Bronchial gene expression signature associated with rate of subsequent FEV1 decline in individuals with and at risk of COPD

Becker, Elizabeth J; Faiz, Alen; van den Berge, Maarten; Timens, Wim; Hiemstra, Pieter S; Clark, Kristopher; Liu, Gang; Xiao, Xiaohui; Alekseyev, Yuriy O; O'Connor, George

Published in:
Thorax

DOI:
10.1136/thoraxjnl-2019-214476

Original research

Bronchial gene expression signature associated with rate of subsequent FEV_1 decline in individuals with and at risk of COPD

Elizabeth J Becker,[1,2] Alen Faiz,[3,4] Maarten van den Berge,[4] Wim Timens ,[5] Pieter S Hiemstra ,[6] Kristopher Clark,[7] Gang Liu,[1] Xiaohui Xiao,[1] Yuriy O Alekseyev,[8] George O'Connor,[9] Stephen Lam,[10] Avrum Spira,[1,2,8] Marc E Lenburg,[1,2,8] Katrina Steiling[1,2,9]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2019-214476).

For numbered affiliations see end of article.

Correspondence to
Dr Katrina Steiling, Division of Computational Biomedicine, Boston University School of Medicine, Boston, MA 02118, USA; steiling@bu.edu

MEL and KS contributed equally.

MEL and KS are joint senior authors.

Received 19 December 2019
Revised 10 February 2021
Accepted 8 April 2021
Published Online First 10 May 2021

ABSTRACT

Background COPD is characterised by progressive lung function decline. Leveraging prior work demonstrating bronchial airway COPD-associated gene expression alterations, we sought to determine if there are alterations associated with differences in the rate of FEV_1 decline.

Methods We examined gene expression among ever smokers with and without COPD who at baseline had bronchial brushings profiled by Affymetrix microarrays and had longitudinal lung function measurements (n=134; mean follow-up=6.38±2.48 years). Gene expression profiles associated with the rate of FEV_1 decline were identified by linear modelling.

Results Expression differences in 171 genes were associated with rate of FEV_1 decline (false discovery rate <0.05). The FEV_1 decline signature was replicated in an independent dataset of bronchial biopsies from patients with COPD (n=46; p=0.018; mean follow-up=6.76±1.32 years). Genes elevated in individuals with more rapid FEV_1 decline are significantly enriched among the genes altered by modulation of XBP1 in two independent datasets (Gene Set Enrichment Analysis (GSEA) p<0.05) and are enriched in mucin-related genes (GSEA p<0.05).

Conclusion We have identified and replicated an airway gene expression signature associated with the rate of FEV_1 decline. Aspects of this signature are related to increased expression of XBP1-regulated genes, a transcription factor involved in the unfolded protein response, and genes related to mucin production. Collectively, these data suggest that molecular processes related to the rate of FEV_1 decline can be detected in airway epithelium, identify a possible indicator of FEV_1 decline and make it possible to detect, in an early phase, ever smokers with and without COPD most at risk of rapid FEV_1 decline.

Key messages

What is the key question?

- ► Can gene expression profiling of the airway epithelium be used to identify molecular processes associated with the rate of FEV_1 decline?

What is the bottom line?

- ► An airway gene expression signature associated with the rate of subsequent decline in FEV_1 is identified and replicated. This signature is enriched for genes that are regulated by XBP1, a key transcription factor involved in the unfolded protein response.

Why read on?

- ► The present study highlights that molecular processes associated with the rate of FEV_1 decline can be detected by bronchial epithelial gene expression profiles. This work identifies a possible indicator of FEV_1 decline and demonstrates the potential for bronchial gene expression to serve as an intermediate endpoint for studying the rate of FEV_1 decline.

INTRODUCTION

COPD is the third leading cause of death in the world.[1] In 2016, 3 million people died of COPD, which accounted for 6% of all deaths globally.[1] Accelerated lung function decline is considered a feature of COPD and is most commonly measured by change in FEV_1. Lower FEV_1 is associated with an increased risk of death,[2] and even smokers who do not yet meet the clinical definition of COPD may experience more rapid FEV_1 decline.[3] The rate of FEV_1 decline is highly variable between individuals.[4] Though some risk factors for rapid FEV_1 decline have been identified, such as cigarette smoking,[5] higher blood neutrophil counts,[6] albuminuria[7] and alpha 1-antitrypsin deficiency,[8] these do not fully explain the heterogeneity in COPD and have not yet been useful in predicting FEV_1 decline for individual patients. The ability to predict FEV_1 decline would enable clinicians to stratify at-risk patients towards more aggressive management. It might also facilitate clinical trials of therapies to modify the natural history of COPD, specifically targeting individuals more likely to experience greater decline in FEV_1. Finally, it could lead to further indications for finding therapeutic targets to slow disease progression.

Previous studies have demonstrated that bronchial epithelial gene expression is altered both by cigarette smoking and in diseases associated with

Check for updates

To cite: Becker EJ, Faiz A, van den Berge M, *et al*. *Thorax* 2022;**77**:31–39.

cigarette smoking.[9–11] We have previously described the Steiling *et al*[10] bronchial airway gene expression signature of COPD and disease severity as measured by FEV_1. Importantly, these gene expression alterations in the more proximal airway were similar to disease-associated changes present in more distal diseased lung tissue, suggesting that bronchial airway gene expression in COPD can be used to study factors in its pathobiology.

Based on these observations, we hypothesised that bronchial airway epithelial gene expression might reflect molecular processes associated with accelerated FEV_1 decline. In this study, we identified and replicated a baseline gene expression signature associated with the rate of FEV_1 decline observed during subsequent follow-up. We found this signature to be significantly enriched for genes with binding sites for the transcription factor encoded by XBP1, which is involved in the unfolded protein response (UPR) to endoplasmic reticulum (ER) stress.

Some of the results reported here have been previously published in abstract form.[12–15]

METHODS

Primary dataset and longitudinal FEV_1

The individuals included in the primary dataset were recruited as part of the British Columbia Lung Health Study[16] and Pan-Canadian Lung Health Study. Additional information about the study population can be found in the supplement. We previously profiled 267 bronchial airway brushings obtained from current and former smokers in this cohort using Affymetrix Human Gene 1.0 ST Arrays.[10] In the current study, we used this existing gene expression data together with spirometry data that were collected during longitudinal follow-up subsequent to bronchoscopy. FEV_1, FEV_1% predicted and FVC were measured using a flow-sensitive spirometer. The ratio between FEV_1 and FVC were used to determine COPD status as previously described in Steiling *et al*.[10] We excluded samples from individuals who did not have a spirometry recording within 1 year of their bronchoscopy (n=8), did not have at least two spirometry measurements at least 4 years apart (n=104) or who developed cancer (n=19). As previously reported, two samples were excluded due to sample labelling errors.[10] Data from the remaining 134 current and former smokers were included in the analysis. Because spirometry was not performed at regular intervals, the rate of FEV_1 decline (ΔFEV_1) for each study participant was estimated using linear regression with all available spirometry measurements from that individual subsequent bronchoscopy. The relationship between the rate of FEV_1 decline and other clinical variables was evaluated by analysis of variance (ANOVA) of linear models.

Identification of a rate of FEV_1 decline gene expression signature

Genes associated with the future rate of FEV_1 decline were identified using the following linear models calculated using the *lm* function and the *anova* function using R statistical software V.3.4.0[17] and RStudio V.1.0.143.[18]

$$\text{ge} \sim \beta_0 + \beta_1 X_{\text{age}} + \beta_2 X_{\text{Smoke\_Status}} + \beta_3 X_{\text{pack\_years}} + \beta_4 X_{\text{Sex}} + \beta_5 X_{\text{baseline\_FEV1}} + \epsilon \tag{1}$$

$$\text{ge} \sim \beta_0 + \beta_1 X_{\text{age}} + \beta_2 X_{\text{Smoke\_Status}} + \beta_3 X_{\text{pack\_years}} + \beta_4 X_{\text{Sex}} + \beta_5 X_{\text{baseline\_FEV1}} + \beta_6 X_{\triangle \text{lFEV1}} + \epsilon \tag{2}$$

where ge is the expression level of a single gene; age is the age at the time of bronchoscopy, pack years is the calculated cumulative cigarette smoke exposure at the time of bronchoscopy and smoke status is the smoking status at the time of bronchoscopy (participants were considered former smokers if they had quit for at least a year). Baseline FEV_1 is the FEV_1 within 1 year of bronchoscopy. The rate of FEV_1 decline (ΔFEV_1) is calculated as described above. ε is an error term. The false discovery rate (FDR) was calculated from the ANOVA p values.[19] Genes with FDR <0.05 were considered to be associated with the rate of FEV_1 decline and included in the signature. The signature was divided into genes that are increased or decreased with more rapid FEV_1 decline by hierarchical clustering which segregated genes according to the sign of the linear model coefficient. The rate of FEV_1 decline signature was compared with the previously published Steiling *et al*[10] airway gene expression signature of COPD severity by determining the number of genes overlapping between the signatures, and by using Gene Set Enrichment Analysis (GSEA).[20]

To evaluate the extent to which the FEV_1 decline signature is a reflection of other factors that correlate with COPD, we first summarised the expression of the FEV_1 decline signature per sample as the sample loading on the first principal component which we refer to as the FEV_1 decline signature. We then tested the association between the signature score and the rate of FEV_1 decline in several subsets of the data. A linear model adjusting for age, sex, smoking status, pack years and baseline FEV_1 was performed to test the association between the FEV_1 decline signature score and the rate of lung function decline in current smokers who continued to be current smokers throughout the follow-up period, former smokers, individuals with COPD, individuals without COPD and individuals who were not using inhaled medications.

Replication of the gene expression signature of rate of FEV_1 decline in GLUCOLD

We investigated the association between the expression of genes in the rate of FEV_1 decline signature and the observed rate of FEV_1 decline in a previously published independent dataset of individuals with COPD who were enrolled in the GLUCOLD trial, a placebo controlled randomised double-blind clinical trial of fluticasone with or without salmeterol[21 22] (GSE36221). Briefly, these participants underwent bronchoscopy with endobronchial biopsy followed by spirometry every 3 months during the 2.5-year trial. After the 2.5-year drug treatment trial, participants performed spirometry every year up to a total of 7.5 years (mean=6.91). The rate of FEV_1 decline was estimated by the coefficient from a linear model for each individual using their baseline spirometry measurement (t=year 0), excluding their time on treatment (t=0.25 to t=2.5 years) and including measurements from 3.5 years forward, to control for treatment effect. In the GLUCOLD participants, the gene expression signature associated with subsequent FEV_1 decline was calculated using principal component analysis (PCA). First, the eigenvector for the first principal component of the signature genes in the z-score normalised discovery set was calculated using the *prcomp* function in R. A summarised signature score for each sample in the discovery set and the GLUCOLD dataset was then calculated from the eigenvector and the z-score normalised expression data using the *predict* method of *prcomp*. The relationship between summarised signature score at baseline and the rate of FEV_1 decline was evaluated using the linear model and ANOVA strategy outlined above for the gene expression analysis.

Identification of enriched biological pathways

To identify transcription factors enriched in the FEV_1 decline signature, we used the Molecular Signature Database (MSigDB)[20] to search computationally derived datasets of transcription factor binding sites. We divided the genes into two clusters when searching: genes that increase with worse FEV_1 decline and genes that decrease with worse FEV_1 decline. For each cluster of genes, we identified transcription factor binding sites with an FDR <0.05. The transcription factor binding sites were based on Xie *et al's* work[23] and TRANSFAC V.7.4.[24]

The transcription factor XBP1, which was identified by the above method, was selected for in silico validation because it has previously been implicated in COPD[25] and because it is a well-studied transcription factor with several publicly available knockout and overexpression datasets. The Gene Expression Omnibus (GEO) was searched using the key terms "XBP1 knock out" and "XBP1 overexpression" to identify potentially useful datasets. This identified 38 datasets.

After searching GEO to identify publicly available datasets investigating the gene expression effects of modulating XBP1 activity, we identified two datasets which we explored further: a dataset examining the effects of XBP1 overexpression in mouse adipocytes (GSE46178)[26] and a dataset examining the effects of XBP1 knockout in mouse hepatocytes (GSE64824).[27] Using t-statistics from a linear model, we ranked genes by their change in expression following XBP1 overexpression in mouse adipocytes (n controls=4, n overexpression=4) using data from Affymetrix Mouse Genome 430A 2.0 Arrays. For the XBP1 knockout study, we ranked genes according to their change in expression between wild-type (WT) and XBP1-knockout hepatocytes by subtracting the gene expression of the controls (n=2) from the knockout hepatocytes (n=2) which had been profiled by RNA-seq using an Illumina HiSeq2000.[27] Before sequencing, the replicates were pooled. We explored the distribution of the two gene clusters from the rate of FEV_1 decline signature in these ranked lists using GSEA.[20] Importantly, neither of these in silico datasets were included in the TRANSFAC data.[28–31]

Because previous work has also identified association of a T helper type 2 cell (Th2) signature in a subset of individuals with COPD,[32] we sought to evaluate the association of this Th2 signature with the FEV_1 decline signature. We used a three-gene Th2 signature derived in patients with asthma,[33] which included POSTN, SERPINB2 and CLCA1. Expression levels of these genes were z-score normalised, and the first principal component was computed to summarise their expression. The association between the Th2 score and FEV_1 decline was assessed using a linear model controlling for age, sex, smoking status, pack years and baseline FEV_1. We also tested the association between the first principal component of the FEV_1 decline signature and the Th2 score.

To determine if genes related to mucus hypersecretion were associated with FEV_1 decline, we used GSEA to evaluate the enrichment of two mucin-related gene sets among a ranked list of all genes ranked by association with FEV_1 decline. The gene sets used included the mucin type O-glycan biosynthesis gene set from KEGG 2019,[34] and the mucin granule gene set from Jensen Compartments.[35] In order to determine if there was an association between ER stress and mucin hypersecretion, Gene Set Variation Analysis[36] scores were calculated for the XPB1 target gene set, the mucin-related gene sets and the FEV_1 signature score. Linear models were used to test pairwise associations between the three scores.

To determine if genes previously identified in genome-wide association studies (GWAS) of COPD were enriched in the FEV_1 decline signature, we used GSEA to evaluate the enrichment of a gene set comprised of COPD GWAS genes among the ranked list of all genes ordered by association with FEV_1 decline. The gene set of COPD GWAS genes was derived from a review of GWAS in COPD.[37]

Table 1 Characteristics of the study participants

	N=134		
	Mean	**SD**	**Range**
Age (years)	64	±6	49.33–77.17
Pack years (missing 8)	46	±16	12–102
Baseline FEV_1 (L)	2.48	±0.78	0.95–4.52
Baseline FEV_1 % predicted	82.16	±20.08	31–123
ΔFEV_1 (mL/year)	–33.72	±47.78	–170 to 170
Follow-up time (years)	6.38	±2.48	4.08–12.64
	N	**Per cent**	
Sex—male	75	55.97	
Inhaled medications—yes	20	14.93	
COPD status—yes	49	36.57	
Smoking status—current*	59	44.03	

All participants were current or former smokers. All participants had at least two spirometry measurements at least 4 years apart. The mean, SD and range are reported for continuous measures.
*Five current smokers quit during the period of follow-up spirometry.

RESULTS

Participant demographics

A total of 134 current and former smokers with (n=49) and without COPD (n=85) were included in this analysis. Clinical and demographic characteristics of the study cohort are provided in table 1. Demographic characteristics separated by GOLD status are available in online supplemental table 1. The correlations between demographic variables are listed in online supplemental table 2.

The average baseline FEV_1 for participants with COPD was significantly lower than in participants without COPD. The rate of FEV_1 decline is significantly higher in individuals with lower baseline FEV_1 and/or COPD ($p<0.05$) (online supplemental table 1). Only 5 of the 59 current smokers quit smoking during follow-up. The initial spirometry measurements were performed within 1 year of bronchoscopy, with 85% performed within 6 months of bronchoscopy and 70% within 90 days (online supplemental figure 1 and online supplemental table 3).

Bronchial airway gene expression signature of FEV_1 rate of decline

The expression levels of 171 genes were significantly associated with the rate FEV_1 decline (FDR <0.05) (figure 1 and online supplemental table 4). A total of 120 genes had higher expression in individuals with faster FEV_1 decline (cluster 1), while 51 genes had lower expression in individuals with faster FEV_1 decline (cluster 2).

We next wanted to determine the sensitivity of the results we obtained to using individuals with 4 or more years of longitudinal spirometry. Therefore, we performed the same analysis using individuals with a shorter duration of follow-up spirometry to determine if the increase in the number of individuals in the analysis resulting from requiring less follow-up would outweigh the potentially less accurate estimates of the rate of FEV_1 decline. No genes were significantly associated with FEV_1 decline when we required a minimum of 2 years or 3 years of spirometry follow-up. We compared the t-statistics obtained with the 4-year estimate of FEV_1 decline and the shorter estimates for the 171

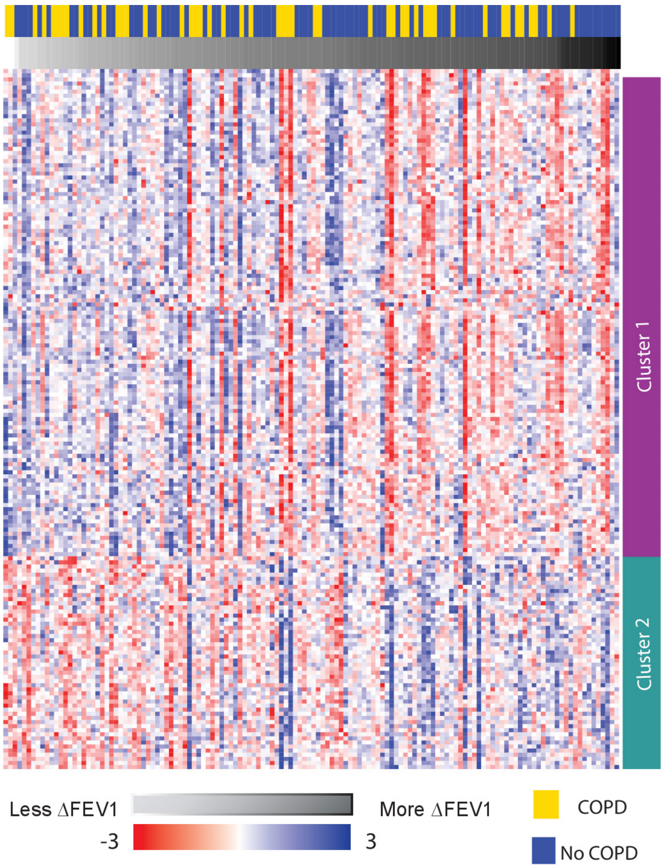


Figure 1 Heatmap of 171 genes associated with change in FEV_1. One hundred and seventy-one genes were associated with the change in the rate of FEV_1 decline using a linear model controlling for age, sex, smoking status, pack years and baseline FEV_1 (FDR <0.05). The participants (columns) are arranged from slowest FEV_1 decline (white) to most rapid FEV_1 decline (black). These genes were grouped into two clusters based on unsupervised hierarchical clustering. Cluster 1 consists entirely of genes expressed at higher levels in patients with more rapid FEV_1 decline, and cluster 2 consists entirely of genes expressed at lower levels in patients with more rapid FEV_1 decline. FDR, false discovery rate.

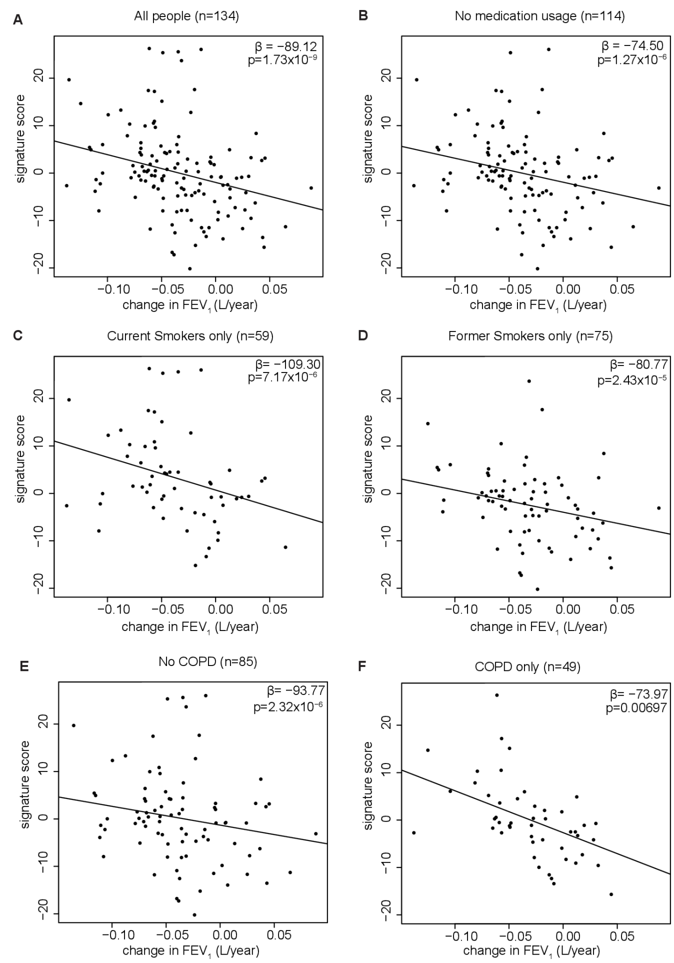


Figure 2 Sensitivity analysis of the FEV_1 decline signature. The expression levels of the 171 genes associated with FEV_1 decline were summarised into a single value using the eigenvector for the first principal component in the discovery dataset. This summarised expression is termed the FEV_1 decline signature score. The scatter plot and best-fit line of each panel show the association between the FEV_1 decline signature score and the change in FEV_1 in various subsets of the dataset. (A) All the individuals included in the training analysis, (B) only former smokers, (C) only current smokers, (D) only individuals not on inhaled medications, (E) only individuals without COPD, (F) only individuals with COPD.

genes in the FEV_1 decline signature. We did not detect a significant correlation of the t-statistics using the estimates of FEV_1 decline based on 4 years and 2 years of follow-up (p=0.369), but did identify a significant association between the t-statistics obtained using the 4-year and 3-year estimates of FEV_1 decline ($p=2.0\times10^{-16}$; online supplemental figure 2).

We also explored the sensitivity of our results to subsetting the cohort based on potential confounders. Toward this end, we summarised the expression of the genes associated with FEV_1 decline per sample as the first principal component from PCA and termed this value the FEV_1 decline signature score. Not surprisingly, using all samples, there is a strong association between the FEV_1 decline signature score and FEV_1 decline across the entire cohort ($p=1.73\times10^{-9}$; figure 2A). We also observed significant associations between the FEV_1 decline signature score and FEV_1 decline in current smokers, smokers who remained current smokers during follow-up, former smokers, individuals with COPD, individuals without COPD and individuals not on inhaled medication (figure 2). We also found that the association between the FEV_1 decline signature score and FEV_1 decline in former smokers remained even after correcting for the years of smoking cessation. Similarly, the association of the FEV_1 signature score and FEV_1 decline remained after correcting for time interval between spirometry and bronchoscopy.

We next examined the overlap between the longitudinal FEV_1 decline signature and the Steiling *et al* bronchial airway 98-gene expression signature of COPD severity we had previously identified in a cross-sectional analysis.[10] A total of 10 genes were shared by the two signatures. Nine genes that were increased in association with worse FEV_1 decline were also increased in COPD. One gene that was decreased with more rapid FEV_1 decline was also decreased in COPD airway (online supplemental table 4). Next, GSEA was also used to evaluate for enrichment between these signatures. There was significant concordant enrichment of the 171 gene signatures associated with rate of FEV_1 decline and COPD status ranked list (positive FDR q<0.0001, negative FDR q<0.0001).

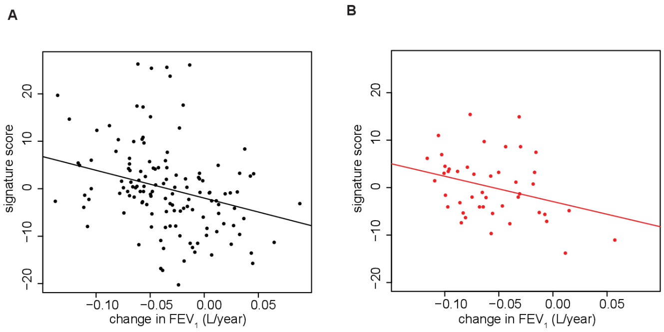


Figure 3 Airway gene expression signature associated with rate of FEV_1 decline replicates in an independent dataset of patients with COPD. The principal component 1 eigenvector used to generate the FEV_1 decline signature score in the training data was also used to generate signature scores from baseline bronchial biopsies of patients with COPD who were followed for subsequent change in FEV_1. The signature scores in the discovery dataset (A) and the independent dataset (B) are each significantly correlated with the rate of FEV_1 decline ($p=1.73\times10^{-9}$ and p=0.018, respectively).

Replication of the airway gene expression signature of rate of FEV_1 decline in the GLUCOLD trial

We next sought to determine whether the rate of FEV_1 decline signature is significantly associated with rate of FEV_1 change in an independent dataset using the signature score derived from PCA. Demographics of this replication cohort are listed in online supplemental table 5. In the discovery dataset, higher signature scores are associated with a more rapid decrease in FEV_1 ($p=1.73\times10^{-9}$; figure 3A). Signature scores generated in an independent dataset of individuals with COPD who were enrolled in a placebo controlled study of inhaled fluticasone±salmeterol[21] showed that the scores are significantly associated with future FEV_1 decline in the independent dataset (p=0.018; figure 3B). These findings suggest that the airway gene expression signature for the rate of decline in FEV_1 is similarly associated with rate of FEV_1 decline in an independent dataset of participants with COPD.

Enrichment of transcription factor binding sites in the FEV_1 rate of decline signature

To explore the potential regulators of the genes associated with the rate of FEV_1 decline, we queried MSigDB to identify transcription factors whose predicted binding sites are overrepresented among the signature genes. Genes whose expression levels are increased in individuals with more rapid FEV_1 decline are enriched for genes with binding sites for XBP1 (FDR q=0.0302) among other transcription factors. A full list of all the MSigDB results can be found in online supplemental table 6.

We further investigated XBP1 due to its role in the UPR which has been implicated in COPD,[25] and we are able to identify publicly available datasets profiling the gene expression effects of modulating XBP1 activity. Using GEO, we identified a dataset examining the effect of XBP1 overexpression in mouse adipocytes (GSE46178)[26] and a dataset examining the effect of XBP1 knockout in mouse hepatocytes (GSE64824)[27] that we used in further analysis.

We ranked the genes that change with XBP1 overexpression in mouse adipocytes (n=4 controls, n=4 XBP1 overexpression) and used GSEA to examine the distribution of genes in the FEV_1 decline signature in this ranked list. Genes whose expression is increased in individuals with more rapid decline in FEV_1 are

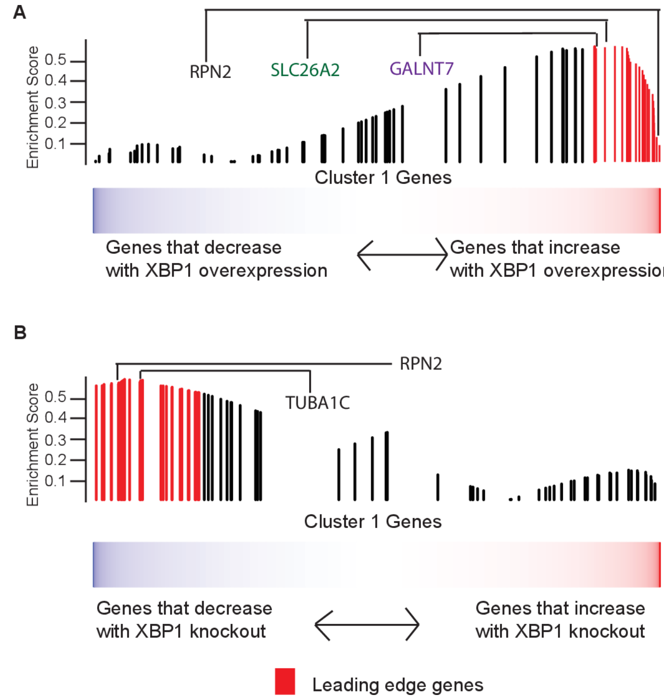


Figure 4 Genes increased in individuals with faster FEV_1 decline are among the genes most induced by XBP1 overexpression and genes increased in individuals with faster FEV_1 decline are among the genes most decreased by XBP1 knockout. (A) Genes that increase in individuals with more rapid FEV_1 decline are significantly enriched among the genes that are most induced by XBP1 overexpression in murine adipocytes (GSEA p<0.0001). The vertical lines are the position of the genes with increased expression in individuals with more rapid FEV_1 decline in a list of all genes ranked from most induced by XBP1 overexpression to most repressed. The height of the vertical line represents the running enrichment score, and the lines highlighted in red represent the leading edge. (B) Genes that increase in individuals with more rapid FEV_1 decline are significantly enriched among the genes that are most repressed in murine hepatocytes deleted for XBP1 (GSEA p=0.025). We previously found the gene highlighted in green to have increased expression in bronchial epithelium from individuals with COPD.[10] The gene highlighted in purple is involved in mucin type O-glycan biosynthesis. The genes in black are predicted targets of XBP1. GSEA, Gene Set Enrichment Analysis.

enriched among the genes that are induced by XBP1 overexpression (p<0.0001) (figure 4A). We next selected the subset of genes contributing the most to this significant enrichment, which are also known as the leading edge genes. We used these leading edge genes to plot a heatmap across the human bronchial airway gene expression data and the mouse adipocyte data. The leading edge genes that increased with more rapid FEV_1 decline were also increased with XBP1 overexpression (figure 4A).

We also created a ranked list of genes based on expression changes in XBP1-knockout versus WT-control hepatocytes from 16-week-old mice (n=2 controls, n=2 XBP1 knockout).[26] Genes whose expression is increased in individuals with more rapid FEV_1 decline are enriched among the genes that decreased in XBP1 knockout (p=0.025; figure 4B). There were 22 leading edge genes from the XBP1 overexpression analysis and 35 in the knockout analysis. Fifteen of these genes overlapped between the two sets. We also compared the leading edge genes to the signature genes that were in the XBP1 transcription binding site list (n=133). Of the four genes that overlap between the FEV_1

decline signature and the XBP1 transcription factor binding site gene set, three of them were also in the leading edge of the analyses (GALE, SEC61A1 and ARMCX3). Together, these data suggest that XBP1-regulated genes are among the genes with increased expression in bronchial epithelial cells of individuals with more rapid FEV_1 decline.

In addition to investigating the potential role of XBP1 in the FEV_1 decline signature, we used a similar approach to evaluate the potential role of other significantly enriched transcription factors. We focused on ATF6, SP1 and FOS/JUN because they have been previously implicated in COPD, and we found datasets examining the gene expression effects of modulating these transcription factors (GSE124797, GSE87298, GSE37935, GSE31628, GSE97226, GSE7742). We examined the perturbation of genes either increased or decreased with more rapid FEV_1 decline among the genes whose expression was most altered by transcription factor perturbation and found no significant association in the datasets we examined which modulated ATF6, SP1, c-FOS or JNK (data not shown).

Exploration of other biological pathways

Given the possibility that elevated XBP1 activity might be due to ER stress resulting from elevated protein secretion, we next sought to determine if the FEV_1 decline signature might be associated with genes involved in mucin production and secretion, given that airway mucus hypersecretion is a well-established feature of COPD.[37] We evaluated two curated gene sets consisting of genes involved in mucin biosynthesis, including the KEGG 2019 mucin type O-glycan biosynthesis pathway[35] and the Jensen Compartments mucin granule gene set.[34] We identified significant enrichment of both gene sets among the genes expressed more highly in individuals with more rapid FEV_1 decline (KEGG 2019 FDR $q<1\times10^{-4}$; Jensen FDR $q=0.008$). The leading edge genes contributing most to the enrichment of these gene sets included several polypeptide N-acetylglucosaminyltransferase genes, several mucin genes and carcinoembryonic antigen cell adhesion molecule 5 (online supplemental table 7). We also examined the potential associations between ER stress (using the expression of predicted targets of XBP1 as a surrogate of ER stress-driven XBP1 activity), mucus hypersecretion and the signature of FEV_1 decline. We found that the association between the summarised expression of the mucin gene sets and the summarised expression of the predicted targets of XBP1 is less pronounced ($p=0.00234$) than the associations between the summarised expression of the predicted targets of XBP1 or the summarised expression of the mucin-related genes and the FEV_1 decline signature ($p=2.0\times10^{-16}$ and $p=4.82\times10^{-12}$, respectively).

We also examined the potential association between markers of Th2 inflammation and FEV_1 decline based on previous work describing an association of a Th2 signature in a subset of individuals with COPD and without a clinical history of asthma.[32] Using a three-gene Th2 score that had previously been developed in patients with asthma,[33] we found a significant association between Th2 inflammation and rate of change in FEV_1 when controlling for age, sex, smoking status, pack years and baseline FEV_1 ($p=0.045$). We also found a significant correlation between the FEV_1 decline signature and the three-gene Th2 score ($r^2=0.1187$; $p=2.71\times10^{-5}$).

We were also interested whether the genes in the FEV_1 decline signature included genes previously identified in GWAS of COPD. We found that a gene set composed of genes previously identified in COPD GWAS[38] was significantly enriched among genes that are decreased in individuals with more rapid FEV_1 decline (GSEA $p=0.035$). Genes in the leading edge of this analysis included ARMC2, DLG2, THSD4, KLHL7, CCDC101, ANKH and CHRNA3.

DISCUSSION

We have identified gene expression differences at baseline in bronchial epithelium from current and former smokers that are associated with the subsequent rate of change in FEV_1. We have replicated the airway gene expression signature of FEV_1 decline in an independent dataset of participants with COPD. There is significant enrichment of the FEV_1 decline signature among genes ranked according to the presence or severity of COPD, suggesting that the FEV_1 decline signature is related to the Steiling *et al*[10] airway gene expression signature of COPD. Interestingly, we found that a subset of the airway gene expression changes associated with more rapid FEV_1 decline may be in part explained by increased activity of the transcription factor XBP1 and mucus hypersecretion. There are also significant associations between the FEV_1 decline signature and a signature of Th2 inflammation.

The replication of the gene expression signature of FEV_1 decline in endobronchial biopsies from participants with COPD in the GLUCOLD trial is notable in two regards. First is the replication of the signature in a different sample type (bronchial biopsies vs brushes). Second, as GLUCOLD is comprised of only patients with COPD, replication in this cohort suggests that the signature can be relevant to COPD progression. Further studies should be conducted to determine whether the gene expression signature of FEV_1 decline can be used to predict which individuals are at risk of more rapid disease development or progression.

The transcription factor binding sites enriched in the rate of FEV_1 decline signature support protein response[39] is hypothesised to play a role in the development of COPD.[25 40–42] Bronchial airway expression of XBP1 targets is increased in advance of FEV_1 decline in our data, even when controlled for smoking status.[43] XBP1 has been previously shown to increase in human bronchial epithelial cells derived from patients with COPD compared with cells derived from smokers without COPD and non-smokers.[44] Elevated XBP1 increases the expression of cytokines (interleukin 8 (IL-8) and IL-1β) and Th1 chemokines.[25] These cytokines have been shown to be elevated in COPD.[45]

XBP1's potential role to contribute to the gene expression changes associated with more rapid FEV_1 decline is interesting given the role of XBP1 in the UPR to ER stress.[25 41 42 46] The UPR is triggered by protein misfolding and/or ER overload, that may result from oxidative stress caused by cigarette smoking. The UPR is mediated by three families of signal transducers, including the IRE1/XBP1 pathway.[47] XBP1 binds to ER stress elements and increases the transcription of chaperone proteins that assist in protein folding and reducing protein synthesis.[39] Both ER stress and XBP1 expression levels are increased by cigarette smoking.[40 41] Dysregulation of proteostasis and the UPR have previously been described in patients with smoking-associated COPD and have been implicated in COPD progression.[48] While potential reducers of XBP1 such as IRE1α inhibitors[49] and a synthetic analogue of TRPC1[50] have previously been investigated in the setting of ER stress in cancer, this is the first study to our knowledge to identify XBP1 as a potential regulator of gene expression changes associated with the rate of FEV_1 decline. We have found that gene expression consequences of XBP1 perturbation in a cell line and a mouse model recapitulate components of the FEV_1 decline signature, supporting a potential regulatory role for XBP1 in the processes that contribute to the rate of FEV_1 decline. However, further studies are needed to evaluate the effects of perturbing XBP1 in human airway epithelial cells, and the possible role of XBP1 in COPD pathogenesis and disease progression.

We also found the expression of several mucin-related genes (GALNT4, GALNT5, GALNT12, MUC2) to be increased in association with more rapid FEV_1 decline.[51 52] The observation of increased mucin-associated gene expression with more rapid FEV_1 decline is further supported by the enrichment of two mucin-related gene sets[34 35] among genes increased with more rapid FEV_1 decline. Further studies are needed to determine if the apparent increase in XBP1 activity associated with more rapid FEV1 decline might be a consequence of increased mucin production. We also identified enrichment of COPD GWAS genes among genes whose airway expression decreases with more rapid FEV_1 decline. These genes include CCDC101, also known as SGF29, which is a subunit of a histone acetyltransferase complex; THSD4, which has metalloendopeptidase activity; and CHRNA3, which is a member of the nicotinic acetylcholine receptor family of proteins.[53–55] Together, these findings support the biological relevance of the FEV_1 decline signature.

We have also identified a significant association between a Th2 score and the FEV_1 decline signature in current and former smokers with and without COPD. Th2 cells mediate the inflammatory response that drives a subtype of asthma which is associated with a more favourable response to corticosteroids.[33] While Th2 inflammation is traditionally associated with asthma, previous work has shown that a subset of patients with COPD and without a clinical history of asthma have increased expression of Th2-associated genes.[32] This suggests that a similar process leads to airflow obstruction in asthma, and in a subgroup of patients with COPD without a clinical history of asthma. In this study, we show that expression of Th2-associated genes is also associated with more rapid decline in FEV_1. This finding lends additional support to the hypothesis that a subgroup of patients with COPD have more 'asthma-like' molecular features which also place them at risk of faster decline in FEV_1.

Though the data described here support a replicable gene expression signature of FEV_1 decline, there are limitations to the cohort and samples from which it was derived. First, the signature was derived in an older population of individuals with and at risk of COPD and may not represent a wider spectrum of disease or normal airway biology. Furthermore, it is possible that FEV_1 decline may be quicker in the earlier stages of COPD, and that our FEV_1 decline signature reflects disease severity (despite our having controlled for FEV_1 in the analysis), or that factors other than those represented by our gene expression signature might influence the rate of FEV_1 decline at other points on the disease-severity spectrum. Second, the signature was derived from a cohort that was initially recruited as part of cancer-related studies. As a result, details about some COPD-related traits are unavailable, and the frequency of spirometry was variable. We found that a minimum of 4 years of follow-up spirometry was required to detect gene expression differences associated with the rate of FEV_1 decline, presumably because of variability in the measurement of FEV_1, and this may have limited our study power, especially in subgroup analyses. For example, we were unable to evaluate the effect of smoking cessation on the signature due to the low number of individuals who quit smoking during follow-up, or gene expression differences associated with the rate of change of other lung function metrics. Similarly, we were unable to evaluate the association of bronchodilator responsiveness, symptoms or exacerbations as these data were not available for the cohort. Race is known to impact FEV_1 decline,[56] but we were unable to evaluate the effect of race on FEV_1 decline as the majority of samples were obtained from individuals who were white which may limit the generalisability of the signature. Additionally, while bronchoscopy brushings are composed of predominantly epithelial cells,[21 57] we do not know the exact cell types that lead to the observed gene expression patterns, and variations in the cellular composition of the bronchial airway epithelium might contribute to the FEV_1 decline signature. Finally, the signature was also derived using data from microarrays generated as part of a previous study. Gene expression profiling by RNA-seq could potentially improve signal-to-noise characteristics and allow for the identification of novel transcripts.

While we have derived a signature of FEV_1 decline from smokers with and at risk of COPD and validated these findings in an independent cohort of individuals with COPD, our analysis indicates that there are individuals who have a pattern of gene expression that is not entirely consistent with their observed rate of FEV_1 decline. Beyond the potential issues of confounders and other limitations of our study described above, it is also likely that the biological process or processes that contribute to the observed gene expression signature may not be perfect predictors of FEV_1 decline. FEV_1 has been found to associate with the rate of FEV_1 decline.[58] Because we have previously described a significant number of genes associated with baseline FEV_1,[21] our analysis included baseline FEV_1 as a covariate. However, low baseline FEV_1 may reflect low peak FEV_1 in early adulthood, or accelerated FEV_1 decline up to the time of baseline FEV_1 measurement and controlling for baseline FEV_1 may obscure the impact of FEV_1 on the rate of FEV_1 decline. The FEV_1 decline signature was derived using samples from smokers with and at risk of COPD, and was validated in a population of subjects with COPD. While this suggests a specific utility of the signature for assessing progression of COPD, we were unable to assess whether the signature reflects FEV_1 decline in other clinical contexts.

In exploring the implications of the gene expression alterations associated with FEV_1 decline, we largely focused our analysis on targets of known transcription factors given the potential to generate straightforward hypotheses about the molecular regulation of the observed gene expression changes. However, there is the potential to gain additional insights from future studies focused on other molecular enrichments. The XBP1 knockout and overexpression datasets where we observed significant enrichment of the FEV1 decline signature genes are derived from adipose cells and hepatocytes. Transcription factors have been shown to have effects that are similar in non-lung cells[59] and these changes have been used to identify new targets.[60] Other studies have suggested that there are cell-specific effects.[61] The changes from knockout or overexpression in these datasets may not apply fully in lung cells.

Despite these limitations, we have identified and replicated gene expression differences associated with the rate of subsequently observed FEV_1 decline using baseline gene expression profiling of bronchoscopy brushings. Further studies using larger sample sizes are needed to determine whether airway gene expression profiling can prospectively identify individuals who will experience more rapid FEV_1 decline and whether this would also apply in more severe COPD, which may present with more parenchymal emphysema than primary airway disease. A previous study from Boudewijn *et al* identified a nasal gene expression signature associated with severe COPD versus controls that was also similar to the previously identified bronchial airway gene expression signature of COPD from Steiling *et al*.[10 62] It may therefore be possible to identify a nasal gene expression signature of FEV_1 decline, which would be less invasive than a bronchial signature. Such markers could be used to stratify patients with and at risk of COPD, and to potentially evaluate the response to therapies aimed at diminishing the rate of FEV_1 decline.

Author affiliations
[1]Division of Computational Biomedicine, Boston University School of Medicine, Boston, Massachusetts, USA
[2]Bioinformatics Program, Boston University, Boston, Massachusetts, USA
[3]Respiratory Bioinformatics and Molecular Biology (RBMB), School of Life Sciences, University of Technology Sydney, Sydney, New South Wales, Australia
[4]Department of Pulmonary Diseases, University Medical Center Groningen, Groningen, The Netherlands
[5]Department of Pathology and Medical Biology, University Medical Centre Groningen, Groningen, The Netherlands
[6]Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands
[7]Internal Medicine Residency Program, Boston Medical Center, Boston, Massachusetts, USA
[8]Department of Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts, USA
[9]Division of Pulmonary, Allergy, Sleep, and Critical Care Medicine, Boston University School of Medicine, Boston, Massachusetts, USA
[10]British Columbia Cancer Agency, Vancouver, British Columbia, Canada

Contributors KS, AS and MEL conceived the idea. KS, AS, MEL, MvdB, AF and GO provided guidance during the analysis. EJB, AF and KC performed the computational analysis. WT, PSH, SL, MvdB and AF collected the original data. GL, XX and YOA profiled the data. EJB drafted the manuscript and all the authors read and provided feedback.

Funding National Institutes of Health/National Heart, Lung, and Blood Institute (R01HL095388 and R01HL118542-01) and Dutch Longfonds Foundation (4.2.16.132JO).

Competing interests GO reports unrelated personal fees from AstraZeneca and grants from Janssen Pharmaceuticals. PSH reports unrelated grants from Boehringer Ingelheim and Galapagos. MvdB reports unrelated research grants from GlaxoSmithKline, TEVA Pharmaceuticals and Chiesi. AS reports unrelated founder's equity from Metera Pharmaceuticals as well as grants and personal fees from Janssen Research and Development. KS reports grants from CHEST Foundation, unrelated grants from Lungevity Foundation Early Detection Award and royalties from UpToDate. MEL reports unrelated founder's equity from Metera Pharmaceuticals and unrelated grants from Janssen Research and Development. WT reports unrelated personal fees from Pfizer, GSK, Roche Diagnostics/Ventana, Merck Sharp Dohme, Novartis, Lilly Oncology, Boehringer Ingelheim, AstraZeneca, Bristol-Myers-Squibb and AbbVie. KS, MEL and AS have US patent 9,677,138 issued. EJB, KS, MEL and AS have a relevant patent pending (application no. 62/916,431).

Patient consent for publication Not required.

Ethics approval This study was approved by the ethics committees and institutional review boards of all participating study sites and informed written consent was acquired from all individuals included in the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The data are available on GEO as GSE37147.

ORCID iDs
Wim Timens http://orcid.org/0000-0002-4146-6363
Pieter S Hiemstra http://orcid.org/0000-0002-0238-5982

REFERENCES

1 WHO. The top 10 causes of death, 2020. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
2 Mannino DM, Buist AS, Petty TL, *et al*. Lung function and mortality in the United States: data from the first National health and nutrition examination survey follow up study. *Thorax* 2003;58:388–93.
3 Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of lung health study participants after 11 years. *Am J Respir Crit Care Med* 2002;166:675–9.
4 Nishimura M, Makita H, Nagai K, *et al*. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185:44–52.
5 Scanlon PD, Connett JE, Waller LA, *et al*. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. the lung health study. *Am J Respir Crit Care Med* 2000;161:381–90.
6 Higashimoto Y, Iwata T, Okada M, *et al*. Serum biomarkers as predictors of lung function decline in chronic obstructive pulmonary disease. *Respir Med* 2009;103:1231–8.
7 Oelsner EC, Balte PP, Grams ME, *et al*. Albuminuria, lung function decline, and risk of incident chronic obstructive pulmonary disease. The NHLBI pooled cohorts study. *Am J Respir Crit Care Med* 2019;199:321–32.
8 Esquinas C, Serreri S, Barrecheguren M, *et al*. Long-Term evolution of lung function in individuals with alpha-1 antitrypsin deficiency from the Spanish registry (REDAAT). *Int J Chron Obstruct Pulmon Dis* 2018;13:1001–7.
9 Beane J, Vick J, Schembri F, *et al*. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-seq. *Cancer Prev Res* 2011;4:803–17.
10 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;187:933–42.
11 AEGIS Study Team. Shared gene expression alterations in nasal and bronchial epithelium for lung cancer detection. *J Natl Cancer Inst* 2017;109. doi:10.1093/jnci/djw327. [Epub ahead of print: 01 07 2017].
12 Becker EJ, Faiz A, Van Den Berge M. *A bronchial airway gene expression signature of future lung function decline is enriched in XBP1-Regulated Genes. In: D107 mitochondria and er stress in homeostasis and repair*. New York: American Thoracic Society, 2019: A7234.
13 Becker EJ, Faiz A, Alekseyev O. Derivation of a bronchial airway gene expression signature associated with FEV1 decline. *Am Thorac Soc Int Conf Meet Abstr*:A7159.
14 Becker EJ, Faiz A, Lenburg M. A bronchial airway gene expression signature of lung function decline. *Am Thorac Soc Int Conf Meet Abstr* 2017:A7036.
15 Becker E, Lam S, van den Berge M, den BMvan, *et al*. Predicting lung function decline in COPD using bronchial airway gene expression. *Chest* 2016;150:912A.
16 Tammemagi MC, Lam SC, McWilliams AM, *et al*. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. *Cancer Prev Res* 2011;4:552–61.
17 R Documentation. R: fitting linear models, 2017. Available: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/lm.html
18 RStudio Team. *RStudio: integrated development for R*. Boston, MA: RStudio, Inc, 2016. www.rstudio.com
19 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* 1995;57:289–300.
20 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 U S A* 2005;102:15545–50.
21 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.
22 Lapperre TS, Snoeck-Stroband JB, Gosman MME, *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.
23 Xie X, Lu J, Kulbokas EJ, *et al*. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 2005;434:338–45.
24 Matys V, Kel-Margoulis OV, Fricke E, *et al*. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006;34:D108–10.
25 Kelsen SG. The unfolded protein response in chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2016;13 Suppl 2:S138–45.
26 Gregor MF, Misch ES, Yang L, *et al*. The role of adipocyte XBP1 in metabolic regulation during lactation. *Cell Rep* 2013;3:1430–9.
27 Liu X, Henkel AS, LeCuyer BE, *et al*. Hepatocyte X-box binding protein 1 deficiency increases liver injury in mice fed a high-fat/sugar diet. *Am J Physiol Gastrointest Liver Physiol* 2015;309:G965–74.
28 Vilen BJ, Cogswell JP, Ting JP. Stereospecific alignment of the X and Y elements is required for major histocompatibility complex class II DRA promoter function. *Mol Cell Biol* 1991;11:2406–15.
29 Ono SJ, Liou HC, Davidon R, *et al*. Human X-box-binding protein 1 is required for the transcription of a subset of human class II major histocompatibility genes and forms a heterodimer with c-fos. *Proc Natl Acad Sci U S A* 1991;88:4309–12.
30 Liou HC, Boothby MR, Finn PW, *et al*. A new member of the leucine zipper class of proteins that binds to the HLA DR alpha promoter. *Science* 1990;247:1581–4.
31 Yoshimura T, Fujisawa J, Yoshida M. Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent enhancer of HTLV-1: all contain a leucine zipper structure and basic amino acid domain. *Embo J* 1990;9:2537–42.
32 Christenson SA, Steiling K, van den Berge M, *et al*. Asthma-COPD overlap. clinical relevance of genomic signatures of type 2 inflammation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2015;191:758–66.
33 Bhakta NR, Solberg OD, Nguyen CP, *et al*. A qPCR-based metric of Th2 airway inflammation in asthma. *Clin Transl Allergy* 2013;3:24.

34 Kanehisa M, Sato Y, Furumichi M, *et al*. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019;47:D590–5.
35 Binder JX, Pletscher-Frankild S, Tsafou K, *et al*. Compartments: unification and visualization of protein subcellular localization evidence. *Database* 2014;2014:bau012.
36 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 2013;14:7.
37 Cerveri I, Brusasco V. Revisited role for mucus hypersecretion in the pathogenesis of COPD. *Eur Respir Rev* 2010;19:109–12.
38 Ranjan A, Singh A, Walia GK, *et al*. Genetic underpinnings of lung function and COPD. *J Genet* 2019;98.
39 Yoshida H, Matsui T, Yamamoto A, *et al*. Xbp1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 2001;107:881–91.
40 Jorgensen E, Stinson A, Shan L, *et al*. Cigarette smoke induces endoplasmic reticulum stress and the unfolded protein response in normal and malignant human lung cells. *BMC Cancer* 2008;8:229.
41 Geraghty P, Wallace A, D'Armiento JM. Induction of the unfolded protein response by cigarette smoke is primarily an activating transcription factor 4-C/EBP homologous protein mediated process. *Int J Chron Obstruct Pulmon Dis* 2011;6:309–19.
42 Kitaguchi Y, Taraseviciene-Stewart L, Hanaoka M, *et al*. Acrolein induces endoplasmic reticulum stress and causes airspace enlargement. *PLoS One* 2012;7:e38038.
43 Martinon F, Chen X, Lee A-H, *et al*. TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. *Nat Immunol* 2010;11:411–8.
44 Geraghty P, Baumlin N, Salathe MA, *et al*. Glutathione peroxidase-1 suppresses the unfolded protein response upon cigarette smoke exposure. *Mediators Inflamm* 2016;2016:1–16.
45 Chung KF. Cytokines in chronic obstructive pulmonary disease. *Eur Respir J* 2001;18:50–9.
46 Kelsen SG, Duan X, Ji R, *et al*. Cigarette smoke induces an unfolded protein response in the human lung. *Am J Respir Cell Mol Biol* 2008;38:541–50.
47 Walter P, Ron D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* 2011;334:1081–6.
48 Min T, Bodas M, Mazur S, *et al*. Critical role of proteostasis-imbalance in pathogenesis of COPD and severe emphysema. *J Mol Med* 2011;89:577–93.
49 Mimura N, Fulciniti M, Gorgun G, *et al*. Blockade of XBP1 splicing by inhibition of IRE1α is a promising therapeutic option in multiple myeloma. *Blood* 2012;119:5772–81.
50 Overley-Adamson B, Artlett CM, Stephens C, *et al*. Targeting the unfolded protein response, XBP1, and the NLRP3 inflammasome in fibrosis and cancer. *Cancer Biol Ther* 2014;15:452–62.
51 Wang G, Xu Z, Wang R, *et al*. Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. *BMC Med Genomics* 2012;5:21.
52 Gosens R, Zaagsma J, Meurs H, *et al*. Muscarinic receptor signaling in the pathophysiology of asthma and COPD. *Respir Res* 2006;7:73.
53 Stelzer G, Rosen N, Plaschkes I, *et al*. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1–1.30.33.
54 Fishilevich S, Nudel R, Rappaport N, *et al*. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017;2017. doi:10.1093/database/bax028. [Epub ahead of print: 01 01 2017].
55 Rappaport N, Twik M, Plaschkes I, *et al*. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res* 2017;45:D877–87.
56 Molfino NA. Genetic predisposition to accelerated decline of lung function in COPD. *Int J Chron Obstruct Pulmon Dis* 2007;2:117–9.
57 Spira A, Beane J, Shah V, *et al*. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 2004;101:10143–8.
58 Tashkin DP. Variations in FEV_1 decline over time in chronic obstructive pulmonary disease and its implications. *Curr Opin Pulm Med* 2013;19:116–24.
59 GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, *et al*. Genetic effects on gene expression across human tissues. *Nature* 2017;550:204.
60 Campbell JD, McDonough JE, Zeskind JE, *et al*. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med* 2012;4:67.
61 Lamb J, Crawford ED, Peck D, *et al*. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;313:1929–35.
62 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.

A bronchial gene expression signature associated with rate of subsequent FEV1 decline in individuals with and at risk of COPD

Elizabeth J Becker, Alen Faiz, Maarten van den Berge, Wim Timens, Pieter S. Hiemstra, Kristopher Clark, Gang Liu, Xiaohui Zhang, Yuriy O Alekseyev, George O'Connor, Stephen Lam, Avrum Spira, Marc E Lenburg, Katrina Steiling

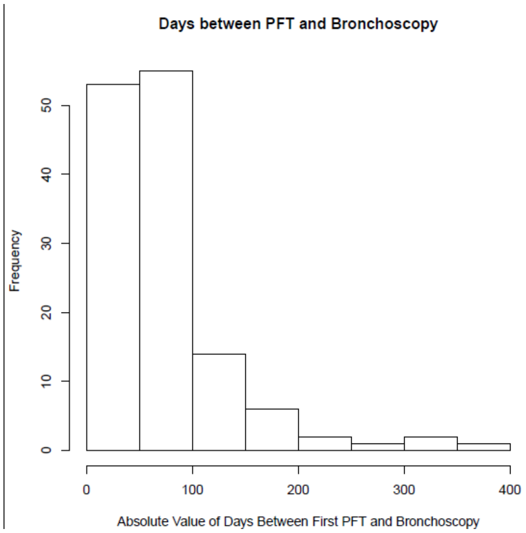
Supplemental Methods

Bronchial airway brushings were obtained during bronchoscopy from individuals recruited as part of the British Columbia Lung Health Study [1] and the Pan-Canadian Lung Health Study. As part of a prior cross-section study of airway gene expression in COPD, a total of 267 bronchial brushing samples were selected to ensure matching for covariates between individuals with and without COPD, and analyzed using Affymetrix Human Gene 1.0 ST Arrays[2]. All subjects provided written informed consent.

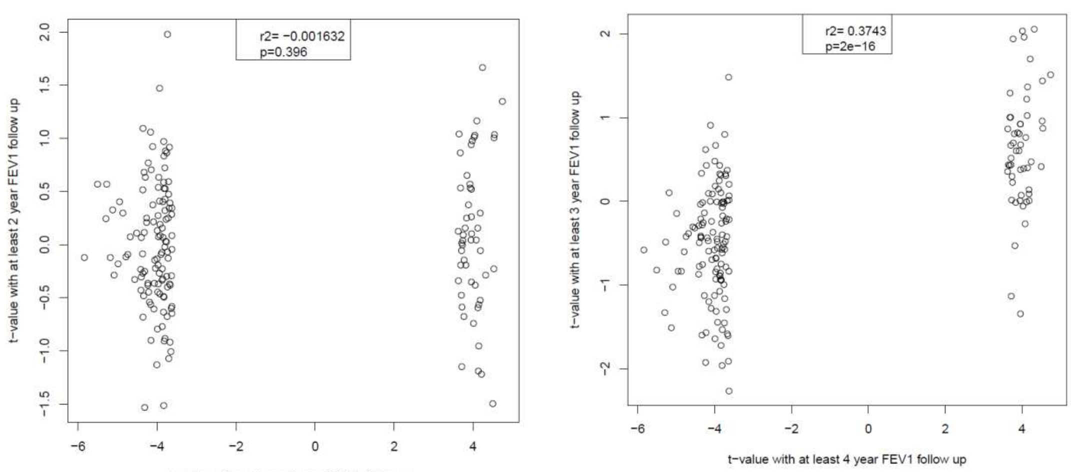
For this secondary analysis, we used the existing publicly available microarray data from the prior study (GSE37147). As in the prior study, we excluded individuals who developed lung cancer or who did not have spirometry measurements within one year of bronchoscopy. We restricted the secondary analysis to individuals with longitudinal spirometry measurements available as part of the parent study. Institutional Review Board approval for this study was obtained from all participating institutions.

Supplemental Figures

Supplemental Figure 1: Histogram showing the number of days between baseline spirometry and bronchoscopy.



Supplemental Figure 2: Determining the FEV_1 measurement interval required to identify gene expression differences associated with FEV_1 decline. Left: The correlation between the t-values of the 171 FEV_1 decline signature genes obtained using change in FEV_1 ascertained over 2 years vs. the t-values obtained using change in FEV_1 ascertained over 4 years. Right: The correlation between the t-values of the signature genes obtained using change in FEV_1 ascertained over 3 years and t-values obtained using change in FEV_1 ascertained over 4 years.



Supplemental Tables

Supplemental Table 1: Demographics of the training cohort separated by GOLD and COPD status. The mean, standard deviation, and range are shown for continuous traits. * missing 1, + missing 3, § missing 4.

	GOLD1			GOLD2			GOLD3			No COPD		
N	22			40			9			63		
	mean	**standard deviation**	**range**	**mean**	**standard deviation**	**range**	**mean**	**standard deviation**	**range**	**mean**	**standard deviation**	**range**
age (years)	64	6	55.08 - 72.33	65	6	54.17 - 75.17	61	5	51.92 - 66.92	64	5	49.33 - 77.17
pack years	43*	20	12 - 102	47+	18	20 - 95.8	53	13	34.02 - 72	46§	15	13 - 102
Change in FEV_1 (mL)	-55.33	43.31	140 - 30	-30.81	36.35	-100 - 40	21.48	-72.41	-70 - 170	-35.91	46.23	-170 - 60
Change in FEV_1 % predicted	-0.73	1.80	-3.92 – 2.53	-0.39	1.31	-2.50 – 2.77	0.92	2.36	-1.77 – 5.09)	-0.12	1.63	-4.03 – 3.38
Baseline FEV_1 (mL)	2811.82	756.03	1190 - 4190	1999.25	425.91	1240 - 2840	1294.44	236.07	950 - 1700	2846.03	687.89	1730 - 4520
Baseline FEV_1 % predicted	91.27	8.07	81-112	65.4	7.37	50-78	40.67	5.54	31-48	95.56	12.51	71-123
Follow up time	5.51	1.72	4.11 - 9.13 years	8.01	3	4.24 - 12.64	7.4	3.09	4.14 - 11.62	5.51	1.56	4.08 - 9.22
	n	**Percent**		**n**	**Percent**		**n**	**Percent**		**n**	**percent**	
Smoking Status = current	11	50.00%		14	35.00%		2	22.22%		32	50.79%	
Sex= male	13	59.09%		24	60.00%		6	66.67%		32	50.79%	
inhaled medication = yes	4	18.18%		11	27.50%		1	11.11%		4	6.35%	
Race= Asian	0	0.00%		2	5.00%		0	0.00%		2	3.17%	

Supplemental Table 2A: The coefficients from a Pearson's correlation test between each variable.

	ΔFEV1	age	sex	Smoking status	Pack Years	Baseline FEV1	COPD status	Follow up time
ΔFEV1	1.0000	-0.1196	0.0801	0.0124	0.0629	-0.2997	0.1996	0.0859
age	-0.1196	1.0000	-0.0332	0.2497	0.0752	-0.2098	0.0611	0.1118
sex	0.0801	-0.0332	1.0000	-0.0612	-0.0895	-0.4964	-0.0804	0.0186
Smoking status	0.0124	0.2497	-0.0612	1.0000	0.1239	-0.1358	0.1740	0.0693
Pack years	0.0629	0.0752	-0.0895	0.1239	1.0000	-0.0902	0.0742	0.0719
Baseline FEV1	-0.2997	-0.2098	-0.4964	-0.1358	-0.0902	1.0000	-0.5970	-0.3535
COPD status	0.1996	0.0611	-0.0804	0.1740	0.0742	-0.5970	1.0000	0.4656
Follow up time	0.0859	0.1118	0.0186	0.0693	0.0719	-0.3535	0.4656	1.000

Supplemental Table 2B: The p-values from a Pearson's correlation test between each variable.

	ΔFEV1	age	sex	Smoking status	Pack Years	Baseline FEV1	COPD status	Follow up time
ΔFEV1	NA	0.1688	0.3576	0.8873	0.4843	0.0004	0.0208	0.3234
age	0.1688	NA	0.7033	0.0036	0.4026	0.0150	0.4834	0.1986
sex	0.3576	0.7033	NA	0.4821	0.3189	0.0000	0.3560	0.8310
Smoking status	0.8873	0.0036	0.4821	NA	0.1668	0.1178	0.0444	0.4262
Pack years	0.4843	0.4026	0.3189	0.1668	NA	0.3151	0.4090	0.4234
Baseline FEV1	0.0004	0.0150	<0.0001	0.1178	0.3151	NA	<0.0001	<0.0001
COPD status	0.0208	0.4834	0.3560	0.0444	0.4090	<0.0001	NA	<0.0001
Follow up time	0.3234	0.1986	0.8310	0.4262	0.4234	<0.0001	<0.0001	NA

Supplemental Table 3: Range of the number of days between the baseline spirometry measurement and bronchoscopy. The median number of days was 63.5 days before bronchoscopy.

Days between spirometry and bronchoscopy	Number of people (% of total) N=134
0-90 days	94 (70.14%)

91-180 days	31 (15.67%)
181-270 days	5 (3.73 %)
271-365 days	4 (2.99%)

Supplemental Table 4: Genes associated with lung function decline (FDR <0.05).

Cluster	Gene Name	Coefficient	t-value	p-value	FDR	Overlap with Steiling *et al* COPD signature [2]
cluster 1	TCN1	-10.464	-5.8369	4.70E-08	0.0009	
cluster 1	AHCYL2	-2.4444	-5.50634	2.14E-07	0.0021	
cluster 1	MFSD4	-2.8478	-5.28954	5.66E-07	0.003	
cluster 1	GALNT4	-2.2838	-5.26957	6.18E-07	0.003	
cluster 1	PLA2G4A	-4.0452	-5.18292	9.04E-07	0.0036	
cluster 1	TM9SF3	-1.6493	-5.12016	1.19E-06	0.0039	
cluster 1	TSPAN13	-2.8593	-5.0864	1.38E-06	0.0039	yes
cluster 1	GALNT5	-4.763	-4.98636	2.12E-06	0.0052	
cluster 1	GALNT7	-3.2562	-4.94712	2.50E-06	0.0055	
cluster 1	PARM1	-4.2185	-4.86524	3.54E-06	0.007	
cluster 1	SURF4	-2.0655	-4.78617	4.93E-06	0.0088	
cluster 1	GALNT12	-2.6976	-4.74198	5.93E-06	0.0091	
cluster 2	KIF13A	1.4077	4.73841	6.02E-06	0.0091	
cluster 1	ARMCX3	-2.5391	-4.67984	7.66E-06	0.0108	
cluster 1	RDH10	-3.4605	-4.56512	1.22E-05	0.0157	
cluster 1	CEACAM5	-10.39	-4.51293	1.51E-05	0.0157	yes
cluster 2	CCDC69	1.9568	4.52184	1.46E-05	0.0157	
cluster 2	PRKCE	1.918	4.53047	1.41E-05	0.0157	
cluster 2	TBC1D22B	1.288	4.53979	1.36E-05	0.0157	
cluster 2	BBS1	1.4065	4.49646	1.62E-05	0.0159	
cluster 1	GNPNAT1	-2.6483	-4.4101	2.28E-05	0.019	
cluster 1	TSPAN8	-3.189	-4.40552	2.32E-05	0.019	
cluster 1	MAGT1	-2.3007	-4.39711	2.40E-05	0.019	
cluster 1	PDXDC1	-1.3108	-4.39564	2.41E-05	0.019	
cluster 1	LRRC8A	-1.9195	-4.36609	2.71E-05	0.019	
cluster 1	FUT6	-2.2112	-4.36209	2.76E-05	0.019	
cluster 1	ENTPD4	-1.4813	-4.35809	2.80E-05	0.019	
cluster 1	SLC26A2	-4.1649	-4.3533	2.85E-05	0.019	
cluster 1	S100A16	-3.7393	-4.35288	2.86E-05	0.019	
cluster 1	SLC39A8	-3.4425	-4.33926	3.02E-05	0.019	

cluster 1	TMEM165	-1.5627	-4.32997	3.13E-05	0.019	
cluster 1	CTSC	-2.622	-4.32131	3.24E-05	0.019	
cluster 1	ASRGL1	-3.9515	-4.31406	3.33E-05	0.019	
cluster 1	PTHLH	-1.9531	-4.31073	3.37E-05	0.019	
cluster 2	FAM53B	1.5573	4.31818	3.28E-05	0.019	
cluster 1	ATP13A5	-5.9505	-4.29322	3.61E-05	0.0198	
cluster 1	ENTPD3	-2.7279	-4.2653	4.03E-05	0.0209	
cluster 1	CLDN10	-4.2711	-4.26487	4.03E-05	0.0209	
cluster 1	FER1L6	-3.2417	-4.23308	4.56E-05	0.0222	
cluster 1	PPAPDC1B	-1.3953	-4.23007	4.61E-05	0.0222	
cluster 2	LINC00341	2.3221	4.23778	4.48E-05	0.0222	
cluster 1	AP2B1	-2.6973	-4.2219	4.76E-05	0.0224	
cluster 1	S100A14	-3.8189	-4.20974	4.99E-05	0.0224	
cluster 2	MAML2	1.5244	4.20987	4.99E-05	0.0224	
cluster 1	SCEL	-4.9864	-4.18833	5.42E-05	0.023	
cluster 2	SRGAP2	1.7241	4.17927	5.61E-05	0.023	
cluster 2	CCDC170	1.7573	4.17981	5.60E-05	0.023	
cluster 2	CNNM2	1.4465	4.18472	5.49E-05	0.023	
cluster 1	SEC31A	-1.0296	-4.16213	5.99E-05	0.0237	
cluster 1	GALE	-2.4574	-4.15894	6.06E-05	0.0237	
cluster 1	AZGP1	-4.1775	-4.1514	6.24E-05	0.0237	
cluster 1	NNT	-1.9732	-4.14685	6.35E-05	0.0237	
cluster 2	WWC1	1.4476	4.14138	6.48E-05	0.0237	
cluster 2	CD38	1.826	4.14279	6.45E-05	0.0237	
cluster 2	HIST3H2BB	2.0805	4.12968	6.78E-05	0.0243	
cluster 2	HEY2	2.4324	4.11715	7.11E-05	0.0246	
cluster 2	KCNB1	2.2671	4.12018	7.03E-05	0.0246	
cluster 1	ARMCX6	-1.5124	-4.11248	7.24E-05	0.0246	
cluster 1	UPK1B	-7.24	-4.10726	7.38E-05	0.0247	
cluster 1	PRSS23	-2.5324	-4.08732	7.96E-05	0.0257	
cluster 1	MORC4	-1.3045	-4.08122	8.14E-05	0.0257	
cluster 2	C11orf63	1.8924	4.07885	8.22E-05	0.0257	
cluster 2	PCNT	1.0629	4.09117	7.84E-05	0.0257	
cluster 1	EIF2AK3	-1.6608	-4.06547	8.64E-05	0.0266	
cluster 1	TXNDC11	-1.1497	-4.06009	8.82E-05	0.0268	
cluster 2	GAB2	1.6751	4.05317	9.05E-05	0.027	
cluster 1	SLC31A1	-2.1688	-4.04251	9.42E-05	0.0277	
cluster 2	KIF24	1.5694	4.03829	9.57E-05	0.0278	

cluster 2	ZNF709	1.9479	4.02668	1.00E-04	0.0286	
cluster 1	S100P	-4.1727	-4.00659	0.000108	0.0299	
cluster 2	CYP27A1	3.2116	4.00764	0.000107	0.0299	
cluster 1	EPT1	-2.0385	-3.99705	0.000112	0.03	
cluster 1	FAM177B	-7.5999	-3.99366	0.000113	0.03	yes
cluster 1	WDR72	-3.3747	-3.98683	0.000116	0.03	
cluster 1	CREB3L1	-3.1	-3.98494	0.000117	0.03	
cluster 1	HMGCS2	-5.4842	-3.97451	0.000121	0.03	
cluster 1	MUC2	-3.8157	-3.97049	0.000123	0.03	
cluster 1	FZD5	-1.7413	-3.96905	0.000124	0.03	
cluster 1	TMPRSS4	-2.7115	-3.96629	0.000125	0.03	
cluster 1	MIA3	-1.8609	-3.96036	0.000128	0.03	
cluster 1	SLC44A3	-1.7244	-3.95725	0.000129	0.03	yes
cluster 2	ZNF382	4.2946	3.95381	0.000131	0.03	
cluster 2	ZNF473	1.1667	3.95404	0.000131	0.03	
cluster 2	BRF1	1.0322	3.95619	0.00013	0.03	
cluster 2	FHAD1	1.4909	3.96247	0.000127	0.03	
cluster 2	SCAI	1.4052	3.97566	0.000121	0.03	
cluster 2	ZNF544	1.1378	3.94993	0.000133	0.0301	
cluster 1	MTHFD2	-3.043	-3.9405	0.000138	0.0304	yes
cluster 1	PRRC1	-1.2498	-3.93466	0.000141	0.0304	
cluster 1	VTCN1	-3.2947	-3.93389	0.000141	0.0304	
cluster 1	STK38L	-1.5853	-3.93262	0.000142	0.0304	
cluster 2	SNRK	1.2327	3.94403	0.000136	0.0304	
cluster 1	DISP1	-2.0634	-3.92329	0.000147	0.0307	
cluster 1	KDELR2	-1.815	-3.9206	0.000148	0.0307	yes
cluster 2	CEP250	1.0346	3.92537	0.000146	0.0307	
cluster 2	USP2	1.3642	3.91639	0.00015	0.0309	
cluster 1	SEC24A	-1.3454	-3.89593	0.000162	0.033	
cluster 1	KCNK6	-1.6357	-3.87926	0.000172	0.0343	
cluster 2	STXBP1	1.4889	3.88067	0.000171	0.0343	
cluster 1	RPN2	-1.3912	-3.87084	0.000178	0.0345	
cluster 1	ATP6V0E1	-1.7274	-3.86919	0.000179	0.0345	
cluster 1	MYO1C	-1.7383	-3.86711	0.00018	0.0345	
cluster 1	DGKA	-3.105	-3.86674	0.00018	0.0345	
cluster 1	ACBD3	-0.8868	-3.85338	0.000189	0.0358	
cluster 1	FUT2	-3.26	-3.85115	0.000191	0.0358	
cluster 1	B4GALT4	-2.693	-3.84516	0.000195	0.0358	

cluster 1	CYP2C18	-3.3432	-3.83689	0.000201	0.0358	
cluster 1	VIPR1	-1.7688	-3.83287	0.000204	0.0358	
cluster 1	ANO10	-1.6109	-3.83196	0.000205	0.0358	
cluster 1	PRSS8	-2.5934	-3.83105	0.000205	0.0358	
cluster 1	MFSD1	-1.5788	-3.83078	0.000205	0.0358	
cluster 2	ENO4	1.6146	3.8362	0.000201	0.0358	
cluster 2	RGAG4	1.6659	3.84485	0.000195	0.0358	
cluster 1	SLC1A5	-2.3667	-3.82351	0.000211	0.036	
cluster 1	TSPAN5	-3.0642	-3.82344	0.000211	0.036	
cluster 1	GNE	-1.9163	-3.82084	0.000213	0.036	
cluster 1	TMEM39A	-1.2691	-3.82034	0.000213	0.036	
cluster 2	STPG1	1.4569	3.81535	0.000217	0.0363	
cluster 1	TMEM167A	-1.9301	-3.80489	0.000226	0.0369	
cluster 1	IDH1	-2.2404	-3.80073	0.000229	0.0369	
cluster 1	C12orf23	-1.8034	-3.79807	0.000231	0.0369	
cluster 1	FUT3	-2.9334	-3.79734	0.000232	0.0369	
cluster 1	LOC100128816	-2.4998	-3.797	0.000232	0.0369	yes
cluster 2	IL5RA	1.7962	3.79672	0.000232	0.0369	
cluster 1	AP4B1	-1.1531	-3.7931	0.000235	0.0371	
cluster 1	PDK1	-1.6119	-3.78851	0.000239	0.0374	
cluster 1	SPTSSA	-2.0225	-3.78291	0.000244	0.0379	
cluster 1	SMPDL3A	-2.6578	-3.77223	0.000254	0.0391	
cluster 2	ULK2	1.2915	3.7693	0.000256	0.0392	
cluster 2	ZBTB44	1.3698	3.76431	0.000261	0.0396	
cluster 1	ADAM9	-1.9957	-3.7589	0.000266	0.04	
cluster 1	SLC16A9	-2.6703	-3.75655	0.000268	0.0401	yes
cluster 1	SEC61A1	-1.5947	-3.75383	0.000271	0.0402	yes
cluster 2	NSUN7	1.821	3.75086	0.000274	0.0403	
cluster 1	ANP32E	-2.2045	-3.74436	0.00028	0.0409	
cluster 2	KCNJ2	2.0652	3.7407	0.000284	0.0411	
cluster 1	TMEM211	-3.1984	-3.73378	0.000291	0.0414	
cluster 1	CEACAM6	-2.9929	-3.73256	0.000292	0.0414	
cluster 1	OPN1LW	-2.3305	-3.73252	0.000292	0.0414	
cluster 2	NEK4	1.0945	3.72762	0.000297	0.0419	
cluster 1	SLC12A8	-1.6888	-3.71954	0.000306	0.0426	
cluster 1	FGFBP1	-4.7744	-3.71881	0.000307	0.0426	
cluster 2	CEP104	0.7202	3.71602	0.00031	0.0427	
cluster 1	SERPINB8	-2.3042	-3.70542	0.000322	0.0428	

cluster 1	CTTNBP2	-2.5074	-3.70515	0.000322	0.0428	
cluster 2	ZFP82	2.0472	3.70405	0.000323	0.0428	
cluster 2	NEK11	1.4543	3.70571	0.000321	0.0428	
cluster 2	CCDC151	1.6148	3.71089	0.000315	0.0428	
cluster 2	LAMC2	1.6012	3.71366	0.000312	0.0428	
cluster 1	BCL2L15	-3.357	-3.70114	0.000327	0.0429	
cluster 1	ALG14	-1.4318	-3.69193	0.000337	0.044	
cluster 1	FKBP14	-2.2071	-3.68887	0.000341	0.0442	
cluster 2	ZNF391	2.4311	3.68616	0.000344	0.0444	
cluster 1	DHX15	-1.0895	-3.67935	0.000353	0.0449	
cluster 2	EZH1	1.1255	3.67881	0.000353	0.0449	
cluster 2	PCSK6	2.0747	3.67481	0.000358	0.0453	
cluster 1	MOB4	-2.6686	-3.67044	0.000364	0.0454	
cluster 1	TUBA1C	-2.637	-3.6702	0.000364	0.0454	
cluster 1	PGBD2	-1.7254	-3.66542	0.00037	0.0459	
cluster 1	KCNK1	-1.4739	-3.65087	0.00039	0.048	
cluster 2	PRSS12	1.5289	3.64465	0.000398	0.0488	
cluster 1	MAL2	-1.2395	-3.6406	0.000404	0.0492	
cluster 1	PDIA5	-1.8686	-3.63485	0.000412	0.0498	
cluster 1	TIMP1	-2.9324	-3.62684	0.000424	0.0499	
cluster 1	GPT2	-2.4736	-3.62484	0.000427	0.0499	
cluster 1	TMED3	-1.8777	-3.62363	0.000428	0.0499	
cluster 1	NIPA2	-1.0976	-3.62323	0.000429	0.0499	
cluster 1	TPBG	-1.3092	-3.62306	0.000429	0.0499	
cluster 1	CANT1	-1.598	-3.62093	0.000433	0.0499	
cluster 2	FXYD6	1.9572	3.62206	0.000431	0.0499	
cluster 2	PYGO1	3.2936	3.62903	0.00042	0.0499	

Supplemental Table 5: Demographics of the GLUCOLD cohort. The mean and standard deviation are shown for continuous variables.

	N=46		
	Mean	**Standard deviation**	**Range**
Age (years)	61.11	+/-7.97	(46-74)
Baseline FEV$_1$ (L)	2.09	+/- 0.49	(1.33-3.28)
ΔFEV$_1$ (mL/year)	-60.53	+/- 44.29	(-211.34 – 57.44)
Follow up time (years)	6.76	+/- 1.32	(3.5-7.5)
	N	**Percent**	

Smoking Status= Current	26	65.00%	
Sex = Male	40	86.96%	

Supplemental Table 6A: Transcription Factor binding sites enriched in the genes that increase with more severe lung function decline (cluster 1)

Gene Set Name	# Genes in Gene Set (K)	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
TTGTTT_FOXO4_01	2061	Genes having at least one occurence of the highly conserved motif M60 TTGTTT sites. The motif matches transcription factor binding site V$FOXO4_01 (v7.4 TRANSFAC).	19	0.0092	8.78E-07	3.68E-04
CAGGTG_E12_Q6	2485	Genes having at least one occurence of the highly conserved motif M12 CAGGTG sites. The motif matches transcription factor binding site V$E12_Q6 (v7.4 TRANSFAC).	21	0.0085	8.80E-07	3.68E-04
HTF_01	72	Genes having at least one occurence of the transcription factor binding site V$HTF_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0556	3.36E-05	8.37E-03
GGGCGGR_SP1_Q6	2940	Genes having at least one occurence of the highly conserved motif M6 GGGCGGR sites. The motif matches transcription factor binding site V$SP1_Q6 (v7.4 TRANSFAC).	20	0.0068	4.00E-05	8.37E-03
TGANNYRGCA_TCF11MAFG_01	301	Genes having at least one occurence of the highly conserved motif M67 TGANNYRGCA sites. The motif matches transcription factor binding site V$TCF11MAFG_01 (v7.4 TRANSFAC).	6	0.0199	1.16E-04	1.47E-02
CTTTGT_LEF1_Q2	1972	Genes having at least one occurence of the highly conserved motif M13 CTTTGT sites. The motif matches transcription factor binding site V$LEF1_Q2 (v7.4 TRANSFAC).	15	0.0076	1.23E-04	1.47E-02

CTTTAAR_UNKNOWN	972	Genes having at least one occurence of the highly conserved motif M29 CTTTAAR in the region spanning up to 4 kb around their transcription start sites. The motif does not match any known transcription factor binding site (v7.4 TRANSFAC).	10	0.0103	1.73E-04	1.80E-02
ATF6_01	123	Genes having at least one occurence of the transcription factor binding site V$ATF6_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0325	2.69E-04	2.49E-02
XBP1_01	133	Genes having at least one occurence of the transcription factor binding site V$XBP1_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0301	3.61E-04	3.02E-02
GGGAGGRR_MAZ_Q6	2274	Genes having at least one occurence of the highly conserved motif M24 GGGAGGRR sites. The motif matches transcription factor binding site V$MAZ_Q6 (v7.4 TRANSFAC).	15	0.0066	5.61E-04	4.26E-02

Supplemental Table 6B: Transcription Factor binding sites enriched in the genes that decrease with more severe lung function decline (cluster 2)

Gene Set Name	# Genes in Gene Set (K)	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
CAGGTG_E12_Q6	2485	Genes having at least one occurence of the highly conserved motif M12 CAGGTG sites. The motif matches transcription factor binding site V$E12_Q6 (v7.4 TRANSFAC).	14	0.0056	1.49E-07	1.24E-04
GGGAGGRR_MAZ_Q6	2274	Genes having at least one occurence of the highly conserved motif M24 GGGAGGRR sites. The	12	0.0053	2.76E-06	7.69E-04

		motif matches transcription factor binding site V$MAZ_Q6 (v7.4 TRANSFAC).				
TEF1_Q6	226	Genes having at least one occurence of the transcription factor binding site V$TEF1_Q6 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	5	0.0221	3.97E-06	8.29E-04
TTGTTT_FOXO4_01	2061	Genes having at least one occurence of the highly conserved motif M60 TTGTTT sites. The motif matches transcription factor binding site V$FOXO4_01 (v7.4 TRANSFAC).	11	0.0053	6.99E-06	1.17E-03
SRY_01	224	Genes having at least one occurence of the transcription factor binding site V$SRY_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0179	9.04E-05	1.22E-02
E2A_Q2	243	Genes having at least one occurence of the transcription factor binding site V$E2A_Q2 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0165	1.24E-04	1.22E-02
S8_01	245	Genes having at least one occurence of the transcription factor binding site V$S8_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0163	1.28E-04	1.22E-02
ZIC2_01	247	Genes having at least one occurence of the transcription factor binding site V$ZIC2_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0162	1.32E-04	1.22E-02

E12_Q6	262	Genes having at least one occurence of the transcription factor binding site V$E12_Q6 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0153	1.65E-04	1.29E-02
LMO2COM_01	264	Genes having at least one occurence of the transcription factor binding site V$LMO2COM_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0152	1.70E-04	1.29E-02
AP1_Q2_01	275	Genes having at least one occurence of the transcription factor binding site V$AP1_Q2_01 (v7.4 TRANSFAC) in the regions spanning up to 4 kb around their transcription starting sites.	4	0.0145	1.98E-04	1.38E-02
TGACAGNY_MEIS1_01	827	Genes having at least one occurence of the highly conserved motif M41 TGACAGNY sites. The motif matches transcription factor binding site V$MEIS1_01 (v7.4 TRANSFAC).	6	0.0073	2.15E-04	1.38E-02
GGGYGTGNY_UNKNOWN	664	Genes having at least one occurence of the highly conserved motif M31 GGGYGTGNY in the region spanning up to 4 kb around their transcription start sites. The motif does not match any known transcription factor binding site (v7.4 TRANSFAC).	5	0.0075	6.36E-04	3.32E-02
TGCCAAR_NF1_Q6	722	Genes having at least one occurence of the highly conserved motif M47 TGCCAAR sites. The motif matches transcription factor binding site V$NF1_Q6 (v7.4 TRANSFAC).	5	0.0069	9.25E-04	4.30E-02

Supplemental Table 7: List of the leading edge genes contributing the most to the enrichment of mucin gene sets in genes ranked by association with FEV1 decline.

Gene Set	Leading edge genes
KEGG 2019 Mucin type O-glycan biosynthesis	GALNT2, GALNT3, GCNT1, GALNT14, ST6GALNAC1, GALNT1, GALNT4, GALNT3, B3GNT6, GALNT6, GALNT5, GALNT7, GALNT12, BGALT5
Jensen components Mucin granule	SLC17A9, P2RY2, TFF3, CLCA1, MUC2, MUC16, MYO5C, TFF1, UNC13B, AGR2, MUC1, and CEACAM5

Citations

1 Tammemagi MC, Lam SC, McWilliams AM, *et al.* Incremental Value of Pulmonary Function and Sputum DNA Image Cytometry in Lung Cancer Risk Prediction. *Cancer Prev Res (Phila Pa)* 2011;**4**:552–61. doi:10.1158/1940-6207.CAPR-10-0183

2 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**:933–42. doi:10.1164/rccm.201208-1449OC