

Cell-Specific Transcriptome of the COPD Alveolar Niche

Maor Sauler (✉ maor.sauler@yale.edu)

Yale University <https://orcid.org/0000-0001-5240-7978>

John McDonough

Yale School of Medicine

Taylor Adams

Yale University

Neeharika Kotahpalli

Yale School of Medicine

Jonas Schupp

Yale University <https://orcid.org/0000-0002-7714-8076>

Thomas Barnthaler

Yale School of Medicine

Matthew Robertson

Baylor College of Medicine

Cristian Coarfa Coarfa

Baylor College of Medicine <https://orcid.org/0000-0002-4183-4939>

Tao Yang

Yale School of Medicine

Mauricio Chioccioli

Yale School of Medicine

Norihito Omote

Yale School of Medicine

Carlos Cosme

Yale University School of Medicine

Sergio Poli

Mount Sinai Medical Center <https://orcid.org/0000-0001-5442-3189>

Ehab Ayaub

Brigham and Women's Hospital

Sarah Chu

Brigham and Women's Hospital

Klaus Jensen

Intomics

Pascal Timshel

Intomics

Jose Gomez

Yale University <https://orcid.org/0000-0002-6521-6318>

Clemente Britto

Yale University

Micha Sam Raredon

Yale University <https://orcid.org/0000-0003-1441-6122>

Laura Niklason

Yale University

Jessica Nouws

Yale School of Medicine

Naftali Kaminski

Yale University <https://orcid.org/0000-0001-5917-4601>

Ivan Rosas

Baylor College of Medicine

Article

Keywords: Chronic Obstructive Pulmonary Disease (COPD), translational research, pathobiology

Posted Date: March 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-276195/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on January 25th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-28062-9>.

1 **Characterization of the COPD Alveolar Niche Using Single-Cell RNA Sequencing.**

2
3 **Authors:** Maor Sauler^{1*}, John E. McDonough^{1*}, Taylor S. Adams¹, Neeharika Kothapalli¹, Jonas
4 C. Schupp¹, Jessica Nouws¹, Thomas Barnthaler^{1,2}, Matthew J. Robertson³, Cristian Coarfa³, Tao
5 Yang¹, Mauricio Chioccioli¹, Norihito Omote¹, Carlos Cosme Jr¹, Sergio Poli⁴, Ehab A. Ayaub⁵,
6 Sarah G. Chu⁵, Klaus H. Jensen⁶, Jose L. Gomez¹, Clemente J. Britto¹, Micha SB Raredon^{7,8},
7 Laura E. Niklason⁷, Pascal N. Timshel⁶, Naftali Kaminski^{1*}, Ivan O. Rosas^{3*}

8
9 **Affiliations:**

10 ¹ Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA

11 ² Division of Pharmacology, Otto Loewi Research Center, Medical University of Graz, Austria

12 ³ Pulmonary, Critical Care and Sleep Medicine, Baylor College of Medicine, Houston, TX, USA

13 ⁴ Department of Internal Medicine, Mount Sinai Medical Center, Miami, FL, USA

14 ⁵ Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard
15 Medical School, Boston, MA, USA

16 ⁶ Intomics A/S, Lyngby, Denmark

17 ⁷ Department of Biomedical Engineering, Yale University, New Haven, CT, USA

18 ⁸ Medical Scientist Training Program, Yale School of Medicine, New Haven, CT, USA

19 * Contributed equally to this work

20
21 **Corresponding Authors:** 1) Maor Sauler MD, 300 Cedar Street, New Haven, CT 06519, email:
22 maor.sauler@yale.edu phone: 203-737-2486 2) John E. McDonough PhD, 300 Cedar Street,
23 New Haven, CT 06519, email: john.e.mcdonough@yale.edu. Phone: 203-785-4162

24
25 **Author Contribution:** M.S. and J.M. analyzed data, generated figures, and wrote the manuscript.
26 T.S.A., N.Koth., J.C.S, C.Cos, and M.J.R. analyzed data. J.N., T.B., T.Y., M.C., and N.O.,
27 performed experiments. S.P., E.A.E., and S.G.C. procured biospecimens and clinical data. K.H.J.
28 and P.N.T. developed algorithms and performed data analysis. J.L.G. and C.J.B. provided clinical
29 insight and edited manuscript. M.S.B.R., L.E.N., K.H.J., C. Coa., and P.N.T. developed analytic
30 approaches and/or supervised data analysis. N.Kam. and I.O.R. conceived project, supervised
31 data analysis, and edited paper. All authors assisted in manuscript preparation and provided final
32 approval of the submitted work.

33
34 **Funding:** This work was supported by NIH grants K08HL135402, Flight Attendant Medical
35 Research Institute, and Claude D. Pepper Older Americans Independence Center, Yale School
36 of Medicine to M.S. NIH grants R01HL127349, RO1HL141852, U01HL145567, UH3TR002445 to
37 N.K, NHLBI P01 HL114501 and support from the Pulmonary Fibrosis Fund to I.O.R., an
38 unrestricted gift from Three Lake Partners to I.O.R and N.K, Department of Defense (W81XWH-
39 19-1-0131) and German Research Foundation (SCHU 3147/1) to J.C.S. R01 HL138540 to L.E.N.;
40 and an unrestricted research gift from Humacyte Inc. to L.E.N.

41 **Short Running Header:** single cell RNA sequencing of COPD lung

42
43
44
45
46
47
48
49
50

51 **ABSTRACT**

52

53 Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of death worldwide. To
54 identify cell-specific mechanisms underlying COPD pathobiology, we analysed single-cell RNA
55 sequencing (scRNAseq) profiles of explanted lung tissue from subjects with advanced COPD or
56 control lungs. Findings were validated with scRNAseq of lungs from mice exposed to 10 months
57 of cigarette smoke (CS), isolated human alveolar epithelial cells, and immunostaining of human
58 lung tissue samples. We identified a subpopulation of alveolar epithelial type II cells with
59 transcriptional evidence for aberrant cellular metabolism and reduced cellular stress tolerance,
60 exemplified by decreased expression of the stress-response gene *NUPR1*. Network analyses
61 identified an important role for inflamed capillary endothelial cells in COPD, particularly through
62 CXCL-motif chemokine signalling. Finally, we detected a metallothionein expressing macrophage
63 subpopulation unique to COPD. Collectively, these findings highlight cell-specific mechanisms
64 involved in the pathobiology of advanced COPD.

65

66 **INTRODUCTION**

67

68 Chronic Obstructive Pulmonary Disease (COPD) is characterized by persistent
69 inflammation and parenchymal tissue destruction, particularly in advanced disease stages¹. While
70 COPD is commonly caused by aerosolized pollutants such as cigarette smoke (CS), the clinical
71 and biologic effects of CS are heterogenous. Therefore, complex interactions between genetic
72 factors, host responses, and environmental exposures underlie COPD pathogenesis.

73 Our understanding of COPD pathogenesis has grown over the past decades through the
74 use of murine models, genome-wide association studies(GWAS), and expression analyses using
75 bulk human lung tissue²⁻⁵. The pathogenic cascade of COPD is initiated by repetitive insults to
76 epithelial and endothelial cells within the distal airways and alveolar niche. This engenders
77 activation of diverse cellular processes including immune cell infiltration, extracellular matrix
78 proteolysis, cellular metabolic dysfunction, loss of proteostasis, DNA damage, autophagy, cellular
79 senescence, and activation of regulated cell death pathways^{3,6,7}. Consequently, there is an
80 inability to maintain alveolar homeostasis, leading to chronic inflammation and alveolar septal
81 destruction, particularly in advanced COPD³. However, detailed knowledge of cell type-specific
82 mechanisms and the complex interactions amongst multiple lung cell types in COPD is lacking.

83 Recent studies have used single-cell RNA sequencing (scRNAseq) to obtain single-cell
84 resolution and identify novel disease mechanisms, cellular phenotypes, and changes in alveolar
85 niche crosstalk⁸⁻¹⁴. Herein, we analyzed scRNAseq profiles of parenchymal tissue obtained from
86 explanted lungs of patients with advanced COPD requiring lung transplant and control donor
87 lungs. We focused our analysis on three cell types commonly implicated in COPD pathogenesis:
88 epithelial cells, endothelial cells, and alveolar macrophages. Among epithelial cells, we identified
89 a subpopulation of *HHIP*-expressing alveolar epithelial type II (AT2) cells that mediate genetic
90 susceptibility to COPD and have aberrant expression of metabolic, antioxidant, and cellular stress
91 response genes in COPD. Analyses of endothelial cells suggested capillary CXLC-motif
92 chemokine signaling is an important cause of alveolar inflammation in COPD. Finally, we
93 identified a previously undescribed subpopulation of metallothionein-expressing alveolar
94 macrophages in the COPD lung.

95

96

97 RESULTS

98
99 We reanalyzed scRNAseq profiles that we previously obtained from explanted lung tissue
100 parenchyma¹². Our analysis focused on 17 patients with advanced COPD, and 15 age-matched
101 controls (**Figure 1A**). Demographics and pulmonary function test results are shown in
102 **Supplemental Table 1**. There were eight females in both groups and the median age of all
103 subjects was 62 years old (range 41–80). All COPD subjects had radiographic evidence of
104 advanced emphysema and were former smokers; four of the donors were either current or former
105 smokers. The final dataset consisted of 49,976 cells from control lungs and 61,564 cells from
106 COPD lungs with 37 distinct cell types identified in both control and COPD lungs based on
107 representative marker genes (**Figures 1B**). Canonical markers of identified cell types are shown
108 (**Figure 1C–E**). Data was previously deposited in the Gene Expression Omnibus (GSE136831)
109 and can be explored using our online portal (www.copdcellatlas.com). To validate findings, we
110 performed scRNAseq of isolated lung cells from SftpcCre^{ERT2}-m^Tm^G (AT2 cell) reporter mice
111 exposed to 10 months of cigarette smoke (CS) (**Figure 1A**). The final dataset for this analysis
112 consisted of 19,311 cells from 4 CS-exposed mice (2 male and 2 female) and 20,410 cells from
113 room air (RA)-exposed mice (2 male and 2 female). Cell identities were manually annotated using
114 previously described cellular markers (**Supplemental Figure 1**). Our scRNAseq findings in
115 alveolar epithelial cells were independently validated using flow-sorted epithelial cells from single
116 cell suspensions of 10 subjects with advanced COPD and 16 controls¹⁵. Additional validation was
117 carried out using a publicly available microarray dataset generated by the Lung Genomics
118 Research Consortium (LGRC) (GSE47460)^{16,17}.

119 120 **Analysis of epithelial cells in COPD reveals an HHIP⁺ AT2 subpopulation enriched for** 121 **genes known to genetically associated with COPD:**

122 Because of their importance in COPD pathobiology, we focused our evaluation on
123 epithelial cells in our human scRNAseq dataset. We identified all major epithelial cell types
124 including alveolar type I (AT1) and II (AT2) cells, ciliated cells, goblet cells, and club cells (**Figure**
125 **2A**). Amongst these major cell types, there were no significant cell proportion differences between
126 COPD and control lung tissue (**Figure 2B**). We also identified 33 recently described aberrant
127 basaloid cells amongst 9 subjects with COPD (1 – 10 per subject)¹². AT2 cells clustered into two
128 groups with analogous characteristics to recently described AT2 “bulk” and AT2 alveolar epithelial
129 progenitor (AEP) cell populations¹⁸⁻²⁰, and with expression profile similar to those recently
130 described in the scRNAseq analysis of the normal human lung by Travaglini *et al* (**Figure 2C**)²¹.
131 Both AT2 clusters expressed *SFTPC*, but one cluster (hereon called AT2_s) expressed genes and

132 regulons (imputed transcription factors) implicated in AEP function including EGF family receptor
133 (ERBB4) and WNT signaling (TNIK, TCF7L2, LRP5) (**Figure 2C, Supplemental Figure 2**). The
134 other AT2 cluster (hereon called AT2_B) has increased expression of classic AT2 markers including
135 ETV5 and SFTPA1. We confirmed SFTPA1 expression in only a subset of *SFTPC*⁺ cells using *in*
136 *situ* hybridization (**Supplemental Figure 3**).

137 AT2_B cells had the largest number of differentially expressed genes (DEGs) between
138 control and advanced COPD amongst all epithelial cells (**Figure 2D**). AT2_B cells uniquely
139 expressed HHIP, one of the most well-associated genes in COPD GWAS which has also been
140 demonstrated to modulate susceptibility to oxidative stress in murine models of CS exposure²²⁻²⁴.
141 AT2_B cells were also major expressors of other commonly identified COPD GWAS² genes
142 including SERPINA1^{25,26} and SFTPD²⁷ (**Supplemental Figure 4**). A key role for AT2_B cells as
143 mediators of COPD heritability was further supported by the use of CELL-type Expression-specific
144 integration for Complex Traits (CELLECT)²⁸. CELLECT integrates scRNAseq data with GWAS
145 data to prioritize cell types enriched for the expression of genes with disease-associated
146 polymorphism. Using CELLECT, we colocalized our scRNAseq findings with summary statistics
147 from two UKBiobank GWAS of COPD (defined as FEV₁/FVC < 70) and lung function (continuous
148 FEV₁/FVC variable)^{29,30}. AT2_B cells had the greatest enrichment amongst epithelial cells for the
149 expression of genes with polymorphisms associated with COPD-related traits (**Figure 2E**). We
150 also identified high enrichment scores in other epithelial and stromal cell populations, particularly
151 smooth muscle cells, but not immune cells.

152

153 **Altered expression of metabolic and cellular stress response genes in COPD AT2_B cells:**

154 We focused our subsequent epithelial cell analysis on AT2_B cells because of the high
155 enrichment for COPD-related genetic variants, the unique expression of *HHIP*, and the large
156 number of DEGs. We identified decreased expression of 182 genes and increased expression of
157 35 genes (**Figure 3A**) in AT2_B cells in COPD. Decreased DEGs included multiple genes
158 associated with the electron transport chain including subunits of complex I (NDUFAB1, NDUF4A,
159 NDUF58), complex III (UQCRCQ, UQCRCB), complex IV (COX8A, COX7B, COX6B1, COX5B
160 COX7C, COX6C) and ATP synthase (ATP5MD, ATP5F1D, ATP5F1E) . There were multiple
161 decreased DEGs encoding antioxidant genes that function in extracellular (SOD3), cellular
162 (GSTO1, GSTP1, TXNDC17), and mitochondrial (MGST1, MGST3) compartments. We also
163 observed reduced expression of S100A family proteins and genes that can mitigate toxin-induced
164 stress (AKR1A1, ALDH2). Amongst increased DEGs were genes related to oxidative stress and
165 apoptosis previously implicated in COPD pathogenesis, including PTPN1³¹ and EGR1³².

166 Comparison to DEGs between RA- and CS-exposed mouse AT2 cells identified two overlapping
167 genes, NUPR1 and CD74. NUPR1, a stress response gene^{33,34}, had the largest decrease in AT2_B
168 expression in COPD (fold change (fc)=0.35) and was also decreased in AT2 cells from mice
169 exposed to CS (fc = 0.51) (**Figure 3B, C**). CD74, the receptor for the cytokine MIF³⁵, was
170 increased in both human AT2_B (fc= 1.82) and mouse AT2 cells (fc = 1.30) (**Supplemental Figure**
171 **5**). We independently validated decreased *NUPR1* expression in human AT2 epithelial cells (fc =
172 0.70) obtained by fluorescence-activated cell sorting (EpCAM^{high}/PDPN⁻)¹⁵ of single-cell
173 suspensions from COPD and control patients (**Figure 3D**). NUPR1 expression obtained through
174 microarray measurements of gene expression in bulk lung tissue from the LGRC cohort of 208
175 individuals with COPD also inversely correlated with percent radiographic emphysema
176 (Spearman $\rho = -0.167$) (**Figure 3E**). We inhibited *NUPR1* to determine its functional relevance in
177 an alveolar epithelial cell line using silencing RNA (**Supplemental Figure 6**) and found a marked
178 increase in cell death susceptibility in response to cigarette smoke extract (CSE) (fc = 2.84)
179 (**Figure 3F,G**). Collectively, these findings highlight aberrant expression of key metabolic and
180 stress response genes in the AT2_B cell population that may contribute to the pathogenesis of
181 COPD.

182

183 **Endothelial cells exhibit enhanced inflammatory gene expression patterns in COPD:**

184 We identified all major endothelial populations that have been previously described
185 including arterial, venous, lymphatic, and systemic peri-bronchial endothelial cells, as well as two
186 recently described populations of capillary types, aerocytes and general capillaries (gCaps)^{36,37}
187 (**Figure 4A**). Aerocytes uniquely express CA4 and HPGD and are specialized for gas exchange
188 and leukocyte trafficking, while endothelial gCaps uniquely express FCN3 and promote vascular
189 homeostasis by functioning as progenitor cells, modulating vasomotor tone, and regulating
190 immune activation. No significant differences in the proportions of endothelial cells between
191 disease and control were detected (**Supplemental Figure 7**). Across all endothelial cell
192 populations, DEGs increased in COPD were enriched for multiple GO pathways including cellular
193 responses to cytokines (FDR = 2.49×10^{-11}), cellular responses to stress (FDR = 2.03×10^{-8}), and
194 cytokine signaling pathways (FDR = 2.52×10^{-7}), while DEGs decreased in COPD were enriched
195 blood vessel development (FDR = 4.4×10^{-4}). We observed many DEGs overlapping across
196 multiple endothelial cell types (**Figure 4B**), including increased expression of AP-1 subunit genes
197 (FOS, FOSB, JUND) and decreased expression of genes related to angiogenesis (ID1, ID3,
198 LDB2) (**Figure 4C**). Capillary endothelial cells had the largest number of DEGs and revealed an
199 enhanced inflammatory response with common DEGs involving inflammatory signaling (CX3CL1,

200 IL32), cellular stress responses (GADD45B), and vesicular trafficking (WASHC2C, WASHC2A),
201 and decreased expression of genes that promote endothelial repair (SEMA6A, WNT2B). While
202 *TNFRSF10D* and *IRF1* were specifically increased in aerocytes, *TNFAIP3*, *IFI6*, and *IL6* were
203 specifically increased in gCaps. These findings demonstrate previously unrecognized unique and
204 overlapping features of cellular stress and inflammation amongst capillary endothelial cells in
205 advanced COPD.

206

207 **Network analyses identify a key role for endothelial CXCL signaling in COPD:**

208 We proceeded to identify capillary endothelial cells as major contributors to alveolar
209 inflammation in COPD using lung connectome analyses¹¹. Here, we generated network-level
210 maps of cell-cell signaling across 24 alveolar and immune cell types based on computational
211 assessments of predicted ligand-receptor interactions (**Figure 5A**). Each node represents a cell
212 population and internodal edges reflect nondirected ligand-receptor interactions. We calculated
213 network centrality metrics and edge weights to quantify changes in alveolar signaling topology.
214 Fibroblasts, AT1, and AT2 cells had high degrees of connectivity within the network as reflected
215 by measurements of Kleinberg centrality which prioritize cell types responsible for incoming
216 (“authority”) and outgoing (“hub”) cell-cell signaling³⁸. In contrast, B and T cells had the lowest
217 Kleinberg centrality scores in both control and COPD lung. To further explore changes in cellular
218 communication we assessed ligand-receptor interactions that were preassigned to canonical
219 signalling pathways, which allowed us to identify changes in specific cell-cell signaling pathways
220 between control and COPD. We identified changes in outgoing or incoming edge weight amongst
221 cell types with the top 3 Kleinberg centrality scores in 14 out of 22 pathways (minimum $fc > 0.3$)
222 (**Supplemental Figure 8, Supplemental Table 3**). The largest change in edge weight was
223 increased outgoing gCap CXCL-motif signaling ($fc = 234.6$) (**Figure 5B**). This increase in outgoing
224 gCap CXCL signaling was also observed in CS-exposed mice (**Supplemental Figure 9,**
225 **Supplemental Table 3**). In both mouse and human connectomes, we found the increase in
226 outgoing gCap CXCL signaling was predominantly due to increased gCap *CXCL12* and multiple
227 *CXCL12*-interacting molecules including increased *CXCR3* and *CXCR4* (**Figure 5C**). Collectively,
228 these findings suggest a novel potential role for capillary endothelial cells as key mediators of
229 alveolar inflammation in COPD.

230

231 **A population of alveolar macrophages with increased expression of metallothioneins and** 232 **HMOX1 is observed in COPD:**

233 Analysis of immune cell populations revealed significant differences in cell compositions
234 amongst interstitial macrophages (fc=0.48; FDR = 2.7×10^{-3}), plasmacytoid dendritic cells
235 (fc=3.46; FDR = 2.7×10^{-3}), conventional dendritic cells (fc=2.79; FDR =0.025), and mast cells
236 (fc=7.70; FDR = 2.7×10^{-3}) (**Supplemental Figure 10**). The largest population of cells were
237 alveolar macrophages which have been implicated in alveolar inflammation and tissue destruction
238 in COPD³⁹. To characterize the heterogeneity of alveolar macrophages, we re-embedded the
239 alveolar macrophage cluster in UMAP space and re-clustered these cells into eight clusters
240 (**Figure 6A**). We then compared the relative abundance of the eight clusters between control and
241 COPD and identified changes in alveolar macrophage population composition between COPD
242 and control; cluster-0 macrophages were enriched amongst controls while cluster-5 macrophages
243 were enriched amongst patients with COPD (FDR <0.05) (**Figure 6B**). The corresponding cluster
244 markers for cluster-5 and cluster-0 are shown in **Figure 6C**. The top cluster-5 markers were
245 metallothioneins (*MT1G*, *MT1X*, *MT1E*, *MT2A*, *MT1M*, *MT1F*, *MT1H*, *MT1A*, *MT1L*) suggesting
246 an enrichment of macrophages expressing this class of cysteine-rich metal-binding antioxidant
247 proteins as well as *HMOX1*, a target gene of the antioxidant transcription factor NRF2 previously
248 implicated in COPD^{40,41}. We confirmed increased *MT2A* expression in a subset of cells co-
249 expressing the phagocyte cell marker CD68, particularly in COPD lung tissue samples (**Figure**
250 **6D, Supplemental Figure 11**). There were also multiple DEGs between COPD and control
251 macrophages that were not represented in any specific cluster. The most significant DEGs were
252 associated with macrophages chemotaxis and inflammation. These include THBS (fc=1.66),
253 PELI1 (fc=1.55), and CDC42 (fc=1.34) (**Figure 6E, Supplemental Figure 12**). These results
254 highlight changes in macrophage population composition, including the presence of a
255 metallothionein expressing alveolar macrophage population in advanced COPD, as well as DEGs
256 across multiple macrophages subpopulations.

257

258

259

260 **DISCUSSION**

261
262 COPD pathogenesis involves diverse and heterogenous biologic processes that vary
263 across cell types, but culminate in the loss of alveolar homeostasis, ultimately causing chronic
264 inflammation and tissue destruction. The extent to which pathologic COPD-related mechanisms
265 manifest in specific cell types remains uncertain. In this study, we analyzed single-cell
266 transcriptional profiles of lung alveolar cells in advanced COPD. We identified transcriptional
267 evidence for altered bioenergetics and impaired cellular stress tolerance in AT2_B cells, including
268 decreased *NUPR1* expression. We found endothelial cells had increased inflammatory
269 expression profiles and our lung connectome analyses suggested the largest change in alveolar
270 signaling in COPD is increased outgoing gCap CXCL-chemokine signaling. We also identified a
271 distinct metallothionein-expressing alveolar macrophage subpopulation enriched in COPD.
272 Together, these findings provide single cell resolution of COPD pathobiology within the alveolar
273 niche.

274 Our findings highlight a previously unrecognized role for the newly described AT2_B cell
275 population in COPD. Within these cells, we identified aberrant mitochondrial metabolism and
276 redox regulation as evidenced by decreased expression of genes encoding electron transport
277 chain complexes I, III, and IV, and ATP synthetase, as well as decreased expression of genes
278 mitigating oxidative stress. Previous studies of COPD pathogenesis have demonstrated alveolar
279 epithelial mitochondria are dysmorphic, have aberrant function, biogenesis, mitophagy,
280 mitochondrial fission, and mitochondrial fusion⁴²⁻⁴⁴. Such changes impair ATP production and
281 increase mitochondrial reactive oxygen species generation, and in the setting of reduced
282 antioxidant production, promote hallmark COPD pathobiological findings including oxidative
283 stress, cellular senescence, regulated cell death, and tissue destruction⁶. Therefore, it is not
284 surprising that we also identified increased AT2_B expression of genes associated with oxidative
285 stress and apoptosis (*PTPR1*, *EGR1*) previously implicated in COPD pathogenesis. While
286 aberrant bioenergetics, impaired redox regulation, and increased cellular injury are well-
287 associated with COPD pathogenesis, our analyses localized these findings predominantly to
288 AT2_B. Additionally, we found a novel role for *NUPR1*, a cellular stress response transcriptional
289 factor that can mitigate susceptibility to DNA damage, apoptosis and cellular senescence. The
290 role of *NUPR1* in non-malignant lung disease has not been previously described, however
291 *NUPR1* promotes chemoresistance in multiple cancer cell lines and protects against toxin-
292 induced renal epithelial injury^{34,45}. We found a dramatic increase in epithelial cell death in
293 response to CSE upon *NUPR1* depletion, consistent with a growing understanding that the
294 pathobiology of COPD involves impaired cellular stress^{39,46}. While the specific form of regulated

295 cell death will require further investigation, the collective findings suggest altered bioenergetics,
296 reduced antioxidant gene expression, and reduced NUPR1 expression may contribute to impaired
297 AT2_B cells homeostasis in advanced COPD.

298 Additional support for the important role of AT2_B cells in COPD comes from our CELLECT
299 analysis that indicated that among epithelial cells, AT2_B cells were most enriched for the
300 expression of genes with COPD-associated variants. Amongst those genes was HHIP which was
301 primarily expressed in AT2_B cells. HHIP is a gene that has been repeatedly and robustly
302 associated with COPD-related traits in GWAS²³, and previous studies using haploinsufficient *Hhip*
303 mice demonstrated *Hhip* protected against oxidative stress, alterations in cellular metabolism,
304 and emphysema^{24,47}. In adult mice, *Hhip* is predominantly expressed in a subset of fibroblasts
305 rather than AT2 cells⁴⁸, which is a limitation of the findings in mice. Additionally, one has to
306 consider transient HHIP expression in other human lung cell types during development or in
307 response to injury may also influence COPD susceptibility^{49,50}. Yet, the finding of HHIP in AT2_B
308 cells together with the results of the CELLECT analysis and the gene expression changes
309 mentioned above should draw attention to the role of AT2_B cells as a key cell type involved in
310 COPD pathobiology. Future studies will be necessary to dissect the detailed relationship between
311 COPD genetics and altered AT2_B bioenergetics and stress responses, either through HHIP or
312 other genes with COPD-associated variants expressed AT2_B cells, including SERPINA1 and
313 SFTPD1. In addition to AT2_B cells, our CELLECT analysis found enrichment for genetic variants
314 associated with decreased lung function amongst other epithelial and stromal cells but not
315 immune cells. This suggests aberrant immunologic phenotypes in COPD are largely acquired
316 rather than genetically programmed. Therefore, despite the inflammatory nature of COPD,
317 targeting epithelial (particularly AT2_B cells) and stromal cells in at-risk individuals with early COPD
318 may be an effective therapeutic strategy.

319 We also found a novel role for endothelial cells as key contributors to persistent alveolar
320 inflammation in advanced COPD. Endothelial injury is a well-established pathogenic mechanism
321 of advanced COPD⁵¹. Serologic and radiographic evidence of endothelial injury is associated with
322 emphysema severity and there is increased endothelial senescence and apoptosis in the
323 emphysematous lung^{51,52}. In mice, emphysema can be induced by inhibiting VEGF signaling,
324 which is critical for endothelial homeostasis⁵³, and multiple studies have shown endothelial injury
325 is necessary and sufficient to cause emphysema⁵⁴. Our transcriptome analysis provides further
326 evidence for reduced endothelial homeostasis in advanced COPD as evidenced by reduced
327 expression of endothelial maintenance factors and increased endothelial expression of cellular
328 stress response genes and AP-1 transcription factors. However, the most intriguing finding was

329 the marked increase in capillary endothelial inflammatory. While most studies have focused on
330 the interaction between epithelial and immune cells as key drivers of persistent inflammation⁵⁵,
331 we identified increased outgoing gCap CXCL-motif chemokine signaling as the largest change in
332 cell-cell communication in the lung with advanced COPD. Increased CXCL chemokine signaling
333 has been previously implicated in COPD pathogenesis, and a recent study demonstrated that
334 AMD3100, a drug that inhibits the CXCL12 receptor CXCR4, can ameliorate CS-induced
335 emphysema in a mouse model, but did not identify the predominant source(s) of CXCR4 ligands
336 in COPD⁵⁶. Our analysis prioritizes gCap endothelial cells as a key source of CXCL12 with the
337 largest increase in COPD. These data support the possibility of an expanded role for capillary
338 endothelial cells in COPD pathogenesis as a major source of pro-inflammatory signaling and
339 suggests targeting outgoing gCap CXCL12 signaling may represent a novel therapeutic approach
340 to this disease.

341 In contrast to solely identifying phenotypic shifts amongst endothelial cells with advanced
342 COPD, we observed both phenotypic shifts and population composition changes amongst
343 alveolar macrophages. A subset of macrophages in control lungs was enriched for heat shock
344 proteins while lungs from subjects with advanced COPD were enriched with a cluster expressing
345 metallothioneins. Increased metallothionein expression was also identified in COPD endothelial
346 cells. Metallothioneins are induced by oxidative stress and inflammation, and protect against
347 cellular injury by sequestering intracellular metals such as zinc and copper⁵⁷. Despite their
348 homeostatic function, little is known about the consequences of chronic metallothionein
349 upregulation in the setting of advanced COPD where redox homeostasis and heavy metal
350 metabolism are disrupted⁵⁸. For instance, Menkes diseases, a congenital disorder of copper
351 deficiency, causes emphysema⁵⁹. Dysregulated zinc homeostasis can cause impaired
352 phagocytosis and an abnormal inflammatory response in macrophages⁶⁰. Therefore,
353 understanding metallothionein regulation of zinc and copper in COPD may improve our
354 understanding of disease pathogenesis.

355 Despite the use of independent validations for observed findings in the human scRNAseq
356 data, there are a few limitations to consider. First, COPD is a heterogenous disorder, yet the
357 subjects in this study represented a distinct subtype of COPD patients with advanced disease
358 requiring transplant and who were not actively smoking; therefore, our findings cannot be
359 generalized to all COPD sub-phenotypes or earlier disease stages. Additionally, while all subjects
360 had radiographic emphysema, we do not know the degree of emphysema within the tissue sample
361 obtained from a given subject. However, given the large number of subjects included in this study,
362 the spectrum of pathologic changes in the lung parenchyma with advanced emphysema were

363 likely captured in our analysis. While we identified two AT subpopulations in human lungs,
364 Louvain clustering did not identify two AT2 cell populations in mice. Other limitations are common
365 to scRNAseq experiments, including dissociation bias due to variable cellular sensitivities to
366 digestive enzymes and differences in cellular embedding in the extracellular matrix. However, all
367 samples were processed using previously validated protocols¹⁵. Additionally, transcriptional
368 changes are informative but do not always reflect protein concentration or function. Finally, *in*
369 *silico* findings using CELLECT and connectome analyses will require further validation.

370 Collectively, our findings provide a high-resolution single cell atlas of the COPD lung. This
371 atlas highlights previously unrecognized changes in gene expression and cellular interactions in
372 distinct epithelial, endothelial and macrophage cell populations in COPD. Highlighting the
373 complexity and diversity of cellular injury and inflammation in COPD, and the potential importance
374 of injury to lung epithelial and stromal cells populations in COPD pathogenesis. Future studies
375 evaluating outcomes of specific transcriptional dysregulations identified herein will provide novel
376 insights into mechanisms that contribute to disease.

377

378 **METHODS**

379 *Human tissue sample scRNAseq:* We reanalyzed scRNAseq of parenchymal lung tissue
380 previously published by Adams *et al*, focusing on individuals with and without COPD and
381 excluding samples from individuals with pulmonary fibrosis¹². One COPD sample was excluded
382 due to a reported history of no cigarette smoke exposure. In order to age-match our control and
383 COPD samples, we excluded control samples from individuals < 40 years of age. Study protocols
384 were approved by Partners Healthcare Institutional Board Review (IRB Protocol 2011P002419).
385 Tissue procurement, sample processing, and single-cell sequencing methods have been
386 previously described^{12,15}. Briefly, explanted lungs were procured from donors with end-stage lung
387 disease undergoing transplant or rejected control donor lungs. Lung tissue was digested and
388 cryopreserved. Single cell barcoding of thawed samples and complementary DNA (cDNA) library
389 preparation was performed according to the manufacturer's protocol (Single Cell 3' Reagent Kits
390 v2, 10x Genomics, USA). Quality control was maintained using an Agilent Bioanalyzer High
391 sensitivity DNA chip. The cDNA libraries were sequenced on a HiSeq 4000 Illumina platform, with
392 a goal of 150 million reads per library and a sequencing configuration of 26 base pairs on read1
393 and 98 base pairs on read2. Full deidentified sequencing data for all subjects is available in the
394 gene expression omnibus (GEO) under accession number GSE136831. Base call files were
395 demultiplexed into FASTQ files using the mkfastq pipeline in Cell Ranger's (v3.0.2) mkfastq
396 pipeline. The Read2 files were trimmed using cutadapt (v2.7) and reads shorter than 20bp were
397 removed. Read processing was conducted with zUMIs v2.0 pipeline and trimmed reads were
398 aligned to the GRCh38 release 91 (GRCh38.p12) using STAR (v2.6.0c). We removed barcoded
399 cells with <12% of transcripts arriving from unspliced mRNA, cells with <1000 transcripts profiled,
400 and cells with >20% of their transcriptome of mitochondrial origin. Expression values were
401 normalized to 10,000 transcripts per cell and log transformed using a pseudocount of 1.

402 *Mouse scRNAseq:* Animal protocols were approved by the Animal Care and Use Committee at
403 Yale University. We obtained Sftpc-CreER^{T2} and Rosa26-mTmG C57Bl/6 mice from Jackson
404 Laboratories and bred them together to generate Sftpc-Cre/Rosa26-mTmG mice. Sftpc-Cre mice
405 were previously described by Rock *et al.*⁶¹ and Rosa26-m^{TmG} mice were previously described by
406 Muzumdar *et al.*⁶². Male and female 8-10-week-old mice received tamoxifen (T5648; Sigma-
407 Aldrich) (20mg/mL stock solution in corn oil) at 150 mg/kg x 4 days given via i.p. injection. Two
408 weeks later, littermates were randomly assigned to begin exposure to room air or cigarette smoke
409 from 3R4F research cigarettes (University of Kentucky) in a Teague TE-10 smoking machine
410 (Teague Enterprises) for 6 hours of exposure per day, 5 days/week for 10 months. One day after

411 the last exposure, single-cell suspensions were obtained by placing right lungs from PBS perfused
412 mice in digestion media (DMEM containing 1mg/ml Collagenase/Dispase (Roche), 3U/mL of
413 Elastase (Worthington), and 20 U/mL of DNAase (Qiagen) and incubating at 37°C for 45 minutes.
414 Digested lung tissue was meshed through a 100 µm cell strainers using a plunger and
415 resuspended in 20 mL ice-cold DMEM + 10% FBS. ACK lysis buffer was used to removed red
416 blood cells. To enrich for non-immune cells, single cell suspensions were MACS sorted on a LS
417 column following incubation with CD45+ microbeads for 15 min at 4°C per protocol. Samples were
418 filtered through 40 µm filter, and viable cells were counted using a Countess II automated cell
419 counter (Thermo Fisher). Cell populations were reconstituted to achieve a final concentration of
420 10⁶ cells/ml consisting of 10% CD45⁺ and 90% CD45⁻ populations. Single cell barcoding and
421 complementary DNA (cDNA) library preparation was performed according to the manufacturer's
422 protocol (Single Cell 3' Reagent Kits v3, 10x Genomics, USA). Briefly, cell suspensions, beads,
423 master mix, and portioning oil were loaded on to single cell "A" chip for a targeted output of 10,000
424 cells per library and run on the Chromium Controller. Reverse Transcription was performed at
425 53°C for 45 min and cDNA was amplified for 12 cycles using a BioRad C1000 Touch thermocycler.
426 We performed cDNA size selection using SpriSelect beads (Beckman Coulter, USA) and cDNA
427 quality was confirmed with an Agilent Bioanalyzer High Sensitivity DNA chip. DNA fragmentation,
428 end-repair, A-tailing, and ligation of sequencing adapters were performed per manufacturer's
429 protocol (10x Genomics, USA). The cDNA libraries were sequenced on a HiSeq 4000 Illumina
430 platform aiming for 150 million reads per library and a sequencing configuration of 28 base pairs
431 on read1 and 98 base pairs on read2. Base call files were demultiplexed into FASTQ files using
432 the mkfastq pipeline in Cell Ranger's (v3.0.2) Adaptor contamination
433 (AAGCAGTGGTATCAACGCAGAGTACATGGG 10x 3prime samples and 20bps or longer
434 poly(A) sequences were removed using cutadapt (v2.9). and reads shorter than 25bp were
435 removed. Read processing was performed using STAR (v2.7.3a). and aligned to the mouse
436 reference genome GRCm38 release M22 (GRCm38.p6) downloaded from GENCODE.⁶³ A
437 modified genome index for the STAR alignment was created by adding the nucleotide sequence
438 information of the transgenes *eGFP* and *tdTomato*, to which sequencing reads were then aligned.
439 Collapsed unique molecular identifiers (UMIs) with reads originating from spliced as well as
440 unspliced RNA were retained. We removed barcoded cells with <7.5% of transcripts arriving from
441 unspliced mRNA, cells with <1000 transcripts profiled, and cells with >5% of their transcriptome
442 of mitochondrial origin. Background contamination from cell free mRNA was removed using
443 SoupX software (v1.2.2)⁶⁴.

444 *Clustering, differential cell expression, and cell population composition:* Clustering and differential
445 analysis of cells were performed using the Seurat package (v.3.2.0) in R¹². Expression values
446 were normalized to 10,000 transcripts per cell and log transformed using a pseudocount of 1.
447 Regression during scaling was performed to adjust for percent mitochondrial genes. Louvain
448 clustering was used to group cells and cell type clusters were identified using canonical marker
449 genes. Overall marker genes for each cell type were identified by applying the Seurat
450 FindAllMarkers implementation of the Wilcoxon rank sum test or by calculating the diagnostics
451 odds ratio (DOR) for each gene per cell type as previously described¹². Cells annotated as
452 “Multipler” were removed prior to downstream analyses. Differentially expressed genes between
453 COPD and control cells were identified by using the FindMarkers function test in Seurat, with
454 statistical test and adjustment for multiple comparison testing as described in the methods and
455 figure legend. For endothelial cells, Gene Ontology analysis was performed with AmiGO^{65,66}
456 (Fisher Exact Test with FDR correction) using 203 unique increased DEGs and 167 unique
457 decreased DEGs across all endothelial subtypes ($p < 0.05$, Wilcoxon rank sum with Bonferroni
458 correction). For endothelial cells, UpsetR plots were generated using the UpsetR package as
459 previously described⁶⁷.

460 *Regulon Analysis for identifying transcription factor regulators of gene expression:* Regulons,
461 transcription factors predicted to regulate gene expression, were identified using pySCENIC
462 (v.0.10.2) with default settings and databases: cisTarget databases (hg38__refseq-
463 r80__500bp_up_and_100bp_down_tss.mc9nr.feather, hg38__refseq-
464 r80__10kb_up_and_down_tss.mc9nr.feather), and the transcription factor motif annotation
465 database (motifs-v9-nr.hgnc-m0.001-o0.0.tbl) were downloaded from
466 resources.aertslab.org/cistarget/, and the list of human transcription factors (hs_hgnc_tfs.txt)
467 was downloaded from github.com/aertslab/pySCENIC/tree/master/resources⁶⁸.

468 *CELL-type Expression-specific integration for Complex Traits (CELLECT):* We used CELLECT
469 to quantify associations between cell-type specificity of expressed genes and findings from
470 genome-wide association studies (GWAS) of lung function and COPD^{29,30}. CELLECT has been
471 previously described²⁸. Briefly, CELLECT generates a genetic prioritization scores for each
472 gene based on cell type-expression specificity and GWAS summary statistics. As input to
473 CELLECT, we used summary statistics derived from genetic studies of UK Biobank data for
474 presence of COPD²⁹ and lung function³⁰. First, cellular expression specificity scores were
475 calculated for each gene using CELLEX [CELL-type EXpression-specificity] (v.1.2.1) using
476 recommended normalization method and preprocessing steps (common transcript count

477 normalization log-transformation). Using CELLEX, an expression specificity likelihood ($ES\mu$)
478 was then computed for each gene across different cell types. We then integrated these results
479 with summary statistics derived from genetic studies of UK Biobank data for presence of COPD
480 and lung function. In these studies, COPD was defined using pre-bronchodilator spirometry
481 according to modified Global Initiative for Chronic Obstructive Lung Disease criteria. We ran
482 CELLECT with the recommended workflow (CELLECT-LDSC) and default parameters (100 kb
483 window size around each gene).

484 *Lung connectome to identify cell-cell interactions:* Methods to generate the lung connectome have
485 been previously described¹¹. Average expression values for every gene within cell types were
486 calculated and mapped against the FANTOM5 database of known ligand-receptor pairs to create
487 a global connectome using the R software Connectome (v0.2.2)
488 (<https://msraredon.github.io/Connectome/>). Nodes were defined as cellular clusters. A directed
489 edge was created connecting two nodes if >5% of cells within the two cell types expressed the
490 cognate molecules of the ligand-receptor pair. Directionality was therefore defined as outgoing
491 signals from ligands and incoming signals to receptors. In general, edge weights represent the
492 product of the average expression values of the ligand and receptor within their respective cell
493 types, and cumulative edge weights are defined as the sum of the weights of all edges connecting
494 pairs of cell types (i.e. nodes). Kleinberg hub and authority scores are metrics of outgoing and
495 incoming centrality, respectively, that take into account the number and weights of edges
496 connecting a node in a network; as such, a large hub score represents a node (cell cluster) that
497 is highly connected or “central” to a network because it sends many outgoing signals with large
498 edge-weights, and a large authority score represents a node that receives many incoming signals
499 with large edge weights. Kleinberg hub and authority scores were calculated for each node using
500 the igraph package in R. The connectome was filtered to edges between predefined cell types for
501 the analysis. For Figure 5A, we used scaled gene expression values to define the edge weights
502 in which the weight of each edge was calculated as the product of average z-scores of the ligand
503 and receptor within their respective clusters. A discussion about the use of scaled vs. unscaled
504 expression values within the connectome analysis is further detailed by Raredon *et al*¹¹. We
505 plotted non-directional cumulative edge weights in the network graphs. For the pathway centrality
506 analysis, the global connectome was filtered to ligand-receptor pairs that were preassigned to
507 specific signaling modes. Edge weights were calculated using average unscaled gene expression
508 values. Cumulative outgoing and incoming edge weights were computed for each node within
509 every signaling mode; these values were scaled by mode and direction of signaling. Kleinberg

510 hub and authority scores were also computed for each signaling mode. The Durbin test was used
511 to assess global differences in signaling between control and COPD for each mode. A differential
512 connectome was generated by computing the log-fold change of normalized expression values
513 of the ligand and receptor for each edge between control and COPD. A perturbation score was
514 then computed as the absolute value of the product of these fold changes for each edge. Ligand-
515 receptor interactions for CXCL signaling were visualized in circos plots using the R package
516 *circize* after filtering the differential connectome based on the following criteria: the outgoing cell-
517 type being gCaps; both ligands and receptors being increased in COPD; ligands and receptors
518 being expressed in at least 5% of the cells of their respective cell types; and omitting edges with
519 perturbation scores less than 0.10.

520 *RNA sequencing of isolated AT2 cells:* The isolation of AT2 cells from cryopreserved single cell
521 suspensions of lung tissues samples used in this study has been previously described¹⁵. Briefly,
522 cryopreserved single cell suspensions of explanted lung tissue were obtained in the same manner
523 as described for scRNAseq above. Cells were stained using the following antibodies or isotype-
524 matched antibodies as negative controls: PE anti-human CD326 (EpCAM) (eBioscience), FITC
525 anti-human CD45 (BD bioscience 340664), Alexa Fluor® 647 anti-human podoplanin (PDPN)
526 (BioLegend), and DAPI (BioLegend) or propidium iodide. Cells were sorted using a BD FACSAria
527 II (BD Biosciences). EpCAM^{high}/PDPN⁻ sorted cells were enriched for AT2 cell markers and used
528 for further RNA sequencing. Library construction for RNA sequencing was performed as
529 previously described using the Illumina TRuSeq RNA Access Library Prep kit (San Diego, CA) for
530 library preparation and sequencing was performed on a 75 bp paired end flowcell using a HiSeq
531 2500 System. Each lane was spiked with 5% PhiX control libraries. Fastq files were trimmed using
532 TrimGalore! (v0.6.6), before aligning to the human genome (GRCh38 p13) using the STAR aligner
533 (2.7.5c). After alignment genes were quantified using featureCounts (v2.0.1). Samples were
534 filtered to remove genes with low expression across all samples before normalization using the
535 trimmed to means method. Differential gene expression was determined using generalized linear
536 models, and after fitting differential expression was determined using a quasi-likelihood F-test. In
537 the generalized linear models, the sequencing lane was treated as a blocking variable and the
538 phenotype (COPD or Control) was treated as the main predictor.

539 *LGRC cohort:* Cyclic loess normalized *NUPR1* gene expression was measured at the probe-level
540 (A_24_P270728) from the Affymetrix Human Gene 1.0 ST Array (Affymetrix) in 208 patients with
541 COPD in the previously described LGRC cohort^{5,17}.

542 *NUPR1 siRNA and apoptosis assay:* A549 cells (American Type Culture Collection) were cultured
543 in DMEM supplemented with 10% heat inactivated fetal bovine serum (FBS). The cells were
544 passaged < 20 times. RNA duplexes for silencing NUPR1 (siNUPR1) and non-targeting control
545 were obtained from Dharmacon. Cells were transfected in 12-well plates using RNAiMAX
546 transfection reagent (Life Technologies) and OptiMEM media according to manufacturer's
547 protocols. Cells were re-transfected after 72 hours and used for experimentation 24-48 hours after
548 the second transfection, and after being incubated with 0% or 8% CSE for 20 hours. For CSE,
549 mainstream smoke from one 3RF4 research cigarette (University of Kentucky, Lexington,
550 Kentucky) was suctioned through 10 mL of cell culture media and filtered using a 0.22 µm filter
551 (MilliporeSigma) as previously described, with the obtained filtrate considered 100% CSE⁶⁹. Flow
552 cytometry for Annexin V and propidium iodide was performed per the manufacturer's protocol (BD
553 Biosciences, 556547) using a Cytoflex LX flow cytometer and data was analyzed using Flow Jo
554 10.6 software.

555 *CD68 and MT2A Immunofluorescence:* Paraffin embedded blocks of lung tissue samples were
556 deparaffinized in xylene and decreasing concentrations of ethanol in distilled water. They were
557 then placed in EDTA pH 9 epitope retrieval buffer at 95°C for 30 min then cooled, rinsed, and
558 washed with TBS with 0.1% tween. Slides were incubated in blocking buffer (Agilent) for 30
559 minutes. The primary antibodies applied overnight at 4 deg C were mouse anti-MT2A
560 (MAB10176) (R&D systems) and rabbit anti-CD68 (PA5-83940) (Invitrogen). Slides were washed
561 with TBS with 0.1% tween and then incubated for 1 hour at room temperature with secondary
562 antibodies donkey anti-rabbit Alexa-488 (Thermo Fisher) and donkey anti-mouse Alexa-555
563 (Invitrogen). Slides were washed with TBS with 0.1% tween and coverslip was mounted using
564 anti-fade mounting media with DAPI (Vectashield). Images were acquired with Nikon eclipse Ti.

565 *RNA in situ hybridization (RNA-ISH):* RNA in situ hybridization was performed using BaseScope
566 (Advanced Cell Diagnostics (ACD), Newark, CA). Human lungs were fixed with 10% Neutral
567 buffered formalin for 18-32hr. Lungs were paraffin embedded, and 5 um tissue sections were
568 mounted on slides. Slides were heated for 1 h at 60°C, deparaffinized in xylene rehydrated with
569 decreasing concentrations of ethanol in distilled water. Sections were treated with hydrogen
570 peroxide (ACD 322381) for 10 min at room temperature, then heated to mild boil (98-102°C) in
571 1x target retrieval reagent buffer (ACD 322001) for 15 min. Protease III (ACD 322340) was
572 applied to sections for 20 min at 40°C in HybEZ Oven (ACD). Hybridization with target probe (SPA
573 891331), preamplifier, amplifier, labels and wash buffer (ACD 323910) was done according to
574 ACD instructions. Parallel sections were incubated with negative control probes. Subsequent

575 immunodetection of pro-SPC was performed as previously described⁷⁰. After visualization with
576 FastRed detection reagent, slides were washed and blocked in 10% goat serum in PBS and Triton
577 X 0.3%. Incubation with primary antibody (rabbit anti pro-SPC, AB3786, Millipore, 1:500) was
578 performed overnight at 4°C, followed by detection with goat anti rabbit fluorescein antibody
579 (Invitrogen, 31635 1:500). To reduce autofluorescence, sections were incubated with TrueView
580 reagent (Vectorlabs, SP-8500) for 3 minutes and mounted. Images were acquired with a Zeiss
581 LSM710 confocal microscope equipped with a 100x oil immersion objective and analysis was
582 performed using ImageJ.

583 *Statistics:* Statistical analyses, adjustment for multiple comparison testing, biological replicates,
584 and p-values are indicated in the results, methods, and figure legends. We used two-sided
585 statistical tests for all comparisons as appropriate. All error bars are defined in figure legends. No
586 statistical methods were used to predetermine sample size. Littermates were randomly assigned
587 to treatment groups.

588

589 **Disclosures:** Dr. Kaminski reports personal fees from Biogen Idec, Boehringer Ingelheim, Third
590 Rock, Samumed, Numedii, Astra Zeneca, Life Max, Tervnce, RohBar, and Pliant. Equity in Pliant.
591 Collaboration with Miragen, Astra Zeneca. Grant from Veracyte, all outside the submitted work;
592 In addition, Dr. Kaminski has a patent New Therapies in Pulmonary Fibrosis, and a patent for
593 Peripheral Blood Gene Expression licensed to Biotech. Dr. Niklason reports grants from
594 Humacyte Inc., outside the submitted work. Ehab A. Ayaub and Sarah G. Chu Dr. Chu report
595 personal fees from Novartis Institutes of BioMedical Research, outside the submitted work. Klaus
596 H. Jensen and Pascal N. Timshel are employed by Intomics.

597 **FIGURE LEGENDS**

598

599 **Figure 1: Profiling of cell types in COPD using scRNAseq data.** (A) Tissue from 17 lungs with
600 advanced COPD and 15 control donor lungs were dissociated into single-cell suspensions for
601 barcoding and sequencing. Similarly, lung tissue from 2 male (M) and 2 female (F) mice exposed
602 to 10 months of cigarette smoke (CS) and 2 M and 2 F mice exposed only to room air (RA) was
603 dissociated into single-cell suspensions for barcoding and sequencing. (B) Uniform Manifold
604 Approximation and Projection (UMAP) representation of 111,540 single cells grouped into 37
605 distinct cell types (*left*) with identification of COPD and control cells (*right*). Violin plots of
606 normalized expression values for canonical cell-specific marker genes in (C) epithelial, (D)
607 endothelial, and (E) stromal cells. AT1=alveolar epithelial type I, AT2=alveolar epithelial type II,
608 PNEC=pulmonary neuroendocrine cells, SMC=smooth muscle cells, gCap=general capillary,
609 cMonocyte=classical monocytes, ncMonocyte=non-classical monocyte, Macs.=macrophages,
610 DC=dendritic cells, cDC=conventional dendritic cells, pDC=plasmacytoid dendritic cells, NK =
611 natural killer cells, ILC = innate lymphoid cells.

612

613 **Figure 2: COPD Epithelial Cells.** (A) UMAP of epithelial cells from COPD and control donor
614 lungs. Samples are color labelled by cell type (*left*) and disease category (*right*). Alveolar type II
615 (AT2) cells can be distinguished as two clusters denoted as AT2_S and AT2_B. (B) Distribution of
616 subject-specific epithelial cell types as a fraction of the total number of epithelial cells per disease
617 category. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. (C) Dot plot of z-
618 scores for marker gene expression values. Dot size reflects percentage of cells with gene
619 expression; color corresponds to degree of regulon activity. (D) Number of differentially expressed
620 genes between control and COPD across epithelial cells types with $p < 0.05$ using Wilcoxon rank
621 sum test with Bonferroni correction. (E) Plot of negative log adjusted p-values for cell type-
622 specific enrichment for GWAS-identified genes with polymorphisms associated with lung
623 function (continuous FEV₁/FVC) (*x-axis*) and presence of COPD (FEV₁/FVC < 70) (*y-axis*).

624

625 **Figure 3: Aberrant AT2_B Cellular Stress Response in COPD.** (A) Heatmaps of z-scores for
626 differentially expressed genes (yellow/purple) between control and COPD for AT2_B cells
627 (Wilcoxon rank sum test, Bonferroni corrected $p < 0.05$). Each column represents expression values
628 for an individual cell. Columns are hierarchically ordered by disease phenotype and subject, in

629 which disease category and individual subject are represented by unique colors. z-scores were
630 calculated across all epithelial cells. **(B)** *NUPRI* in AT2_B cells from human subjects from non-
631 smokers (n=11), former/current smokers without COPD (n=4), and former smokers with COPD
632 (n=17). **(C)** *Nupr1* in AT2 cells from mice exposed to room air for 10 months vs. mice exposed to
633 10 months of cigarette smoke (CS). **(D)** *NUPRI* in isolated AT2 cells from control and COPD lung
634 tissue. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. **(E)** Spearman
635 correlation of *NUPRI* with the square root (sqrt) of radiographic emphysema in lung tissue
636 samples from the LGRC cohort (n=208). **(F)** Flow cytometric detection of propidium iodide (*y*-
637 *axis*) and Annexin V (*x*-*axis*) cells treated with *NUPRI* silencing RNA (si*NUPRI*) vs. silencing
638 control RNA (siCTRL) and exposed to 0% or 8% cigarette smoke extract (CSE). **(G)** Percent cell
639 death determined by flow cytometry for Annexin V/ Propidium iodide in cells treated with
640 si*NUPRI* vs. siCTRL and exposed to 0% or 8% CSE (n= 5/group). n.s. = non-significant, **
641 p<0.001, *** p<0.0001 using Wilcoxon rank sum test with Bonferroni correction **(B,C)**,
642 unadjusted Wilcoxon rank sum test **(D)**, or one-way ANOVA with Tukey post-hoc test **(G)**.

643
644 **Figure 4: COPD endothelial cell types demonstrate universal and cell type-specific**
645 **transcriptional aberrations.** **(A)** UMAPs of all vascular endothelial (VE) and lymphatic
646 endothelial cells from control and COPD subjects. UMAPs are color labelled by cell type (*top*)
647 and disease status (*bottom*). **(B)** UpSet plot visualizing the properties of intersecting and unique
648 sets of differentially expressed (DE) genes between COPD and control amongst endothelial
649 (Wilcoxon rank sum test, unadjusted p < 0.001, minimal fold change > 0.5). **(C)** Heatmap of
650 corresponding differentially expressed genes between COPD and control amongst six subtypes of
651 endothelial cells. Each column represents expression values for an individual cell. Columns are
652 hierarchically ordered by endothelial subtype, disease phenotype, and then subject. Grey row (*top*):
653 expression values for marker genes are unity normalized between 0 and 1 across all endothelial
654 subtypes. Orange row (*middle*): z-scores of differentially expressed genes in three or more
655 endothelial cell types between control and COPD. Yellow row (*middle*): z-scores of differentially
656 expressed genes in both aerocytes and gCaps between control and COPD. Blue and green row
657 (*bottom*): z-scores of differentially expressed genes unique to aerocytes (*blue*) or gCaps (*green*).
658 Unity normalization and z-score calculations were performed using all endothelial subtypes.

659 **Figure 5: Alveolar niche networks and pathway centrality analyses.** (A) Network plots of the
660 alveolar niche in control (*left*) and COPD (*right*). Each node represents a cell population and each
661 internodal edge reflects ligand-receptor interactions between cell types. The edge weight
662 (thickness) between nodes reflects the sum of individual edge weights (non-directed) which are
663 based on the product of ligand-receptor gene expression values, while the size of the node reflects
664 measurements of Kleinberg centrality which prioritizes cell types responsible for incoming
665 (authority) and outgoing (hub) cell-cell signaling. Individual cell types are labelled by color and
666 number. (B) Centrality analysis of the alveolar connectome for CXCL signaling between control
667 and COPD. Dot size is proportional to the Kleinberg scores for each cell type within CXCL
668 signaling. Panel shows outgoing edge weights and Kleinberg hub scores (*left*) and incoming edge
669 weights and Kleinberg authority scores (*right*). Individual cell types are color labelled as in Figure
670 **5A**, and numbers shown identify cell types with the largest Kleinberg centrality score. ***
671 $p < 0.0001$ using the Durbin test to compare control and COPD across cell types. (C) Differential
672 circos plots for outgoing gCap CXCL signaling from human and mouse connectomes. Edge
673 thickness is proportional to perturbation scores, defined as the product of the absolute values of
674 the log-fold change for both the receptor and ligand. CXCL differential network analysis limited
675 to edges in which both ligand and receptor expression are increased.

676
677 **Figure 6: Changes in alveolar macrophage population composition in COPD.** (A) UMAPs of
678 control and COPD alveolar macrophage cells, color labelled by Louvain cluster (*left*) and disease
679 status (*right*). (B) Percent makeup of alveolar macrophage across all nine Louvain clusters per
680 subject, grouped by disease state. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x
681 IQR. The percent of cluster 0 and cluster 5 alveolar macrophages are different between control
682 and COPD (Wilcoxon rank sum test, FDR < 0.05). (C) Heatmap of the distribution of z-scores of
683 marker genes for cluster 0 and cluster 5 alveolar macrophages. Columns represent expression
684 values from individual cells and are hierarchically ordered by macrophages cluster, disease status,
685 and subject. (D) Immunofluorescence images of MT2A (*red*) expression in CD68+ cells (*green*)
686 in control and COPD lung tissue samples (*arrows*). Arrows point to colocalization of MT2A and
687 CD68 (*yellow*). Scale bar = 50 μm . (E) Violin plots of *THBS1*, *PEL11*, and *CDC42* gene expression
688 in alveolar macrophages in control and COPD subjects. *** $p < 0.0001$ using Wilcoxon rank sum
689 test with Bonferroni correction.

690 **REFERENCES**

- 691 1 Hogg, J. C. *et al.* The nature of small-airway obstruction in chronic obstructive
692 pulmonary disease. *N Engl J Med* **350**, 2645-2653, doi:10.1056/NEJMoa032158 (2004).
693 2 Silverman, E. K. Genetics of COPD. *Annu Rev Physiol* **82**, 413-431,
694 doi:10.1146/annurev-physiol-021317-121224 (2020).
695 3 Tudor, R. M. & Petrache, I. Pathogenesis of chronic obstructive pulmonary disease. *J*
696 *Clin Invest* **122**, 2749-2755, doi:10.1172/JCI60324 (2012).
697 4 Barnes, P. J. *et al.* Chronic obstructive pulmonary disease. *Nat Rev Dis Primers* **1**,
698 15076, doi:10.1038/nrdp.2015.76 (2015).
699 5 Kusko, R. L. *et al.* Integrated Genomics Reveals Convergent Transcriptomic Networks
700 Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis.
701 *Am J Respir Crit Care Med* **194**, 948-960, doi:10.1164/rccm.201510-2026OC (2016).
702 6 Sauler, M., Bazan, I. S. & Lee, P. J. Cell Death in the Lung: The Apoptosis-Necroptosis
703 Axis. *Annu Rev Physiol* **81**, 375-402, doi:10.1146/annurev-physiol-020518-114320
704 (2019).
705 7 Sauler, M. *et al.* The DNA repair transcriptome in severe COPD. *Eur Respir J* **52**,
706 doi:10.1183/13993003.01994-2017 (2018).
707 8 Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a
708 transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172, doi:10.1038/s41590-018-
709 0276-y (2019).
710 9 Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct
711 epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* **6**,
712 eaba1972, doi:10.1126/sciadv.aba1972 (2020).
713 10 Reyfman, P. A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides
714 Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med*,
715 doi:10.1164/rccm.201712-2410OC (2018).
716 11 Raredon, M. S. B. *et al.* Single-cell connectomic analysis of adult mammalian lungs. *Sci*
717 *Adv* **5**, eaaw3851, doi:10.1126/sciadv.aaw3851 (2019).
718 12 Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell
719 populations in idiopathic pulmonary fibrosis. *Science Advances* **6**, eaba1983,
720 doi:10.1126/sciadv.aba1983 (2020).
721 13 Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct
722 localization and relevance to fibrosis. *Nat Commun* **11**, 1920, doi:10.1038/s41467-020-
723 15647-5 (2020).
724 14 Vieira Braga, F. A. *et al.* A cellular census of human lungs identifies novel cell states in
725 health and in asthma. *Nat Med* **25**, 1153-1163, doi:10.1038/s41591-019-0468-5 (2019).
726 15 Chu, S. G. *et al.* Biobanking and cryopreservation of human lung explants for omic
727 analysis. *Eur Respir J* **55**, doi:10.1183/13993003.01635-2018 (2020).
728 16 Yang, I. V. *et al.* Relationship of DNA methylation and gene expression in idiopathic
729 pulmonary fibrosis. *Am J Respir Crit Care Med* **190**, 1263-1272,
730 doi:10.1164/rccm.201408-1452OC (2014).
731 17 Kim, S. *et al.* Integrative phenotyping framework (iPF): integrative clustering of multiple
732 omics data identifies novel lung disease subphenotypes. *BMC Genomics* **16**, 924,
733 doi:10.1186/s12864-015-2170-4 (2015).
734 18 Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved
735 epithelial progenitor. *Nature* **555**, 251-255, doi:10.1038/nature25786 (2018).
736 19 Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in
737 lung development, renewal and cancer. *Nature* **507**, 190-194, doi:10.1038/nature12930
738 (2014).

739 20 Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-
740 cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118-
741 1123, doi:10.1126/science.aam6603 (2018).

742 21 Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA
743 sequencing. *Nature* **587**, 619-625, doi:10.1038/s41586-020-2922-4 (2020).

744 22 Lao, T. *et al.* Haploinsufficiency of Hedgehog interacting protein causes increased
745 emphysema induced by cigarette smoke through network rewiring. *Genome Med* **7**, 12,
746 doi:10.1186/s13073-015-0137-3 (2015).

747 23 Zhou, X. *et al.* Identification of a chronic obstructive pulmonary disease genetic
748 determinant that regulates HHIP. *Hum Mol Genet* **21**, 1325-1335,
749 doi:10.1093/hmg/ddr569 (2012).

750 24 Lao, T. *et al.* Hhip haploinsufficiency sensitizes mice to age-related emphysema. *Proc*
751 *Natl Acad Sci U S A* **113**, E4681-4687, doi:10.1073/pnas.1602342113 (2016).

752 25 Strnad, P., McElvaney, N. G. & Lomas, D. A. Alpha1-Antitrypsin Deficiency. *N Engl J*
753 *Med* **382**, 1443-1455, doi:10.1056/NEJMra1910234 (2020).

754 26 Alam, S. *et al.* Z alpha1-antitrypsin confers a proinflammatory phenotype that contributes
755 to chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **189**, 909-931,
756 doi:10.1164/rccm.201308-1458OC (2014).

757 27 Wert, S. E. *et al.* Increased metalloproteinase activity, oxidant production, and
758 emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci U S A* **97**,
759 5972-5977, doi:10.1073/pnas.100448997 (2000).

760 28 Timshel, P. N., Thompson, J. J. & Pers, T. H. Mapping heritability of obesity by brain cell
761 types. *bioRxiv*, 2020.2001.2027.920033, doi:10.1101/2020.01.27.920033 (2020).

762 29 Sakornsakolpat, P. *et al.* Genetic landscape of chronic obstructive pulmonary disease
763 identifies heterogeneous cell-type and phenotype associations. *Nat Genet* **51**, 494-505,
764 doi:10.1038/s41588-018-0342-2 (2019).

765 30 Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic
766 obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**,
767 481-493, doi:10.1038/s41588-018-0321-7 (2019).

768 31 Geraghty, P. *et al.* The glutathione peroxidase 1-protein tyrosine phosphatase 1B-
769 protein phosphatase 2A axis. A key determinant of airway inflammation and alveolar
770 destruction. *Am J Respir Cell Mol Biol* **49**, 721-730, doi:10.1165/rcmb.2013-0026OC
771 (2013).

772 32 Chen, Z. H. *et al.* Egr-1 regulates autophagy in cigarette smoke-induced chronic
773 obstructive pulmonary disease. *PLoS One* **3**, e3316, doi:10.1371/journal.pone.0003316
774 (2008).

775 33 Santofimia-Castano, P. *et al.* Inactivation of NUPR1 promotes cell death by coupling ER-
776 stress responses with necrosis. *Sci Rep* **8**, 16999, doi:10.1038/s41598-018-35020-3
777 (2018).

778 34 Santofimia-Castano, P. *et al.* Ligand-based design identifies a potent NUPR1 inhibitor
779 exerting anticancer activity via necroptosis. *J Clin Invest* **129**, 2500-2513,
780 doi:10.1172/JCI127223 (2019).

781 35 Sauler, M. *et al.* Macrophage migration inhibitory factor deficiency in chronic obstructive
782 pulmonary disease. *Am J Physiol Lung Cell Mol Physiol* **306**, L487-496,
783 doi:10.1152/ajplung.00284.2013 (2014).

784 36 Gillich, A. *et al.* Capillary cell-type specialization in the alveolus. *Nature* **586**, 785-789,
785 doi:10.1038/s41586-020-2822-7 (2020).

786 37 Vila Ellis, L. *et al.* Epithelial Vegfa Specifies a Distinct Endothelial Population in the
787 Mouse Lung. *Dev Cell* **52**, 617-630 e616, doi:10.1016/j.devcel.2020.01.009 (2020).

788 38 Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *J Acm* **46**, 604-632,
789 doi:Doi 10.1145/324133.324140 (1999).

790 39 Ueno, M. *et al.* Alendronate inhalation ameliorates elastase-induced pulmonary
791 emphysema in mice by induction of apoptosis of alveolar macrophages. *Nat Commun* **6**,
792 6332, doi:10.1038/ncomms7332 (2015).

793 40 Ryter, S. W. & Choi, A. M. Targeting heme oxygenase-1 and carbon monoxide for
794 therapeutic modulation of inflammation. *Transl Res* **167**, 7-34,
795 doi:10.1016/j.trsl.2015.06.011 (2016).

796 41 Lee, P. J., Alam, J., Wiegand, G. W. & Choi, A. M. Overexpression of heme oxygenase-
797 1 in human pulmonary epithelial cells results in cell growth arrest and increased
798 resistance to hyperoxia. *Proc Natl Acad Sci U S A* **93**, 10393-10398,
799 doi:10.1073/pnas.93.19.10393 (1996).

800 42 Mizumura, K. *et al.* Mitophagy-dependent necroptosis contributes to the pathogenesis of
801 COPD. *J Clin Invest* **124**, 3987-4003, doi:10.1172/JCI74985 (2014).

802 43 Zhang, X. *et al.* Cathepsin E promotes pulmonary emphysema via mitochondrial fission.
803 *Am J Pathol* **184**, 2730-2741, doi:10.1016/j.ajpath.2014.06.017 (2014).

804 44 Cloonan, S. M. *et al.* Mitochondrial iron chelation ameliorates cigarette smoke-induced
805 bronchitis and emphysema in mice. *Nat Med* **22**, 163-174, doi:10.1038/nm.4021 (2016).

806 45 Galichon, P. *et al.* Stress Response Gene Nupr1 Alleviates Cyclosporin A Nephrotoxicity
807 In Vivo. *J Am Soc Nephrol* **28**, 545-556, doi:10.1681/ASN.2015080936 (2017).

808 46 Tuder, R. M. Bringing Light to Chronic Obstructive Pulmonary Disease Pathogenesis
809 and Resilience. *Ann Am Thorac Soc* **15**, S227-S233, doi:10.1513/AnnalsATS.201808-
810 583MG (2018).

811 47 Wan, E. S. *et al.* Metabolomic profiling in a Hedgehog Interacting Protein (Hhip) murine
812 model of chronic obstructive pulmonary disease. *Sci Rep* **7**, 2504, doi:10.1038/s41598-
813 017-02701-4 (2017).

814 48 Xie, T. *et al.* Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary
815 Fibrosis. *Cell Rep* **22**, 3625-3640, doi:10.1016/j.celrep.2018.03.010 (2018).

816 49 Peng, T. *et al.* Hedgehog actively maintains adult lung quiescence and regulates repair
817 and regeneration. *Nature* **526**, 578-582, doi:10.1038/nature14984 (2015).

818 50 Wang, C. *et al.* Expansion of hedgehog disrupts mesenchymal identity and induces
819 emphysema phenotype. *J Clin Invest* **128**, 4343-4358, doi:10.1172/JCI99435 (2018).

820 51 Casanova, C. *et al.* Microalbuminuria and hypoxemia in patients with chronic obstructive
821 pulmonary disease. *Am J Respir Crit Care Med* **182**, 1004-1010,
822 doi:10.1164/rccm.201003-0360OC (2010).

823 52 Barr, R. G. *et al.* Percent emphysema, airflow obstruction, and impaired left ventricular
824 filling. *N Engl J Med* **362**, 217-227, doi:10.1056/NEJMoa0808836 (2010).

825 53 Kasahara, Y. *et al.* Inhibition of VEGF receptors causes lung cell apoptosis and
826 emphysema. *J Clin Invest* **106**, 1311-1319, doi:10.1172/JCI10259 (2000).

827 54 Petrache, I. *et al.* Ceramide upregulation causes pulmonary cell apoptosis and
828 emphysema-like disease in mice. *Nat Med* **11**, 491-498, doi:10.1038/nm1238 (2005).

829 55 Barnes, P. J. Inflammatory mechanisms in patients with chronic obstructive pulmonary
830 disease. *J Allergy Clin Immunol* **138**, 16-27, doi:10.1016/j.jaci.2016.05.011 (2016).

831 56 Barwinska, D. *et al.* AMD3100 ameliorates cigarette smoke-induced emphysema-like
832 manifestations in mice. *Am J Physiol Lung Cell Mol Physiol* **315**, L382-L386,
833 doi:10.1152/ajplung.00185.2018 (2018).

834 57 Coyle, P., Philcox, J. C., Carey, L. C. & Rofe, A. M. Metallothionein: the multipurpose
835 protein. *Cell Mol Life Sci* **59**, 627-647, doi:10.1007/s00018-002-8454-2 (2002).

836 58 Rahman, I. & MacNee, W. Oxidant/antioxidant imbalance in smokers and chronic
837 obstructive pulmonary disease. *Thorax* **51**, 348-350, doi:10.1136/thx.51.4.348 (1996).

838 59 Mizuno, S. *et al.* Copper deficiency induced emphysema is associated with focal
839 adhesion kinase inactivation. *PLoS One* **7**, e30678, doi:10.1371/journal.pone.0030678
840 (2012).

841 60 Gao, H. *et al.* Metal transporter Slc39a10 regulates susceptibility to inflammatory stimuli
842 by controlling macrophage survival. *Proc Natl Acad Sci U S A* **114**, 12940-12945,
843 doi:10.1073/pnas.1708018114 (2017).

844 61 Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without
845 evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A* **108**, E1475-
846 1483, doi:10.1073/pnas.1117988108 (2011).

847 62 Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent
848 Cre reporter mouse. *Genesis* **45**, 593-605, doi:10.1002/dvg.20335 (2007).

849 63 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
850 *Nucleic Acids Res* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).

851 64 Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-
852 based single-cell RNA sequencing data. *Gigascience* **9**,
853 doi:10.1093/gigascience/giaa151 (2020).

854 65 Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics*
855 **25**, 288-289, doi:10.1093/bioinformatics/btn615 (2009).

856 66 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene
857 Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

858 67 Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of
859 intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940,
860 doi:10.1093/bioinformatics/btx364 (2017).

861 68 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat*
862 *Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

863 69 Nouws, J. *et al.* MicroRNA miR-24-3p reduces DNA damage responses, apoptosis, and
864 susceptibility to chronic obstructive pulmonary disease. *JCI Insight* **6**,
865 doi:10.1172/jci.insight.134218 (2021).

866 70 Barnthaler, T. *et al.* Inhibiting eicosanoid degradation exerts antifibrotic effects in a
867 pulmonary fibrosis mouse model and human tissue. *J Allergy Clin Immunol* **145**, 818-833
868 e811, doi:10.1016/j.jaci.2019.11.032 (2020).

869

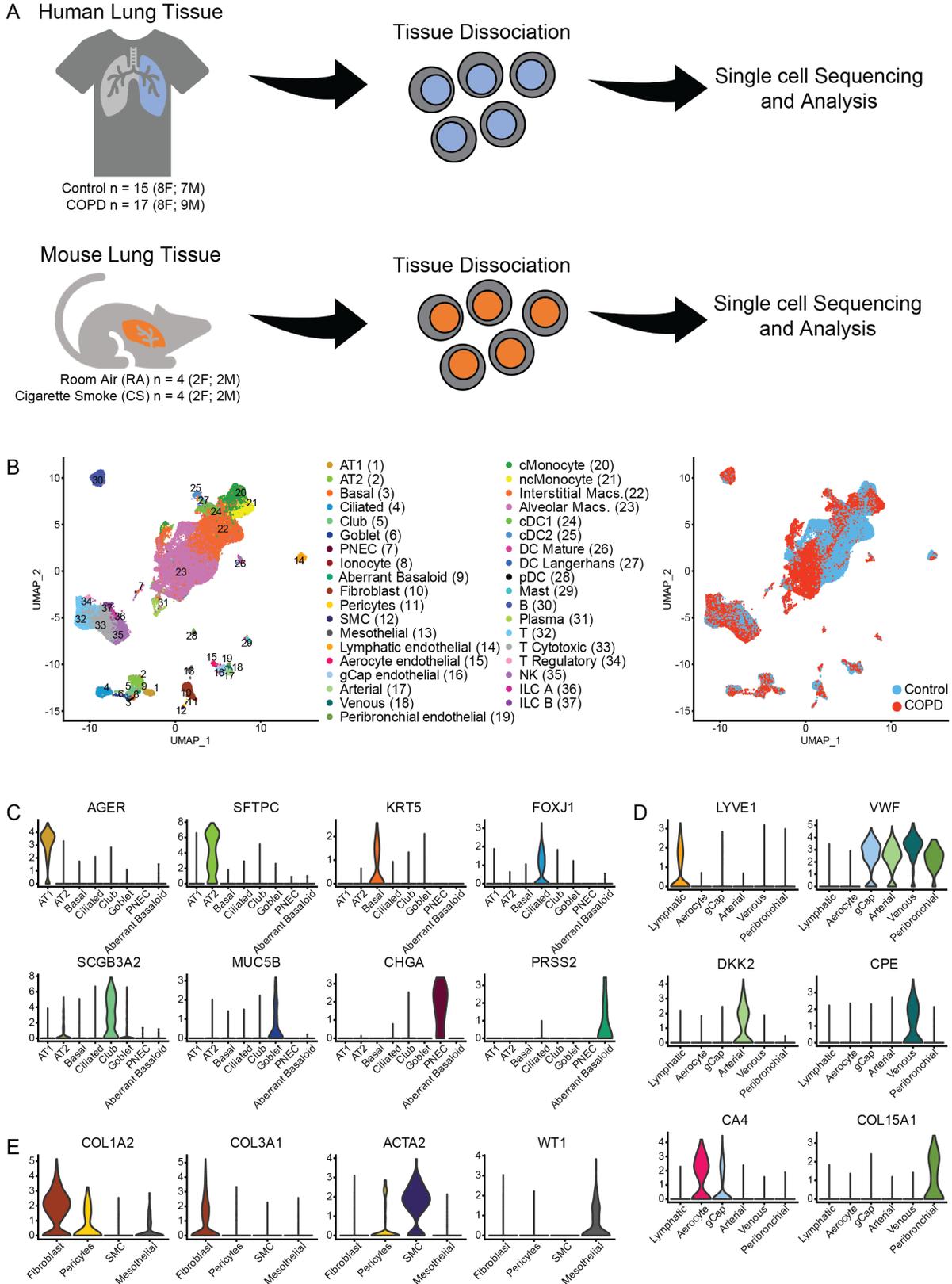


Figure 1: Profiling of cell types in COPD using scRNAseq data. (A) Tissue from 17 lungs with advanced COPD and 15 control donor lungs were dissociated into single-cell suspensions for barcoding and sequencing. Similarly, lung tissue from 2 male (M) and 2 female (F) mice exposed to 10 months of cigarette smoke (CS) and 2 M and 2 F mice exposed only to room air (RA) was dissociated into single-cell suspensions for barcoding and sequencing. (B) Uniform Manifold Approximation and Projection (UMAP) representation of 111,540 single cells grouped into 37 distinct cell types (left) with identification of COPD and control cells (right). Violin plots of normalized expression values for canonical cell-specific marker genes in (C) epithelial, (D) endothelial, and (E) stromal cells. AT1=alveolar epithelial type I, AT2=alveolar epithelial type II, PNEC=pulmonary neuroendocrine cells, SMC=smooth muscle cells, gCap=general capillary, cMonocyte=classical monocytes, ncMonocyte=non-classical monocyte, Macs.=macrophages, DC=dendritic cells, cDC=conventional dendritic cells, pDC=plasmacytoid dendritic cells, NK = natural killer cells, ILC = innate lymphoid cells.

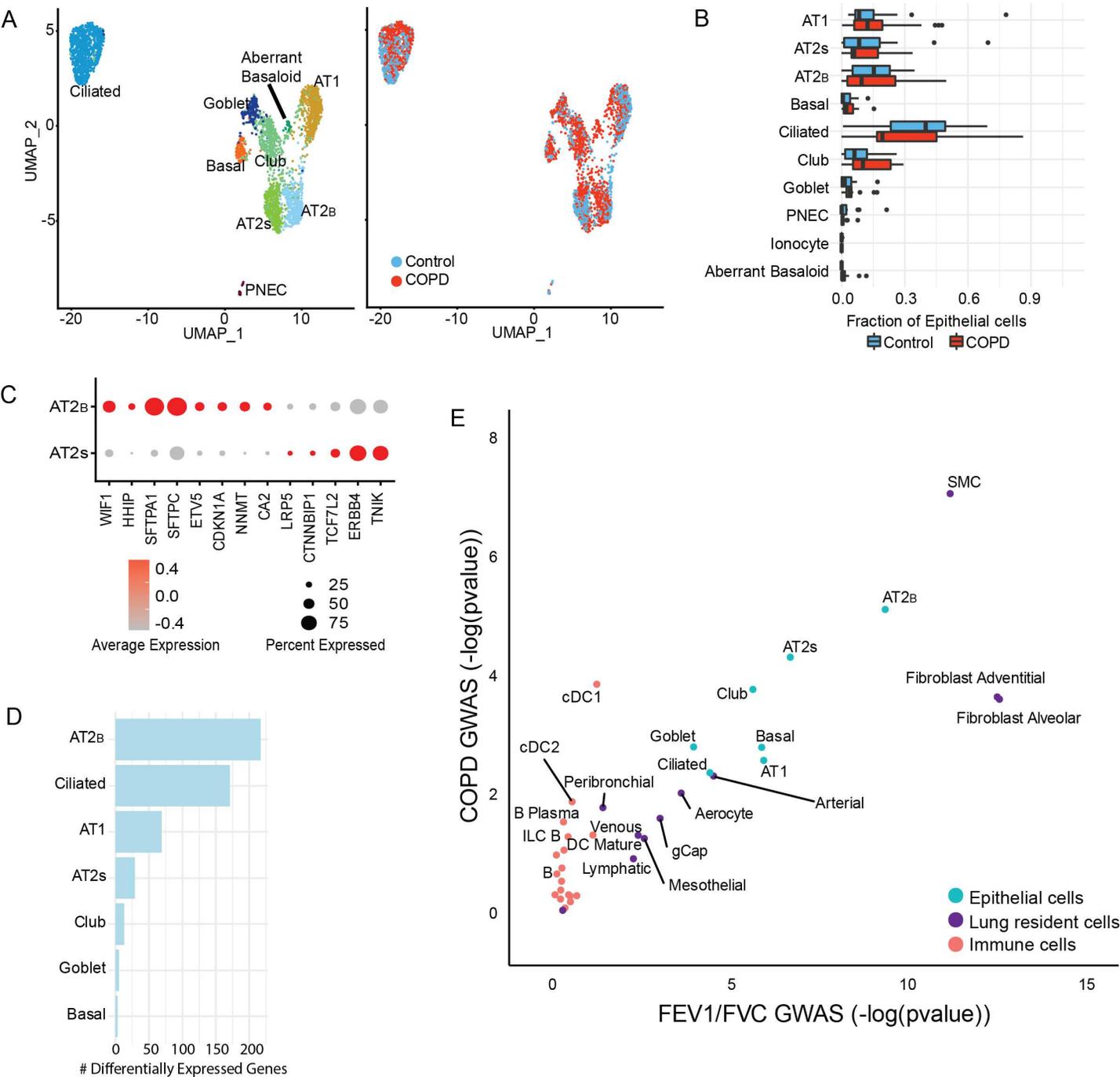


Figure 2: COPD Epithelial Cells. (A) UMAP of epithelial cells from COPD and control donor lungs. Samples are color labelled by cell type (*left*) and disease category (*right*). Alveolar type II (AT2) cells can be distinguished as two clusters denoted as AT2_S and AT2_B. (B) Distribution of subject-specific epithelial cell types as a fraction of the total number of epithelial cells per disease category. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. (C) Dot plot of z-scores for marker gene expression values. Dot size reflects percentage of cells with gene expression; color corresponds to degree of regulon activity. (D) Number of differentially expressed genes between control and COPD across epithelial cells types with $p < 0.05$ using Wilcoxon rank sum test with Bonferroni correction. (E) Plot of negative log adjusted p-values for cell type-specific enrichment for GWAS-identified genes with polymorphisms associated with lung function (continuous FEV₁/FVC) (x -axis) and presence of COPD (FEV₁/FVC < 70) (y -axis).

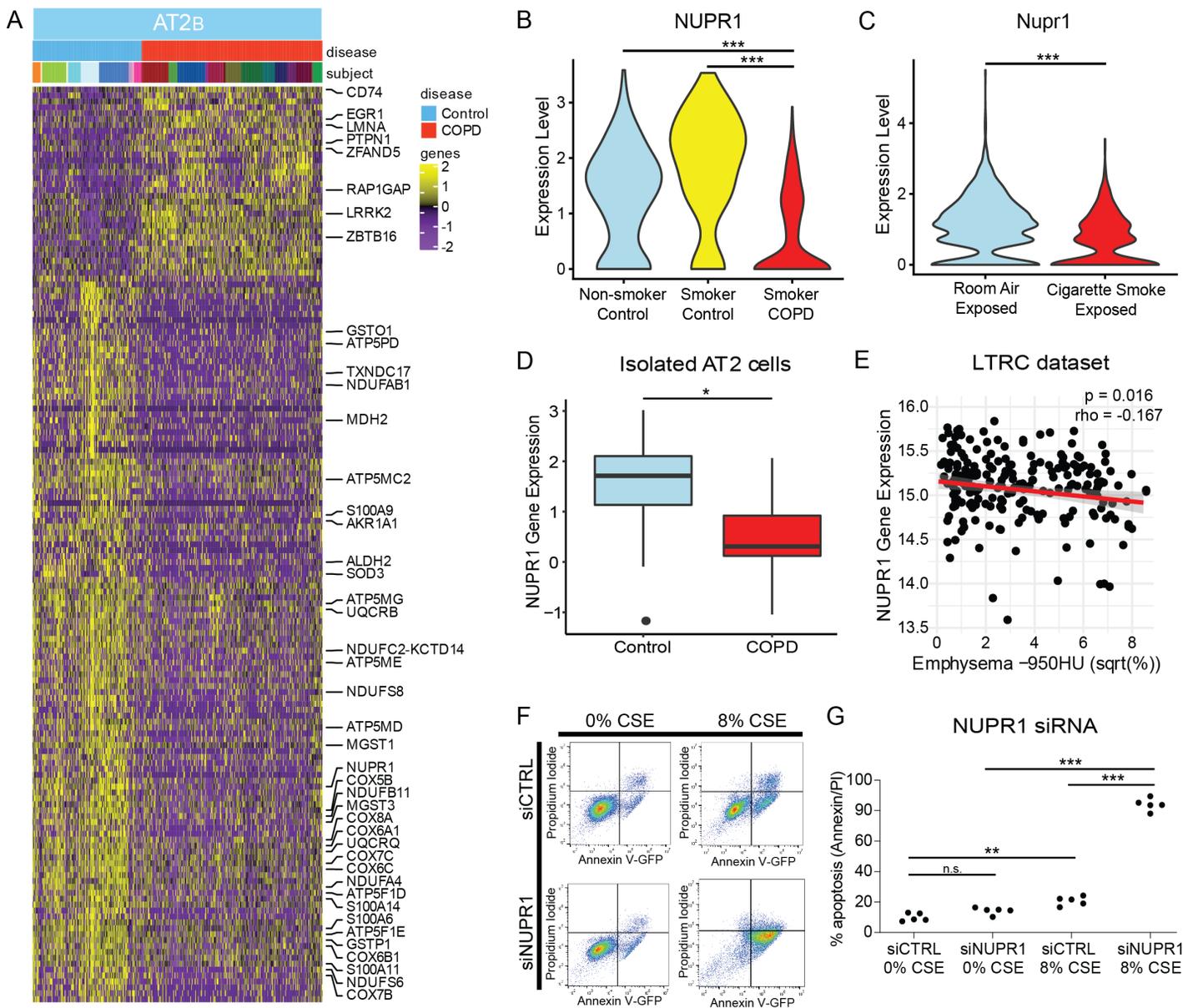
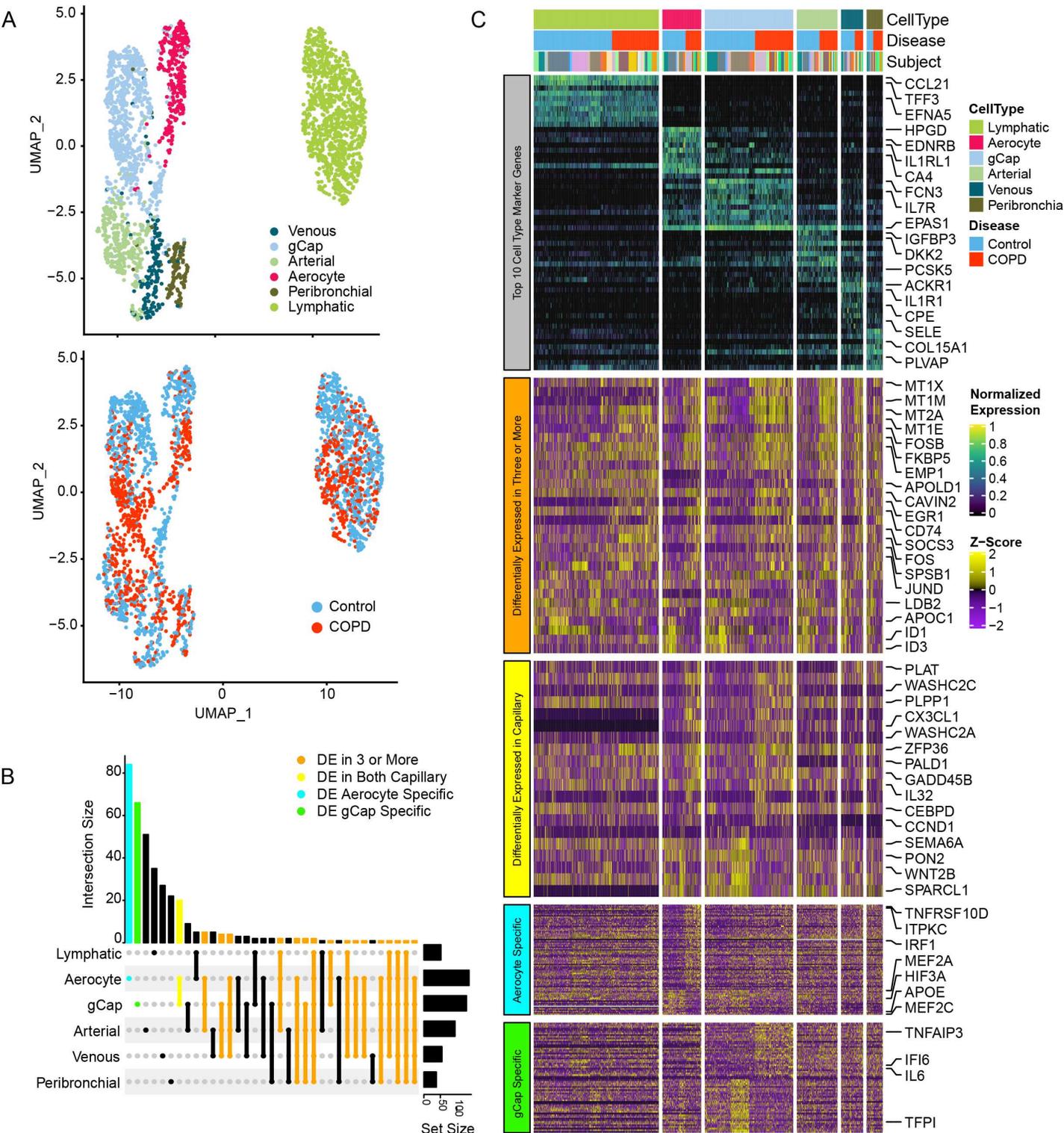


Figure 3: Aberrant AT2_b Cellular Stress Response in COPD. (A) Heatmaps of z-scores for differentially expressed genes (yellow/purple) between control and COPD for AT2_b cells (Wilcoxon rank sum test, Bonferroni corrected $p < 0.05$). Each column represents expression values for an individual cell. Columns are hierarchically ordered by disease phenotype and subject, in which disease category and individual subject are represented by unique colors. Z-scores were calculated across all epithelial cells. (B) *NUPR1* in AT2_b cells from human subjects from non-smokers ($n=11$), former/current smokers without COPD ($n=4$), and former smokers with COPD ($n=17$). (C) *Nupr1* in AT2 cells from mice exposed to room air for 10 months vs. mice exposed to 10 months of cigarette smoke (CS). (D) *NUPR1* in isolated AT2 cells from control and COPD lung tissue. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. (E) Spearman correlation of *NUPR1* with the square root (sqrt) of radiographic emphysema in lung tissue samples from the LGRC cohort ($n=208$). (F) Flow cytometric detection of propidium iodide (y-axis) and Annexin V (x-axis) in cells treated with NUPR1 silencing RNA (siNUPR1) vs. silencing control RNA (siCTRL) and exposed to 0% or 8% cigarette smoke extract (CSE). (G) Percent cell death determined by flow cytometry for Annexin V/ Propidium iodide in cells treated with siNUPR1 vs. siCTRL and exposed to 0% or 8% CSE ($n=5$ /group). n.s. = non-significant, ** $p < 0.01$, *** $p < 0.0001$ using Wilcoxon rank sum test with Bonferroni correction (B,C), unadjusted Wilcoxon rank sum test (D), or one-way ANOVA with Tukey post-hoc test (G).



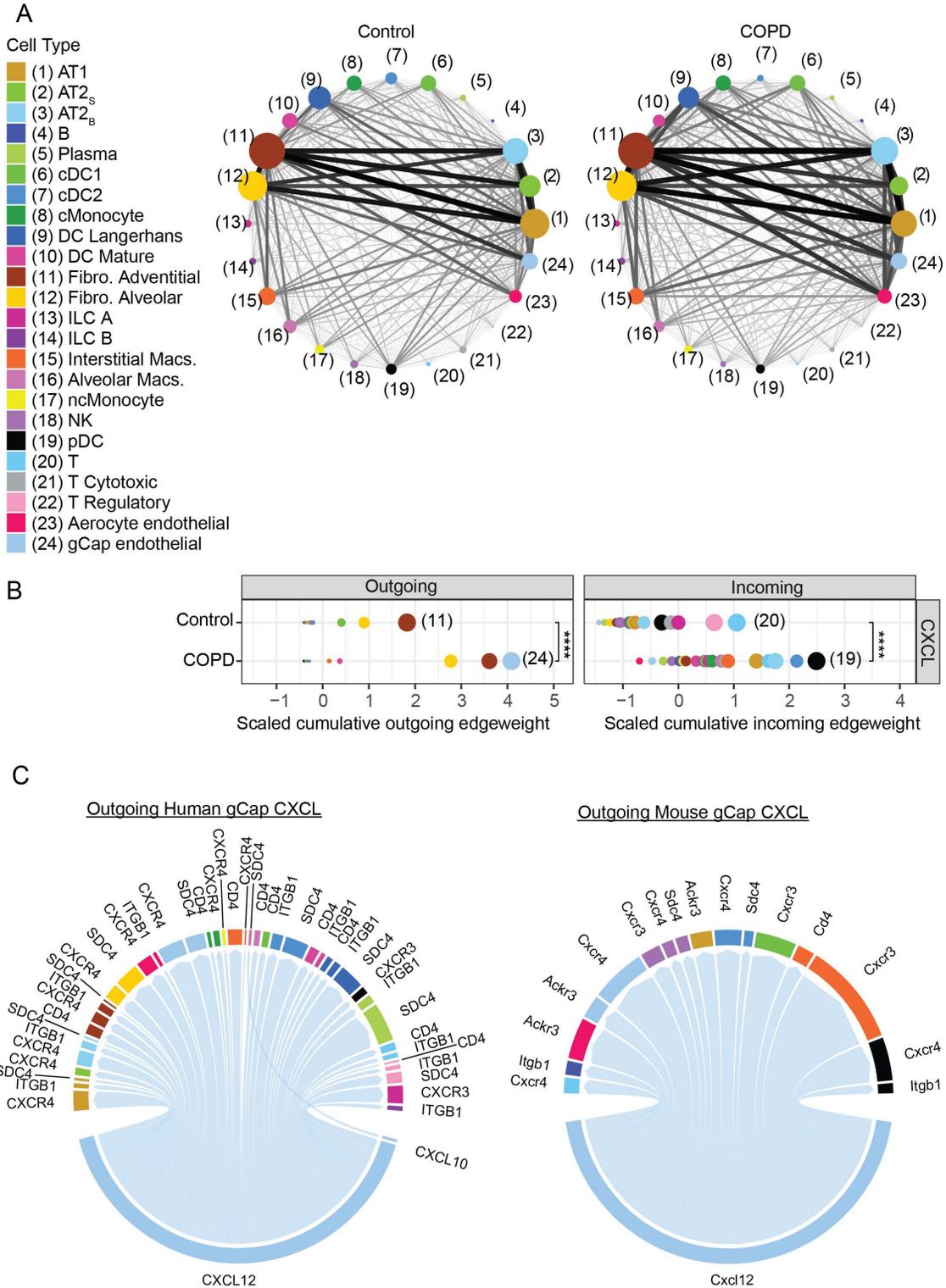


Figure 5: Alveolar niche networks and pathway centrality analyses. (A) Network plots of the alveolar niche in control (*left*) and COPD (*right*). Each node represents a cell population and each internodal edge reflects ligand-receptor interactions between cell types. The edge weight (thickness) between nodes reflects the sum of individual edge weights (non-directed) which are based on the product of ligand-receptor gene expression values, while the size of the node reflects measurements of Kleinberg centrality which prioritizes cell types responsible for incoming (authority) and outgoing (hub) cell-cell signaling. Individual cell types are labelled by color and number. **(B)** Centrality analysis of the alveolar connectome for CXCL signaling between control and COPD. Dot size is proportional to the Kleinberg scores for each cell type within CXCL signaling. Panel shows outgoing edge weights and Kleinberg hub scores (*left*) and incoming edge weights and Kleinberg authority scores (*right*). Individual cell types are color labelled as in Figure 5A, and numbers shown identify cell types with the largest Kleinberg centrality score. *** $p < 0.0001$ using the Durbin test to compare control and COPD across cell types. **(C)** Differential circos plots for outgoing gCap CXCL signaling from human and mouse connectomes. Edge thickness is proportional to perturbation scores, defined as the product of the absolute values of the log-fold change for both the receptor and ligand. CXCL differential network analysis limited to edges in which both ligand and receptor expression are increased.

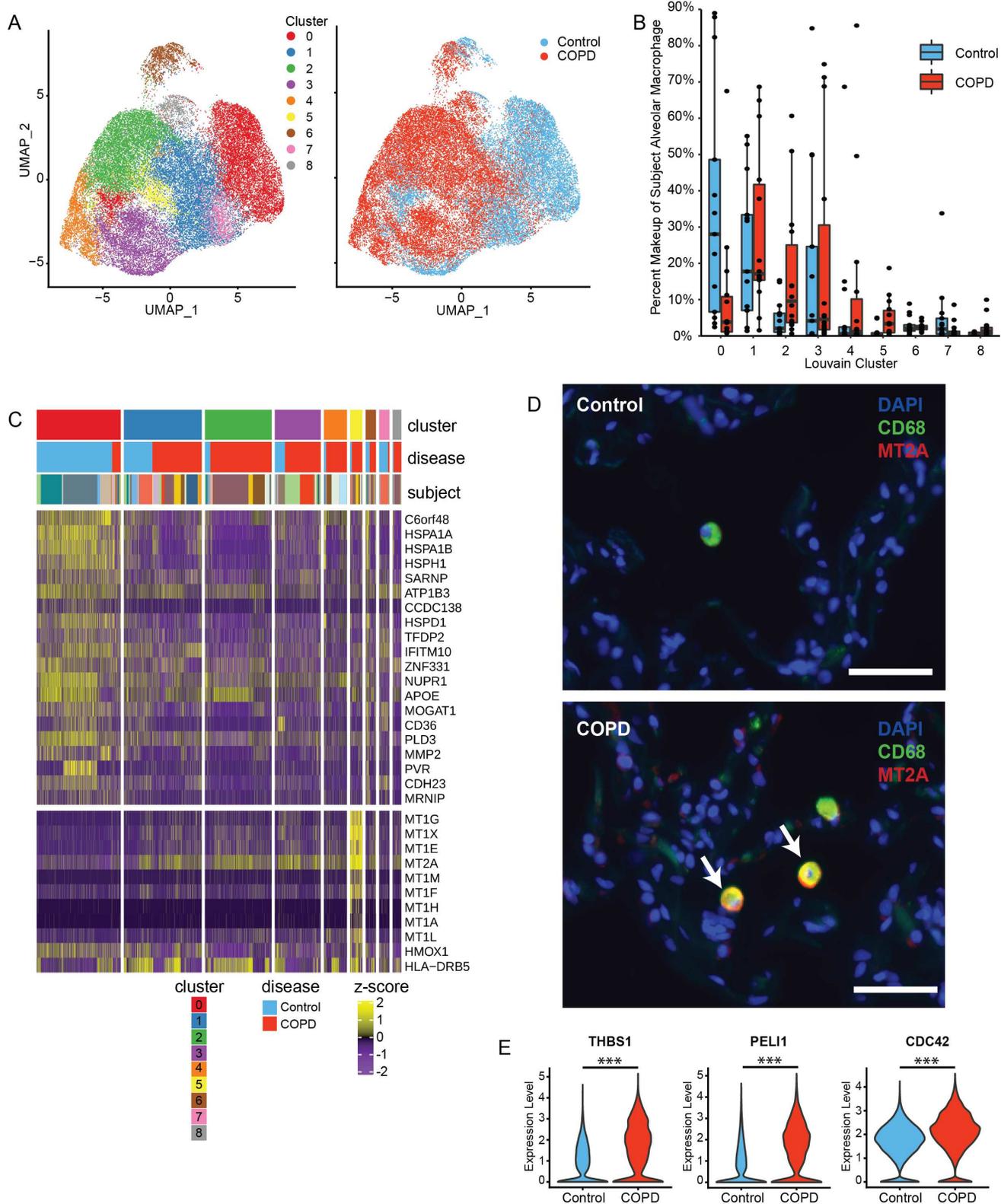


Figure 6: Changes in alveolar macrophage population composition in COPD. (A) UMAPs of control and COPD alveolar macrophage cells, color labelled by Louvain cluster (*left*) and disease status (*right*). (B) Percent makeup of alveolar macrophage across all nine Louvain clusters per subject, grouped by disease state. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. The percent of cluster 0 and cluster 5 alveolar macrophages are significantly different between control and COPD (Wilcoxon rank sum test, FDR < 0.05). (C) Heatmap of the distribution of z-scores of marker genes for cluster 0 and cluster 5 alveolar macrophages. Columns represent expression values from individual cells and are hierarchically ordered by macrophages cluster, disease status, and subject. (D) Immunofluorescence images of MT2A (*red*) expression in CD68+ cells (*green*) in control and COPD lung tissue samples (*arrows*). Arrows point to colocalization of MT2A and CD68 (*yellow*). Scale bar = 50 μ m. (E) Violin plots of *THBS1*, *PELI1*, and *CDC42* gene expression in alveolar macrophages in control and COPD subjects. *** $p < 0.0001$ using Wilcoxon rank sum test with Bonferroni correction.

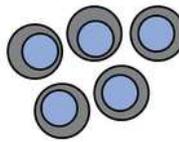
Figures

A Human Lung Tissue



Control n = 15 (8F; 7M)
COPD n = 17 (8F; 9M)

Tissue Dissociation



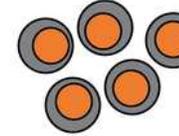
Single cell Sequencing and Analysis

Mouse Lung Tissue



Room Air (RA) n = 4 (2F; 2M)
Cigarette Smoke (CS) n = 4 (2F; 2M)

Tissue Dissociation



Single cell Sequencing and Analysis

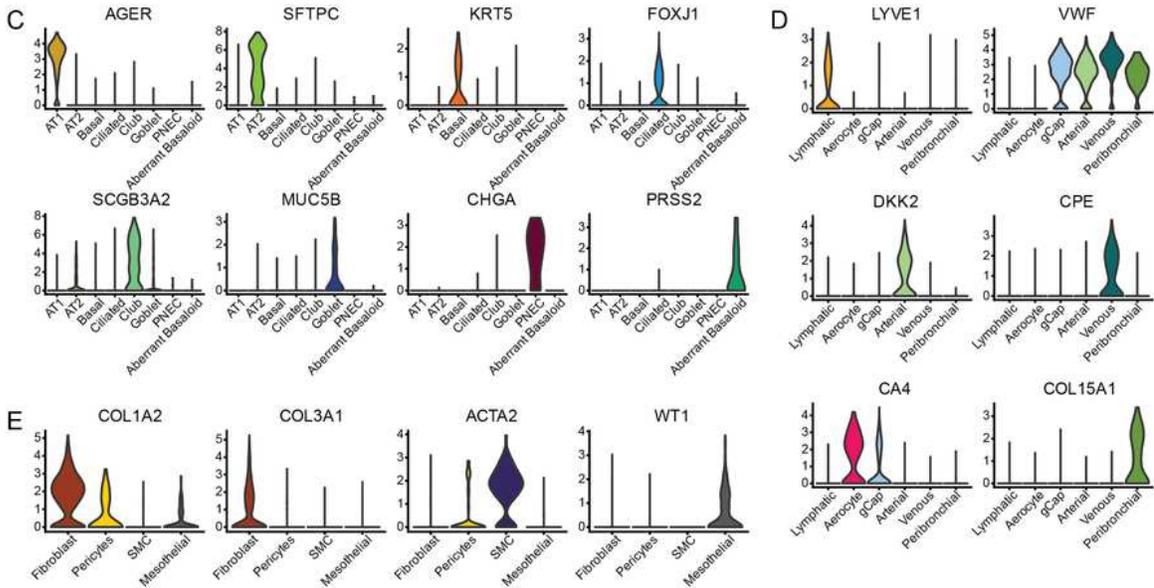
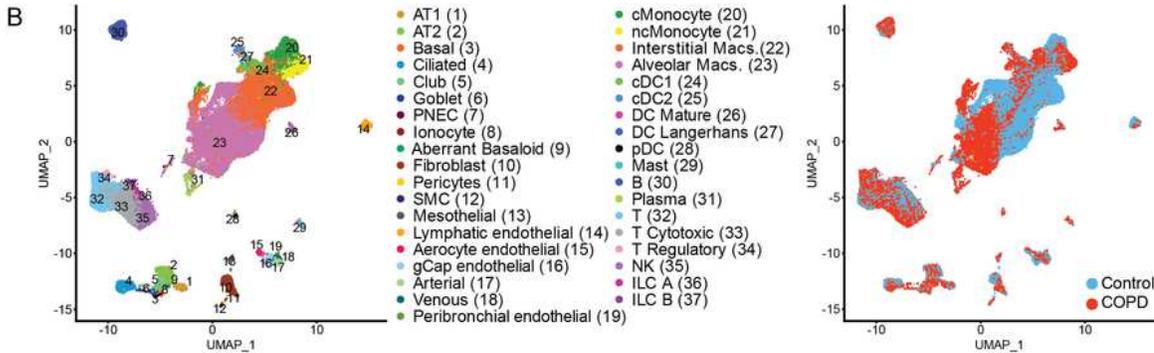


Figure 1

Profiling of cell types in COPD using scRNAseq data. (A) Tissue from 17 lungs with advanced COPD and 15 control donor lungs were dissociated into single-cell suspensions for barcoding and sequencing. Similarly, lung tissue from 2 male (M) and 2 female (F) mice exposed to 10 months of cigarette smoke

(CS) and 2 M and 2 F mice exposed only to room air (RA) was dissociated into single-cell suspensions for barcoding and sequencing. (B) Uniform Manifold Approximation and Projection (UMAP) representation of 111,540 single cells grouped into 37 distinct cell types (left) with identification of COPD and control cells (right). Violin plots of normalized expression values for canonical cell-specific marker genes in (C) epithelial, (D) endothelial, and (E) stromal cells. AT1=alveolar epithelial type I, AT2=alveolar epithelial type II, PNEC=pulmonary neuroendocrine cells, SMC=smooth muscle cells, gCap=general capillary, cMonocyte=classical monocytes, ncMonocyte=non-classical monocyte, Macs.=macrophages, DC=dendritic cells, cDC=conventional dendritic cells, pDC=plasmacytoid dendritic cells, NK = natural killer cells, ILC = innate lymphoid cells.

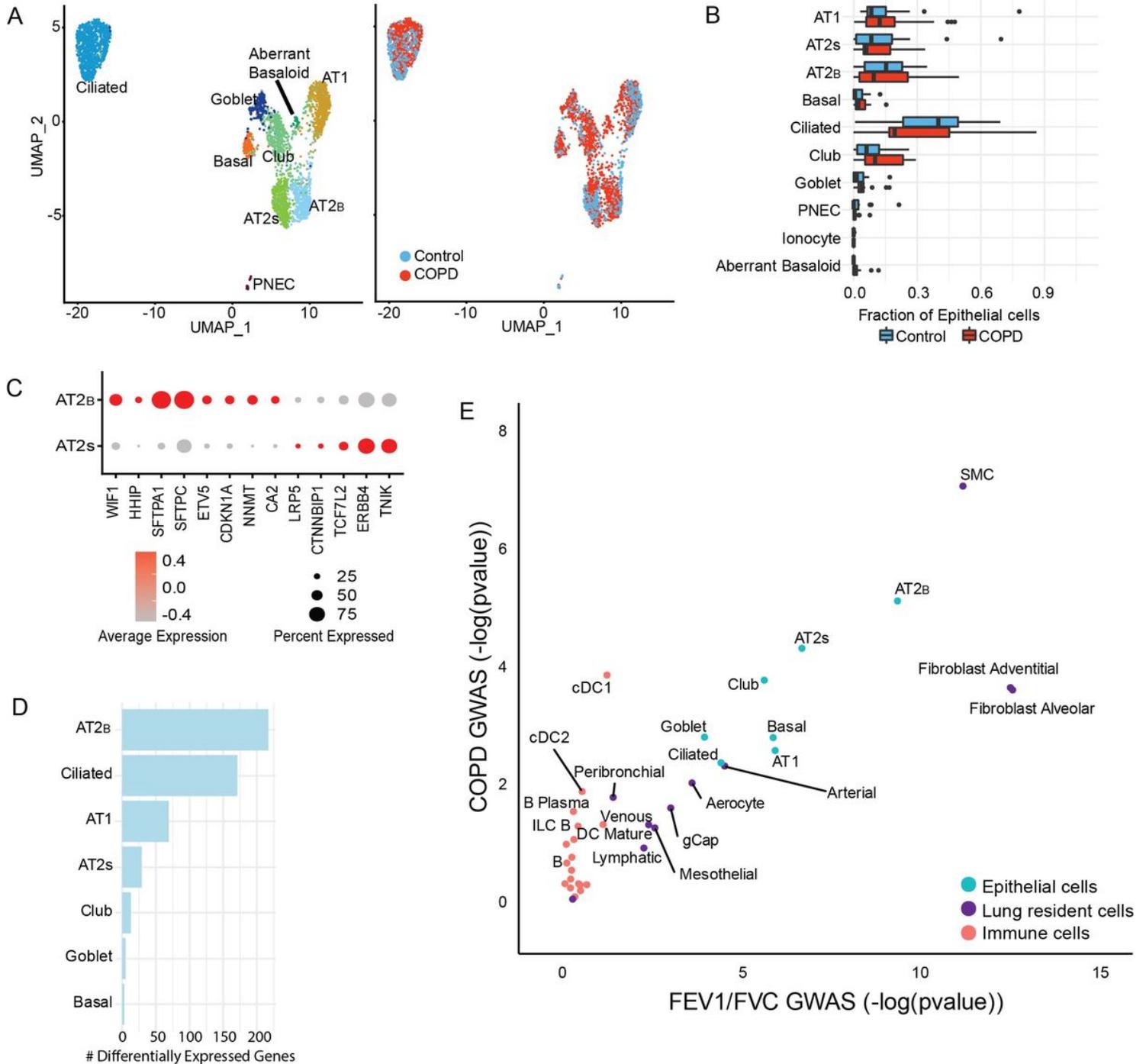


Figure 2

COPD Epithelial Cells. (A) UMAP of epithelial cells from COPD and control donor lungs. Samples are color labelled by cell type (left) and disease category (right). Alveolar type II (AT2) cells can be distinguished as two clusters denoted as AT2S and AT2B. (B) Distribution of subject-specific epithelial cell types as a fraction of the total number of epithelial cells per disease category. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. (C) Dot plot of z-scores for marker gene expression values. Dot size reflects percentage of cells with gene expression; color corresponds to degree of regulon activity. (D) Number of differentially expressed genes between control and COPD across epithelial cells types with $p < 0.05$ using Wilcoxon rank sum test with Bonferroni correction. (E) Plot of negative log adjusted p-values for cell type-specific enrichment for GWAS-identified genes with polymorphisms associated with lung function (continuous FEV1/FVC) (x-axis) and presence of COPD (FEV1/FVC < 70) (y-axis).

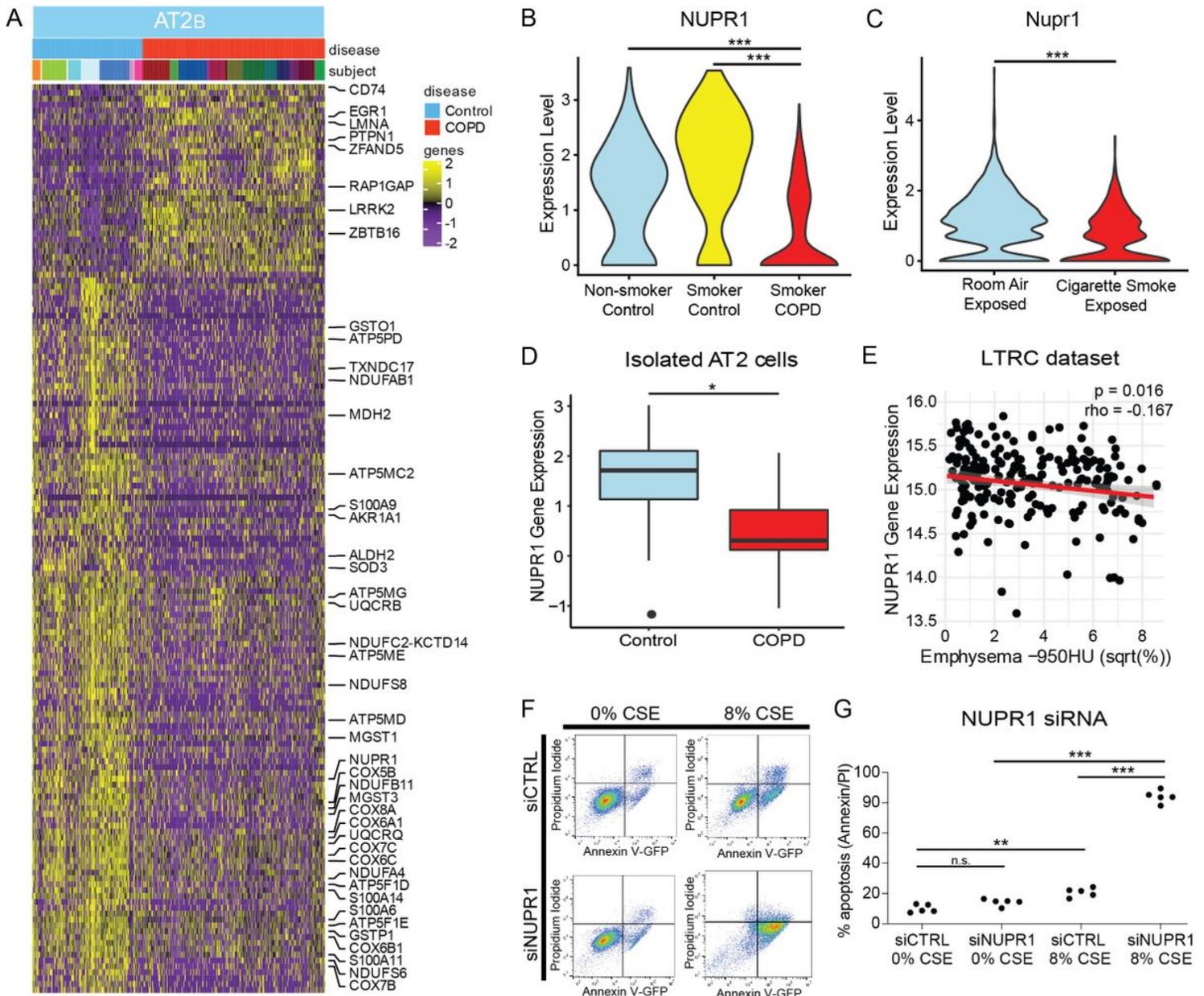


Figure 3

Aberrant AT2B Cellular Stress Response in COPD. (A) Heatmaps of z-scores for 625 differentially expressed genes (yellow/purple) between control and COPD for AT2B cells (Wilcoxon rank sum test, Bonferroni corrected $p < 0.05$). Each column represents expression values for an individual cell. Columns are hierarchically ordered by disease phenotype and subject, in which disease category and individual subject are represented by unique colors. z-scores were calculated across all epithelial cells. (B) NUPR1 in AT2B cells from human subjects from non-smokers ($n=11$), former/current smokers without COPD ($n=4$), and former smokers with COPD ($n=17$). (C) Nupr1 in AT2 cells from mice exposed to room air for 10 months vs. mice exposed to 10 months of cigarette smoke (CS). (D) NUPR1 in isolated AT2 cells from control and COPD lung tissue. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. (E) Spearman correlation of NUPR1 with the square root (sqrt) of radiographic emphysema in lung tissue samples from the LGRC cohort ($n=208$). (F) Flow cytometric detection of propidium iodide (y-axis) and Annexin V (x-axis) cells treated with NUPR1 silencing RNA (siNUPR1) vs. silencing control RNA (siCTRL) and exposed to 0% or 8% cigarette smoke extract (CSE). (G) Percent cell death determined by flow cytometry for Annexin V/ Propidium iodide in cells treated with siNUPR1 vs. siCTRL and exposed to 0% or 8% CSE ($n= 5$ /group). n.s. = non-significant, ** $p < 0.001$, *** $p < 0.0001$ using Wilcoxon rank sum test with Bonferroni correction (B,C), unadjusted Wilcoxon rank sum test (D), or one-way ANOVA with Tukey post-hoc test (G).

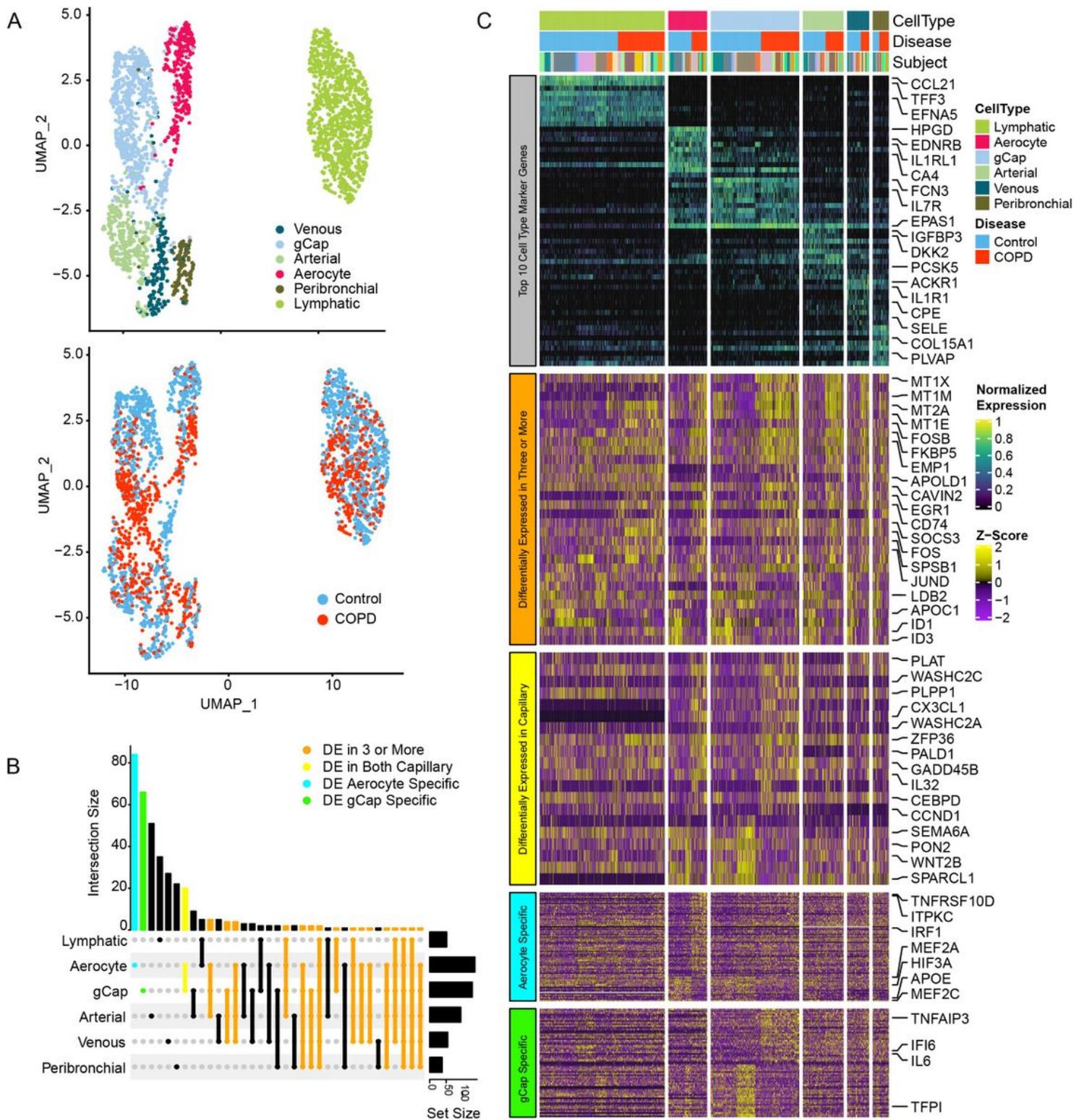


Figure 4

COPD endothelial cell types demonstrate universal and cell type-specific transcriptional aberrations. (A) UMAPs of all vascular endothelial (VE) and lymphatic endothelial cells from control and COPD subjects. UMAPs are color labelled by cell type (top) and disease status (bottom). (B) UpSet plot visualizing the properties of intersecting and unique sets of differentially expressed (DE) genes between COPD and control amongst endothelial (Wilcoxon rank sum test, unadjusted $p < 0.001$, minimal fold change > 0.5).

(C) Heatmap of corresponding differentially expressed genes between COPD and control amongst six subtypes of endothelial cells. Each column represents expression values for an individual cell. Columns are hierarchically ordered by endothelial subtype, disease phenotype, and then subject. Grey row (top): expression values for marker genes are unity normalized between 0 and 1 across all endothelial subtypes. Orange row (middle): z-scores of differentially expressed genes in three or more endothelial cell types between control and COPD. Yellow row (middle): z-scores of differentially expressed genes in both aerocytes and gCaps between control and COPD. Blue and green row (bottom): z-scores of differentially expressed genes unique to aerocytes (blue) or gCaps (green). Unity normalization and z-score calculations were performed using all endothelial subtypes.

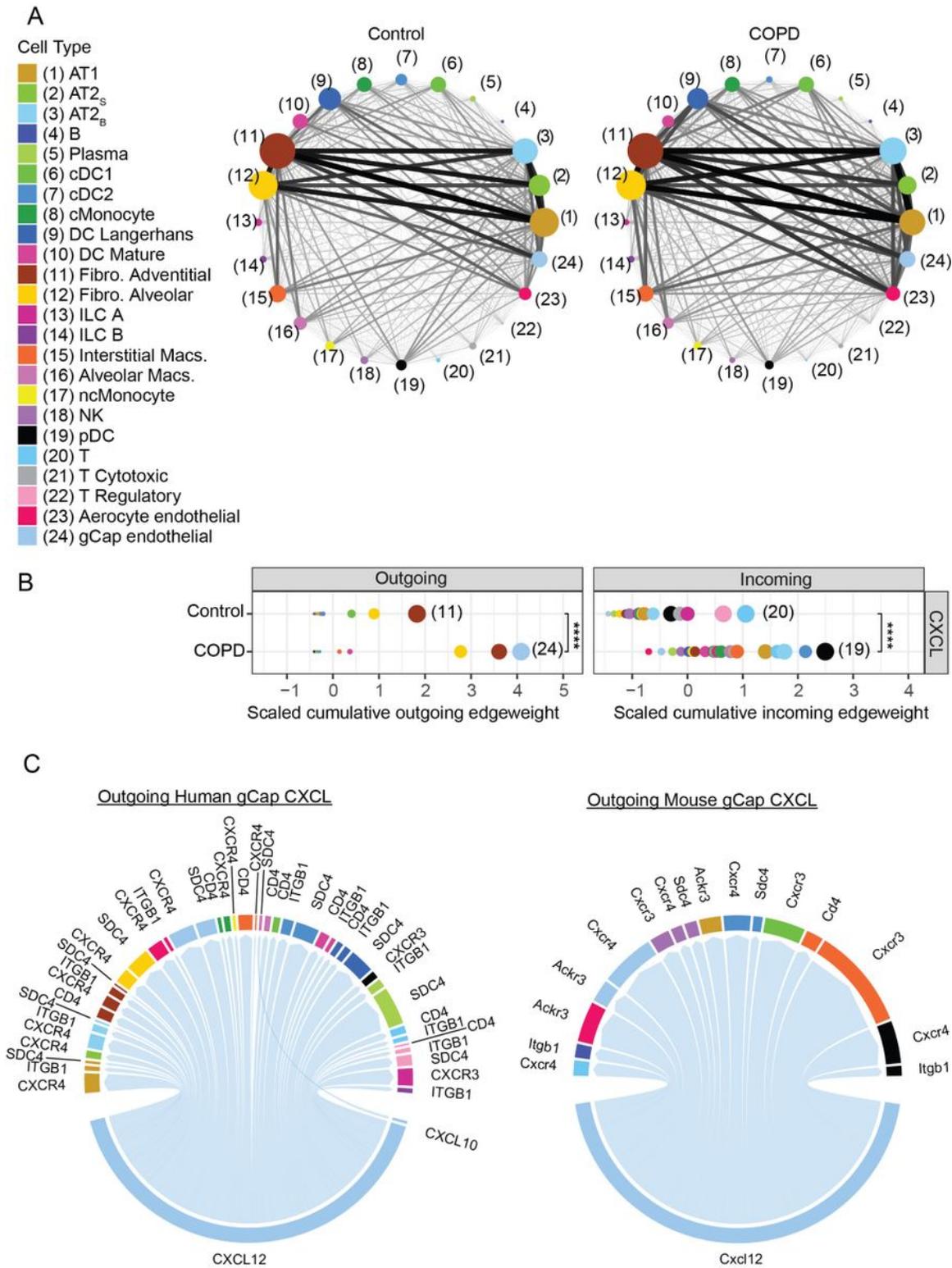


Figure 5

Alveolar niche networks and pathway centrality analyses. (A) Network plots of the alveolar niche in control (left) and COPD (right). Each node represents a cell population and each internodal edge reflects ligand-receptor interactions between cell types. The edge weight (thickness) between nodes reflects the sum of individual edge weights (non-directed) which are based on the product of ligand-receptor gene expression values, while the size of the node reflects measurements of Kleinberg centrality which

prioritizes cell types responsible for incoming 664 (authority) and outgoing (hub) cell-cell signaling. Individual cell types are labelled by color and number. (B) Centrality analysis of the alveolar connectome for CXCL signaling between control and COPD. Dot size is proportional to the Kleinberg scores for each cell type within CXCL signaling. Panel shows outgoing edge weights and Kleinberg hub scores (left) and incoming edge weights and Kleinberg authority scores (right). Individual cell types are color labelled as in Figure 5A, and numbers shown identify cell types with the largest Kleinberg centrality score. *** $p < 0.0001$ using the Durbin test to compare control and COPD across cell types. (C) Differential circos plots for outgoing gCap CXCL signaling from human and mouse connectomes. Edge thickness is proportional to perturbation scores, defined as the product of the absolute values of the log-fold change for both the receptor and ligand. CXCL differential network analysis limited to edges in which both ligand and receptor expression are increased.

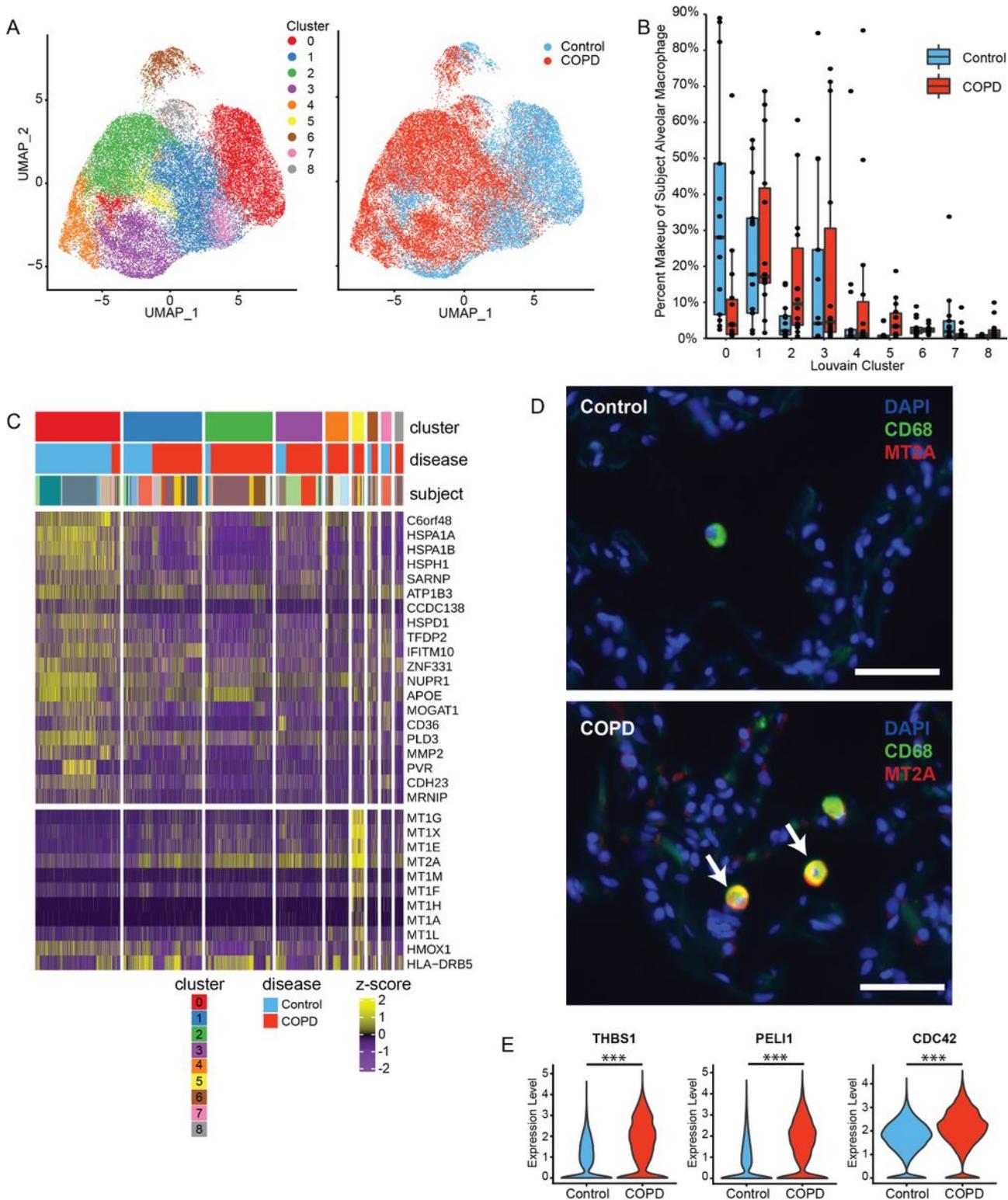


Figure 6

Changes in alveolar macrophage population composition in COPD. (A) UMAPs of 6 control and COPD alveolar macrophage cells, color labelled by Louvain cluster (left) and disease status (right). (B) Percent makeup of alveolar macrophage across all nine Louvain clusters per subject, grouped by disease state. Boxes represent interquartile ranges (IQRs); whiskers are 1.5 x IQR. The percent of cluster 0 and cluster 5 alveolar macrophages are different between control and COPD (Wilcoxon rank sum test, FDR <0.05). (C)

Heatmap of the distribution of z-scores of 682 marker genes for cluster 0 and cluster 5 alveolar macrophages. Columns represent expression values from individual cells and are hierarchically ordered by macrophages cluster, disease status, and subject. (D) Immunofluorescence images of MT2A (red) expression in CD68+ cells (green) in control and COPD lung tissue samples (arrows). Arrows point to colocalization of MT2A and CD68 (yellow). Scale bar = 50 μ m. (E) Violin plots of THBS1, PELI1, and CDC42 gene expression in alveolar macrophages in control and COPD subjects. *** $p < 0.0001$ using Wilcoxon rank sum test with Bonferroni correction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFiguresandTables1and2.docx](#)
- [NCsinglecellTable3.xlsx](#)