Title

Bronchial gene expression alterations associated with radiological bronchiectasis

Journal

European Respiratory Journal, 61(1)

ISSN

0903-1936

Authors

Xu, Ke
Diaz, Alejandro A
Duan, Fenghai
et al.

Publication Date

2023

DOI

10.1183/13993003.00120-2022

Bronchial gene expression alterations associated with radiological bronchiectasis

Ke Xu[1,6], Alejandro A. Diaz[2,6], Fenghai Duan[3], Minyi Lee[1], Xiaohui Xiao[1], Hanqiao Liu[1], Gang Liu[1], Michael H. Cho[2,4], Adam C. Gower[1], Yuriy O. Alekseyev[1], Avrum Spira[1], Denise R. Aberle[5], George R. Washko[2], Ehab Billatos[1,7] and Marc E. Lenburg[1,7], on behalf of the DECAMP investigators

[1]Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [2]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [3]Department of Biostatistics and Center for Statistical Sciences, Brown University School of Public Health, Providence, RI, USA. [4]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. [5]Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. [6]K. Xu and A.A. Diaz contributed equally to this work. [7]E. Billatos and M.E. Lenburg contributed equally to this article as lead authors and supervised the work.

Corresponding author: Marc E. Lenburg (mlenburg@bu.edu)

Shareable abstract (@ERSpublications)
Within the bronchial epithelium, a gene expression signature associated with radiological bronchiectasis was discovered, suggesting an inflammatory airway environment that may correlate with decreased basal cells, but increased ciliated cells https://bit.ly/3KoNXYX

Cite this article as: Xu K, Diaz AA, Duan F, *et al*. Bronchial gene expression alterations associated with radiological bronchiectasis. *Eur Respir J* 2023; 61: 2200120 [DOI: 10.1183/13993003.00120-2022].

This article has an editorial commentary: https://doi.org/10.1183/13993003.01733-2022

Received: 18 Jan 2022
Accepted: 15 Aug 2022

Abstract

Objectives Discovering airway gene expression alterations associated with radiological bronchiectasis may improve the understanding of the pathobiology of early-stage bronchiectasis.

Methods Presence of radiological bronchiectasis in 173 individuals without a clinical diagnosis of bronchiectasis was evaluated. Bronchial brushings from these individuals were transcriptomically profiled and analysed. Single-cell deconvolution was performed to estimate changes in cellular landscape that may be associated with early disease progression.

Results 20 participants have widespread radiological bronchiectasis (three or more lobes). Transcriptomic analysis reflects biological processes associated with bronchiectasis including decreased expression of genes involved in cell adhesion and increased expression of genes involved in inflammatory pathways (655 genes, false discovery rate <0.1, $\log_2$ fold-change >0.25). Deconvolution analysis suggests that radiological bronchiectasis is associated with an increased proportion of ciliated and deuterosomal cells, and a decreased proportion of basal cells. Gene expression patterns separated participants into three clusters: normal, intermediate and bronchiectatic. The bronchiectatic cluster was enriched by participants with more lobes of radiological bronchiectasis ($p<0.0001$), more symptoms ($p=0.002$), higher SERPINA1 mutation rates ($p=0.03$) and higher computed tomography derived bronchiectasis scores ($p<0.0001$).

Conclusions Genes involved in cell adhesion, Wnt signalling, ciliogenesis and interferon-γ pathways had altered expression in the bronchus of participants with widespread radiological bronchiectasis, possibly associated with decreased basal and increased ciliated cells. This gene expression pattern is not only highly enriched among individuals with radiological bronchiectasis, but also associated with airway-related symptoms in those without discernible radiological bronchiectasis, suggesting that it reflects a bronchiectasis-associated, but non-bronchiectasis-specific lung pathophysiological process.

Introduction

Bronchiectasis is a pathological dilation of bronchi [1]. Once considered an orphan disease, bronchiectasis has become increasingly common in the United States (USA), with >70 000 new cases diagnosed, and between 340 000 and 522 000 adults treated for this condition in 2013 [2]. Bronchiectasis is diagnosed based on both imaging and clinical features. The imaging features include airway-to-arterial ratio >1.5, lack of tapering of bronchi and airway visibility within 1 cm of a costal pleural surface or touching the mediastinal pleura on computed tomography (CT) [3]. The clinical features of severe bronchiectasis

include chronic cough and sputum production [4]. Bronchiectasis is frequently identified as an incidental finding on CT in asymptomatic or mildly symptomatic individuals scanned for unrelated reasons. In lung cancer screening and COPD studies, 12–30% of the participants had bronchiectasis on CT [5, 6]. In the present study, we used the term radiological bronchiectasis to refer to participants with CT features compatible with bronchiectasis and without physician-diagnosed disease. Although the clinical and epidemiological aspects remain to be determined, we have shown previously that smokers with radiological bronchiectasis have an increased burden of emphysema on CT and lower airflow and exercise capacity [7]. Furthermore, in smokers with radiological bronchiectasis, structural damage to lung parenchyma and small airways correlated with a higher number of exacerbations over time [8]. Molecular profiling of the airway epithelium in individuals with radiological bronchiectasis may provide additional insights into the pathogenesis and biological pathways that are not yet targeted by current treatments [9–11].

Studies of bronchoalveolar lavage fluid and sputum have detected gene expression alterations or protein level changes related to bronchiectasis [12–14]. The utility of transcriptomic analysis in bronchiectasis has not been sufficiently explored, possibly due to a paucity of pertinent endobronchial tissue. Leveraging bronchial tissue at the mainstem bronchus, our group has previously discovered gene expression alterations in normal-appearing airway epithelial cells associated with smoking [15], lung cancer [16] and COPD [17, 18]. Not only may these gene expression alterations serve as biomarkers, but they have provided insights into disease pathophysiology.

We hereby aimed to characterise gene expression alterations in normal appearing bronchial epithelial cells in association with radiological bronchiectasis. A secondary objective was to explore whether these transcriptomic changes may represent a changing microenvironment by computationally deconvolving bulk samples into estimated proportions of each cell type with aid from single-cell RNA sequencing of the airway.

Methods

Study participants and sample analysis

The details of the Detection of Early Lung Cancer Among Military Personnel (DECAMP) protocols have been published previously [19]. Briefly, it consists of 15 military treatment facilities, Veterans Affairs hospitals and academic centres, conducting two prospective studies with the goals of developing an integrated panel of airway and blood-based molecular biomarkers that discriminate benign and malignant indeterminate nodules detected on CT scan. DECAMP collects biospecimens including bronchial brushings and biopsies from normal-appearing airway epithelium. In the DECAMP1 study (clinicaltrials.gov identifier NCT01785342), samples were collected from current and former smokers with indeterminate pulmonary nodules (7–30 mm). The DECAMP2 study (NCT02504697) included participants at high risk of developing lung cancer (heavy smoker, history of COPD, emphysema or at least one first-degree relative with a diagnosis of lung cancer, *etc.*) undergoing lung cancer screening. None of the DECAMP participants reported physician-diagnosed bronchiectasis at the time of sample collection. Per DECAMP protocol, each participant was asked to fill a Lung Health Questionnaire, a screening tool that includes demographics, personal medical history, family medical history, medications, smoking history, alcohol and recreational drug history and symptom history such as cough, dyspnoea and sputum production. This study was approved by the Human Research Protection Office (HRPO) for the United States Department of Defense and the institutional review board of every participating site. All subjects have given written informed consent to participate in the study.

From the large cohort of DECAMP of 365 subjects (138 from DECAMP-1 and 227 from DECAMP-2), 173 participants (DECAMP-1 n=129, DECAMP-2 n=44) with matching RNA-sequencing (seq) from the right mainstem bronchus and CT scans were available for analysis (figure 1).

Radiological bronchiectasis ascertainment

The detection of bronchiectasis was performed visually by a pulmonologist (A.A. Diaz) with >10 years of experience in lung imaging, blinded to the gene expression profiles and participants' clinical data. Radiological bronchiectasis was defined with one or more of the following criteria: 1) airway dilation (airway lumen diameter greater than adjacent pulmonary vessel diameter); 2) abnormal airway tapering of any extent (no decrease in or increase in lumen moving from proximal to distal airways); and 3) visualisation of a bronchus within 1 cm of the pleura. The lingula was considered a separate lobe. Severity of radiological bronchiectasis was determined using a bronchiectasis CT score, as reported previously [20]. Briefly, the score included the following items: 1) type of bronchiectasis (tubular, varicose and cystic); 2) bronchial dilation severity based on the bronchial lumen diameter to artery diameter ratio; 3) number of bronchopulmonary segments with bronchiectasis; 4) presence of mucus plugs in

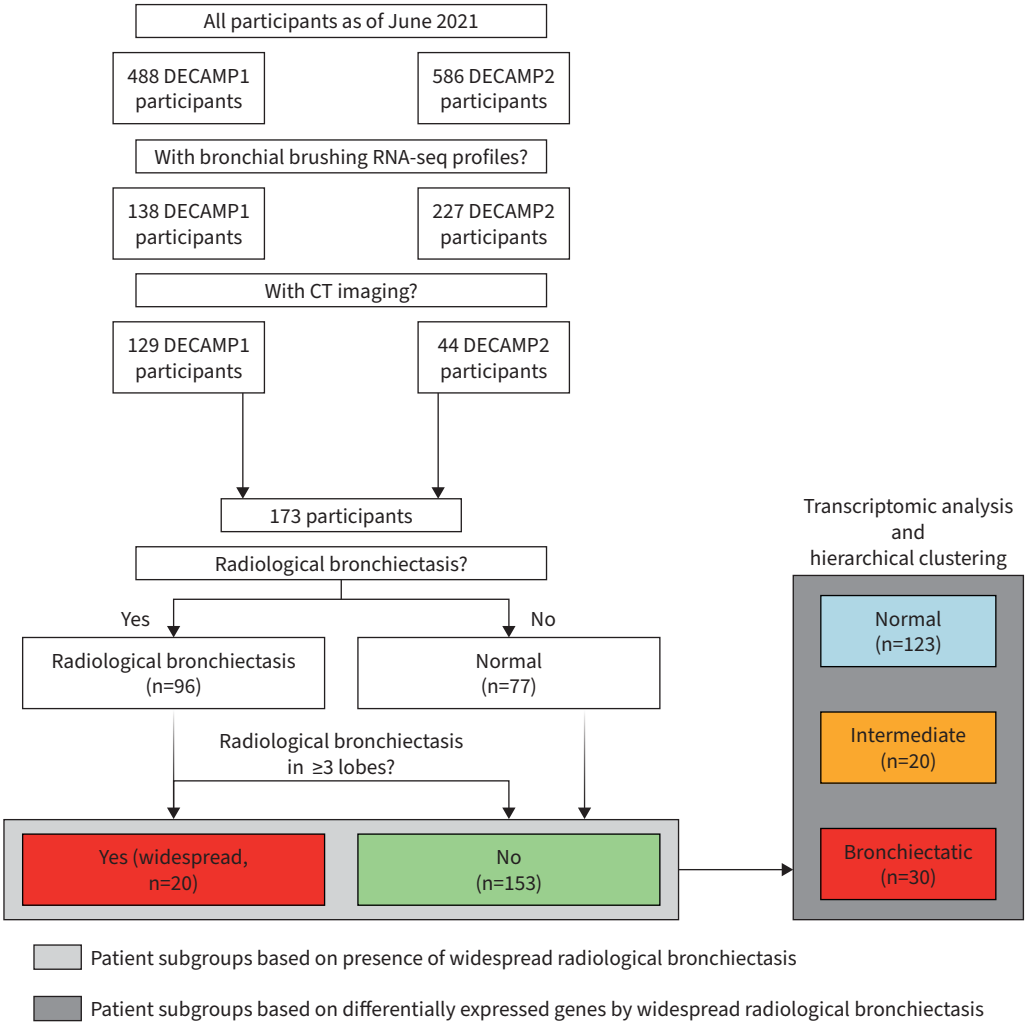

FIGURE 1 Schematic representation of participant clustering based on imaging and gene expression. DECAMP: Detection of Early Lung Cancer Among Military Personnel; RNA-seq: RNA sequencing; CT: computed tomography.

bronchiectatic airways; and 5) presence of wall thickness in bronchiectatic airways. The score ranges from 1 to 40, with greater values indicating greater severity of radiological bronchiectasis. Additionally, traction bronchiectasis was identified as a separate feature. A definition for traction bronchiectasis was adapted from the Fleischner Society criteria [21]: 1) presence of pulmonary fibrosis in any lung zone; and 2) presence of dilated airways of any size in at least one area of pulmonary fibrosis. If the quality of CT was inadequate for this assessment, traction bronchiectasis was deemed as indeterminate. Among participants with radiological bronchiectasis, those reporting frequent cough (at least four times a day and 4 days a week) and significant sputum production (at least twice a day and 4 days a week) were labelled as "meet clinical criteria for bronchiectasis". For the analysis of differential gene expression, we identified individuals as having widespread radiological bronchiectasis when three or more lobes were involved.

Single-cell RNA-seq analysis and deconvolution of bulk RNA-seq datasets

The Seurat R package (version 3.1) [22] was used for downstream analyses including normalisation, scaling, clustering of cells and identifying cluster marker genes on a dataset from Deprez *et al.* [23] including 20 519 tracheal epithelial cells from nine healthy volunteers. Cells were filtered out if they met any of the following criteria: 1) bottom quartile for a total number of genes detected; 2) bottom quartile for total library size; and 3) >30% of counts mapped to the mitochondrial genome. 13 176 cells were kept for further analysis. Uniform Manifold Approximation and Projection dimensionality reduction was performed using the first 15 principal components with a resolution setting of 0.8. Cell types were previously

assigned and individually validated with reported cell markers. Reference gene expression profiles (GEPs) were derived by identifying the marker genes of each cell type using the FindMarkers function (false discovery rate (FDR) q<0.05, log fold change >0.25). We applied the AutoGeneS package [24] to further identify the top 1000 most informative genes from the GEPs before applying the deconvolve function to estimate cell proportion from the bulk RNA-seq data with model parameter set to Nu Support Vector Regression (nusvr).

Results

Participant demographics, pulmonary function and imaging measurements

Representative CT images of radiological bronchiectasis are shown in figure 2. Of the 173 evaluated participants, 96 participants had radiological bronchiectasis in at least one lobe. Participants with and without radiological bronchiectasis had similar clinical characteristics (supplementary table S1). We had quantitative assessment of emphysema for 102 participants. In contrast with Martinez *et al.* [7], we did not see increased emphysema in patients with radiological bronchiectasis. Radiological bronchiectasis was predominantly present in the lower lobes (47%), and less likely to be presented in the lingula (5.5%) and the right middle lobe (13.2%) (supplementary table S2). Of the 173 participants, 20 had widespread radiological bronchiectasis (three females and 17 males; mean±SD age 66±5 years), and 153 did not (29 females and 124 males; mean±SD 67±8 years), including 76 with radiological bronchiectasis in fewer than three lobes and 77 without radiological bronchiectasis. In addition, we identified 26 participants that meet clinical criteria for bronchiectasis. We found that participants with and without widespread radiological bronchiectasis had similar clinical features, except that those with widespread radiological bronchiectasis were more likely to report shortness of breath (p=0.01; table 1).

Identification of three clusters of participants based on gene expression profiles

No genes were differentially expressed between the bronchial epithelium of participants with (>0 lobe) and without (=0 lobe) radiological bronchiectasis. Between participants with (three or more lobes) and without (fewer than three lobes) widespread radiological bronchiectasis, 1192 genes were significantly differentially expressed (FDR q-value <0.1) and 655 genes were significantly differentially expressed with absolute log fold change >0.25 controlling for biological sex and smoking status (figure 3 and supplementary table S3). Including COPD status in the model had little effect on the genes significantly associated with widespread radiological bronchiectasis: 639 genes were differentially expressed, of which 634 were part of the 655 genes derived from the model not including COPD status. Unsupervised clustering using the 655 genes first separated the 173 participants into two genomic clusters. The predominant cluster (n=123; light blue, left branch in figure 3) was primarily composed of participants without widespread radiological bronchiectasis. The smaller cluster on the right branch of the dendrogram contained two subgroups of participants, one that included most of the participants with widespread bronchiectasis (n=30; red), and one that was composed of participants without widespread bronchiectasis yet demonstrated gene expression patterns similar to those with widespread bronchiectasis (n=20; orange).

In addition to different number of lobes with radiological bronchiectasis (p<0.0001), these three clusters of participants differ by the severity of radiological bronchiectasis quantified by the bronchiectasis CT score (p<0.0001) and by the likelihood of having cardinal symptoms associated with bronchiectasis (cough and phlegm production). Based on differences both in the gene expression and clinical characteristics that correlated with an increasing presentation of symptoms related to bronchiectasis, we named these three

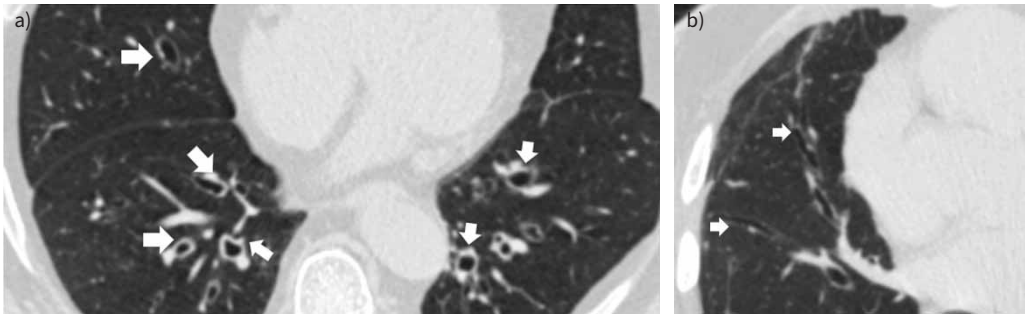

FIGURE 2 Computed tomography axial image of two participants with white arrows showing radiological bronchiectasis with a) airway dilation and b) abnormal airway tapering.

TABLE 1 Clinical characteristics of the participants with and without widespread bronchiectasis

	Participants with widespread radiological bronchiectasis	Participants without widespread radiological bronchiectasis	p-value[#]
Participants	20	153	
Number of lobes with radiological bronchiectasis median (interquartile range)	3 (1)	0 (2)	$<2\times10^{-16}$
Meet clinical criteria for bronchiectasis[¶]	6 (30)	20 (13)	0.08
Age years	66±5	67±8	0.62
Sex			1
Male	17 (85)	124 (81)	
Female	3 (15)	29 (19)	
Race			0.63
White	15 (75)	114 (74.5)	
Black	3 (15)	18 (11.8)	
Asian	1 (5)	4 (2.6)	
Other/unknown	1 (5)	17 (11.1)	
Smoking status			1
Current	9 (45)	66 (43.1)	
Former	10 (50)	80 (52.3)	
Unknown	1 (5)	7 (4.6)	
Smoking pack-years[+]	45±19	51±26	0.37
FEV_1 % predicted	69±22	75±20	0.23
FEV_1/FVC	0.6±0.2	0.6±0.1	1
COPD	14 (70)	105 (69)	1
Malignancy			0.31
Yes	4 (20)	71 (46)	
No	6 (30)	45 (31)	
Unknown	10 (50)	37 (24)	
Radiological bronchiectasis			<0.001
Yes	20 (100)	76 (50)	
No	0 (0)	77 (50)	
Traction bronchiectasis			0.081
Yes	7 (35)	23 (15)	
No	12 (60)	116 (76)	
Indeterminate	1 (5)	14 (9)	
Emphysema[§] %	16.2±14.7	13.8±16.7	0.71
CT bronchiectasis score	7.7±5.5	1.3±2.0	<0.001
Cough			1
Yes	9 (45)	68 (44)	
No	9 (45)	72 (47)	
Unknown	2 (10)	13 (9)	
Phlegm			0.46
Yes	11 (55)	65 (43)	
No	8 (40)	74 (48)	
Unknown	1 (5)	14 (9)	
Both cough and phlegm			0.804
Yes	8 (40)	47 (31)	
No	12 (60)	90 (59)	
Unknown	0 (0)	16 (11)	
Shortness of breath			0.01
Yes	17 (85)	81 (53)	
No	2 (10)	59 (39)	
Unknown	1 (5)	13 (9)	

Data are presented as n, mean±SD or n (%), unless otherwise stated. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; CT: computed tomography. [#]: p-values calculated using ANOVA F-test or Fisher's exact test; [¶]: missing values for 12 participants without widespread radiological bronchiectasis and two participants with widespread radiological bronchiectasis; [+]: missing values for two participants without widespread radiological bronchiectasis; [§]: missing values for 13 participants with widespread radiological bronchiectasis and 58 participants without widespread radiological bronchiectasis.

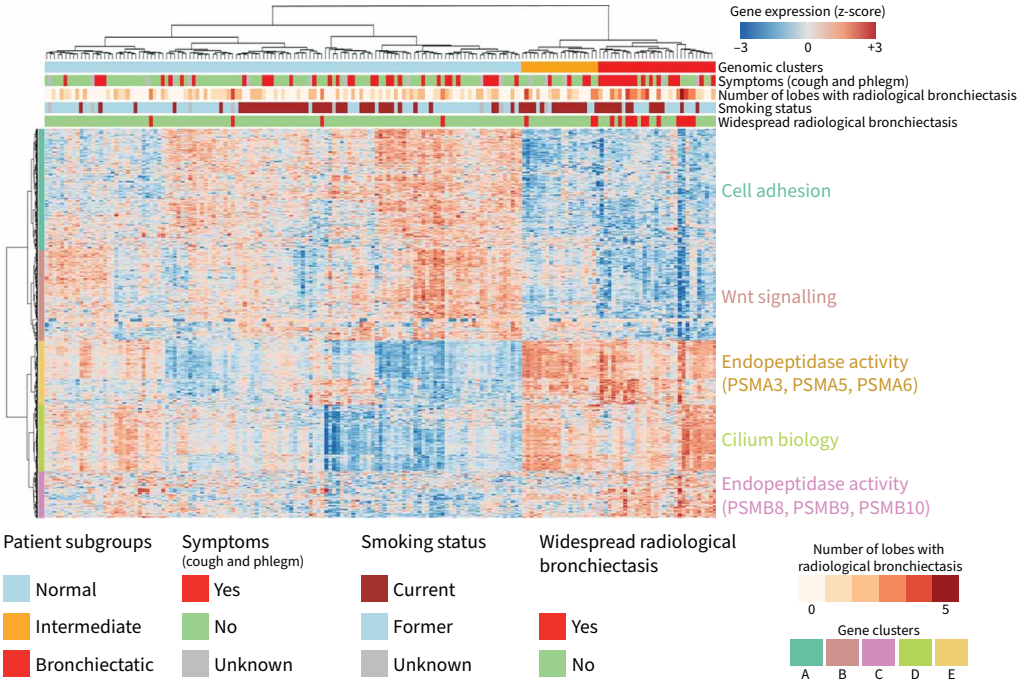

FIGURE 3 Unsupervised heatmap of the 655 genes associated with widespread radiological bronchiectasis (presence of radiological bronchiectasis in at least three lobes). Based on hierarchical clustering, participants were grouped into three genomic clusters, while genes were grouped into give gene clusters. Biological pathways in which these clusters of gene were enriched in were shown on the side. False discovery rate <0.1; absolute log fold change >0.25.

participant clusters normal (light blue), intermediate (orange) and bronchiectatic (red) (figure 1). The bronchiectatic cluster has the highest average of number lobes with radiological bronchiectasis, the highest proportion of participants reporting both cough and phlegm (n=20, 67%; p=0.002) and the highest proportion of participants who meet clinical criteria for bronchiectasis (n=10, 33%; p=0.009). Interestingly, the intermediate cluster had a higher proportion of current smokers (n=15, 75%; p=0.006) and intermediate bronchiectasis CT score (2.1, p<0.0001), but otherwise consists of individuals with similar clinical characteristics to the participants in the normal cluster with the lowest average number of lobes with radiological bronchiectasis (figure 3 and table 2). Of note, the three clusters were not correlated with COPD or lung cancer (table 2).

Interestingly, participants without widespread bronchiectasis within the bronchiectatic cluster were similar to those with widespread bronchiectasis in terms of clinical characteristics (supplementary table S4). When compared to the participants without widespread bronchiectasis in the normal and intermediate clusters, nonbronchiectasis participants in the bronchiectatic gene expression cluster exhibit an increased likelihood of cough (n=14, 82%; p=0.003) and co-presence of cough and phlegm production (n=10, 59%; p=0.02) (supplementary table S5).

Bulk RNA-seq analysis reveals molecular pathways associated with widespread radiological bronchiectasis

To better characterise the differentially expressed genes, we divided the 655 genes into five co-expression clusters (A–E) based on hierarchical clustering. Gene clusters A and B were significantly enriched for genes involved in cell adhesion (protocadherin-γ subunit (PCDHG)A4, PCDHGA7 and PCDHGA9, as well as seven other protocadherin-γ related transcripts (PCDHG)) and Wnt signalling (WNT2B, WNT3A and WNT5A), respectively (supplementary table S6), whereas gene clusters C and E were part of the endopeptidase activity pathway (PSMA3, PSMA5 and PSMA6, as well as four other proteasome-related transcripts) and genes in cluster D were involved in cilium organisation (TUBA1A, CC2D2A and CCDC176).

TABLE 2 Clinical characteristics of the participants of the three clusters based on gene expression profiles

	Normal	Intermediate	Bronchiectatic	p-value[#]
Participants	123	20	30	
Number of lobes with radiological bronchiectasis median (interquartile range)	1 (2)	1 (2)	2 (3)	<0.0001
Meet clinical criteria for bronchiectasis[¶]	15 (12)	1 (5)	10 (33)	0.009
Widespread radiological bronchiectasis				1.18×10^{-7}
Yes	4 (3)	3 (15)	13 (43)	
No	119 (97)	17 (85)	17 (57)	
Age years	67±8	65±6	67±6	0.29
Sex				0.14
Male	98 (80)	15 (75)	28 (93)	
Female	25 (20)	5 (25)	2 (7)	
Race				0.22
White	88 (72)	15 (75)	26 (87)	
Black	17 (14)	2 (10)	2 (7)	
Asian	2 (2)	2 (10)	1 (3)	
Others/unknown	16 (13)	1 (5)	1 (3)	
Smoking status				0.006
Current	46 (37)	15 (75)	14 (47)	
Former	70 (57)	4 (20)	16 (53)	
Unknown	7 (6)	1 (5)	0 (0)	
Smoking pack-years[+]	50±26	61±31	44±20	0.09
FEV_1 % predicted	75±20	78±20	68±21	0.16
FEV_1/FVC	0.6±0.1	0.7±0.1	0.6±0.2	0.07
COPD	83 (67.5)	13 (65)	23 (76.7)	0.59
Malignancy				0.18
Yes	56 (46)	11 (55)	8 (27)	
No	36 (29)	4 (20)	11 (37)	
Unknown	31 (25)	5 (25)	11 (37)	
Radiological bronchiectasis				0.42
Yes	65 (53)	11 (55)	20 (67)	
No	58 (47)	9 (45)	10 (33)	
Traction bronchiectasis				0.54
Yes	19 (15)	4 (20)	7 (23)	
No	93 (76)	13 (65)	22 (73)	
Indeterminate	11 (9)	3 (15)	1 (3)	
Emphysema[§] %	12.9±15.5	10.9±13.9	21.6±21.9	0.14
CT bronchiectasis score	1.6±2.6	2.1±3.4	3.7±5.2	<0.0001
Cough				0.0003
Yes	48 (39)	6 (30)	23 (77)	
No	66 (54)	10 (50)	5 (17)	
Unknown	9 (7)	4 (30)	2 (7)	
Phlegm				0.02
Yes	49 (40)	7 (35)	20 (67)	
No	63 (51)	11 (55)	8 (27)	
Unknown	11 (9)	2 (10)	2 (7)	
Both cough and phlegm				0.002
Yes	33 (27)	4 (20)	18 (60)	
No	79 (64)	13 (65)	10 (33)	
Unknown	11 (9)	3 (15)	2 (7)	
Shortness of breath				0.15
Yes	64 (52)	13 (65)	21 (70)	
No	49 (40)	5 (25)	7 (23)	
Unknown	10 (8)	2 (10)	2 (7)	

Data are presented as n, % or mean±SD, unless otherwise stated. FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; CT: computed tomography. [#]: p-values calculated using an ANOVA F-test or Fisher's exact test; [¶]: missing values for eight participants in the normal cluster, three participants in the intermediate cluster and three participants in the bronchiectatic cluster; [+]: missing values for one participant in the intermediate cluster and one participant in the bronchiectatic cluster; [§]: missing values for 48 participants in the normal cluster, eight participants in the intermediate cluster and 15 participants in the bronchiectatic cluster.

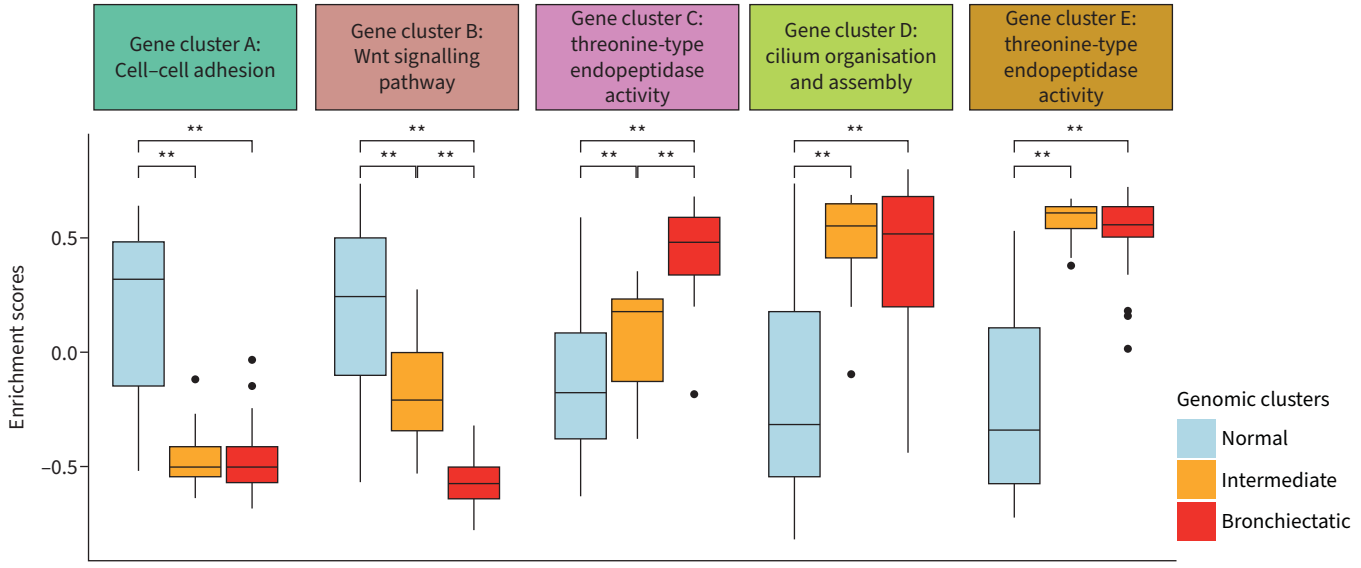

FIGURE 4 Modular gene expressions in genomic clusters. **: p<0.01 (Tukey-adjusted).

Enrichment scores for each cluster for each participant were calculated (figure 4). The normal and the bronchiectatic clusters showed clear distinctions in all gene clusters; gene clusters A and B were at higher levels among participants of the normal cluster; and gene clusters C, D and E were at higher levels among participants of the bronchiectatic cluster. However, participants in the intermediate cluster displayed two patterns of gene expression: one in which the intermediate and bronchiectatic clusters exhibited levels of gene expression different from the normal cluster (clusters A, D and E); and another in which the intermediate cluster exhibited gene expression intermediate between the normal and the bronchiectatic clusters (clusters B and C). In addition, we observed that overall, the expression of the differentially expressed genes was similar among participants without (0–2 lobes) and with (3–5 lobes) widespread radiological bronchiectasis, and the differences were significant between two and three lobes (supplementary figure S1).

To further explore the possible biological processes contributing to the bronchiectasis-associated gene expression differences, we performed gene set enrichment analysis using canonical gene sets [25, 26] and found that genes upregulated among participants with widespread radiological bronchiectasis were enriched in interferon (IFN)-γ, oxidative phosphorylation and IFN-α pathways, while genes upregulated in participants without widespread radiological bronchiectasis were enriched in genes downregulated by KRAS activation, genes involved in the epithelial–mesenchymal transition and pancreas β-cell genes (supplementary table S7).

Moreover, we compared the differentially expressed genes to a signature of ciliogenesis, which contains a list of 310 genes involved in cilia organisation and associated with primary ciliary dyskinesia [27], a significant risk factor for bronchiectasis. 42 of the 310 genes were upregulated in patients with widespread radiological bronchiectasis (supplementary table S8). We found significant enrichment of ciliogenesis-associated genes among the genes expressed at higher levels in participants with widespread radiological bronchiectasis (q<0.001; supplementary figure S2).

Widespread radiological bronchiectasis correlates with SERPINA1 mutations

We inferred the presence of ZZ and SZ, as well as the Ile74Asn SERPINA1 genotype [28]. We found that there was a trend toward more participants in the bronchiectatic cluster having the ZZ variant (four out of 30, 13.3%; p=0.077) and that there were significantly more patients in the bronchiectatic cluster having either the ZZ or SZ variant (five out of 30, 16.7%; p=0.031; supplementary table S9).

Widespread radiological bronchiectasis correlates with increased proportions of ciliated and deuterosomal cells and decreased proportions of basal cells

To explore whether the observed gene expression alterations might be cell type specific, we leveraged a published single-cell RNA-seq dataset of tracheal epithelial cells [23] and calculated per-cell enrichment

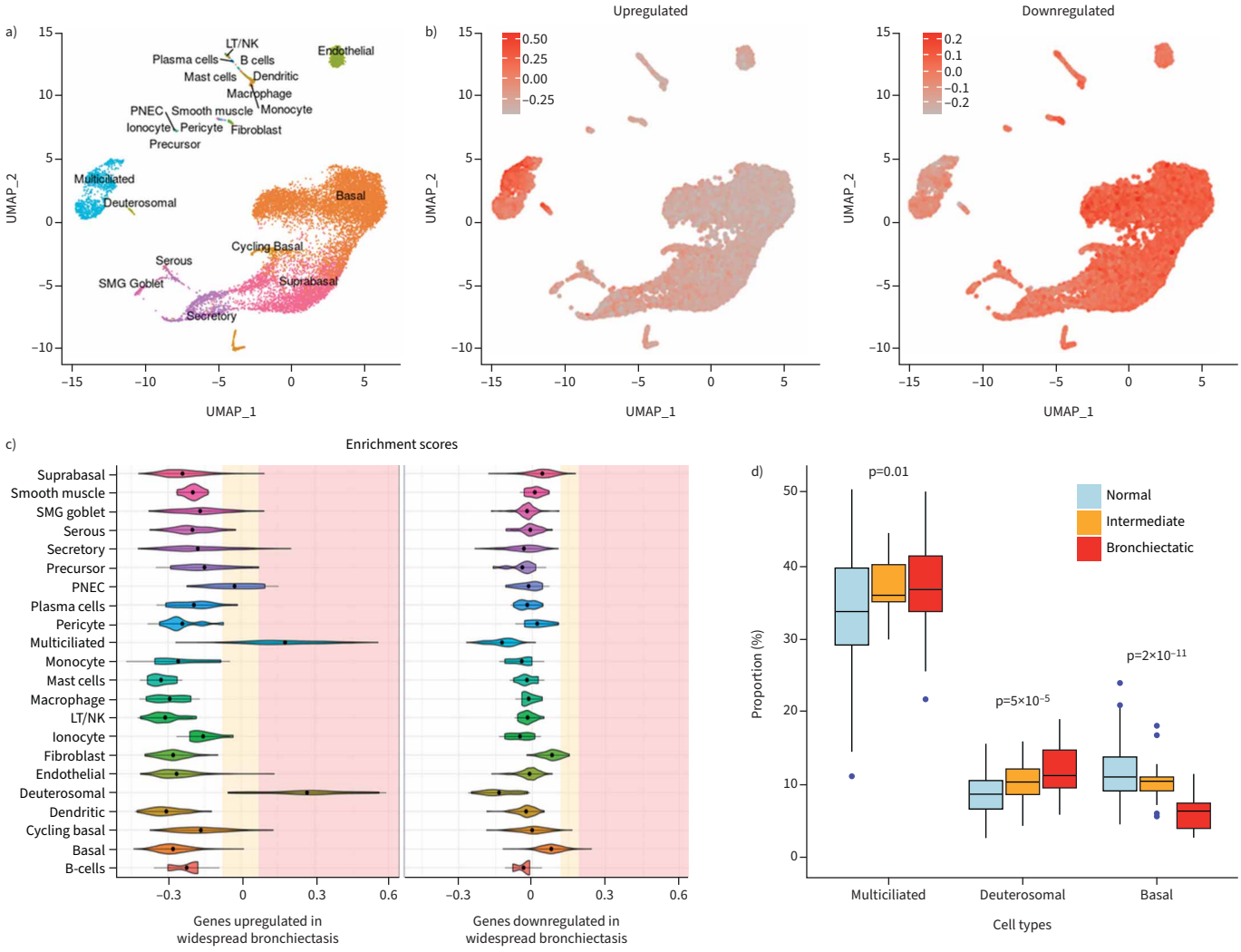

FIGURE 5 a) Single-cell RNA-sequencing (seq) of bronchial brushings from nine subjects (n=13 176 cells) were clustered. Cell types were previously assigned and reported by Deprez *et al.* [23]. b) Uniform Manifold Approximation and Projections (UMAP) showing the expression pattern of genes upregulated and downregulated in widespread radiological bronchiectasis across different cell types. The cells are coloured grey for low expression and red for high expression of metagene scores of each set of genes. c) Violin plot showing the metagene score for each set of gene modules across the cell types. For each violin plot, metagene expression is designated as elevated (light yellow) or highly elevated (pink) if it is greater than one or two standard deviations above the mean metagene score, respectively. d) Boxplots of cell type proportions estimated by AutoGeneS in bulk RNA-seq data from the bronchial brushings (n=173) obtained from the Detection of Early Lung Cancer Among Military Personnel cohort. Significant cell proportion differences between the normal, intermediate and bronchiectatic clusters determined by Kruskal–Wallis test. LT/NK: lymphotoxin/natural killer; PNEC: pulmonary neuroendocrine cell; SMG: submucosal gland.

scores for the genes upregulated or downregulated in participants with widespread radiological bronchiectasis. We found that the genes upregulated in individuals with widespread radiological bronchiectasis were expressed almost exclusively by deuterosomal cells and multiciliated cells, whereas the genes upregulated in individuals without widespread radiological bronchiectasis were expressed diffusely across cell types, but at highest levels among basal cells (figure 5a–c). To further examine the possibility that the gene expression alterations observed at the bulk level may be consistent with altered epithelial cell prevalence, we estimated cell type proportions in the bulk RNA-seq data (supplementary table S10). The predicted proportions of multiciliated and deuterosomal cells increased from the normal to the intermediate and then to the bronchiectatic cluster ($p<0.01$ and $p<0.001$, respectively; figure 5d), with the shift being most pronounced for the immature deuterosomal cells. In contrast, the proportion of basal cells incrementally decreased from the normal to the intermediate and then to the bronchiectatic cluster ($p<0.001$). Of note, the proportion of goblet cells, previously shown to be correlated with cigarette smoking [29], was increased in the smoker-predominant intermediate cluster (supplementary figure S3).

Discussion

Current knowledge about the pathogenesis of bronchiectasis has been summarised as a "vicious cycle" model [30] in which cyclical epithelial dysfunction, chronic infection, recurring inflammation and structural damage promote the enlargement of bronchi. Our analysis of individuals without previous physician-diagnosed bronchiectasis but who show signs of bronchiectasis in multiple lobes on CT ("widespread radiological bronchiectasis") potentially offers insights into early molecular changes associated with bronchiectasis development. Consistent with the proposed vicious cycle model, a loss of gene expression related to cell adhesion and an increased level of gene expression related to inflammation is detected in participants with widespread radiological bronchiectasis. Our analysis reveals two additional biological pathways that may play important roles in the initiation of bronchiectasis: a decreased expression of genes in the Wnt signalling pathway and an increased expression of genes involved in ciliogenesis.

The predominant presence of radiological bronchiectasis in the lower lobes, the decreased expression of genes involved in cell adhesion, the increased expression of genes involved in inflammatory processes and an increased prevalence of cough and phlegm production suggest similarities between widespread radiological bronchiectasis and clinical bronchiectasis [31]. This similarity is further supported by the increased prevalence of participants meeting the radiological and clinical criteria for bronchiectasis in the bronchiectatic patient cluster defined by the expression of genes associated with widespread radiological bronchiectasis.

With regard to cell adhesion, we were intrigued to find decreased protocadherin gene expression associated with widespread radiological bronchiectasis. PCDHGA7 and PCDHGB4 are members of the protocadherin-γ gene cluster, initially found as adhesion proteins in the cell–cell connections of the brain [32]. PCDHGA7 was downregulated in colorectal cancer that is associated with epithelial damage [33]. Moreover, protocadherin genes are strongly conserved between mouse and human, and are critical in epithelial maintenance and repair of the asthmatic mouse lung [34]. Therefore, the decreased expression of these protocadherin genes may be related to airway dilation in early bronchiectasis.

Large immunohistological studies have reported the presence of $CD8^+$ T-cells [35], $CD4^+$ T-cells, macrophages, neutrophils and interleukin-8 positive cells in the airways of patients with bronchiectasis [36, 37]. Our findings highlight the potential role of proteasome-related proteins in mediating the inflammatory processes. Constitutively expressed proteasome 20S subunit-β (PSMB)5, and the IFN-γ inducible immunoproteasome subunits PSMB8, PSMB9, PSMB10 [38, 39] were upregulated in individuals with widespread bronchiectasis. The immunoproteasome has been shown to enhance the generation of major histocompatibility complex I associated peptide and facilitate the antiviral immune response in the lung [40–42]. Furthermore, IFN-γ has previously been shown in mice to play an essential role in protective immunity against tuberculosis and mycobacterial infection, both of which are relevant to infection in bronchiectatic airway [43, 44]. Heightened IFN response in the bronchiectatic cluster could reflect immune responses to infectious processes that lead to more severe bronchiectasis.

Activation of the Wnt signalling pathway is important in maintaining the epithelial niche *via* the balance of epithelial/mesenchymal pairing [45–47]. LGR5, the gene most strongly downregulated in participants with widespread radiological bronchiectasis, maintains the epithelial progenitor niche [48, 49]. Genes downregulated in participants with widespread bronchiectasis are highly expressed in basal cells, suggesting that these cells may be less prevalent or less active in individuals with widespread radiological bronchiectasis, which may alter the structure of the airway or the ability to repair damaged airway epithelium. Previously, Chandrasekaran *et al.* [50] suggested that Wnt signalling pathway activity decreases with age and may explain age-associated severity in bronchiectasis. Participants with and without widespread radiological bronchiectasis were of similar chronological age in our cohort, but decreased WNT-related gene expression may still reflect a biological ageing process associated with elevated bronchiectasis risk.

Because bronchiectasis is often accompanied by loss of cilia [51], we were intrigued to observe increased expression of cilia-related genes in participants with widespread radiological bronchiectasis. Genes previously found to be upregulated in ciliogenesis [27] are upregulated among participants with widespread bronchiectasis. Additionally, single-cell RNA-seq of the bronchial epithelium shows that genes upregulated in participants with widespread radiological bronchiectasis are enriched in the multiciliated and deuterosomal cells. We hypothesise that the increased expression of ciliogenesis genes could be a response to ciliary damage, supported by the deconvolution result demonstrating significant increases in the proportion of both the multiciliated cells and deuterosomal cells in samples from the bronchiectatic cluster. The predicted increase in the proportion of deuterosomal cells in samples in the bronchiectatic cluster is

the most dramatic. These cells, characterised by high expression of DEUP1, FOXN4 and CDC20B, have been reported as a precursor of multiciliated cells [52]. An alternative hypothesis is that the overproduction of certain ciliary proteins contributes to faulty cilia assembly, reduced clearing capacity of the lung [37] and bronchiectasis pathogenesis.

Lastly, the increased prevalence of SERPINA1 mutations within the bronchiectatic cluster is consistent with the association between these genetic variants and α1-antitrypsin deficiency, which can lead to multiple pulmonary pathologies including chronic bronchitis and bronchiectasis [28, 53].

Limitations and future directions

Our analysis was performed on a predominantly male population of heavy current and former smokers at high risk of developing lung cancer. Thus, caution should be used when generalising the results to other patient populations, especially those who are never-smoking and/or female. In addition, CT reconstruction was not uniform and does not conform to the recent recommendation [3] of 1 mm slice thickness (⩾1.25 mm was used in DECAMP), which may result in decreased sensitivity for more subtle bronchiectasis. Subsequent imaging studies would allow us to gain more insight into whether the genomic clusters reflect risk of developing more radiological or clinical bronchiectasis, or if they are correlated with specific symptoms that are bronchiectasis-related but not bronchiectasis-exclusive, namely cough and sputum production. The later hypothesis is based on the observation that while 13 of the 30 participants in the bronchiectatic gene expression cluster have widespread radiological bronchiectasis, the other 17 do not. These 17 participants differ from the other participants without widespread bronchiectasis in that they are significantly more likely to have cough and phlegm production. It remains to be determined if the gene expression profile that defines the bronchiectatic cluster may be the common consequence of two separate mechanisms (one associated with bronchiectasis and another associated with cough and phlegm production), or if the gene expression profile reflects a single mechanism, perhaps one that can lead to a spectrum of airway dysfunction.

The identification of the intermediate genomic cluster could suggest a previously unappreciated risk of bronchiectasis among smokers. Alternatively, those in the intermediate cluster may progress to develop symptoms such as cough and phlegm production, becoming more similar to those within the bronchiectatic group but without widespread bronchiectasis. Longitudinal follow-up of participants of the intermediate cluster may validate whether smoking indeed leads to an increased risk for developing radiological evidence of bronchiectasis.

Conclusion

In conclusion, gene expression differences in individuals with radiological bronchiectasis in multiple lobes reflect biological processes that have previously been associated with bronchiectasis, including an increase of inflammatory processes and a decrease in cell adhesion. Novel mechanisms that may be associated with bronchiectasis initiation were also discovered: an increase of ciliary process and a reduction of the Wnt signalling pathway. The gene expression alterations were also detected in a subpopulation of participants who present with cough and phlegm production but did not have widespread radiological bronchiectasis, and an intermediate pattern of gene expression was present in a group predominantly composed of current smokers. Longitudinal clinical follow-up and molecular profiling of the participants will provide an opportunity to explore the potential for progression and the molecular risk factors for developing radiological bronchiectasis.

Author contributions: K. Xu and A.A. Diaz contributed equally to this work (co-first authors); E. Billatos and M.E. Lenburg contributed equally to this work (co-senior authors). K. Xu is the guarantor of this paper and takes full responsibility for the integrity of the work as a whole. A. Spira, M.E. Lenburg, E. Billatos and G.R. Washko initiated the study design. F. Duan and A.C. Gower prepared patient characteristics for the DECAMP cohort. X. Xiao, H. Liu and G. Liu collected and processed samples. K. Xu, M. Lee and A.A. Diaz analysed data. K. Xu prepared the figures. K. Xu, A.A. Diaz, G.R. Washko, E. Billatos and M.E. Lenburg interpreted results. K. Xu drafted the manuscript. A.A. Diaz analysed radiological scans. K. Xu, A.A. Diaz, M.H. Cho, E. Billatos and M.E. Lenburg edited and revised the manuscript. All the other authors approved the final version of the manuscript.

Conflict of interest: A.A. Diaz reports a provisional patent for genetic signatures to identify bronchiectasis filed in the United States Patent and Trademark Office (Patent of Novel Essays to Detect Respiratory Diseases and Disorders) outside the submitted work. M.H. Cho reports grants from GSK and Bayer; consulting fees from AstraZeneca and Genentech; lecture honoraria from Illumina; outside the submitted work. A.C. Gower reports grants from Novartis and Johnson & Johnson (JNJ) Immunology; outside the submitted work. A. Spira reports

being a current employee of JNJ. D.R. Aberle reports grants from American College of Radiology, Boston University, NIH/NCI (R01 CA226079, U01 CA233370, R01 CA210360, U01 CA214182) and Kaiser Foundation Research Institute/PCORI (prime); travel support from American Institute for Medical and Biological Engineering (AIMBE), Cleveland Clinic, Specialized Programs of Research Excellence (SPOREs) Workshop and International Association for the Study of Lung Cancer (IASLC); outside the submitted work. G.R. Washko reports grants from NIH and DoD; consulting fees from Boehringer Ingelheim, CSL Behring, GlaxoSmithKline, Novartis, Philips and Vertex Pharmaceuticals; travel support from Philips; participation on advisory board with Pulmonx; outside the submitted work; is a co-founder and equity share holder in Quantitative Imaging Solutions, a company that provides consulting services for image and data analytics; and G.R. Washko's spouse works for Biogen. E. Billatos reports grants from NIH (1R01HL149861-01A1), Novartis Institutes of Biomedical Research and International Association for the Study of Lung Cancer; a patent application filed stemming from research described in this manuscript from Boston University; outside the submitted work. M.E. Lenburg reports Johnson and Johnson, Novartis Institute for Biomedical Research; consulting fees from Johnson and Johnson; a patent application filed stemming from research described in this manuscript from Boston University; outside the submitted work. All other authors have nothing to disclose.

Support statement: The DECAMP study is supported by funds from the Department of Defense (W81XWH-11-2-0161), the National Cancer Institute (U01CA196408), and Johnson and Johnson Services, Inc (JJSI). Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Chalmers JD, Chang AB, Chotirmall SH, *et al.* Bronchiectasis. *Nat Rev Dis Primer* 2018; 4: 45.
2 Weycker D, Hansen GL, Seifer FD. Prevalence and incidence of noncystic fibrosis bronchiectasis among US adults in 2013. *Chron Respir Dis* 2017; 14: 377–384.
3 Tiddens HAWM, Meerburg JJ, van der Eerden MM, *et al.* The radiological diagnosis of bronchiectasis: what's in a name? *Eur Respir Rev* 2020; 29: 190120.
4 Aliberti S, Goeminne PC, O'Donnell AE, *et al.* Criteria and definitions for the radiological and clinical diagnosis of bronchiectasis in adults for use in clinical trials: international consensus recommendations. *Lancet Respir Med* 2022; 10: 298–306.
5 Tan WC, Hague CJ, Leipsic J, *et al.* Findings on thoracic computed tomography scans and respiratory outcomes in persons with and without chronic obstructive pulmonary disease: a population-based cohort study. *PLoS One* 2016; 11: e0166745.
6 Diaz AA, Maselli DJ, Rahaghi F, *et al.* Pulmonary vascular pruning in smokers with bronchiectasis. *ERJ Open Res* 2018; 4: 00044-2018.
7 Martinez CH, Okajima Y, Yen A, *et al.* Paired CT measures of emphysema and small airways disease and lung function and exercise capacity in smokers with radiographic bronchiectasis. *Acad Radiol* 2021; 28: 370–378.
8 Maselli DJ, Yen A, Wang W, *et al.* Small airway disease and emphysema are associated with future exacerbations in smokers with CT-derived bronchiectasis and COPD: results from the COPDGene cohort. *Radiology* 2021; 300: 706–714.
9 Metersky M, Chalmers J. Bronchiectasis insanity: doing the same thing over and over again and expecting different results? *F1000Res* 2019; 8: 293.
10 Barker AF, O'Donnell AE, Flume P, *et al.*. Aztreonam for inhalation solution in patients with non-cystic fibrosis bronchiectasis (AIR-BX1 and AIR-BX2): two randomised double-blind, placebo-controlled phase 3 trials. *Lancet Respir Med* 2014; 2: 738–749.
11 De Soyza A, Aksamit T, Bandel TJ, *et al.* RESPIRE 1: a phase III placebo-controlled randomised trial of ciprofloxacin dry powder for inhalation in non-cystic fibrosis bronchiectasis. *Eur Respir J* 2018; 51: 1702052.
12 Chen AC-H, Pena OM, Nel HJ, *et al.* Airway cells from protracted bacterial bronchitis and bronchiectasis share similar gene expression profiles. *Pediatr Pulmonol* 2018; 53: 575–582.
13 Chalmers JD, Moffitt KL, Suarez-Cuartin G, *et al.* Neutrophil elastase activity is associated with exacerbations and lung function decline in bronchiectasis. *Am J Respir Crit Care Med* 2017; 195: 1384–1393.
14 Guan W, Gao Y, Xu G, *et al.* Sputum matrix metalloproteinase-8 and -9 and tissue inhibitor of metalloproteinase-1 in bronchiectasis: clinical correlates and prognostic implications. *Respirology* 2015; 20: 1073–1081.
15 Sridhar S, Schembri F, Zeskind J, *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008; 9: 259.
16 Spira A, Beane JE, Shah V, *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007; 13: 361–366.
17 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.
18 Billatos E, Ash SY, Duan F, *et al.* Distinguishing smoking-related lung disease phenotypes *via* imaging and molecular features. *Chest* 2021; 159: 549–563.

19 Billatos E, Duan F, Moses E, *et al.* Detection of Early Lung Cancer Among Military Personnel (DECAMP) consortium: study protocols. *BMC Pulm Med* 2019; 19: 59.
20 Diaz AA, Young TP, Maselli DJ, *et al.* Quantitative CT measures of bronchiectasis in smokers. *Chest* 2017; 151: 1255–1262.
21 Hansell DM, Bankier AA, MacMahon H, *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697–722.
22 Butler A, Hoffman P, Smibert P, *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; 36: 411–420.
23 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.
24 Aliee H, Theis F. AutoGeneS: automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Cell Cyst* 2021; 12: 706–715.
25 Liberzon A, Subramanian A, Pinchback R, *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; 27: 1739–1740.
26 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545–15550.
27 Paff T, Kooi IE, Moutaouakil Y, *et al.* Diagnostic yield of a targeted gene panel in primary ciliary dyskinesia patients. *Hum Mutat* 2018; 39: 653–665.
28 Carpagnano GE, Santacroce R, Palmiotti GA, *et al.* A new SERPINA-1 missense mutation associated with alpha-1 antitrypsin deficiency and bronchiectasis. *Lung* 2017; 195: 679–682.
29 Duclos GE, Teixeira VH, Autissier P, *et al.* Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Sci Adv* 2019; 5: eaaw3413.
30 Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet* 2018; 392: 880–890.
31 King PT, Holdsworth SR, Freezer NJ, *et al.* Characterisation of the onset and presenting clinical features of adult bronchiectasis. *Respir Med* 2006; 100: 2183–2189.
32 Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 1999; 97: 779–790.
33 Liu Y, Peng K, Xie R, *et al.* Protocadherin γ-A7 is down-regulated in colorectal cancer and associated with the prognosis in patients with wild-type KRAS. *Hum Pathol* 2019; 83: 14–21.
34 Koning H, van Oosterhout AJ, Brouwer U, *et al.* Mouse protocadherin-1 gene expression is regulated by cigarette smoke exposure *in vivo*. *PLoS One* 2014; 9: e98197.
35 Keller IE, Vosyka O, Takenaka S, *et al.* Regulation of immunoproteasome function in the lung. *Sci Rep* 2015; 5: 10230.
36 Lapa e Silva JR, Guerreiro D, Noble B, *et al.* Immunopathology of experimental bronchiectasis. *Am J Respir Cell Mol Biol* 1989; 1: 297–304.
37 Gaga M, Bentley A, Humbert M, *et al.* Increases in $CD4^+$ T lymphocytes, macrophages, neutrophils and interleukin 8 positive cells in the airways of patients with bronchiectasis. *Thorax* 1998; 53: 685–691.
38 Guo Z, Chen W, Wang L, *et al.* Clinical and genetic spectrum of children with primary ciliary dyskinesia in China. *J Pediatr* 2020; 225: 157–165.
39 Rouette A, Trofimov A, Haberl D, *et al.* Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers. *Sci Rep* 2016; 6: 34019.
40 Basler M, Kirk CJ, Groettrup M. The immunoproteasome in antigen processing and other immunological functions. *Curr Opin Immunol* 2013; 25: 74–80.
41 de Verteuil D, Muratore-Schroeder TL, Granados DP, *et al.* Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Mol Cell Proteomics* 2010; 9: 2034–2047.
42 de Verteuil DA, Rouette A, Hardy MP, *et al.* Immunoproteasomes shape the transcriptome and regulate the function of dendritic cells. *J Immunol* 2014; 193: 1121–1132.
43 Sadikot RT, Blackwell TS, Christman JW, *et al.* Pathogen–host interactions in *Pseudomonas aeruginosa* pneumonia. *Am J Respir Crit Care Med* 2005; 171: 1209–1223.
44 Vankayalapati R, Wizel B, Samten B, *et al.* Cytokine profiles in immunocompetent persons infected with *Mycobacterium avium* complex. *J Infect Dis* 2001; 183: 478–484.
45 Samaha E, Vierlinger K, Weinhappel W, *et al.* Expression profiling suggests loss of surface integrity and failure of regenerative repair as major driving forces for chronic obstructive pulmonary disease progression. *Am J Respir Cell Mol Biol* 2021; 64: 441–452.
46 Zepp JA, Zacharias WJ, Frank DB, *et al.* Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* 2017; 170: 1134–1148.
47 Nabhan AN, Brownfield DG, Harbury PB, *et al.* Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* 2018; 359: 1118–1123.

48 Yan KS, Janda CY, Chang J, *et al.* Non-equivalence of Wnt and R-spondin ligands during Lgr5^{+} intestinal stem-cell self-renewal. *Nature* 2017; 545: 238–242.

49 Barker N, van Es JH, Kuipers J, *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 2007; 449: 1003–1007.

50 Chandrasekaran R, Mac Aogáin M, Chalmers JD, *et al.* Geographic variation in the aetiology, epidemiology and microbiology of bronchiectasis. *BMC Pulm Med* 2018; 18: 83.

51 Goddard M. Histopathology of bronchiectasis. *In*: Bronchiectasis (ERS Monograph). Sheffield, European Respiratory Society, 2011; pp. 22–31.

52 Ruiz García S, Deprez M, Lebrigand K, *et al.* Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* 2019; 146: dev177428.

53 DeMeo DL, Silverman EK. α_1-Antitrypsin deficiency. 2: genetic aspects of α_1-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. *Thorax* 2004; 59: 259–264.