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 >200mLof 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.

- RNA-Seq
- Single cell reference dataset
  - Cellular Deconvolution
  - Cell proportions
  - High positive correlation
  - Relative abundance
  - Genes that were affected by cell populations were considered not relevant since these genes could be cell type specific

Severe vs control
- Linear Discriminant Analysis

Control vs Moderate
- Linear Discriminant Analysis

Severe vs Moderate
- Linear Discriminant Analysis

Genes that changed in control vs severe should also change between moderate vs severe COPD. If not, they were not unique for severe COPD.

ICS effect
- Genes that were previously found to be affected by inhaled Corticosteroids (ICS) were considered not relevant since all severe COPD patients were using high dosages of ICS.
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.