

1 **High-resolution transcriptomic and epigenetic profiling identifies novel regulators of COPD  
2 phenotypes in human lung fibroblasts**

3 Uwe Schwartz<sup>1,2</sup>, Maria Llamazares Prada<sup>1,3</sup>, Stephanie T. Pohl<sup>1,4</sup>, Mandy Richter<sup>1</sup>, Raluca Tamas<sup>1</sup>,  
4 Michael Schuler<sup>5</sup>, Corinna Keller<sup>6</sup>, Vedrana Mijosek<sup>1</sup>, Thomas Muley<sup>7,8</sup>, Marc A. Schneider<sup>7,8</sup>, Karsten  
5 Quast<sup>9</sup>, Joschka Hey<sup>3,10</sup>, Claus P. Heußen<sup>8,11,12</sup>, Arne Warth<sup>7,8,13/</sup>, Hauke Winter<sup>8,14</sup>, Özdemirhan  
6 Serçin<sup>1</sup>, Harry Karmouty-Quintana<sup>15</sup>, Felix Herth<sup>7,8,16</sup>, Ina Koch<sup>17</sup>, Giuseppe Petrosino<sup>1^\*</sup>, Balca R.  
7 Mardin<sup>1</sup>, Dieter Weichenhan<sup>3</sup>, Tomasz P. Jurkowski<sup>18</sup>, Charles D. Imbusch<sup>19</sup>, Benedikt Brors<sup>19</sup>,  
8 Vladimir Benes<sup>20</sup>, Brigit Jung<sup>6</sup>, David Wyatt<sup>21</sup>, Heiko Stahl<sup>6</sup>, Christoph Plass<sup>3</sup> and Renata Z.  
9 Jurkowska<sup>1,4\*</sup>

10 *US and MLP contributed equally to this work*

11

12 <sup>1</sup> BioMed X Institute, Heidelberg, Germany

13 <sup>2</sup> NGS Analysis Center Biology and Pre-Clinical Medicine, University of Regensburg, Regensburg,  
14 Germany

15 <sup>3</sup> Division of Cancer Epigenomics, German Cancer Research Center (DKFZ) and Translational Lung  
16 Research Center, Member of the German Center for Lung Research (DZL), Heidelberg, Germany,  
17 Heidelberg, Germany

18 <sup>4</sup> Division of Biomedicine, School of Biosciences, Cardiff University, Cardiff, UK

19 <sup>5</sup> Drug Discovery Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Germany

20 <sup>6</sup> Immunology and Respiratory Disease Research, Boehringer Ingelheim Pharma GmbH & Co. KG –  
21 Biberach, Germany

22 <sup>7</sup> Translational Research Unit, Thoraxklinik, University Hospital Heidelberg, University of Regensburg  
23 Germany

24 <sup>8</sup> Translational Lung Research Center (TLRC), Member of the German Center for Lung Research  
25 (DZL), Heidelberg, Germany

26 <sup>9</sup> Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG,  
27 Germany

- 28 <sup>10</sup> Ruprecht Karl University of Heidelberg, Heidelberg, Germany
- 29 <sup>11</sup> Diagnostic and Interventional Radiology with Nuclear Medicine, Thoraxklinik, University of  
30 Heidelberg, Heidelberg, Germany
- 31 <sup>12</sup> Diagnostic and Interventional Radiology, University Hospital Heidelberg, Heidelberg, Germany
- 32 <sup>13</sup> Pathological Institute, University Hospital Heidelberg, Heidelberg, Germany
- 33 <sup>14</sup> Dept. of Surgery, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany
- 34 <sup>15</sup> Department of Biochemistry and Molecular Biology, McGovern Medical School, University of Texas  
35 Health Science Center at Houston, Houston, USA
- 36 <sup>16</sup> Department of Pneumology and Critical Care Medicine and Translational Research Unit,  
37 Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany
- 38 <sup>17</sup> Asklepios Biobank for Lung Diseases, Department of Thoracic Surgery, Asklepios Fachkliniken  
39 München-Gauting, German Center for Lung Research (DZL), Munich, Germany
- 40 <sup>18</sup> Division of Molecular Biology, School of Biosciences, Cardiff, UK
- 41 <sup>19</sup> Division of Applied Bioinformatics, German Cancer Research Center, Germany
- 42 <sup>20</sup> Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
- 43 <sup>21</sup> Biotherapeutics Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG, Germany
- 44 <sup>1</sup> Current address: Institute of Pathology, Cytopathology and Molecular Pathology MVZ UEGP  
45 Gießen/Wetzlar/Limburg/Bad Hersfeld, Germany
- 46 <sup>1</sup> Current address: Institute of Molecular Biology (IMB), Mainz, Germany
- 47 \* Correspondence should be addressed to [jurkowskar@cardiff.ac.uk](mailto:jurkowskar@cardiff.ac.uk), telephone +44 29 2087 9067
- 48
- 49 **Keywords:** COPD, human lung fibroblasts, epigenome, DNA methylation, WGBS, RNA sequencing,
- 50 phenotypic assays, biomarkers
- 51
- 52 **Running title:**
- 53 Omics profiling identifies new regulators of COPD.

54 **Abstract**

55 Patients with chronic obstructive pulmonary disease (COPD) are still waiting for curative treatments.  
56 Considering the environmental cause of COPD (e.g., cigarette smoke) and disease phenotypes,  
57 including stem-cell senescence and impaired differentiation, we hypothesized that COPD will be  
58 associated with altered epigenetic signaling in lung cells. We generated genome-wide DNA  
59 methylation maps at single CpG resolution of primary human lung fibroblasts (HLFs) isolated from  
60 distal parenchyma of ex-smoker controls and COPD patients, with both mild and severe disease. The  
61 epigenetic landscape is markedly changed in lung fibroblasts across COPD stages, with DNA  
62 methylation changes occurring predominantly in regulatory regions, including promoters and  
63 enhancers. RNA sequencing of matched fibroblasts demonstrated dysregulation of genes involved in  
64 proliferation, DNA repair, and extracellular matrix organization. Notably, we identified epigenetic and  
65 transcriptional dysregulation already in mild COPD patients, providing unique insights into early  
66 disease. Integration of profiling data identified 110 candidate regulators of disease phenotypes,  
67 including epigenetic factors. Using phenotypic screens, we verified the regulator capacity of multiple  
68 candidates and linked them to repair processes in the human lung.

69 Our study provides first integrative high-resolution epigenetic and transcriptomic maps of human lung  
70 fibroblasts across stages of COPD. We reveal novel transcriptomic and epigenetic signatures  
71 associated with COPD onset and progression and identify new candidate regulators involved in the  
72 pathogenesis of chronic respiratory diseases. The presence of various epigenetic factors among the  
73 candidates demonstrates that epigenetic regulation in COPD is an exciting research field that holds  
74 promise for novel therapeutic avenues for patients.

75

76

77

78

79 **Introduction**

80 Chronic obstructive pulmonary disease (COPD) is a prevalent, smoke-related disease characterized  
81 by persistent inflammation of the lung epithelium, irreversible airway remodeling, and destruction of  
82 the alveolar tissue (emphysema) (Barnes et al. 2015; Rabe and Watz 2017; GOLD 2021). COPD-  
83 related mortality is increasing and it already affects more than 3 million people worldwide every year  
84 (WHO 2019). However, despite its prevalence, there is currently no treatment to halt progression of  
85 COPD, as none of the existing drugs can modify the long-term decline in lung function. COPD is a  
86 heterogeneous disease with variable clinical manifestations and responses to therapy where patient  
87 stratification remains challenging (Woodruff et al. 2015; Agusti et al. 2017; Garudadri and Woodruff  
88 2018; Barnes 2019a).

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

102 Numerous genetic loci have been associated with COPD and lung function (Wilk et al. 2009; Hancock  
103 et al. 2010; Soler Artigas et al. 2011; Cho et al. 2014; Wain et al. 2015; Hobbs et al. 2017; Wyss et al.  
104 2018; Sakornsakolpat et al. 2019), yet they explain only a small fraction of COPD risk. Transcriptional

105 programs in cells are regulated by a landscape of epigenetic modifications that modulate chromatin  
106 structure and thereby control gene expression. Smoking is the most prominent risk factor for COPD,  
107 and its impact on epigenetic landscape remodeling is well established (Belinsky et al. 2002; Chen et  
108 al. 2013; Zeilinger et al. 2013; Wan et al. 2015). Earlier studies also provided strong evidence for the  
109 association of dysregulated DNA methylation and COPD in blood (Qiu et al. 2012; Busch et al. 2016;  
110 Carmona et al. 2018), sputum (Sood et al. 2010), oral mucosa (Wan et al. 2015), lung tissue (Sood et  
111 al. 2010; Yoo et al. 2015; Morrow et al. 2016; Sundar et al. 2017), bronchial brushings (Vucic et al.  
112 2014), fibroblasts (Clifford et al. 2018) and macrophages from a mouse model of muco-obstructive  
113 disease (Hey et al. 2021). Notably, DNA methylation changes were associated with altered expression  
114 of genes and pathways important to COPD pathology. However, these studies were either performed  
115 on material encompassing mixed cell populations or/and used low-resolution approaches and could  
116 therefore not resolve differential gene expression and methylation changes caused by a specific cell  
117 type during COPD development and progression. To date, the full epigenomic landscape of purified  
118 COPD cells remains uncharted and thus, the precise epigenetic changes and their contribution to  
119 altered transcriptional patterns in COPD are still unknown.

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

128

129

130

131 **Results**

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

133 To assess the extent of epigenetic remodeling in COPD genome-wide, we generated high-resolution  
134 DNA methylomes of primary lung fibroblasts isolated from well-matched control donors (no COPD,  
135 n=3) and patients with established COPD (stage II-IV according to Global Initiative for Chronic  
136 Obstructive Lung Disease (GOLD) (GOLD 2021), n=5), which can be classified based on lung function  
137 (**Figure 1A, Suppl. Fig. 1A-B**). Isolated cells displayed a typical fibroblast morphology, and their high  
138 purity was confirmed by fluorescence-activated cell sorting (FACS) and immunofluorescence (IF)  
139 staining (**Suppl. Fig. 1C-D**). We used tagmentation-based whole-genome bisulfite sequencing (T-  
140 WGBS) for DNA methylation profiling, allowing genome-scale assessment of DNA methylation at  
141 single CpG resolution from low cell numbers (Wang et al. 2013) (**Figure 1B**).

142 We observed no significant differences in the global levels of methylated cytosines between COPD  
143 and control samples (**Figure 1C**), suggesting that, in contrast to cancer cells (Esteller 2008), COPD is  
144 not associated with a global drop or gain of methylation in lung fibroblasts. When we looked at local  
145 aberrant DNA methylation between no COPD and COPD (II-IV), we found numerous distinct regions.  
146 Using CpG sites covered at least 4x in all samples and a  $\geq 10\%$  methylation-difference cutoff, we  
147 identified 6,279 differentially methylated regions (DMRs) (p-value < 0.1, see Methods part for details)  
148 (**Figure 1D-F, Table 1**), indicating widespread methylation alterations in primary human lung  
149 fibroblasts of COPD patients. The distribution of methylation differences across DMRs demonstrated a  
150 more prominent loss of methylation, suggestive of a more permissive chromatin state in COPD (58%  
151 of DMRs, 3,615 hypomethylated regions, **Figure 1D**). The remaining 2,664 regions showed increased  
152 methylation (42% of the DMRs, 2,664 hypermethylated regions, **Figure 1D**). Called DMRs contained 8  
153 CpG sites on average and showed a median size of 479 bp (**Suppl. Fig. 1E-F**), indicating that specific  
154 regions are altered.

155 To investigate whether DNA methylation changes occur early in COPD development and identify  
156 alterations associated with disease progression, we additionally performed T-WGBS on fibroblasts  
157 isolated from mild COPD patients (GOLD I, n=3), with noticeable obstruction (FEV<sub>1</sub>/FVC<70%) but  
158 preserved FEV<sub>1</sub> (>80%, **Suppl. Fig. 1A**), and integrated them into the analysis. Notably, as  
159 demonstrated by the principal component analysis (PCA) on all 6,279 previously identified DMRs,  
160 COPD (I) samples grouped with the COPD (II-IV) samples on the first principal component (**Figure**  
161 **1G**), confirming the COPD specific DMR calling using independent test samples (COPD (I) samples  
162 were not used for the initial DMR selection). Furthermore, the COPD (I) samples were separated from  
163 COPD (II-IV) on the second principal component, indicating that DNA methylation data might provide  
164 information about disease progression (**Figure 1G**). Consistent with the PCA, hierarchical clustering  
165 using the identified DMRs showed that COPD (I) samples group with COPD (II-IV) samples,  
166 demonstrating that COPD-associated methylation changes occur early in the disease pathogenesis  
167 (**Suppl. Fig. 1G**). This important discovery suggests that DNA methylation might provide a sensitive  
168 biomarker to separate early COPD patients from smokers with preserved lung function.

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

178 To shed light on the cellular processes and pathways affected by aberrant DNA methylation changes  
179 in COPD, we linked DMRs to the nearest gene and performed gene ontology (GO) enrichment  
180 analysis using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010). As

181 hypermethylated DMRs did not reveal any significant enrichment in GREAT, we focused on the  
182 hypomethylated DMRs. Among the top categories, we identified cellular response to hypoxia,  
183 regulation of focal adhesion assembly, negative regulation of transforming growth factor beta (TGF $\beta$ )  
184 receptor signaling, and epithelial to mesenchymal transition (**Suppl. Fig. 1H**). These biological  
185 processes are relevant for COPD development and progression (Konigshoff et al. 2009; Barnes et al.  
186 2015; Rabe and Watz 2017; Barnes et al. 2019), indicating that DNA methylation changes occur near  
187 genes that are critically involved in COPD pathogenesis and might therefore contribute to disease  
188 phenotypes in lung fibroblasts. Specific examples include hypomethylation of genes implicated in  
189 negative regulation of TGF $\beta$  receptor signaling (e.g., SMAD6, SMAD7, VASN, SKI, PMEPA1, **Table 1**)  
190 (Miyazono 2000), potentially explaining the reduced response to TGF $\beta$  and decreased repair capacity  
191 of lung fibroblasts observed in emphysema (Togo et al. 2008).

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

195

## 196 **DNA methylation changes occur at regulatory regions in COPD lung fibroblasts**

197 Methylome analysis identified genome-wide changes of DNA methylation in primary HLFs of COPD  
198 patients. To better understand the functional role of aberrant methylation in COPD, we investigated  
199 the distribution of DMRs across the genome. We observed a different distribution of regions displaying  
200 loss or gain of methylation, with hypomethylated DMRs predominately located in intronic sequences  
201 and hypermethylated DMRs preferentially found in intergenic regions (**Figure 2A**). Notably, both types  
202 of DMRs were overrepresented at regulatory and gene coding sequences compared to the genomic  
203 background (**Figure 2B**), with a stronger enrichment for hypomethylated DMRs. Thus,  
204 hypomethylated DMRs are located 4 times more often at promoter sequences than expected by  
205 chance (**Figure 2B**). Further intersection with known regulatory genomic features annotated by the

206 ENCODE Chromatin States (Ernst et al. 2011) revealed a strong enrichment of hypomethylated DMRs  
207 in active promoters and enhancers, indicating their potential regulatory role (**Figure 2C-D, Table 1**).  
208 The significant association with active enhancer elements was confirmed by the local increase of  
209 enhancer defining chromatin marks (H3K4me1 and H3K27ac) (Heintzman et al. 2009; Rada-Iglesias  
210 et al. 2010) in the center of the hypomethylated DMRs (**Figure 2D-E**). Conversely, the  
211 hypermethylated DMRs were overrepresented at Polycomb-repressed regions, defined by the  
212 presence of H3K27me3 (**Figure 2C and 2E**).

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

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

226

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

228 The hypomethylated regions identified in WGBS data might reflect the binding of transcription factors  
229 and can be therefore used for foot-printing their binding sites (Stadler et al. 2011). We observed a  
230 significant enrichment of binding motifs of several transcription factors in the hypomethylated DMRs at

231 strong enhancers, with the highest enrichment of TCF21 motif in the early DMR cluster (cluster 1, p-  
232 value:  $1 \times 10^{-17}$ ) and FOSL2/FRA2 in the progressive DMR cluster (cluster 3, p-value:  $1 \times 10^{-8}$ ) (**Figure**  
233 **2G**). TCF21 mediates fibroblast fate specification in multiple organs and is required for alveolar  
234 development (Quaggin et al. 1999). In turn, FOSL2/FRA2 is a known regulator of wound repair and  
235 TGF $\beta$ -mediated fibrosis (Eferl et al. 2008). Our data establish TCF21 and FOSL2/FRA2 as potential  
236 mediators of aberrant epigenetic changes at strong enhancers in COPD fibroblasts.

237 Since DNA methylation can also directly interfere with the binding of transcriptional regulators to DNA  
238 (Yin et al. 2017), we performed motif analysis in the identified hypo- and hypermethylated DMRs to  
239 identify transcription factors reported to change their binding affinity upon methylation of their motifs  
240 (Yin et al. 2017). At hypomethylated DMRs, which are overrepresented at regulatory sites, numerous  
241 methylation-sensitive transcription-factor motifs were significantly enriched, suggesting either  
242 increased (methyl-minus, red dots) or attenuated (methyl-plus, blue triangles) DNA binding in COPD  
243 (**Figure 2H**). Among the transcription factors exhibiting higher binding affinities towards methylated  
244 DNA (methyl-plus, blue triangles), we identified motifs of nuclear receptors, known regulators of  
245 cellular homeostasis, development, and metabolism (**Figure 2H**). Our data suggest that their DNA  
246 binding might be abrogated at regulatory sites in COPD due to loss of methylation. A few motifs of  
247 methylation-sensitive transcription factors were enriched in the hypermethylated DMRs, consistent  
248 with their location in repressive regions of the genome. The strongest enrichment was observed for  
249 ZBTB7A, a known repressor associated with the TGF $\beta$  signaling pathway (Shen et al. 2017). ZBTB7A  
250 preferentially binds unmethylated DNA (Yin et al. 2017) (methyl-minus, **Figure 2H**) indicating that its  
251 DNA binding in COPD might be hindered due to motif hypermethylation.

252 In summary, our data link aberrant DNA methylation in COPD fibroblasts to imbalanced transcription  
253 factor binding and provides insights into potentially disturbed regulatory networks in COPD.

254

255 **Gene expression changes accompany epigenetic modifications in COPD**

256 Our genome-wide DNA methylation analysis identified methylation changes at promoter and enhancer  
257 regions, suggesting that DMRs may have regulatory effects on gene expression. To assess whether  
258 epigenetic changes are associated with gene expression changes in COPD, we performed RNA-seq  
259 analysis on fibroblast samples matching those used for T-WGBS. This analysis identified 333 up-  
260 regulated and 287 down-regulated genes between no COPD (n=3) and COPD (II-IV) (n=5; FDR <  
261 0.05 and  $|\log_2(\text{fold-change})| > 0.5$ , **Figure 3A**, **Table 2**), including several long non-coding RNAs  
262 (**Figure 3A** orange labels, **Suppl. Fig. 3A**), providing a transcriptional signature of COPD. Enrichment  
263 analysis of the differentially expressed genes revealed that genes up-regulated in COPD are involved  
264 in cell cycle regulation, DNA replication and DNA repair (**Suppl. Fig. 3B**). In turn, down-regulated  
265 genes are associated with extracellular matrix (ECM) organization and cholesterol biosynthesis  
266 (**Suppl. Fig. 3C**).

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

282 To assess the extent of epigenetic remodeling in COPD, we analyzed the expression changes of  
283 epigenetic enzymes and readers(Medvedeva et al. 2015). 38 epigenetic factors were differentially  
284 expressed, with the majority (76%) showing up-regulation in COPD (**Table 2**). Examples of  
285 dysregulated epigenetic players include histone methyltransferases (e.g., SETD1B, SUV39H2,  
286 KMT2D, EZH2), histone demethylases (KDM6B) and chromatin remodeling factors (e.g., CHAF1A,  
287 ATAD2 CHAF1B), indicating that in addition to DNA methylation and histone acetylation (Ito et al.  
288 2005; Szulakowski et al. 2006) other epigenetic layers may also be dysregulated in COPD.

289

290 **Integrative data analysis reveals epigenetically regulated genes in COPD fibroblasts**

291 To further dissect the association between alterations in DNA methylation and changes in gene  
292 expression, we assigned DMRs to genes in their proximity. In total, we detected 4,059 genes  
293 associated with at least one DMR (in total 4,424 DMRs) within 4 kb upstream of the transcriptional  
294 start site (TSS) to 4 kb downstream of the transcriptional termination site (TTS) (**Figure 4A**). About  
295 45% of the gene-associated DMRs are located close to the TSS, mainly in the promoter and first  
296 intron (**Figure 4B, Suppl. Fig. 4A-B**). To further decipher the interplay between expression and  
297 methylation changes in COPD, we focused our analysis on DMRs within 4 kb surrounding the TSS  
298 (**Figure 4C**). We observed an overrepresentation at differentially expressed genes compared to genes  
299 whose expression is not significantly changed in COPD (Fisher's exact test: p-value =  $1.3 \times 10^{-7}$ ,  
300 **Suppl. Fig. 4C**). In total, 77 differentially expressed genes were associated with at least one DMR  
301 (**Figure 4C**), which was mainly located in the promoter or downstream of the TSS (**Suppl. Fig. 4B**).  
302 Spearman correlation between DNA methylation and expression change was calculated to assess  
303 how epigenetic variations are related to transcriptional differences. In contrast to unchanged genes  
304 which exhibited the expected normal distribution (**Figure 4D**, blue line), differentially expressed genes  
305 displayed a bimodal curve with enrichment at high positive and negative correlation rates, suggesting  
306 that some genes might be dysregulated by aberrant methylation in COPD (**Figure 4D**, red line).  
307 Examples of genes showing correlated DNA methylation and gene expression changes include

308 UMPS, STEAP3, GABRR1, GLI4, AQP3, and LPXN (**Figure 4E, Suppl. Fig. 4D**), suggesting potential  
309 regulation of their expression by DNA methylation.

310

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

312 Integration of DNA methylation and gene expression data, together with upstream regulator analysis  
313 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  
314 candidate regulators of the observed epigenetic and transcriptional changes in COPD HLFs. Overall,  
315 110 candidates were manually selected for functional validation using phenotypic screens (**Suppl. Fig.**  
316 **5A, Table 3**), 78% of them were not linked to COPD before. To determine the function of the selected  
317 candidates in key fibroblast processes related to COPD, high-content image-based phenotypic assays  
318 using small-interfering RNAs (siRNA)-mediated gene knockdown (KD) were carried out in primary  
319 human lung fibroblasts isolated from two normal healthy (NHLFs) and three COPD (DHLFs) donors  
320 (**Figure 5A**).

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

333 The technical performance of the assays was demonstrated by the robust effect of the siRNAs  
334 targeting assay controls. As expected, siRNA targeting TGF $\beta$  receptor 1 (TGF $\beta$ R1) showed a strong  
335 impact on the FMT assay (on both  $\alpha$ SMA and col1 readouts) (**Figure 5B**). Similarly, siRNA against  
336 ACTA2 (which encodes  $\alpha$ SMA) showed a specific effect only in the  $\alpha$ SMA readout. Furthermore, the  
337 strong correlation between the results obtained from different donors among normal fibroblasts (e.g.,  
338 R=0.95, p-value:  $3.4 \times 10^{-64}$  for  $\alpha$ SMA) and fibroblasts derived from COPD patients (e.g., R=0.91, p-  
339 value:  $2.7 \times 10^{-48}$  for  $\alpha$ SMA) confirmed the robustness of the assays (**Figure 5D** and **Suppl. Fig. 5B**).  
  
340 To evaluate the differences between non-targeting siRNA controls (NTC) and gene-targeted  
341 knockdowns, we calculated strictly standardized mean differences (SSMD) (Zhang 2007; Zhang et al.  
342 2007). Overall, a high hit-rate was observed, as 61 out of the selected 110 candidates showed an  
343 effect in at least one assay after applying the strict cutoff of  $|3 \text{ SSMD}|$  and 87 when using a SSMD  
344 cutoff of  $|2|$ , demonstrating the power of multimodal analysis in identifying candidates with regulatory  
345 potential (**Figure 5C, 5E, Suppl. Fig. 5C, Table 4**). Three genes, leupaxin (LPX), aquaporin 3 (AQP3)  
346 and GLI4, showed strong effects in all three readouts in both normal and diseased cells, indicating  
347 their critical function in fibroblast biology (**Figure 5B-C and 5E**). Among the positive hits, multiple  
348 epigenetic factors were also present (**Figure 5C, Suppl. Fig. 5C, Table 4**). For example, we observed  
349 strong effects on lung fibroblast proliferation and differentiation upon targeted knockdown of different  
350 epigenetic enzymes (DNA methyltransferase DNMT3B, histone methyltransferases KMT2A, MKT2B,  
351 KMT3C, SETD1B, and EZH2, histone acetyltransferase EP300), chromatin remodeling factors  
352 (CHAF1A and CHAF1B) as well as epigenetics readers (BAZ2, CBX3), identifying these factors as key  
353 regulators of fibroblasts and COPD phenotypes. The imbalance of histone acetyltransferase (HAT)  
354 and deacetylase (HDAC) activities has previously been linked to COPD (Ito et al. 2005), providing a  
355 scientific basis for the potential use of bromodomain (BET) and HDAC inhibitors in COPD (van den  
356 Bosch et al. 2017), however the potential of targeting other dysregulated epigenetic activities in COPD  
357 remains to be explored.

358 Interestingly, we observed potential disease-specific effects for some of the tested candidates that  
359 were preserved between different donors. Here, the effect after siRNA-mediated gene knockdown  
360 varied between normal and diseased fibroblasts (**Figure 5F, Suppl. Fig. 5D**). For example,  
361 knockdown of CHAF1A increased, while knockdown of CHAF1B reduced the expression of  $\alpha$ SMA,  
362 respectively, and both effects were stronger in diseased fibroblasts compared to normal cells. In turn,  
363 knockdown of AQP3 in normal fibroblasts had a larger effect on  $\alpha$ SMA levels upon TGF $\beta$ 1 stimulation  
364 compared to diseased cells (**Figure 5F**). All 3 genes were dysregulated in our RNA-seq in COPD  
365 (CHAF1A and CHAF1B were upregulated, whereas AQP3 was downregulated in COPD, **Table 2**),  
366 indicating that their dysregulation may be linked to COPD phenotypes in fibroblasts.  
  
367 The cell-based assays in primary normal and COPD fibroblasts confirmed the functional role of  
368 numerous candidates identified from profiling data, indicating that integrating genome-wide epigenetic  
369 and transcriptomic profiling of purified normal and diseased human lung cells is a powerful approach  
370 for the identification of novel regulators of disease phenotypes. In addition, the presence of various  
371 epigenetic factors among the positive hits demonstrates that epigenetic regulation in COPD is an  
372 exciting research field that should be explored in-depth, as it holds promise for novel therapeutic  
373 avenues for patients with COPD.

374

375 **Discussion**

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

379 Earlier studies using various patient material consisting of mixed-cell populations provided evidence of  
380 dysregulated DNA methylation patterns in COPD and identified CpG sites and pathways associated  
381 with smoking and COPD (Sood et al. 2010; Qiu et al. 2012; Vucic et al. 2014; Wan et al. 2015; Yoo et  
382 al. 2015; Busch et al. 2016; Morrow et al. 2016; Sundar et al. 2017; Carmona et al. 2018). Two recent  
383 publications also suggested that DNA methylation changes may originate in early life (Kachroo et al.  
384 2020) and be linked to the severity of airflow limitation (Casas-Recasens et al. 2021). However, they  
385 all used low-resolution approaches, covering a representation of the genome only, mostly gene  
386 promoters. Hence, the full epigenomic landscape of COPD cells remains uncharted. Overall, there has  
387 been a limited consistency between different studies, likely coming from the cellular heterogeneity of  
388 the starting material, diverse donor selection criteria and different statistical models used. The  
389 dissection of cell-type specific mechanisms associated with COPD requires epigenetic profiling of  
390 defined cell populations. Only one study investigated DNA methylation changes in COPD patients with  
391 cell-type resolution (Clifford et al. 2018). Using Illumina 450K BeadChip Array (focusing on gene  
392 promoters), Clifford *et al.* identified 887 and 44 differentially methylated CpG sites in parenchymal and  
393 airway fibroblasts of COPD patients, respectively (Clifford et al. 2018). Our study, providing a much  
394 higher resolution of previously unexplored regions (e.g., enhancers) significantly extends these  
395 observations and demonstrates pronounced, genome-wide DNA methylation and gene expression  
396 changes in parenchymal fibroblasts in COPD, in both mild and severe disease.

397 Little is known about the correlation of DNA methylation with disease severity. Methylation changes in  
398 13 genes have been identified in the lung tissue of COPD GOLD I and II patients compared to non-  
399 smoker controls (Casas-Recasens et al. 2021). However, it is unclear whether they represent  
400 smoking- or COPD-related changes, as ex-smoker controls were not investigated in this study (Casas-

401 Recasens et al. 2021). Our data reveal that genome-wide DNA methylation changes are present in  
402 lung fibroblasts of COPD (I) patients compared to controls with matched smoking status and history  
403 (all ex-smokers), demonstrating that epigenetic changes occur early in disease development. Notably,  
404 COPD (I) samples clustered with COPD (II-IV) rather than no COPD samples, indicating that DNA  
405 methylation may provide a sensitive biomarker for early disease detection. This hypothesis awaits  
406 further validation in larger patient cohorts.

407 Currently, it is unclear how altered DNA methylation patterns in COPD translate into biological effects  
408 in COPD fibroblasts. DNA methylation in regulatory regions can modulate the binding of transcriptional  
409 factors to DNA (Stadler et al. 2011), hence, methylation profiling allows identifying transcriptional  
410 regulators potentially mediating the epigenetic alterations observed. We detected a significant  
411 enrichment of binding sites for TCF21 and FOSL2/FRA2 transcription factors in the DMRs overlapping  
412 with strong enhancers in COPD. TCF21 is a mesenchyme-specific basic helix-loop-helix transcription  
413 factor regulating multiple processes, including proliferation, extracellular matrix assembly, as well as  
414 secretion of pro-inflammatory mediators (Akama and Chun 2018). It is required for lung development  
415 in mice, mesenchymal-epithelial crosstalk (Quaggin et al. 1999) and specification of fibroblast cell fate  
416 in different organs (Acharya et al. 2012; Braitsch et al. 2012). Recently, TCF21 has been identified as  
417 a specific marker of lipofibroblasts in mouse (Park et al. 2019) and human lung (Liu et al. 2021), a  
418 subpopulation of fibroblasts essential for alveolar niche homeostasis and repair. Despite its central  
419 roles in alveolar development and maintenance, TCF21 function in human lung fibroblasts is largely  
420 unknown. Our phenotypic screens demonstrate that TCF21 is required for lung fibroblast proliferation  
421 and differentiation upon TGF $\beta$ 1 stimulation, providing first insights into its molecular function. Notably,  
422 the effects of TCF21 on proliferation and  $\alpha$ SMA were stronger in diseased cells, indicating that COPD  
423 cells may be more sensitive to TCF21 loss than healthy lung fibroblasts. Consistent with the  
424 documented role of TCF21 in regulating cell proliferation in cancer (Lotfi et al. 2021), its differential  
425 binding at strong enhancers in COPD fibroblasts could provide a molecular mechanism for the  
426 differential expression of DNA replication genes observed in our RNA-seq analysis.

427 FOSL2/FRA2 belongs to the activator-protein (AP)-1 family of transcription factors and is a known  
428 regulator of wound repair, and TGF $\beta$  mediated fibrosis (Eferl et al. 2008). Increased FOSL2/FRA2  
429 expression is detected in several chronic lung diseases, including pulmonary fibrosis, COPD, and  
430 asthma (Birnhuber et al. 2019). Notably, ectopic expression of FOSL2/FRA2 in mice results in fibrosis  
431 of several organs, including the lung, highlighting a potential profibrotic role of FOSL2/FRA2 (Eferl et  
432 al. 2008). Furthermore, our high-content screens demonstrate that FOSL2/FRA2 is required for  
433 myofibroblast differentiation, consistent with its postulated profibrotic role. Collectively, our data  
434 suggest that TCF21 and FOSL2/FRA2, whose binding sites are enriched in DMRs at strong  
435 enhancers, may mediate downstream biological effects in COPD fibroblasts and contribute to disease  
436 phenotypes, linking epigenetic changes to gene regulatory networks.

437 Transient TGF $\beta$ 1 activity is required for lung tissue regeneration and repair upon injury, however its  
438 persistent activation in lung fibroblasts leads to aberrant repair and fibrosis (Fernandez and Eickelberg  
439 2012). In turn, reduced fibroblast proliferation and response to TGF $\beta$ 1 has been proposed as a key  
440 mechanism driving the development of emphysema (Plantier et al. 2007; Togo et al. 2008; Konigshoff  
441 et al. 2009; Kulkarni et al. 2016). How these phenotypes are controlled at the molecular level is not  
442 well understood. We identified and functionally validated numerous, previously unknown regulators of  
443 lung fibroblast function in COPD. Among the top candidates, silencing of the water/glycerol channel  
444 aquaporin 3 (AQP3), Hedgehog transcription factor GLI4 and focal adhesion protein leupaxin (LPXN)  
445 had the most drastic effects on both fibroblast proliferation and TGF $\beta$ 1-mediated differentiation, establishing  
446 these three proteins as new regulators of lung fibroblast repair and remodeling. AQP3 contribution to wound  
447 healing, ECM remodeling and cell proliferation has been well documented in other cellular systems (Xu et al.  
448 2011; Ryu et al. 2012; Chen et al. 2014; Huang et al. 2015; Hou et al. 2016; Luo et al. 2016; Xiong et al.  
449 2017) and the role of LPXN in cancer cell proliferation and migration through regulation of focal adhesion  
450 sites is recognized (Kaulfuss et al. 2008; Dierks et al. 2015). Hence, downregulation of AQP3 in COPD  
451 fibroblasts could contribute, at least in part, to the decreased proliferation and contractility, manifesting in reduced  
452 fibroblast activity and impaired

453 response to injury in emphysema (Togo et al. 2008). In support of our data, change in expression of  
454 AQP3 (Heinbockel et al. 2018) and LPXN (Spira et al. 2004) in COPD/emphysematous lung tissue  
455 has been observed previously, consistent with their dysregulation in COPD fibroblasts in our RNA-seq.  
456 Notably, we showed that changes in expression of AQP3, LPXN and GLI4 in COPD fibroblasts were  
457 also associated with aberrant methylation in proximity to their TSS, indicating that their epigenetic  
458 regulation may be one of the factors contributing to the reduced repair capacity of lung fibroblasts in  
459 emphysema (Muller et al. 2006; Togo et al. 2008).

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

467

#### 468 **Data availability**

469 The WGBS and RNA-seq data generated in this study have been deposited at the European  
470 Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession  
471 **EGAXXXXXXX**.

472

#### 473 **Methods**

#### 474 **Study approval**

475 The protocol for tissue collection was approved by the ethics committees of the University of  
476 Heidelberg (S-270/2001) and Ludwig-Maximilians-Universität München (projects 333-10 and 17-166)

477 the University of Texas Health Science Center at Houston (HSC-MS-08-0354 and HSC-MS-15-1049)  
478 and followed the guidelines of the Declaration of Helsinki. All patients gave written informed consent  
479 before inclusion in the study and remained anonymous in the context of this study.

480 **Patient samples**

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

489 **Collection of lung tissue samples for profiling**

490 To identify molecular changes associated with COPD development and progression we collected  
491 distal lung tissue from patients across COPD stages and divided them into three groups: 1) no COPD,  
492 2) mild COPD (stage I, according to GOLD classification(2021)) and 3) established COPD (GOLD  
493 stages II-IV, **Suppl. Fig. 1A**). Strict patient inclusion criteria for a prospective tissue collection were  
494 established to ensure best possible matching of control and disease groups. To avoid direct smoking  
495 effects (van der Vaart et al. 2004), all included donors were ex-smokers. In addition, lung function  
496 results, as well quantitative emphysema score index (ESI) based on chest CT and whenever possible,  
497 medical history were collected for each patient for their best possible characterization. Patients'  
498 characteristics and representative images from hematoxylin and eosin staining of the tissue are  
499 provided in **Suppl. Fig. 1A and 1B**, respectively. Each tissue sample was reviewed by an experienced  
500 lung pathologist, who confirmed that all samples were tumor-free and evaluated COPD relevant  
501 phenotypes, like emphysema, airway thickening and immune infiltration. Only ex-smokers with

502 preserved lung function and no indication of emphysema or fibrosis in the test results or patient history  
503 were included as control samples. Importantly, we included 2 samples from COPD (GOLD II-IV)  
504 donors (HLD38 and HLD39), which originated from lung resections, ensuring that the observed  
505 changes are present in COPD tissue without cancer background.

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

#### 514 **Emphysema score index (ESI) determination**

515 Lung and emphysema segmentation were performed to calculate the ESI from clinically indicated  
516 preoperative CT scans taken with mixed technical parameters. After automated lung segmentation  
517 using the YACTA-software, a threshold of -950 HU was used with a noise-correction range between -  
518 910 and -950 HU to calculate the relative amount of emphysema in % of the respective lung portion  
519 (Lim et al. 2016). While usually global ESI was measured, only the contralateral non-affected lung side  
520 was used if one lung was severely affected by the tumor.

#### 521 **FFPE and H&E**

522 Representative slices from different areas of the tissue were fixed O/N with 10% neutral buffered  
523 formalin (Sigma-Aldrich). Next, fixed tissue samples were washed with PBS (Fisher Scientific) and  
524 kept in 70% ethanol at 4°C. Sample dehydration, paraffin embedding, and hematoxylin and eosin  
525 (H&E) staining was performed at Morphisto (Morphisto GmbH, Frankfurt, Germany). Per sample, two  
526 4 µm thick sections were cut on a Leica RM2255 microtome with an integrated cooling station and

527 water basin and transferred to adhesive glass slides (Superfrost Plus, Thermo Fisher). Subsequently,  
528 the sections were dried O/N in a 40°C oven to remove excess water and enhance adhesion. H&E-  
529 stained slides were evaluated by an experienced lung pathologist at the Thorax Clinic in Heidelberg.

530 **Cryopreservation of lung parenchyma**

531 Lung tissue was cryopreserved upon reception as described previously (Llamazares-Prada et al.  
532 2021). Briefly, specimens were transported in CO<sub>2</sub>-idepenent medium (Thermo Fisher Scientific)  
533 supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1%  
534 Amphotericin B (Fisher Scientific). Upon reception, tissue pieces were carefully inflated with ice-cold  
535 HBSS (Fisher Scientific), supplemented with 2mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl  
536 Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific).  
537 Exemplary samples of the different areas of the lung piece were collected for subsequent histological  
538 analysis. The pleura was removed from the remaining tissue, and airways and vessels separated from  
539 the parenchyma as much as possible. The parenchymal airway and vessel-free fractions were  
540 minced, transferred to cryo-tubes, covered with ice-cold freezing medium [70% DMEM, high glucose  
541 with GlutaMAX™ (Thermo Fisher Scientific), 20% FBS (Gibco) and 10% DMSO (Carl Roth)], kept on  
542 ice for 15 min, and transferred to -80°C in Mr. Frosty™ containers (Nalgene) to ensure a gradual  
543 temperature decrease (1°C/min). For long-term storage, samples were kept in liquid nitrogen.

544 **Fibroblast isolation from human lung tissue**

545 As no universal fibroblast markers for FACS are available, primary human lung fibroblasts were  
546 isolated by explant outgrowth from tumor-free, distal parenchymal lung tissue that has been depleted  
547 from visible airways and vessels, as previously described (Llamazares-Prada et al. 2021). Briefly, 7-8  
548 micro-dissected lung parenchyma pieces were placed per well into 6-well plates, let for 30 min at RT  
549 without medium to improve explant attachment, and carefully covered with 1 mL of growth medium:  
550 DMEM, high glucose, GlutaMAX™ (Thermo Fisher Scientific) supplemented with 2% FBS (Gibco) and  
551 1% penicillin & streptomycin (Thermo Fisher Scientific). Explants were left undisturbed for 4-7 days,

552 afterwards the medium was exchanged every 2 days and the outgrowth of fibroblasts from the  
553 explants was followed daily. Cells were collected from multiple explant pieces when reaching 70%  
554 confluence to preserve the fibroblast heterogeneity. Possible epithelial contamination was prevented  
555 by short trypsinization (0.05% trypsin with EDTA (Gibco), 3 min at 37°C) and keeping cells in the  
556 growth medium indicated above, suitable for fibroblasts enrichment.

557 **Immunofluorescence of human lung fibroblasts**

558 The purity of the isolated fibroblasts was assessed by immunofluorescence using mesenchymal  
559 markers vimentin (VIM) and alpha smooth muscle actin ( $\alpha$ SMA) as follows.  $10^4$  human lung fibroblasts  
560 in passage 3 were seeded per well in a 96-well plate for imaging (Zell-kontakt). 48h later, cells were  
561 washed with 1X PBS (Fisher Scientific), fixed for 10 min with 4% PFA (Sigma-Aldrich) at RT, washed  
562 and permeabilized for 10 min with 0.3% Triton-X-100 (Carl Roth) at RT. Unspecific staining was  
563 blocked by incubating 1h at RT with blocking buffer: 5% BSA (Carl Roth), 2% Normal Donkey Serum  
564 (Abcam) in 1X PBS (Fisher Scientific). Cells were incubated with primary antibodies against VIM (sc-  
565 7557, Santa Cruz Biotechnology, 1:200) and  $\alpha$ SMA (ab7817, Abcam, 1:100) overnight at 4°C and  
566 labelled with respective secondary antibodies [donkey anti-goat IgG Alexa Fluor 488 (A-11055,  
567 Thermo Fisher Scientific, 1:500) and donkey anti-mouse IgG Alexa Fluor 568 (A-10037, Thermo  
568 Fisher Scientific, 1:500)] for 40 min at RT in the dark. After washing with 1X PBS (Fisher Scientific),  
569 the nuclei were counterstained with DAPI (Thermo Fisher Scientific, 1:5,000) for 10 min at RT and  
570 washed with 1X PBS. Stained and fixed cells were kept in 1X PBS (Fisher Scientific) at 4°C in the dark  
571 until imaging. Imaging was conducted at the ZMBH imaging facility (Heidelberg, Germany) using the  
572 Zeiss LSM780 confocal fluorescent microscope.

573 **FACS analysis of isolated fibroblasts and lung suspension**

574 Cryopreserved lung tissues were thawed for 2 min in a 37°C water-bath, collected in 50mL Falcon  
575 tubes and washed with wash buffer: HBSS supplemented with 2mM EDTA (Thermo Fisher Scientific),  
576 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher

577 Scientific). Tissue was minced into smaller pieces prior to mechanical and enzymatic dissociation as  
578 indicated previously(Llamazares-Prada et al. 2021). Briefly, the minced tissue (1g) was introduced in  
579 GentleMACS C-tubes (Miltenyi Biotec) containing 10  $\mu$ M ROCK inhibitor (Y-27632, Adooq  
580 Bioscience), 10  $\mu$ g DNaseI (ProSpec-Tany TechnoGene), the enzyme mix from the human tumor  
581 tissue dissociation kit (Miltenyi Biotec) and 4.5 mL of CO<sub>2</sub>-independent media (Thermo Fisher  
582 Scientific) supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher Scientific) and  
583 1% Amphotericin B (Fisher Scientific). Tubes were closed tightly, introduced into the GentleMACS  
584 dissociator (Miltenyi Biotec) for mechanic disruption and the following program was performed:  
585 program h\_tumor\_01, followed by 15 min incubation at 37°C on a rotator; h\_tumor\_01, 15 min at 37°C  
586 on a rotator; h\_tumor\_02, and 15 min at 37°C on a rotator for a final enzymatic dissociation and a last  
587 mechanical shearing using the program h\_tumor\_02. The samples were pipetted up and down to help  
588 disaggregating. Finally, the enzymatic reaction was stopped by adding 20% FBS (Gibco) and single  
589 cells were collected by sequential filtering through 100  $\mu$ m, 70  $\mu$ m, and 40  $\mu$ m cell strainers (BD  
590 Falcon). Cells were centrifuged, resuspended in ACK lysis buffer (Sigma Aldrich) and incubated for 3  
591 min at RT to lyse erythrocytes. Lung single-cell suspensions were washed with HBSS (Fisher  
592 Scientific) supplemented with 2mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1%  
593 penicillin & streptomycin (Fisher Scientific) and 1% Amphotericin B (Fisher Scientific).  
  
594 To generate fibroblast single-cell suspensions, passage 3 fibroblasts were trypsinized 5 min at 37°C  
595 using 0.05% trypsin with EDTA (Gibco), centrifuged at 1000 rpm for 5 min at RT and resuspended in  
596 HBSS (Fisher Scientific) supplemented with 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher  
597 Scientific) and 1% Amphotericin B (Fisher Scientific).  
  
598 Lung and fibroblast single-cell suspensions were incubated with human TruStain FcX (Biolegend) for  
599 30 min on ice to block Fc receptors. Immune and epithelial cells were labelled using CD45 (CD45-  
600 Bv605, BD Bioscience) and EpCAM (anti-human CD326 -PE, Affymetrix eBioscience) antibodies  
601 respectively for 30 min in the dark at 4°C following manufacturer instructions. Stained samples were  
602 washed with PBS 1X (Fisher Scientific) and resuspended in HBSS (Fisher Scientific) supplemented

603 with 2mM EDTA (Thermo Fisher Scientific), 1% BSA (Carl Roth), 1% penicillin & streptomycin (Fisher  
604 Scientific) and 1% Amphotericin B (Fisher Scientific). Stained cells were added to Falcon 5 mL  
605 polystyrene tubes with 40 µm cell strainer caps (Neolab Migge). To discriminate between live and  
606 dead cells, we used SyTOX blue (Thermo Fisher Scientific) as recommended by manufacturer.

607 **RNA isolation and RNA-seq**

608  $10^5$  HLFs were harvested at passage 3 for RNA-seq studies by scraping in ice-cold PBS. Total RNA  
609 was isolated using RNeasy plus micro kit (Qiagen, Hilden, Germany) following manufacturer's  
610 instructions. DNA was removed by passing the lysate through the gDNA eliminator column and by an  
611 additional on-column DNase I treatment (Qiagen) before the elution. RNA was eluted using nuclease-  
612 free water (Thermo Fisher Scientific) and the concentration measured with Qubit HS Kit (Thermo  
613 Fisher Scientific). RNA integrity was assessed using the Bioanalyzer 2100 (Agilent, model G2939A)  
614 and the RNA 6000 pico kit (Agilent). Only samples with RIN > 8.5 were processed.

615 Libraries were prepared at the Genomics core facility (GeneCore) at EMBL using 200 ng of total RNA  
616 as input. Ribosomal RNA was removed by Illumina Ribo-Zero Gold rRNA Removal Kit  
617 Human/Mouse/Rat (Illumina, San Diego, CA, USA) and stranded total RNA-seq libraries were  
618 prepared using the Illumina TruSeq RNA Sample Preparation v2 Kit (Illumina, San Diego, CA, USA)  
619 implemented on the liquid handling robot Beckman FXP2. Obtained libraries were pooled in equimolar  
620 amounts. 1.8 pM solution of each library was pooled and loaded on the Illumina sequencer NextSeq  
621 500 High output and sequenced uni-directionally, generating ~450 million reads per run, each 75  
622 bases long.

623 **RNA-seq read alignment and transcript abundance quantification**

624 Single-end reads were mapped to the human genome version 37 (hg19) and the reference gene  
625 annotation (release 70, Ensembl) using STAR v2.5.0a (Dobin et al. 2013) with following parameters:

626 --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --  
627 alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --alignIntronMin 20 --alignIntronMax 100000 -

628 -outFilterMismatchNoverReadLmax 0.04 --outSAMtype BAM SortedByCoordinate --outSAMmultNmax  
629 1 --outMultimapperOrder Random.

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

#### 634 **Differential Gene Expression Analysis**

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

#### 644 **DNA isolation and T-WGBS**

645 Genomic DNA was extracted from  $2 \times 10^5$  primary human lung fibroblasts harvested in passage 3 using  
646 QIAamp Micro Kit (Qiagen, Hilden, Germany) following manufacturer's protocol, with an additional  
647 RNase A treatment step. T-WGBS was essentially performed as described previously (Wang et al.  
648 2013) using 30 ng genomic DNA as input. 15 pg unmethylated DNA of phage lambda was used as  
649 control for bisulfite conversion. Four sequencing libraries were generated per sample using 11  
650 amplification cycles. For each sample, equimolar amounts of all four libraries were pooled and  
651 sequenced on two lanes of a HiSeq2500 (Illumina, San Diego, California, US) machine at NGX Bio  
652 (San Francisco), resulting in 100 bp, paired-end reads.

653 **Read Alignment**

654 The whole genome bisulfite sequencing mapping pipeline MethylCtools with modifications to adapt for  
655 the T-WGBS data was used (<https://github.com/hovestadt/methylCtools>) (Hovestadt et al. 2014).  
656 Briefly, the hg19 reference genome (37d5) was transformed *in silico* for both the top strand (C to T)  
657 and bottom strand (G to A). Before alignment, adaptor sequences were trimmed using Trimmomatic  
658 (release 0.35) (Bolger et al. 2014). The first read in each read pair was then C-to-T converted and the  
659 2nd read in the pair was G-to-A converted. The converted reads were aligned to a combined reference  
660 of the transformed top (C to T) and bottom (G to A) strands using BWA MEM (bwa-0.7.8) with default  
661 parameters, yet, disabling the quality threshold for read output (-T 0) (Li and Durbin 2009). After  
662 alignment, reads were converted back to the original states, and reads mapped to the antisense  
663 strand of the respective reference were removed. Duplicate reads were marked, and the complexity  
664 determined using Picard *MarkDuplicates* (<http://picard.sourceforge.net/>). Total genome coverage was  
665 calculated using the total number of bases aligned from uniquely mapped reads over the total number  
666 of mappable bases in the genome.

667 **Methylation calling**

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

673 **DMR calling**

674 Differences in CpG methylation profiles of no COPD donors and patients diagnosed with COPD were  
675 analyzed using the R/Bioconductor package bsseq (Hansen et al. 2012). First, the data was smoothed  
676 using the built-in *BSmooth* function with default settings. Only CpG sites with a coverage of at least 4x  
677 were kept for subsequent analysis. A t-statistic was calculated between no COPD and COPD (II-IV)

678 samples using the *BSmooth.tstat* function with following parameters: local.correct=TRUE,  
679 maxGap=300, estimate.var="same". Differentially methylated regions (DMRs) were called by 1)  
680 selecting the regions with the 5% most extreme t-statistics in the data (lower and upper 2.5% quantile;  
681 default paramters of the dmrFinder function), 2) filtering for regions exhibiting at least 10% methylation  
682 difference between no COPD and COPD (II-IV) and containing at least 3 CpGs. 3) Finally, a non-  
683 parametric Wilcoxon test was applied using the average methylation level of the region to remove  
684 potentially false positive regions, since the t-statistic is not well-suited for not normally distributed  
685 values, as expected at very low/high (close to 0% / 100%) methylation levels. A significance level of  
686 0.1 was used.

687 **Gene ontology analysis**

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

690 **ChIP-seq Data**

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

695 Chromatin states for adult human lung fibroblasts (accession: ENCFF001TDQ) were obtained from  
696 ENCODE data base (Ernst and Kellis 2017). The number of bp of each DMR coinciding with a  
697 chromatin state was calculated and the chromatin state with the largest overlap was assigned to the  
698 DMR. To assess the genomic background, 10.000 regions with matching size and CpG distribution  
699 were randomly selected.

700 **Profile Plots of DMRs and Identification of Super-enhancers**

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

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

#### 709 **Transcription factor motif analysis**

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

713 In order to obtain information about methylation dependent binding for transcription factor motifs which  
714 are enriched at DMRs, the results of a recent SELEX study (Yin et al. 2017) were integrated in the  
715 analysis and a motif database of 1787 binding motifs with associated methylation dependency was  
716 constructed. The log odds detection threshold was calculated for the HOMER motif search as  
717 following:

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

#### 723 **siRNA-based phenotypic assays in primary human lung fibroblasts**

724 Normal human lung fibroblasts (NHLFs) from 2 donors (donor IDs: 608197; 543644) and diseased  
725 COPD human lung fibroblasts (DHLFs) from 3 donors (donor IDs: OF3353, OF3418 and OF3238)  
726 were purchased from Lonza and tested for their response to FGF2 and TGF $\beta$  stimulation.

727 **Fibroblast to myofibroblast transition (FMT) assay**

728 Cells in passage 5 were plated in a poly-D-lysine coated 384 CellCarrier microtiter plate from  
729 PerkinElmer in fibroblast basal medium (FBM) with FGM-2TM Single Quots (Lonza) at a density of  
730 2000 cells per well. Six hours after cell seeding, cells were transfected with siRNAs (Horizon ON-  
731 Target Plus siRNA pools) as previously described (Weigle et al. 2019). 24h later, the medium was  
732 replaced by FBM containing 0,1% fetal calf serum (starvation medium). 24h later, fibroblast to  
733 myofibroblast differentiation was initiated by adding fresh starvation medium containing a mixture of  
734 Ficoll 70 and 400 (GE Healthcare; 37,5mg/mL and 25mg/mL, respectively), 200 $\mu$ M vitamin C and  
735 5ng/mL TGF $\beta$ 1. After 72h the medium was removed, cells were fixed with 100% ice-cold methanol for  
736 30 min, washed with PBS, permeabilized 20 min using 1% Triton-X-100 (Sigma), washed, and  
737 blocked for 30 min with 3% BSA in PBS. After an additional wash step, cell nuclei were stained using  
738 1 $\mu$ M Hoechst 33342 (Molecular Probes). Alpha smooth-muscle actin ( $\alpha$ SMA) and collagen I (col1)  
739 were stained using monoclonal antibodies (1:1000 diluted, Sigma, A2547 and SAB4200678,  
740 respectively). For detection of primary antibodies, cells were washed and incubated for 30 min at 37°C  
741 with AF647-goat-anti-mouse IgG2b ( $\alpha$ SMA) and AF568 goat-anti-mouse IgG1 (col1) antibodies. After  
742 removal of secondary antibodies, cells were stained with HCS Cell Mask Green stain (Invitrogen,  
743 1:50,000). Following a final PBS 1x wash step, images were acquired in a GE Healthcare InCell 2200  
744 Analyzer, using 2D-deconvolution for nuclei (Hoechst channel), cells (FITC channel),  $\alpha$ SMA (Cy5  
745 channel) and collagen I (TexasRed channel), and images were transferred to and analyzed using  
746 Perkin Elmer's Columbus<sup>TM</sup> Image Storage as previously described (Aumiller et al. 2017; Weigle et al.  
747 2019). Briefly, the building blocks (BB) of the Columbus<sup>TM</sup> Image Analysis system were used, first  
748 nuclei (Hoechst channel acquisition) were detected using the BB "nuclei". Second, cells were defined  
749 with the BB "find cytoplasm" from the FITC channel image.  $\alpha$ SMA fibers and col1 area were defined

750 by two individual BBs “find simple image region” based on images acquired in the Cy5 and TexasRed  
751 channels, respectively. Both  $\alpha$ SMA fibers and col1 readouts were normalized to the number of cells  
752 per image field. The FMT assay was performed in 2 NHLFs and 3 DHLFs independent donors. For  
753 each donor, siRNA transfection for every gene was performed in 4 technical replicates. In addition, the  
754 FMT screen was performed in each donor twice independently.

755 **Proliferation assay (nuclei count)**

756 To analyze the effects of gene knockdown on FGF2-mediated fibroblast proliferation, 2000 cells at  
757 passage 5 were plated in a poly-D-lysine coated 384-CellCarrier microtiter plate (PerkinElmer) in FBM  
758 with FGM-2TM Single Quots (Lonza). Six hours after seeding, cells were transfected with siRNAs  
759 (Horizon ON-Target Plus siRNA pools) as previously described(Weigle et al. 2019). 24h later, the  
760 medium was replaced by FBM containing 0,1% fetal calf serum (starvation medium). 24h later the  
761 medium was replaced by starvation medium containing 20ng/mL basic FGF (R&D Systems). After 72h  
762 the medium was removed, cells washed with PBS and treated with 3.7% formaldehyde containing  
763 1 $\mu$ M Hoechst 33342 for 30 min. Cells were washed with PBS and images were acquired in a GE  
764 Healthcare InCell 2200 Analyzer, using 2D-deconvolution for nuclei (Hoechst channel). Nuclei  
765 numbers were determined using the Columbus<sup>TM</sup> image analysis software as described above (BB  
766 “nuclei”). The proliferation assay was performed in 5 independent donors: 2 NHLFs and 3 DHLFs. For  
767 each donor, siRNA transfection to knockdown the selected candidate genes was performed in 4  
768 technical replicates. In addition, the proliferation screen was performed in each donor twice  
769 independently.

770 **Analysis of phenotypic screen data**

771 siRNA transfection was performed in each phenotypic screen in 4 technical replicates and repeated 2  
772 times independently for each donor. For statistical analysis of both the FMT and proliferation data,  
773 each readout (nuclei for both FMT and proliferation,  $\alpha$ SMA and col1 for FMT) was first normalized  
774 within each plate, based on the negative control wells, corresponding to cells transfected with non-

775 target siRNA control (NTC) (40 wells per plate). After plate-based normalization, the normalized  
776 values for the specific readout (e.g., nuclei,  $\alpha$ SMA and col1) were averaged for the independent  
777 replicates. To measure the siRNA effect as the magnitude of the difference between an individual  
778 siRNA and the negative control (NTC siRNA), the previously described strictly standardized mean  
779 difference (SSMD) was applied(Zhang 2007; Zhang et al. 2007). The following formula was used for  
780 the SSMD calculation:  $SSMD = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$  , where  $\mu_1$  is the normalized mean of all NTC siRNAs,  $\mu_2$  is  
781 the mean of the normalized values of siRNA for a given gene,  $\sigma_1$  is the variance of all normalized NTC  
782 siRNAs values and  $\sigma_2$  is the variance of all normalized values transfected with siRNA for a given  
783 gene.

#### 784 **Statistical analysis**

785 An unpaired non-parametric *t* test (Mann-Whitney test, GraphPad Prism software, version 8.0.1) was  
786 employed to compare the lung function (FEV1 and FEV1/FVC values) between control and COPD  
787 donors (**Fig 1A**). For the **Suppl. Fig. 1A** displaying all the patient metadata of the three groups studied  
788 (control, COPD I and COPD II-IV), one-way ANOVA non-parametric unpaired test was used (Kruskal-  
789 Wallis test, GraphPad Prism software, version 8.0.1) followed by correction for multiple comparisons  
790 using Dunn's test. The adjusted p-value for each comparison is shown. The significance level was set  
791 to 0.05.

792

#### 793 **Author contributions**

794 RZJ, US & MLP contributed to the design and conception of the study. MLP, SP and MR performed  
795 experiments with the help from VM, RT, MR, OS and DWe. MSchu and CK performed the phenotypic  
796 screens. US performed bioinformatic analysis and integration of the RNA-seq and WGBS data, with  
797 the help of KQ, JH and GP. TMu, MS, FH, HW, FH, HKQ and IK provided lung tissue and patient data.  
798 AW performed the pathological analysis of the H&E lung specimens. CPH determined the emphysema  
799 score index of the patients. HS, BJ, VB, DWy, TPJ, BM, BB, CI and CP provided critical input, analysis

800 software and materials. RZJ, US and MLP wrote the manuscript with input from all authors. All authors  
801 contributed to scientific discussions and approved the final version of the manuscript.

802

803 **Acknowledgments**

804 We would like to thank Lung Biobank (Heidelberg, Germany) – a member of the Biomaterial bank  
805 Heidelberg (BMBH), the tissue bank of the National Center for Tumor Diseases (NCT) and the  
806 Biobank platform of the German Center for Lung Research (DZL), as well as the Asklepios Biobank for  
807 Lung Diseases, member of the German Center for Lung Research (DZL) for providing Biomaterials  
808 and Data. We also thank Christa Stolp for help with collecting primary material. We acknowledge  
809 excellent sequencing service and helpful discussions from the Genomics core facility (GeneCore,  
810 EMBL, Germany) for RNA-seq and from NGX Bio (San Francisco, USA) for T-WGBS sequencing, as  
811 well as support from the ZMBH imaging facility (Heidelberg, Germany) for immunofluorescence. We  
812 thank Morphisto GmbH (Frankfurt, Germany) for excellent histological service. We also thank  
813 Christian Tidona (BioMed X Innovation Center) and Markus Koester (Boehringer Ingelheim) for helpful  
814 project discussions. We thank the ENCODE Consortium for providing and the Bradly Bernstein Lab for  
815 producing the ChIP datasets of human fibroblasts used in this study.

816

817 **Potential Conflict of Interest**

818 RZJ, MLP, VM, US, RT, MR, SP, OS and BM as employees of BioMed X Institute received research  
819 funding by Boehringer Ingelheim Pharma GmbH & Co KG. HS, BJ, MSchu, CK, KQ and DWy are  
820 employees of Boehringer Ingelheim Pharma GmbH & Co KG and receive compensation as such. TM  
821 received a research grant, non-financial support and has patent applications with Roche Diagnostics  
822 GmbH outside of the described work. CPH has stock ownership in GSK; received research funding  
823 from Siemens, Pfizer, MeVis and Boehringer Ingelheim; consultation fees from Schering-Plough,  
824 Pfizer, Basilea, Boehringer Ingelheim, Novartis, Roche, Astellas, Gilead, MSD, Lilly Intermune and

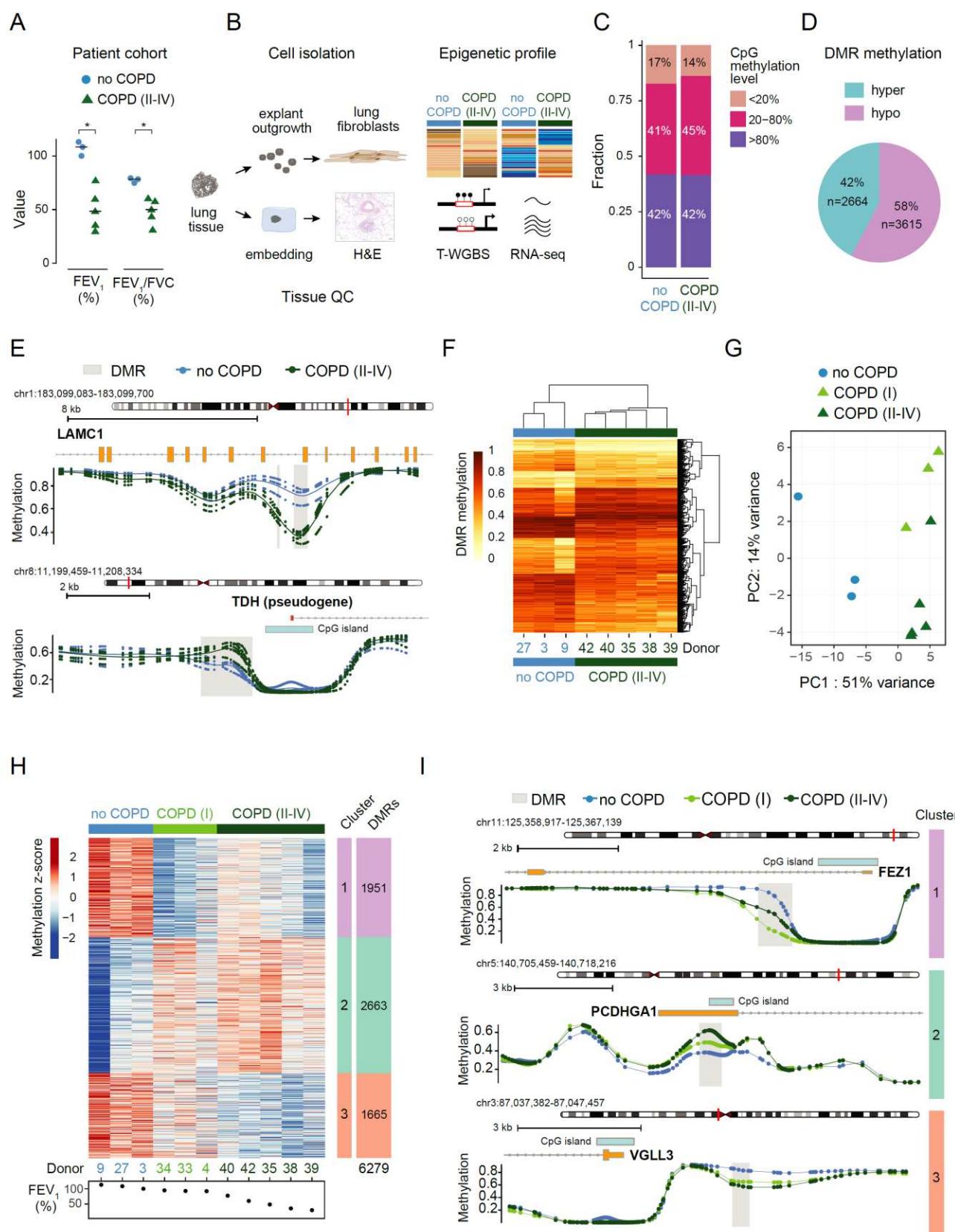
825 Fresenius, and speaker fees from Gilead, Essex, Schering-Plough, AstraZeneca, Lilly, Roche, MSD,  
826 Pfizer, Bracco, MEDA Pharma, Intermune, Chiesi, Siemens, Covidien, Boehringer Ingelheim, Grifols,  
827 Novartis, Basilea and Bayer, outside the submitted work. HW received consultation fees from Intuitive  
828 and Roche.

829

830 **Funding**

831 This study was supported by Boehringer Ingelheim. The work was partly funded by the School of  
832 Biosciences (Cardiff University) to RZJ and the German Center for Lung Research (DZL) to CP, TM,  
833 MS, FH, CPH, HW, AW and MLP. Asklepios Biobank as part of the DZL is partly funded by BMBF.

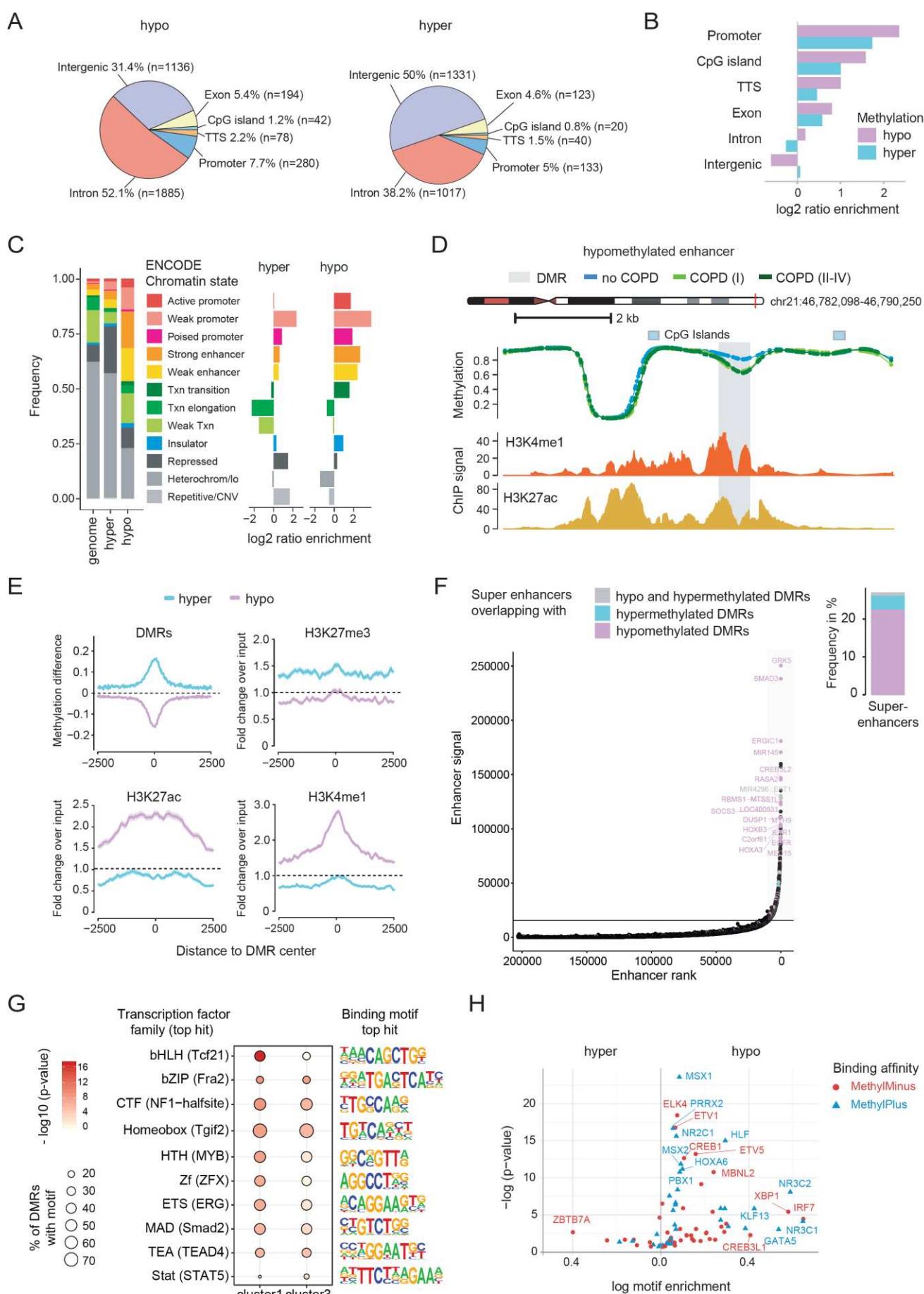
834



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

838 **A** Lung function data of COPD (II-IV) and no COPD (ex-smoker control group) donors used in this  
839 study. Lung function between the two groups is significantly different (\*p-value<0.05, unpaired non-  
840 parametric Mann-Whitney t-test). FEV<sub>1</sub>, forced expiratory volume in 1 s; FVC, forced vital capacity. **B**  
841 Schematic diagram illustrating the experimental approach used for epigenetic (T-WGBS) and  
842 transcriptomic (RNA-seq) profiling of purified primary parenchymal lung fibroblasts. **C-I** Data of  
843 tagmentation-based WGBS (T-WGBS) of primary fibroblasts from no COPD and COPD (II-IV) patients  
844 were analyzed on single CpGs level (**C**) and on differentially methylated regions (DMRs) (**D-I**). **C**  
845 Genome wide CpG methylation statistic. Bar plot showing the fraction of high (>80%), moderate (20-  
846 80%) and low (<20%) methylated CpGs in no COPD and COPD (II-IV) samples. **D** Number of hyper-  
847 or hypomethylated DMRs in COPD (II-IV). **E** Detailed view of a representative hypo- (top) and  
848 hypermethylated (bottom) DMR (grey box). CpG methylation levels of each individual donor (dots) and  
849 the group average (lines) methylation profile of three no COPD (blue) and five COPD (II-IV) (dark  
850 green) donors are displayed. RefSeq annotated genes and CpG islands are indicated. **F** Heatmap of  
851 6,279 DMRs identified in COPD (II-IV). Statistically significant DMRs (at significance level=0.1; see  
852 methods for DMR calling details) with at least three CpGs and a mean difference in methylation  
853 between no COPD and COPD (II-IV) of  $\geq 10\%$  were selected. Color shades indicate low (light) or high  
854 (dark) DMR methylation. **G** Principal component analysis (PCA) of COPD (II-IV) (dark green), no  
855 COPD (blue) and mild COPD (I) (light green, samples not used for initial DMR calling) on identified  
856 6,279 DMRs. **H** K-means clustering of all DMRs identified between no COPD and COPD (II-IV) across  
857 all samples, including COPD (I). Three clusters were identified. Cluster 1 shows early hypomethylation  
858 in COPD (I), clusters 2 and 3, gradual hyper- and hypomethylation, respectively. Donors are sorted  
859 according to their FEV<sub>1</sub> value as indicated at the bottom. **I** Representative methylation profiles at  
860 selected DMRs from each cluster. Group median CpG methylation is shown for no COPD (blue),

861 COPD (I) (light green) and COPD (II-IV) (dark green). RefSeq annotated genes and CpG islands are  
862 indicated. FEV<sub>1</sub>, forced expiratory volume in 1 s.



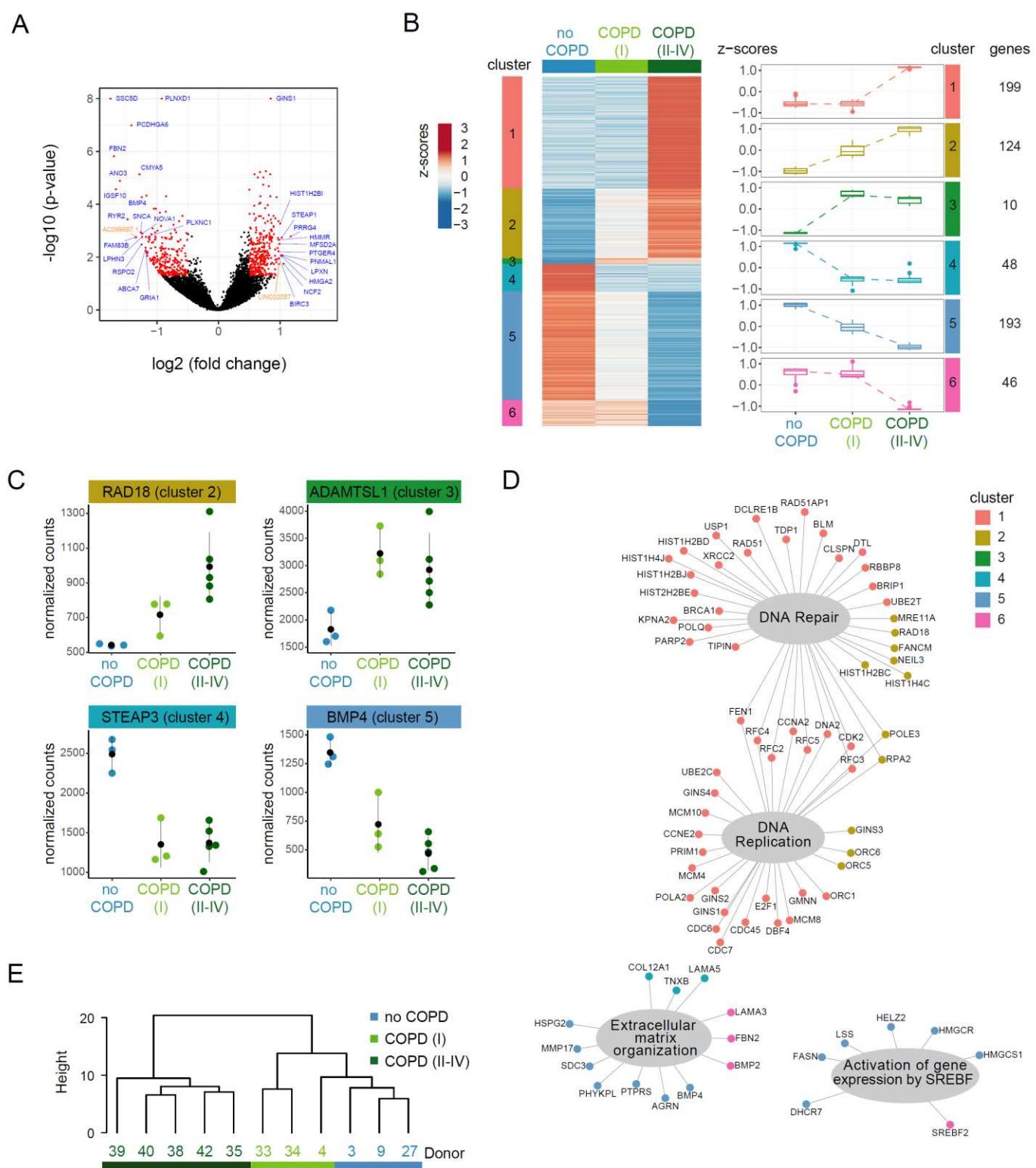
864 **Figure 2. DNA methylation changes occur at regulatory regions in primary human lung**  
865 **fibroblasts cells during COPD.**

866 **A-B** Genomic location of identified DMRs. **A** Distribution of genomic features overlapping with hypo-  
867 (left) and hypermethylated (right) DMRs. **B** Enrichment of genomic features at hypo- (purple) and  
868 hypermethylated (cyan) DMRs compared to a sampled background of 10,000 regions exhibiting no  
869 significant change in methylation. TTS, transcription termination site. **C** Distribution of human lung  
870 fibroblast specific chromatin states (ENCODE accession: ENCF001TDQ) at hypo- and  
871 hypermethylated DMRs. Fraction of DMRs overlapping with specific chromatin states is shown on the  
872 left panel. The genome background was sampled using 10,000 regions with matching GC content  
873 exhibiting no significant change in methylation. Chromatin state enrichment relative to the genome  
874 background is illustrated in the right panel. **D** Genome browser view of an example DMR at a putative  
875 enhancer region. Group median CpG methylation is shown for no COPD (blue), COPD (I) (light green)  
876 and COPD (II-IV) (dark green). At the bottom the level of enhancer marks is depicted as fold-change  
877 over control: H3K4me1 (ENCODE accession: ENCF102BGI) and H3K27ac (ENCODE accession:  
878 ENCF386FDQ). **E** Alterations of DNA methylation and selected histone marks around DMRs. Solid  
879 lines represent the mean profile and shaded lines the standard error of the mean across all  
880 summarized regions. **F** Ranking of enhancer elements, defined by the co-occurrence of H3K4me1 and  
881 H3K27ac signals in human lung fibroblasts. The horizontal line defines the signal and corresponding  
882 rank threshold used to identify super enhancers (SE). Selected SE overlapping with DMRs are  
883 annotated and the nearest gene to the SE is indicated. The fraction of SEs overlapping with hypo-  
884 (purple), hypermethylated (cyan) or both (grey) is illustrated in the bar plot on the right panel. **G**  
885 Transcription factor motifs most enriched at DMRs overlapping with strong enhancers (ENCODE  
886 chromatin states) from cluster 1 and 3 (see Figure 1H). **H** Enrichment of methylation sensitive  
887 transcription factor motifs at hypo- (right) and hypermethylated (left) DMRs. Methylation sensitive  
888 motifs were derived from the study of Yin *et al.* (Yin et al. 2017). Transcription factors, whose binding

889 affinity was impaired upon methylation of their corresponding DNA motif are shown in red  
890 (MethylMinus) and transcription factors, whose binding affinity was increased, in blue (MethylPlus).

891

892



893

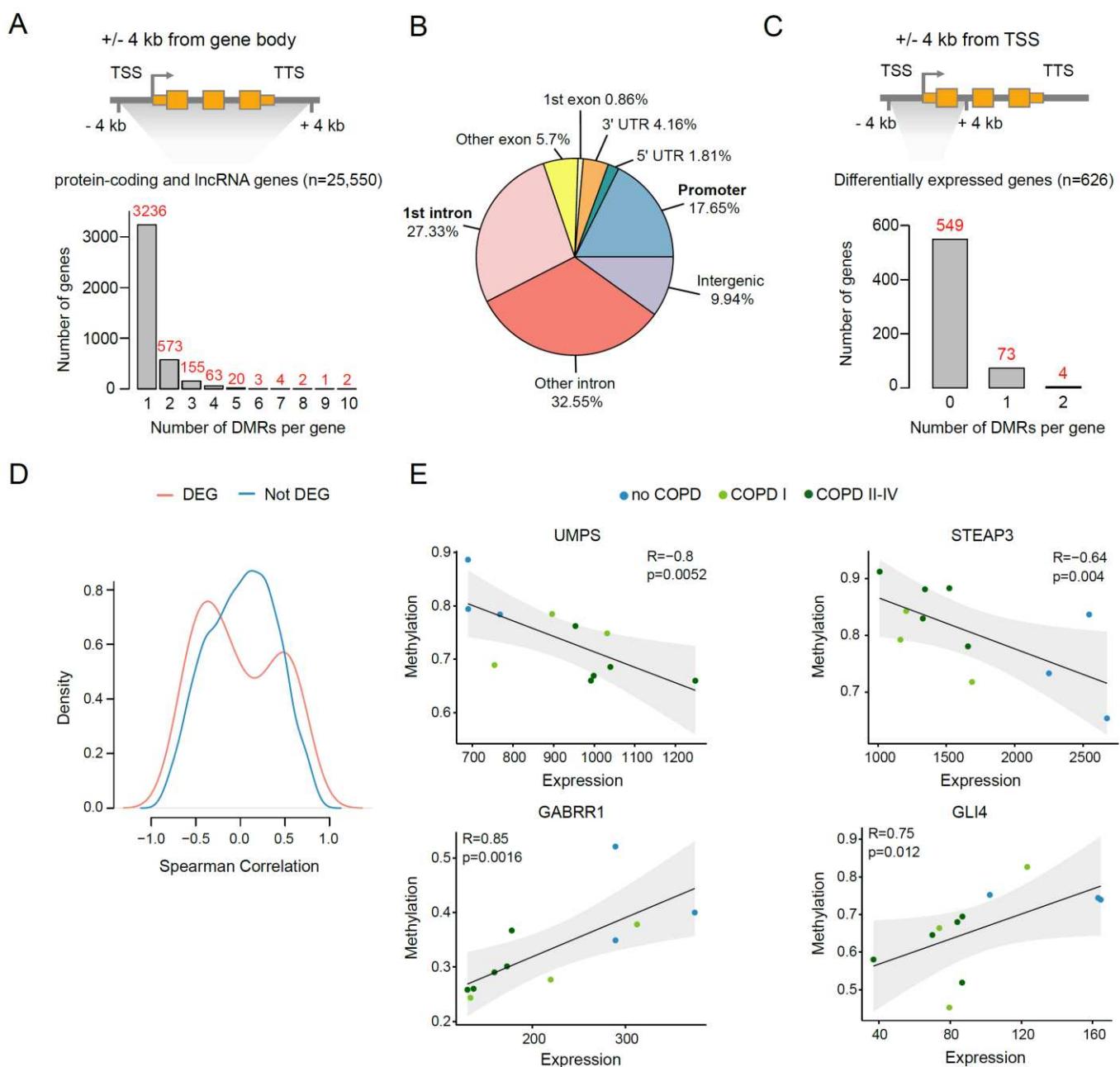
894 **Figure 3. DNA methylation changes in primary human lung fibroblasts are accompanied by**  
 895 **gene expression changes in COPD.**

896 **A** Volcano plot of differentially expressed genes (DEG) (red dots; FDR < 0.05 and  $|\log_2(\text{fold change})|$   
897 > 0.5) in COPD (II-IV) compared to no COPD controls. Protein-coding (blue) and lincRNA (orange)  
898 with the highest expression change or lowest p-values are labeled. **B** Self-organizing maps (SOM)  
899 clustering based on the scaled median expression level per group of the 620 DEG identified between  
900 COPD (II-IV) and no COPD samples. DEGs were grouped into 6 distinct clusters showing different  
901 kinetics in COPD progression: Clusters 1 and 6 show late changes. Clusters 2 and 5 display changes  
902 gradually progressing with disease severity. Clusters 3 and 4 correspond to early changes observed  
903 already in COPD (I). **C** Selected examples of DEG across disease stages from clusters 2-5. **D** DEGs  
904 associated with altered biological processes (grey bubbles) in COPD. DEG nodes are colored  
905 according to their corresponding gene expression kinetic in COPD (clusters defined in **3B**). **E**  
906 Unsupervised hierarchical clustering of all samples, including COPD (I) based on DEGs (n=620)  
907 identified between no COPD and COPD (II-IV) samples.

908

909

910



911

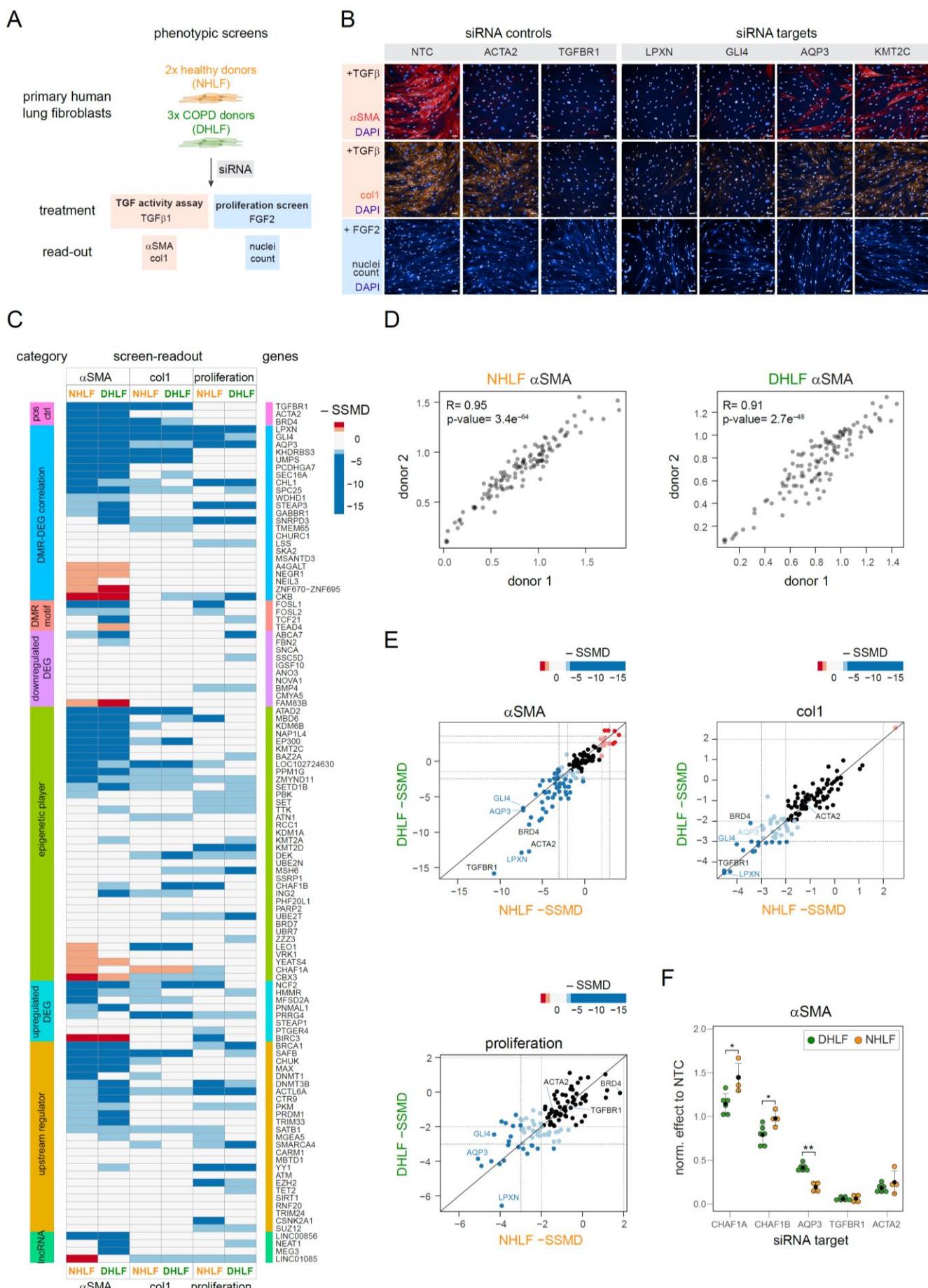
912 **Figure 4. Integrative data analysis reveals epigenetically regulated genes in COPD fibroblasts.**

913 **A** DMRs located in the proximity of annotated protein-coding and lincRNA genes. DMRs within +- 4  
 914 kb from gene body were assigned to their corresponding gene. TSS, transcription start site; TTS,  
 915 transcription termination site. **B** Gene features of gene-associated DMRs. Promoter is defined as the  
 916 region of - 1 kb to + 100 bp around the TSS. **C** DMRs located in the proximity of the TSS of DEGs.  
 917 DMRs within +- 4 kb from TSS of DEG were assigned to their corresponding gene. **D** Spearman

918 correlation between gene expression and DMR methylation. DMRs within +/- 4 kb from TSS were  
919 considered. Gene-DMR pairs were split into DEGs (red) and not significantly changed genes (no DEG,  
920 blue). **E** Scatter plots showing examples of correlations between gene expression and methylation of  
921 promoter associated DMRs. Each dot represents an individual donor. Dots are color coded according  
922 to disease state. Gene expression is illustrated as normalized counts. Methylation is illustrated as  
923 average beta value of the corresponding DMR. Linear regression analysis was performed (black line)  
924 and the 95% confidence interval is indicated (grey area). P-value and Spearman correlation coefficient  
925 (R) are indicated.

926

927

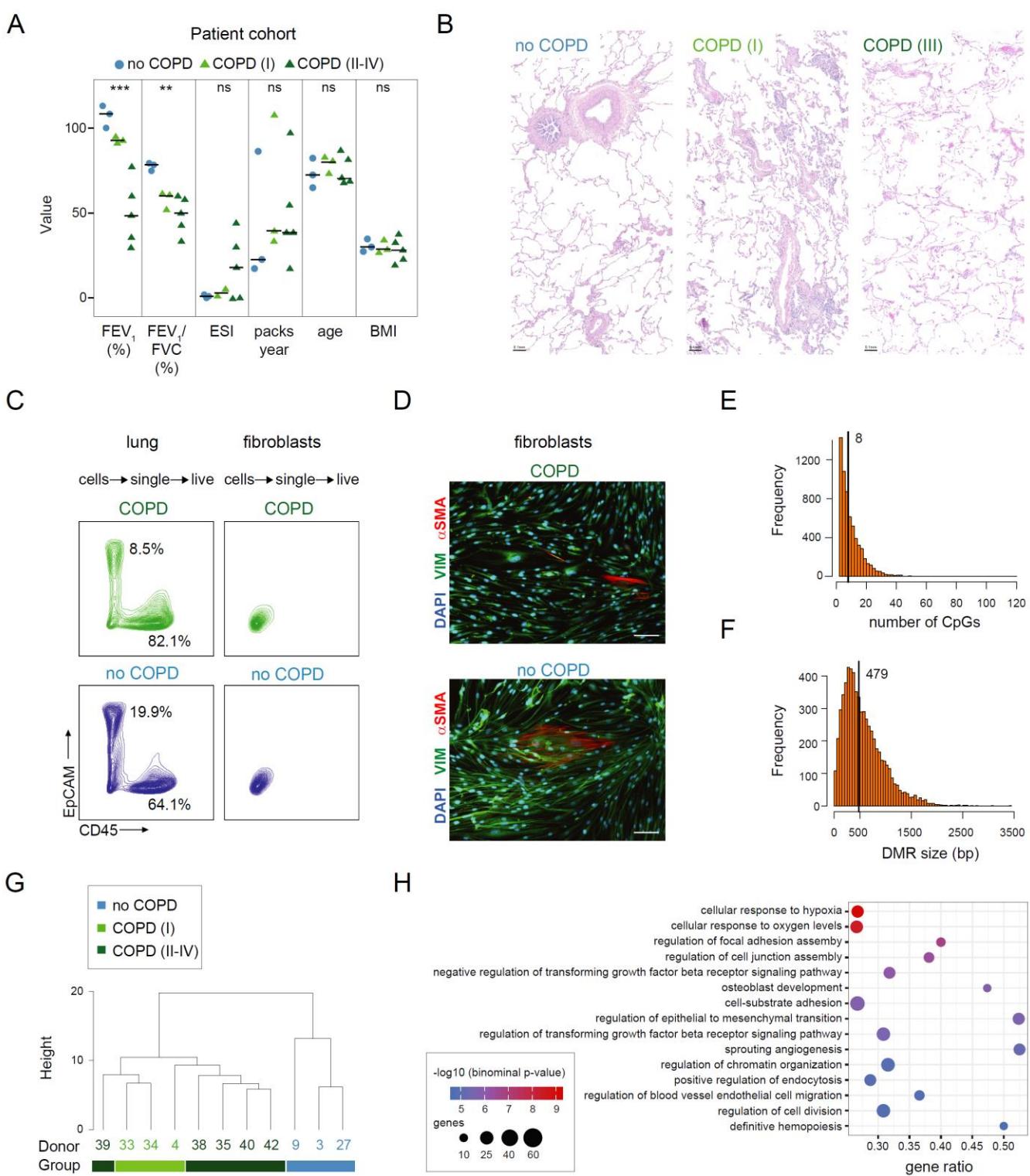


929 **Figure 5. siRNA-based phenotypic screens in normal and COPD primary human lung**  
930 **fibroblasts identify multiple candidate genes regulating COPD phenotypes.**

931 **A** Schematic representation of the siRNA-based phenotypic assays performed in primary normal  
932 human lung fibroblasts (NHLFs, 2 donors) and diseased/COPD human lung fibroblasts (DHLFs, 3  
933 donors). KD, knockdown; FGF, fibroblast growth factor; TGF $\beta$ , transforming growth factor beta;  $\alpha$ SMA,  
934 alpha smooth-muscle actin; col1, collagen 1. **B** Examples of primary pictures obtained in the screens  
935 showing the performance of the siRNA controls as well as positive hits upon KD. NTC, non-targeting  
936 siRNA control. **C** Heatmap showing the effect of the KD of each candidate gene on the three  
937 measured readouts ( $\alpha$ SMA, col1 and proliferation) in primary normal (NHLF) or COPD (DHLF) human  
938 lung fibroblasts, red: readout higher than NTC, blue: readout lower than NTC; SSMD = strictly  
939 standardized mean difference.  $|\text{SSMD values}| \geq 2$  are shown in lighter shade and  $|\text{SSMD values}| \geq 3$   
940 in stronger shade. **D** Scatterplots showing the correlation of the screen data from 2 different NHLFs  
941 donors (left) and 2 different DHLFs (right) for the  $\alpha$ SMA readout. **E** Comparison of the KD effect of  
942 each candidate relative to NTCs in NHLFs and DHLFs. Each dot represents a unique candidate  
943 tested, blue and red dots represent significant hits ( $|\text{SSMD values}| \geq 2$  are shown in lighter shade and  
944  $|\text{SSMD values}| \geq 3$  in stronger shade). Assay controls are labeled in black and examples of strong hits  
945 regulating fibroblast to myofibroblast transition and cell proliferation processes are labeled in blue. **F**  
946 Dotplots showing examples of positive hits with significant differences between NHLFs (2 donors, 2  
947 biological replicates each, shown in green) and DHLFs (3 donors, 2 biological replicates each, shown  
948 in orange) in  $\alpha$ SMA readout. Screen readout was normalized to the corresponding non NTC. TGF $\beta$ R1  
949 and ACTA2 represent positive screen controls. The results of all replicates are shown. Statistical  
950 evaluation was performed with unpaired two-tailed student's t-test. \*: p-value < 0.05; \*\*: p-value <  
951 0.01. Black dots denote the means and error bars represent the standard deviation.

952

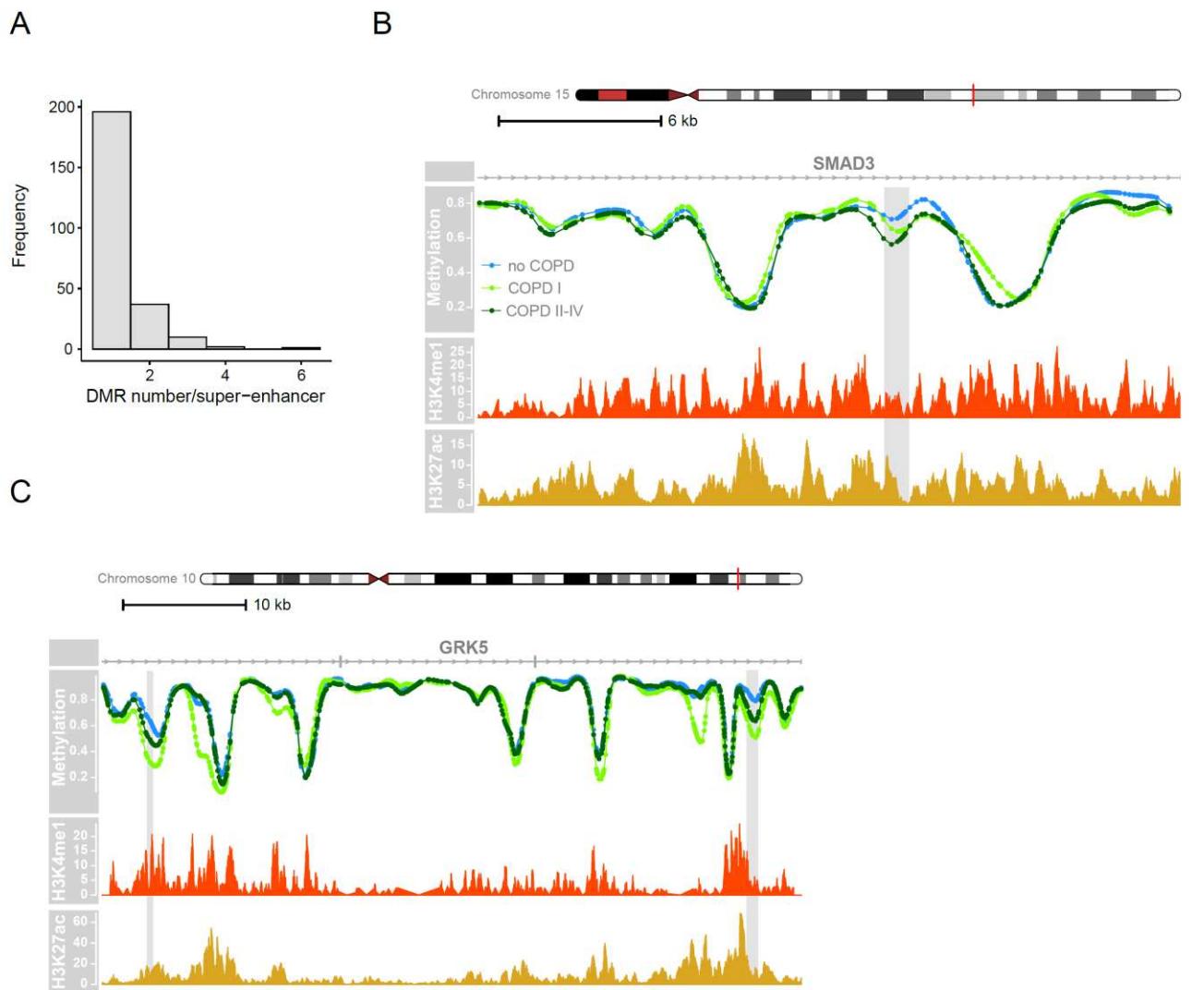
953 **Extended data figures:**



955 **Supplementary Figure 1. Genome-wide DNA methylation changes occur early in human lung  
956 fibroblasts in COPD and progress with disease development (supporting information Figure 1)**

957 **A** Characteristics of the lung tissue donors used in this study for fibroblast isolation (BMI, body mass  
958 index; FEV<sub>1</sub>, forced expiratory volume in 1 s; FVC, forced vital capacity; ESI, emphysema score  
959 index). Lung function between COPD and smoker control group (no COPD) is significantly different  
960 (\*p-value < 0.05; non-parametric unpaired Kruskal-Wallis test). **B** Examples of hematoxylin and eosin  
961 (H&E) stainings of no COPD, COPD (I) and COPD (II-IV) tissue samples. **C-D** Validation of the purity  
962 of isolated fibroblast by FACS (**C**) and immunofluorescence (**D**). **C** Left, contour plot of lung live-single  
963 cell suspensions from COPD (green) and no COPD (blue) donors gated as indicated above. Right,  
964 contour plot of trypsinized fibroblasts from COPD (green) and no COPD (blue) donors. EpCAM-PE  
965 was used as an epithelial marker and CD45-Bv605 as an immune marker. **D** Immunofluorescence  
966 staining of fibroblasts isolated from the lungs of COPD (top) and no COPD (bottom) donors and  
967 stained with antibodies against vimentin (VIM, mesenchymal marker) and alpha smooth-muscle actin  
968 (αSMA, myofibroblast marker). DAPI was used to counterstain the nuclei. Scale bar: 0.1mm. **E-F**  
969 Histograms showing the number of CpGs per DMR (**E**) and the DMR size distribution (**F**). Median  
970 values are indicated and highlighted by a vertical line. **G** Hierarchical clustering of all samples based  
971 on 6,279 DMRs identified between no COPD and COPD (II-IV). **H** Functional annotation of genes  
972 located next to hypomethylated DMRs using GREAT. Hits were sorted according to the binomial p-  
973 value and the top 15 hits are shown. The adjusted p-value is indicated by the color code and the  
974 number of DMR associated genes is indicated by the node size.

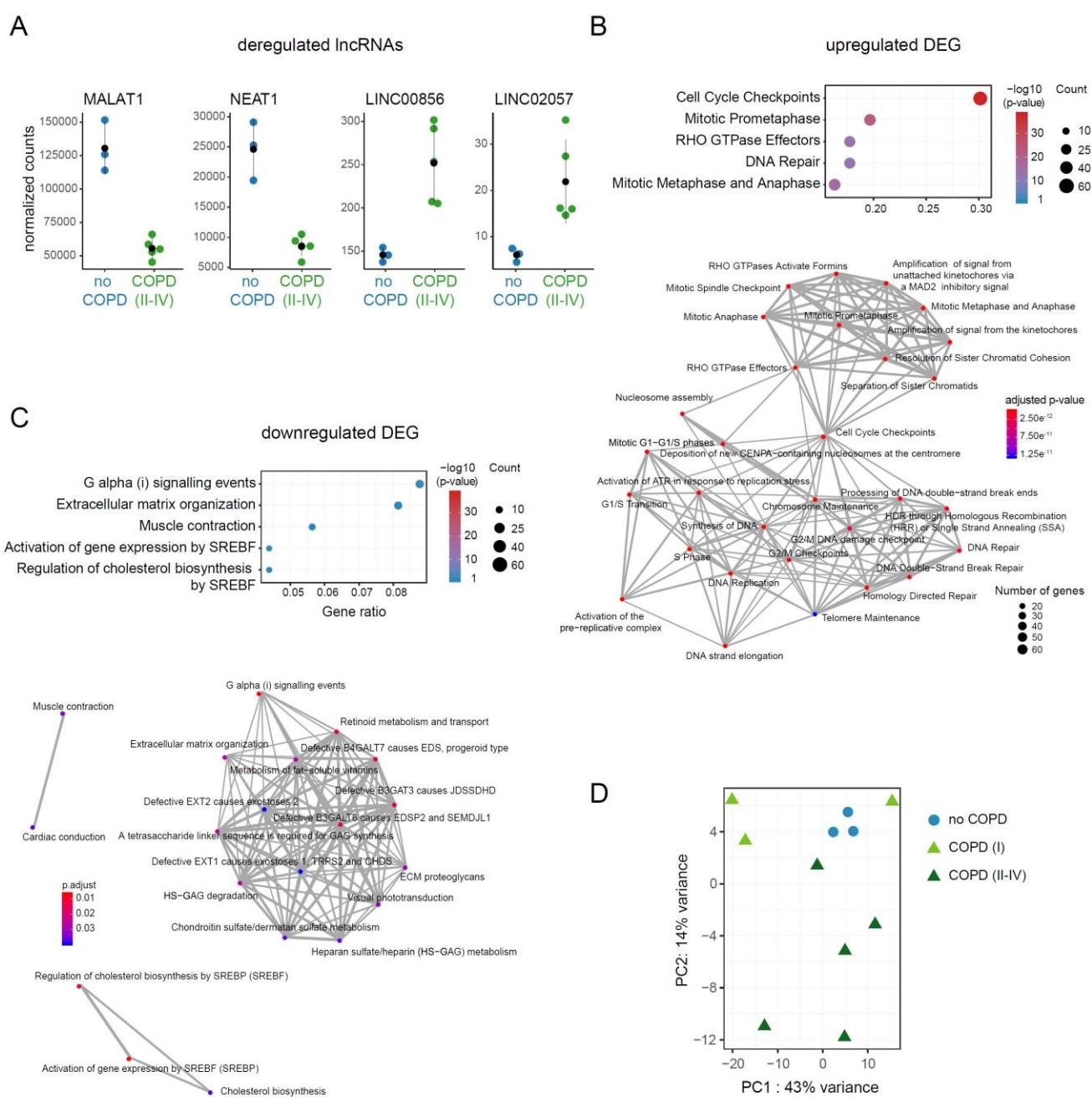
975



976

977 **Supplementary Figure 2. DNA methylation changes occur at regulatory regions in primary**  
978 **human lung fibroblasts cells during COPD (supporting information Figure 2)**

979 **A** Number of DMRs associated with super-enhancers **B-C** Genome browser views of two super-  
980 enhancer regions overlapping with identified DMRs (shaded in grey). Group median CpG methylation  
981 is shown for no COPD (blue), COPD (I) (light green) and COPD (II-IV) (dark green). At the bottom, the  
982 level of enhancer marks: H3K4me1 (ENCODE accession: ENCFF102BGI; red track) and H3K27ac  
983 (ENCODE accession: ENCFF386FDQ; orange track) is depicted as fold change over control.



984

985 **Supplementary Figure 3 DNA methylation changes in primary human lung fibroblasts are**  
 986 **accompanied by gene expression changes in COPD (supporting information Figure 3)**

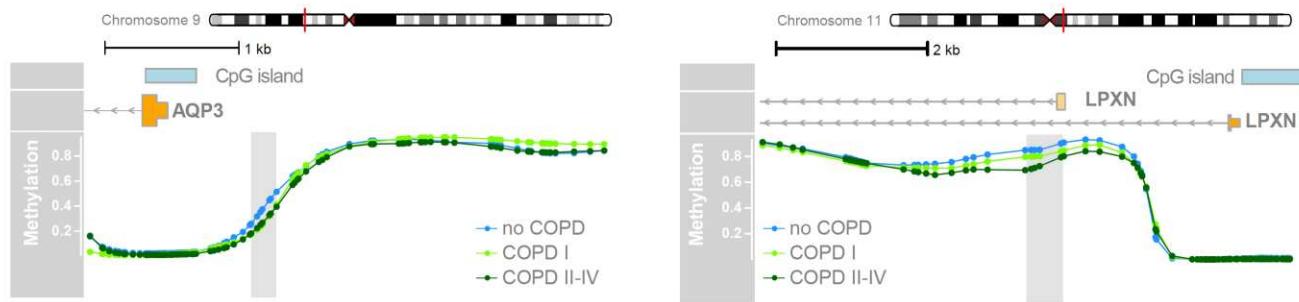
987 **A** Normalized RNA-seq read counts of four exemplary differentially expressed lncRNAs **B-C**  
 988 Functional annotation of upregulated (**B**) and downregulated (**C**) DEGs. Top 5 enriched GO biological  
 989 process terms are shown on top of each panel. Bottom panels show a more detailed overview of the

990 biological processes enriched for upregulated (**B**) and downregulated (**C**) DEGs. Biological processes  
991 are connected, if DEGs associated with both processes are in common. (**D**) Unsupervised principal  
992 component analysis (PCA) of all samples based on the 500 most variable expressed genes.

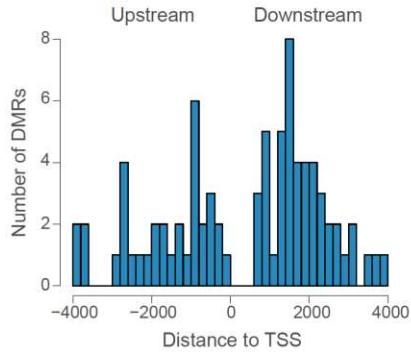
993

994

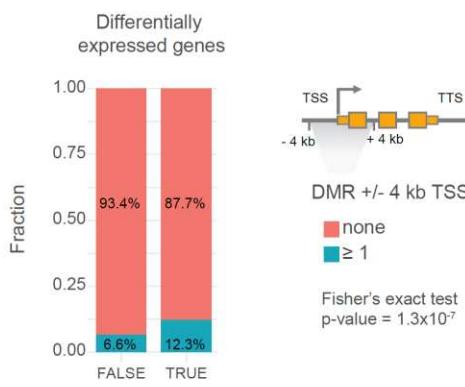
A



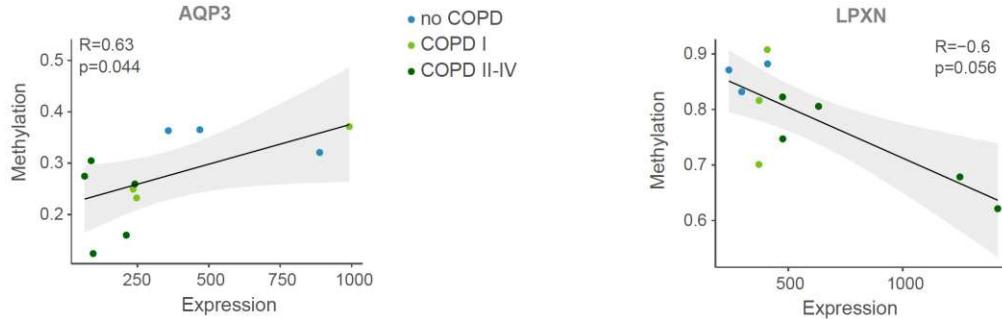
B



C



D



995

996 **Supplement Figure 4. Integrative data analysis reveals epigenetically regulated genes in COPD**  
997 **fibroblasts (supporting information Figure 4).**

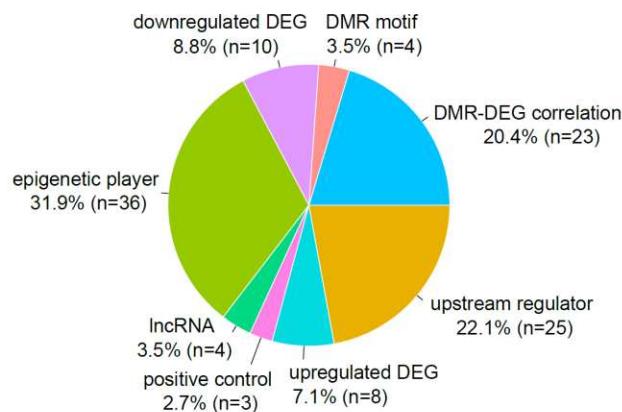
998 **A** Representative methylation profiles and DMRs (shaded in grey). Group median CpG methylation is  
999 shown for no COPD (blue), COPD (I) (light green) and COPD (II-IV) (dark green). RefSeq annotated  
1000 genes and CpG islands are indicated. **B** Location of DMRs relative to the TSS of DEGs. **C** Fraction of  
1001 genes associated with at least one DMR in the proximity ( $\pm 4$  kb) of the TSS are indicated in blue.  
1002 Genes were split into not significantly changed genes (left bar, FALSE) and DEGs (right bar, TRUE).

1003 DMRs are significantly enriched at DEGs (Fisher's exact test: p-value =  $1.3 \times 10^{-7}$ ). **D** Examples of  
1004 genes showing correlation between gene expression and methylation. Scatter plots showing positive  
1005 (left, AQP3) and negative (right, LPXN) correlation between gene expression and methylation of  
1006 promoter-associated DMR. Each dot represents an individual donor. Dots are color coded according  
1007 to disease state. Gene expression is illustrated as normalized counts. Methylation is illustrated as  
1008 average beta value of the corresponding DMR. Linear regression analysis was performed (black line)  
1009 and the 95% confidence interval is indicated (grey area). P-value and Spearman correlation coefficient  
1010 (R) are indicated.

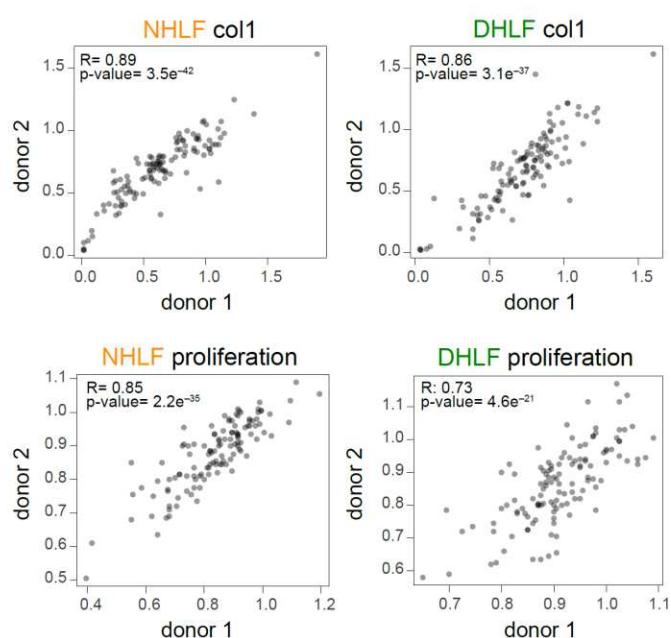
1011

1012

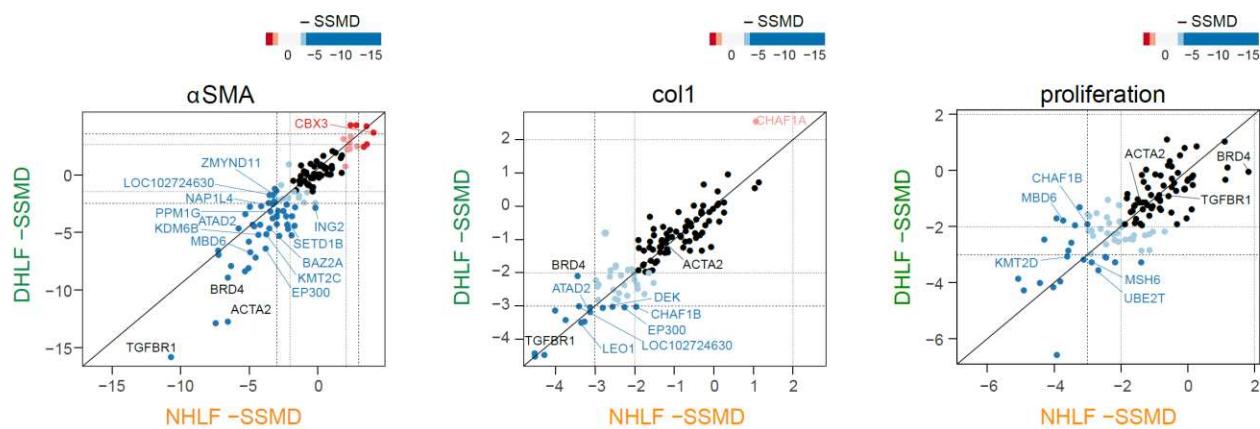
A



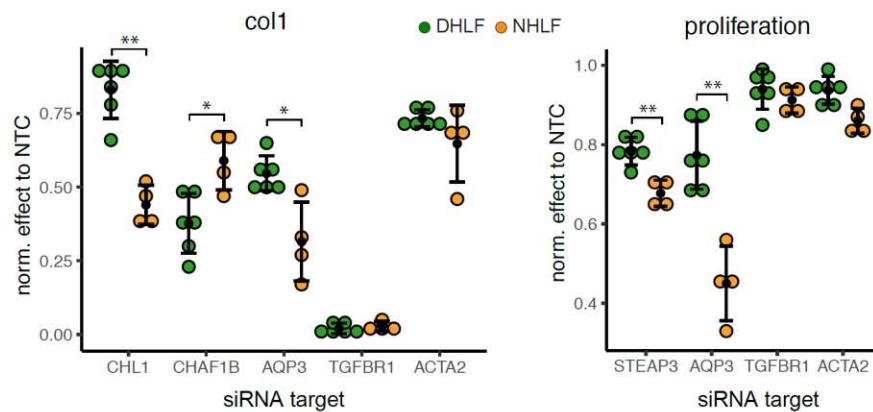
B



C



D



1015 **Supplement Figure 5. siRNA-based phenotypic screens in normal and COPD primary human**  
1016 **lung fibroblasts identify multiple candidate genes regulating COPD phenotypes (supporting**  
1017 **information Figure 5).**

1018 **A** Distribution of candidates selected for functional validation among different categories (related to  
1019 Figure 5C). The selected candidates were divided into 7 categories: (1) epigenetic players  
1020 dysregulated on the transcriptional level, (2) DMR motif includes transcription factors with binding sites  
1021 overrepresented in DMRs, (3) upstream regulators identified by the Ingenuity Pathway Analysis, (4)  
1022 DMR-DEG correlation represents differentially expressed genes correlated with epigenetic changes in  
1023 their promoter, (5) top downregulated genes, (6) top upregulated DEG, and (7) lncRNA for the top  
1024 dysregulated long non-coding RNAs. In addition, 3 assay controls (ACTA2, TGF $\beta$ R1 and BRD4) were  
1025 included. **B** Scatterplots showing the correlation of the data obtained from 2 different NHLF (left) and 2  
1026 DHLF (right) donors in the phenotypic screens using col1 and proliferation/nuclei count as readouts. **C**  
1027 Examples of epigenetic hits from the siRNA screen for each indicated readout. Each dot represents a  
1028 unique candidate tested, blue and red dots represent significant hits ( $|SSMD \text{ values}| \geq 2$  are shown in  
1029 lighter shade and  $|SSMD \text{ values}| \geq 3$  in stronger shade). Assay controls are labeled in black and  
1030 epigenetic factors identified as strong positive hits ( $|SSMD \text{ values}| \geq 3$ ) in fibroblast to myofibroblast  
1031 transition and cell proliferation processes are labeled in blue or red depending on the effect on each  
1032 readout upon KD. **D** Dotplots showing examples of positive hits with significant differences between  
1033 NHLFs (2 donors, 2 biological replicates each, shown in green) and DHLFs (3 donors, 2 biological  
1034 replicates each, shown in orange) in col1 and proliferation/nuclei count readout. Screen readout was  
1035 normalized to the corresponding non targeting siRNA control (NTC). TGF $\beta$ R1 and ACTA2 represent  
1036 screen controls. The results of all replicates are shown. Statistical evaluation was performed with  
1037 unpaired two-tailed student's t-test. \*: p-value < 0.05; \*\*: p-value < 0.01. Black dots denote the means  
1038 and error bars represent standard deviation.

1039

1040 **References**

- 1041 Acharya A, Baek ST, Huang G, Eskiocak B, Goetsch S, Sung CY, Banfi S, Sauer MF, Olsen GS, Duffield JS et al.  
1042 2012. The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast  
1043 progenitors. *Development* **139**: 2139-2149.
- 1044 Agusti A, Celli B, Faner R. 2017. What does endotyping mean for treatment in chronic obstructive pulmonary  
1045 disease? *Lancet* **390**: 980-987.
- 1046 Akama T, Chun T-H. 2018. Transcription factor 21 (TCF21) promotes proinflammatory interleukin 6 expression  
1047 and extracellular matrix remodeling in visceral adipose stem cells. *Journal of Biological Chemistry* **293**:  
1048 6603-6610.
- 1049 Aumiller V, Strobel B, Romeike M, Schuler M, Stierstorfer BE, Kreuz S. 2017. Comparative analysis of lysyl  
1050 oxidase (like) family members in pulmonary fibrosis. *Scientific Reports* **7**.
- 1051 Barnes PJ. 2019a. Inflammatory endotypes in COPD. *Allergy* **74**: 1249-1256.
- 1052 Barnes PJ. 2019b. Small airway fibrosis in COPD. *Int J Biochem Cell Biol* **116**: 105598.
- 1053 Barnes PJ, Baker J, Donnelly LE. 2019. Cellular Senescence as a Mechanism and Target in Chronic Lung Diseases.  
1054 *Am J Respir Crit Care Med* **200**: 556-564.
- 1055 Barnes PJ, Burney PG, Silverman EK, Celli BR, Vestbo J, Wedzicha JA, Wouters EF. 2015. Chronic obstructive  
1056 pulmonary disease. *Nat Rev Dis Primers* **1**: 15076.
- 1057 Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS,  
1058 Smith TM et al. 2002. Aberrant promoter methylation in bronchial epithelium and sputum from current  
1059 and former smokers. *Cancer Res* **62**: 2370-2377.
- 1060 Birnhuber A, Biasin V, Schnoegl D, Marsh LM, Kwapiszewska G. 2019. Transcription factor Fra-2 and its  
1061 emerging role in matrix deposition, proliferation and inflammation in chronic lung diseases. *Cell Signal*  
1062 **64**: 109408.
- 1063 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.  
1064 *Bioinformatics* **30**: 2114-2120.
- 1065 Braitsch CM, Combs MD, Quaggin SE, Yutzy KE. 2012. Pod1/Tcf21 is regulated by retinoic acid signaling and  
1066 inhibits differentiation of epicardium-derived cells into smooth muscle in the developing heart.  
1067 *Developmental Biology* **368**: 345-357.
- 1068 Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D. 2016. Differential DNA methylation marks and gene  
1069 comethylation of COPD in African-Americans with COPD exacerbations. *Respir Res* **17**: 143.
- 1070 Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, Colicino E, Karrasch S, Wahl S,  
1071 Kunze S et al. 2018. Metastable DNA methylation sites associated with longitudinal lung function  
1072 decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. *Epigenetics* **13**:  
1073 1039-1055.
- 1074 Casas-Recasens S, Noell G, Mendoza N, Lopez-Giraldo A, Garcia T, Guirao A, Agusti A, Faner R. 2021. Lung DNA  
1075 Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow  
1076 Limitation Severity. *American Journal of Respiratory and Critical Care Medicine* **203**: 129-134.
- 1077 Chen A, Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J et  
1078 al. 2013. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS ONE* **8**:  
1079 e63812.
- 1080 Chen J, Wang T, Zhou YC, Gao F, Zhang ZH, Xu H, Wang SL, Shen LZ. 2014. Aquaporin 3 promotes epithelial-  
1081 mesenchymal transition in gastric cancer. *J Exp Clin Cancer Res* **33**: 38.
- 1082 Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Ziniti J, Laird NM et  
1083 al. 2014. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and  
1084 meta-analysis. *Lancet Respir Med* **2**: 214-225.
- 1085 Clifford RL, Fishbane N, Patel J, MacIsaac JL, McEwen LM, Fisher AJ, Brandsma CA, Nair P, Kobor MS, Hackett TL  
1086 et al. 2018. Altered DNA methylation is associated with aberrant gene expression in parenchymal but  
1087 not airway fibroblasts isolated from individuals with COPD. *Clin Epigenetics* **10**: 32.

- 1088 Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymurdov UK, Narayanan AK  
1089 et al. 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**:  
1090 D794-D801.
- 1091 Dierks S, von Hardenberg S, Schmidt T, Bremmer F, Burfeind P, Kaulfuss S. 2015. Leupaxin stimulates adhesion  
1092 and migration of prostate cancer cells through modulation of the phosphorylation status of the actin-  
1093 binding protein caldesmon. *Oncotarget* **6**: 13591-13606.
- 1094 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR:  
1095 ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
- 1096 Eferl R, Hasselblatt P, Rath M, Popper H, Zenz R, Komnenovic V, Idarraga MH, Kenner L, Wagner EF. 2008.  
1097 Development of pulmonary fibrosis through a pathway involving the transcription factor Fra-2/AP-1.  
1098 *Proceedings of the National Academy of Sciences* **105**: 10525-10530.
- 1099 Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**:  
1100 2478-2492.
- 1101 Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al.  
1102 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43-49.
- 1103 Esteller M. 2008. Molecular origins of cancer: Epigenetics in cancer. *New Engl J Med* **358**: 1148-1159.
- 1104 Fernandez IE, Eickelberg O. 2012. The impact of TGF-beta on lung fibrosis: from targeting to biomarkers. *Proc  
1105 Am Thorac Soc* **9**: 111-116.
- 1106 Garudadri S, Woodruff PG. 2018. Targeting Chronic Obstructive Pulmonary Disease Phenotypes, Endotypes,  
1107 and Biomarkers. *Ann Am Thorac Soc* **15**: S234-S238.
- 1108 GOLD 2021. The Global Strategy for Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary  
1109 Disease (updated 2021). doi:<https://goldcopd.org/2021-gold-reports/>.
- 1110 Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen  
1111 TH, Barr RG et al. 2010. Meta-analyses of genome-wide association studies identify multiple loci  
1112 associated with pulmonary function. *Nat Genet* **42**: 45-52.
- 1113 Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to  
1114 differentially methylated regions. *Genome Biology* **13**.
- 1115 Heinbockel L, Marwitz S, Schromm AB, Watz H, Kugler C, Ammerpohl O, Schnepp K, Rabe KF, Droemann D,  
1116 Goldmann T. 2018. Identification of novel target genes in human lung tissue involved in chronic  
1117 obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* **13**: 2255-2259.
- 1118 Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW et al.  
1119 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression.  
1120 *Nature* **459**: 108-112.
- 1121 Hey J, Paulsen M, Toth R, Weichenhan D, Butz S, Schatterny J, Liebers R, Lutsik P, Plass C, Mall MA. 2021.  
1122 Epigenetic reprogramming of airway macrophages promotes polarization and inflammation in muco-  
1123 obstructive lung disease. *Nat Commun* **12**: 6520.
- 1124 Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, Wain LV, Hall IP, Jackson VE, Wyss AB et al.  
1125 2017. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung  
1126 function and pulmonary fibrosis. *Nat Genet* **49**: 426-432.
- 1127 Holz O, Zuhlike I, Jaksztat E, Muller KC, Welker L, Nakashima M, Diemel KD, Branscheid D, Magnussen H, Jorres  
1128 RA. 2004. Lung fibroblasts from patients with emphysema show a reduced proliferation rate in culture.  
1129 *Eur Respir J* **24**: 575-579.
- 1130 Hou SY, Li YP, Wang JH, Yang SL, Wang Y, Wang Y, Kuang Y. 2016. Aquaporin-3 Inhibition Reduces the Growth  
1131 of NSCLC Cells Induced by Hypoxia. *Cell Physiol Biochem* **38**: 129-140.
- 1132 Hovestadt V, Jones DTW, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz  
1133 H-J et al. 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation  
1134 sequencing. *Nature* **510**: 537-541.
- 1135 Huang YT, Zhou J, Shi S, Xu HY, Qu F, Zhang D, Chen YD, Yang J, Huang HF, Sheng JZ. 2015. Identification of  
1136 Estrogen Response Element in Aquaporin-3 Gene that Mediates Estrogen-induced Cell Migration and  
1137 Invasion in Estrogen Receptor-positive Breast Cancer. *Sci Rep* **5**: 12484.

- 1138 Ito K, Ito M, Elliott WM, Cosio B, Caramori G, Kon OM, Barczyk A, Hayashi S, Adcock IM, Hogg JC et al. 2005.  
1139       Decreased histone deacetylase activity in chronic obstructive pulmonary disease. *N Engl J Med* **352**:  
1140       1967-1976.
- 1141 Kachroo P, Morrow JD, Kho AT, Vyhildal CA, Silverman EK, Weiss ST, Tantisira KG, Demeo DL. 2020. Co-  
1142       methylation analysis in lung tissue identifies pathways for fetal origins of COPD. *European Respiratory  
1143       Journal* **56**.
- 1144 Kaulfuss S, Grzmil M, Hemmerlein B, Thelen P, Schweyer S, Neesen J, Bubendorf L, Glass AG, Jarry H, Auber B et  
1145       al. 2008. Leupaxin, a novel coactivator of the androgen receptor, is expressed in prostate cancer and  
1146       plays a role in adhesion and invasion of prostate carcinoma cells. *Mol Endocrinol* **22**: 1606-1621.
- 1147 Konigshoff M, Kneidinger N, Eickelberg O. 2009. TGF-beta signaling in COPD: deciphering genetic and cellular  
1148       susceptibilities for future therapeutic regimen. *Swiss Med Wkly* **139**: 554-563.
- 1149 Kramer A, Green J, Pollard J, Jr., Tugendreich S. 2014. Causal analysis approaches in Ingenuity Pathway Analysis.  
1150       *Bioinformatics* **30**: 523-530.
- 1151 Kulkarni T, O'Reilly P, Antony VB, Gaggar A, Thannickal VJ. 2016. Matrix Remodeling in Pulmonary Fibrosis and  
1152       Emphysema. *Am J Respir Cell Mol Biol* **54**: 751-760.
- 1153 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*  
1154       **25**: 1754-1760.
- 1155 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence  
1156       reads to genomic features. *Bioinformatics* **30**: 923-930.
- 1157 Lim HJ, Weinheimer O, Wielputz MO, Dinkel J, Hielscher T, Gompelmann D, Kauczor HU, Heussel CP. 2016. Fully  
1158       Automated Pulmonary Lobar Segmentation: Influence of Different Prototype Software Programs onto  
1159       Quantitative Evaluation of Chronic Obstructive Lung Disease. *PLoS One* **11**: e0151498.
- 1160 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  
1161       lung mesenchymal cells in development and fibrosis. *iScience* **24**: 102551.
- 1162 Llamazares-Prada M, Espinet E, Mijosek V, Schwartz U, Lutsik P, Tamas R, Richter M, Behrendt A, Pohl ST, Benz  
1163       NP et al. 2021. Versatile workflow for cell type-resolved transcriptional and epigenetic profiles from  
1164       cryopreserved human lung. *Jci Insight* **6**.
- 1165 Lotfi CFP, Passaia BS, Kremer JL. 2021. Role of the bHLH transcription factor TCF21 in development and  
1166       tumorigenesis. *Braz J Med Biol Res* **54**: e10637.
- 1167 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with  
1168       DESeq2. *Genome Biol* **15**: 550.
- 1169 Luo J, Liu X, Liu J, Jiang M, Luo M, Zhao J. 2016. Activation of TGF-beta1 by AQP3-Mediated H2O2 Transport  
1170       into Fibroblasts of a Bleomycin-Induced Mouse Model of Scleroderma. *J Invest Dermatol* **136**: 2372-  
1171       2379.
- 1172 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves  
1173       functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**: 495-501.
- 1174 Medvedeva YA, Lennartsson A, Ehsani R, Kulakovskiy IV, Vorontsov IE, Panahandeh P, Khimulya G, Kasukawa T,  
1175       Drabløs F. 2015. EpiFactors: a comprehensive database of human epigenetic factors and complexes.  
1176       *Database* **2015**: bav067.
- 1177 Miyazono K. 2000. Positive and negative regulation of TGF-beta signaling. *J Cell Sci* **113 ( Pt 7)**: 1101-1109.
- 1178 Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Glass K et al.  
1179       2016. DNA methylation profiling in human lung tissue identifies genes associated with COPD.  
1180       *Epigenetics* **11**: 730-739.
- 1181 Muller KC, Welker L, Paasch K, Feindt B, Erpenbeck VJ, Hohlfeld JM, Krug N, Nakashima M, Branscheid D,  
1182       Magnussen H et al. 2006. Lung fibroblasts from patients with emphysema show markers of senescence  
1183       in vitro. *Respir Res* **7**: 32.
- 1184 Nobukuni S, Watanabe K, Inoue J, Wen FQ, Tamaru N, Yoshida M. 2002. Cigarette smoke inhibits the growth of  
1185       lung fibroblasts from patients with pulmonary emphysema. *Respirology* **7**: 217-223.

- 1186 Noordhoek JA, Postma DS, Chong LL, Vos JT, Kauffman HF, Timens W, van Straaten JF. 2003. Different  
1187 proliferative capacity of lung fibroblasts obtained from control subjects and patients with emphysema.  
1188 *Exp Lung Res* **29**: 291-302.
- 1189 Park J, Ivey MJ, Deana Y, Riggsbee KL, Sørensen E, Schwabl V, Sjöberg C, Hjertberg T, Park GY, Swonger JM et al.  
1190 2019. The Tcf21 lineage constitutes the lung lipofibroblast population. *American Journal of Physiology-  
1191 Lung Cellular and Molecular Physiology* **316**: L872-L885.
- 1192 Phan SH. 2008. Biology of fibroblasts and myofibroblasts. *Proc Am Thorac Soc* **5**: 334-337.
- 1193 Plantier L, Boczkowski J, Crestani B. 2007. Defect of alveolar regeneration in pulmonary emphysema: role of  
1194 lung fibroblasts. *Int J Chron Obstruct Pulmon Dis* **2**: 463-469.
- 1195 Qiu W, Baccarelli A, Carey VJ, Bautaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W,  
1196 Lomas DA et al. 2012. Variable DNA methylation is associated with chronic obstructive pulmonary  
1197 disease and lung function. *Am J Respir Crit Care Med* **185**: 373-381.
- 1198 Quaggin SE, Schwartz L, Cui SY, Igarashi P, Deimling J, Post M, Rossant J. 1999. The basic-helix-loop-helix  
1199 protein Pod1 is critically important for kidney and lung organogenesis. *Development* **126**: 5771-5783.
- 1200 Rabe KF, Watz H. 2017. Chronic obstructive pulmonary disease. *Lancet* **389**: 1931-1940.
- 1201 Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2010. A unique chromatin signature  
1202 uncovers early developmental enhancers in humans. *Nature* **470**: 279-283.
- 1203 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  
1204 expression is up-regulated by TGF-beta1 in rat peritoneal mesothelial cells and plays a role in wound  
1205 healing. *Am J Pathol* **181**: 2047-2057.
- 1206 Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, Shrine N, Qiao D, Bartz TM,  
1207 Kim DK et al. 2019. Genetic landscape of chronic obstructive pulmonary disease identifies  
1208 heterogeneous cell-type and phenotype associations. *Nat Genet* **51**: 494-505.
- 1209 Sayols S, Scherzinger D, Klein H. 2016. dupRadar: a Bioconductor package for the assessment of PCR artifacts in  
1210 RNA-Seq data. *BMC Bioinformatics* **17**: 428.
- 1211 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  
1212 loop between ZBTB7A and TGF-beta in breast cancer. *Oncol Lett* **14**: 1403-1410.
- 1213 Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE et al. 2011.  
1214 Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat  
1215 Genet* **43**: 1082-1090.
- 1216 Sood A, Petersen H, Blanchette CM, Meek P, Picchi MA, Belinsky SA, Tesfaigzi Y. 2010. Wood smoke exposure  
1217 and gene promoter methylation are associated with increased risk for COPD in smokers. *Am J Respir  
1218 Crit Care Med* **182**: 1098-1104.
- 1219 Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS. 2004. Gene expression profiling of  
1220 human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* **31**: 601-610.
- 1221 Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, Nimwegen Ev, Wirbelauer C, Oakeley EJ, Gaidatzis  
1222 D et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*  
1223 **480**: 490-495.
- 1224 Sundar IK, Yin Q, Baier BS, Yan L, Mazur W, Li D, Susiarjo M, Rahman I. 2017. DNA methylation profiling in  
1225 peripheral lung tissues of smokers and patients with COPD. *Clin Epigenetics* **9**: 38.
- 1226 Szulakowski P, Crowther AJ, Jimenez LA, Donaldson K, Mayer R, Leonard TB, MacNee W, Drost EM. 2006. The  
1227 effect of smoking on the transcriptional regulation of lung inflammation in patients with chronic  
1228 obstructive pulmonary disease. *Am J Respir Crit Care Med* **174**: 41-50.
- 1229 Togo S, Holz O, Liu X, Sugiura H, Kamio K, Wang X, Kawasaki S, Ahn Y, Fredriksson K, Skold CM et al. 2008. Lung  
1230 fibroblast repair functions in patients with chronic obstructive pulmonary disease are altered by  
1231 multiple mechanisms. *Am J Respir Crit Care Med* **178**: 248-260.
- 1232 van den Bosch T, Kwiatkowski M, Bischoff R, Dekker FJ. 2017. Targeting transcription factor lysine acetylation in  
1233 inflammatory airway diseases. *Epigenomics* **9**: 1013-1028.
- 1234 van der Vaart H, Postma DS, Timens W, ten Hacken NH. 2004. Acute effects of cigarette smoke on inflammation  
1235 and oxidative stress: a review. *Thorax* **59**: 713-721.

- 1236 Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, Zhang M, Lonergan KM, Steiling K, Brown CJ et al.  
1237 2014. DNA methylation is globally disrupted and associated with expression changes in chronic  
1238 obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* **50**: 912-922.
- 1239 Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP et  
1240 al. 2015. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive  
1241 pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**: 769-  
1242 781.
- 1243 Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, Hokanson JE, Bowler RP, Crapo JD, DeMeo DL.  
1244 2015. Smoking-Associated Site-Specific Differential Methylation in Buccal Mucosa in the COPDGene  
1245 Study. *Am J Respir Cell Mol Biol* **53**: 246-254.
- 1246 Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, Bähr M, Wolf S, Shendure J, Eils R et al. 2013.  
1247 Tagmentation-based whole-genome bisulfite sequencing. *Nature Protocols* **8**: 2022-2032.
- 1248 Wehrens R, Kruisselbrink J. 2018. Flexible Self-Organizing Maps in kohonen 3.0. *J Stat Softw* **87**: 1-18.
- 1249 Weigle S, Martin E, Voegtle A, Wahl B, Schuler M. 2019. Primary cell-based phenotypic assays to  
1250 pharmacologically and genetically study fibrotic diseases in vitro. *Journal of Biological Methods* **6**: 115
- 1251 WHO. 2019. World health statistics overview 2019: monitoring health for the SDGs, sustainable development  
1252 goals. World Health Organization In *Global Health Observatory (GHO) data*. World Health Organization
- 1253 Whyte Warren A, Orlando David A, Hnisz D, Abraham Brian J, Lin Charles Y, Kagey Michael H, Rahl Peter B, Lee  
1254 Tong I, Young Richard A. 2013. Master Transcription Factors and Mediator Establish Super-Enhancers at Key  
1255 Cell Identity Genes. *Cell* **153**: 307-319.
- 1256 Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Bandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST  
1257 et al. 2009. A genome-wide association study of pulmonary function measures in the Framingham  
1258 Heart Study. *PLoS Genet* **5**: e1000429.
- 1259 Woodruff PG, Agusti A, Roche N, Singh D, Martinez FJ. 2015. Current concepts in targeting chronic obstructive  
1260 pulmonary disease pharmacotherapy: making progress towards personalised management. *Lancet*  
1261 **385**: 1789-1798.
- 1262 Wyss AB Sofer T Lee MK Terzikhan N Nguyen JN Lahousse L Latourelle JC Smith AV Bartz TM Feitosa MF et al.  
1263 2018. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary  
1264 function. *Nat Commun* **9**: 2976.
- 1265 Xiong G, Chen X, Zhang Q, Fang Y, Chen W, Li C, Zhang J. 2017. RNA interference influenced the proliferation  
1266 and invasion of XWLC-05 lung cancer cells through inhibiting aquaporin 3. *Biochem Biophys Res  
1267 Commun* **485**: 627-634.
- 1268 Xu H, Xu Y, Zhang W, Shen L, Yang L, Xu Z. 2011. Aquaporin-3 positively regulates matrix metalloproteinases via  
1269 PI3K/AKT signal pathway in human gastric carcinoma SGC7901 cells. *J Exp Clin Cancer Res* **30**: 86.
- 1270 Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F et al.  
1271 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors.  
1272 *Science* **356**: eaaj2239.
- 1273 Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A et al. 2015.  
1274 Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of  
1275 COPD. *PLoS Genet* **11**: e1004898.
- 1276 Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A  
1277 et al. 2013. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *Plos One*  
1278 **8**.
- 1279 Zhang J, Wu L, Qu JM, Bai CX, Merrilees MJ, Black PN. 2012. Pro-inflammatory phenotype of COPD fibroblasts  
1280 not compatible with repair in COPD lung. *J Cell Mol Med* **16**: 1522-1532.
- 1281 Zhang XD. 2007. A pair of new statistical parameters for quality control in RNA interference high-throughput  
1282 screening assays. *Genomics* **89**: 552-561.
- 1283 Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, Crackower MA, Holder DJ, Heyse JF, Strulovici B. 2007.  
1284 The Use of Strictly Standardized Mean Difference for Hit Selection in Primary RNA Interference High-  
1285 Throughput Screening Experiments. *Journal of Biomolecular Screening* **12**: 497-509.

1286