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

van Nijnatten, Jos; Faiz, Alen; Timens, Wim; Guryev, Victor; Slebos, Dirk-Jan; Klooster, Karin; Hartman, Jorine E; Kole, Tessa; Choy, David F; Chakrabarti, Arindam

Published in:
ERJ Open Research

DOI:
10.1183/23120541.00354-2023

Publication date:
2023

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

Jos van Nijnatten[1,2,3], Alen Faiz[2,3,4], Wim Timens[1,2], Victor Guryev[5], Dirk-Jan Slebos[2,4], Karin Klooster[2,4], Jorine E. Hartman[2,4], Tessa Kole[2,4], David F. Choy[6], Arindam Chakrabarti[6], Michele Grimbaldeston[6], Carrie M. Rosenberger[6], Huib Kerstjens[4], Corry-Anke Brandsma[1,2,7] and Maarten van den Berge[2,4,7]

[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, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands. [3]University of Technology Sydney, Respiratory Bioinformatics and Molecular Biology, Sydney, NSW, Australia. [4]University of Groningen University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands. [5]University of Groningen University Medical Center Groningen, European Research Institute for the Biology of Ageing, Groningen, the Netherlands. [6]Genentech Inc., South San Francisco, CA, USA. [7]These authors contributed equally.

Corresponding author: Jos van Nijnatten (j.l.l.van.nijnatten@umcg.nl)

Shareable abstract (@ERSpublications)
Severe COPD has a unique bronchial gene signature, with key roles for *VEGFA* and *FN1*, which is retained in the upper airways. https://bit.ly/459KNlv

Cite this article as: van Nijnatten J, Faiz A, Timens W, *et al*. A bronchial gene signature specific for severe COPD that is retained in the nose. *ERJ Open Res* 2023; 9: 00354-2023 [DOI: 10.1183/23120541.00354-2023].

This article has an editorial commentary: https://doi.org/10.1183/23120541.00647-2023

Received: 30 May 2023
Accepted: 9 Aug 2023

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. The aim of the present study was 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 10 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

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 ∼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-COPD. 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, ClinicalTrails.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 (Global Initiative for Chronic Obstructive Lung Disease (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 2 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 [18, 19]. All patients were fully characterised, *i.e.*, lung function, computed tomography 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-test 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 $\leqslant 0.05$ 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 sex, age and smoking pack-years. An FDR of <0.05 and an absolute 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 *versus* non-COPD controls 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 inhaled corticosteroid (ICS) doses, we removed steroid-sensitive genes, as previously identified in the GLUCOLD

study, from our analysis [20]. 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 pack-years (non-COPD: 31.1±20.6 pack-years (mean±SD), 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 1a and supplementary 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 supplementary table S1). A volcano plot and a heatmap are depicted in figure 1b and supplementary 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 1c and supplementary figure S4, respectively. The top three 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 (median: 86.5% (IQR: 74.2–93.7)), followed by ciliated cells (11.7% (5.9–19.9)) and the basal cells (0.0% (0.0–2.6)) (figure 2a). When comparing sCOPD to non-COPD controls, we found that there was a

TABLE 1 Patient demographics

	Non-COPD controls	COPD patients	
		mCOPD	sCOPD
Patients n	23	23	123
Sex, male, n (%)	12 (52)	18 (78)	36 (29)
Age years, median (IQR)	60 (52–63)	62 (57–65)	60 (56–64)
Ex-smokers, n (%)	23 (100)	23 (100)	122 (99)
Pack-years, median (IQR)	28 (14–48)	56 (23–74)	36 (30–46)
Years of cessation, median (IQR)	9 (2–21)	9 (5–18)	5 (2–9)
FEV_1 % pred, median (IQR)	102 (90–110)	77 (68–89)	24 (20–29)
FEV_1/FVC %, median (IQR)	0.73 (0.70–0.77)	0.55 (0.5–0.59)	0.28 (0.24–0.32)
RV % pred, median (IQR)	106 (95–111)	110 (102–115)	239 (215–260)
ICS before inclusion, n (%, NA)	0 (0.0, 0)	11 (48, 0)	108 (99, 14)
SGRQ Part 1 Q2, median (IQR)	5 (5–5)	5 (3–5)	4 (2–5)
History of exacerbations in last year, median (IQR)	N/A	0 (0–0)	2 (1–3)

Values were calculated excluding the missing values. mCOPD: mild to moderate COPD; sCOPD: severe COPD; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; RV: residual volume; ICS: inhaled corticosteroids; NA: not available; SGRQ: St George's Respiratory Questionnaire.

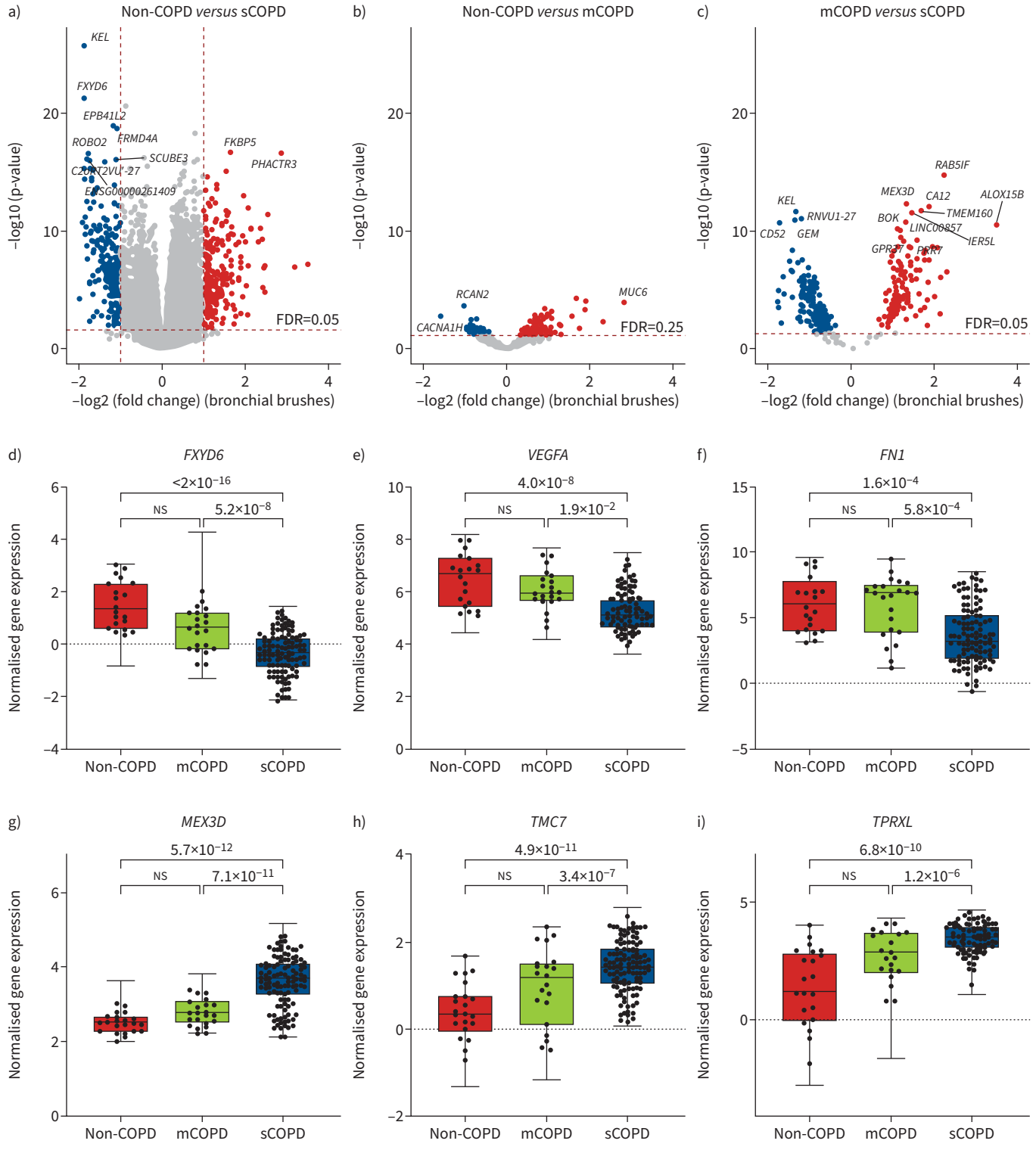

FIGURE 1 Differential gene expression of bronchial brushes of severe COPD (sCOPD) and non-COPD controls. a) Differential expression analysis between sCOPD compared to non-COPD controls. Biased differential analysis between b) mild to moderate COPD (mCOPD) compared with non-COPD controls and c) sCOPD and mCOPD. d–f) The three higher and g–i) lower expressed genes between sCOPD compared to non-COPD controls.

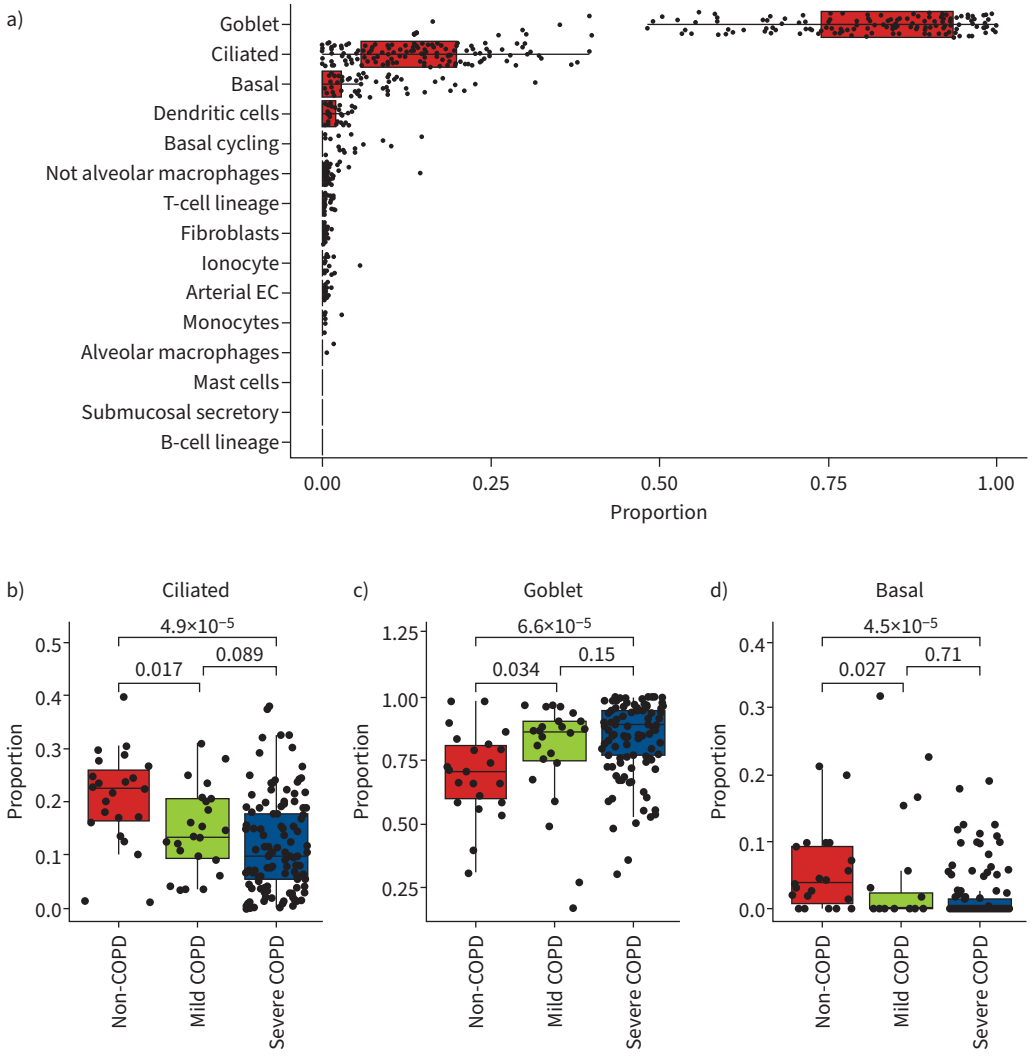

FIGURE 2 Cellular deconvolution and cell proportions. Cellular deconvolution of bronchial brushings using non-negative least squares regression (NNLS). a) Overall composition of bronchial brushings. Separation of cell-type composition by disease status for b) ciliated, c) goblet and d) basal. A Kruskal–Wallis test was conducted over all the groups and a Mann–Whitney U-test was conducted between the groups. EC: endothelial cells.

significantly lower proportion of ciliated and basal and a higher proportion of goblet cells (figure 2b–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 supplementary table S2.)

The effect of inhaled corticosteroids

We used an existing dataset of bronchial biopsies obtained from COPD patients before and after 6 months of treatment with either ICS ± long-acting β_2-agonists (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 supplementary table S3. These 43 genes were therefore removed from our sCOPD gene signature (see supplementary 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*, *LINC00857*, *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 supplementary 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 gene set variation (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 3a and 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 supplementary 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 3c and 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 five genes: *EPB41L2*, *FRMD4A*, *GGTA1P*, *GEM* and *CPED1*), and 41 genes were higher expressed (top five genes: *MEX3D*, *TMEM132B*, *PCSK1N*, *PRR7* and *MESP1*) (see supplementary 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 supplementary 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 supplementary figure S5).

TABLE 2 Top 20 genes more and less expressed in severe COPD (sCOPD) that were specific for sCOPD, not influenced by basal, ciliated and secretory cell types and 6 months of inhaled corticosteroid usage

Ensembl gene ID	HGNC symbol	Log(FC)	Log(CPM)	F	p-value	FDR
ENSG00000237523	*LINC00857*	1.41	9.11	80.38	1.35×10^{-15}	1.24×10^{-13}
ENSG00000181588	*MEX3D*	1.44	11.90	81.52	9.28×10^{-16}	1.24×10^{-13}
ENSG00000105388	*CEACAM5*	1.87	16.44	78.04	2.90×10^{-15}	2.00×10^{-13}
ENSG00000228594	*FNDC10*	1.23	9.11	74.92	8.19×10^{-15}	3.77×10^{-13}
ENSG00000170537	*TMC7*	1.35	9.72	74.97	8.06×10^{-15}	3.77×10^{-13}
ENSG00000180438	*TPRXL*	1.92	11.73	72.28	1.99×10^{-14}	7.86×10^{-13}
ENSG00000171208	*NETO2*	1.67	10.67	69.99	4.35×10^{-14}	1.33×10^{-12}
ENSG00000206075	*SERPINB5*	1.34	14.10	68.89	6.35×10^{-14}	1.75×10^{-12}
ENSG00000178363	*CALML3*	2.48	11.34	68.40	7.52×10^{-14}	1.89×10^{-12}
ENSG00000205277	*MUC12*	1.50	11.07	66.74	1.34×10^{-13}	3.08×10^{-12}
ENSG00000137726	*FXYD6*	−1.58	8.63	66.43	1.49×10^{-13}	3.16×10^{-12}
ENSG00000204136	*GGTA1P*	−1.34	8.11	57.48	3.62×10^{-12}	4.17×10^{-11}
ENSG00000164949	*GEM*	−1.57	10.22	55.19	8.41×10^{-12}	8.60×10^{-11}
ENSG00000106034	*CPED1*	−1.24	8.67	47.87	1.32×10^{-10}	9.87×10^{-10}
ENSG00000120457	*KCNJ5*	−1.33	8.34	42.97	8.93×10^{-10}	5.48×10^{-9}
ENSG00000112715	*VEGFA*	−1.23	13.97	39.65	3.35×10^{-9}	1.52×10^{-8}
ENSG00000176049	*JAKMIP2*	−1.18	8.64	39.43	3.67×10^{-9}	1.63×10^{-8}
ENSG00000147443	*DOK2*	−1.28	9.16	39.02	4.32×10^{-9}	1.86×10^{-8}
ENSG00000117009	*KMO*	−1.33	8.93	37.56	7.82×10^{-9}	3.13×10^{-8}
ENSG00000147138	*GPR174*	−1.45	9.32	37.38	8.41×10^{-9}	3.32×10^{-8}

HGNC: HUGO Gene Nomenclature Committee gene symbol; FC: fold change; CPM: counts per million; FDR: false discovery rate.

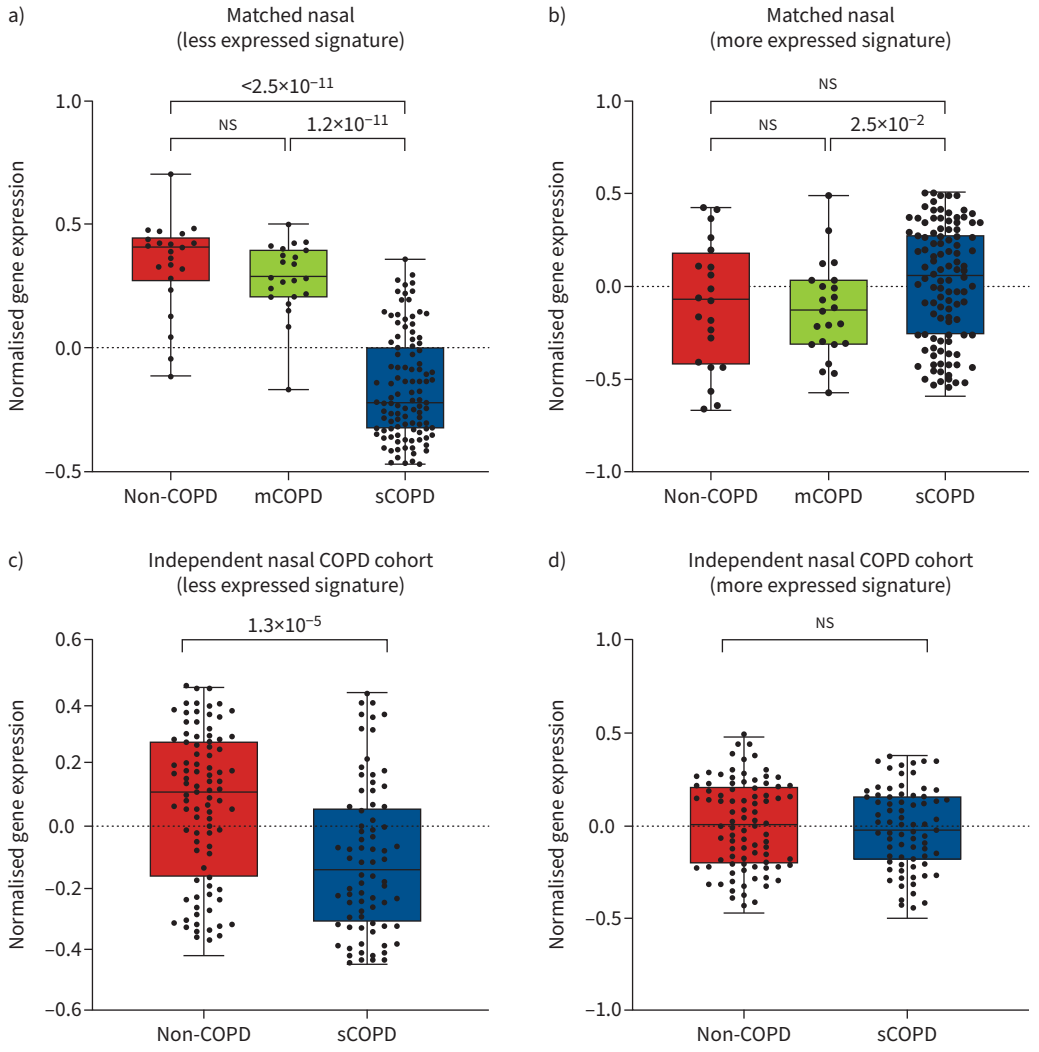

FIGURE 3 Reflection of the bronchial severe COPD (sCOPD) signature in the nose. Comparison of the sCOPD signature a) less and b) more expressed in matched nasal brushings. Reflection of the sCOPD signature c) less and d) more expressed in nasal brushing from sCOPD (n=76) and non-COPD controls (n=92). Reflection of the sCOPD signature in normal COPD (n=38) and non-COPD controls (n=30). A Mann–Whitney U-test was conducted ($p<0.05$ was considered significant). mCOPD: mild to moderate COPD.

Within the StringDB network, *FN1* and *VEGFA* were key regulatory genes, both with 25 connections (see supplementary table S9). 15 connecting genes overlapped between the two key regulatory genes (*SPARC*, *TWIST1*, *LIF*, *SEMA3E*, *FOS*, *PTHLH*, *PECAM1*, *ABCB1*, *BDNF*, *CEACAM5*, *CX3CR1*, *CYR61*, *DCN*, *DKK1* and *EGR1*). Both key regulatory networks were related to tissue development, while 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 eight 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 [21]. *FN1* is expressed by many cells in the lung, including (alveolar) fibroblasts, (alveolar) macrophages, and 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 (supplementary figure S6).

Other regulatory genes in the network with >10 connections were *FOS* (15 connections), *EGR1* (14 connections), *FCGR3A* (13 connections), and *BDNF* and *NR4A1* (both 12 connections). The StringDB

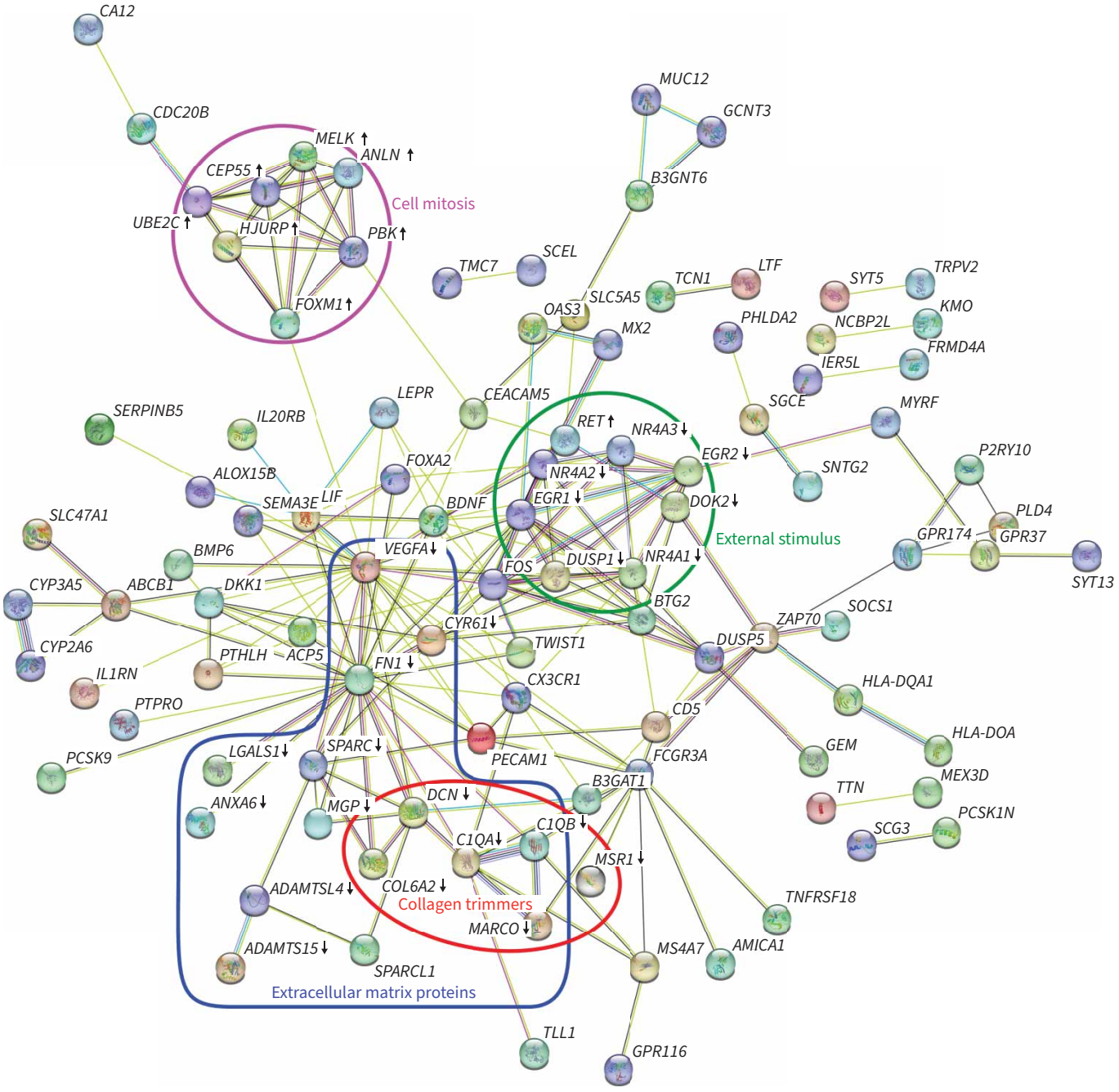

FIGURE 4 StringDB analysis of the genes unique for severe COPD (sCOPD). Arrows indicate if the gene was more (↑) or less (↓) expressed in sCOPD. The purple grouping contains genes involved in cell mitosis; the green grouping contains genes involved in external stimuli; the blue grouping contains genes involved in the extracellular matrix; the red grouping contains genes involved with collagen trimmers.

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 (supplementary 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 [21], is a key growth factor for building lung architecture [22] that needs *VEGFR2*, expressed in endothelial cells, to form pulmonary capillaries [23], VEGFA is less expressed in the lower respiratory tract of smokers and even lower in smokers with COPD [24]. It was previously shown that loss of *VEGFA* leads to endothelial cell apoptosis and is associated with emphysema [24–28], 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 [24], 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, and endothelial and smooth muscle cells, and is lowly expressed in epithelial cells [21]. *FN1* is important during the development of the lung, barely detectable during adulthood in healthy lungs [29], but highly upregulated during tissue repair [29]. 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 [30]. 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* [31–33], 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 [21]. It binds to *VEGFA* and interferes with its binding to *VEGFR1* [34] and, in doing so, inhibits the proliferation of endothelial cells. *SPARC* also has a role in the regulation of secretion rates of fibronectin [35, 36]. 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 [21]. It is a protein with an important role in collagen cross-linking 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* [37], 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 [38, 39]. 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 [40]. 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 [41]. 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.

Provenance: Submitted article, peer reviewed.

Conflict of interest: W. Timens declares consultancy payments to their institution from Merck Sharp & Dohme and Bristol Myers Squibb, in the 36 months prior to manuscript submission; as well as board membership of the Dutch Society of Pathology and membership of the Council for Research and Innovation of the Federation of Medical Specialists. D-J. Slebos declares grants or contracts to their institution from the following US companies in the 36 months prior to manuscript submission: PulmonX Corp., PneumRx/BTG/Boston Scientific, FreeFlowMedical, Nuvaira, PulmAir, GALA, CSA Medical and Apreo; consulting fees paid to their institution from PulmonX Corp., PneumRx/BTG/Boston Scientific, Nuvaira and Apreo; payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from PulmonX Corp, PneumRx/BTG/Boston Scientific and Nuvaira; support for attending meeting and/or travel from PulmonX Corp., PneumRx/BTG/Boston Scientific and Nuvaira; and receipt by their institution of study material/devices from PulmonX Corp., PneumRx/BTG/Boston Scientific, FreeFlowMedical, Nuvaira, PulmAir, GALA and CSA Medical, all in the 36 months prior to manuscript submission. D.F. Choy is an employee of Genentech, Inc., a member of the Roche Group, and is a co-inventor, as part of that employment, for pending patents related to the diagnosis and treatment of chronic respiratory diseases. They also hold stocks and/or options in Roche. A. Chakrabarti is an employee of Genentech, Inc., a member of the Roche Group. M. Grimbaldeston is an employee of Genentech, Inc., a member of the Roche Group, and has received stock options. C.M. Rosenberger is an employee of Genentech, Inc., a member of the Roche Group, and has received support for attending conferences from their employer. They hold stocks and stock options in Roche. C-A. Brandsma declares a research grant from Genentech, related to the current manuscript. M. van den Berge declares a research grant from Genentech, related to the current manuscript. All other authors declare no competing interests.

References

1 Vogelmeier CF, Román-Rodríguez M, Singh D, *et al.* Goals of COPD treatment: focus on symptoms and exacerbations. *Respir Med* 2020; 166: 105938.
2 Eisner MD, Anthonisen N, Coultas D, *et al.* An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 182: 693–718.
3 Alwan A Global Status Report on Non-Communicable Diseases. Geneva, World Health Organization, 2010.
4 Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* 2004; 364: 709–721.
5 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977; 1: 1645–1648.
6 Goldklang MP, Marks SM, D'Armiento JM. Second hand smoke and COPD: lessons from animal studies. *Front Physiol* 2013; 4: 30.

7 Kurmi OP, Semple S, Simkhada P, *et al.* COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax* 2010; 65: 221–228.
8 Hoffmann D, Hoffmann I. Letters to the Editor – Tobacco smoke components. *Beitr Tab Int/Contrib Tob Res* 1998; 18: 49–52.
9 Chang JT, Meza R, Levy DT, *et al.* Prediction of COPD risk accounting for time-varying smoking exposures. *PLoS One* 2021; 16: e0248535.
10 Raherison C, Girodet PO. Epidemiology of COPD. *Eur Respir Rev* 2009; 18: 213–221.
11 Lange P, Groth S, Nyboe GJ, *et al.* Effects of smoking and changes in smoking habits on the decline of FEV_1. *Eur Respir Soc* 1989; 2: 811–816.
12 Silverman EK, Chapman HA, Drazen JM, *et al.* Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 1998; 157: 1770–1778.
13 Silverman EK. Genetic Epidemiology of COPD. *Chest* 2002; 121: 3 Suppl, 1S–6S.
14 Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur Respir J* 2003; 22: 672–688.
15 Wouters EFM. The burden of COPD in the Netherlands: results from the confronting COPD survey. *Respir Med* 2003; 97: Suppl C, S51–S59.
16 Fromer L, Cooper CB. A review of the GOLD guidelines for the diagnosis and treatment of patients with COPD. *Int J Clin Pract* 2008; 62: 1219–1236.
17 Miravitlles M, Vogelmeier C, Roche N, *et al.* A review of national guidelines for management of COPD in Europe. *Eur Respir J* 2016; 47: 625–637.
18 Imkamp K, Berg M, Vermeulen CJ, *et al.* Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci. *J Allergy Clin Immunol* 2018; 142: 314–317.e15.
19 Reddy KD, Lan A, Boudewijn IM, *et al.* Current smoking alters gene expression and DNA methylation in the nasal epithelium of patients with asthma. *Am J Respir Cell Mol Biol* 2021; 65: 366–377.
20 van den Berge M, Steiling K, Timens W, *et al.* Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity. *Thorax* 2014; 69: 14–23.
21 Deprez M, Zaragosi LE, Truchi M, *et al.* A single-cell atlas of the human healthy airways. *Am J Respir Crit Care Med* 2020; 202: 1636–1645.
22 Laddha AP, Kulkarni YA. VEGF and FGF-2: promising targets for the treatment of respiratory disorders. *Respir Med* 2019; 156: 33–46.
23 Yamamoto H, Yun EJ, Gerber HP, *et al.* Epithelial-vascular cross talk mediated by VEGF-A and HGF signaling directs primary septae formation during distal lung morphogenesis. *Dev Biol* 2007; 308: 44–53.
24 Kawamoto T, Kanazawa H, Tochino Y, *et al.* Evaluation of the severity of small airways obstruction and alveolar destruction in chronic obstructive pulmonary disease. *Respir Med* 2018; 141: 159–164.
25 Soltani A, Walters EH, Reid DW, *et al.* Inhaled corticosteroid normalizes some but not all airway vascular remodeling in COPD. *Int J Chron Obstruct Pulmon Dis* 2016; 11: 2359–2367.
26 Kasahara Y, Tuder RM, Cool CD, *et al.* Endothelial cell death and decreased expression of vascular endothelial growth factor and vascular endothelial growth factor receptor 2 in emphysema. *Am J Respir Crit Care Med* 2001; 163: 737–744.
27 Kasahara Y, Tuder RM, Taraseviciene-Stewart L, *et al.* Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J Clin Invest* 2000; 106: 1311–1319.
28 Breen EC, Malloy JL, Tang K, *et al.* Impaired pulmonary defense against *Pseudomonas aeruginosa* in VEGF gene inactivated mouse lung. *J Cell Physiol* 2013; 228: 371–379.
29 Dean DC. Expression of the fibronectin gene. *Am J Respir Cell Mol Biol* 1989; 1: 5–10.
30 Wijelath ES, Rahman S, Namekata M, *et al.* Heparin-II domain of fibronectin is a vascular endothelial growth factor-binding domain: enhancement of VEGF biological activity by a singular growth factor/matrix protein synergism. *Circ Res* 2006; 99: 853–860.
31 Bing DH, Almeda S, Isliker H, *et al.* Fibronectin binds to the C1q component of complement. *Proc Natl Acad Sci USA* 1982; 79: 4198–4201.
32 Iruela-Arispe ML, Vernon RB, Wu H, *et al.* Type I collagen-deficient Mov-13 mice do not retain SPARC in the extracellular matrix: implications for fibroblast function. *Dev Dyn* 1996; 207: 171–183.
33 Schmidt G, Hausser H, Kresse H. Interaction of the small proteoglycan decorin with fibronectin. Involvement of the sequence NKISK of the core protein. *Biochem J* 1991; 280: 411–414.
34 Kupprion C, Motamed K, Sage EH. SPARC (BM-40, osteonectin) inhibits the mitogenic effect of vascular endothelial growth factor on microvascular endothelial cells. *J Biol Chem* 1998; 273: 29635–29640.
35 Kamihagi K, Katayama M, Ouchi R, *et al.* Osteonectin/SPARC regulates cellular secretion rates of fibronectin and laminin extracellular matrix proteins. *Biochem Biophys Res Commun* 1994; 200: 423–428.

36 Yusuf N, Inagaki T, Kusunoki S, *et al.* SPARC was overexpressed in human endometrial cancer stem-like cells and promoted migration activity. *Gynecol Oncol* 2014; 134: 356–363.

37 Garcia-Melero A, Reverter M, Hoque M, *et al.* Annexin A6 and late endosomal cholesterol modulate integrin recycling and cell migration. *J Biol Chem* 2016; 291: 1320–1335.

38 Boudewijn IM, Faiz A, Steiling K, *et al.* Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression. *Respir Res* 2017; 18: 213.

39 Steiling K, van den Berge M, Hijazi K, *et al.* A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* 2013; 187: 933–942.

40 Lapperre TS, Snoeck-Stroband JB, Gosman MM, *et al.* Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: a randomized trial. *Ann Intern Med* 2009; 151: 517–527.

41 Hardin M, Cho MH, Sharma S, *et al.* Sex-based genetic association study identifies CELSR1 as a possible chronic obstructive pulmonary disease risk locus among women. *Am J Respir Cell Mol Biol* 2017; 56: 332–341.