A bronchial gene signature specific for severe COPD that is retained in the nose


This manuscript has recently been accepted for publication in the *ERJ Open Research*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJOR online.

Copyright ©The authors 2023. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org
A bronchial gene signature specific for severe COPD that is retained in the nose

Authors: Jos van Nijnatten\textsuperscript{1,3,5}; Alen Faiz\textsuperscript{2,3,5}; Wim Timens\textsuperscript{1,3}; Victor Guryev\textsuperscript{4}; Dirk-Jan Slebos\textsuperscript{2,3}; Karin Klooster\textsuperscript{2,3}; Jorine E. Hartman\textsuperscript{2,3}; Tessa Kole\textsuperscript{2,3}; David F. Choy\textsuperscript{6}; Arindam Chakrabarti\textsuperscript{8}; Michele Grimbaldeston\textsuperscript{6}; Carrie M. Rosenberger\textsuperscript{6}; Genentech partners; Huib Kerstjens\textsuperscript{2}; Corry-Anke Brandsma\textsuperscript{*1,3}; Maarten van den Berge\textsuperscript{*2,3};

Author affiliations:
1) University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands
2) University of Groningen University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands
3) University of Groningen University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands
4) University of Groningen University Medical Center Groningen, European Research Institute for the Biology of Ageing, Groningen, the Netherlands
5) University of Technology Sydney, Respiratory Bioinformatics and Molecular Biology, Sydney, New South Wales, Australia
6) Genentech Inc, South San Francisco, CA, USA

* Co-senior authors
Abstract

Introduction: A subset of COPD patients develops advanced disease with severe airflow obstruction, hyperinflation, and extensive emphysema. We propose that the pathogenesis in these patients differs from mild-moderate COPD and is reflected by bronchial gene expression.

Aims: To identify a unique bronchial epithelial gene signature for severe COPD patients.

Methods: We obtained RNA sequencing data from bronchial brushes from 123 ex-smokers with severe COPD, 23 with mild-moderate COPD and 23 non-COPD controls. We identified genes specific to severe COPD by comparing severe COPD to non-COPD controls, followed by removing genes that were also differentially expressed between mild-moderate COPD and non-COPD controls. Next, we performed a pathway analysis on these genes and evaluated whether this signature is retained in matched nasal brushings.

Results: We identified 219 genes uniquely differentially expressed in severe COPD. Interaction network analysis identified VEGFA and FN1 as the key genes with the most interactions. Genes were involved in extracellular matrix regulation, collagen binding and the immune response. Of interest were ten genes (VEGFA, DCN, SPARC, COL6A2, MGP, CYR61, ANXA6, LGALS1, C1QA, and C1QB) directly connected to fibronectin 1 (FN1). Most of these genes were lower expressed in severe COPD and showed the same effect in nasal brushings.

Conclusions: We found a unique severe COPD bronchial gene signature with key roles for VEGFA and FN1, which was retained in the upper airways. This supports the hypothesis that severe COPD, at least partly, comprises a different pathology and supports the potential for biomarker development based on nasal brushes in COPD.
Introduction

Chronic obstructive pulmonary disease (COPD) is a common, chronic inflammatory lung disease affecting over hundreds of millions of people worldwide.\(^1\,^2\). It was the third leading cause of death worldwide in 2019 and is expected to become even more prevalent in the upcoming years.\(^3\). Characteristics of COPD include irreversible airflow limitation, hypersecretion of mucus and alveolar destruction (emphysema).\(^4\).

One of the most common risk factors for COPD development is the inhalation of noxious particles. This includes cigarette smoking, second-hand smoke, biomass smoke and air pollution.\(^5\,^7\). Cigarette smoke exposes the lung, and specifically the bronchial epithelium, to over 4000 different components,\(^8\), directly causing irritation, mucus hypersecretion and inflammation in the airway. Duration and intensity of smoking have previously been associated with COPD incidence, increasing the risk of developing COPD five-fold so that approximately 30% of smokers develop COPD.\(^9\,^12\). Importantly, some people develop COPD faster and to a much more severe extent than others, suggesting an underlying individual susceptibility to the disease.\(^13\).

Most patients develop COPD later in life and only suffer from mild to moderate airflow obstruction (mCOPD), whereas a small subset of patients progresses to advanced disease with severe airflow obstruction, hyperinflation, and extensive emphysema or small airways disease (sCOPD).\(^14\). This severe group accounts for a majority of the personal burden as well as societal and economic burden attributed to COPD via healthcare and time lost from work.\(^15\). Next, to help with smoking cessation, treatment of the remaining sCOPD patients aims at alleviating the symptoms by providing temporary bronchodilation, while this does not alter the progressive lung function decline characteristic of the disease.\(^16\,^17\). More insight into the mechanisms leading to sCOPD and progressive lung function decline is needed to find novel targets for therapeutic intervention.
The aim of the present study was to investigate whether sCOPD represents a clinically distinct disease phenotype. To this end, we aimed to identify unique differentially expressed genes in bronchial brushes of sCOPD patients compared to mCOPD and non-COPDs. Additionally, since bronchial brushings are an invasive way to diagnose and phenotype disease, we also investigated whether the identified bronchial gene signature is also represented in the nasal epithelium, providing potential biomarkers for sCOPD in the nose.

**Methods**

*Patients and Study Design*

SHERLOCK (An integrative genomic approach to Solve tHe puzzle of sevEre earLy-Onset COPD, ClinicalTrials.gov: NCT04263961 and NCT04023409) is a cross-sectional study without pharmacological intervention performed by the University of Groningen, the Netherlands. We enrolled 23 non-COPD controls, 23 mCOPD patients (GOLD stages 1 and 2) and 123 patients with sCOPD (GOLD stages 3 and 4, with both extensive hyperinflation and emphysema). Participants did not smoke for at least two months prior to inclusion in the study and did not have an exacerbation or lung infection 4 weeks before the study. Subjects underwent bronchoscopy, during which bronchial and nasal brushes were obtained. All patients were fully characterized, i.e., lung function, CT scans, blood, and questionnaire data. RNA isolation and the RNA-Seq procedure are outlined in the Supplementary Methods.

The local medical ethics committees approved the study, and all subjects gave their written informed consent (the SHERLOCK study was approved by the medical ethics committee of the University of Groningen/University Medical Center Groningen, METc 2016/572 and METc 2014/102).

*Statistics*

All analyses were performed using R statistical software (version 4.0.2). The normality of the distribution of the data was established using histogram plots. A Kruskal-Wallis test was conducted across groups, and a Mann-Whitney U or Wilcoxon signed-rank test was conducted between the
groups for non-paired and paired data, respectively. A Benjamini-Hochberg false discovery rate (FDR) was calculated where appropriate. A p-value or FDR less than or equal to 0.05 was considered significant unless specified otherwise.

Selection of unique genes for sCOPD

To identify genes that are uniquely changed in bronchial brushes of sCOPD patients, we first identified genes that were differentially expressed between sCOPD patients compared to the non-COPD controls. We conducted a linear model using the EdgeR package (version 3.30.3, dependent upon Limma version 3.44.3), correcting for gender, age and smoking packyears. A n FDR of less than 0.05 and a fold change of > |2| was considered statistically significant. To assess if the sCOPD-associated genes also differ in mCOPD compared to non-COPD controls, we conducted a candidate gene approach for the sCOPD genes. Genes were identified as common COPD genes and removed from the sCOPD list when they were also differentially expressed in mCOPD vs non-COPD control in the same direction. We here used a more lenient FDR significance cut-off of 0.25 to avoid false-negative outcomes. Finally, we checked if the remaining sCOPD genes were also differentially expressed when directly comparing the sCOPD to mCOPD patients using an FDR cut-off of 0.05. Since most sCOPD subjects used high ICS doses, we removed steroid-sensitive genes, as previously identified in the GLUCOLD study, from our analysis. Additionally, we calculated the cell proportions using cellular deconvolution and adjusted for this in our model. These selection criteria for cell types and ICS genes, as well as replication in the nasal brushings and an independent study, and the pathway analysis, are described in the supplementary materials. Figure S1 presents an outline of the study.

Results

Clinical characteristics

In the current study, we investigated the differences in bronchial gene expression between patients with mCOPD (n=23), sCOPD (n=123) and non-COPD controls (n=23). There was no significant
difference in age across the groups. However, there was a difference in the male/female ratio; (non-COPD: 52% male, mCOPD: 78%, sCOPD: 29%) and packyears (non-COPD: 31.1±20.6 packyears, mCOPD: 66.6±62.9, sCOPD 39.1±18.3). We corrected for these two confounding factors in our models. Additionally, non-COPD participants did not use ICS, while mCOPD participants used 291±527 µg beclomethasone equivalent, and sCOPD used 620±877 µg. Clinical characteristics of included subjects are presented in Table 1.

Identification of genes common for COPD and unique for sCOPD

We identified 435 genes differentially expressed between sCOPD patients and non-COPD controls (FDR < 0.05, fold change (FC) > ±2). Of these, 213 genes showed a higher and 222 genes showed a lower expression in sCOPD patients. A volcano plot and heatmap are depicted in Figure 1 A and Figure S2, respectively.

Next, we performed a differential gene expression between non-COPD controls and mCOPD participants. Here we took a more lenient FDR cut-off of FDR < 0.25 and identified 123 genes differentially expressed in both mCOPD and sCOPD, which should thus not be considered unique for sCOPD (See Table S1). A volcano plot and a heatmap are depicted in Figure 1 B and Figure S3, respectively. After removing these 123 genes from the 435 sCOPD gene list, we were left with 312 potentially unique genes for sCOPD.

Next, we directly compared sCOPD versus mCOPD and found that 285 of the 312 genes were differentially expressed between mCOPD and sCOPD (FDR < 0.05). Of these, 118 genes were higher expressed in sCOPD compared to mCOPD, while 167 genes were lower expressed. A volcano plot and a heatmap are depicted in Figure 1 C and Figure S4, respectively. The top 3 remaining higher and lower expressed genes are shown in Figure 1D-I.
Effects of cell-type proportions

We then investigated whether there was a difference in cellular composition between the three patient groups. Within our bronchial brushings, we found goblet cells to be the most common cell type (average [74.2, 93.7]), followed by ciliated cells (average 11.7 [5.9, 19.9]) and the basal cells (average 0.0 [0.0, 2.6]) (figure 2 A). When comparing sCOPD to non-COPD controls, we found that there was a significantly lower proportion of ciliated and basal and a higher proportion of goblet cells (Figure 2 B, C & D). Adjusting for basal, ciliated and goblet cell proportions in our first differential expression analysis showed that 262 out of 285 sCOPD associated genes (92%) were not affected by cell proportions. (See Table S2.)

The effect of inhaled corticosteroids

We used an existing dataset of bronchial biopsies obtained from COPD patients before and after six months of treatment with either ICS+- LABA or placebo to define ICS-responsive genes. This resulted in a list of 2691 ICS-response genes, of which 43 were present in our sCOPD gene signature. The complete list of ICS-responsive genes can be found in Table S3. These 43 genes were therefore removed from our sCOPD gene signature (See Table S4).

Specific bronchial epithelium gene signature for sCOPD

Our final specific bronchial epithelium gene signature for sCOPD consisted of 219 genes that are uniquely differentially expressed in sCOPD compared to non-COPD controls. Of these 219 genes, 104 genes were higher expressed in sCOPD, with the top 10 most significant genes being: *MEX3D*, *LINCO0857*, *CEACAM5*, *TMC7*, *FNDC10*, *TPRXL*, *NETO2*, *SERPINB5*, *CALML3*, and *MUC12*. A total of 115 genes were lower in sCOPD, with the 10 most significant genes being *FXYD6*, *GGTA1P*, *GEM*, *CPED1*, *KCNJ5*, *VEGFA*, *JAKMIP2*, *DOK2*, *KMO*, and *GPR174*. A list of the top 10 genes more and less expressed in sCOPD can be found in Table 2; the complete list of 219 genes can be found in Table S5.
**Representation of the sCOPD signature in nasal brushings**

We assessed whether our bronchial epithelium sCOPD gene signature was also present in nasal brushings. Here we first compared the bronchial signature in matched nasal brushings that were collected at the same visit. Using GSVA analysis, we demonstrated a significant representation of the bronchial gene signature in the nasal brushes that were lower in sCOPD compared to non-COPD controls and mCOPD, while no significant representation was observed for the gene set that was higher in sCOPD (Figure 3 A & B).

Next, we checked whether our results showed similar representation in an independent nasal gene expression dataset comparing sCOPD to control. (Clinical characteristics of included subjects are presented in Table S6.) The bronchial gene signature that was lower in sCOPD, was similarly represented in that dataset, whereas the bronchial gene set that was higher was not (Figure 3 C & D).

We then performed a meta-analysis on the 219 genes unique for sCOPD in the matched nasal brushings and the independent nasal brushings dataset to assess similar representation in the nose on the gene level. In this meta-analysis, 83 genes in both datasets were significantly associated with sCOPD (meta-FDR<0.05) in the same direction in both nasal cohorts. 42 genes were lower expressed in the sCOPD patients compared to controls (top 5 genes: EPB41L2, FRMD4A, GGTA1P, GEM and CPED1), and 41 genes were higher expressed (top 5 genes: MEX3D, TMEM132B, PCSK1N, PRR7 and MESP1) (See Table S7).

**Pathway analysis**

StringDB analysis using default settings, of the 219 unique sCOPD genes demonstrated significant enrichment of pathways related to the extracellular matrix (ECM), ECM binding and collagen (See Table S8). Spearman correlation of the GSVA enrichment scores of these pathways with single cell
proportions showed a positive correlation between basal and ciliated cells with ECM and collagen related pathways, and a negative correlation with goblet cells (see Figure S5).

Within the StringDB network, FN1 and VEGFA were key regulatory genes, both with 25 connections (See Table S9). Fifteen connecting genes overlapped between the two key regulatory genes (SPARC, TWIST1, LIF, SEMA3E, FOS, PTTLH, PECAM1, ABCB1, BDNF, CEACAMS, CX3CR1, CYR61, DCN, DKK1 and EGR1). Both key regulatory networks were related to tissue development, whilst the network surrounding FN1 was also involved with collagen and ECM binding. The overlapping networks showed enrichment in the regulation of developmental processes, tissue development, and cell division.

Functionally related to FN1 was a network surrounding DCN. DCN, lower expressed in sCOPD, was at the centre of a cluster of 8 genes (COL6A2, SPARCL1, C1QA, B3GAT1, FAGFA, FN1, SPARC and MGP), together involved in collagen binding and the ECM (Figure 4). We used the human single cell lung atlas to determine which cell types expressed FN1 and VEGFA most. FN1 is expressed by many cells in the lung, including (alveolar) fibroblasts, (alveolar) macrophages, endothelial, and smooth muscle cells, and is lowly expressed in epithelial cells, whereas VEGFA is mostly expressed in airway basal cells, goblet cells and to a lesser extent in other cells. (Figure S6).

Other regulatory genes in the network with more than 10 connections were FOS (15 connections), EGR1 (14 connections), FCGR3A (13 connections), BDNF and NR4A1 (both 12 connections). The StringDB analysis of the 83 genes also represented in the nose showed a very similar network with again FN1 and VEGFA as the key genes with the most connections (both 16 connected genes) and enrichment of pathways involved in the ECM, and collagen binding, cell adhesion and cell signalling, all in the same direction as the same enriched pathway in the bronchus (Figure S7).

**Bootstrapping in the sCOPD subgroup**

Since the number of subjects in the sCOPD group was much larger compared to mCOPD and non-COPD controls, we performed bootstrapping in the sCOPD subgroup. To this end, we randomly
sampled 23 sCOPD cases with sCOPD and compared their expression of the 435 sCOPD-associated genes to the expression in non-COPD controls. We performed 1000 iterations and found that on average, 63.2% of the 435 genes were replicated.

Discussion

We identified a specific bronchial epithelial gene expression signature for sCOPD, consisting of 219 genes. This sCOPD signature is different from mCOPD, supporting our hypothesis that sCOPD represents a distinct disease phenotype. Pathway analyses demonstrated that sCOPD-associated genes are mainly involved in immune response, developmental processes, and ECM binding. Protein-interaction networks indicate VEGFA and FN1 as potential key drivers in sCOPD. Additionally, the gene signature that was lower in sCOPD in bronchial brushes was also represented in matched nasal brushings as well as nasal samples from an independent sCOPD cohort. Of interest, the signature-related gene set that was present in both nasal cohorts was again centred around VEGFA and FN1.

The two key genes driving the sCOPD gene signature were VEGFA and FN1, both lower expressed in sCOPD. VEGFA, expressed in basal and goblet cells, is a key growth factor for building lung architecture that needs VEGFR2, expressed in endothelial cells, to form pulmonary capillaries, VEGFA is less expressed in the lower respiratory tract of smokers and even lower in smokers with COPD. It was previously shown that loss of VEGFA leads to endothelial cell apoptosis and is associated with emphysema, a hallmark of sCOPD. This might be due to reduced blood supply from small capillaries associated with loss of alveolar septa. Our findings suggest a role for VEGFA in the development of sCOPD, not only in the alveoli, as shown by previous studies, but also in the bronchus and the nose. Therefore, the lower VEGFA expression in bronchial and nasal brushes could reflect similar changes in the parenchyma leading to emphysematous destruction and lack of alveolar and endothelial repair due to the lack of VEGFA.
The other key regulatory gene identified in our sCOPD gene signature was FN1. FN1 is a glycoprotein expressed by many cells in the lung, including fibroblasts, monocytes, endothelial, and smooth muscle cells, and is lowly expressed in epithelial cells. FN1 is important during the development of the lung, barely detectable during adulthood in healthy lungs, but highly upregulated during tissue repair. Therefore, lower FN1 expression in the bronchial and nasal brushes in sCOPD patients may reflect a disturbed epithelial repair response and may reflect similar events in the parenchyma leading to a lack of repair leading to emphysematous tissue destruction present in these patients. Of interest is the connection between FN1 and VEGFA in our network since it has been shown that FN1, when bound to VEGFA, is necessary to promote VEGFA-induced endothelial cell proliferation and migration. Next to FN1 and VEGFA, the protein-interaction network included 13 genes involved in ECM organisation and 11 specifically involved in collagen binding (HMCN2, LGALS1, C1QB, C1QA, ANXA6, SPARCL1, DCN, MGP, SPARC, COCH and FN1). Five of these genes (C1QB, ANXA6, DCN, MGP, and SPARC) were previously found to interact with FN1, and two genes, SPARC and DCN, also interacted with VEGFA. SPARC, lower expressed in sCOPD, encodes for the secreted protein osteonectin and is lowly expressed in basal cells and mainly in endothelial cells, fibroblasts, and macrophages. It binds to VEGFA and interferes with its binding to VEGFR1 and, in doing so, inhibits the proliferation of endothelial cells. SPARC also has a role in the regulation of secretion rates of fibronectin. One of the genes that interact with SPARC, FN1 and VEGFA is DCN. This gene encodes for decorin, is lower expressed in sCOPD and is mainly expressed in basal cells and fibroblasts. It is a protein with an important role in collagen crosslinking and fibrillogenesis. Decreased DCN expression in the lung might affect collagen tensile strength and binding of ECM proteins, resulting in changed cell fate and function. Lastly, ANXA6, lower expressed in sCOPD, was previously found to inhibit the secretion of FN1, adding another mechanism to control fibronectin function and angiogenesis.
Clearly, part of the sCOPD-related gene expression changes is also present in the nose, providing support for the use of nasal brushes as a proxy for the lung. Here we identified a gene signature in the nose that reflects bronchial changes specific to sCOPD. This study confirms and extends what previous findings show that severe COPD can be differentiated from non-COPD controls using nasal gene expression. Furthermore, our finding of an overlapping signature in the bronchus and the nose supports the theory of a single transcriptional profile throughout the airways. However, it should be noted that not all genes are concordantly expressed. We only replicated this finding for genes with a lower expression in sCOPD. Our findings are important as it suggests that the nose may serve as an easily accessible biomarker in COPD, at least for a subset of genes and biological pathways.

There were some limitations to the current study. One of the limitations was the size of the non-COPD and mCOPD groups, which could have influenced the results. However, bootstrapping of the sampling of the sCOPD group showed similar results as, on average, 63.2% of the genes could be replicated. Furthermore, most patients with sCOPD used high ICS doses, which may affect gene expression levels. Because of the uneven distribution, with non-COPD controls not using ICS and mild-moderate COPD patients using much lower doses, it is difficult to adjust for this variable in the statistical analyses. To account for this possible confounding factor, we removed all genes previously shown to be sensitive to treatment with corticosteroids, as previously identified in a placebo-controlled longitudinal study. This way, we made sure that the higher ICS dose in sCOPD did not lead to the identification of sCOPD-associated differentially expressed genes. However, we could not rule out the possibility that we removed genes relevant to the development of sCOPD. Additionally, in line with previous observations, a higher percentage of the sCOPD patients was female. Although we adjusted for this in our linear models, we cannot entirely rule out the possibility that some of the observed sCOPD-associated gene expression differences are due to an imbalance in sex. Lastly, although transcriptomic data are a good indicator for changes on the protein level in most
cases, it is not a replacement. Future studies are needed to investigate the protein levels of FN1 and VEGFA, which could be done in nasal epithelial lining fluids collected from severe COPD patients and controls. This exploration would further clarify how FN1 and VEGFA are involved in the pathogenesis of severe COPD. In addition, it would be of interest to assess their utility as biomarkers for disease severity and progression.

In conclusion, we found a unique sCOPD gene signature that was indicative of an abnormal epithelial repair response, impaired fibroblast function and decreased angiogenesis, which was retained throughout the airways in the nose. This supports the hypothesis that sCOPD comprises a partly different pathology compared to the majority of patients with mCOPD driving the specific disease phenotype. Moreover, as part of the sCOPD-related gene expression, changes are also present in the nose, supporting the potential for biomarker development based on nasal brushes in COPD.
References


<table>
<thead>
<tr>
<th></th>
<th>Non-COPD controls</th>
<th>COPD Patients</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mCOPD</td>
<td>sCOPD</td>
</tr>
<tr>
<td>Gender = male (%)</td>
<td>12 (52)</td>
<td>23</td>
<td>123</td>
</tr>
<tr>
<td>Age (Years) (median (IQR))</td>
<td>60 (52,63)</td>
<td>62 (57,65)</td>
<td>60 (56,64)</td>
</tr>
<tr>
<td>Smoking status = Ex-smoker (%)</td>
<td>23 (100)</td>
<td>23 (100)</td>
<td>122 (99)</td>
</tr>
<tr>
<td>Packyears (median (IQR))</td>
<td>28 (14,48)</td>
<td>56 (23,74)</td>
<td>36 (30,46)</td>
</tr>
<tr>
<td>Years of cessation (median (IQR))</td>
<td>9 (2,21)</td>
<td>9 (5,18)</td>
<td>5 (2,9)</td>
</tr>
<tr>
<td>FEV1 %pred (median (IQR))</td>
<td>102 (90,110)</td>
<td>77 (68,89)</td>
<td>24 (20,29)</td>
</tr>
<tr>
<td>FEV1/FVC (%) (median (IQR))</td>
<td>0.73 (0.70,0.77)</td>
<td>0.55 (0.5,0.59)</td>
<td>0.28 (0.24,0.32)</td>
</tr>
<tr>
<td>RV %pred (median (IQR))</td>
<td>106 (95,111)</td>
<td>110 (102,115)</td>
<td>239 (215,260)</td>
</tr>
<tr>
<td>ICS before inclusion = true (%)</td>
<td>0 (0.0, 0)</td>
<td>11 (48,0)</td>
<td>108 (99, 14 NA)</td>
</tr>
<tr>
<td>SGRQ Part 1 Q2 (median (IQR))</td>
<td>5 (5,5)</td>
<td>5 (3,5)</td>
<td>4 (2,5)</td>
</tr>
<tr>
<td>History of exacerbations in last year</td>
<td>N/A</td>
<td>0 (0,0)</td>
<td>2 (1,3)</td>
</tr>
</tbody>
</table>

FEV₁: forced expiratory volume in 1 second; FVC: forced vital capacity; ICS: Inhaled Corticoid Steroids. Values were calculated excluding the missing values.

Table 1
<table>
<thead>
<tr>
<th>Ensembl gene ID</th>
<th>HGNC symbol</th>
<th>Log(FC)</th>
<th>Log(CPM)</th>
<th>F</th>
<th>P-Value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000237523</td>
<td>LINCO0857</td>
<td>1.41</td>
<td>9.11</td>
<td>80.38</td>
<td>1.35E-15</td>
<td>1.24E-13</td>
</tr>
<tr>
<td>ENSG00000181588</td>
<td>MEX3D</td>
<td>1.44</td>
<td>11.90</td>
<td>81.52</td>
<td>9.28E-16</td>
<td>1.24E-13</td>
</tr>
<tr>
<td>ENSG00000105388</td>
<td>CEACAM5</td>
<td>1.87</td>
<td>16.44</td>
<td>78.04</td>
<td>2.90E-15</td>
<td>2.00E-13</td>
</tr>
<tr>
<td>ENSG00000228594</td>
<td>FNDC10</td>
<td>1.23</td>
<td>9.11</td>
<td>74.92</td>
<td>8.19E-15</td>
<td>3.77E-13</td>
</tr>
<tr>
<td>ENSG00000170537</td>
<td>TMC7</td>
<td>1.35</td>
<td>9.72</td>
<td>74.97</td>
<td>8.06E-15</td>
<td>3.77E-13</td>
</tr>
<tr>
<td>ENSG00000171208</td>
<td>NETO2</td>
<td>1.67</td>
<td>10.67</td>
<td>69.99</td>
<td>4.35E-14</td>
<td>1.33E-12</td>
</tr>
<tr>
<td>ENSG00000206075</td>
<td>SERPINB5</td>
<td>1.34</td>
<td>14.10</td>
<td>68.89</td>
<td>6.35E-14</td>
<td>1.75E-12</td>
</tr>
<tr>
<td>ENSG00000178363</td>
<td>CALML3</td>
<td>2.48</td>
<td>11.34</td>
<td>68.40</td>
<td>7.52E-14</td>
<td>1.89E-12</td>
</tr>
<tr>
<td>ENSG00000205277</td>
<td>MUC12</td>
<td>1.50</td>
<td>11.07</td>
<td>66.74</td>
<td>1.34E-13</td>
<td>3.08E-12</td>
</tr>
<tr>
<td>ENSG00000137726</td>
<td>FXYD6</td>
<td>-1.58</td>
<td>8.63</td>
<td>66.43</td>
<td>1.49E-13</td>
<td>3.16E-12</td>
</tr>
<tr>
<td>ENSG00000204136</td>
<td>GGTA1P</td>
<td>-1.34</td>
<td>8.11</td>
<td>57.48</td>
<td>3.62E-12</td>
<td>4.17E-11</td>
</tr>
<tr>
<td>ENSG00000164949</td>
<td>GEM</td>
<td>-1.57</td>
<td>10.22</td>
<td>55.19</td>
<td>8.41E-12</td>
<td>8.60E-11</td>
</tr>
<tr>
<td>ENSG00000106034</td>
<td>CPED1</td>
<td>-1.24</td>
<td>8.67</td>
<td>47.87</td>
<td>1.32E-10</td>
<td>9.87E-10</td>
</tr>
<tr>
<td>ENSG00000120457</td>
<td>KCNJ5</td>
<td>-1.33</td>
<td>8.34</td>
<td>42.97</td>
<td>8.93E-10</td>
<td>5.48E-09</td>
</tr>
<tr>
<td>ENSG00000112715</td>
<td>VEGFA</td>
<td>-1.23</td>
<td>13.97</td>
<td>39.65</td>
<td>3.35E-09</td>
<td>1.52E-08</td>
</tr>
<tr>
<td>ENSG00000176049</td>
<td>JAKMIP2</td>
<td>-1.18</td>
<td>8.64</td>
<td>39.43</td>
<td>3.67E-09</td>
<td>1.63E-08</td>
</tr>
<tr>
<td>ENSG00000147443</td>
<td>DOK2</td>
<td>-1.28</td>
<td>9.16</td>
<td>39.02</td>
<td>4.32E-09</td>
<td>1.86E-08</td>
</tr>
<tr>
<td>ENSG00000117009</td>
<td>KMO</td>
<td>-1.33</td>
<td>8.93</td>
<td>37.56</td>
<td>7.82E-09</td>
<td>3.13E-08</td>
</tr>
<tr>
<td>ENSG00000147138</td>
<td>GPR174</td>
<td>-1.45</td>
<td>9.32</td>
<td>37.38</td>
<td>8.41E-09</td>
<td>3.32E-08</td>
</tr>
</tbody>
</table>

Table 2
Figure 1
Figure 2
Figure 3

A  Matched Nasal (Less expressed signature)
   
B  Matched Nasal (More expressed signature)
   
C  Independent nasal COPD cohort (Less expressed signature)
   
D  Independent nasal COPD cohort (More expressed signature)
Supplementary Methods

The local medical ethics committees approved the study, and all subjects gave their written informed consent (the SHERLOCK study was approved by the medical ethics committee of the University of Groningen/University Medical Center Groningen, METc 2016/572 and METc 2014/102).

RNA-Sequencing

Total RNA was isolated from the bronchial and nasal brushings using QIAgen AllPrep DNA/RNA/miRNA Universal Mini Kit per manufacturer’s instructions. Ribosomal RNA was removed using the RiboZero Magnetic Gold kit. RNA sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA method. Libraries were paired-end sequenced. FastQC (version 0.11.7) was used to determine the raw RNA sequencing data quality. StarAligner (version 2.73a) was used to index and align the raw sequencing data to the human reference genome (version GrCh38). Ensembl (release 100) was used as the gene annotation database. We checked if the expected sexes and recorded sexes matched and if the mapped read counts were similar across samples of the same type. Lastly, we checked for any outliers in a principal component analysis.

Removal of cell effect using Cellular Deconvolution

To estimate the proportions of cell types in the bronchial brushes, we performed cellular deconvolution using scRNA-Seq signatures from 600 informative genes from a single cell dataset obtained from bronchial biopsies, including 15 cell types (alveolar macrophages, not alveolar macrophages, arterial endothelial cells, proliferating basal cells, not proliferating basal cells, ciliated cells, dendritic cells, fibroblasts, goblet cells, ionocytes, B cells, T cells, mast cells, monocytes, and submucosal secretory cells) as previously described (40). We estimated the cell proportions using the non-negative least squares (NNLS) method on our counts per million (CPM) normalised gene expression data.
A new differential expression analysis was performed on the uniquely differentially expressed genes for sCOPD, comparing non-COPD controls versus (very) sCOPD, in which we also corrected for the cell proportions that made up more than two per cent of the cell populations in the samples on average in addition to the previous parameters and were different between COPD groups.

**Classification of ICS-sensitive genes**

Patients with more sCOPD often used high doses of inhaled corticosteroids and/or systemic prednisolone, providing a potential confounder in our study. To account for this confounder, we identified ICS responsive genes in the airways of COPD patients using the Groningen Leiden Universities and Corticosteroids in Obstructive Lung Disease study (GLUCOLD, ClinicalTrials.gov NCT00158847) as described previously (41) and removed these ICS-responsive genes from our sCOPD gene signature. Bronchial biopsy RNA-Seq data of patients (n=19) at baseline and after six months of ICS treatment was compared using EdgeR while correcting for age, sex, current smoking status and packyears. A nominal p-value less than 0.05 was considered statistically significant.

**Replication in nasal brushes**

To investigate if the identified genes could be replicated in nasal brushes, we used matching nasal brushes of the same patients. We used gene set variation analysis (GSVA, version 1.42) to estimate the changes of the negatively and positively correlated genes separately per disease group (42). Next, we checked if the same genes could be replicated using the same method in an independent sCOPD cohort of nasal brushings (FAIR and NORM cohorts, ClinicalTrials.gov: NCT01351792 and NCT00848406, see below) (43).

We then performed an unbiased differential expression analysis using the matched nasal samples and the independent nasal brushings cohort separately, comparing sCOPD and non-COPD controls for the 219 genes unique for sCOPD, and performed a meta-analysis.
Pathway analysis and protein interaction on the genes unique for sCOPD and replicated genes

Protein interaction analysis and pathway enrichment within our sCOPD gene signature in the bronchus, as well as in the replicated in the nose, was performed using StringDB. We used the default settings of the program to construct the gene network and used the following databases for functional enrichment analysis: Gene Ontology (biological processes, molecular functions, cellular components) and Kegg pathways. The original 219 genes unique for sCOPD were used as input for the StringDB analysis. Additionally, we performed a separate StringDB analysis on the genes that were FDR significant in the meta-analysis of the nasal brushings.

Sampling of the initial differential expression analysis

Our primary dataset, the SHERLOCK study, had approximately five times more sCOPD subjects than mCOPD or non-COPD controls. To make sure that the size of these groups did not affect the outcome of the study, we chose to redo the first differentially expression analysis between non-COPD controls and sCOPD subject, with randomly sampled equally sized groups (n=23). We did this a thousand times, and calculated how many genes on average replicated.

Description of the COPD replication cohort for nasal brushings

The FAIR cohort (ClinicalTrials.gov: NCT01351792) consists of COPD patients who participated in a multicenter, randomised, longitudinal study. Participants were excluded if they were diagnosed with asthma, were pregnant, were treated with long-term oxygen therapy, had a clinically unstable concurrent disease (as judged by the investigator), had a COPD exacerbation within two months prior to the first study visit and reversibility of the FEV1>15% and >200mL of initial FEV1.

Description of the non-COPD replication cohort for the nasal brushings

The NORM cohort (ClinicalTrials.gov: NCT00848406) consists of non-COPD participants who participated in a cross-sectional study and had a pulmonary function, defined by an FEV1/FVC >
lower limit of normal, absence of bronchial hyperresponsiveness and reversibility of the FEV1% predicted to salbutamol <10%. Exclusion criteria were the use of inhaled or oral corticosteroids, upper respiratory tract infection, clinically unstable concurrent disease (as judged by the investigator) or pregnancy.
Supplementary Figures

**Figure S1** Flowchart of statistical methods.
Figure S2) Heatmap of differential gene expression between non-COPD vs sCOPD.
Figure S3) Heatmap of differential gene expression between non-COPD vs mCOPD.
Figure S4) Heatmap of differential gene expression between mCOPD vs sCOPD.
Figure S5) Spearman correlation estimate and p-value of the GSVA score of the significant pathways versus the cellular proportions in bronchial brushes and nasal brushes.
Figure S6) Gene expression of FN1 and VEGFA in an independent single-cell dataset (human lung cell atlas).
Figure S7) StringDB analysis of the genes unique for sCOPD that were replicated after a meta-analysis of matched nasal brushings and an independent nasal brushings cohort.