005;2(1):75-79.

30. Bafadhel M, McKenna S, Terry S, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. *Am J Respir Crit Care Med*. 2011;184(6):662-671. doi: https://doi.org/10.1164/rccm.201104-0597OC.

31. Blair P, Flaumenhaft R. Platelet alpha-granules: basic biology and clinical correlates. *Blood Rev*. 2009;23(4):177-189. doi: https://doi.org/10.1016/j.blre.2009.04.001.

32. Agusti A, Edwards LD, Rennard SI, et al. Persistent systemic inflammation is associated with poor clinical outcomes in COPD: A novel phenotype. *PLoS ONE*. 2012;7(5):e37483. doi: https://doi.org/10.1371/journal.pone.0037483.

33. Miller BE, Tal-Singer R, Rennard SI, et al. Plasma Fibrinogen Qualification as a Drug Development Tool in Chronic Obstructive Pulmonary Disease. Perspective of the Chronic Obstructive Pulmonary Disease Biomarker Qualification Consortium. *Am J Respir Crit Care Med*. 2016;193(6):607-613. doi: https://doi.org/10.1164/rccm.201509-1722PP.

34. Stoll P, Wuertemberger U, Bratke K, Zingler C, Virchow JC, Lommatzsch M. Stage-dependent association of BDNF and TGF-beta1 with lung function in stable COPD. *Respir Res*. 2012;13:116. doi: https://doi.org/10.1186/1465-9921-13-116.

35. Pinto-Plata V, Toso J, Lee K, et al. Profiling serum biomarkers in patients with COPD: associations with clinical parameters. *Thorax*. 2007;62(7):595-601. doi: https://doi.org/10.1136/thx.2006.064428.

36. Elwing J, Panos RJ. Pulmonary hypertension associated with COPD. *Int J Chron Obstruct Pulmon Dis*. 2008;3(1):55-70.

37. Churg A, Tai H, Coulthard T, Wang R, Wright JL. Cigarette smoke drives small airway remodeling by induction of growth factors in the airway wall. *Am J Respir Crit Care Med*. 2006;174(12):1327-1334. doi: https://doi.org/10.1164/rccm.200605-585OC.

38. Takizawa H, Tanaka M, Takami K, et al. Increased expression of transforming growth factor-beta1 in small airway epithelium from tobacco smokers and patients with chronic obstructive pulmonary disease (COPD). *Am J Respir Crit Care Med*. 2001;163(6):1476-1483. doi: https://doi.org/10.1164/ajrccm.163.6.9908135.

39. de Boer WI, Hau CM, van Schadewijk A, Stolk J, van Krieken JH, Hiemstra PS. Expression of epidermal growth factors and their receptors in the bronchial epithelium of subjects with chronic obstructive pulmonary disease. *Am J Clin Pathol*. 2006;125(2):184-192. doi: https://doi.org/10.1309/W1AX-KGT7-UA37-X257.

40. Chung KF. Cytokines in chronic obstructive pulmonary disease. *Eur Respir J Suppl*. 2001;34:50s-9s.

41. Ingram JL, Bonner JC. EGF and PDGF receptor tyrosine kinases as therapeutic targets for chronic lung diseases. *Curr Mol Med*. 2006;6(4):409-421.

42. Yang YC, Zhang N, Van Crombruggen K, Hu GH, Hong SL, Bachert C. Transforming growth factor-beta1 in inflammatory airway disease: a key for understanding inflammation and remodeling. *Allergy*. 2012;67(10):1193-1202. doi: https://doi.org/10.1111/j.1398-9995.2012.02880.x.

43. Kanazawa H. Role of vascular endothelial growth factor in the pathogenesis of chronic obstructive pulmonary disease. *Med Sci Monit*. 2007;13(11):RA189-95.

44. Tuder RM, Yun JH. Vascular endothelial growth factor of the lung: friend or foe. *Curr Opin Pharmacol*. 2008;8(3):255-260. doi: https://doi.org/10.1016/j.coph.2008.03.003.

45. Hieshima K, Imai T, Baba M, et al. A novel human CC chemokine PARC that is most homologous to macrophage-inflammatory protein-1 alpha/LD78 alpha and chemotactic for T lymphocytes, but not for monocytes. *J Immunol*. 1997;159(3):1140-1149.

46. Schraufstatter I, Takamori H, Sikora L, Sriramarao P, DiScipio RG. Eosinophils and monocytes produce pulmonary and activation-regulated chemokine, which activates cultured monocytes/macrophages. *Am J Physiol Lung Cell Mol Physiol*. 2004;286(3):L494-501. doi: https://doi.org/10.1152/ajplung.00323.2002.

47. Zhang L, Cheng Z, Liu W, Wu K. Expression of interleukin (IL)-10, IL-17A and IL-22 in serum and sputum of stable chronic obstructive pulmonary disease patients. *COPD*. 2013;10(4):459-465. doi: https://doi.org/10.3109/15412555.2013.770456.