Online Data Supplement

Peripheral blood gene expression signatures of eosinophilic COPD

Jeong H. Yun M.D., M.P.H., Robert Chase M.S., Margaret M. Parker Ph.D., Aabida Saferali Ph.D., Peter J. Castaldi M.D, Edwin K. Silverman M.D., Ph.D., Craig P. Hersh M.D., M.P.H. for the COPDGene Investigators

Supplementary Methods

RNA sequencing

At the COPDGene Phase 2 (5-year) visit, blood samples were collected into PAXgene blood RNA tubes and total RNA was extracted using PAXgene blood miRNA kit (Preanalytix). TruSeq Stranded Total RNA library prep kit (Illumina) was used to generate cDNA libraries depleted for globin genes using Ribo-Zero Globin kit. Sequencing was performed as 75 base pair paired-end runs on an Illumina HiSeq2000. Paired end reads were mapped to the human reference GRCh38 using STAR 2.4.0h, annotated to genes and exons using Ensembl version 81.

Differential gene expression analysis

Differential gene expression analyses comparing eosinophilic vs. non-eosinophilic COPD were performed with voom transformation in the limma R package[1]. Age, gender, race, pack-years of smoking, current smoking status, white blood cell count, library batch and surrogate variables were included as covariates in the linear model. Differentially expressed genes were defined based on a false discovery rate< 0.05. Pathway analysis was performed using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit, http://webgestalt.org) using differentially expressed genes with false discovery rate < 0.20[2]. For identification of eosinophil associated genes, eosinophil count was included as an independent variable in addition to the covariates (age, gender, race, pack-years of smoking, current smoking status, library batch and surrogate variables) in the multivariate linear regression within limma. Statistical significance of overlap between gene lists were calculated by hypergeometric test. Euler diagrams were generated using eulerr R package[3]. Expression data has been deposited to Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE97531, and the Database of Genotypes and Phenotypes (dbGaP) (https://www.ncbi.nlm.nih.gov/gap/), accession phs000765.v6.p2.

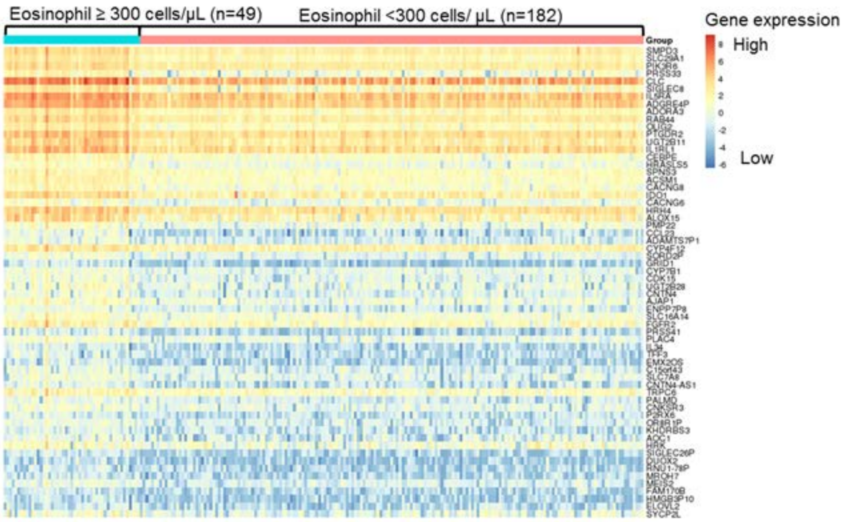
Supplementary Table E1. Characteristics of study participants

	Control (n= 224)	COPD (n= 231)	P values
Age (mean(SD))	62.31 (8.36)	67.20 (8.64)	<0.01
Gender (Female) (%)	114 (50.9)	98 (42.4)	0.09
Race (White) (%)	151 (67.4)	177 (76.6)	0.04
Post-bronchodilator FEV1 % predicted (mean(SD))	97.25 (11.08)	51.66 (17.03)	<0.01
GOLD grade (%)			
0	224 (100)	0	
2	0	124 (53.7)	
3	0	76 (32.9)	
4	0	19 (8.3)	
Any asthma history (%)	24 (10.7)	50 (21.6)	<0.01
BMI (kg/m2) (mean(SD))	29.12 (6)	27.72 (6.83)	0.02
Pack-Years of smoking (mean(SD))	37.89 (18.46)	50.86 (23.58)	<0.01
Current smokers (%)	108 (48.2)	97 (42)	0.22
Inhaled corticosteroid use (%)	19 (8.4)	115 (49.1)	<0.001
WBC (mean(SD))	7.17 (2.23)	7.48 (2.28)	0.14
Eosinophil % (mean(SD))	2.52 (1.89)	2.69 (2.54)	0.43
Neutrophil % (mean(SD))	58.00 (10.27)	60.64 (10.97)	0.01
Lymphocyte % (mean(SD))	31.14 (9.46)	27.64 (10.02)	<0.01
Monocyte % (mean(SD))	7.72 (2.20)	8.44 (2.54)	<0.01
Basophil % (mean(SD))	0.60 (0.60)	0.58 (0.58)	0.80
Eosinophil ≥ 300/μL	33 (14.7)	49 (21.2)	0.09

Reference

1. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. PMID: 25605792
2. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 2017;4:278.
3. J L. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. 2019. https://cran.r-project.org/package=eulerr

Supplementary Figure E1. Heatmap of log2 CPM normalized counts of top differentially expressed genes in eosinophilic COPD compared to non-eosinophilic COPD.


Eosinophil ≥ 300 cells/μL (n=49)
Eosinophil <300 cells/ μL (n=182)
Group
SMPD3
SLC29A1
PIK3R6
PRSS33
CLC
SIGLEC8
IL5RA
ADGRE4P
ADORA3
RAB44
OLIG2
PTGDR2
UGT2B11
IL1RL1
CEBPE
HRASLS5
SPNS3
ACSM1
CACNG8
IDO1
CACNG6
HRH4
ALOX15
PMP22
CCL23
ADAMTS7P1
CYP4F12
SORD2P
GRID1
CYP7B1
CDK15
UGT2B28
CNTN4
AJAP1
ENPP7P8
SLC16A14
FGFR2
PRSS41
PLAC4
IL34
TFF3
EMX2OS
C15orf43
SLC7A8
CNTN4-AS1
TRPC6
PALMD
CNKSR3
P2RX6
OR8R1P
KHDRBS3
AOC1
HRK
SIGLEC26P
DUOX2
RNU1-78P
MROH7
MEIS2
FAM170B
HMGB3P10
ELOVL2
SYCP2L
Gene expression
High
Low