Systemic Markers of Adaptive and Innate Immunity Are Associated with COPD Severity and Spirometric Disease Progression

Eitan Halper-Stromberg[1,2], Jeong H. Yun[3,4], Margaret M. Parker[3], Ruth Tal Singer[5], Amit Gaggar[6], Edwin K. Silverman[3,4], Sonia Leach[2], Russell P. Bowler[1,2]*, Peter J. Castaldi[3,7]*

[1]University of Colorado School of Medicine, Aurora, CO; [2]National Jewish Health, Denver, CO; [3]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; [4]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA; GSK, [5]King of Prussia, Pennsylvania; [6]Division of Division of Pulmonary, Allergy & Critical Care Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL; [7]Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA

*authors contributed equally

Corresponding Authors:

Peter Castaldi, MD, MSc

Channing Division of Network Medicine

Brigham and Women's Hospital

Boston, MA 02115

peter.castaldi@channing.harvard.edu

Russell P. Bowler, MD, PhD

National Jewish Health

Denver, CO

BowlerR@NJHealth.org

Word Count: 3,472

ABSTRACT

The progression of chronic obstructive pulmonary disease (COPD) is associated with marked alterations of circulating immune cell populations, but no large studies have comprehensively characterized alterations in circulating immune cells across the full spectrum of lung function impairment in current and former smokers. In 6,299 subjects from the COPDGene and ECLIPSE studies, we related Coulter blood counts and proportions to cross-sectional FEV_1 adjusting for current smoking status. We also related cell count measures to three-year change in FEV_1 in the ECLIPSE subjects. In a subset of subjects with genome-wide blood gene expression data, we used two cell type deconvolution methods to infer the proportions of immune cell subpopulations, and we related these proportions to COPD status. We observed that FEV_1 levels are positively correlated with lymphocytes and negatively correlated with neutrophils, monocytes, and eosinophils. In multivariate models, both cell counts and proportions were significantly associated with FEV_1. We also observed that lymphocytes, monocytes, and eosinophil counts were predictive of three year change in lung function. Using cell type deconvolution to study immune cell subpopulations, we observed that subjects with COPD had a lower proportion of CD4+ resting memory cells and naïve B cells compared to non-COPD smokers. Alterations in circulating immune cells in COPD support a mixed pattern of lymphocyte suppression and an enhanced innate immune response. Cell counts and proportions contribute independent information to models predicting lung function. Cell type deconvolution is a promising method for the simultaneous study of multiple immune cell subpopulations.

Abstract Word Count: 245

Keywords: COPD, cell type deconvolution, gene expression, lymphocytes

Background

Chronic obstructive pulmonary disease (COPD) is associated with profound alterations in immune cells within the lung and in the systemic circulation. The systemic inflammation may reflect "spill-over" of inflammatory processes within the lung, primary alterations in the extra-pulmonary immune response, or a combination of both processes(1). These alterations affect cell types involved in both the innate(2-6) and adaptive(7-15) immune response. In population studies(16, 17) and a systematic review(18), total counts of peripheral leucocytes were associated with cross-sectional and prospective changes in lung function, but few studies have been performed in large cohorts with detailed cell count data to observe relationships across the full spectrum of lung function. In addition, current smoking has an independent effect on immune cells(8, 19) and often serves as an important potential confounder of immunologic studies of current and former smokers with COPD.

Cell type quantification by flow cytometry is rarely available from large, population-based studies of COPD. However, novel, cell-type "deconvolution" approaches have been shown to infer accurately the relative proportions of immune cell types from genome-wide blood gene expression data(20, 21). Thus, cell-type deconvolution is a potentially powerful approach to enable the simultaneous study of many different cell types in large cohorts of subjects with available blood gene expression, but it has not yet been applied to cohorts of subjects with COPD.

We hypothesized that 1) peripheral immune cell types quantified through Coulter complete blood counts (CBC) have significant associations to cross-sectional FEV_1 and prospective FEV_1 decline; and that 2) cell-type deconvolution methods can enable the simultaneous study of multiple immune cell subpopulations in cohorts of smokers with COPD and blood gene expression data. We explored the first hypothesis in two large cohorts of smokers enriched for COPD, the COPDGene and ECLIPSE studies, which enabled the characterization of immune cell profiles across the full spectrum of lung function impairment while accounting for current smoking effects. The large number of study subjects allowed for detailed modeling of the relationship between

multiple cell types, current smoking status, and lung function. To explore the second hypothesis, we used two cell type deconvolution methods to infer immune cell subpopulation proportions in a subset of smokers with blood gene expression data in the ECLIPSE Study, and we validated these inferred cell type proportions against measured CBC data. We then compared levels of inferred circulating immune cell subpopulations by COPD status, confirming that inferred estimates of circulating immune cell types such as monocytes, naïve B cells, and resting T memory cells are altered in the COPD state.

Methods

Study Populations

Recruitment criteria and study protocols for the ECLIPSE and COPDGene studies have been previously reported. COPDGene enrolled 10,192 subjects across the entire GOLD spectrum between the ages of 45 and 80 with at least a 10 pack-year smoking history(22). These subjects completed their Phase 1 study visit between 2007-2011. As of September 24, 2016, 5000 subjects had completed their Phase 2 five-year follow-up visit which included all of the data items collected in Phase 1 as well as complete blood count data, which was not collected at the Phase 1 visit.

The Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study was a multicenter study that enrolled subjects aged 40-75 with COPD and at least a 10-year smoking history (COPD defined by FEV_1< 80% of predicted and FEV_1/FVC <= 0.7) or who were smokers without COPD (FEV_1>85% and FEV_1/FVC > 0.7). Details of this study have been previously published(23). Gene expression analyses were performed in a subset of subjects in this study from whom genome-wide gene expression data were generated on the Affymetrix Human U133 Plus2 chip as previously reported(24). For both COPDGene and ECLIPSE, the institutional review boards of all participating centers approved these studies, and written informed consent was obtained from all subjects.

Phenotype and Covariate Definitions

In COPDGene and ECLIPSE, spirometry was performed before and after administration of 180 mcg of albuterol according to international guidelines(25). COPD cases and GOLD stages were defined according to GOLD spirometric criteria (FEV1 % of predicted < 80% and FEV1/FVC < 0.7) (26). Subjects with preserved ratio impaired spirometry (PrISM) were defined by post-bronchodilator FEV_1 % of predicted <80% and FEV_1/FVC > 0.7(27). COPD blood gene expression subtypes were previously defined by Chang et al. using network-based stratification (NBS)(24). Of the four NBS subtypes identified in the original publication, the two most prevalent subtypes were analyzed. These subtypes are referred to as the less impaired lung function (LI-NBS) and the more impaired lung function (MI-NBS) subtypes. Current smoking status, inhaled corticosteroid use, and oral corticosteroid use were ascertained by questionnaire. In ECLIPSE, only 8 subjects reported using oral steroids at baseline, and these were removed from subsequent analyses.

Association of CBC Cell Types with COPD GOLD Stage, Cross-Sectional FEV_1, and Prospective Change in FEV_1

Using 4,558 subjects with complete CBC and spirometric data from the COPDGene Phase 2 visit, we plotted the distribution of neutrophil and lymphocyte counts and proportions against GOLD Stage after removing outlying cell count observations greater than +/- 4 SD from the mean. We tested for association between cell counts and proportions with post-bronchodilator FEV_1 % of predicted using Wald tests from linear regression models.

In 1,741 smokers from the ECLIPSE Study with complete covariate, CBC, baseline and prospective FEV_1 measurements, we related absolute cell counts and proportions to post-bronchodilator FEV_1 % of predicted levels as described above. For the analysis of three-year change in FEV_1 % of predicted levels, we calculated the difference between the first and last available post-bronchodilator FEV_1 % of predicted measurement in all study subjects (calculated as last measurement - first measurement, i.e. negative values represent decline in lung function). To determine the association between CBC measurements and change in FEV_1, we used

multivariate linear regression models with change in FEV_1 as the response variable adjusting for baseline FEV_1, days of follow up, and smoking status at the first and last study visit. Smoking status was represented in four groups, i.e. current smokers at first and last visit, former smokers at first and last visit, current smoker at first visit and former smoker at last visit, and former smoker at first visit and current smoker at last visit. Subjects with less than 1000 days between their first and last spirometric measurements were excluded from analysis.

Gene Expression

Sample preparation and quality control procedures for genome-wide gene expression data in ECLIPSE have been previously described(28). Standard quality control and quantile normalization were performed. Gene expression data are accessible via GEO [ECLIPSE GSE76705].

Cell Type Deconvolution and Association of Inferred Cell Types with COPD Status

Cell type deconvolution was performed in 221 ECLIPSE subjects with complete genome-wide gene expression and covariate data. Cell type proportions were inferred using two methods - CIBERSORT(21) and the method of Abbas et al. using linear least squares regression (LR)(20). Cell type reference expression profiles were used from the LM22 pure-cell dataset obtained on [12/21/2015] from the CIBERSORT website (https://cibersort.stanford.edu). Detailed description is provided in the Online Supplement.

After obtaining cell type estimates of the 22 cell types from each of the three methods, we organized groups of similar cell types into broader categories to create estimates of an additional 9 aggregated groups: CD4+ cells, T cells, B cells, lymphocytes, and monocytes/macrophages. We performed this aggregation by summing individual cell type values for cell types within each category. The 22 inferred cell type proportions and aggregated cell type estimates were tested for association with COPD status and COPD molecular subtypes using the Wilcoxon-Mann-Whitney test. Significant cell type associations were considered to be those with a Wilcoxon-Mann-Whitney test p-val < 0.05 for both CIBERSORT and LR.

Prediction Models for COPD Status and COPD Molecular Subtypes

Classification of subjects according to COPD status or NBS molecular subtype using estimated cell-type quantities, CBC quantities, and clinical covariates was performed in R, using the support vector machine implementation in the e1071 package(29). Validation within ECLIPSE involved performing one round of partitioning in which half of the subjects were used in the training set and the other half were used in the validation set. Probabilities were returned from the SVM and used with the R package ROCR to generate ROC plots and calculate AUCs(30).

Additional details regarding study cohorts and statistical methods are included in the Supplemental Materials.

Results

Relating Circulating Immune Cells to Cross-sectional FEV_1 and Current Smoking

We examined complete blood count (CBC) data from 4,558 smokers from the COPDGene phase 2 visit and an additional 1,741 smokers with >1000 days of spirometric follow-up data in the ECLIPSE Study. The clinical characteristics and cell type distributions of analyzed subjects in both studies are shown in Table 1. The CONSORT diagram for the analyses of cross-sectional and longitudinal data in ECLIPSE is shown in Supplemental Figure 1, and comparison of characteristics of ECLIPSE analyzed and excluded subjects is shown in Table E1.

Linear regression relating the absolute amount and percentage of five cell types to FEV_1 % of predicted indicated that 1) neutrophils, lymphocytes, monocytes, and eosinophils are strongly correlated with FEV_1, 2) the absolute counts of eosinophils and monocytes are more strongly correlated with FEV_1 than the relative percentages of these cell types, and 3) there are differences in the patterns of association between absolute cell counts and cell count proportions with COPD severity (Tables 2 and E2). Boxplots showing the amount of each cell type by GOLD stage for COPDGene and ECLIPSE are shown in Supplemental Figures 2 and 3.

Given the predominance of neutrophils and lymphocytes in blood, we examined the absolute counts and percentages of these cell types across GOLD stages, and we observed two phenomena. First, with increasing COPD severity, the proportion of neutrophils increases and lymphocytes decreases. However, in terms of absolute cell counts the number of neutrophils increases while the total number of lymphocytes remains relatively stable (Figure 1, Panels A and C), suggesting that the observed changes in neutrophil and lymphocyte proportions associated with COPD severity are primarily driven by an increase in the number of circulating neutrophils. The same pattern is present in ECLIPSE subjects (Figure 1, Panels B and D).

We evaluated these relationships in a series of models in COPDGene relating cell count, cell proportion, and current smoking (CS) to FEV_1 % of predicted while adjusting for inhaled and oral steroid use, and these models demonstrated the following relationships (Table 3). First, the models explaining the largest proportion of variance in FEV_1, after adjusting for model complexity, included both cell counts and proportions demonstrating that both measures have independent association to FEV_1. Both lymphocyte and neutrophil absolute counts and percentages were significantly associated with FEV_1 across most models.

Relating Circulating Immune Cells to Prospective, Three-Year Change in Lung Function

In 1,741 smokers from the ECLIPSE Study with CBC data and three-year change in FEV_1, we related peripheral blood counts and proportions at baseline to change in lung function. In an analysis relating single cell type measures to change in FEV_1, monocyte and eosinophil counts and proportions were significantly associated with change in FEV_1 (Table 4). Higher monocyte levels at baseline were associated with greater FEV_1 decline, and the opposite pattern was observed for eosinophils. Neutrophil and lymphocyte proportions, but not counts, were significantly associated with lung function decline. Larger neutrophil proportions were associated with more lung function decline, with the opposite relationship observed for lymphocyte proportion.

Table 5 demonstrates that for multivariate models including counts and proportions of these four cell types, cell counts, but not proportions, showed significant associations to change in FEV_1. Absolute counts of

monocytes, eosinophils, and lymphocytes were significantly associated with FEV1 decline (p=0.0003, 0.0004, and p=0.02, respectively). Higher levels of monocytes were associated with larger amounts of FEV_1 decline, and the opposite pattern was present for lymphocytes and eosinophils.

Association of Inferred Lymphocyte Subpopulations to COPD and COPD Subtypes

In a subset of 221 subjects from ECLIPSE with complete genome-wide blood gene expression and covariate data (subject characteristics shown in Table E3), we used cell type deconvolution to estimate the proportion of immune cell subpopulations in each study subject, and we related these proportions to COPD case-control status.

To first assess the performance of cell type deconvolution in blood gene expression from smokers, we examined results from applying two methods that have been previously validated for the detection of immune cell types, CIBERSORT and the linear regression (LR) method of Abbas et al.(20, 21) To benchmark these algorithms against known cell type quantifications, we compared their neutrophil, aggregated lymphocyte, aggregated monocyte, eosinophil, and basophil quantifications against concurrently drawn CBC counts (Figure 2). Both methods showed high correlation to neutrophils and lymphocytes (Spearman *r* ranges from 0.7-0.8, p<0.001), with weaker correlations for eosinophils and monocytes. Correlation with basophils was low for both methods. For the inferred proportions of neutrophils and lymphocytes, the correlation between methods was high (0.86 and 0.83, respectively).

We compared the inferred cell type proportions by COPD case/control status and observed that, relative to smoker controls, subjects with COPD had significantly lower levels of aggregated lymphocytes, aggregated T-cells, CD4+ resting memory cells, naïve B-cells, and increased levels of monocytes (Table 6).

Inferred Cell Type Proportions Predict COPD Blood Gene Expression Subtypes

In a previous publication, the ECLIPSE blood gene expression data had been used to define COPD molecular subtypes that differed in clinical characteristics and blood gene expression patterns, and we previously demonstrated that these subtypes could not be recovered using CBC data alone(24). To determine whether these molecular subgroups could be accurately predicted from inferred cell count proportions, we trained support vector machine classifiers to predict NBS subtype and COPD case/control status using CBC data, clinical covariates, or inferred cell type proportions. Figure 3 demonstrates that predictive models for NBS subtypes using inferred cell type proportions classified subjects by COPD molecular subtype with high accuracy (AUC = 0.95) and demonstrated better performance than models using only CBC cell-type quantities (AUC = 0.53) or clinical covariates (AUC = 0.65). Predictive models for COPD case/control status using inferred cell type proportions also showed statistically significant, but less powerful, predictive performance (AUC = 0.71), with the cell type subpopulation models still outperforming the models using CBC data (Table E4).

We compared levels of the inferred cell type proportions between the NBS-MI and NBS-LI COPD molecular subtypes and observed that the list of specific immune cell types that were significantly different between COPD molecular subtypes was more extensive than between COPD cases and controls. This list included T-regulatory cells, CD4 resting and activated memory T-cells, memory B-cells, aggregated T-cells, aggregated B-cells, and monocytes. (Table E5, Figures E4 and E5).

Discussion

Using two large cohorts enriched for subjects with COPD, we characterized alterations in circulating immune cell types associated with cross-sectional FEV_1, longitudinal FEV_1 decline, COPD status, and COPD molecular subtypes. Using both CBC and cell type deconvolution methods, we studied general classes of immune cell types as well as specific immune cell subpopulations in current and former smokers across the full spectrum of lung function impairment. The main findings are: 1) the predominant peripheral immune cell type alteration associated with increasing COPD severity is an increase in the absolute count of neutrophils, which

drives the apparent association between FEV_1 and lymphocyte proportions. 2) Monocytes and eosinophils have strong multivariate associations to prospective change in FEV_1. 3) Cell-type estimates from gene expression deconvolution methods show good accuracy for some, but not all, cell types and show promise for future studies of immune cell subpopulations in large cohorts with gene expression data.

Prior immunologic studies of COPD have demonstrated alterations in specific innate and adaptive immune cell populations(31, 32). Most of these studies have been performed in small to moderate sized study samples with limited ability to control for the effects of current smoking. Our findings complement and extend previous results by demonstrating that 1) the decrease in overall lymphocyte proportion in COPD is primarily driven by an increase in absolute neutrophil counts, 2) absolute counts and proportions of immune cell types have independent, statistically significant associations to lung function, and 3) absolute lymphocyte, monocyte, and eosinophil counts are predictive of COPD disease progression. This point extends previous observations relating total peripheral leucocyte count to cross-sectional and longitudinal lung function(16-18) by implicating specific immune cell types.

To study immune cell subpopulations not quantified by CBC, we explored the use of cell type deconvolution approaches to quantify 22 distinct cell subpopulations in a subset of ECLIPSE subjects with blood genome-wide gene expression data. Both methods we employed have been previously validated against flow cytometry quantifications(20, 21), and these methods are a promising alternative for estimating cell type proportions in large study samples with available expression data. However, these approaches have not been widely applied in smokers with COPD. Our findings demonstrate that in smokers enriched for COPD, the deconvolution approaches studied yielded consistent and reasonably reliable estimates of neutrophil and lymphocyte cell proportions with mixed performance for other cell types. Inferred cell type proportions enabled significantly better prediction of externally defined COPD molecular subtypes than CBC and clinical data alone, providing indirect evidence that these inferred proportions capture meaningful information on the cell type composition of bulk blood expression samples. In the future, direct comparisons with flow cytometry

data obtained from blood in active and former smokers with COPD can provide additional validation for this promising approach beyond the flow cytometry benchmarking that has already been performed in populations without COPD, and our data provides proof-of-concept of the feasibility of using cell type deconvolution to study immune cell subpopulations in large cohorts of smokers with available blood gene expression data.

Analysis of the cell type deconvolution data supports previous observations of an overall decrease in lymphocytes and T-cells in the COPD state, coupled with an increase in monocytes. When we examined cell type differences in previously defined COPD molecular subtypes, the differences in immune cell profiles was more pronounced, with the more severely affected subtype characterized by increased monocytes, T-regulatory cells, memory T-cells, and memory B-cells as well as decreased total lymphocytes. These observations are consistent with previous reports describing a Th1-like cytokine response in COPD(12, 13) and a study identifying increased T-regulatory cells in the blood of subjects with COPD(33). This pattern is consistent with the expected behavior of T-regulatory cells, which play a critical immunomodulatory role by suppressing other lymphocyte populations in part through IL-10 and TGF-β(34). Overall, these findings provide additional support for the model that the circulating immune response to CS and COPD is characterized by distinct aspects of suppression of the adaptive immune response(7) and a chronic increase in innate immune cell types, and it suggests that within smokers there are patterns of coordinated immune response that can be used to identify clinically distinct subgroups of subjects.

The main strength of this study is that peripheral cell type quantifications were available from a large number of smokers with a broad range of lung function from two independent studies. The study design also enabled the study of immune cell alterations adjusting for CS, a major confounder of the relationship between immune cell alterations and COPD severity. The use of cell type deconvolution to study the immune response in COPD is novel and enabled the simultaneous study of a large number of lymphocyte subpopulations. Because CBC quantifications and blood gene expression were available in the same ECLIPSE subjects from the same time point, we could benchmark our deconvolution approaches against a known standard.

This study also has important limitations. We did not have access to immune cells in the lung or specific lung compartments in our study subjects, thus we could not relate the blood observations to the lung compartment. We also were not able to characterize immune cell functional states through cytokine profiling or quantification of response to antigenic stimulation. While we observed significant associations between circulating immune cell subpopulations and COPD, further study is required to determine the pathophysiological significance of these observations. We did not have flow cytometry values against which we could benchmark our cell type deconvolution estimates, but we did have Coulter counter data available, and our deconvolution results were generated using methods that has been previously validated against flow cytometry for immune cell populations.

In conclusion, analysis of CBC counts and proportions in >6,000 subjects from the COPDGene and ECLIPSE studies demonstrated that cross-sectional FEV_1 is associated with alterations in multiple circulating immune cell types, including total neutrophil count. COPD disease progression, as quantified by decline in FEV_1, is associated with increased absolute monocyte counts and decreased lymphocytes and eosinophil counts at baseline. Cell type deconvolution is a viable approach to simultaneously study multiple immune cell populations in smokers with COPD, and our findings support a model of suppressed adaptive immune responses coupled with enhanced innate immunity. Future studies to characterize COPD-related alterations in more fine-grained immune cell types will benefit from quantification of both cell type proportions and absolute counts.

Abbreviations

AUC – area under the curve

CBC – complete blood count

COPD – chronic obstructive pulmonary disease

CT – computed tomography

FEV_1 – forced expiratory volume in one second

GOLD – Global Initiative for Obstructive Lung Disease

LAA950 – low attenuation area less than 950 Hounsfield units (computed tomorgraphy based emphysema measure)

LR – linear regression

NBS – network-based stratification

ROC – receiver operating characteristic

SVM – support vector machine

Declarations

Ethics approval and consent to participate

For both COPDGene and ECLIPSE, the institutional review boards of all participating centers approved these studies, and written informed consent was obtained from all subjects.

Consent for publication

Not applicable.

Availability of data and material

Gene expression data used in this study are accessible via GEO [ECLIPSE GSE76705, COPDGene GSE42057].

Competing interests

PJC receives research support and has served on an Advisory Board for GSK. All other authors have no conflicts to disclose.

Funding

This work was funded by the University of Colorado School of Medicine Medical Student Research Track, NHLBI R01HL089897, R01HL089856, R01HL124233, and R01HL126596. The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of

AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, GSK, Siemens and Sunovion. The ECLIPSE study (GSK code SCO104960) was funded by GSK.

Authors' contributions

Study concept and design: Stromberg, Castaldi, Bowler

Acquisition, analysis, or interpretation of data: Stromberg, Castaldi, Bowler

Drafting of the manuscript: Stromberg, Castaldi, Bowler

Critical revision of the manuscript for important intellectual content: All authors

Statistical analysis: Stromberg, Castaldi, Bowler

Obtained funding: Castaldi and Bowler

1 **Table 1. Characteristics of Analyzed Subjects in COPDGene and ECLIPSE**

	COPDGene	ECLIPSE
N	4,568	1741
Age	65.44 (8.71)	61.89 (7.85)
Gender, % Female	50	36
Race, % African-American	27	0
FEV1 % of predicted	78.32 (24.93)	55.0 (26.13)
FEV1/FVC	0.67 (0.15)	0.66 (0.21)
COPD, % GOLD 2-4	57	84
Pack Years	43.96 (23.95)	45.40 (26.48)
Current Smoking	37	39
BMI	28.91 (6.28)	26.59 (5.34)
Oral Steroids, %	2	0
neutrophil , %	59.38 (10.01)	63.88 (8.17)
neutrophil , 1000 cells/ul	4.29 (1.62)	4.95 (1.64)
lymphocyte , %	29.37 (9.36)	26.84 (7.62)
lymphocyte , 1000 cells/ul	2.02 (0.73)	2.00 (0.64)
monocyte , %	8.07 (2.37)	6.15 (2.13)
monocyte , 1000 cells/ul	0.56 (0.19)	0.46 (0.18)
eosinophil , %	2.57 (1.65)	2.79 (1.73)
eosinophil , 1000 cells/ul	0.18 (0.12)	0.21 (0.14)
basophil , %	0.6 (0.53)	0.34 (0.2)
basophil , 1000 cells/ul	0.03 (0.04)	0.03 (0.02)
Data are mean (SE) unless otherwise indicated.		

2

3 **Table 2. Relationship of Cell Type Counts and Proportions to FEV_1 % of Predicted in 4,558 Smokers**

4 **in COPDGene**

	Cell Type Count (1000 cells/uL)		Cell Type Proportion	
	Beta	P value	Beta	P value
Neutrophils	-3.53 (0.22)	<0.001	-0.48 (0.04)	<0.001
Lymphocytes	2.34 (0.51)	<0.001	0.58 (0.04)	<0.001
Monocytes	-21.84 (1.88)	<0.001	-0.29 (0.16)	0.06
Eosinophils	-21.25 (3.06)	<0.001	-0.54 (0.22)	0.02
Basophils	-33.28 (9.81)	0.001	0.09 (0.7)	0.89
Beta - change in FEV_1 per unit change in cell count or proportion.				

5

6

7

8 **Table 3. Multivariate Models of the Relationship between FEV_1 % of Predicted and Neutrophil and Lymphocyte Quantifications**

	Neutrophil Count	Neutrophil %	Lymphocyte Count	Lymphocyte %	Current Smoking	Neutrophil Count x Current Smoking	Neutrophil % x Current Smoking	Lymphocyte Count x Current Smoking	Lymphocyte % x Current Smoking	Adjusted R2
Cell Proportions	---	0.39 (0.10) *	---	0.75 (0.11) *	---	---	---	---	---	0.274
Cell Counts	-2.19 (0.20) *	---	1.72 (0.44) *	---	---	---	---	---	---	0.275
Cell Counts + Current Smoking	-2.73 (0.27) *	---	1.79 (0.58) *	---	-5.74 (2.48) *	1.21 (0.40) *	---	`-0.06 (0.90)	---	0.276
Cell Proportions + Counts + Current Smoking	-2.47 (0.62) *	0.74 (0.14) *	-0.41 (1.32)	0.94 (0.16) *	-10.99 (18.91)	0.24 (1.03)	0.13 (0.23)	1.85 (2.07)	-0.1 (0.27)	0.285

* $p<0.05$

Adjusted R2 - variance explained adjusted for model complexity. Cell counts are in units of 1000 cells/ul.

9

0

1

2

3

4

Table 4. Relationship of Cell Type Counts and Proportions at Baseline to Three-Year Change in FEV_1 % of Predicted in 1,741 Smokers in ECLIPSE

	Cell Type Count (1000 cells/uL)		Cell Type Proportion	
	Beta	P value	Beta	P value
Neutrophils	-0.11 (0.12)	0.37	-0.06 (0.02)	0.017
Lymphocytes	0.58 (0.32)	0.07	0.07 (0.03)	0.006
Monocytes	-3.36 (1.09)	0.002	-0.30 (0.09)	0.001
Eosinophils	4.31 (1.44)	0.003	0.37 (0.11)	0.001
Basophils	-5.07 (12.91)	0.69	-0.10 (1.01)	0.92

Beta - change in FEV_1 over three years per unit change in cell count or proportion.
Change in FEV_1 calculated as (last visit value - first visit value), i.e. negative values indicate decline in FEV_1
Models analyze one cell type at a time. All models are adjusted for FEV1 % of predicted at baseline and smoking status at baseline and 3-year follow-up.

Table 5. Multivariate Models Relating Cell Type Counts and Proportions at Baseline to Three-Year Change in FEV_1 % of Predicted in 1,741 Smokers in ECLIPSE

	Neutrophil Count	Neutrophil %	Lymphocyte Count	Lymphocyte %	Monocyte Count	Monocyte %	Eosinophil Count	Eosinophil %
Cell Proportions	---	1.03 (1.03)	---	1.11 (1.03)	---	0.69 (1.04)	---	1.43 (1.05)
Cell Counts	0 (0.13)	---	0.76 (0.32) *	---	-4.64 (1.18) *	---	4.98 (1.46) *	---

* $p<0.05$

Cell counts are in units of 1000 cells/ul.

Models adjusted for baseline FEV1 % of predicted, days of follow-up, and current smoking status at baseline and last measurement.

Table 6. Inferred Immune Cell Types Associated with COPD Status

	Linear Regression			CIBERSORT		
	Controls	Cases	pvalue	Controls	Cases	pvalue
Aggregated monocytes	-0.41 (0.93)	0.25 (0.96)	5.4E-07	-0.35 (0.98)	0.20 (0.94)	6.3E-05
Monocytes	-0.33 (0.99)	0.21 (0.95)	4.9E-05	-0.32 (0.97)	0.19 (0.94)	2.1E-04
Naïve B cells	0.21 (0.89)	-0.13 (1.06)	2.4E-03	0.15 (0.92)	-0.11 (1.04)	1.5E-03
Aggregated T cells	0.24 (0.99)	-0.15 (0.99)	5.1E-03	0.4 (0.98)	-0.22 (0.94)	3.7E-05
CD4 memory resting T cells	0.25 (1.05)	-0.16 (0.95)	6.4E-03	0.23 (1.25)	-0.15 (0.79)	0.01
Aggregated lymphocytes	0.24 (0.94)	-0.13 (1.01)	0.02	0.25 (0.97)	-0.13 (0.98)	0.02

Cell count values normalized to mean 0 and standard error of 1. See Figures E4 and E5 for boxplot distributions.
Values are mean (SD) unless otherwise indicated.
P-values calculated using Wilcoxon rank sum test.
The following cell types were not analyzed because they were detected in less than 10% of samples: M1 Macrophages, activated dendritic cells, activated Mast cells, gamma-delta T-cells, and T follicular helper cells.

Figure 1: Neutrophil and Lymphocyte Counts and Proportions Stratified by GOLD Spirometric Stage. As GOLD Stage increases, the relative proportions of peripheral neutrophils and lymphocytes increase and decrease, respectively (Panel A, COPDGene and Panel C, ECLIPSE). This phenomenon is driven by an increase in absolute neutrophil count, while the absolute amount of lymphocytes remains relatively stable across GOLD Stages (Panel B, COPDGene and Panel D, ECLIPSE). NS = non-smoker. PRISm = Preserved ratio impaired spirometry, i.e. subjects with $FEV_1/FVC > 0.7$ but FEV_1 % predicted < 80%.

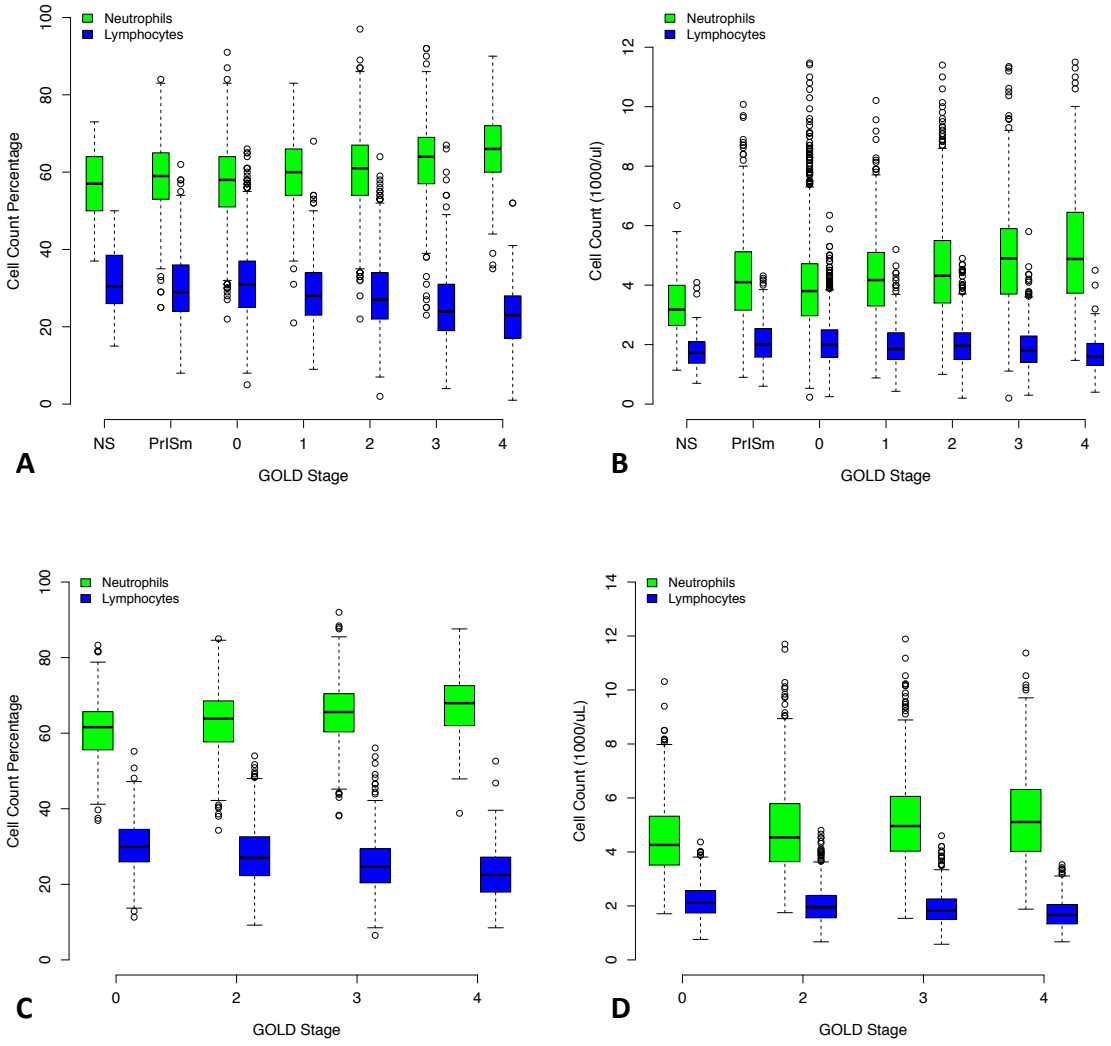


Figure 2. Correlation Between Inferred Cell Subpopulation Proportions and Complete Blood Count Cell Type Proportions. Spearman correlation between estimated cell subpopulations proportions from three deconvolution methods and CBC proportions. Abbreviations: LR- Linear Regression. CI- CIBERSORT. CBC- Complete Blood Count. Neuts- Neutrophils. Lymph- Lymphocytes. Eos- Eosinophils. Mons- Monocytes. Basos- Basophils.

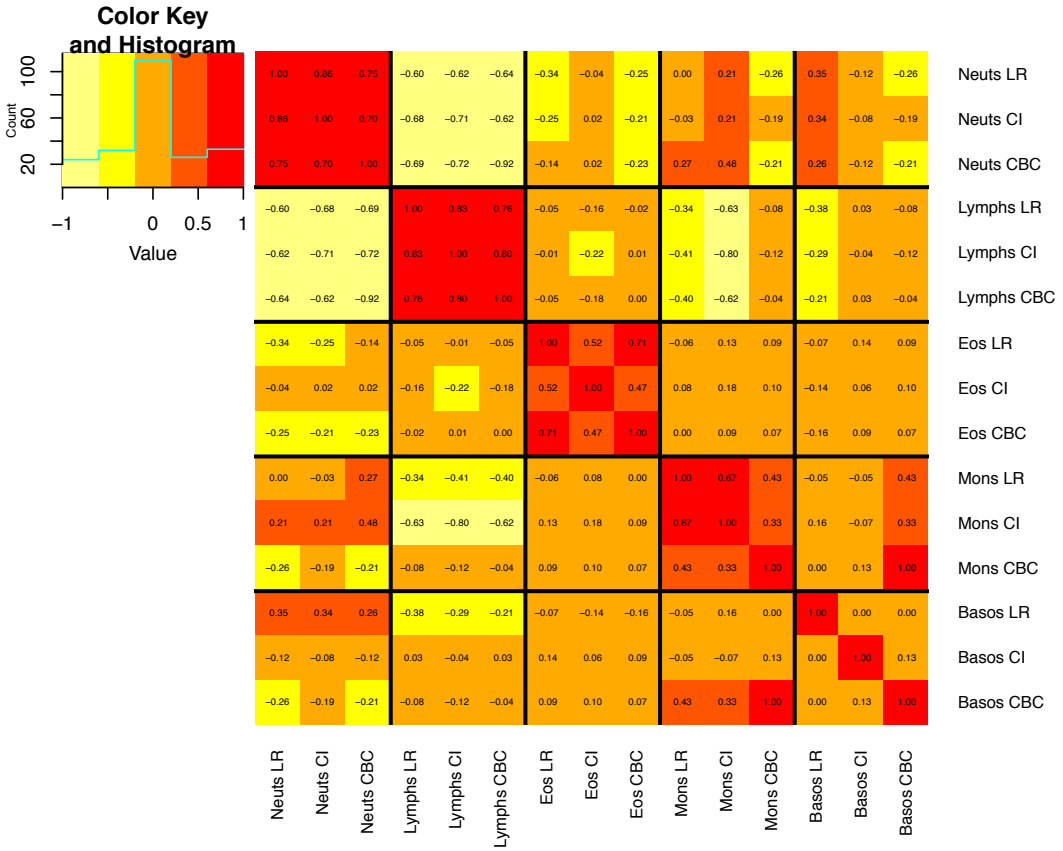
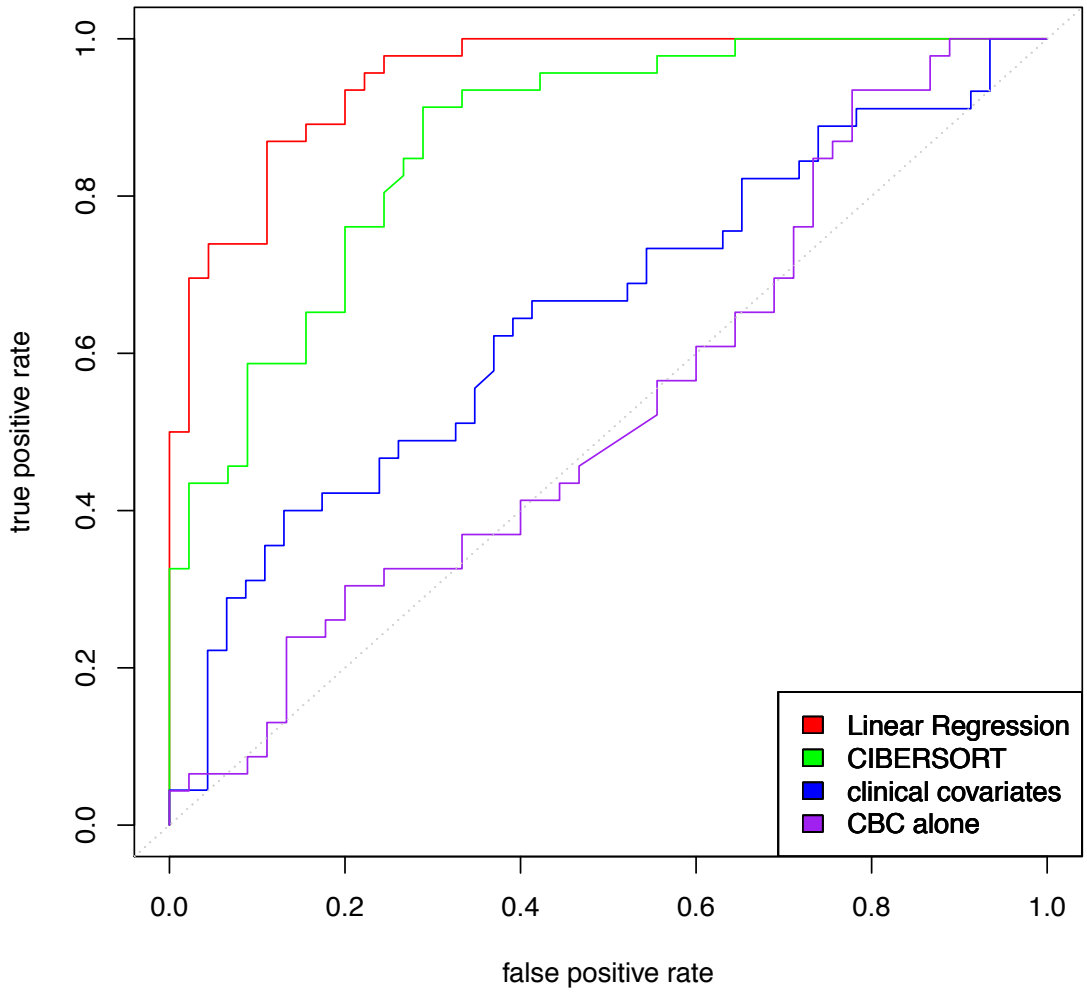


Figure 3. Performance of Predictive Models for COPD Molecular Subtypes and COPD Case-control Status Using Complete Blood Counts, Inferred cell Subpopulation Proportions, and Clinical Covariates. Receiver operating characteristic curves demonstrative the predictive performance of SVM classifiers using complete blood count data, inferred cell subpopulation data, and clinical covariates for COPD molecular subtypes (Panel A) and COPD status (Panel B). Results are shown for models learned and validated in ECLIPSE. Clinical covariates are age, sex, and pack-years.



References

1. Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *European Respiratory Journal* 2009;33:1165–1185.
2. Oudijk E-JD, Nijhuis EHJ, Zwank MD, van de Graaf EA, Mager HJ, Coffer PJ, Lammers J-WJ, Koenderman L. Systemic inflammation in COPD visualised by gene profiling in peripheral blood neutrophils. *Thorax* 2005;60:538–544.
3. Vaguliene N, Zemaitis M, Lavinskiene S, Miliauskas S, Sakalauskas R. Local and systemic neutrophilic inflammation in patients with lung cancer and chronic obstructive pulmonary disease. *BMC Immunol* 2013;14:36.
4. Sapey E, Stockley JA, Greenwood H, Ahmad A, Bayley D, Lord JM, Insall RH, Stockley RA. Behavioral and structural differences in migrating peripheral neutrophils from patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* 2011;183:1176–1186.
5. Donaldson GC, Seemungal TAR, Patel IS, Bhowmik A, Wilkinson TMA, Hurst JR, Maccallum PK, Wedzicha JA. Airway and systemic inflammation and decline in lung function in patients with COPD. *Chest* 2005;128:1995–2004.
6. van Eeden SF, Lawrence E, Sato Y, Kitagawa Y, Hogg JC. Neutrophils released from the bone marrow by granulocyte colony-stimulating factor sequester in lung microvessels but are slow to migrate. *Eur Respir J* 2000;15:1079–1086.
7. Arnson Y, Shoenfeld Y, Amital H. Effects of tobacco smoke on immunity, inflammation and autoimmunity. *J Autoimmun* 2010;34:J258–65.
8. de Jong JW, van der Belt-Gritter B, Koëter GH, Postma DS. Peripheral blood lymphocyte cell subsets in subjects with chronic obstructive pulmonary disease: association with smoking, IgE and lung function. *Respiratory Medicine* 1997;91:67–76.
9. Zhu X, Gadgil AS, Givelber R, George MP, Stoner MW, Sciurba FC, Duncan SR. Peripheral T cell functions correlate with the severity of chronic obstructive pulmonary disease. *J Immunol* 2009;182:3270–3277.
10. Silver JS, Kearley J, Copenhaver AM, Sanden C, Mori M, Yu L, Pritchard GH, Berlin AA, Hunter CA, Russell P Bowler MP, Erjefalt JS, Kolbeck R, Humbles AA. Inflammatory triggers associated with exacerbations of COPD orchestrate plasticity of group 2 innate lymphoid cells in the lungs. *Nat Immunol* 2016;17:626–635.
11. Mustimbo E P Roberts P, Brandon W Higgs P, Philip Brohawn BS, Xiang Guo P, Fernanda Pilataxi BA, Michael Kuziora P, Russell P Bowler MP, Wendy I White P. CD4+ T-Cell Profiles and Peripheral Blood Ex-Vivo Responses to T-Cell Directed Stimulation Delineate COPD Phenotypes. *Chronic Obstructive Pulmonary Diseases:Journal of the COPD Foundation* 2014;2:268–280.
12. Majori M, Corradi M, Caminati A, Cacciani G. Predominant T h1 cytokine pattern in peripheral blood from subjects with chronic obstructive pulmonary disease. *Journal of Allergy and Clinical Immunology* 1999;
13. Hodge G, Nairn J, Holmes M, Reynolds PN, Hodge S. Increased intracellular T helper 1 proinflammatory cytokine production in peripheral blood, bronchoalveolar lavage and intraepithelial T cells of COPD subjects. *Clinical & Experimental Immunology* 2007;150:22–29.
14. Morissette MC, Parent J, Milot J. Perforin, granzyme B, and FasL expression by peripheral blood T lymphocytes in emphysema. *Respiratory Research* 2007;8:62.
15. Wang J, Urbanowicz RA, Tighe PJ, Todd I, Corne JM, Fairclough LC. Differential Activation of Killer Cells in the Circulation and the Lung: A Study of Current Smoking Status and Chronic Obstructive Pulmonary Disease (COPD). In: Allen RL, editor. *PLoS ONE* 2013;8:e58556.
16. Sparrow D, Glynn RJ, Cohen M, Weiss ST. The relationship of the peripheral leukocyte count and

cigarette smoking to pulmonary function among adult men. *Chest* 1984;86:383–386.
17. Chan-Yeung M, Abboud R, Buncio AD, Vedal S. Peripheral leucocyte count and longitudinal decline in lung function. *Thorax* 1988;43:462–466.
18. Gan WQ, Man SFP, Senthilselvan A, Sin DD. Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. *Thorax* 2004;59:574–580.
19. Hoonhorst SJM, Timens W, Koenderman L, Tam Loi Lo AT, Lammers J-WJ, Boezen HM, van Oosterhout AJM, Postma DS, Hacken Ten NHT. Increased activation of blood neutrophils after cigarette smoking in young individuals susceptible to COPD. *Respiratory Research* 2014;15:121.
20. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. In: Tan P, editor. *PLoS ONE* 2009;4:e6098.
21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Meth* 2015;12:453–457.
22. Regan EA, Silverman E, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.
23. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, Hagan G, Knobil K, Lomas DA, MacNee W, Silverman E, Tal-Singer R, ECLIPSE Investigators. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *European Respiratory Journal* 2008;31:869–873.
24. Chang Y, Glass K, Liu Y-Y, Silverman E, Crapo JD, Tal-Singer R, Russell P Bowler MP, Dy JG, Cho M, Castaldi P. COPD subtypes identified by network-based clustering of blood gene expression. *Genomics* 2016;107:51–58.
25. Standardization of Spirometry, 1994 Update. American Thoracic Society. *American Journal of Respiratory and Critical Care Medicine* 1995;
26. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DMG, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Agustí AGN. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *American Journal of Respiratory and Critical Care Medicine* 2017;
27. Wan ES, Hokanson JE, Murphy JR, Regan EA, Make BJ, Lynch DA, Crapo JD, Silverman E, COPDGene Investigators. Clinical and radiographic predictors of GOLD-unclassified smokers in the COPDGene study. *American Journal of Respiratory and Critical Care Medicine* 2011;184:57–63.
28. Singh D, Fox SM, Tal-Singer R, Bates S, Riley JH, Celli B. Altered gene expression in blood and sputum in COPD frequent exacerbators in the ECLIPSE cohort. In: Eickelberg O, editor. *PLoS ONE* 2014;9:e107381.
29. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2011;2:27–27.
30. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics* 2005;21:3940–3941.
31. Hoenderdos K, Condliffe A. The neutrophil in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2013;48:531–539.
32. Barnes PJ. Immunology of asthma and chronic obstructive pulmonary disease. *Nat Rev Immunol* 2008;8:183–192.
33. Brandsma C-A, Hylkema MN, Geerlings M, van Geffen WH, Postma DS, Timens W, Kerstjens HAM. Increased levels of (class switched) memory B cells in peripheral blood of current smokers. *Respiratory Research* 2009;10:108.
34. Vignali DAA, Collison LW, Workman CJ. How regulatory T cells work. *Nat Rev Immunol* 2008;8:523–532.