Resource

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

THE EMBO JOURNAL

High-resolution transcriptomic and epigenetic profiling identifies novel regulators of COPD

Uwe Schwartz[1,2,†], Maria Llamazares Prada[1,3,4,†], Stephanie T Pohl[1,5], Mandy Richter[1], Raluca Tamas[1], Michael Schuler[6], Corinna Keller[7], Vedrana Mijosek[1], Thomas Muley[4,8], Marc A Schneider[4,8], Karsten Quast[9], Joschka Hey[3,10], Claus P Heußel[4,11,12], Arne Warth[4,8,13,‡], Hauke Winter[8,14], Özdemirhan Serçin[1], Harry Karmouty-Quintana[15], Soma SK Jyothula[16], Manish K Patel[16], Felix Herth[4,8,17], Ina Koch[18], Giuseppe Petrosino[1,§], Alexandru Titimeaua[5], Balca R Mardin[1], Dieter Weichenhan[3], Tomasz P Jurkowski[19], Charles D Imbusch[20], Benedikt Brors[20], Vladimir Benes[21], Birgit Jung[7], David Wyatt[22], Heiko F Stahl[7], Christoph Plass[3,4] & Renata Z Jurkowska[1,5,*]

Abstract

Patients with chronic obstructive pulmonary disease (COPD) are still waiting for curative treatments. Considering its environmental cause, we hypothesized that COPD will be associated with altered epigenetic signaling in lung cells. We generated genome-wide DNA methylation maps at single CpG resolution of primary human lung fibroblasts (HLFs) across COPD stages. We show that the epigenetic landscape is changed early in COPD, with DNA methylation changes occurring predominantly in regulatory regions. RNA sequencing of matched fibroblasts demonstrated dysregulation of genes involved in proliferation, DNA repair, and extracellular matrix organization. Data integration identified 110 candidate regulators of disease phenotypes that were linked to fibroblast repair processes using phenotypic screens. Our study provides high-resolution multi-omic maps of HLFs across COPD stages. We reveal novel transcriptomic and epigenetic signatures associated with COPD onset and progression and identify new candidate regulators involved in the pathogenesis of chronic lung diseases. The presence of various epigenetic factors among the candidates demonstrates that epigenetic regulation in COPD is an exciting research field that holds promise for novel therapeutic avenues for patients.

Keywords COPD; DNA methylation; human lung fibroblasts; RNA sequencing; WGBS

1 BioMed X Institute, Heidelberg, Germany
2 NGS Analysis Center Biology and Pre-Clinical Medicine, University of Regensburg, Regensburg, Germany
3 Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany
4 Translational Lung Research Center, Member of the German Center for Lung Research (DZL), Heidelberg, Germany
5 Division of Biomedicine, School of Biosciences, Cardiff University, Cardiff, UK
6 Drug Discovery Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
7 Immunology and Respiratory Disease Research, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
8 Translational Research Unit, Heidelberg Lung Biobank, Thoraxklinik,, University Hospital Heidelberg, Heidelberg, Germany
9 Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
10 Ruprecht Karl University of Heidelberg, Heidelberg, Germany
11 Diagnostic and Interventional Radiology with Nuclear Medicine, Thoraxklinik, University of Heidelberg, Heidelberg, Germany
12 Diagnostic and Interventional Radiology, University Hospital Heidelberg, Heidelberg, Germany
13 Pathological Institute, University Hospital Heidelberg, Heidelberg, Germany
14 Department of Surgery, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany
15 Department of Biochemistry and Molecular Biology, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA
16 Center for Advanced Cardiopulmonary Therapies and Transplantation, University of Texas Health Science Center at Houston, Houston, TX, USA
17 Department of Pneumology and Critical Care Medicine and Translational Research Unit, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany
18 Asklepios Biobank for Lung Diseases, Department of Thoracic Surgery, Asklepios Fachkliniken München-Gauting, German Center for Lung Research (DZL), Munich, Germany
19 Division of Molecular Biology, School of Biosciences, Cardiff, UK
20 Division of Applied Bioinformatics, German Cancer Research Center, Heidelberg, Germany
21 Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
22 Biotherapeutics Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
*Corresponding author. Tel: +44 29 2087 9067; E-mail: jurkowskar@cardiff.ac.uk
†These authors contributed equally to this work
‡Present address: Institute of Pathology, Cytopathology and Molecular Pathology MVZ UEGP Gießen/Wetzlar/Limburg/Bad Hersfeld, Wetzlar, Germany
§Present address: Institute of Molecular Biology (IMB), Mainz, Germany

Subject Categories Cancer; Chromatin, Transcription & Genomics; Methods & Resources
DOI 10.15252/embj.2022111272 | Received 26 March 2022 | Revised 14 March 2023 | Accepted 27 March 2023
The EMBO Journal (2023) e111272

Introduction

Chronic obstructive pulmonary disease (COPD) is a prevalent, smoke-related disease characterized by persistent inflammation of the lung epithelium, irreversible airway remodeling, and destruction of the alveolar tissue (emphysema) (Barnes *et al*, 2015; Rabe & Watz, 2017; GOLD, 2021). COPD-related mortality is increasing, and it already affects more than 3 million people worldwide every year (Safiri *et al*, 2022). However, despite its prevalence, there is currently no treatment to halt the progression of COPD, as none of the existing drugs can modify the long-term decline in lung function. COPD is a heterogeneous disease with variable clinical manifestations and responses to therapy where patient stratification remains challenging (Woodruff *et al*, 2015; Agusti *et al*, 2017; Garudadri & Woodruff, 2018; Barnes, 2019a).

Fibroblasts are ubiquitous mesenchymal cells found in the parenchyma and the outer layer of airways and vessels in the adult lung (Phan, 2008). They have essential functions in lung homeostasis, maintenance of stem cells, wound healing, and tissue repair. In COPD, airway fibroblasts are the key cells contributing to the excessive deposition of extracellular matrix, small-airway fibrosis, and airway remodeling (Barnes, 2019b). In turn, parenchymal fibroblasts from patients with COPD/emphysema show reduced proliferation (Nobukuni *et al*, 2002; Holz *et al*, 2004), contractility and migration *in vitro* (Togo *et al*, 2008), are senescent (Muller *et al*, 2006), display altered growth factor response (Noordhoek *et al*, 2003; Togo *et al*, 2008), and express increased levels of pro-inflammatory cytokines (Zhang *et al*, 2012), indicative of a reduced tissue-repair capacity. The altered function of alveolar fibroblasts also contributes to epithelial progenitor dysfunction, establishing fibroblasts as a critical cell type contributing to the development of emphysema (Plantier *et al*, 2007; Kulkarni *et al*, 2016). However, it remains unknown how these phenotypic changes in parenchymal fibroblasts are encoded at the molecular level.

Numerous genetic loci have been associated with COPD and lung function (Wilk *et al*, 2009; Hancock *et al*, 2010; Soler Artigas *et al*, 2011; Cho *et al*, 2014; Wain *et al*, 2015; Hobbs *et al*, 2017; Wyss *et al*, 2018; Sakornsakolpat *et al*, 2019), yet they explain only a small fraction of COPD risk. Transcriptional programs in cells are regulated by a landscape of epigenetic modifications that modulate chromatin structure and thereby control gene expression. Smoking is the most prominent risk factor for COPD, and its impact on epigenetic landscape remodeling is well established (Belinsky *et al*, 2002; Zeilinger *et al*, 2013; Wan *et al*, 2015). Earlier studies also provided strong evidence for the association of dysregulated DNA methylation and COPD in the blood (Qiu *et al*, 2012; Busch *et al*, 2016; Carmona *et al*, 2018), sputum (Sood *et al*, 2010), oral mucosa (Wan *et al*, 2015), lung tissue (Sood *et al*, 2010; Yoo *et al*, 2015; Morrow *et al*, 2016; Sundar *et al*, 2017), bronchial brushings (Vucic *et al*, 2014), fibroblasts (Clifford *et al*, 2018), and macrophages from a mouse model of muco-obstructive disease (Hey *et al*, 2021). Notably, DNA methylation changes were associated with altered expression of genes and pathways important to COPD pathology. However, these studies were either performed on material encompassing mixed cell populations or/and used low-resolution approaches and could therefore not resolve differential gene expression and methylation changes caused by a specific cell type during COPD development and progression. To date, the full epigenomic landscape of purified COPD cells remains uncharted, and thus, the precise epigenetic changes and their contribution to altered transcriptional patterns in COPD are still unknown.

To identify the epigenetic and functional alterations associated with COPD in parenchymal fibroblasts, we used tagmentation-based whole-genome bisulfite sequencing (T-WGBS) to profile DNA methylation and RNA sequencing to measure gene expression changes in primary fibroblasts from patients with COPD and matched ex-smoker controls. Importantly, we hypothesized that epigenetic modifications would arise early during COPD development; thus, we analyzed cells from patients at different COPD stages. Our data provide integrative epigenetic and transcriptomic maps of fibroblasts at high resolution. It reveals pathways and novel candidate regulators, including epigenetic factors, that might be involved in the pathogenesis of chronic respiratory diseases.

Results

Genome-wide epigenetic changes occur early in primary lung fibroblasts during COPD

To assess the extent of epigenetic remodeling in COPD genome-wide, we generated high-resolution DNA methylomes of primary lung fibroblasts isolated from the lung parenchyma of well-matched control donors (no COPD, $n = 3$) and patients with established COPD (stage II–IV according to Global Initiative for Chronic Obstructive Lung Disease; GOLD, 2021, $n = 5$), which can be classified based on lung function (Figs 1A and EV1A and B, Dataset EV1). Isolated cells displayed a typical fibroblast morphology, and their purity was confirmed by fluorescence-activated cell sorting (FACS) and immunofluorescence (IF) staining (Fig EV1C and D). We used T-WGBS for DNA methylation profiling, allowing genome-scale assessment of DNA methylation at single CpG resolution from low cell numbers (Wang *et al*, 2013) (Fig 1B). 17.9 million CpG sites were covered at least 4× in all samples and used for further analysis, illustrating the high resolution of our generated data set (Fig EV1E).

We first looked at global changes in DNA methylation by dividing single CpG sites by their methylation level into lowly (< 20%), medium (20–80%), and highly (> 80%) methylated sites (Fig 1C). We observed no significant differences in the global levels of methylated cytosines between COPD and no COPD (control) samples, suggesting that, in contrast to cancer cells (Esteller, 2008), COPD is not associated with a global drop or gain of methylation in lung fibroblasts. We have not performed differential analysis on individual CpG sites, but for further analysis, we focused on differentially methylated regions (DMRs), which comprise at least three consecutively methylated CpG sites, between no COPD and COPD (II–IV). We decided to focus on larger regions rather than individual sites when looking at differential methylation as DNA methylation is spatially correlated and methylation changes in larger regions are more

likely to have a biological function. Using CpG sites covered at least 4× in all samples and a ≥ 10% methylation-difference cutoff, we identified 6,279 DMRs (P-value < 0.1, see Materials and Methods part for details; Fig 1D–F, Dataset EV2), indicating numerous methylation alterations in primary human lung fibroblasts of COPD patients. We define hypomethylation as the state showing lower DNA methylation levels in COPD samples as compared to no COPD samples, whereas hypermethylation describes higher DNA methylation levels in COPD samples. The distribution of methylation differences across DMRs demonstrated a more prominent loss of methylation, suggestive of a more permissive chromatin state in COPD (58% of DMRs, 3,615 hypomethylated regions, Fig 1D). The remaining 2,664 regions showed increased methylation (42% of the DMRs, 2,664 hypermethylated regions, Fig 1D). Identified DMRs contained eight CpG sites on average and showed a median size of 479 bp (Fig EV1F and G), indicating that specific larger regions are altered. An earlier study used Illumina Infinium HumanMethylation450 BeadChip array to determine DMRs in parenchymal fibroblasts in COPD (Clifford *et al*, 2018). 76% of our identified DMRs were not covered by any probe on the array illustrating the higher resolution of T-WGBS data. To compare the results of both studies, we selected DMRs from our study which were covered by at least three CpG probes on the array ($n = 136$) and correlated the average methylation differences in COPD (Fig EV1H). Despite the low coverage of CpGs overlapping with our DMRs, we observe a moderate positive correlation between both data sets ($R = 0.3$, P-value = 0.0005), confirming the robustness of our DMR calling strategy.

To investigate whether DNA methylation changes occur early in COPD development and identify alterations associated with disease progression, we integrated into the analysis T-WGBS data of fibroblasts from mild COPD patients (GOLD I, $n = 3$), with noticeable obstruction ($FEV_1/FVC < 70\%$) but preserved FEV_1 (> 80%, Fig EV1A, Dataset EV1). Notably, as demonstrated by the principal component analysis (PCA) on all 6,279 previously identified DMRs, COPD (I) samples grouped with the COPD (II-IV) samples on the first principal component (Fig 1G), confirming the COPD-specific DMR calling using independent test samples (as COPD (I) samples were not used for the initial DMR selection). Furthermore, the COPD (I) samples were separated from COPD (II–IV) on the second principal component, indicating that DNA methylation data might provide information about disease progression (Fig 1G). Consistent with the PCA, hierarchical clustering using the identified DMRs showed that COPD (I) samples group with COPD (II–IV) samples, demonstrating that COPD-associated methylation changes occur early in the disease pathogenesis (Fig EV1I). This important discovery suggests that DNA methylation might provide a sensitive biomarker to separate early COPD patients from smokers with preserved lung function but needs to be validated in a larger cohort.

Since COPD is a progressive lung disease, we wanted to gain more insights into the kinetics of DNA methylation changes between the three donor groups with different disease severity (no COPD, COPD (I), COPD (II–IV)). For this, we performed k-means clustering on the 6,279 identified DMRs using all samples. This analysis defined three main clusters displaying DNA methylation changes that progressed with increasing disease severity denoted by the decline of lung function of patients (decreasing FEV_1, Fig 1H). Clusters 1 and 3 showed loss of methylation at 1,951 DMRs and 1,665 DMRs, respectively, while cluster 2 (2,663 DMRs) displayed a progressive gain of methylation in COPD. Cluster 1 reveals regions with pronounced demethylation occurring already in COPD (I), while cluster 3 shows gradual loss of methylation as disease severity increases (Fig 1H and I).

To shed light on the cellular processes and pathways affected by aberrant DNA methylation changes in COPD, we linked DMRs to the nearest gene and performed gene ontology (GO) enrichment analysis using the Genomic Regions Enrichment of Annotations Tool (GREAT; McLean *et al*, 2010; Dataset EV3). All DMRs were included in the enrichment pathway analysis, but we looked separately for hypomethylated and hypermethylated DMRs. As hypermethylated DMRs did not reveal any significant enrichment with our GREAT settings (see Materials and Methods), we focused on the hypomethylated DMRs. Among the top categories, we identified cellular response to hypoxia and oxygen levels, regulation of focal adhesion assembly, negative regulation of transforming growth factor beta (TGFβ) receptor signaling, and epithelial-to-mesenchymal transition (Fig EV1J). These biological processes are relevant for COPD development and progression (Konigshoff *et al*, 2009; Barnes *et al*, 2015, 2019; Rabe & Watz, 2017), indicating that DNA methylation changes occur near genes that are critically involved in COPD pathogenesis and might therefore contribute to disease phenotypes in lung fibroblasts. Specific examples include hypomethylation of genes implicated in the negative regulation of the TGFβ receptor signaling (e.g., SMAD6, SMAD7, VASN, SKI, PMEPA1, Dataset EV2) (Miyazono, 2000), providing a potential mechanism for explaining the reduced response to TGFβ and decreased repair capacity of lung fibroblasts observed in emphysema (Togo *et al*, 2008). To further characterize processes that are seen early in the disease versus those that develop later, we extended the enrichment analysis by adding the proportions of the identified DMR cluster to the respective pathways (Fig EV1J, right panel). Interestingly, DMRs from cluster 1 showing early methylation changes (Fig 1H) were associated with the "regulation of chromatin organization (72.5%)" and "cell-substrate adhesion (70,3%)" (Fig EV1J, right panel), indicating that these cellular processes may be perturbed early in COPD development.

In summary, DNA methylation profiling of COPD samples across disease stages demonstrates that genome-wide epigenetic changes occur already early in COPD development and many of them progress with disease development.

DNA methylation changes occur at regulatory regions in COPD lung fibroblasts

Methylome analysis identified genome-wide changes in DNA methylation in primary HLFs of COPD patients. To better understand the functional role of aberrant methylation in COPD, we investigated the distribution of DMRs across the genome. We observed a different distribution of regions displaying loss and gain of methylation, with hypomethylated DMRs predominately located in intronic sequences and hypermethylated DMRs preferentially found in intergenic regions (Fig 2A). Notably, both types of DMRs were overrepresented at regulatory and gene coding sequences compared to the genomic background (Fig 2B), with a stronger enrichment observed for hypomethylated DMRs. Thus, hypomethylated DMRs are located four times more often at promoter sequences than expected by chance (Fig 2B). Further intersection with known regulatory

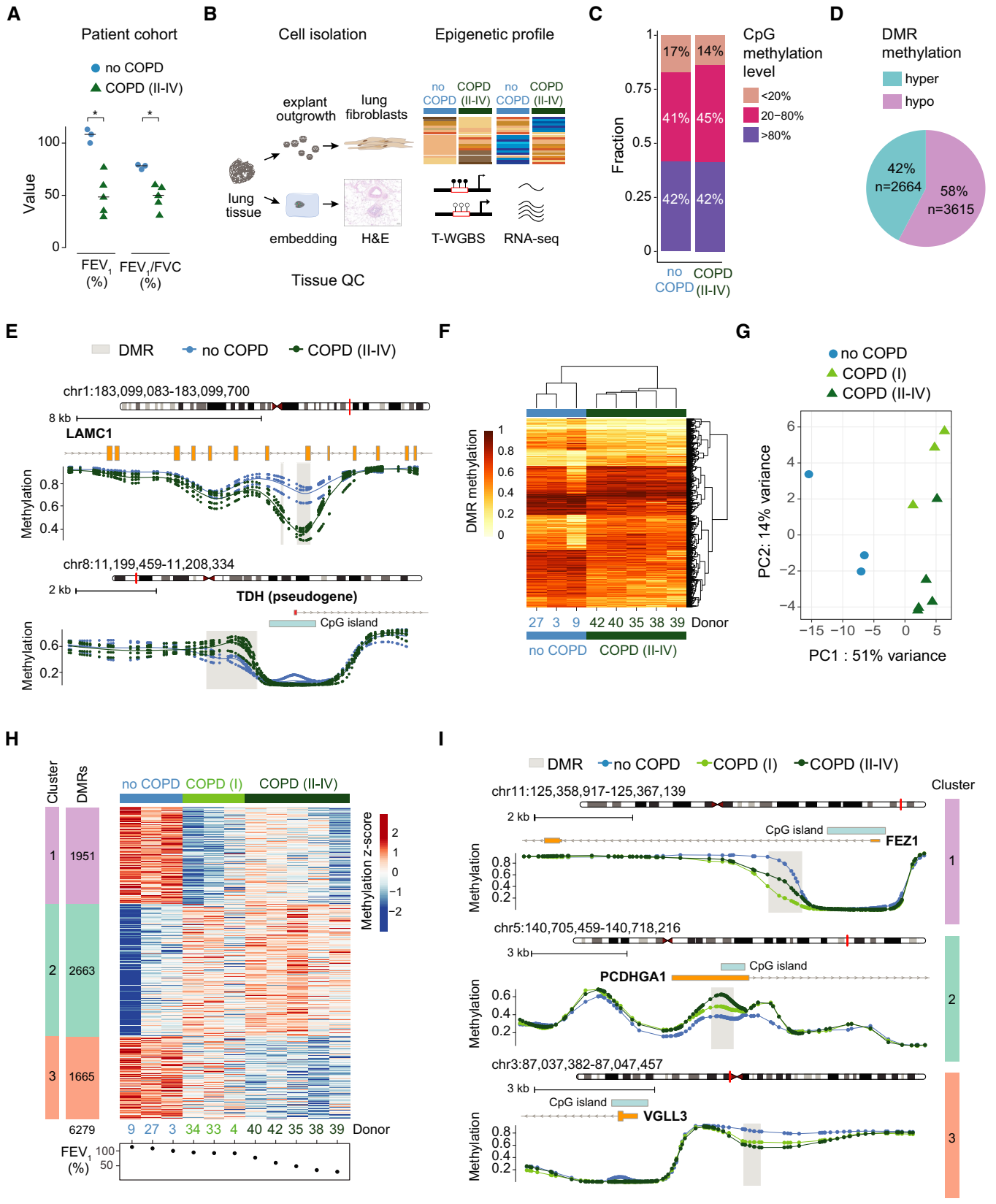

Figure 1.

Figure 1. Genome-wide DNA methylation changes occur early in human lung fibroblasts during COPD and progress with disease development.

A Lung function data of COPD (II-IV) and no COPD (ex-smoker controls) donors used in this study. The lung function between the two groups is significantly different.

B Schematic diagram illustrating the experimental approach used for epigenetic (T-WGBS) and transcriptomic (RNA-seq) profiling of purified primary parenchymal lung fibroblasts.

C–I T-WGBS data of primary fibroblasts from no COPD and COPD (II–IV) patients were analyzed at single CpGs level (C) and on DMRs (D–I).

C Genome wide CpG methylation statistics. Bar plot showing the fraction of high (> 80%), moderate (20–80%), and low (< 20%) methylated CpGs in no COPD and COPD (II–IV) samples.

D Number of hyper- or hypomethylated DMRs in COPD (II-IV).

E Detailed view of a representative hypo- (top) and hypermethylated (bottom) DMR (gray box). CpG methylation levels of each individual donor (dots) and the group average (lines) methylation profile of three no COPD (blue) and five COPD (II–IV) (dark green) donors are displayed. RefSeq annotated genes and CpG islands are indicated.

F Heatmap of 6,279 DMRs identified in COPD (II-IV). Statistically significant DMRs (at significance level = 0.1; see methods for DMR calling details) with at least three CpGs and a mean difference in methylation between no COPD and COPD (II–IV) of ≥ 10% were selected. Color shades indicate low (light) or high (dark) DMR methylation.

G PCA of COPD (II–IV) (dark green), no COPD (blue), and mild COPD (I) (light green, samples not used for initial DMR calling) on identified 6,279 DMRs.

H K-means clustering of all DMRs identified between no COPD and COPD (II–IV) across all samples, including COPD (I). Three clusters were identified. Cluster 1 shows early hypomethylation in COPD (I), clusters 2 and 3, gradual hyper- and hypomethylation, respectively. Donors are sorted according to their FEV_1 value as indicated at the bottom.

I Representative methylation profiles at selected DMRs from each cluster. Group median CpG methylation is shown for no COPD (blue), COPD (I) (light green), and COPD (II–IV) (dark green). RefSeq annotated genes and CpG islands are indicated.

Data information: In (A), data points represent each donor values and horizontal bars the group median. An unpaired non-parametric *t*-test (Mann–Whitney test, GraphPad Prism software, version 8.0.1) was employed to compare the lung function (FEV_1 and FEV_1/FVC values) between control (no COPD, $n = 3$) and COPD II-IV donors ($n = 5$), **P*-value < 0.05. Exact *P*-values are: FEV_1, *P*-value = 0.0357; FEV_1/FVC, *P*-value = 0.0357. FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; T-WGBS, tagmentation-based whole-genome bisulfite sequencing; DMR, differentially methylated regions; PCA, principal component analysis.

genomic features annotated by the ENCODE Chromatin States (Ernst *et al*, 2011) revealed strong enrichment of hypomethylated DMRs in active promoters and enhancers, indicating their potential regulatory role (Fig 2C and D, Dataset EV2). The significant association with active enhancer elements was confirmed by the local increase of enhancer-defining chromatin marks (H3K4me1 and H3K27ac) (Heintzman *et al*, 2009; Rada-Iglesias *et al*, 2010) in the center of the hypomethylated DMRs (Fig 2D and E). Conversely, the hypermethylated DMRs were overrepresented at Polycomb-repressed regions, defined by the presence of H3K27me3 (Fig 2C and E).

Since we detected an enrichment of hypomethylated DMRs residing in regions broadly marked by H3K4me1 and H3K27ac (Fig 2D and E), we used the intensity of the H3K4me1 and H3K27ac signals in the ENCODE ChIP-seq data (Davis *et al*, 2018) to classify super-enhancers (SE) in human lung fibroblasts. Next, we tested the overlap of the identified SE with the DMRs identified in COPD HLFs. About a quarter of all SE contained at least one DMR (Fig 2F, right panel, Fig EV2A, examples shown in Figs 2D and EV2B and C). Consistent with the chromatin state analysis, hypomethylated DMRs were preferentially associated with SE and coincided with the best scoring SE (Fig 2F, purple bar and labels), indicating that these SE may become differentially regulated in COPD. Finally, SE were assigned to nearby genes that they may regulate. SMAD3, GRK5, ERGIC1, CREB3L2, and RASA2 genes were associated with the most active super-enhancers overlapping with hypomethylated DMRs (Fig 2F, hockey plot, purple labels, and Fig EV2B and C).

We conclude that methylation changes identified in COPD HLFs, especially hypomethylation, occur at regulatory regions, including strong enhancers.

DMRs in COPD show enrichment of binding motifs for key lung transcription factors

The hypomethylated regions identified in WGBS data might reflect the binding of transcription factors and can be therefore used for foot-printing their binding sites (Stadler *et al*, 2011). We observed a significant enrichment of binding motifs of several transcription factors in the hypomethylated DMRs at strong enhancers, with the highest enrichment of TCF21 motif in the early DMR cluster (cluster 1, *P*-value: 1×10^{-17}) and FOSL2/FRA2 in the progressive DMR cluster (cluster 3, *P*-value: 1×10^{-8}) (Fig 2G). TCF21 mediates fibroblast fate specification in multiple organs and is required for alveolar development (Quaggin *et al*, 1999). In turn, FOSL2/FRA2 is a known regulator of wound repair and TGFβ-mediated fibrosis (Eferl *et al*, 2008). Our data indicate that TCF21 and FOSL2/FRA2 may be potential mediators of aberrant epigenetic changes at strong enhancers in COPD fibroblasts.

Since DNA methylation can also directly interfere with the binding of transcriptional regulators to DNA (Yin *et al*, 2017), we performed motif analysis in the identified hypo- and hypermethylated DMRs to identify transcription factors reported to change their binding affinity upon methylation of their motifs (Yin *et al*, 2017). At hypomethylated DMRs, which are overrepresented at regulatory sites, numerous methylation-sensitive transcription-factor motifs were significantly enriched, suggesting either increased (methyl-minus, red dots) or attenuated (methyl-plus, blue triangles) DNA binding in COPD (Fig 2H). Among the transcription factors exhibiting higher binding affinities toward methylated DNA (methyl-plus, blue triangles), we identified motifs of nuclear receptors, known regulators of cellular homeostasis, development, and metabolism (Fig 2H). Our data suggest that their DNA binding might be abrogated at regulatory sites in COPD due to loss of methylation, but this prediction needs to be experimentally validated. A few motifs of methylation-sensitive transcription factors were enriched in the hypermethylated DMRs, consistent with their location in repressive regions of the genome. The strongest enrichment was observed for ZBTB7A, a known repressor associated with the TGFβ signaling pathway (Shen *et al*, 2017). ZBTB7A preferentially binds unmethylated DNA (Yin *et al*, 2017) (methyl-minus, Fig 2H) indicating that its DNA binding in COPD might be hindered due to motif

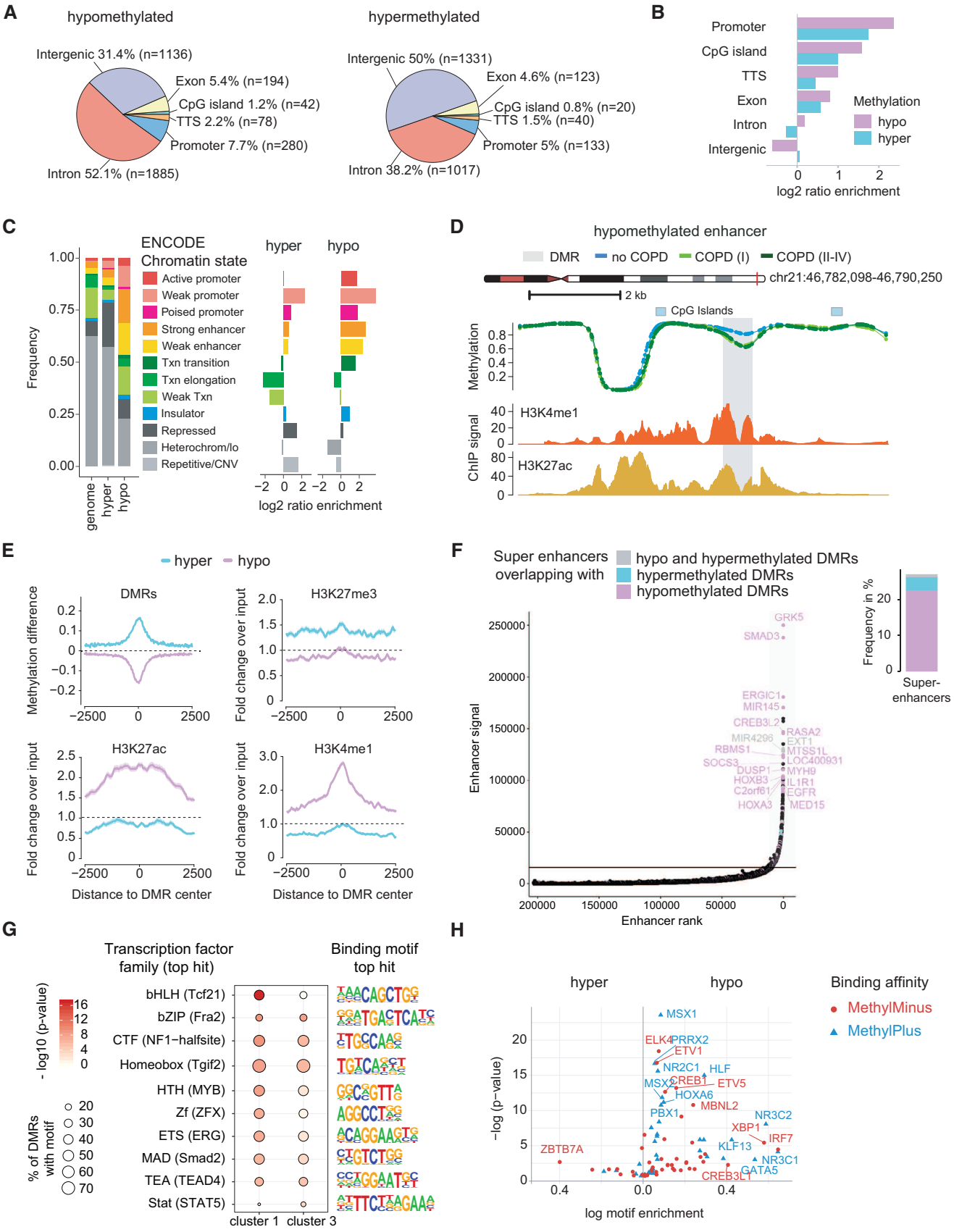

Figure 2.

Figure 2. DNA methylation changes occur at regulatory regions in primary human lung fibroblasts cells during COPD.

A, B Genomic location of identified DMRs. (A) Distribution of genomic features overlapping with hypo- (left) and hypermethylated (right) DMRs. (B) Enrichment of genomic features at hypo- (purple) and hypermethylated (cyan) DMRs compared to a sampled background of 10,000 regions exhibiting no significant change in methylation.

C Distribution of human lung fibroblast specific chromatin states (ENCODE accession: ENCFF001TDQ) at hypo- and hypermethylated DMRs. Fraction of DMRs overlapping with specific chromatin states is shown on the left panel. The genome background was sampled using 10,000 regions with matching GC content exhibiting no significant change in methylation. Chromatin state enrichment relative to the genome background is illustrated in the right panel.

D Genome browser view of an examplary DMR at a putative enhancer region. Group median CpG methylation is shown for no COPD (blue), COPD (I) (light green), and COPD (II–IV) (dark green). At the bottom the level of enhancer marks is depicted as fold-change over control: H3K4me1 (ENCODE accession: ENCFF102BGI) and H3K27ac (ENCODE accession: ENCFF386FDQ).

E Alterations of DNA methylation and selected histone marks around DMRs. Solid lines represent the mean profile and shaded lines the standard error of the mean across all summarized regions. Dashed lines indicate the expected signal value at sites, which are not changed, i.e. 0 for methylation difference in COPD (II–IV) to no COPD and 1 for the fold change to input in ChIP-seq experiments.

F Ranking of enhancer elements, defined by the co-occurrence of H3K4me1 and H3K27ac signals in human lung fibroblasts. The horizontal line defines the signal and corresponding rank threshold used to identify super enhancers (SE). Selected SE overlapping with DMRs are annotated and the nearest gene to the SE is indicated. The fraction of SEs overlapping with hypo- (purple), hypermethylated (cyan) or both (gray) is illustrated in the bar plot on the right panel.

G Transcription factor motifs most enriched at DMRs overlapping with strong enhancers (ENCODE chromatin states) from cluster 1 and 3 (see Fig 1H).

H Enrichment of methylation-sensitive transcription factor motifs at hypo- (right) and hypermethylated (left) DMRs. Methylation-sensitive motifs were derived from the study of Yin *et al* (2017). Transcription factors, whose binding affinity was impaired upon methylation of their corresponding DNA motif are shown in red (MethylMinus) and transcription factors, whose binding affinity was increased, in blue (MethylPlus). TTS, transcription termination site.

Data information: In G and H, Motif enrichment was calculated using HOMER, which uses ZOOPS scoring (zero or one occurrence per sequence) coupled with the hypergeometric enrichment calculations.

hypermethylation. To validate our motif enrichment, we used transcription factor footprints in active chromatin elements (TRACE) data from human lung fibroblasts available in ENCODE (Ouyang & Boyle, 2020) for TCF21, FOSL2, and ZBTB7A. Indeed, TCF21 and FOS2L binding was significantly enriched at strong enhancer DMRs and ZBTB7A at hypermethylated DMRs, respectively (P-value < 0.001, one-sided hypergeometric test).

In summary, our data suggest that aberrant DNA methylation in COPD fibroblasts may be linked to imbalanced transcription factor binding, providing insights into potentially disturbed regulatory networks in COPD.

Gene expression changes accompany epigenetic modifications in COPD

Our genome-wide DNA methylation analysis identified methylation changes at promoter and enhancer regions, suggesting that DMRs may have regulatory effects on gene expression. To assess whether epigenetic changes are associated with gene expression changes in COPD, we performed RNA-seq analysis on fibroblast samples matching those used for T-WGBS (Fig 1B). This analysis identified 333 up-regulated and 287 down-regulated genes between no COPD ($n = 3$) and COPD (II–IV) ($n = 5$; FDR < 0.05 and |log2(fold-change)| > 0.5, Fig 3A, Dataset EV4), including several long non-coding RNAs (Fig 3A orange labels, Fig EV3A), providing the transcriptional signature of COPD. Enrichment analysis of the differentially expressed genes (DEGs) revealed that genes up-regulated in COPD are involved in cell cycle regulation, DNA replication, and DNA repair (Fig EV3B). In turn, down-regulated genes are associated with extracellular matrix (ECM) organization and cholesterol biosynthesis (Fig EV3C). An earlier study in a larger cohort of samples ($n = 198$) used RNA seq to identify differentially expressed genes in lung tissue of smokers and COPD patients (Kim *et al*, 2015). Expression changes associated with COPD in whole tissue did not correlate with the changes we observed in isolated fibroblasts. In the GSEA analysis (log2FC sorted), we could not find a significant overlap of our differentially up-regulated genes. In contrast, although we observed a significant enrichment of our down-regulated genes in the Kim study, the expression change was reversed for most genes, and they were up-regulated (Fig EV3D). This may reflect the depletion of fibroblasts in emphysema, as the fibroblasts show reduced proliferation in COPD/emphysema (Nobukuni *et al*, 2002; Holz *et al*, 2004).

As COPD (I) samples ($n = 3$) were not used to identify differentially expressed genes, we used them as an independent test set to validate the obtained results and gain insights into gene expression kinetics in disease progression. Hierarchical clustering and PCA of all samples on the identified DEGs or 500 most variable genes, respectively, revealed that COPD (I) samples cluster together and show higher similarity to the no COPD group (Figs 3B and EV3E). To further resolve gene expression signatures of COPD states, we performed self-organizing map (SOM) clustering (Wehrens & Kruisselbrink, 2018) using all DEGs ($n = 620$). We identified six clusters showing different kinetics related to COPD progression (Fig 3C–E, Dataset EV4). Clusters 1 and 6 encompass genes whose expression is not yet changed in COPD (I) but gets dysregulated at later stages of COPD development (Fig 3C). Here, multiple genes involved in DNA replication (e.g., MCM10, ORC5, GINS3) or DNA double-strand break repair (e.g., BRCA1, FANCM, USP1, RAD51) are present (Fig 3E). Clusters 2, 3, 4, and 5 feature gene subsets already dysregulated in COPD (I) and may serve as early disease markers (Fig 3C, examples displayed in Fig 3D). Notably, the identification of early gene expression changes associated with COPD development is of high clinical relevance, as it might offer a unique advantage for the future development of disease-modifying therapies.

To assess the extent of epigenetic remodeling in COPD, we analyzed the expression changes of epigenetic enzymes and readers (Medvedeva *et al*, 2015). Thirty eight epigenetic factors were differentially expressed, with the majority (76%) showing up-regulation in COPD (Dataset EV4). Examples of dysregulated epigenetic players include histone methyltransferases (e.g., SETD1B, SUV39H2, KMT2D, EZH2), histone demethylases (KDM6B), and chromatin remodeling factors (e.g., CHAF1A, ATAD2 CHAF1B), indicating that in addition to DNA methylation and histone acetylation (Ito

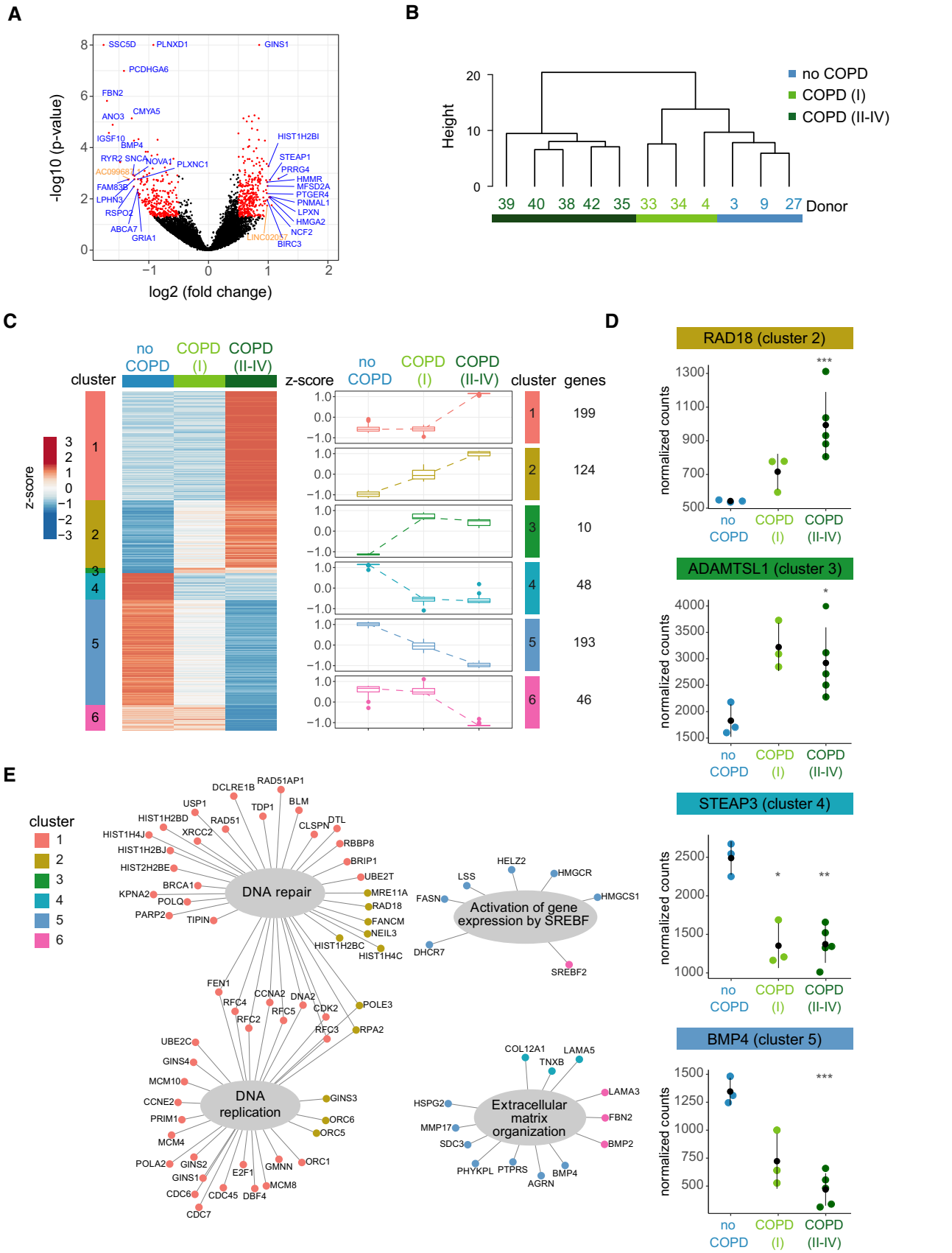

Figure 3.

Figure 3. DNA methylation changes in primary human lung fibroblasts are accompanied by gene expression changes in COPD.

A Volcano plot of differentially expressed genes (DEGs) (red dots; FDR < 0.05 and |log2(fold change)| > 0.5) in COPD (II–IV, $n = 3$) compared to no COPD controls ($n = 5$). Protein-coding (blue) and lincRNA (orange) with the highest expression change or lowest *P*-values are labeled.

B Unsupervised hierarchical clustering of all samples, including COPD (I) based on DEGs ($n = 620$) identified between no COPD and COPD (II–IV) samples.

C Self-organizing maps (SOM) clustering based on the scaled median expression level per group of the 620 DEG identified between COPD (II–IV) and no COPD samples. DEGs were grouped into six distinct clusters showing different kinetics in COPD progression: Clusters 1 and 6 show late changes. Clusters 2 and 5 display changes gradually progressing with disease severity. Clusters 3 and 4 correspond to early changes observed already in COPD (I). Left panel shows the row scaled (*z*-scores) expression levels per group as heatmap; right panel illustrates the distribution of the scaled expression levels in each group across all cluster as boxplots. The box represents the interquartile range (IQR). The line inside the box represents the median of the dataset. The whiskers extend from the edges of the box to the first and third quartiles, but no further than 1.5 times of the IQR. Any data points outside this range are plotted individually as dots. The dashed line connects the medians of the groups in each cluster. The number of each genes within the corresponding cluster is indicated on the right.

D Selected examples of DEG across disease stages from clusters 2–5 defined in (C).

E DEGs associated with altered biological processes (gray bubbles) in COPD. DEG nodes are colored according to their corresponding gene expression kinetic in COPD (clusters defined in C).

Data information: In (A), data points represent each gene, DEGs are indicated in red and defined by the following cutoffs: FDR < 0.05 and |log2(fold-change)| > 0.5. In (D) each donor is represented by an individual point (no COPD $n = 3$, COPD I $n = 3$, COPD II-IV $n = 5$), the group mean is shown as a black dot and the error bar represents the standard deviation. In (A) and (D) FDRs and log2(fold-changes) (in A) were calculated using DESeq2, which uses a negative binominal GLM (generalized linear model) and Wald statistics. *FDR < 0.05; **FDR < 0.01; ***FDR < 0.001. The specific statistically significant FDR values are the following: RAD18 in COPD II–IV = 0.000434; ADAMTSL1 in COPD II–IV = 0.02771; STEPA3 in COPD I = 0.04219 and in COPD II-IV = 0.001272; BMP4 in COPD II-IV = 0.0000535.

et al, 2005; Szulakowski *et al*, 2006) other epigenetic layers may also be dysregulated in COPD.

Integrative data analysis reveals epigenetically regulated genes in COPD fibroblasts

To further dissect the association between alterations in DNA methylation and changes in gene expression, we assigned DMRs to genes in their proximity. In total, we detected 4,059 genes associated with at least one DMR (in total 4,424 DMRs) within 4 kb upstream of the transcriptional start site (TSS) to 4 kb downstream of the transcriptional termination site (TTS; Fig 4A). About 45% of the gene-associated DMRs are located close to the TSS, mainly in the promoter and first intron (Figs 4B and EV4A and B). To further decipher the interplay between gene expression and DNA methylation changes in COPD, we focused our analysis on DMRs within 4 kb surrounding the TSS (Fig 4C). We observed an overrepresentation of DMRs at the promoter of differentially expressed genes compared to genes whose expression is not significantly changed in COPD (Fisher's exact test: $P\text{-value} = 1.3 \times 10^{-7}$, Fig EV4C). In total, 77 DEGs were associated with at least one DMR (Fig 4C, Dataset EV5), which was mainly located around or downstream of the TSS (Fig EV4B). In contrast to unchanged genes which exhibited the expected normal distribution (Fig 4D, blue line), differentially expressed genes displayed a bimodal curve with an enrichment at high positive and negative correlation rates, suggesting that some genes might be dysregulated by aberrant methylation in COPD (Fig 4D, red line). Pathway enrichment analysis of the DMR-associated DEGs (DMR:DEG) revealed genes involved in diseases of DNA repair ($FDR = 4 \times 10^{-5}$) and cell cycle checkpoints (FDR = 0.0008) (Fig EV4D). To identify common transcription factors at DMR:DEGs, we conducted a motif enrichment analysis. PAX5 (found in 6 DMR:DEGs) and TEAD2 (found in 15 DMR:DEGs) motifs were significantly enriched (*P*-value = 0.01, one-sided hypergeometric test). Additionally, we checked for the binding of TCF21, ZBTB7A and FOS2L, which we found enriched in our initial DMR motif analyses (Fig 2G and H). Here, we used TRACE data in human lung fibroblasts available in ENCODE (Ouyang & Boyle, 2020). For TCF21 and ZBTB7A, enriched in distal regulatory elements, we did not find any overlap between their putative binding sites and our DMR:DEGs. In contrast, FOSL2 binding was detected at GLI4, NEIL3, and BRCA1-associated DMRs.

We observed that early hypomethylated DMRs (Fig 1H, cluster 1) are enriched among the set of DMR:DEGs (*P*-value = 0.00007, one-sided hypergeometric test). To further investigate the relationship between gene expression and methylation changes, we classified DMR:DEGs using k-means clustering (Fig 4E, Dataset EV5). This analysis provides a direct connection between DNA methylation changes and accompanied expression changes in proximity. In clusters 2, 4, and 5, changes in methylation and expression occur at the same time. Interestingly, in clusters 1 and 3, we can already detect methylation changes in COPD (I) samples, whereas the expression is not changed yet, suggesting that some methylation changes may precede transcriptional changes. Next, we focused on DMR:DEGs pairs, which are directly correlated (as indicated by clusters 4 and 5 in Fig 4E) and exhibited an absolute correlation coefficient ≥ 0.5 (Fig 4D). Examples of genes showing correlation between DNA methylation and gene expression include UMPS, STEAP3, GABRR1, GLI4, AQP3, and LPXN (Figs 4F and EV4E), suggesting potential regulation of their expression by DNA methylation.

Functional siRNA screens identify novel regulators of COPD phenotypes in lung fibroblasts

Integration of DNA methylation and gene expression data, together with upstream regulator analysis using Ingenuity Pathway Analysis (Kramer *et al*, 2014) (https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/) allowed us to select candidate regulators of the observed epigenetic and transcriptional changes in COPD HLFs. Specifically, depending on the data and analysis type used, we divided the candidates into six categories (as detailed in Dataset EV6, selection criteria). These included the following: (i) top 10 upregulated and top 10 downregulated genes with the strongest log2 fold change (and FDR < 0.05) between no COPD and COPD (II–IV) samples in RNA-seq, (ii) top 4 dysregulated long non-coding RNAs with the

strongest log2 fold change (and FDR < 0.05) in RNA-seq, (iii) epigenetic factors that showed significant dysregulation in RNA-seq (FDR < 0.05), (iv) top 4 transcription factors with the strongest enrichment of their binding sites motifs in the identified DMRs, (v) top differentially expressed genes (FDR < 0.05) containing DMRs in promoters with the strongest Spearman correlation (> 0.5 or < −0.5), and (vi) strongest upstream epigenetic regulators from the Ingenuity Pathway Analysis. Three genes (TGFBR1, ACTA2, and BRD4) were used as positive controls for the redouts of the phenotypic assays. Overall, 110 candidates were manually selected for characterization using phenotypic screens (Fig EV5A, Dataset EV6), and 78% of them were not linked to COPD before. To determine the function of the selected candidates in key fibroblast processes related to COPD, high-content image-based phenotypic assays using small-interfering RNAs (siRNA)-mediated gene knockdown (KD) were carried out in primary human lung fibroblasts isolated from two normal healthy (NHLFs) and three COPD (DHLFs) donors (Fig 5A).

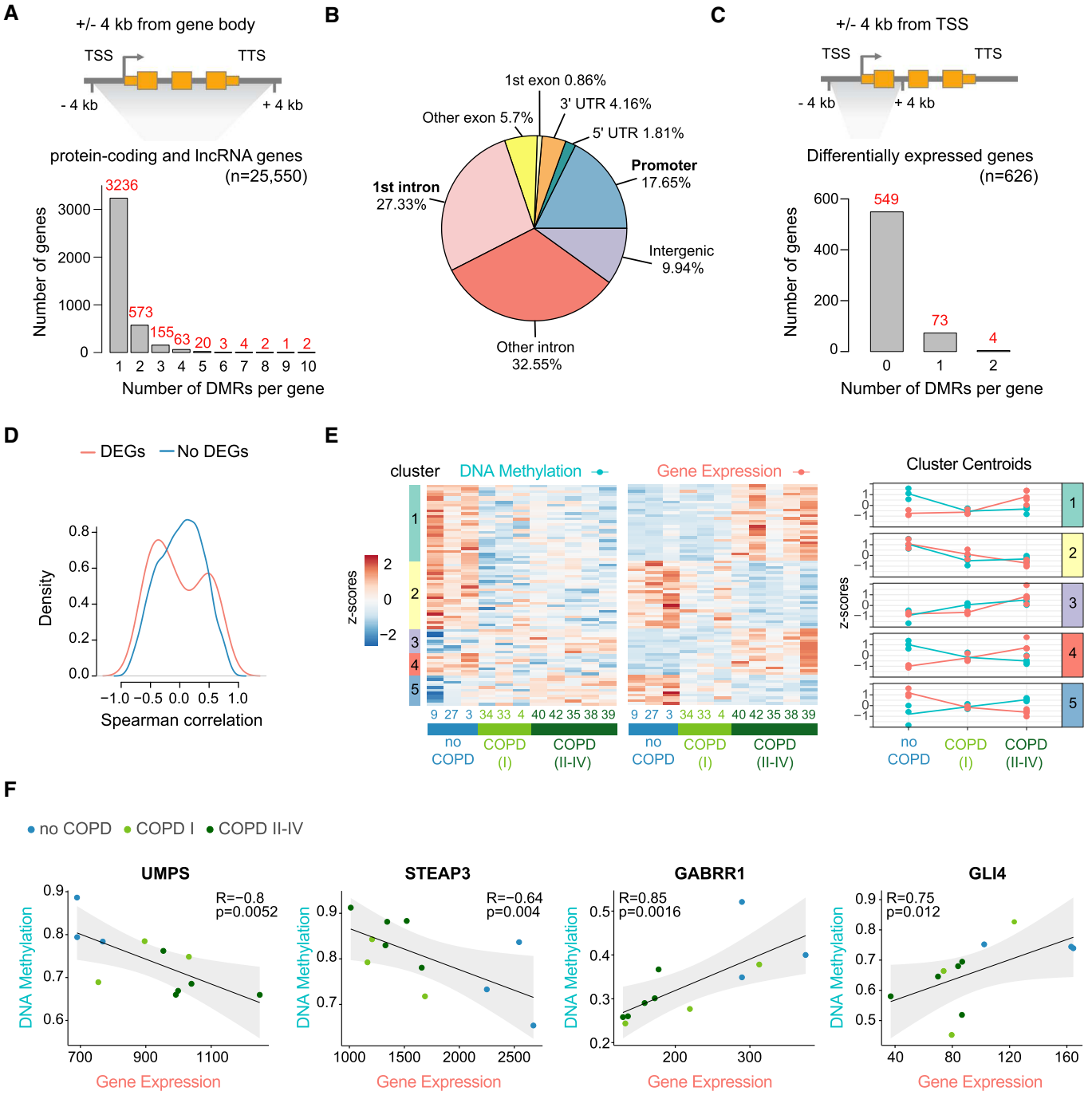

Figure 4.

Figure 4. Integrative data analysis reveals epigenetically regulated genes in COPD fibroblasts.

A DMRs located in the proximity of annotated protein-coding and lincRNA genes. DMRs within ± 4 kb from gene body were assigned to their corresponding gene. TSS, transcription start site; TTS, transcription termination site.

B Gene features of gene-associated DMRs. Promoter is defined as the region of −1 kb to +100 bp around the TSS.

C DMRs located in the proximity of the TSS of DEGs. DMRs within ± 4 kb from TSS of DEG were assigned to their corresponding gene.

D Spearman correlation between gene expression and DMR methylation. DMRs within ± 4 kb from TSS were considered. Gene-DMR pairs were split into DEGs (red) and not significantly changed genes (no DEG, blue).

E K-means clustering of DMRs and associated DEGs. The heatmaps show the *z*-scores of either rlog normalized expression counts (gene expression) or beta values (DNA methylation). Line diagram on the right side depicts the cluster centroids of each sample for gene expression (salmon) and DNA methylation data (turquoise).

F Scatter plots showing examples of correlations between gene expression and methylation of promoter associated DMRs. Each dot represents an individual donor. Dots are color coded according to disease state. Gene expression is illustrated as normalized counts. Methylation is illustrated as average beta value of the corresponding DMR.

Data information: In (F) linear regression analysis was performed (black line) and the 95% confidence interval is indicated (gray area). Correlation coefficients and *P*-values were calculated by the Spearman correlation method.

In COPD/emphysema, reduced proliferation, migration, and response to TGFβ1 of lung fibroblasts have been documented (Nobukuni *et al*, 2002; Holz *et al*, 2004; Muller *et al*, 2006; Togo *et al*, 2008), indicating the impaired tissue-repair capacity of fibroblasts in the COPD lung. Reduced fibroblast activity in the injured alveolar microenvironment has been proposed as a critical mechanism driving the development of emphysema (Plantier *et al*, 2007; Kulkarni *et al*, 2016). Thus, to evaluate fibroblast response to TGFβ1 upon candidate gene knockdown, we performed a TGFβ1-induced fibroblast-to-myofibroblast transition assay (FMT) (Weigle *et al*, 2019). High-content image-based quantification of collagen 1 deposition (col1) and α-smooth-muscle fibers (αSMA), two TGFβ1-responsive genes, were used as readouts. Additionally, to assess the proliferation capacity of primary fibroblasts upon candidate gene knockdown, we quantified the number of nuclei upon fibroblast growth factor 2 (FGF2) stimulation (Fig 5A and B).

The technical performance of the assays was demonstrated by the robust effect of the siRNAs targeting assay controls. As expected, siRNA targeting TGFβ receptor 1 (TGFβR1) showed a strong impact on the FMT assay (on both αSMA and col1 readouts) (Fig 5B), whereas the siRNA against ACTA2 (which encodes αSMA) showed a specific effect only in the αSMA readout.

To evaluate the differences between the non-targeting siRNA controls (NTC) and gene-targeted knockdowns, we calculated strictly standardized mean differences (SSMD) (Zhang, 2007; Zhang *et al*, 2007). Overall, a high hit rate was observed, as 61 out of the selected 110 candidates showed an effect in at least one assay after applying the strict cutoff of $|SSMD| > 3$ and 87 when using a cutoff of 2, demonstrating the power of multimodal analysis in identifying candidates with regulatory potential (Figs 5C and E, and EV5C, Dataset EV7). Furthermore, the strong correlation between the results obtained from different donors among normal fibroblasts (e.g., $R = 0.95$, *P*-value: 3.4×10^{-64} for αSMA) and fibroblasts derived from COPD patients (e.g., $R = 0.91$, *P*-value: 2.7×10^{-48} for αSMA) confirmed the robustness of the assays (Figs 5D and EV5B). Three genes, leupaxin (LPX), aquaporin 3 (AQP3), and GLI4, showed strong effects in all three readouts in both normal and diseased cells, indicating their critical function in fibroblast biology (Figs 5B, C, and E). Their knockdown efficiency was validated by RT–qPCR (Fig EV5E). Notably, among the positive hits, multiple epigenetic factors were also present (Figs 5C and EV5C, Dataset EV7). For example, we observed strong effects on lung fibroblast proliferation and differentiation upon targeted knockdown of different epigenetic enzymes (DNA methyltransferase DNMT3B, histone methyltransferases KMT2A, MKT2B, KMT3C, SETD1B, and EZH2, histone acetyltransferase EP300), chromatin remodeling factors (CHAF1A and CHAF1B) as well as epigenetic readers (BAZ2, CBX3), identifying these factors as key regulators of fibroblasts and COPD repair phenotypes.

Interestingly, we observed potential disease-specific effects for some of the tested candidates that were preserved between different donors. Here, the effect after siRNA-mediated gene knockdown varied between normal and diseased fibroblasts (Figs 5F and EV5D). For example, the knockdown of CHAF1A increased, while the knockdown of CHAF1B reduced the expression of αSMA, respectively, and both effects were stronger in diseased fibroblasts compared to normal cells. In turn, the knockdown of AQP3 in normal fibroblasts had a larger effect on αSMA levels upon TGFβ1 stimulation compared to diseased cells (Fig 5F). All three genes were dysregulated in our RNA-seq in COPD (CHAF1A and CHAF1B were upregulated, whereas AQP3 was downregulated in COPD, Dataset EV4), indicating that their dysregulation may be linked to COPD phenotypes in fibroblasts.

The cell-based assays in primary normal and COPD fibroblasts confirmed the functional role of numerous candidates identified from profiling data, indicating that integrating genome-wide epigenetic and transcriptomic profiling of purified normal and diseased human lung cells is a powerful approach for the identification of novel regulators of disease phenotypes. In addition, the presence of various epigenetic factors among the positive hits demonstrates that epigenetic regulation in COPD is an exciting research field that should be explored in-depth, as it may hold promise for novel therapeutic avenues for patients with COPD.

Discussion

In this study, we reveal novel transcriptomic and epigenetic signatures associated with COPD onset and progression, establishing a roadmap for further dissection of molecular mechanisms driving COPD phenotypes in lung fibroblasts.

Earlier studies using various patient materials consisting of mixed-cell populations provided evidence of dysregulated DNA methylation patterns in COPD and identified CpG sites and pathways associated with smoking and COPD (Sood *et al*, 2010; Qiu *et al*, 2012; Vucic *et al*, 2014; Wan *et al*, 2015; Yoo *et al*, 2015;

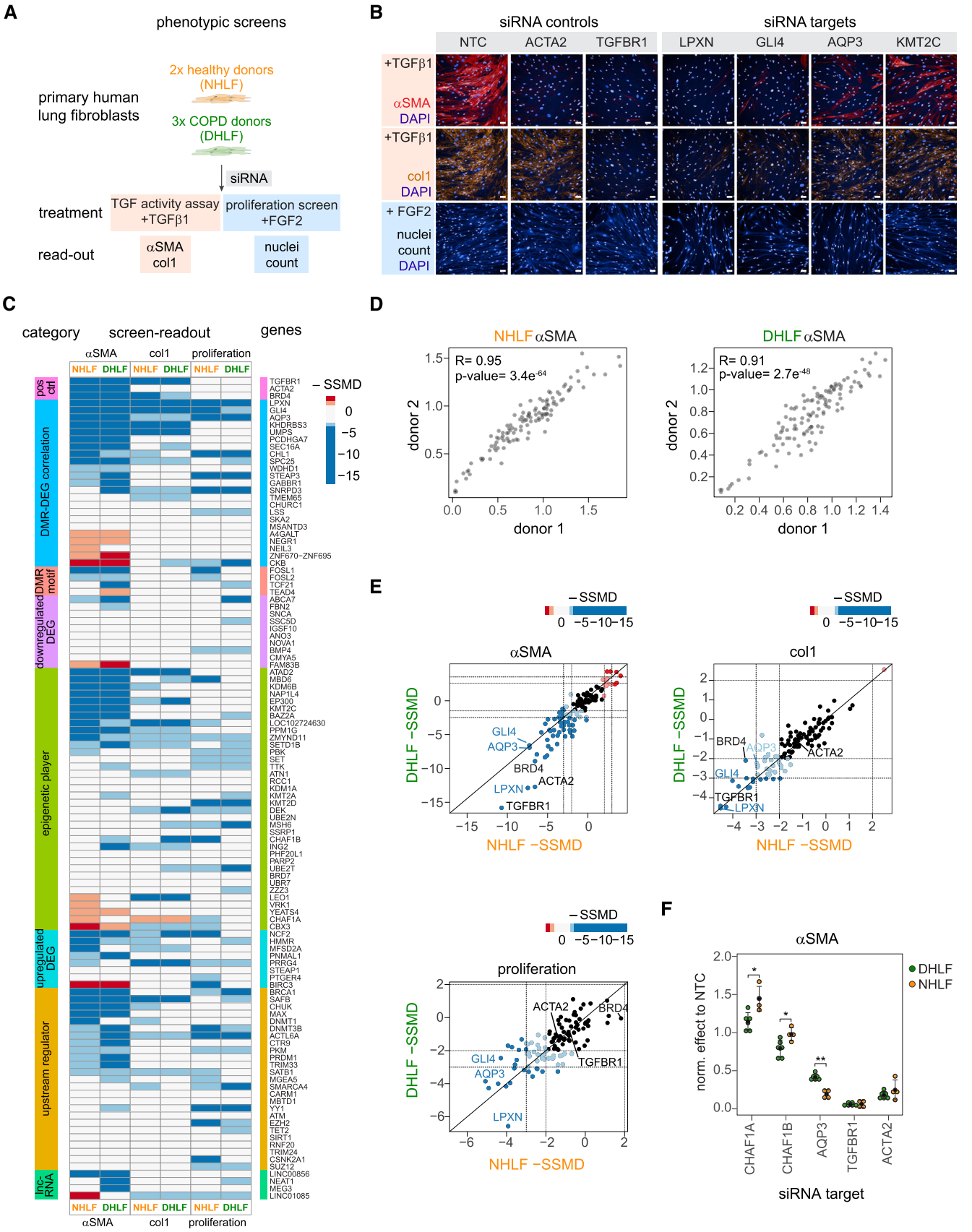

Figure 5.

Figure 5. siRNA-based phenotypic screens in normal and COPD primary human lung fibroblasts identify multiple candidate genes regulating COPD phenotypes.

A Schematic representation of the siRNA-based phenotypic assays performed in primary normal human lung fibroblasts (NHLFs, two donors) and diseased/COPD human lung fibroblasts (DHLFs, three donors).

B Examples of primary pictures obtained in the screens showing the performance of the siRNA controls as well as positive hits upon KD. Scale bars: 0.05 mm.

C Heatmap showing the effect of the KD of each candidate gene on the three measured readouts (αSMA, col1 and proliferation) in primary normal (NHLF) or COPD (DHLF) human lung fibroblasts, red: readout higher than NTC, blue: readout lower than NTC. ISSMD values| ≥ 2 are shown in lighter shade and ISSMD values| ≥ 3 in stronger shade.

D Scatterplots showing the correlation of the screen data from two different NHLFs donors (left) and two different DHLFs (right) for the αSMA readout.

E Comparison of the KD effect of each candidate relative to NTCs in NHLFs and DHLFs. Each dot represents a unique candidate tested, blue and red dots represent significant hits (ISSMD values| ≥ 2 are shown in lighter shade and ISSMD values| ≥ 3 in stronger shade). Assay controls are labeled in black and examples of strong hits regulating fibroblast to myofibroblast transition and cell proliferation processes are labeled in blue.

F Dot plot showing examples of positive hits with significant differences between NHLFs (2 donors, 2 biological replicates each, shown in green) and DHLFs (3 donors, 2 biological replicates each, shown in orange) in αSMA readout. Screen readout was normalized to the corresponding NTC. TGFβR1 and ACTA2 represent positive screen controls. The results of all replicates are shown.

Data information: In (D) *P*-value was derived from linear regression analysis and the Pearson correlation coefficient (*R*) is indicated. In (F) statistical evaluation was performed using an unpaired two-tailed student's *t*-test. **P*-value < 0.05; ***P*-value < 0.01. Black dots denote the means and error bars represent the standard deviation. KD, knockdown; FGF, fibroblast growth factor; TGFβ, transforming growth factor beta; αSMA, alpha smooth-muscle actin; col1, collagen 1; NTC, non-targeting siRNA control; SSMD, strictly standardized mean difference.

Busch *et al*, 2016; Morrow *et al*, 2016; Sundar *et al*, 2017; Carmona *et al*, 2018). Two recent publications also suggested that DNA methylation changes may originate in early life (Kachroo *et al*, 2020) and be linked to the severity of airflow limitation (Casas-Recasens *et al*, 2021). However, all these studies used low-resolution approaches, covering a representation of the genome only, mostly gene promoters. Hence, the full epigenomic landscape of COPD cells remained uncharted. Overall, there has been a limited consistency between different studies, likely coming from the cellular heterogeneity of the starting material, diverse donor selection criteria and different statistical models used. The dissection of cell-type specific mechanisms associated with COPD requires epigenetic profiling of defined cell populations. Only one study investigated DNA methylation changes in COPD patients with cell-type resolution (Clifford *et al*, 2018). Using Illumina 450 K BeadChip Array (focusing on gene promoters), Clifford *et al* identified 887 and 44 differentially methylated CpG sites in parenchymal and airway fibroblasts of COPD patients, respectively (Clifford *et al*, 2018). Our study, providing a much higher resolution of previously unexplored regions (e.g., enhancers) significantly extends these observations and demonstrates pronounced, genome-wide DNA methylation and gene expression changes in parenchymal fibroblasts in COPD, in both mild and severe disease. Direct comparison between both studies revealed a moderate positive correlation between the two data sets, which we interpret as validation considering the differences in both CpG detection technologies.

Little is known about the correlation of DNA methylation with disease severity. Methylation changes in 13 genes have been identified in the lung tissue of COPD GOLD I and II patients compared to non-smoker controls (Casas-Recasens *et al*, 2021). However, it is unclear whether they represent smoking- or COPD-related changes, as ex-smoker controls were not investigated in this study (Casas-Recasens *et al*, 2021). Our data reveal that genome-wide DNA methylation changes are present in lung fibroblasts of COPD (I) patients compared to controls with matched smoking status and history (all ex-smokers), demonstrating that epigenetic changes occur early in disease development. We cannot exclude that time since quitting smoking may have contributed to the methylation changes in our donors, as noted in previous studies (Zeilinger *et al*, 2013; Guida *et al*, 2015); however, there was no systematic and statistically significant difference between the three groups with regards to time since quitting; hence, it is unlikely that this factor would bias our identification of COPD-related methylation changes. Notably, COPD (I) samples clustered with COPD (II-IV) rather than no COPD samples, indicating that DNA methylation may provide a sensitive biomarker for early disease detection. As our analysis was performed in a small number of samples, this hypothesis awaits further replication and validation in larger patient cohorts.

Currently, it is unclear how altered DNA methylation patterns in COPD translate into biological effects in COPD fibroblasts. DNA methylation in regulatory regions can modulate the binding of transcriptional factors to DNA (Stadler *et al*, 2011); hence, methylation profiling allows for identifying transcriptional regulators potentially mediating the epigenetic alterations. We detected a significant enrichment of binding sites for TCF21 and FOSL2/FRA2 transcription factors in the DMRs overlapping with strong enhancers in COPD. TCF21 is a mesenchyme-specific basic helix–loop–helix transcription factor regulating multiple processes, including proliferation, extracellular matrix assembly, as well as secretion of pro-inflammatory mediators (Akama & Chun, 2018). It is required for lung development in mice, mesenchymal–epithelial crosstalk (Quaggin *et al*, 1999) and specification of fibroblast cell fate in different organs (Acharya *et al*, 2012; Braitsch *et al*, 2012). Recently, TCF21 has been identified as a specific marker of lipofibroblasts in the mouse (Park *et al*, 2019) and human lung (Liu *et al*, 2021), a subpopulation of fibroblasts essential for alveolar niche homeostasis and repair. Despite its central roles in alveolar development and maintenance, TCF21 function in human lung fibroblasts is largely unknown. Our phenotypic screens demonstrate that TCF21 is required for lung fibroblast proliferation and differentiation upon TGFβ1 stimulation, providing first insights into its molecular function in the human lung. Notably, the effects of TCF21 on proliferation and αSMA were stronger in diseased cells, indicating that COPD cells may be more sensitive to TCF21 loss than healthy lung fibroblasts. Consistent with our results, regulation of cell proliferation by TCF21 has been previously demonstrated in cancer cells (Lotfi *et al*, 2021).

FOSL2/FRA2 belongs to the activator-protein (AP)-1 family of transcription factors and is a known regulator of wound repair, and TGFβ-mediated fibrosis (Eferl *et al*, 2008). Increased FOSL2/FRA2 expression is detected in several chronic lung diseases, including pulmonary fibrosis, COPD, and asthma (Birnhuber *et al*, 2019). Notably, ectopic expression of FOSL2/FRA2 in mice results in fibrosis of several organs, including the lung, highlighting a potential profibrotic role of FOSL2/FRA2 (Eferl *et al*, 2008). Our high-content screens demonstrate that FOSL2/FRA2 is required for myofibroblast differentiation, consistent with its postulated profibrotic role. Collectively, our data suggest that TCF21 and FOSL2/FRA2, whose binding sites are enriched in DMRs at strong enhancers, may mediate some of the downstream biological effects in COPD fibroblasts and contribute to disease phenotypes, linking epigenetic changes to gene regulatory networks. Future work is needed to delineate and experimentally validate the target genes directly bound and regulated by these transcription factors in lung fibroblasts.

Transient TGFβ1 activity is required for lung tissue regeneration and repair upon injury, however its persistent activation in lung fibroblasts leads to aberrant repair and fibrosis (Fernandez & Eickelberg, 2012). In turn, reduced fibroblast proliferation and response to TGFβ1 has been proposed as a key mechanism driving the development of emphysema (Plantier *et al*, 2007; Togo *et al*, 2008; Konigshoff *et al*, 2009; Kulkarni *et al*, 2016). How these phenotypes are controlled at the molecular level is not well understood. We identified and functionally characterized numerous, previously unknown regulators of lung fibroblast function in COPD. Among the top candidates, silencing of the water/glycerol channel aquaporin 3 (AQP3), Hedgehog transcription factor GLI4 and focal adhesion protein leupaxin (LPXN) had the most drastic effects on both fibroblast proliferation and TGFβ1-mediated differentiation, establishing these three proteins as potential regulators of lung fibroblast repair and remodeling. AQP3 contribution to wound healing, ECM remodeling and cell proliferation has been well documented in other cellular systems (Xu *et al*, 2011; Ryu *et al*, 2012; Chen *et al*, 2014; Huang *et al*, 2015; Hou *et al*, 2016; Luo *et al*, 2016; Xiong *et al*, 2017) and the role of LPXN in cancer cell proliferation and migration through regulation of focal adhesion sites is recognized (Kaulfuss *et al*, 2008; Dierks *et al*, 2015). Hence, downregulation of AQP3 in COPD fibroblasts could contribute, at least in part, to the decreased proliferation and contractility, manifesting in reduced fibroblast activity and impaired response to injury in emphysema (Togo *et al*, 2008). In support of our data, change in expression of AQP3 (Heinbockel *et al*, 2018) and LPXN (Spira *et al*, 2004) in COPD/emphysematous lung tissue has been observed previously, consistent with their dysregulation in COPD fibroblasts in our RNA-seq. Notably, we showed that changes in expression of AQP3, LPXN, and GLI4 in COPD fibroblasts were also associated with aberrant methylation in proximity to their TSS, raising the possibility that their epigenetic regulation may be one of the factors contributing to the reduced repair capacity of lung fibroblasts in emphysema (Muller *et al*, 2006; Togo *et al*, 2008).

Notably, we observed strong effects on lung fibroblast proliferation and differentiation upon targeted knockdown of different epigenetic enzymes and readers, identifying these factors as key regulators of fibroblasts and COPD repair phenotypes. The imbalance of histone acetyltransferase (HAT) and deacetylase (HDAC) activities has previously been linked to COPD (Ito *et al*, 2005), providing a scientific basis for the potential use of bromodomain (BET) and HDAC inhibitors in COPD (van den Bosch *et al*, 2017); however, the potential of targeting other dysregulated epigenetic activities in COPD remains to be explored.

Collectively, our results demonstrate that focused, high-resolution profiling of defined cell populations in COPD effectively complements large-cohort epigenetic biomarker studies and can provide important insights into the COPD-driving cell populations and associated mechanisms. Integration of -omics data across disease stages is a powerful tool for the identification of novel candidate disease regulators and potential sensitive biomarkers. Future large-scale profiling of early disease in an extended patient cohort will be crucial for an improved understanding of COPD pathology and will guide the development of new diagnostic strategies and disease-modifying therapies for early disease.

Materials and Methods

Study approval

The protocol for tissue collection was approved by the ethics committees of the University of Heidelberg (S-270/2001), Ludwig-Maximilians-Universität München (projects 333-10 and 17-166) and the University of Texas Health Science Center at Houston (HSC-MS-08-0354 and HSC-MS-15-1049). The experiments followed the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report. All patients gave written informed consent before inclusion in the study and remained anonymous in the context of this study.

Patient samples

Lung tissue samples were obtained through collaborations with the Lung Biobank Heidelberg at the Thoraxklinik (Heidelberg, Germany), the Asklepios Clinic (Gauting, Germany) and the UTHealth Pulmonary Center of Excellence (Houston, TX, USA). Residual lung parenchyma samples were obtained from patients undergoing lung surgery due to primary squamous cell carcinomas (SCC) who had not received chemotherapy or radiation within 4 years before surgery or from COPD patients undergoing lung resection. Normal human lung tissue was only used for protocol optimization and obtained from the International Institute for the Advancement of Medicine (IIAM), from lungs rejected for transplantation due to reasons unrelated to obvious acute or chronic pulmonary disease.

Collection of lung tissue samples for profiling

To identify molecular changes associated with COPD development and progression we collected distal lung tissue from patients across COPD stages and divided them into three groups: (i) no COPD, (ii) mild COPD (stage I, according to GOLD classification (GOLD, 2021) and (iii) established COPD (GOLD stages II–IV, Fig EV1A). Strict patient inclusion criteria for prospective tissue collection were established to ensure best possible matching of control and disease groups. To avoid direct smoking effects (van der Vaart *et al*, 2004), all included donors were ex-smokers. In addition, lung function results, as well quantitative emphysema score index (ESI) based on chest CT and whenever possible, medical history were collected for

each patient for their best possible characterization. Patients' characteristics and exemplary images from hematoxylin and eosin staining of the tissue are provided in Dataset EV1 and Fig EV1B, respectively. Each tissue sample was reviewed by an experienced lung pathologist, who confirmed that all samples were tumor-free and evaluated COPD relevant phenotypes, like emphysema, airway thickening, and immune infiltration. Only ex-smokers with preserved lung function and no indication of emphysema or fibrosis in the test results or patient history were included as control samples. Importantly, we included two samples from COPD (GOLD II-IV) donors (HLD38 and HLD39), which originated from lung resections, ensuring that the observed changes are also present in COPD tissues without cancer background.

There were no significant differences between control and COPD donors regarding age, body mass index, smoking status, and smoking history, but the control and COPD group could be clearly separated based on lung function data (Figs 1A and EV1A, Dataset EV1). Tissue samples that met the inclusion criteria were cryopreserved upon collection to allow their thorough characterization by an experienced lung pathologist before cell isolation and profiling (Figs 1B and EV1B). We have previously shown that this step is crucial to ensure the exclusion of low-quality control samples presenting additional lung pathologies, which may result in confounding effects in sequencing-based analyses (Llamazares-Prada *et al*, 2021).

Emphysema score index (ESI) determination

Lung and emphysema segmentation were performed to calculate the ESI from clinically indicated preoperative CT scans taken with mixed technical parameters. After automated lung segmentation using the YACTA software, a threshold of -950 HU was used with a noise-correction range between -910 and -950 HU to calculate the relative amount of emphysema in % of the respective lung portion (Lim *et al*, 2016). While usually global ESI was measured, only the contralateral non-affected lung side was used if one lung was severely affected by the tumor.

FFPE and H&E

Representative slices from different areas of the tissue were fixed O/N with 10% neutral buffered formalin (Sigma-Aldrich). Next, fixed tissue samples were washed with PBS (Fisher Scientific) and kept in 70% ethanol at 4°C until embedding. Sample dehydration, paraffin embedding, and hematoxylin and eosin (H&E) staining was performed at Morphisto (Morphisto GmbH, Frankfurt, Germany). Two 4 µm thick sections were cut per sample using a Leica RM2255 microtome with an integrated cooling station and water basin and transferred to adhesive glass slides (Superfrost Plus, Thermo Fisher). Subsequently, the sections were dried O/N in a 40°C oven to remove excess water and enhance adhesion. H&E-stained slides were evaluated by an experienced lung pathologist at the Thoraxklinik in Heidelberg (Germany).

Cryopreservation of lung parenchyma

The lung tissues from all donors were cryopreserved on arrival, enabling cell isolation from multiple donors in parallel and minimizing technical bias introduced by multiple batch processing of fresh samples, as documented previously (Llamazares-Prada *et al*, 2021; Pohl *et al*, 2023). Briefly, specimens were transported in CO_2-independent medium (Thermo Fisher Scientific) supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific), and 1% Amphotericin B (Fisher Scientific). Upon reception, tissue pieces were carefully inflated with ice-cold HBSS (Fisher Scientific), supplemented with 2 mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific), and 1% Amphotericin B (Fisher Scientific). Exemplary samples of the different areas of the lung piece were collected for subsequent histological analysis. The pleura was removed from the remaining tissue, and airways and vessels separated from the parenchyma as much as possible. The parenchymal airway and vessel-free fractions were minced, transferred to cryo-tubes, covered with ice-cold freezing medium [70% DMEM, high glucose with GlutaMAX™ (Thermo Fisher Scientific), 20% FBS (Gibco) and 10% DMSO (Carl Roth)], kept on ice for 15 min, and transferred to −80°C in Mr. Frosty™ containers (Nalgene) to ensure a gradual temperature decrease (1°C/min). For long-term storage, samples were kept in liquid nitrogen.

Fibroblast isolation from human lung tissue

Fibroblasts isolation followed a strict, standardized protocol to minimize potential biases related to tissue location and batch effects. We took care to have a representation of different disease stages during cell isolation to avoid introducing disease-stage-related batch effects. As no universal fibroblast markers for FACS are available, primary human lung fibroblasts were isolated by explant outgrowth from tumor-free, distal parenchymal lung tissue that has been depleted from visible airways and vessels, following previously described protocols (Hallgren *et al*, 2010; Dessalle *et al*, 2016; Clifford *et al*, 2018; Llamazares-Prada *et al*, 2021). To have a better representation of the parenchymal fibroblasts and preserve lung heterogeneity, we collected tissue explants from different regions of the parenchyma and combined them for cell isolation. Briefly, 7–8 micro-dissected lung parenchyma pieces were placed per well into 6-well plates, let for 30 min at RT without medium to improve explant attachment, and carefully covered with 1 ml of growth medium: DMEM, high glucose, GlutaMAX™ (Thermo Fisher Scientific) supplemented with 2% FBS (Gibco) and 1% penicillin & streptomycin (Thermo Fisher Scientific). Explants were left undisturbed for 4–7 days, afterward the medium was exchanged every 2 days, and the outgrowth of fibroblasts from the explants was followed daily. Cells were collected from multiple explant pieces when reaching 70% confluency to preserve the fibroblast heterogeneity. Possible epithelial contamination was prevented by short trypsinization of the outgrown cells during the first passage (0.05% trypsin with EDTA (Gibco), 3 min at 37°C) and keeping cells in the growth medium indicated above, suitable for fibroblasts enrichment. Fibroblast cells for RNA-seq and T-WGBS profiling were collected from all donors in passage 3.

Immunofluorescence of human lung fibroblasts

The purity of the isolated fibroblasts was assessed by immunofluorescence using mesenchymal markers vimentin (VIM) and alpha

smooth muscle actin (αSMA) as follows. 10^4 human lung fibroblasts in passage 3 were seeded per well in a 96-well plate for imaging (Zell-kontakt). 48 h later, cells were washed with 1X PBS (Fisher Scientific), fixed for 10 min with 4% PFA (Sigma-Aldrich) at RT, washed and permeabilized for 10 min with 0.3% Triton-X-100 (Carl Roth) at RT. Unspecific staining was blocked by incubating 1 h at RT with blocking buffer: 5% BSA (Carl Roth), 2% Normal Donkey Serum (Abcam) in 1× PBS (Fisher Scientific). Cells were incubated with primary antibodies against VIM (sc-7557, Santa Cruz Biotechnology, 1:200) and αSMA (ab7817, Abcam, 1:100) overnight at 4°C and labeled with respective secondary antibodies [donkey anti-goat IgG Alexa Fluor 488 (A-11055, Thermo Fisher Scientific, 1:500) and donkey anti-mouse IgG Alexa Fluor 568 (A-10037, Thermo Fisher Scientific, 1:500)] for 40 min at RT in the dark. After washing with 1X PBS (Fisher Scientific), the nuclei were counterstained with DAPI (Thermo Fisher Scientific, 1:5,000) for 10 min at RT and washed with 1X PBS. Stained and fixed cells were kept in 1× PBS (Fisher Scientific) at 4°C in the dark until imaging. Imaging was conducted at the ZMBH imaging facility (Heidelberg, Germany) using the Zeiss LSM780 confocal fluorescent microscope.

FACS analysis of isolated fibroblasts and lung suspension

Cryopreserved lung tissues were thawed for 2 min in a 37°C waterbath, collected in 50 ml Falcon tubes, and washed with wash buffer: HBSS supplemented with 2 mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific). Tissue was minced into smaller pieces prior to mechanical and enzymatic dissociation as indicated previously (Llamazares-Prada *et al*, 2021; Pohl *et al*, 2023). Briefly, the minced tissue (1 g) was introduced in GentleMACS C-tubes (Miltenyi Biotec) containing 10 μM ROCK inhibitor (Y-27632, Adooq Bioscience), 10 μg DNase I (ProSpec-Tany TechnoGene), the enzyme mix from the human tumor tissue dissociation kit (Miltenyi Biotec) and 4.5 ml of CO_2-independent media (Thermo Fisher Scientific) supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific). Tubes were closed tightly, introduced into the GentleMACS dissociator (Miltenyi Biotec) for mechanic disruption and the following program was performed: program h_tumor_01, followed by 15 min incubation at 37°C on a rotator; h_tumor_01, 15 min at 37°C on a rotator; h_tumor_02, and 15 min at 37°C on a rotator for a final enzymatic dissociation and a last mechanical shearing using the program h_tumor_02. The samples were pipetted up and down to help disaggregation. Finally, the enzymatic reaction was stopped by adding 20% FBS (Gibco) and single cells were collected by sequential filtering through 100 μm, 70 μm, and 40 μm cell strainers (BD Falcon). Cells were centrifuged, resuspended in ACK lysis buffer (Sigma-Aldrich), and incubated for 3 min at RT to lyse erythrocytes. Lung single-cell suspensions were washed with HBSS (Fisher Scientific) supplemented with 2 mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific), and 1% Amphotericin B (Fisher Scientific).

To generate fibroblast single-cell suspensions, passage 3 fibroblasts were trypsinized 5 min at 37°C using 0.05% trypsin with EDTA (Gibco), centrifuged at 250 *g* for 5 min at RT, and resuspended in HBSS (Fisher Scientific) supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific).

Lung and fibroblast single-cell suspensions were incubated with human TruStain FcX (BioLegend) for 30 min on ice to block Fc receptors. Immune and epithelial cells were labeled using CD45 (CD45-Bv605, BD Bioscience) and EpCAM (anti-human CD326 -PE, Affymetrix eBioscience) antibodies respectively for 30 min in the dark at 4°C following manufacturer instructions. Stained samples were washed with PBS 1X (Fisher Scientific) and resuspended in HBSS (Fisher Scientific) supplemented with 2 mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific). Stained cells were added to Falcon 5 ml polystyrene tubes with 40 μm cell strainer caps (Neolab Migge). To discriminate between live and dead cells, we used SyTOX blue (Thermo Fisher Scientific) as recommended by the manufacturer.

RNA isolation and RNA-seq

10^5 HLFs were harvested at passage 3 for RNA-seq studies 24 h after fresh medium change by scraping in ice-cold PBS. To minimize technical batch effects, cells were washed, pelleted, flash-frozen in liquid nitrogen, and stored at −80°C until all donors were collected. Afterwards, RNA isolation, library preparation, and sequencing of all samples were performed simultaneously. Total RNA was isolated using RNeasy plus micro kit (Qiagen, Hilden, Germany) following manufacturer's instructions. DNA was removed by passing the lysate through the gDNA eliminator column and by an additional on-column DNase I treatment (Qiagen) before the elution. RNA was eluted using nuclease-free water (Thermo Fisher Scientific) and the concentration measured with Qubit HS Kit (Thermo Fisher Scientific). RNA integrity was assessed using the Bioanalyzer 2100 (Agilent, model G2939A) and the RNA 6000 pico kit (Agilent). Only samples with RIN > 8.5 were processed.

Libraries were prepared at the Genomics core facility (GeneCore) at EMBL using 200 ng of total RNA as input. Ribosomal RNA was removed by Illumina Ribo-Zero Gold rRNA Removal Kit Human/Mouse/Rat (Illumina, San Diego, CA, USA) and strand-specific total RNA-seq libraries were prepared using the Illumina TruSeq RNA Sample Preparation v2 Kit (Illumina, San Diego, CA, USA) implemented on the liquid handling robot Beckman FXP2. Obtained libraries were pooled in equimolar amounts. 1.8 pM solution of each library was pooled and loaded on the Illumina sequencer NextSeq 500 High output and sequenced uni-directionally, generating ~450 million reads per run, each 75 bases long. General statistics about the quality of the RNA-seq can be found in Table EV1.

RNA-seq read alignment and transcript abundance quantification

Single-end reads were mapped to the human genome version 37 (hg19) and the reference gene annotation (release 70, Ensembl) using STAR v2.5.0a (Dobin *et al*, 2013) with following parameters: --outFilterType BySJout –outFilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outFilterMismatchNmax 999 –alignIntronMin 20 –alignIntronMax 100000 –outFilterMismatchNoverReadLmax 0.04 –outSAMtype BAM SortedByCoordinate –outSAMmultNmax 1 –outMultimapperOrder Random.

Contamination of PCR duplication artifacts in the RNA-seq data was controlled using the R package dupRadar (Sayols *et al*, 2016). The *featureCounts* script (Liao *et al*, 2014) of the Subread package v1.5.3 was used to assign and count mapped reads to annotated protein-coding and lncRNA genes with default settings.

Differential gene expression analysis

Statistical analysis of differential gene expression was performed with the DESeq2 Bioconductor package (Love *et al*, 2014). For exploratory RNA-seq data analysis, the data needs to be homoscedastic. Therefore, the raw counts were transformed by the regularized-logarithm transformation *rlog*. Genes with less than 32 counts in at least five samples were excluded from further analysis. Ex-smokers with preserved lung function (no COPD, $n = 3$) were considered as ground state and differential gene expression in COPD patients classified as GOLD Grade II-IV ($n = 5$) was identified as a significant change in expression by an FDR (false discovery rate) < 0.05 and an absolute log2 fold change > 0.5 (corresponding to fold change > 1.4), after fold change correction with the built-in *lfcShrink* function.

DNA isolation and T-WGBS

For DNA methylation analysis, DNA isolation and library preparation were performed in two rounds, but for each round, a representation of samples from different disease stages was ensured to avoid potential disease-stage-related batch effects. Genomic DNA was extracted from 2×10^5 primary human lung fibroblasts harvested in passage 3 using QIAamp Micro Kit (Qiagen, Hilden, Germany) following manufacturer's protocol, with an additional RNase A treatment step. T-WGBS was essentially performed as described previously (Wang *et al*, 2013) using 30 ng genomic DNA as input. 15 pg unmethylated DNA of phage lambda was used as control for bisulfite conversion. Four sequencing libraries were generated per sample using 11 amplification cycles. For each sample, equimolar amounts of all four libraries were pooled and sequenced on two lanes of a HiSeq2500 (Illumina, San Diego, California, US) machine at NGX Bio (San Francisco), resulting in 100 bp, paired-end reads. For different sequencing runs, a representation of libraries from disease stages in each flow cell was ensured. General statistics about the quality of the WGBS can be found in Table EV2.

Read alignment

The whole genome bisulfite sequencing mapping pipeline MethylCtools with modifications to adapt for the T-WGBS data was used (https://github.com/hovestadt/methylCtools) (Hovestadt *et al*, 2014). Briefly, the hg19 reference genome (37d5) was transformed *in silico* for both the top strand (C to T) and bottom strand (G to A). Before alignment, adaptor sequences were trimmed using Trimmomatic (release 0.35) (Bolger *et al*, 2014). The first read in each read pair was then C-to-T converted and the 2nd read in the pair was G-to-A converted. The converted reads were aligned to a combined reference of the transformed top (C to T) and bottom (G to A) strands using BWA MEM (bwa-0.7.8) with default parameters, yet, disabling the quality threshold for read output (–T 0) (Li & Durbin, 2009). After alignment, reads were converted back to the original states, and reads mapped to the antisense strand of the respective reference were removed. Duplicate reads were marked, and the complexity determined using Picard *MarkDuplicates* (http://picard.sourceforge.net/). Total genome coverage was calculated using the total number of bases aligned from uniquely mapped reads over the total number of mappable bases in the genome.

Methylation calling

At each cytosine position, reads that maintain the cytosine status were considered methylated, and the reads that have cytosine converted to thymine were considered unmethylated. Only bases with Phred-scaled quality score of ≥ 20 were considered. In addition, the 10 bp at the two ends of the reads were excluded from methylation calling according to M-bias plot quality control. In addition, CpGs located on sex chromosomes were removed from analysis.

DMR calling

Differences in CpG methylation profiles of no COPD donors ($n = 3$) and patients diagnosed with COPD ($n = 5$) were analyzed using the R/Bioconductor package bsseq (Hansen *et al*, 2012). First, the data were smoothed using the built-in *Bsmooth* function with default settings. Only CpG sites with a coverage of at least 4× were kept for subsequent analysis. A *t*-statistic was calculated between no COPD and COPD (II–IV) samples using the *Bsmooth.tstat* function with following parameters: local.correct = TRUE, maxGap = 300, estimate.var = "same". Differentially methylated regions (DMRs) were called by (i) selecting the regions with the 5% most extreme *t*-statistics in the data (lower and upper 2.5% quantile; default parameters of the dmrFinder function), (ii) filtering for regions exhibiting at least 10% methylation difference between no COPD and COPD (II–IV) and containing at least 3 CpGs. (iii) Finally, a non-parametric Wilcoxon test was applied using the average methylation level of the region to remove potentially false positive regions, since the *t*-statistic is not well-suited for not normally distributed values, as expected at very low/high (close to 0% / 100%) methylation levels. A significance level of 0.1 was used. No further FDR correction was performed.

Validation of T-WGBS workflow with mass Array

While establishing the T-WGBS for this project, we initially validated our approach using matrix-assisted time-of-flight mass spectrometry (MassARRAY; Agena Bioscience), a sequencing-independent method. For this, we performed T-WGBS on the commercially available smoker and COPD lung fibroblasts (purchased from Epithelix) and selected nine regions with different levels of methylation for validation using MassARRAY. The MassARRAY assay was performed as described previously (Ehrich *et al*, 2005). As shown in Fig EV1K, we obtained an excellent correlation between both methods, providing technical validation of T-WGBS and our workflow.

Gene ontology analysis

The closest genes were assigned to DMRs and subjected to gene ontology enrichment analysis using GREAT (McLean *et al*, 2010).

To define significant associations with pathways, we used the default settings of the GREAT tool, which are as follows: FDR < 0.05 in both binominal and hypergeometric tests and minimum region-based fold enrichment of 2.

ChIP-seq data

Histone modification ChIP-seq data of human lung fibroblasts were obtained from the ENCODE portal (https://www.encodeproject.org/) (Davis *et al*, 2018) with the following identifiers: ENCFF354IJB, ENCFF070CZY, ENCFF377BNX, ENCFF227WSF, ENCFF208SHP, ENCFF386FDQ, ENCFF102BGI, ENCFF843AYT.

Chromatin states for adult human lung fibroblasts (accession: ENCFF001TDQ) were obtained from ENCODE data base (Ernst & Kellis, 2017). The number of bp of each DMR coinciding with a chromatin state was calculated and the chromatin state with the largest overlap was assigned to the DMR. To assess the genomic background, 10,000 regions with matching size and CpG distribution were randomly selected.

TRACE inferred binding sites of TCF21 (ENCFF506LDV), FOSL2 (ENCFF248CHP) and ZBTB7A (ENCFF063SAY) in human lung fibroblasts were obtained from ENCODE (https://www.encodeproject.org/) (Davis *et al*, 2018). Enrichment at selected DMRs was tested using the enrichPeakOverlap from the Bioconductor package clusterProfiler (Wu *et al*, 2021).

Profile plots of DMRs and identification of super-enhancers

Enrichment of H3K4me1, H3K27ac, and H3K27me3 signals at DMRs, stratified in hypo- and hypermethylated regions, was performed with peakSeason (https://github.com/PoisonAlien/peakseason).

Super Enhancers (SE) were identified using ROSE (Rank Ordering of Super Enhancers) software (v.0.1; https://bitbucket.org/young_computation/rose), by merging closely spaced (< 12.5 kb) enhancer peaks (H3K4me1 peaks overlapping with H3K27ac peaks) (Whyte *et al*, 2013). Further on, all enhancers were ranked by their H3K27ac signals. Separation of SE and enhancers was performed based on the geometrical inflection point.

Transcription factor motif analysis

All hypomethylated DMRs which showed an overlap with the strong enhancer chromatin state were selected and motif enrichment analysis was carried out using the *findMotifsGenome.pl* script of the HOMER software suit omitting CG correction.

In order to obtain information about methylation dependent binding for transcription factor motifs which are enriched at DMRs, the results of a recent SELEX study (Yin *et al*, 2017) were integrated in the analysis and a motif database of 1,787 binding motifs with associated methylation dependency was constructed. The log odds detection threshold was calculated for the HOMER motif search as following.

Bases with a probability > 0.7 get a score of log(base probability/0.25), otherwise the score was set to 0. The final threshold was calculated as the sum of the scores of all bases in the motif. Motif enrichment analysis was carried out against a sampled background of 50,000 random regions with matching GC content using the *findMotifsGenome.pl* script of the HOMER software suit omitting CG correction and setting the generated SELEX motifs as motif database.

For the TF enrichment analysis within the clusters, chromatin states for adult human lung fibroblasts (accession: ENCFF001TDQ) were obtained from ENCODE database (Ernst & Kellis, 2017). The number of bp of each DMR coinciding with a chromatin state was calculated and the chromatin state with the largest overlap was assigned to the DMR. All hypomethylated DMRs which showed an overlap with the strong enhancer chromatin state were selected and split by the DMR k-means cluster analysis into two groups (cluster1 and cluster3). Motif enrichment analysis against known motifs was carried out using the findMotifsGenome.pl script of the HOMER software suit omitting CG correction, as described above.

siRNA-based phenotypic assays in primary human lung fibroblasts

Normal human lung fibroblasts (NHLFs) from two donors (donor IDs: 608197; 543644) and diseased COPD human lung fibroblasts (DHLFs) from three donors (donor IDs: OF3353, OF3418, and OF3238) were purchased from Lonza and tested for their response to FGF2 and TGFβ stimulation.

Fibroblast to myofibroblast transition (FMT) assay

Cells in passage 5 were plated in a poly-D-lysine coated 384 CellCarrier microtiter plate from PerkinElmer in fibroblast basal medium (FBM) with FGM-2TM Single Quots (Lonza) at a density of 2,000 cells per well. Six hours after cell seeding, cells were transfected with siRNAs (Horizon ON-Target Plus siRNA pools, Table EV3) as previously described (Weigle *et al*, 2019). Twenty four hours later, the medium was replaced by FBM containing 0.1% fetal calf serum (starvation medium). Twenty four hours later, fibroblast to myofibroblast differentiation was initiated by adding fresh starvation medium containing a mixture of Ficoll 70 and 400 (GE Healthcare; 37.5 mg/ml and 25 mg/ml, respectively), 200 μM vitamin C and 5 ng/ml TGFβ1. After 72 h the medium was removed, cells were fixed with 100% ice-cold methanol for 30 min, washed with PBS, permeabilized 20 min using 1% Triton-X-100 (Sigma), washed, and blocked for 30 min with 3% BSA in PBS. After an additional wash step, cell nuclei were stained using 1 μM Hoechst 33342 (Molecular Probes). Alpha smooth-muscle actin (αSMA) and collagen I (col1) were stained using monoclonal antibodies (1:1,000 diluted, Sigma, A2547 and SAB4200678, respectively). For detection of primary antibodies, cells were washed and incubated for 30 min at 37°C with AF647-goat-anti-mouse IgG2b (αSMA) and AF568 goat-anti-mouse IgG1 (col1) antibodies. After removal of secondary antibodies, cells were stained with HCS Cell Mask Green stain (Invitrogen, 1:50,000). Following a final PBS 1× wash step, images were acquired in a GE Healthcare InCell 2200 Analyzer, using 2D-deconvolution for nuclei (Hoechst channel), cells (FITC channel), αSMA (Cy5 channel) and collagen I (TexasRed channel), and images were transferred to and analyzed using Perkin Elmer's Columbus™ Image Storage as previously described (Aumiller *et al*, 2017; Weigle *et al*, 2019). Briefly, the building blocks (BB) of the Columbus™ Image Analysis system were used, first nuclei (Hoechst channel acquisition) were detected using the BB "nuclei". Second, cells were defined with the BB "find cytoplasm" from the FITC channel image. αSMA fibers and col1 area were defined by two individual BBs "find simple image region" based on images acquired in the Cy5 and TexasRed channels,

respectively. Both αSMA fibers and col1 readouts were normalized to the number of cells per image field. The FMT assay was performed in 2 NHLFs and 3 DHLFs independent donors. For each donor, siRNA transfection for every gene was performed in four technical replicates. In addition, the FMT screen was performed in each donor twice independently.

Proliferation assay (nuclei count)

To analyze the effects of gene knockdown on FGF2-mediated fibroblast proliferation, 2,000 cells at passage 5 were plated in a poly-D-lysine coated 384-CellCarrier microtiter plate (PerkinElmer) in FBM with FGM-2TM SingleQuots™ (Lonza). Six hours after seeding, cells were transfected with siRNAs (Dharmacon ON-Target Plus siRNA pools, HORIZON discovery, Table EV3) as previously described (Weigle *et al*, 2019). Twenty four hours later, the medium was replaced by FBM containing 0.1% fetal calf serum (starvation medium). Twenty four hours later the medium was replaced by starvation medium containing 20 ng/ml basic FGF (R&D Systems). After 72 h the medium was removed, cells washed with PBS and treated with 3.7% formaldehyde containing 1 μM Hoechst 33342 for 30 min. Cells were washed with PBS and images were acquired in a GE Healthcare InCell 2200 Analyzer, using 2D-deconvolution for nuclei (Hoechst channel). Nuclei numbers were determined using the Columbus™ image analysis software as described above (BB "nuclei"). The proliferation assay was performed in five independent donors: two NHLFs and three DHLFs. For each donor, siRNA transfection to knockdown the selected candidate genes was performed in four technical replicates. In addition, the proliferation screen was performed in each donor twice independently.

Analysis of phenotypic screen data

siRNA transfection was performed in each phenotypic screen in four technical replicates and repeated two times independently for each donor. Three DHLF and two NHLF independent donors were used for each screen. For statistical analysis of both the FMT and proliferation data, each readout (nuclei for both FMT and proliferation, αSMA and col1 for FMT) was first normalized within each plate, based on the negative control wells, corresponding to cells transfected with non-target siRNA control (NTC) (40 wells per plate). After plate-based normalization, the normalized values for the specific readout (e.g., nuclei, αSMA and col1) were averaged for the independent replicates. To measure the siRNA effect as the magnitude of the difference between an individual siRNA and the negative control (NTC siRNA), the previously described strictly standardized mean difference (SSMD) was applied (Zhang, 2007; Zhang *et al*, 2007). The following formula was used for the SSMD calculation: $SMD = \frac{\mu_1 - \mu_2}{\sqrt{\sigma 1^2 + \sigma 2^2}}$, where μ_1 is the normalized mean of all NTC siRNAs, μ_2 is the mean of the normalized values of siRNA for a given gene, σ1 is the variance of all normalized NTC siRNAs values and σ2 is the variance of all normalized values transfected with siRNA for a given gene.

Analysis of siRNA-mediated gene knockdown

For determining siRNA-mediated gene knockdown of ACTA2 and TGFBR1, cells were lysed and RNA prepared using the RNeasy Plus 96 Kit according to the manufacture's protocol. Two micrograms of total RNA was reverse transcribed with the High-Capacity cDNA Reverse Transcription Kit as described in the supplier's protocol. qPCR with 2 μl of cDNA and gene-specific TaqMan Assays was performed as suggested in the manual (TGFBR1: Hs00610320_m1; ACTA2: Hs00909449_m1, RNA-polymerase II amplification primers: GCAAGCGGATTCCATTTGG and TCTCAGGCCCGTAGTCATCCT, probe: AAGCACCGGACTCTTGCCTCACTTCATC). The gene-specific knockdown was calculated using the ($2^{-\Delta\Delta CT}$) compared to control siRNA treated cells (NTC) after normalization to RNA-polymerase II expression.

For determining the siRNA-mediated gene knockdown of LPXN, AQP3 and GLI4, the Luna Universal One-Step Reaction Mix Kit was used (New England Biolabs) with the following amplification primers: LPXN: CCACCACCTTCTAAAACGTCAG and CCCAAGCATTGAGTCC AGGG; AQP3: CTCGTGAGCCCTGGATCAAGC and AAAGCTGGTTG TCGGCGAAGT; GLI4, TCCCGCTCGGAAGGTCC and CTGAATGTC CCCTAGGGCTG; RPLP0: CTCTGCATTCTCGCTTCCTGGAG, CAGATG GATCAGCCAAGAAGG. Specifically, 10 ng of RNA was loaded into MicroAmp™ Fast Optical 96-Well Reaction Plate (Applied Biosystems) together with Luna Universal One-Step Reaction Mix, Luna WarmStart® RT Enzyme Mix (New England Biolabs), 1 μl of 10× SYBR™ Green I Nucleic Acid Gel Stain (Invitrogen), and 2 μl of 10 μM of forward and reverse primers in a final reaction of 20 μl. The plate was loaded in a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems) and ran using the following program as recommended: 10 min 55C, 1 min 95C, followed by 40 cycles of 10 s 95C, 1 min 60C. All reactions were run in duplicates, and the average C_T values between duplicates were used for the analysis. The gene-specific knockdown was calculated using the ($2^{-\Delta\Delta CT}$) compared to control siRNA treated cells (NTC) after normalization to RPLP0 expression.

Statistical analysis

Statistical analysis was performed using GraphPad Prism software, version 8.0.1. The significance level was set to 0.05, unless otherwise specified. An unpaired non-parametric *t*-test (Mann–Whitney test, GraphPad Prism software, version 8.0.1) was employed to compare the lung function (FEV1 and FEV1/FVC values) between control and COPD donors. For the analysis of the patient metadata of the three groups studied (control, COPD I and COPD II–IV), one-way ANOVA non-parametric unpaired test was used (Kruskal–Wallis test, GraphPad Prism software, version 8.0.1) followed by correction for multiple comparisons using Dunn's test. For experimental validation of the data, the number of replicates and the statistical test used are described in figure legends for each of the panels.

Data availability

The WGBS and RNA-seq data generated in this study have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG. The access to the patient data is controlled by the data access committee.

- RNA-seq data: European Genome-phenome Archive (EGA) EGAS 00001006602 (https://ega-archive.org/studies/EGAS00001006602)

- T-WGBS DNA methylation data: European Genome-phenome Archive (EGA) EGAS00001006603 (https://ega-archive.org/studies/EGAS00001006603)

Expanded View for this article is available online.

Acknowledgements

We would like to thank the Lung Biobank (Heidelberg, Germany) – a member of the Biomaterial bank Heidelberg (BMBH), the tissue bank of the National Center for Tumor Diseases (NCT) and the Biobank platform of the German Center for Lung Research (DZL; grant number: DZL3.0), as well as the Asklepios Biobank for Lung Diseases, member of the German Center for Lung Research (DZL) for providing Biomaterials and Data. We also thank Christa Stolp for her help with collecting primary material. We acknowledge the excellent sequencing service and helpful discussions from the Genomics core facility (GeneCore, EMBL, Germany) for RNA-seq and from NGX Bio (San Francisco, USA) for T-WGBS sequencing, as well as the support from the ZMBH imaging facility (Heidelberg, Germany) for immunofluorescence. We thank Morphisto GmbH (Frankfurt, Germany) for the excellent histological service. We also thank Christian Tidona (BioMed X Institute) and Markus Koester (Boehringer Ingelheim) for helpful project discussions and Pavlo Lutsik for discussions regarding methylation analysis. We thank the ENCODE Consortium for providing and the Bradly Bernstein Lab for producing the ChIP datasets of human fibroblasts used in this study. Finally, we acknowledge Marlene Ganslmeier for providing extracted EPIC data from (Clifford *et al*, 2018), Renjiao Li for help with the establishment of the bisulfite conversion protocol and Oliver Mücke for performing the MassARRAY assay. This study was supported by Boehringer Ingelheim. The work was partly funded by the School of Biosciences (Cardiff University) to RZJ, AT and STP, the German Center for Lung Research (DZL) to CP, TM, MS, FH, CPH, HW, AW, and MLP and Helmholtz Association to MLP. Asklepios Biobank as part of the DZL is partly funded by Bundesministerium für Bildung und Forschung (BMBF).

Author contributions

Uwe Schwartz: Conceptualization; data curation; formal analysis; validation; investigation; visualization; writing – original draft; writing – review and editing. **Maria Llamazares Prada:** Conceptualization; formal analysis; validation; investigation; visualization; methodology; writing – original draft; writing – review and editing. **Stephanie T Pohl:** Investigation; writing – review and editing. **Mandy Richter:** Investigation. **Raluca Tamas:** Investigation. **Michael Schuler:** Formal analysis; supervision; investigation; writing – original draft. **Corinna Keller:** Formal analysis; investigation. **Vedrana Mijosek:** Investigation. **Thomas Muley:** Resources. **Marc A Schneider:** Resources. **Karsten Quast:** Formal analysis. **Joschka Hey:** Formal analysis; visualization. **Claus P Heußel:** Formal analysis. **Arne Warth:** Investigation. **Hauke Winter:** Resources. **Özdemirhan Serçin:** Resources. **Harry Karmouty-Quintana:** Resources. **Soma SK Jyothula:** Resources. **Manish K Patel:** Resources. **Felix Herth:** Resources. **Ina Koch:** Resources. **Giuseppe Petrosino:** Resources; formal analysis. **Alexandru Titimeaua:** Investigation. **Balca R Mardin:** Resources. **Dieter Weichenhan:** Resources. **Tomasz P Jurkowski:** Resources. **Charles D Imbusch:** Resources; software. **Benedikt Brors:** Resources; software. **Vladimir Benes:** Resources. **Birgit Jung:** Supervision. **David Wyatt:** Supervision. **Heiko F Stahl:** Supervision. **Christoph Plass:** Resources; supervision; writing – review and editing. **Renata Z Jurkowska:** Conceptualization; formal analysis; supervision; funding acquisition; investigation; visualization; writing – original draft; project administration; writing – review and editing.

Disclosure and competing interests statement

During their employment at BioMed X Institute, RZJ, MLP, VM, US, RT, MR, SP, OS, and BM received research funding by Boehringer Ingelheim Pharma GmbH & Co KG. HS, BJ, MSchu, CK, KQ and DWy are employees of Boehringer Ingelheim Pharma GmbH & Co KG and receive compensation as such. TM received a research grant, non-financial support and has patent applications with Roche Diagnostics GmbH outside of the described work. CPH has stock ownership in GSK; received research funding from Siemens, Pfizer, MeVis, and Boehringer Ingelheim; consultation fees from Schering-Plough, Pfizer, Basilea, Boehringer Ingelheim, Novartis, Roche, Astellas, Gilead, MSD, Lilly Intermune, and Fresenius, and speaker fees from Gilead, Essex, Schering-Plough, AstraZeneca, Lilly, Roche, MSD, Pfizer, Bracco, MEDA Pharma, Intermune, Chiesi, Siemens, Covidien, Boehringer Ingelheim, Grifols, Novartis, Basilea, and Bayer, outside the submitted work. HW received consultation fees from Intuitive and Roche.

References

Acharya A, Baek ST, Huang G, Eskiocak B, Goetsch S, Sung CY, Banfi S, Sauer MF, Olsen GS, Duffield JS *et al* (2012) The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. *Development* 139: 2139–2149

Agusti A, Celli B, Faner R (2017) What does endotyping mean for treatment in chronic obstructive pulmonary disease? *Lancet* 390: 980–987

Akama T, Chun T-H (2018) Transcription factor 21 (TCF21) promotes proinflammatory interleukin 6 expression and extracellular matrix remodeling in visceral adipose stem cells. *J Biol Chem* 293: 6603–6610

Aumiller V, Strobel B, Romeike M, Schuler M, Stierstorfer BE, Kreuz S (2017) Comparative analysis of lysyl oxidase (like) family members in pulmonary fibrosis. *Sci Rep* 7: 149

Barnes PJ (2019a) Inflammatory endotypes in COPD. *Allergy* 74: 1249–1256

Barnes PJ (2019b) Small airway fibrosis in COPD. *Int J Biochem Cell Biol* 116: 105598

Barnes PJ, Burney PG, Silverman EK, Celli BR, Vestbo J, Wedzicha JA, Wouters EF (2015) Chronic obstructive pulmonary disease. *Nat Rev Dis Primers* 1: 15076

Barnes PJ, Baker J, Donnelly LE (2019) Cellular senescence as a mechanism and target in chronic lung diseases. *Am J Respir Crit Care Med* 200: 556–564

Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM *et al* (2002) Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 62: 2370–2377

Birnhuber A, Biasin V, Schnoegl D, Marsh LM, Kwapiszewska G (2019) Transcription factor Fra-2 and its emerging role in matrix deposition, proliferation and inflammation in chronic lung diseases. *Cell Signal* 64: 109408

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120

van den Bosch T, Kwiatkowski M, Bischoff R, Dekker FJ (2017) Targeting transcription factor lysine acetylation in inflammatory airway diseases. *Epigenomics* 9: 1013–1028

Braitsch CM, Combs MD, Quaggin SE, Yutzey KE (2012) Pod1/Tcf21 is regulated by retinoic acid signaling and inhibits differentiation of epicardium-derived cells into smooth muscle in the developing heart. *Dev Biol* 368: 345–357

Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D (2016) Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. *Respir Res* 17: 143

Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, Colicino E, Karrasch S, Wahl S, Kunze S *et al* (2018) Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. *Epigenetics* 13: 1039–1055

Casas-Recasens S, Noell G, Mendoza N, Lopez-Giraldo A, Garcia T, Guirao A, Agusti A, Faner R (2021) Lung DNA methylation in chronic obstructive pulmonary disease: relationship with smoking status and airflow limitation severity. *Am J Respir Crit Care Med* 203: 129–134

Chen J, Wang T, Zhou YC, Gao F, Zhang ZH, Xu H, Wang SL, Shen LZ (2014) Aquaporin 3 promotes epithelial-mesenchymal transition in gastric cancer. *J Exp Clin Cancer Res* 33: 38

Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Ziniti J, Laird NM *et al* (2014) Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2: 214–225

Clifford RL, Fishbane N, Patel J, MacIsaac JL, McEwen LM, Fisher AJ, Brandsma CA, Nair P, Kobor MS, Hackett TL *et al* (2018) Altered DNA methylation is associated with aberrant gene expression in parenchymal but not airway fibroblasts isolated from individuals with COPD. *Clin Epigenetics* 10: 32

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK *et al* (2018) The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46: D794–D801

Dessalle K, Narayanan V, Kyoh S, Mogas A, Halayko AJ, Nair P, Baglole CJ, Eidelman DH, Ludwig MS, Hamid Q (2016) Human bronchial and parenchymal fibroblasts display differences in basal inflammatory phenotype and response to IL-17A. *Clin Exp Allergy* 46: 945–956

Dierks S, von Hardenberg S, Schmidt T, Bremmer F, Burfeind P, Kaulfuss S (2015) Leupaxin stimulates adhesion and migration of prostate cancer cells through modulation of the phosphorylation status of the Actin-binding protein caldesmon. *Oncotarget* 6: 13591–13606

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21

Eferl R, Hasselblatt P, Rath M, Popper H, Zenz R, Komnenovic V, Idarraga MH, Kenner L, Wagner EF (2008) Development of pulmonary fibrosis through a pathway involving the transcription factor Fra-2/AP-1. *Proc Natl Acad Sci U S A* 105: 10525–10530

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 102: 15785–15790

Ernst J, Kellis M (2017) Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12: 2478–2492

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43–49

Esteller M (2008) Molecular origins of cancer: epigenetics in cancer. *N Engl J Med* 358: 1148–1159

Fernandez IE, Eickelberg O (2012) The impact of TGF-beta on lung fibrosis: from targeting to biomarkers. *Proc Am Thorac Soc* 9: 111–116

Garudadri S, Woodruff PG (2018) Targeting chronic obstructive pulmonary disease phenotypes, endotypes, and biomarkers. *Ann Am Thorac Soc* 15: S234–S238

GOLD (2021) The Global Strategy for Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease (updated 2021). https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf

Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, Krogh V, Tumino R, Sacerdote C, Panico S *et al* (2015) Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet* 24: 2349–2359

Hallgren O, Nihlberg K, Dahlback M, Bjermer L, Eriksson LT, Erjefalt JS, Lofdahl CG, Westergren-Thorsson G (2010) Altered fibroblast proteoglycan production in COPD. *Respir Res* 11: 55

Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG *et al* (2010) Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 42: 45–52

Hansen KD, Langmead B, Irizarry RA (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13: R83

Heinbockel L, Marwitz S, Schromm AB, Watz H, Kugler C, Ammerpohl O, Schnepf K, Rabe KF, Droemann D, Goldmann T (2018) Identification of novel target genes in human lung tissue involved in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 13: 2255–2259

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW *et al* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459: 108–112

Hey J, Paulsen M, Toth R, Weichenhan D, Butz S, Schatterny J, Liebers R, Lutsik P, Plass C, Mall MA (2021) Epigenetic reprogramming of airway macrophages promotes polarization and inflammation in muco-obstructive lung disease. *Nat Commun* 12: 6520

Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, Wain LV, Hall IP, Jackson VE, Wyss AB *et al* (2017) Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 49: 426–432

Holz O, Zuhlke I, Jaksztat E, Muller KC, Welker L, Nakashima M, Diemel KD, Branscheid D, Magnussen H, Jorres RA (2004) Lung fibroblasts from patients with emphysema show a reduced proliferation rate in culture. *Eur Respir J* 24: 575–579

Hou SY, Li YP, Wang JH, Yang SL, Wang Y, Wang Y, Kuang Y (2016) Aquaporin-3 inhibition reduces the growth of NSCLC cells induced by hypoxia. *Cell Physiol Biochem* 38: 129–140

Hovestadt V, Jones DTW, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz H-J *et al* (2014) Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* 510: 537–541

Huang YT, Zhou J, Shi S, Xu HY, Qu F, Zhang D, Chen YD, Yang J, Huang HF, Sheng JZ (2015) Identification of estrogen response element in Aquaporin-3 gene that mediates estrogen-induced cell migration and invasion in estrogen receptor-positive breast cancer. *Sci Rep* 5: 12484

Ito K, Ito M, Elliott WM, Cosio B, Caramori G, Kon OM, Barczyk A, Hayashi S, Adcock IM, Hogg JC *et al* (2005) Decreased histone deacetylase activity in chronic obstructive pulmonary disease. *N Engl J Med* 352: 1967–1976

Kachroo P, Morrow JD, Kho AT, Vyhlidal CA, Silverman EK, Weiss ST, Tantisira KG, Demeo DL (2020) Co-methylation analysis in lung tissue identifies pathways for fetal origins of COPD. *Eur Respir J* 56: 1902347

Kaulfuss S, Grzmil M, Hemmerlein B, Thelen P, Schweyer S, Neesen J, Bubendorf L, Glass AG, Jarry H, Auber B *et al* (2008) Leupaxin, a novel coactivator of the androgen receptor, is expressed in prostate cancer and

plays a role in adhesion and invasion of prostate carcinoma cells. *Mol Endocrinol* 22: 1606–1621

Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM (2015) Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. *Int J Genomics* 2015: 206937

Konigshoff M, Kneidinger N, Eickelberg O (2009) TGF-beta signaling in COPD: deciphering genetic and cellular susceptibilities for future therapeutic regimen. *Swiss Med Wkly* 139: 554–563

Kramer A, Green J, Pollard J Jr, Tugendreich S (2014) Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30: 523–530

Kulkarni T, O'Reilly P, Antony VB, Gaggar A, Thannickal VJ (2016) Matrix remodeling in pulmonary fibrosis and emphysema. *Am J Respir Cell Mol Biol* 54: 751–760

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25: 1754–1760

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930

Lim HJ, Weinheimer O, Wielputz MO, Dinkel J, Hielscher T, Gompelmann D, Kauczor HU, Heussel CP (2016) Fully automated pulmonary lobar segmentation: influence of different prototype software programs onto quantitative evaluation of chronic obstructive lung disease. *PLoS One* 11: e0151498

Liu X, Rowan SC, Liang J, Yao C, Huang G, Deng N, Xie T, Wu D, Wang Y, Burman A *et al* (2021) Categorization of lung mesenchymal cells in development and fibrosis. *iScience* 24: 102551

Llamazares-Prada M, Espinet E, Mijosek V, Schwartz U, Lutsik P, Tamas R, Richter M, Behrendt A, Pohl ST, Benz NP *et al* (2021) Versatile workflow for cell type-resolved transcriptional and epigenetic profiles from cryopreserved human lung. *JCI Insight* 6: e140443

Lotfi CFP, Passaia BS, Kremer JL (2021) Role of the bHLH transcription factor TCF21 in development and tumorigenesis. *Braz J Med Biol Res* 54: e10637

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Luo J, Liu X, Liu J, Jiang M, Luo M, Zhao J (2016) Activation of TGF-beta1 by AQP3-mediated H2O2 transport into fibroblasts of a bleomycin-induced mouse model of scleroderma. *J Invest Dermatol* 136: 2372–2379

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495–501

Medvedeva YA, Lennartsson A, Ehsani R, Kulakovskiy IV, Vorontsov IE, Panahandeh P, Khimulya G, Kasukawa T, Drabløs F (2015) EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* 2015: bav067

Miyazono K (2000) Positive and negative regulation of TGF-beta signaling. *J Cell Sci* 113: 1101–1109

Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Glass K *et al* (2016) DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* 11: 730–739

Muller KC, Welker L, Paasch K, Feindt B, Erpenbeck VJ, Hohlfeld JM, Krug N, Nakashima M, Branscheid D, Magnussen H *et al* (2006) Lung fibroblasts from patients with emphysema show markers of senescence in vitro. *Respir Res* 7: 32

Nobukuni S, Watanabe K, Inoue J, Wen FQ, Tamaru N, Yoshida M (2002) Cigarette smoke inhibits the growth of lung fibroblasts from patients with pulmonary emphysema. *Respirology* 7: 217–223

Noordhoek JA, Postma DS, Chong LL, Vos JT, Kauffman HF, Timens W, van Straaten JF (2003) Different proliferative capacity of lung fibroblasts obtained from control subjects and patients with emphysema. *Exp Lung Res* 29: 291–302

Ouyang N, Boyle AP (2020) TRACE: transcription factor footprinting using chromatin accessibility data and DNA sequence. *Genome Res* 30: 1040–1046

Park J, Ivey MJ, Deana Y, Riggsbee KL, Sörensen E, Schwabl V, Sjöberg C, Hjertberg T, Park GY, Swonger JM *et al* (2019) The Tcf21 lineage constitutes the lung lipofibroblast population. *Am J Physiol Lung Cell Mol Physiol* 316: L872–L885

Phan SH (2008) Biology of fibroblasts and myofibroblasts. *Proc Am Thorac Soc* 5: 334–337

Plantier L, Boczkowski J, Crestani B (2007) Defect of alveolar regeneration in pulmonary emphysema: role of lung fibroblasts. *Int J Chron Obstruct Pulmon Dis* 2: 463–469

Pohl ST, Prada ML, Espinet E, Jurkowska R (2023) Practical considerations for complex tissue dissociation for single-cell transcriptomics. *Methods Mol Biol* 2584: 371–387

Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W, Lomas DA *et al* (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med* 185: 373–381

Quaggin SE, Schwartz L, Cui SY, Igarashi P, Deimling J, Post M, Rossant J (1999) The basic-helix-loop-helix protein Pod1 is critically important for kidney and lung organogenesis. *Development* 126: 5771–5783

Rabe KF, Watz H (2017) Chronic obstructive pulmonary disease. *Lancet* 389: 1931–1940

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2010) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279–283

Ryu HM, Oh EJ, Park SH, Kim CD, Choi JY, Cho JH, Kim IS, Kwon TH, Chung HY, Yoo M *et al* (2012) Aquaporin 3 expression is up-regulated by TGF-beta1 in rat peritoneal mesothelial cells and plays a role in wound healing. *Am J Pathol* 181: 2047–2057

Safiri S, Carson-Chahhoud K, Noori M, Nejadghaderi SA, Sullman MJM, Ahmadian Heris J, Ansarin K, Mansournia MA, Collins GS, Kolahi AA *et al* (2022) Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the global burden of disease study 2019. *BMJ* 378: e069679

Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, Shrine N, Qiao D, Bartz TM, Kim DK *et al* (2019) Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet* 51: 494–505

Sayols S, Scherzinger D, Klein H (2016) dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-seq data. *BMC Bioinformatics* 17: 428

Shen YY, Cao RX, Liu W, Zhou YQ, Wu Y, Tan JJ, Jin M, Zhong J, Zhang QH, Liu JH *et al* (2017) Negative feedback loop between ZBTB7A and TGF-beta in breast cancer. *Oncol Lett* 14: 1403–1410

Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE *et al* (2011) Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 43: 1082–1090

Sood A, Petersen H, Blanchette CM, Meek P, Picchi MA, Belinsky SA, Tesfaigzi Y (2010) Wood smoke exposure and gene promoter methylation are

associated with increased risk for COPD in smokers. *Am J Respir Crit Care Med* 182: 1098–1104

Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS (2004) Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 31: 601–610

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, Ev N, Wirbelauer C, Oakeley EJ, Gaidatzis D *et al* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480: 490–495

Sundar IK, Yin Q, Baier BS, Yan L, Mazur W, Li D, Susiarjo M, Rahman I (2017) DNA methylation profiling in peripheral lung tissues of smokers and patients with COPD. *Clin Epigenetics* 9: 38

Szulakowski P, Crowther AJ, Jimenez LA, Donaldson K, Mayer R, Leonard TB, MacNee W, Drost EM (2006) The effect of smoking on the transcriptional regulation of lung inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 174: 41–50

Togo S, Holz O, Liu X, Sugiura H, Kamio K, Wang X, Kawasaki S, Ahn Y, Fredriksson K, Skold CM *et al* (2008) Lung fibroblast repair functions in patients with chronic obstructive pulmonary disease are altered by multiple mechanisms. *Am J Respir Crit Care Med* 178: 248–260

van der Vaart H, Postma DS, Timens W, ten Hacken NH (2004) Acute effects of cigarette smoke on inflammation and oxidative stress: a review. *Thorax* 59: 713–721

Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, Zhang M, Lonergan KM, Steiling K, Brown CJ *et al* (2014) DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 50: 912–922

Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP *et al* (2015) Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK biobank. *Lancet Respir Med* 3: 769–781

Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, Hokanson JE, Bowler RP, Crapo JD, DeMeo DL (2015) Smoking-associated site-specific differential methylation in buccal mucosa in the COPDGene study. *Am J Respir Cell Mol Biol* 53: 246–254

Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, Bähr M, Wolf S, Shendure J, Eils R *et al* (2013) Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc* 8: 2022–2032

Wehrens R, Kruisselbrink J (2018) Flexible Self-Organizing Maps in kohonen 3.0. *J Stat Softw* 87: 1–18

Weigle S, Martin E, Voegtle A, Wahl B, Schuler M (2019) Primary cell-based phenotypic assays to pharmacologically and genetically study fibrotic diseases *in vitro*. *J Biol Methods* 6: 115

Whyte WA, Orlando David A, Hnisz D, Abraham Brian J, Lin Charles Y, Kagey Michael H, Rahl Peter B, Lee Tong I, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307–319

Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST *et al* (2009) A genome-wide association study of pulmonary function measures in the Framingham heart study. *PLoS Genet* 5: e1000429

Woodruff PG, Agusti A, Roche N, Singh D, Martinez FJ (2015) Current concepts in targeting chronic obstructive pulmonary disease pharmacotherapy: making progress towards personalised management. *Lancet* 385: 1789–1798

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L *et al* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2: 100141

Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, Latourelle JC, Smith AV, Bartz TM, Feitosa MF *et al* (2018) Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* 9: 2976

Xiong G, Chen X, Zhang Q, Fang Y, Chen W, Li C, Zhang J (2017) RNA interference influenced the proliferation and invasion of XWLC-05 lung cancer cells through inhibiting aquaporin 3. *Biochem Biophys Res Commun* 485: 627–634

Xu H, Xu Y, Zhang W, Shen L, Yang L, Xu Z (2011) Aquaporin-3 positively regulates matrix metalloproteinases via PI3K/AKT signal pathway in human gastric carcinoma SGC7901 cells. *J Exp Clin Cancer Res* 30: 86

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F *et al* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356: eaaj2239

Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A *et al* (2015) Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* 11: e1004898

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A *et al* (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 8: e63812

Zhang XD (2007) A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* 89: 552–561

Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, Crackower MA, Holder DJ, Heyse JF, Strulovici B (2007) The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J Biomol Screen* 12: 497–509

Zhang J, Wu L, Qu JM, Bai CX, Merrilees MJ, Black PN (2012) Pro-inflammatory phenotype of COPD fibroblasts not compatible with repair in COPD lung. *J Cell Mol Med* 16: 1522–1532