

1 **Acquisition of alveolar fate and differentiation competence by human fetal lung epithelial**
2 **progenitor cells**

3

4 Kyungtae Lim¹, Walfred Tang¹, Dawei Sun¹, Peng He^{2,3}, Sarah A. Teichmann², John C.
5 Marioni^{2,3}, Kerstin B. Meyer² and Emma L. Rawlins^{1,*}

6

7 ¹ Wellcome Trust/CRUK Gurdon Institute, Department of Physiology, Development and
8 Neuroscience, Wellcome Trust/MRC Stem Cell Institute, University of Cambridge, Cambridge,
9 CB2 1QN, UK.

10

11 ² Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK

12

13 ³ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI),
14 Wellcome Genome Campus, Cambridge, UK.

15

16

17 * Contact: e.rawlins@gurdon.cam.ac.uk

18

19 *Key words:* human lung development; distal tip; organoids; alveolar differentiation; NKX2.1;
20 stem cell; Wnt.

21

22 **ABSTRACT**

23 Variation in lung alveolar development is strongly linked to disease susceptibility. However,
24 the cellular and molecular mechanisms underlying alveolar development are difficult to study
25 in humans. Using primary human fetal lungs we have characterized a tip progenitor cell
26 population with alveolar fate potential. These data allowed us to benchmark a self-organising
27 organoid system which captures key aspects of lung lineage commitment and can be efficiently
28 differentiated to alveolar type 2 cell fate. Our data show that Wnt and FGF signalling, and the
29 downstream transcription factors NKX2.1 and TFAP2C, promote human alveolar or airway
30 fate respectively. Moreover, we have functionally validated cell-cell interactions in human lung
31 alveolar patterning. We show that Wnt signalling from differentiating fibroblasts promotes
32 alveolar type 2 cell identity, whereas myofibroblasts secrete the Wnt inhibitor, NOTUM,
33 providing spatial patterning. Our organoid system recapitulates key aspects of human lung
34 development allowing mechanistic experiments to determine the underpinning molecular
35 regulation.

36

37

38

39 INTRODUCTION

40 During human lung development the airway tree is formed by branching between ~5 and 16
41 post conception weeks (pcw) in the pseudoglandular stage of development. At the canalicular
42 stage, ~16 to 26 pcw, the most distal epithelial tubes narrow, come into close proximity to
43 capillaries and start to differentiate as alveolar epithelial cells^{1,2}. Preterm infants born during
44 the late canalicular stage have a rudimentary gas exchange surface and can survive if provided
45 with specialised, neonatal intensive care. However, the molecular mechanisms underlying
46 human alveolar development remain largely unknown.

47

48 During airway branching, the human tip epithelium is SOX9, SOX2 dual-positive and
49 functions as a multipotent progenitor. Pseudoglandular tip epithelium has been cultured as self-
50 renewing organoids which model the airway branching stage^{3,4}. During the canalicular stage,
51 as alveolar differentiation begins, tip progenitors become SOX9 single-positive and more
52 cuboidal in shape³. We hypothesized that growth of tip organoids from canalicular stage lungs
53 would provide an improved model for studying human alveolar differentiation.

54

55 Differentiation of human iPSCs to alveolar lineages suggests Wnt signalling is essential
56 for alveolar fate⁵. NKX2.1 is also implicated in alveolar differentiation. In mouse lungs,
57 Nkx2.1 is essential for alveolar differentiation and maintenance and binds to promoters of
58 alveolar type 1 (AT1) and alveolar type 2 (AT2) cell-specific genes⁶. Heterozygous missense
59 mutations in the *NKX2.1* homeodomain cause brain-lung-thyroid syndrome which includes
60 disrupted surfactant gene expression and interstitial lung disease^{7,8}. However, whether NKX2.1
61 simply promotes surfactant synthesis, or has additional roles in human lung alveolar
62 differentiation is currently unknown.

63

64 We find that SOX9⁺ human lung tip progenitors acquire an AT2 gene expression
65 signature by the canalicular stage of development. We derive, and characterise, a canalicular
66 stage tip self-renewing organoid model. This has allowed us to determine upstream signals and
67 downstream TFs which promote lung lineage commitment and investigate the spatial
68 patterning of the developing alveolus.

69

70 RESULTS

71 Human fetal lung tip progenitor cells acquire alveolar lineage signatures *in vivo*

72 We investigated the distal lung tip epithelium of human fetal lungs at the pseudoglandular
73 and canalicular stages. Consistent with our previous report³, pseudoglandular distal tips were
74 columnar and marked by SOX9 and SOX2. The canalicular stage distal tip epithelium
75 contained more cuboidal SOX2⁻ SOX9⁺ cells which co-expressed the AT2 cell markers, SFTPC
76 and HTII-280 (Fig. 1A,B). We identified surface markers to distinguish between
77 pseudoglandular and canalicular stage tips. CD44 marks tip epithelial cells across all stages of
78 lung development tested. Whereas CD36 is expressed specifically in the canalicular stage tip
79 (~16-21 pcw) where it is co-expressed with CD44, *SFTPC* and SOX9. A lower level of both
80 CD44 and CD36 extends into the SOX9/PDPN⁺ tip-adjacent cells and CD44 extends further
81 proximally into the differentiating stalk region (Figs. 1C-F; Extended Data Fig. 1A,B).

82

83 Distal lung regions were dissected to enrich for the tip and EPCAM⁺ cells were sorted for
84 CD44 and/or CD36. At the pseudoglandular stage (11 pcw), 75% of sorted cells were CD44⁺
85 and CD44⁺CD36⁺ cells were rare (Fig. 1G; *blue*). By contrast, at the canalicular stage (20 pcw)
86 53% of sorted cells were CD44⁺CD36⁺ and only 17% were single CD44⁺ (Fig. 1G; *red*). qRT-
87 PCR showed that the 17-20 pcw CD44⁺CD36⁺ cells robustly expressed *CD36*, *CD44*, *SFTPC*
88 and *SOX9*, but extremely low levels of the airway markers, *TP63* and *SOX2*, consistent with
89 the immunostaining (Fig. 1H). In contrast, single CD44⁺ cells showed a higher level of *SOX2*,
90 but much lower levels of *SFTPC* and *SOX9*, suggesting that they are derived from the
91 CD44⁺SOX9⁻ stalk region (Fig. 1H; Extended Data Fig. 1B). Finally, the CD44⁻CD36⁻ cells
92 had higher levels of *TP63* and *SOX2*, but low *SFTPC* and *SOX9*, indicating they are derived
93 from more proximal airway-lineage cells (Fig. 1H). Therefore, dual expression of CD44 and
94 CD36 marks the tip epithelial population in the canalicular stage lung.

95

96 **Gradual acquisition of tip alveolar lineage signature during human lung development**

97 Flow cytometric analysis showed that the expression of CD36 was robustly acquired
98 between 13 and 15 pcw, prior to the canalicular stage (Fig. 1I). Similarly, in the CD36⁺ cells,
99 the mRNA levels of *SFTPC*, *CD36*, and *LAMP3* began to increase from 13 pcw onward (Fig.
100 1J). We confirmed that CD36 was detectable in the tip epithelium at exactly 14 pcw, moreover
101 the intensity of *SFTPC* transcripts increased while *SOX9* gradually lowered during this
102 transition period (Fig. 1K-M). These results show that the acquisition of AT2 lineage signatures
103 occurs gradually in the tip epithelium prior to the canalicular stage.

104

105 **Organoids derived from distal tip epithelium at the canalicular stage exhibit alveolar** 106 **lineage signatures**

107 To determine the fate potential of the canalicular stage (17-20 pcw) distal tip, CD44⁺CD36⁺
108 cells were cultured for 3 weeks (Fig. 2A). Two morphologically distinct organoids formed:
109 cystic and folded (Fig. 2B). All the folded organoids consisted of cuboidal cells and expressed
110 both progenitor and AT2 markers, including an *SFTPC*-eGFP reporter (Fig. 2C). By contrast,
111 the cystic organoids had a columnar cell shape and expressed tip progenitor markers, but not
112 AT2 markers, resembling the pseudoglandular tip epithelium (Fig. 2B-E; Extended Data Fig.
113 2A,B). For simplicity, we refer to the folded and cystic organoids isolated from 17-21 pcw
114 lungs as lineage positive (Lin^{POS}) and negative (Lin^{NEG}).

115

116 We tested whether any organoids (from the mixed Lin^{POS} and Lin^{NEG} population) retained
117 CD44 and CD36 expression after 3 weeks culture (Fig. 2F). The CD44⁺CD36⁺ cells showed
118 the highest level of *SFTPC* with a moderate level of *SOX9*, but very low levels of *SOX2* and
119 *TP63*. They were located at the tips of the Lin^{POS} organoids where they expressed *SFTPC*,
120 *HTII-280*, *SOX9*, *CD44* and *KI67* (Fig. 2G,H; Extended Data Fig. 2C-F). The CD44⁻CD36⁻
121 cells had the highest levels of *TP63* and *SOX2*. They corresponded to the inner parts of the
122 Lin^{POS} organoids where scattered TP63⁺ cells were found (Fig. 2G,H; Extended Data Fig. 2D).
123 By contrast, the CD44⁺CD36⁻ cells had the highest level of *SOX9* and a moderate level of *SOX2*,
124 but no lineage markers, and corresponded to the Lin^{NEG} organoids which had uniform CD44,
125 *SOX2* and *SOX9* (Fig. 2G,H; Extended Data Fig. 2E). These data suggested that the
126 CD44⁺CD36⁺ canalicular stage tip cells originally plated had self-renewed (at the tips) and

127 differentiated towards airway lineages (in the centre) to form the Lin^{POS} organoids. Moreover,
128 that a fraction of the CD44⁺CD36⁺ tip cells had also grown into Lin^{NEG} organoids, resembling
129 the pseudoglandular stage tips. To test this hypothesis, we infected freshly-isolated
130 CD44⁺CD36⁺ epithelial cells with *SFTPC*-eGFP⁺ lentivirus and sorted for eGFP. We observed
131 that sorted *SFTPC*-eGFP⁺ cells formed both Lin^{POS} and Lin^{NEG} organoids (Extended Data Fig.
132 2G-I). These data suggest that the emergence of Lin^{NEG} organoids, which are similar to
133 pseudoglandular stage tip organoids, is due to dedifferentiation of the epigenetically unstable
134 canalicular tip epithelium. However, we cannot exclude the possibility that the canalicular tip
135 epithelium contains a mixture of cell states.

136

137 We tested if the CD44⁺CD36⁺ cells continued to self-renew upon passaging. P0 organoids
138 (mixed population of Lin^{POS} and Lin^{NEG}) were sorted as CD44⁺CD36⁺, CD44⁺CD36⁻ and
139 CD44⁻CD36⁻ and cultured separately in the self-renewal medium. Only the CD44⁺CD36⁺ cells
140 were able to generate a large proportion of Lin^{POS} organoids with folded structure and
141 progenitor/AT2 gene signature (Figs. 2I-K; Extended Data Fig. 2J-L). In contrast, the
142 CD44⁺CD36⁻ cells, derived from Lin^{NEG} organoids, produced Lin^{NEG} organoids. The CD44⁻
143 CD36⁻ cells, derived from the centre of the Lin^{POS} organoids, largely formed airway-fated
144 spheres expressing a significantly higher level of *TP63*/*TP63* (Fig. 2I-K; Extended Data Fig.
145 2J-L). These data conclusively demonstrate that the CD44⁺CD36⁺ cells are the major tip
146 progenitor subpopulation *in vitro* and can maintain the Lin^{POS} organoids. We have therefore
147 captured the canalicular stage lung tip epithelial population which co-expresses *SOX9* and *AT2*
148 markers in the Lin^{POS} organoids.

149

150 We performed RNAseq to compare the transcriptome of passaged pseudoglandular stage
151 tip organoids derived from 8-9 pcw³ with passaged canalicular stage Lin^{NEG} and Lin^{POS}
152 organoids (Extended Data Fig. 3A). Hierarchical clustering and principal component analysis
153 showed that the Lin^{NEG} organoids were very similar to the pseudoglandular organoids, but
154 distinct from the Lin^{POS} organoids (Fig. 2L; Extended Data Fig. 3B). We identified >280
155 differentially expressed genes between the Lin^{POS} and Lin^{NEG} organoids (Supplementary Table
156 1; Extended Data Fig. 3C; log₂FC>4, P<0.05). Similar to the pseudoglandular organoids, the
157 Lin^{NEG} organoids were highly associated with Gene Ontology (GO) terms related to ion
158 transport and branching morphogenesis, confirming that they resemble the pseudoglandular
159 stage distal tips. Whereas the Lin^{POS} organoids had significant GO terms for respiratory
160 gaseous exchange and lung alveolus development, as well as canonical Wnt pathway signalling
161 (Extended Data Fig. 3D-G). Moreover, the Lin^{POS} organoids were enriched for AT2 markers
162 *SFTPC*, *SLC34A2*, *NAPSA*, *LPCAT1*, *FOXP2*, and *CEBPD*, Wnt signalling-related genes
163 *CTNNB1*, *TCF7L1*, *WNT7B*, *WIF1*, *LRRK2* and low levels of airway genes *TP63*, *SCGB3A1*,
164 *SCGB3A2* (Fig. 2M; Supplementary Table 1). These data confirm that the passaged Lin^{POS}
165 organoids recapitulate key molecular characteristics of the canalicular stage lung tip epithelial
166 progenitors.

167

168 **Coordinated control of the canalicular stage tip epithelial cell fate by Wnt and FGF** 169 **signalling**

170 To determine which signalling cues direct differentiation of the canalicular tips, tip

171 epithelium was isolated from the distal lung and directly exposed to pairwise signal
172 combinations (Fig. 3A; Extended Data Fig. 4A). The cells did not grow in the absence of
173 SMADi (Noggin and SB431542) (Fig. 3B; Extended Data Fig. 4B). However, we observed
174 that two distinct populations of organoids could be obtained by combining SMADi with CHIR
175 (CHIR99021, a Wnt agonist), or with FGFs (FGF7 and FGF10) (Fig. 3B). Organoids grown in
176 SMADi/CHIR had a thin epithelium with a hollow lumen. They could not be passaged and
177 expressed the highest level of *SFTPC*, greater than the Lin^{POS} organoids. In contrast, organoids
178 grown in SMADi/FGF formed spheres with a small lumen, a relatively thick, proliferative
179 epithelial layer and expressed the highest level of TP63 (Fig. 3B-D). In both conditions *SOX9*
180 levels were lower than the Lin^{POS} and Lin^{NEG} organoids (Fig. 3C). These data indicate that Wnt
181 and FGF signalling promote the lineage determination of the canalicular stage tip epithelium
182 to alveolar or airway lineages, in agreement with previously published data^{5,9}. Moreover, when
183 SMADi/CHIR/FGFs were combined (equivalent to our self-renewing medium) organoids
184 displayed a mixture of alveolar and airway characteristics, as in the Lin^{POS} organoids (Fig. 3D).

185

186 We demonstrated that the canalicular stage tips are highly plastic and can switch readily
187 between alveolar and airway differentiation by altering the medium and observing rapid
188 organoid morphology and gene expression changes (Extended Data Fig. 4C-E). Moreover,
189 FGF7 alone, without FGF10, was sufficient to promote airway fate (Extended Data Fig. 4F,G).
190 Freshly-isolated pseudoglandular stage (8 pcw) distal tip epithelial cells did not show similar
191 levels of differentiation when grown in the same conditions (Fig. 4H). These data indicate that
192 the canalicular stage tip progenitors are both highly plastic and in a differentiation-ready state
193 compared to the pseudoglandular tip.

194

195 We reasoned that Wnt and FGF signalling likely control lineage determination of the tip
196 epithelium *in vivo*. In canalicular stage tissue, we found that the *SFTPC*⁺ tips expressed higher
197 levels of the Wnt targets *AXIN2* and *WIF1*, compared with stalk and airway epithelium (Figs.
198 3F-H; Extended Data Figs. 4H,I; 5A-D). We observed that *WNT2* is co-expressed with *FGFR4*
199 in fibroblasts throughout the canalicular stage (Fig. 4A); putative alveolar fibroblasts¹⁰. This
200 led us to question how Wnt-responsive *SFTPC* could be precisely restricted to the tip
201 epithelium in the presence of widespread *WNT2*. A secreted Wnt inhibitor, *NOTUM*¹¹, is
202 expressed in both the distal tip epithelium and the myofibroblast/smooth muscle population
203 which surrounds the differentiating epithelial stalk cells (Fig. 4B,C; Extended Data Fig. 5B,C).
204 The *NOTUM*⁺ myofibroblasts co-express the Wnt targets *LEF1* and *AXIN2*, suggesting that
205 they may also be responding to Wnt (Extended Data Fig. 5C,D). We hypothesised that in
206 response to the WNT2 signal, the stalk myofibroblasts locally secrete NOTUM, preventing the
207 stalk epithelium from experiencing a high level of Wnt activity, allowing them to turn off
208 *SFTPC* and exit the tip fate. We identified surface antigens for the specific isolation of WNT2⁺
209 fibroblasts or NOTUM⁺ myofibroblasts (Fig. 4D; Extended Data Fig. 5E,F). Isolated
210 PDGFRA⁺CD141⁺ myofibroblasts express high levels of *ACTA2*, *NOTUM* and *LEF1*. Whereas
211 PDGFRA⁻CD141⁻ fibroblasts express high levels of *WNT2* and *FGFR4* (Fig. 4E). This specific
212 gene expression is maintained if the cell types are cultured individually for 14 days. When
213 freshly isolated fibroblasts and myofibroblasts were co-cultured, the levels of *LEF1* and
214 *NOTUM* expression increased in the myofibroblasts, suggesting that they are indeed

215 responding to WNT2 from the fibroblasts (Extended Data Fig. 5G).

216

217 We therefore asked whether co-culture with the PDGFRA⁻CD141⁻ fibroblasts, or
218 PDGFRA⁺CD141⁺ myofibroblasts, could affect *SFTPC* expression in the Lin^{POS} organoids
219 (Fig. 4F). Lin^{POS} organoids robustly express the *SFTPC*-GFP reporter when cultured in the
220 self-renewing medium, but not in 2% FBS (Fig. 4G). However, co-culture with PDGFRA⁻
221 CD141⁻ fibroblasts is sufficient to substitute for the self-renewing medium and maintain
222 *SFTPC*-GFP and endogenous *SFTPC* and *LAMP3* (Fig. 4H,I). By contrast, co-culture of the
223 Lin^{POS} organoids with PDGFRA⁺CD141⁺ myofibroblasts, or both PDGFRA⁺CD141⁺
224 myofibroblasts and PDGFRA⁻CD141⁻ fibroblasts, did not support expression of AT2 genes.
225 This leads us to propose that *in vivo* WNT2-expressing alveolar fibroblasts promote *SFTPC*
226 expression in the distal tip. Moreover, that differentiating stalk cells are protected from the Wnt
227 signal by the NOTUM-secreting myofibroblasts allowing them to turn off *SFTPC* and enter a
228 differentiation programme (Fig. 4K).

229

230 Differentiating AT2 cells are also *SFTPC*⁺. We observed that the differentiating *SFTPC*⁺
231 cells were never directly over-lain by the *NOTUM*⁺ myoepithelial cells (Fig. 4J). This further
232 supports the concept that the differentiating alveolar epithelium is patterned by signals from
233 the myofibroblasts (Fig. 4K).

234

235 **NKX2.1 is a major driving force for alveolar differentiation in the tip epithelial organoids.**

236 To identify putative transcription factors for cell differentiation we analysed chromatin
237 accessibility of the pseudoglandular and Lin^{POS} organoids by bulk-ATAC seq. There were ~2-
238 fold more differentially open chromatin regions in the Lin^{POS} than pseudoglandular organoids,
239 consistent with the increased cell type complexity of the Lin^{POS} organoids (Extended Data Fig.
240 6A; Supplementary Table 2). The genomic distribution of the differentially opened chromatin
241 was similar in both organoid types (Extended Data Fig. 6B). GO analysis of the genes nearest
242 to differentially open chromatin was consistent with the RNA-seq data (Fig. 2; Extended Data
243 Fig. 6C-E). However, a much higher proportion of lung development-associated genes had
244 open chromatin at the promoter regions in the Lin^{POS} organoids (Fig. 5A). For example, the
245 promoter regions of lung differentiation genes *SFTPC*, *TP63*, *FOXP2* and *CD36* and Wnt
246 signalling genes, *AXIN2*, *CTNNB1*, *DVL3*, *LRRK2* and *WIF1*, were more accessible in the
247 Lin^{POS} than pseudoglandular organoids (Fig. 5B; Extended Data Fig. 6D,E). These data
248 strongly suggest that the chromatin accessibility of the Lin^{POS} organoids is more favourable for
249 lineage differentiation than the pseudoglandular organoids.

250

251 To predict which transcription factors (TFs) control cell fate specification, we performed
252 TF motif analysis in the differential ATAC seq peaks and compared this with our RNA seq data.
253 The motifs for FOSL1 and GATA6 binding were differentially open, and *FOSL1* and *GATA6*
254 highly expressed, in the pseudoglandular organoids (Fig. 5C). In the Lin^{POS} organoids, NKX2.1
255 and TFAP2C motifs were accessible, and these factors were highly expressed (Fig. 5D).
256 Immunostaining confirmed that NKX2.1 is more strongly expressed in the Lin^{POS} than
257 pseudoglandular organoids. Moreover, TFAP2C was absent in the pseudoglandular organoids,

258 but ubiquitous in the Lin^{POS} organoids (Fig. 5E). *In vivo*, *NKX2-1* transcripts were most highly
259 expressed in the alveolar regions of the lungs, whereas *TFAP2C* was expressed in the airway
260 epithelium (Extended Data Fig. 7A,B).

261

262 We performed overexpression (OE) of *NKX2.1* and *TFAP2C* in the pseudoglandular
263 organoids to test if either factor was sufficient to induce differentiation to the alveolar or airway
264 lineages (Fig. 5F). *NKX2.1* OE resulted in ~60% of the pseudoglandular organoids acquiring
265 an alveolar-like structure with high levels of *SFTPC* expression (Fig. 5G-J; Extended Data Fig.
266 7C). *NKX2.1* also significantly upregulated other AT2 lineage markers including *SCL34A2*,
267 *LAMP3*, *CEBPD*, *HOPX* and *AXIN2*, but downregulated the tip markers *SOX9*, *SOX2* and
268 *CD44* (Fig. 5K). In contrast, *TFAP2C* OE caused around 70% of the organoids to form
269 bronchiolar-like structures and significantly increased basal cell markers including *TP63/P63*,
270 *KRT5* and *NGFR*, (Fig. 5G-J; Extended Data Fig. 7C,D). Therefore, *NKX2.1* and *TFAP2C*
271 function as key regulators of differentiation toward AT2 and basal cell lineages respectively.

272

273 The Lin^{POS} organoid cells co-express high levels of *NKX2-1* and *TFAP2C* (Fig. 5E) yet are
274 comprised of distinct *SOX9/SFTPC*⁺ tip and *TP63*^{Lo} central regions (Fig. 2). We analysed the
275 relationship between *NKX2.1* and *TFAP2C* by overexpressing them together in the
276 pseudoglandular organoids. The *NKX2.1/TFAP2C* OE organoids were highly folded, similar
277 to *NKX2-1* OE. Furthermore, *TP63* was barely detectable, but *SFTPC* was markedly induced
278 (Extended Data Fig. 7E). When Lin^{POS} organoids were cultured in SMADi/CHIR/FGF7,
279 *NKX2.1* and *SFTPC* were high, but *TP63* and *TFAP2C* were low. Whereas in SMADi/FGF7
280 (without the Wnt agonist), *NKX2.1* decreased ~2-fold, *SFTPC* turned off and *TP63* and
281 *TFAP2C* were robustly expressed (Extended Data Fig. 7F). These data clearly demonstrate that
282 high *NKX2.1* expression, in combination with Wnt signalling, suppresses the airway lineage
283 program, also explaining why *TP63* expression is low in the Lin^{POS} organoids although
284 *TFAP2C* is expressed (Extended Data Fig. 7F). Further support for the importance of *NKX2.1*
285 in promoting alveolar and inhibiting airway differentiation came from an *NKX2.1* knock-down
286 experiment in the Lin^{POS} organoids. A small decrease in *NKX2.1* expression was sufficient to
287 decrease AT2-specific gene expression and increase *TP63* (Fig. 5L).

288

289 **Organoid assays can be used to predict the effects of human genetic variation**

290 Overexpression of *NKX2.1* containing a deleted DNA binding homeodomain showed that
291 the homeodomain is essential for AT2 differentiation (Fig. 6A,B). Numerous naturally-
292 occurring human variants in the *NKX2.1* homeodomain have been described¹²⁻¹⁴. Many of
293 these are associated with acute respiratory failure, others are predicted to be pathogenic. We
294 hypothesized that the *NKX2.1* OE assay would be a simple method to determine the effects of
295 these variants on AT2-specific gene expression (Fig. 6C). The variants differentially affected
296 organoid morphology, AT2 gene transcription and surfactant protein production (Fig. 6D-F),
297 with the predicted pathogenic c.485_487 deletion behaving indistinguishably to the wildtype
298 control. Interestingly, expression of the *NKX2.1* variants (c.523G>T and c532C>T) often
299 resulted in the production of mis-localised pro-*SFTPC* which was not processed to the mature
300 form (Fig. 6D,F). This strongly suggests that *NKX2.1* promotes multiple aspects of AT2
301 differentiation, not simply *SFTPC* transcription.

302

303 **Cultured canalicular stage tip cells differentiate readily into alveolar type 2 cells**

304 NKX2-1 OE pseudoglandular organoids had higher levels of SFTPC and ACE2 than Lin^{POS}
305 organoids, consistent with differentiation to AT2 fate (Fig. 7A; Extended Data Fig. 8A). We
306 hypothesized that a medium change would allow passaged Lin^{POS} organoids to differentiate
307 into AT2 cells. In medium containing DAPT (Notch inhibition), DCI (dexamethasone, cAMP,
308 IBMX), CHIR (Wnt agonist) and SB431542 (TGF β inhibition), *NKX2-1*, *SFTPC* and *ACE2*
309 were upregulated and *SOX9*, *SOX2* and *TP63* downregulated (Fig. 7B,C). Moreover, the
310 *SFTPC*-GFP reporter and the AT2-specific proteins LAMP3, HOPX and ACE2 were increased
311 (Fig. 7C,D; Extended Data Fig. 8B-D). The pseudoglandular organoids did not differentiate
312 towards AT2 fate in response to the same medium (Fig. 7E), confirming that the canalicular
313 stage tips are in a distinct differentiation-ready state.

314

315 Electron microscopy revealed that the Lin^{POS} organoids in the SN medium contained rare,
316 immature lamellar bodies usually surrounded by glycogen (Fig. 7F,G). Whereas higher
317 numbers of lamellar bodies with a characteristic surfactant projection core were readily visible
318 in the differentiated organoids (Fig. 7F,G)^{15,16}. NKX2.1 protein levels were increased following
319 AT2 differentiation (Fig. 7H). The differentiated cells also more efficiently processed pro-
320 SFTPC and SFTPB to the mature form (Fig. 7H)¹⁶. Moreover, NKX2.1 binds more strongly to
321 the promoters of AT2-specific genes following differentiation (Fig. 7I). These data demonstrate
322 that the Lin^{POS} organoids are readily differentiated to an AT2-like fate and confirm the
323 importance of NKX2.1 levels in this process.

324

325

326 DISCUSSION

327 We show that the distal tip cells at the canalicular stage of human lung development retain their
328 progenitor status yet exhibit aspects of AT2 gene expression and can be isolated using specific
329 surface proteins. Late-tip cell (Lin^{POS}) organoids self-renew, capture features of the canalicular
330 stage of human lung development, have extensive open chromatin and can be readily
331 differentiated to AT2-like cells. We have used this organoid system to demonstrate that Wnt
332 signalling and NKX2.1 are required for human AT2 cell differentiation. Additionally, we show
333 that antagonistic signalling interactions between differentiating fibroblasts and myofibroblasts
334 provide a spatial component to Wnt activation allowing patterning of the alveolar epithelium
335 into specific lineages. We also demonstrate that our organoid system can be used to study
336 human genetic variation.

337
338 One of major insights of this work is the demonstration that the molecular acquisition of
339 alveolar features precedes morphological changes occurring in the tip epithelium at the
340 canalicular stage. Our Lin^{POS} organoids are derived from the CD44⁺,CD36⁺ canalicular stage
341 tips. The CD44⁺,CD36⁺ cells are located at the tips of the organoids, maintain expression of
342 SOX9 and AT2 markers, self-renew and give rise to airway-fated cells in the centre of the
343 organoids (Fig. 2). When provided with appropriate cues they differentiate to an AT2-like cell
344 (Fig. 7). *In vivo*, tip cells acquire AT2 markers gradually between 13 and 15 pcw (Fig. 1). We
345 hypothesize that during this transition period (~13–15 pcw) the tips are generating the final
346 branch of the airway epithelium and at ~15 pcw switch to generating alveolar fated daughter
347 cells. However, due to well-documented tip progenitor cell plasticity in transplantation
348 assays^{17,18}, it is not yet possible to test this definitively.

349
350 We identify Wnt signalling as a key driver of human AT2 fate and patterning. Wnt
351 signalling promotes AT2 differentiation of the tip epithelium *in vitro* and *in vivo* (Fig. 3,4). This
352 is consistent with previous reports in mouse¹⁹. Similarly, human NKX2.1⁺ lung progenitors
353 derived from PSCs expressed alveolar epithelial markers in response to Wnt⁵. Our experiments
354 with primary tissue support a model in which opposing signals from differentiating alveolar
355 fibroblasts and myofibroblasts spatially restrict late tip and AT2 identity in the canalicular stage
356 human lung (Fig. 4). These data are analogous to a recent mouse report where developing AT1
357 cells are aligned with, and signal to, differentiating myofibroblasts²⁰. It will be interesting to
358 test in the future whether myofibroblast inhibition of AT2 cell fate occurs in pulmonary fibrosis
359 where myofibroblasts are expanded and AT2 cells lost.

360
361 We clearly demonstrate that NKX2.1 is a key upstream TF driving the onset of the alveolar
362 program whilst suppressing the airway program (Fig. 5), consistent with reported roles in lung
363 cancers²¹. Our data also strongly suggest that Wnt is upstream of NKX2.1 during alveolar
364 differentiation. By contrast, FGF signalling and TFAP2C cooperate to promote airway fate.
365 Ectopic expression of TFAP2C promoted the expression of basal cell markers, but not other
366 airway lineages (Fig. 5). This could mean the culture conditions are permissive only for basal
367 cell differentiation. Alternatively, TFAP2C may be specific for basal cell specification. The
368 latter interpretation would be consistent with a report that TFAP2C activates TP63 expression
369 during epidermal lineage maturation²².

370

371 In summary, we have identified a distinct late-tip progenitor cell state in the developing
372 human lung. Culture of these cells as organoids has allowed us to investigate the roles of Wnt
373 signalling and NXX2.1 in human AT2 cell development. Moreover, this differentiating
374 organoid system will be useful for understanding the next phases of the alveolar maturation
375 process at the sacular and alveolar stages in the developing human lung.

376

377 **ACKNOWLEDGEMENTS**

378 We would like to acknowledge the Gurdon Institute Imaging Facility and Dr Karin Mueller of
379 Cambridge Advanced Imaging Centre for microscopy support. Prof Azim Surani for
380 epigenetics advice. KL is supported by Basic Science Research Program through the National
381 Research Foundation of Korea (NRF) funded by the Ministry of Education
382 (2018R1A6A3A03012122). DS is supported by a Wellcome Trust PhD studentship
383 (109146/Z/15/Z) and the Department of Pathology, University of Cambridge ELR. is supported
384 by Medical Research Council (MR/P009581/1; MR/S035907/1). Gurdon Institute Core
385 support from the Wellcome Trust (203144/Z/16/Z) and Cancer Research UK (C6946/A24843).
386 KBM and SAT acknowledge funding from the MRC (MR/S035907/1) and from Wellcome
387 (WT211276/Z/18/Z and Sanger core grant WT206194).

388

389

390 **AUTHOR CONTRIBUTIONS**

391 Conceptualization, KL and ELR; Methodology, Investigation, and Validation, KL, DS ;
392 Software and Formal Analysis, KL, WT, PH; Writing–Original Draft by KL; Writing – Review
393 & Editing, ELR; Funding Acquisition and Supervision: ELR, SAT, JCM, KBM.

394

395

396 **FIGURE LEGENDS**

397 **Fig. 1. Human fetal lung tip progenitor cells acquire alveolar features during normal**
398 **development.**

399 (A and B) Human fetal lung at pseudoglandular and canalicular stages; 11, 19 pcw (A) and 10,
400 20 pcw (B). Tip epithelium (arrowheads) is marked by E-cadherin, SFTPC, HTII-280, and
401 SOX9. SOX2, airway epithelium (A).

402 (C-F) Surface antigens, CD44 and CD36, mark tip epithelium at pseudoglandular and
403 canalicular stages. Lungs at 10 (C), 17 (D, F), and 19 pcw (E) were stained with CD36 and
404 CD44 and/or SOX2 and SOX9 antibodies. Arrows (D) show patterning from distal to proximal
405 regions. Yellow dashed lines (E) indicate separation of SOX9⁺ tip regions from the PDPN⁺
406 stalk. The *SFTPC* transcript was visualised by *in situ* HCR (F) following immunostaining for
407 CD36.

408 (G) Flow cytometry of the human lung tip epithelial population at the pseudoglandular and
409 canalicular stages, 11 (*blue*) and 20 pcw (*red*).

410 (H) qRT-PCR of the freshly purified lung epithelial cells from canalicular stage lungs sorted.
411 Data normalized to fresh EPCAM⁺ cells from 20 pcw distal tissues; mean \pm SD, $n = 7$ (15~21
412 pcw). Significance evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns:
413 not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

414 (I) Proportion of the freshly purified tip epithelium as CD44⁺CD36⁻ or CD44⁺CD36⁺ at 11, 13,
415 15, 18, 20 and 21 pcw; $n = 1$ each time.

416 (J) Relative mRNA levels of the tip progenitor markers, *SOX9* and *SOX2*, and type 2 alveolar
417 lineage markers, *SFTPC*, *CD36* and *LAMP3*, in CD44⁺CD36⁻ tip epithelial population at 11
418 and 13 pcw, and in CD44⁺CD36⁺ tip epithelial population at 15, 16, 18, 20 and 21 pcw, by
419 qRT-PCR. Data was normalized to fresh EPCAM⁺ cells from 20 pcw tip tissues; $n = 1$ at each
420 stage.

421 (K and L) Human fetal lung tissues during the transition from 13 to 15 pcw were stained using
422 antibodies against CD36, CD44, and CD31 (K), or for *SFTPC* and *SOX9* mRNA (L). Three 13
423 pcw, two 14 pcw, and two 15 pcw samples.

424 (M) Signal intensity of *SOX9* and *SFTPC* transcripts in Fig. 1L. Ten tip regions analysed per
425 stage and the intensity represented as mean \pm SD. Significance evaluated by 1-way ANOVA
426 with Tukey multiple comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$,
427 **** $P < 0.001$.

428 DAPI, nuclei. Scale bar, 50 μ m.

429

430 **Extended Data Fig. 1. Characterization of the lung tip epithelium at the canalicular stages.**

431 (A and B) Frozen sections of human fetal lung tissues at 17 pcw. Stained for CD36, E-cadherin
432 and PDPN (A) and CD36, CD44 (B). Arrowheads indicate CD36⁺PDPN⁻ tips. Inset (B) shows
433 a CD44⁺, CD36⁻ stalk epithelial region. DAPI, nuclei. Scale bars, 50 μ m.

434

435 **Fig. 2. CD36, CD44 dual-positive tip cells self-renew and undergo lineage commitment *in***
436 ***vitro* to form canalicular stage lung organoids.**

437 (A) Isolation and viral infection of CD44⁺CD36⁺ tip epithelial cells from human fetal lungs at
438 the canalicular stage and *in vitro* culture in self-renewing (SN) medium.

439 (B and C) Gross morphology (B) of the cultured epithelial tip organoids. Detailed morphology
440 (C) E-cadherin (magenta); *SFTPC*-GFP and *TagRFP*. Arrows and arrowhead indicate folded
441 Lin^{POS} organoids and cystic Lin^{NEG} organoids, respectively. Scale bars, 100 μm .
442 (D) Gene expression profile of the Lin^{POS} and Lin^{NEG} organoids. Data are quantified by qRT-
443 PCR; mean \pm SD of 4 biological replicates. Significance evaluated by unpaired student *t*-test;
444 $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.
445 (E) Immunofluorescence analysis of the Lin^{POS} (arrowheads) and Lin^{NEG} organoids (arrow) at
446 passage 1 cultured in the SN medium, showing co-expression of SOX9 and SOX2. DAPI,
447 nuclei. Scale bar, 50 μm .
448 (F and G) Canalicular stage lung tip organoids sorted into 3 populations at passage zero by
449 FACS using antibodies against CD36 and CD44 (F). The sorted P0 cell populations were
450 analysed by qRT-PCR (G). Data was normalized to total EPCAM⁺ cells freshly sorted from 20
451 pcw tissues; mean \pm SD ($n = 5$). Significance was evaluated by 1-way ANOVA with Tukey
452 multiple comparison post-test; $*P < 0.05$, $**P < 0.01$.
453 (H) Tip organoids at passage zero cultured in self-renewal medium stained with SFTPC, HTII-
454 280 and SOX2 antibodies. Arrowheads indicate the tip-like, SFTPC⁺ subpopulation in the
455 Lin^{POS} organoids. Arrows indicate the Lin^{NEG} organoids.
456 (I-K) Passage 1 organoids were grown from the sorted CD44⁺CD36⁺, CD44⁺CD36⁻ or CD44⁻
457 CD36⁻ populations at Passage 0 and reanalysed for CD44 and CD36 at the end of Passage 1 (I;
458 Extended Data Fig. 2L). Arrows and arrowheads indicate the Lin^{POS} and Lin^{NEG} organoids. The
459 organoid forming efficiency (J) and the proportion (K) of the organoids of each morphological
460 sub-type at passage 1 was measured at 3 weeks after plating. Data was represented as mean \pm
461 SD of 4 biological replicates. Scale bar, 100 μm .
462 (L) Hierarchical clustering analysis of bulk-RNA seq data using pseudoglandular, Lin^{POS} and
463 Lin^{NEG} organoids.
464 (M) Heatmap analysis of selected genes highly enriched in pseudoglandular, Lin^{NEG} and Lin^{POS}
465 organoids.

466

467 **Extended Data Fig. 2. Characterization of the canalicular stage lung tip organoids.**

468 (A and B) Immunofluorescence analysis of the Lin^{POS} (arrowheads) and Lin^{NEG} organoids
469 (arrow) at passage 1 cultured in the self-renewing medium, showing the alveolar lineage
470 markers, SFTPC, HTII-280 (A) and LAMP3 (B, lower panel) were expressed in the Lin^{POS}
471 organoids, but not in the Lin^{NEG} organoids. DAPI, nuclei. Scale bar, 50 μm .
472 (C-E) Immunofluorescence images (C-E) of the Lin^{NEG} and Lin^{POS} organoids originating from
473 sorted CD44⁺CD36⁺ tip epithelium from 20 pcw lung. Antibodies against SOX9, HTII-280 (C),
474 CD44, TP63, SOX2 (D) and CD44, KI67 (E). Arrow (E) indicates a Lin^{NEG} organoid. DAPI
475 indicates nuclei. Scale bar, 50 μm .
476 (F) The percentage of KI67⁺ cells in the CD44⁺ tip and CD44⁻ non-tip regions are represented
477 as mean \pm SD of biological 4 replicates.
478 (G-I) Diagram (G) illustrating isolation and *in vitro* culture of the tip epithelial cells infected
479 with lentivirus harbouring *SFTPC* promoter-driven eGFP and EF1a promoter-driven TagRFP.
480 The cells expressing *SFTPC-eGFP* were sorted at 48h post-infection and cultured in the self-
481 renewing medium for 3 weeks (H). The Lin^{NEG} and Lin^{POS} organoids at passage zero derived
482 from the *SFTPC-eGFP* positive-sorted tip epithelial cells (I). Low eGFP signals remained in

483 some Lin^{NEG} organoids (arrowhead) confirming their derivation from SFTPC⁺ tip cells. Scale
484 bar, 50 μ m.

485 (J and K) Expression of lineage markers was investigated by immunostaining (J) and qRT-
486 PCR (K) in the Lin^{POS} organoids and airway-like spheres at passage 1 derived from the
487 CD44⁺CD36⁺ or CD44⁻CD36⁻ passage zero subpopulations respectively. Data was normalized
488 to EPCAM⁺ cells freshly sorted from 20 pcw tip tissues and represented as mean \pm SD of 4
489 biological replicates. Significance was evaluated by 2-way ANOVA with Bonferroni multiple
490 comparison post-test; ns: not significant, * P <0.05.

491 (L) Diagram summarising the organoid experiments performed in Fig. 2. CD44⁺CD36⁺ cells
492 from canalicular stage lung tissues are the major tip progenitor subpopulation *in vitro*, growing
493 into self-renewing Lin^{POS} organoids showing key features of the canalicular stage lung tip cells.

494

495 **Extended Data Fig. 3. Transcriptomic analysis of the lung tip organoids**

496 (A) Morphology of the pseudoglandular organoids from the pseudoglandular stage, and the
497 Lin^{POS} and Lin^{NEG} organoids from the canalicular stage. The Lin^{POS} and Lin^{NEG} organoids
498 grown from EPCAM⁺ tip epithelial cells were manually separated and cultured in the self-
499 renewing culture condition.

500 (B) Principal component analysis of bulk-RNA seq data using pseudoglandular organoids,
501 Lin^{POS} and Lin^{NEG} organoids.

502 (C) Volcano plot showing differentially expressed genes between Lin^{POS} organoids (*red*)
503 versus Lin^{NEG} organoids (*dark blue*); log₂FC > 4.

504 (D) Gene ontology (GO) enrichment analysis performed for biological process (BP)-
505 associated GO terms on the differentially expressed genes between the Lin^{POS} and Lin^{NEG}
506 organoids; log₂FC > 4.

507 (E) KEGG pathway analysis using Enrichr. Length of coloured bars indicates combined
508 enrichment score by adjusted p-value < 0.05.

509 (F and G) Gene set enrichment (GSEA) analysis of the differentially expressed genes of the
510 Lin^{POS} organoids (F) and pseudoglandular organoids (G).

511

512 **Fig. 3. Wnt and FGF signalling coordinate human tip cell maintenance and differentiation** 513 *in vitro*

514 (A) Diagram showing *in vitro* culture of the tip epithelial cells for 2 weeks in single or pairwise
515 combinations of signalling cues: FGFs (FGF7, FGF10), SMADi (Noggin, SB431542), CHIR
516 (CHIR99021) and EGF.

517 (B) Morphology of the tip organoids cultured in different culture conditions for 2 weeks.
518 Representative image from 1 biological replicate is shown; n=4 biological replicates in total.
519 Scale bar, 200 μ m.

520 (C) Relative mRNA levels of *SOX9*, *SFTPC* and *TP63* measured by qRT-PCR. Normalized to
521 a Lin^{POS} organoid line; mean \pm SD of 4 independent biological replicates. Significance was
522 evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant,
523 * P <0.05 and ** P <0.01.

524 (D) Immunofluorescence analysis of the tip organoids at passage zero cultured in the different
525 culture conditions. Antibodies against E-cadherin, SFTPC, TP63 and SOX2 were used. Scale
526 bar, 20 μ m.

527 (E) Immunofluorescence analysis of the pseudoglandular organoids derived from an 8 pcw
528 pseudoglandular stage lung in the different conditions at passage 0. Antibodies against SFTPC,
529 TP63, E-cadherin and SOX2 were used. Scale bar, 50 μm .

530 (F-H) Frozen sections of human fetal lung tissues at 17 pcw (F) and 20 pcw (G) were
531 immunostained for AXIN2, E-cadherin and SOX2 (F), or ACTA2 followed by *in situ* HCR for
532 *SFTPC* and *AXIN2* (G). Arrowheads (F) indicate AXIN2⁺ tip epithelial cells. Red dashed line
533 in the inset (G) indicates SFTPC⁺ tip epithelial cells. *AXIN2* signals were counted from 20 areas
534 of tips and stalks and 10 areas of airway across 3 independent lung tissues at 18-20 pcw (H).
535 Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns:
536 not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$. Scale bar, 50 μm .
537 DAPI indicates nuclei.

538

539 **Extended Data Fig. 4. High plasticity of the canalicular stage tip epithelial cells in**
540 **response to Wnt and FGF signalling.**

541 (A) The population of EPCAM⁺ tip epithelial cells in the distal lung tissue at 20 pcw was
542 isolated using MACS and were analysed by CD44 and CD36 expression using FACS.

543 (B) Tip epithelial cells cultured with, or without, SMAD inhibition for 3 weeks. The self-renew
544 medium condition (w/ SMADi; *lower* panel) was used for positive control. Scale bar, 200 μm .

545 (C, D) After 2 weeks growing in SMADi conditions, cells were transferred to culture medium
546 containing CHIR, or FGFs, or CHIR/FGFs. After 2 weeks of exposure to the different culture
547 conditions, the epithelial organoids were stained with lineage makers including SFTPC or TP63
548 (C) and relative mRNA levels of *SFTPC* and TP63 were measured by qRT-PCR (D). Data were
549 normalized to the Lin^{POS} organoids; mean \pm SD of at least 4 biological replicates. Significance
550 was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant,
551 * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. Scale bar, 200 μm .

552 (E) After 2 weeks growing in SMADi/FGFs (or SMADi/CHIR) the organoids were transferred
553 to SMADi/CHIR (or SMADi/FGFs) for another 2 weeks. After 2 weeks of exposure to the
554 different culture conditions organoid morphology altered and qRT-PCR for lineage markers
555 was performed. Data were normalized to the Lin^{POS} organoids; mean \pm SD of 3 biological
556 replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison
557 post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. Scale bar, 200 μm .

558 (F and G) Tip epithelial cells cultured with FGF7, FGF10, or FGF7 and FGF10 (F); scale bar,
559 200 μm . After 2 weeks of exposure to each culture condition, the organoids were
560 immunostained for SOX2, TP63 and E-cadherin (G); scale bar, 50 μm .

561 (H) Gene expression profile of the freshly isolated lung epithelial cells from the canalicular
562 stage human lung tissues sorted by CD44⁺CD36⁺, CD44⁺CD36⁻ and CD44⁻CD36⁻. Data
563 normalized to freshly isolated EPCAM⁺ cells from 20 pcw tip tissues; mean \pm SD of 5
564 biological replicates aged from 15~20 pcw. Significance was evaluated by 1-way ANOVA
565 with Tukey multiple comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$ and
566 *** $P < 0.001$.

567 (I) Frozen sections of human fetal lung tissues at 20 pcw were stained for ACTA2 followed by
568 *in situ* HCR for *SFTPC* and *AXIN2*.

569

570 **Fig. 4. Spatial patterning of the differentiating alveolar epithelium by NOTUM+**

571 **myofibroblasts.**

572 (A-C) Frozen sections of 15-17 pcw human fetal lung stained by *in situ* HCR and/or antibodies.
573 A. 16 pcw, *SFTPC*, *WNT2*, and *FGFR4* probes. B. 15 pcw, *SFTPC* and *NOTUM* probes,
574 ACTA2 antibody. C. 17 pcw *SFTPC*, *WNT2*, and *NOTUM* probes. Arrows (B) and asterisks
575 (A,C) represent ACTA2⁺*NOTUM*⁺ myofibroblasts in the tissues. Lines and dashed lines
576 indicate the boundaries of epithelial cells and myofibroblasts, respectively. Scale bar, 50 μ m.

577 (D) Isolation of PDGFRa⁺CD141⁺ myofibroblasts and PDGFRa⁻CD141⁻ alveolar fibroblasts
578 from human fetal lung tissues at 17-21 pcw using a combination of PDGFRa-APC and CD141-
579 PE antibodies.

580 (E) qRT-PCR of PDGFRa⁺CD141⁺ myofibroblasts and PDGFRa⁻CD141⁻ alveolar fibroblasts
581 freshly isolated from 17 to 21 pcw human lung tissues. Data was normalized to the total isolated
582 fibroblast population; mean \pm SD of biological 4 replicates. Significance was evaluated by
583 unpaired student *t*-test; **P*<0.05, ***P*<0.01, ****P*<0.001.

584 (F) Diagram illustrating *in vitro* coculture of the isolated PDGFRa⁺CD141⁺ myofibroblasts and
585 PDGFRa⁻CD141⁻ alveolar fibroblast with Lin^{POS} tip organoids expressing *SFTPC-GFP*.

586 (G and H) *SFTPC-GFP* signal of Lin^{POS} tip organoids. G. Cultured alone in self-renewing (SN)
587 or DMEM + 2% FBS medium. H. Cocultured with freshly isolated fibroblast sub-populations.
588 I, insert; P, plate. Scale bar, 200 μ m.

589 (I) qRT-PCR for *SFTPC* and *LAMP3*, 2 weeks after *in vitro* culture. Mean \pm SD of 3 biological
590 replicates. Significance was evaluated by one-way ANOVA; ns: not significant, **P*<0.05,
591 ***P*<0.01, ****P*<0.001.

592 (J) Thick sections of human fetal lung immunostained with ACTA2 followed by *in situ* HCR
593 for *SFTPC*. Lines and dashed lines in the inset indicate *SFTPC*⁺ epithelial cell populations
594 located in tip and stalk regions, respectively. Thickness: 42 μ m. See also Supplementary Video
595 1.

596 (K) Summary diagram showing the spatial regulation of Wnt signalling mediated by
597 ACTA2⁺PDGFRa⁺CD141⁺ myofibroblasts in the distal regions of human lung tissues during
598 the canalicular stage.

599 DAPI indicates nuclei. Scale bar, 50 μ m.

600

601 **Extended Data Fig. 5. Wnt-responsive NOTUM⁺ myofibroblasts in the distal human fetal**
602 **lung.**

603 (A-F) Human fetal lung sections at 17 (A, D, E) and 19 pcw (B, C, F). A. *SFTPC*, *WIF1*,
604 *NOTUM*. B. ACTA2, *NOTUM*. C. *NOTUM*, *LEF1*. D. *SFTPC*, *AXIN2*. E. CD44, ACTA2,
605 PDGFRa. F. ACTA2, CD44, CD141. Arrows and asterisks indicate ACTA2⁺ PDGFRa⁺
606 *NOTUM*⁺ myofibroblasts. Lines and dashed lines indicate the boundaries of epithelial cells and
607 myofibroblasts, respectively.

608 (G) Relative mRNA levels from myofibroblasts and alveolar fibroblasts cultured alone, or
609 cocultured in transwells. Data were normalized to the whole freshly isolated lung fibroblast
610 population; mean \pm SD of biological 4 replicates. Significance was evaluated by unpaired
611 student *t*-test; **P*<0.05, ***P*<0.01, ****P*<0.001.

612 DAPI indicates nuclei. Scale bar, 50 μ m.

613

614 **Fig. 5. Identification of key transcription factors controlling airway and alveolar lineage**

615 **differentiation using the organoid system.**

616 (A) Genomic distribution of differentially accessible chromatin regions associated with human
617 fetal lung development between the pseudoglandular and Lin^{POS} organoids.

618 (B) Representative ATAC-seq tracks visualized in Integrative Genomics Viewer (IGV) at
619 *SFTPC*, *TP63*, *AXIN2* and *CTNNB1*. Red box indicates the promoter.

620 (C and D) HOMER motif analysis coupled with RNA seq data. The top 15 most highly enriched
621 motifs and TF gene expression level (heat map) are shown for the pseudoglandular (C) and
622 Lin^{POS} organoids (D).

623 (E) Pseudoglandular and Lin^{POS} organoids stained with antibodies against NKX2.1, TFAP2C,
624 CD44 and E-cadherin.

625 (F) Diagram showing doxycycline-inducible overexpression of NKX2.1 and/or TFAP2C in the
626 pseudoglandular organoids. Constitutively expressed TagRFP was used for sorting transduced
627 cells.

628 (G) Morphology of the pseudoglandular organoids overexpressing NKX2.1 or TFAP2C for 2
629 weeks. Scale bar, 100 μ m.

630 (H) Relative mRNA levels of *SFTPC* and *TP63* were measured by qRT-PCR in NKX2.1- or
631 TFAP2C-OE pseudoglandular organoids. Data was normalized to EPCAM⁺ cells freshly
632 isolated from 20 pcw tip tissues; mean \pm SD of four biological replicates. Significance was
633 evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant,
634 **** $P < 0.0001$.

635 (I and J) SFTPC and TP63 antibody staining of the pseudoglandular organoids overexpressing
636 NKX2.1 or TFAP2C for 2 weeks (I). The proportion of the organoids positively stained in (J)
637 was measured based on morphology and signal intensity. N= 4 biological replicates.

638 (K) qRT-PCR of the pseudoglandular organoids overexpressing NKX2.1 for 2 weeks. Data was
639 normalized to EPCAM⁺ cells freshly isolated from 20 pcw tip tissues; mean \pm SD of four
640 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple
641 comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

642 (L) Knock-down (KD) of endogenous *NKX2.1* by CRISPR-dCas9-KRAB system. Data was
643 normalized to EPCAM⁺ cells freshly isolated from 20 pcw tip tissues; mean \pm SD of 5
644 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple
645 comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

646 DAPI indicates nuclei. Scale bar, 50 μ m.

647

648 **Extended Data Fig. 6. Chromatin accessibility analysis of the pseudoglandular and Lin^{POS}**
649 **organoids.**

650 (A) Analysis of chromatin accessibility in the pseudoglandular and Lin^{POS} organoids by bulk-
651 ATAC seq. The number of differential accessible chromatin regions which differ between the
652 organoids was identified; fold change > 2 and FDR < 0.05 .

653 (B) Pie charts representing the genomic distribution of global accessible chromatin regions in
654 pseudoglandular organoids and Lin^{POS} organoids.

655 (C) Biological Process-associated GO term analysis using the differential accessible chromatin
656 regions highly enriched in the pseudoglandular organoids and Lin^{POS} organoids.

657 (D and E) IGV image shots of representative ATAC seq tracks at loci showing differential
658 accessible chromatin regions between pseudoglandular organoids and Lin^{POS} organoids. Red

659 box indicates the promoter regions.

660

661 **Extended Data Fig. 7 NKX2.1 drives the onset of the alveolar program whilst suppressing**
662 **the airway program**

663 (A and B) *In situ* HCR images for detecting transcripts, *SFTPC* or *TFAP2C* with *NKX2.1* at 19
664 pcw.

665 (C) Transgene induction following doxycycline treatment for 2 weeks measured by qRT-PCR.
666 Data was normalized to the untreated group; mean \pm SD of four biological replicates.
667 Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns:
668 not significant, **** $P < 0.0001$.

669 (D) qRT-PCR of the pseudoglandular organoids overexpressing *TFAP2C* for 2 weeks. Data
670 was normalized to *EPCAM*⁺ positive cells freshly isolated from 20 pcw tip tissues; mean \pm SD
671 of biological 4 replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple
672 comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

673 (E) Morphology and fluorescent images of pseudoglandular organoids overexpressing both
674 *NKX2.1* and *TFAP2C* for 2 weeks.

675 (F) qRT-PCR of endogenous *NKX2.1*, *SFTPC*, and *TP63* in the *Lin*^{POS} organoids cultured in
676 medium containing CHIR, FGF7 or CHIR/FGF7. Data was normalized to *EPCAM*⁺ cells
677 freshly isolated from 20 pcw tip tissues; mean \pm SD of three biological replicates. Significance
678 was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant,
679 * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

680 DAPI indicates nuclei. Scale bar, 50 μ m.

681

682 **Fig. 6. Analysis of naturally occurring human genetic variation using organoid assays.**

683 (A and B) Morphology (A) and gene expression profile (B) in the pseudoglandular organoids
684 overexpressing wildtype *NKX2.1*, or a *NKX2.1* lacking a DNA binding domain (DBD
685 deletion), cultured for 2 weeks in the presence, or absence, of doxycycline (\pm DOX).

686 (C) Diagram describing overexpression of wild type and mutant forms of *NKX2.1* in the
687 pseudoglandular organoids using doxycycline inducible lentiviral system. Individually, five
688 different mutations were introduced into the DNA-binding homeobox domain; deletion of
689 Arg¹⁶² (R162del)¹⁴, two nonsense point mutations (Q175*, R178*)¹⁴, and two missense point
690 mutations (I207F, I207M)^{12,13} were tested.

691 (D-F) Morphology and immunostaining (D), qRT-PCR (E), and western blot (F) analysis of the
692 pseudoglandular organoids following overexpression of wildtype or mutant human *NKX2.1*
693 for 1 week. Data were normalized to doxycycline-non-treated lines; mean \pm SD of 4 biological
694 replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison
695 post-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Western blot showing mature *SFTPB* and *SFTPC*.
696 *GAPDH* was used for a loading control. DAPI indicates nuclei. Scale bar, 50 μ m.

697

698 **Fig. 7. Efficient *in vitro* differentiation of *Lin*^{POS} organoids to alveolar type 2 cell fate.**

699 (A) qRT-PCR for *ACE2* in the pseudoglandular organoids overexpressing *NKX2.1* or *TFAP2C*
700 for 1 or 2 weeks. Data was normalized to fresh *EPCAM*⁺ cells from 20 pcw tip tissues; mean
701 \pm SD of four biological replicates. Significance was evaluated by 1-way ANOVA with Tukey
702 multiple comparison post-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

703 (B) qRT-PCR of the Lin^{POS} organoids cultured in SN medium or alveolar-induction culture
704 conditions containing combinations of DAPT, Dexamethasone/cyclic AMP/ IBMX (DCI),
705 CHIR with/without SB431542 (SB) and FGF7, for 1 week. *NKX2.1*, *SFTPC*, and *SOX9* levels
706 were normalized to EPCAM⁺ cells, and *ACE2* was normalized to EPCAM⁻ cells, freshly sorted
707 from 20 pcw tip tissues; mean \pm SD of four biological replicates. Significance was evaluated
708 by 1-way ANOVA with Dunnett multiple comparison post-test; * P <0.05, ** P <0.01,
709 *** P <0.001.

710 (C) Morphology and fluorescent images of the Lin^{POS} organoids cultured in DAPT/DCI/CHIR
711 with SB, or in control SN medium, for 1 week.

712 (D) Immunofluorescent analysis of *SFTPC*-GFP⁺ Lin^{POS} organoids cultured in
713 DAPT/DCI/CHIR with SB or in the SN medium for 1 week. DAPI indicates nuclei. Scale bar,
714 50 μ m.

715 (E) qRT-PCR of *NKX2.1* and *SFTPC* in the pseudoglandular organoids cultured in the SN
716 medium or in DAPT/DCI/CHIR plus SB. Data were normalized to EPCAM⁺ cells freshly
717 isolated from 20 pcw tip tissues; mean \pm SD of three biological replicates. Significance was
718 evaluated by 1-way ANOVA with Tukey multiple comparison post-test; **** P <0.0001.

719 (F) Electron microscopy images of Lin^{POS} organoids cultured in DAPT/DCI/CHIR with SB
720 (*left*) or SN medium (*right*). LBs, lamellar bodies; PC, projection core; MV, microvilli; GC,
721 glycogen; PLM, Primitive lipid membrane within a pool of monoparticulate glycogen at an
722 early stage in the formation of LBs. Scale bar, 3 μ m.

723 (G) Numbers of LBs per cells in Lin^{POS} organoids cultured in SN medium (*red*) or
724 DAPT/DCI/CHIR with SB (*blue*). Immature and mature LBs were measured in total 40 cells
725 from two biological samples for each condition. Significance evaluated by unpaired student *t*-
726 test; * P <0.05, ** P <0.01, *** P <0.001, **** P <0.0001.

727 (H) Western blot showing NKX2.1 levels and SFTPB/SFTPC processing in the Lin^{POS}
728 organoids cultured in the SN medium, or in DAPT/DCI/CHIR/SB and the pseudoglandular
729 organoids cultured in the SN medium for 1 week. GAPDH was used for a loading control.

730 (I) Chromatin immunoprecipitation (ChIP)-qPCR analysis for quantifying relative enrichment
731 of NKX2.1 binding on the promoter regions of type 2 alveolar lineage markers, *SFTPC*,
732 *LAMP3*, and *SLC34A2* in the organoids cultured in the SN medium, or in DAPT/DCI/CHIR
733 plus SB. Data was normalized to the IgG control; mean \pm SD of three biological replicates.

734

735 **Extended Data Fig. 8. Type 2 alveolar differentiation of the cultured canalicular stage tip**
736 **cells.**

737 (A and B) Immunofluorescent analysis of ACE2 in (A) pseudoglandular organoids
738 overexpressing NKX2.1 and (B) Lin^{POS} organoids were cultured in DAPT/DCI/CHIR/SB or in
739 the SN medium for 1 week. Phalloidin (F-actin) marks apical membrane of epithelial cells in
740 the organoids. Scale bar, 50 μ m.

741 (C) qRT-PCR of Lin^{POS} organoids cultured in the SN medium or in DAPT/DCI/CHIR with SB,
742 or A83-01, for 1 week. Data were normalized to EPCAM⁺ cells freshly isolated from 20 pcw
743 tip tissues; mean \pm SD of four biological replicates. Significance was evaluated by 1-way
744 ANOVA with Tukey multiple comparison post-test; * P <0.05, ** P <0.01, *** P <0.001.

745

746

747 **MATERIALS and METHODS**

748 **Human embryonic and foetal lung tissue**

749 Human embryonic and foetal lung tissues were provided from terminations of pregnancy from
750 Cambridge University Hospitals NHS Foundation Trust under permission from NHS Research
751 Ethical Committee (96/085) and the MRC/Wellcome Trust Human Developmental Biology
752 Resource (London and Newcastle, University College London (UCL) site REC reference:
753 18/LO/0822; Newcastle site REC reference: 18/NE/0290; Project 200454; www.hdbr.org).
754 Sample age ranged from 4 to 23 weeks of gestation (post-conception weeks; pcw). Stages of
755 the samples were determined according to their external physical appearance and
756 measurements. All the samples used for the current study had no known genetic abnormalities.

757 ***In vitro* culture of human fetal lung organoids**

758 The isolated tip epithelial cells were embedded in Matrigel (Corning, 356231) and cultured in
759 48-well plates in self-renewal (SN) medium: Advanced DMEM/F12 supplemented with 1x
760 GlutaMax, 1 mM HEPES and Penicillin/Streptomycin, 1X B27 supplement (without Vitamin
761 A), 1X N2 supplement, 1.25 mM n-Acetylcysteine, 50 ng/ml recombinant human EGF
762 (PeproTech, AF-100-15), 100 ng/ml recombinant human Noggin (PeproTech, 120-10C), 100
763 ng/ml recombinant human FGF10 (PeproTech, 100-26), 100 ng/ml recombinant human FGF7
764 (PeproTech, 100-19), 3 μ M CHIR99021 (Stem Cell Institute, University of Cambridge) and 10
765 mM SB431542 (Bio-Techne, 1614). The culture medium was replaced every 2 days and the
766 organoids were usually split 1:3 once per week by breaking them into small fragments.
767 Numbers of replicates are indicated in figure legends. 10 ng/ μ l recombinant human BMP4
768 (PeproTech, 120-05) and 10 ng/ μ l recombinant human TGF- β 1 (PeproTech, 100-21) were added
769 to the medium instead of adding SB431542 and Noggin to activate dual SMAD signalling
770 (Extended Data Fig. 4B).

771 To perform *in vitro* co-culture experiments, freshly sorted 2×10^5 PDGFRA⁺CD141⁺
772 myofibroblasts or PDGFRA⁻CD141⁻ fibroblasts were mixed with *SFTPC*-eGFP⁺ Lin^{POS}
773 organoids in 100 μ l Matrigel and then loaded into an insert of transwell (Merck, CLS3493).
774 On the bottom well plates, coated with Collagen (Merck, CLS3493), 4×10^5 PDGFRA⁻CD141⁻
775 fibroblasts were plated in the culture medium containing 2% fetal bovine serum (FBS; Thermo,
776 10500064) in the Advanced DMEM/F12 supplemented with 1x GlutaMax, 1 mM HEPES and
777 Penicillin/Streptomycin. 100 μ l culture medium was added every 2 days. For coculture of
778 PDGFRA⁺CD141⁺ myofibroblasts with PDGFRA⁻CD141⁻ fibroblasts, 2×10^5 of
779 PDGFRA⁺CD141⁺ myofibroblasts were plated on the insert and 4×10^5 of PDGFRA⁻CD141⁻
780 fibroblasts were plated on the bottom well plate. After 2 weeks of co-cultures the organoids, or
781 the mesenchymal cells, were harvested for further analysis.

782 **Isolation of tip epithelial cells, myofibroblasts, and alveolar fibroblasts**

783 For isolation of CD44⁺, or CD44⁺CD36⁺, tip epithelium directly from the distal lung tissues,
784 the tissues were finely dissected into tiny pieces and enzymatically digested into single cells
785 by incubating them in a dissociation solution containing 0.125 mg/ml Collagenase (Merck,
786 C9891), 1 U/ml Dispase (Thermo Fisher Scientific, 17105041) and 0.1 U/ μ l DNAase (Merck,
787 D4527), in a rotating incubator for 1 h at 37°C. After rinsing in washing buffer containing 2%
788 FBS in cold PBS the cells were filtered by 100 μ m strainer and harvested by centrifugation.

791 The cell pellets were resuspended and treated with RBC lysis buffer (BioLegend, 420301).
792 Next, the cells were rinsed in the washing buffer and then incubated with primary antibodies
793 against CD45 (1:100; PE-Cy7 conjugated, Thermo Fisher Scientific, 25-9459-42), CD31
794 (1:100; PE-Cy7 conjugated, Thermo Fisher Scientific, 25-0319-42), EPCAM (1:100; PE-
795 conjugated; BioLegend, 324206), CD44 (1:200; APC-conjugated; BioLegend, 103012), and
796 CD36 (1:100; FITC, conjugated; Thermo Fisher Scientific, 11-0369-42), with a viability dye,
797 Zombie (Biolegend, 423113) for 25 min on ice. Following removal of dead cells and
798 immune/endothelial cells, the EPCAM⁺ epithelial cells were sorted by CD44 and/or CD36
799 expression by FACS (BD Influx™ Cell Sorter) (Fig. 1G).

800 Alternatively, the tip epithelial cells were isolated by EPCAM⁺ magnetic-activated cell
801 sorting (MACS) beads according to the manufacturer's instruction (CD326 MicroBeads,
802 human, Miltenyi Biotec) from the distal lung tissues. Then, the enriched EPCAM⁺ epithelial
803 cells were sorted by CD44 and/or CD36 expression by FACS (SH800S Cell Sorter) to more
804 purely enrich the tip cell population (Fig. 2A).

805 To purify myofibroblasts and alveolar fibroblasts, the single cells dissociated from the
806 distal lung tissues were incubated with the following primary antibodies: CD45 (1:100; PE-
807 Cy7 conjugated, Thermo Fisher Scientific, 25-9459-42), CD31 (1:100; PE-Cy7 conjugated,
808 Thermo Fisher Scientific, 25-0319-42), CD9 (1:100; PE-Cy7 conjugated, BioLegend, 312115),
809 EPCAM (1:100; FITC-conjugated, 324204), PDGFRA (1:100; APC-conjugated, BioLegend,
810 313511), CD141 (1:100; PE-conjugated, BioLegend, 344104), with the viability dye, Zombie
811 (Biolegend, 423113). After removing dead cells, immune/endothelial cells, airway smooth
812 muscle cells, and epithelial cells, the cells are sorted by PDGFRA and/or CD141 expression
813 using BD Influx Cell Sorter. The sorted cells were directly applied to an organoid coculture or
814 a gene expression analysis.

815

816 **Type 2 alveolar differentiation of Lin^{POS} organoids**

817 The Lin^{POS} organoids were embedded in Matrigel and cultured for 7 days in alveolar type 2
818 (AT2) differentiation medium: Advanced DMEM/F12 supplemented with 1x GlutaMax, 1 mM
819 HEPES and Penicillin/Streptomycin, 1X B27 supplement (without Vitamin A), 1x N2
820 supplement, 1.25 mM n-Acetylcysteine, 10mM CHIR99021, 50 μM Dexamethasone (Merck,
821 D4902), 0.1 M 8-Bromoadenosine 3'5'-cyclic monophosphate (cAMP; Merck, B5386), 0.1 M
822 3-Isobutyl-1-methylxanthine (IBMX; Merck, 15679), 50 mM DAPT (Merck, D5942) with 10
823 mM SB431542 or 10 mM A83-01 (Tocris, 2939). The culture medium was replaced every 2
824 days without passaging.

825

826 **Lentiviral transduction**

827 To introduce a reporter system into the tip epithelial cells, the lentiviral vector pHAGE hSPC-
828 eGFP-W given from Darrell Kotton (Addgene plasmid # 36450; <http://n2t.net/addgene:36450>;
829 RRID: Addgene_36450) was modified by inserting EF1a-promoter TagRFP cassette. The tip
830 epithelial cells were infected with the modified lentiviral vector for 24 h at 37°C in a single
831 cell suspension in the SN medium containing 10 μM Y-27632 (Merck, 688000). After 24 h,
832 the cells were embedded to the Matrigel and cultured in the SN medium containing 10 μM Y-
833 27632 for another 48 h to support single cell survival. The cultured cells were further sorted by
834 eGFP/TagRFP signals to enrich the infected cells.

835 For overexpressing NKX2.1 and/or TFAP2C, Tet-ON 3G doxycycline (Dox)-
836 inducible lentiviral vector (Takara, 631337) was modified by inserting EF1a-TagRFP-2A-
837 tet3G with tetON-NKX2-1 CDS, or by inserting EF1a-mNeonGreen-2A-tet3G with tetON-
838 TFAP2C CDS. For generating NKX2.1 variants, naturally occurring mutations in NKX2.1
839 binding domain region was selected from Leiden Open Variation Database 3.0¹⁴
840 (www.lovd.nl/3.0) and two previously reported clinical cases^{12,13} – 1 amino acid deletion¹⁴
841 (p.R162del), two nonsense point mutations¹⁴ (p.Q175* and p.R178*), and two missense point
842 mutations^{12,13} (p.I207F and p.I207M). NKX2.1 CDS harbouring each mutation was amplified
843 and inserted by Infusion (638909, Takara) cloning into the tetON-NKX2.1/EF1a-TagRFP-2A-
844 tet3G Dox-inducible lentiviral vector. NKX2.1 CDS lacking the entire DNA binding domain
845 was inserted into the EF1a-TagRFP-2A-tet3G Dox-inducible lentiviral vector by Infusion
846 cloning.

847 For the NKX2.1 knock-down experiment, a modified Dox-inducible CRISPRi vector was
848 gifted²³; N-terminal KRAB-dCas9 (a gift from Bruce Conklin, Addgene plasmid # 73498)
849 fused with a destabilising domain, dihydrofolate reductase (DHFR) sequence that is only
850 stabilised by trimethoprim (TMP) treatment, was sub-cloned into the EF1a-TagRFP-2A-tet3G
851 Dox-inducible lentiviral vector²³. Treatment of 2 µg/ml Dox (Merck, D9891) with 10 nmol/L
852 TMP (Merck, 92131) in the SN medium stabilize the functional KRAB-dCas9 protein. Three
853 gRNAs targeting NKX2.1²⁴ were individually subcloned into gRNA lentivirus as follows:
854 gRNA-1; 5'-GTCTGACGGCGGCAGAAGAG-3', gRNA-2; 5'-
855 GGACCAACAGTGCGGCCCA-3', gRNA-3; 5'-GAAATGAGCGAGCGAGTCTG-3'.
856 Single cells dissociated from organoids were infected and the infected cells were sorted by
857 TagRFP and/or mNeonGreen fluorescent signal using FACS (Fig. 5) after 48 h of infection.
858 The sorted TagRFP⁺ and/or mNeonGreen⁺ cells were cultured in the Matrigel in the absence
859 of Dox or TMP for 1 week. After the cells were grown into a typical organoid, the Dox and/or
860 TMP were added and culture continued for additional 2 weeks.

861

862 **Immunostaining of organoids and lung tissues**

863 For immunostaining of human lung tissue sections, the lungs were fixed in 4%
864 paraformaldehyde (PFA; Merck, 158127) overnight, washed in PBS and 15%, 20% and 30%
865 sucrose (w/v) in PBS before embedding in Optimum Cutting Temperature (OCT) medium
866 (Merck, F4680). 12 µm thick frozen sections were collected and permeabilised using 0.3%
867 Triton-X in PBS for 15 min. Antigen retrieval was performed by heating the slides in 10 mM
868 Na-Citrate buffer at pH 6.0 in a microwave for 5 min. Then slides were treated with blocking
869 solution containing 5% NDS, 1% Bovine Serum Albumin (BSA), 0.1% Triton-X in PBS at
870 room temperature for 1 h.

871 For whole-mount immunostaining of lung organoids, the Matrigel was completely
872 removed from the cultured organoids using Cell Recovery Solution (Corning, 354253) and
873 fixed in 4% PFA for 30 min on ice. After rinsing in PBS washing solution containing 0.2%
874 (v/v) Triton X-100 and 0.5% (w/v) BSA, the samples were transferred to a round-bottom 96
875 well plate and incubated in permeabilization/blocking solution containing 0.2% (v/v) Triton X-
876 100, 1% (w/v) BSA, and 5% normal donkey serum (NDS) in PBS, overnight at 4°C.

877 For primary antibody treatment, the following antibodies were treated to the organoids
878 and the tissue slices at 4°C overnight: proSFTPC (1:200; Merck, AB3786), E-cadherin (1: 500;

879 Thermo Fisher Scientific, 13-1900), NKX2.1 (1:200; Merck, 07-601), TFAP2C (1:200; Abcam,
880 ab218107), CD44 (1:200; Thermo Fisher Scientific, 17-0441-82), CD36 (1:200; Proteintech,
881 18836-1-AP), alpha-smooth muscle actin (1:500; Thermo Fisher Scientific, MA1-06110),
882 ACE2 (1:100; Abcam, ab108252), AXIN2 (1:200; R&D Systems, MAB6078), PDPN (1:200;
883 R&D Systems, AF3670), CD31 (1:200; Abcam, ab9498), PDGFRA (1:200; Cell Signaling
884 Technology, 3174), TP63 (1:200; Cell Signaling Technology, 13109), SOX2 (1: 500, Bio-
885 techne, AF2018), SOX9 (1: 500, Merck, AB5535), LAMP3 (1:100; Atlas Antibodies,
886 HPA051467), HTII-280 (1:200; Terracebiotech, TB-27AHT2-280), CD141 (1:100; PE-
887 conjugated; BioLegend, 344104), ZO-1 (1:200; Invitrogen, 40-2200) and KI67 (1:200; BD
888 Biosciences, 550609). After three washes with PBS, 97% (v/v) 2'-2'-thio-diethanol (TDE,
889 Sigma, 166782) was treated for clearing. Images were collected under Leica SP8 confocal
890 microscope.

891

892 ***In situ* hybridization chain reaction (*in situ* HCR)**

893 *In situ* HCR v3.0 was performed according to the manufacturer's procedure (Molecular
894 Instruments²⁵). Probes were designed according to the protocol and amplifiers with buffers
895 were purchased from Molecular Technologies. Sequence information of the probes for
896 detecting *SFTPC*, *WNT2*, *NOTUM*, *AXIN2*, *SOX9*, *FGFR4* and *TFAP2C* mRNA targets is
897 listed in Supplementary Table 3. Briefly, frozen human tissue sections were cut at 20 μ m from
898 lungs fixed overnight in 4% PFA in DEPC-treated PBS and processed to cryoblocks. Lung
899 sections were carefully rinsed in nuclease-free water, followed by 10 μ g/mL proteinase K
900 treatment (Thermo Fisher Scientific, AM2546), and 2 pmol of each probe was treated at 37°C
901 overnight. After washing, the tissue was incubated with 6 pmol of the amplifiers at room
902 temperature overnight for amplification. The amplifiers, consisting of a pair of hairpins
903 conjugated to fluorophores, Alexa 546, 647, or 488, were snap-cooled separately and added at
904 final 0.03 μ M to the tissue. After removing excess hairpins in 5X SSC (sodium chloride sodium
905 citrate) buffer containing 0.1% Triton X-100, nuclei were counter-stained with DAPI.

906 To combine *in situ* HCR with antibody immunostaining, the frozen human tissue sections
907 from 20 μ m up to 100 μ m thickness were permeabilised using 0.3% Triton-X in DEPC-treated
908 PBS for 20 min at room temperature. Then the tissues were treated with blocking solution
909 containing 5% NDS, 1% BSA, 0.1% Triton-X in DEPC-treated PBS at 4°C for 3 h. After
910 rinsing with cold DEPC-treated PBS, treated with a primary antibody against ACTA2 (1:500;
911 Thermo Fisher Scientific, MA1-06110) for 24 h, followed by a secondary antibody treatment
912 (1:500; Thermo Fisher Scientific, A10036) at 4°C overnight. After the tissue was washed three
913 times in the DEPC-treated PBS at room temperature, 2 pmol of each *in situ* HCR probes for
914 targeting *SFTPC* and *NOTUM* was treated at 37°C overnight without 10 μ g/mL proteinase K
915 treatment to preserve the antibody immunofluorescence. After washing, the tissues were
916 incubated with 6 pmol of the amplifiers at room temperature overnight. Following three times
917 of rinsing in in 5X SSC buffer containing 0.1% Triton X-100, nuclei were stained with DAPI.
918 Finally, the tissues were processed to 2'-2'-thio-diethanol (TDE, Sigma, 166782) for clearing
919 and mounting: 10 %, 25 %, 50 % (v/v) TDE in 1x DEPC-treated PBS for 1 hr and 97% TDE
920 overnight at 4°C. Images were collected under Leica SP8 confocal microscope.

921

922 **RNA extraction, cDNA synthesis, qRT-PCR analysis, and bulk RNA-seq**

923 Organoids were removed from the Matrigel and lysed. Total RNA was extracted according to
924 the RNeasy Mini Kit (Qiagen, 74004) protocol. For cells freshly purified from human lung
925 tissues were directly lysed using 100 µl lysis buffer from PicoPure™ RNA Isolation Kit
926 (Thermo Fisher Scientific, KIT0204). First Strand cDNA synthesis was performed using High-
927 Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4368814). Then, cDNA was
928 diluted 1:50 for qRT-PCR reaction (SYBR Green PCR Master Mix; Applied Biosystems,
929 4309155). Primer sequence information is listed in Supplementary Table 4. Data is presented
930 as fold change, calculated by ddCt method, using *ACTB* as housekeeping reference gene. For
931 bulk RNA-seq, RNA quality was validated on Agilent 2200 TapeStation. The RNA-seq
932 libraries were generated at the Cancer Research UK Cambridge Institute and sequenced on an
933 Illumina HiSeq 4000. A list of differentially expressed genes was extracted using the counted
934 reads and R package edgeR²⁶ version 3.16.5 for the 3 pairwise comparisons (Supplementary
935 Table 1). GO biological processes term enrichment, KEGG pathway, and gene set enrichment
936 analysis were performed using DAVID²⁷, Enrichr²⁸, and R package fgsea package²⁹,
937 respectively.

938

939 **Immunoblotting**

940 The organoid samples were harvested and lysed (RIPA buffer; Merck, R0278) after complete
941 removal of the Matrigel and run on 12.5 ~ 20 % SDS PAGE gels. Proteins on the gels were
942 transferred onto PVDF membrane using BioRad Mini Trans-Blot system (BioRad, Mini Trans-
943 Blot® Cell). The membranes were washed with pure water and blocked with 5% skimmed milk
944 in 0.1% Tween-20/PBS (PBST) for 30 min at room temperature. Membranes were incubated
945 with primary antibodies against NKX2.1 (1:200; Merck, 07-601), proSFTPC (1:1000; Merck,
946 AB3786), mature SFTPC (1:1000; Seven Hills Rioreagents, WRAB-76694), mature SFTPB
947 (1:1000; Seven Hills Rioreagents, WRAB-48604), and GAPDH (1:5000; Abcam, ab8245) in
948 the blocking buffer overnight at 4 °C. After washing with PBST, secondary antibodies
949 conjugated with fluorescence dyes (1:5000; anti-mouse IRDye® 800CW and anti-rabbit
950 IRDye® 680RD; Abcam, ab216774 and ab216779, respectively) were treated at room
951 temperature for 3 h. The membranes were washed in PBST and developed using Li-Cor
952 Odyssey imaging system.

953

954 **Chromatin immunoprecipitation**

955 Chromatin immunoprecipitation (ChIP) was performed according to SimpleChIP® Chromatin
956 immunoprecipitation protocol (Cell Signaling Technology, 9002). In brief, the organoids were
957 harvested and enzymatically dissociated into single cells using TrypLE Express Enzyme
958 (Thermo Fisher Scientific, 12605010). Then the cells were crosslinked with 1% formaldehyde
959 for 15 min at room temperature and the reaction was quenched by glycine at a final
960 concentration of 0.125 M. Chromatin was digested with 1 µl MNase (Cell Signaling
961 Technologies, 10011S) for 20 min at 37°C, followed by sonication for 12 cycles of 30 seconds
962 on and 30 seconds off using Biorupter (Diagenode, UCD-300), to length of an average size of
963 150-900 bp. 5 µg of digested chromatin samples was treated with antibodies against rabbit IgG
964 (1:100; Cell Signaling Technology, 2729) or NKX2.1 (1:100; Merck, 07-601). The amount of
965 immunoprecipitated DNA was quantified by qPCR using primers specific for promoter regions
966 of SFTPC, LAMP3, and SLC34A2. Fold enrichment values are presented as the fold-change

967 over the level of ChIP with negative control IgG antibody (ChIP signal/IgG signal). Sequence
968 information of the primers for targeting SFTPC, LAMP3, and SLC34A2 promoter regions is
969 listed in Supplementary Table 4.

970

971 **Bulk ATAC-seq**

972 Genome-wide chromatin accessibility of lung organoids was assessed as previously
973 described³⁰. In brief, 50,000 cells were harvested from organoids and lysed in lysis buffer (10
974 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) IGEPAL CA-630). The lysate
975 was treated in 50 µL reactions with Nextera TDE1 transposase (Illumina, 15027865) for 30
976 min at 37°C. The purified DNA was amplified and indexed using Nexera DNA CD Indexes
977 (Illumina, 20018707), and size distribution of the DNA libraries was analysed using High-
978 sensitivity Qubit dsDNA Assay Kit (ThermoFisher, Q32851) and Agilent 2200 TapeStation.
979 The libraries were sequenced on an Illumina HiSeq 4000. Peak calling was done using MACS2
980 algorithm³¹ (version 2.1.1) and further processed to extract differential peaks (Supplementary
981 Table 2). Then, the differential peak data was further used for analysing transcription factor
982 motifs using HOMER³² software in combined with RNA-seq data.

983

984 **Electron microscopy imaging**

985 The organoid samples were fixed in 2 % formaldehyde/2 % glutaraldehyde in 0.05 M sodium
986 cacodylate buffer (NaCAC), pH 7.4, containing 2 mM calcium chloride (Merck, C27902)
987 overnight at 4°C. After washing in 0.05 M NaCAC at pH 7.4, the samples were osmicated for
988 3 days at 4°C. After washing in deionised water (DIW), the samples were treated twice with
989 0.1 % (w/v) thiocarbohydrazide (Merck, 223220) in DIW for each 20 min and 1 h at room
990 temperature in the dark, followed by block-staining with uranyl acetate (2 % uranyl acetate in
991 0.05 M maleate buffer pH 5.5) for 3 days at 4°C. Then, the samples were dehydrated in a graded
992 series of ethanol (50%/70%/95%/100%/100% dry) 100% dry acetone and 100% dry
993 acetonitrile, three times in each for at least 5 min. Next, the samples were infiltrated with a
994 50:50 mixture of 100% dry acetonitrile/Quetol resin (TAAB, Q005) without BDMA (TAAB,
995 B008) overnight, followed by 3 days in 100% Quetol without BDMA. The sample was
996 infiltrated for 5 days in 100% Quetol resin with BDMA, exchanging the resin each day. The
997 Quetol resin mixture is: 12 g Quetol 651, 15.7 g NSA (TAAB, N020), 5.7 g MNA (TAAB,
998 M012) and 0.5 g BDMA. Samples were placed in embedding moulds and cured at 60°C for 3
999 days.

1000 Thin sections were cut using an Ultracut E ultramicrotome (Leica) and mounted on
1001 melinex plastic coverslips. The coverslips were mounted on aluminium SEM stubs using
1002 conductive carbon tabs and the edges of the slides were painted with conductive silver paint.
1003 Then, the samples were sputter coated with 30 nm carbon using a Quorum Q150 T E carbon
1004 coater and imaged in a Verios 460 scanning electron microscope (FEI, Thermo Fisher
1005 Scientific) at 4 keV accelerating voltage and 0.2 nA probe current in backscatter mode using
1006 the concentric backscatter detector in immersion mode at a working distance of 3.5-4 mm;
1007 1,536 x 1,024 pixel resolution, 3 µs dwell time, 4 line integrations. Stitched maps were acquired
1008 using FEI MAPS software using the default stitching profile and 10% image overlap.

1009

1010 **Data availability**

1011 All bulk RNA-seq and ATAC-seq data generated have been deposited in NCBI's Gene
1012 Expression Omnibus and are accessible through GEO Series accession number GSE178529.

1013

1014 **Statistical analysis**

1015 Data are expressed as average \pm standard deviation (SD). Statistical significance was evaluated
1016 by unpaired student's *t* test, 1- or 2-way ANOVA with Tukey/Bonferroni/ Dunnett comparison
1017 multiple comparison post-test; ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and
1018 **** $P < 0.0001$.

1019

1020

1021 **REFERENCES**

- 1022 1. Burri, P. H. Fetal and postnatal development of the lung. *Annu. Rev. Physiol.* **46**, 617–628
1023 (1984).
- 1024 2. Nikolić, M. Z., Sun, D. & Rawlins, E. L. Human lung development: recent progress and
1025 new challenges. *Development* **145**, (2018).
- 1026 3. Nikolić, M. Z. *et al.* Human embryonic lung epithelial tips are multipotent progenitors
1027 that can be expanded in vitro as long-term self-renewing organoids. *Elife* **6**, (2017).
- 1028 4. Miller, A. J. *et al.* In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor
1029 Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports* **10**, 101–119 (2018).
- 1030 5. McCauley, K. B. *et al.* Efficient Derivation of Functional Human Airway Epithelium from
1031 Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell* **20**, 844-
1032 857.e6 (2017).
- 1033 6. Little, D. R. *et al.* Differential chromatin binding of the lung lineage transcription factor
1034 NKX2-1 resolves opposing murine alveolar cell fates in vivo. *Nat. Commun.* **12**, 2509
1035 (2021).
- 1036 7. Guillot, L. *et al.* NKX2-1 mutations leading to surfactant protein promoter dysregulation
1037 cause interstitial lung disease in “Brain-Lung-Thyroid Syndrome.” *Hum. Mutat.* **31**,
1038 E1146-62 (2010).
- 1039 8. Attarian, S. J. *et al.* Mutations in the thyroid transcription factor gene NKX2-1 result in
1040 decreased expression of SFTPB and SFTPC. *Pediatr. Res.* **84**, 419–425 (2018).
- 1041 9. Jacob, A. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Lung
1042 Alveolar Epithelial Cells. *Cell Stem Cell* **21**, 472-488.e10 (2017).
- 1043 10. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA
1044 sequencing. *Nature* **587**, 619–625 (2020).
- 1045 11. Kakugawa, S. *et al.* Notum deacylates Wnt proteins to suppress signalling activity. *Nature*
1046 **519**, 187–192 (2015).
- 1047 12. Gillett, E. S., Deutsch, G. H., Bamshad, M. J., McAdams, R. M. & Mann, P. C. Novel
1048 NKX2.1 mutation associated with hypothyroidism and lethal respiratory failure in a full-
1049 term neonate. *J. Perinatol.* **33**, 157–160 (2013).
- 1050 13. Maquet, E. *et al.* Lethal respiratory failure and mild primary hypothyroidism in a term girl
1051 with a de novo heterozygous mutation in the TITF1/NKX2.1 gene. *J. Clin. Endocrinol.*
1052 *Metab.* **94**, 197–203 (2009).
- 1053 14. Fokkema, I. F. A. C. *et al.* LOVD v.2.0: the next generation in gene variant databases.
1054 *Hum. Mutat.* **32**, 557–563 (2011).
- 1055 15. Ochs, M. The closer we look the more we see? Quantitative microscopic analysis of the
1056 pulmonary surfactant system. *Cell. Physiol. Biochem.* **25**, 27–40 (2010).
- 1057 16. Weaver, T. E., Na, C.-L. & Stahlman, M. Biogenesis of lamellar bodies, lysosome-related
1058 organelles involved in storage and secretion of pulmonary surfactant. *Semin. Cell Dev.*
1059 *Biol.* **13**, 263–270 (2002).
- 1060 17. Laresgoiti, U. *et al.* Lung epithelial tip progenitors integrate glucocorticoid- and STAT3-
1061 mediated signals to control progeny fate. *Development* **143**, 3686–3699 (2016).
- 1062 18. Shannon, J. M., Nielsen, L. D., Gebb, S. A. & Randell, S. H. Mesenchyme specifies
1063 epithelial differentiation in reciprocal recombinants of embryonic lung and trachea. *Dev.*
1064 *Dyn.* **212**, 482–494 (1998).

- 1065 19. Frank, D. B. *et al.* Emergence of a Wave of Wnt Signaling that Regulates Lung
1066 Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. *Cell Rep.* **17**,
1067 2312–2325 (2016).
- 1068 20. Zepp, J. A. *et al.* Genomic, epigenomic, and biophysical cues controlling the emergence
1069 of the lung alveolus. *Science* **371**, (2021).
- 1070 21. Mollaoglu, G. *et al.* The Lineage-Defining Transcription Factors SOX2 and NKX2-1
1071 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment.
1072 *Immunity* **49**, 764-779.e9 (2018).
- 1073 22. Li, L. *et al.* TFAP2C- and p63-Dependent Networks Sequentially Rearrange Chromatin
1074 Landscapes to Drive Human Epidermal Lineage Commitment. *Cell Stem Cell* vol. 24 271-
1075 284.e8 (2019).
- 1076 23. Sun, D., Evans, L., Lim, K. & Rawlins, E. L. A functional genetic toolbox for human
1077 tissue-derived organoids. *bioRxiv* 2020.05.04.076067 (2021)
1078 doi:10.1101/2020.05.04.076067.
- 1079 24. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-
1080 mediated gene repression and activation. *Elife* **5**, (2016).
- 1081 25. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed,
1082 quantitative, sensitive, versatile, robust. *Development* **145**, (2018).
- 1083 26. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1084 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–
1085 140 (2010).
- 1086 27. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of
1087 large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
- 1088 28. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).
- 1089 29. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021)
1090 doi:10.1101/060012.
- 1091 30. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for
1092 assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-
1093 21.29.9 (2015).
- 1094 31. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using
1095 MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
- 1096 32. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1097 cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–
1098 589 (2010).

Fig. 1. Human fetal lung tip progenitor cells acquire alveolar features during normal development.

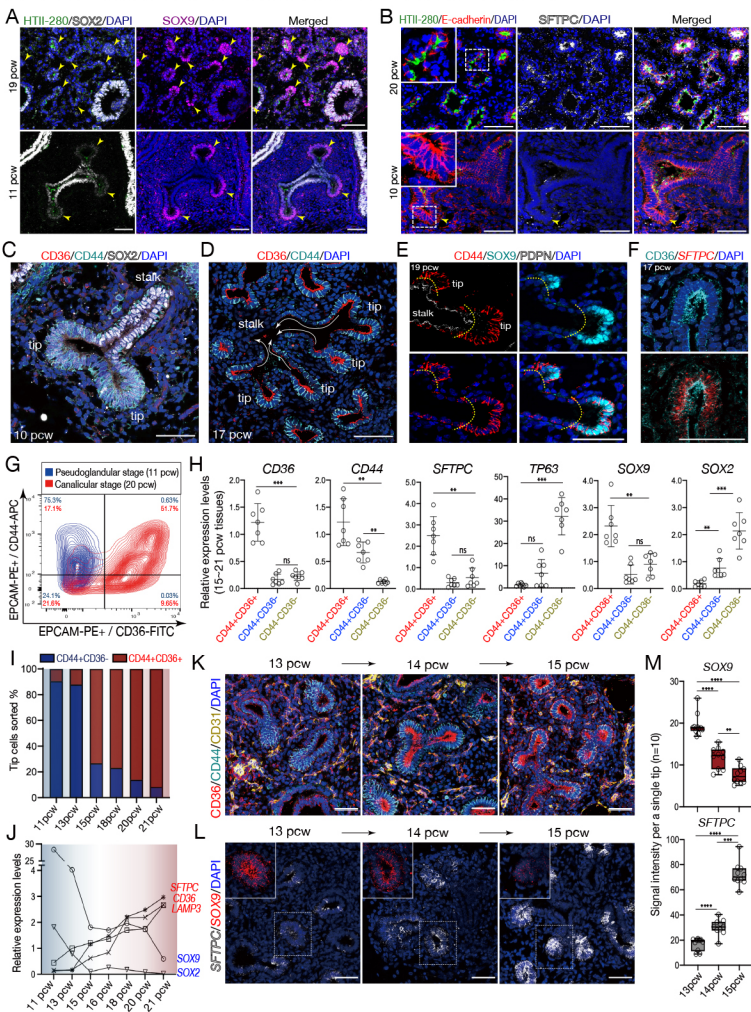


Fig. 2. CD36, CD44 dual-positive tip cells self-renew and undergo lineage commitment *in vitro* to form lineal stage lung organoids

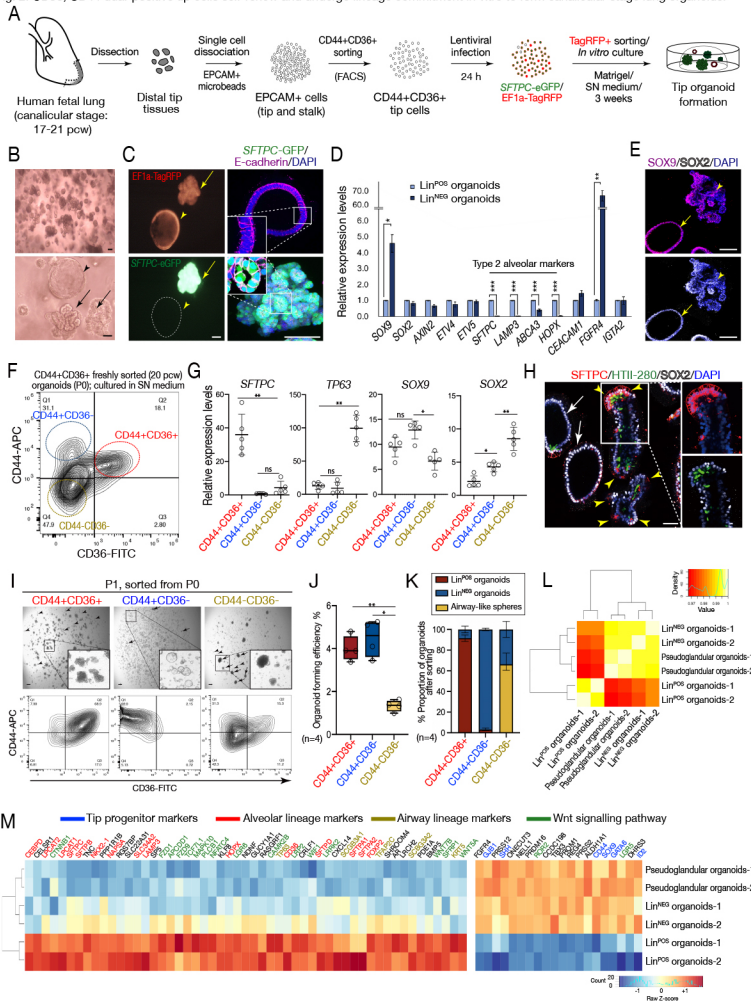


Fig. 3. Wnt and FGF signalling coordinate human tip cell maintenance and differentiation in vitro

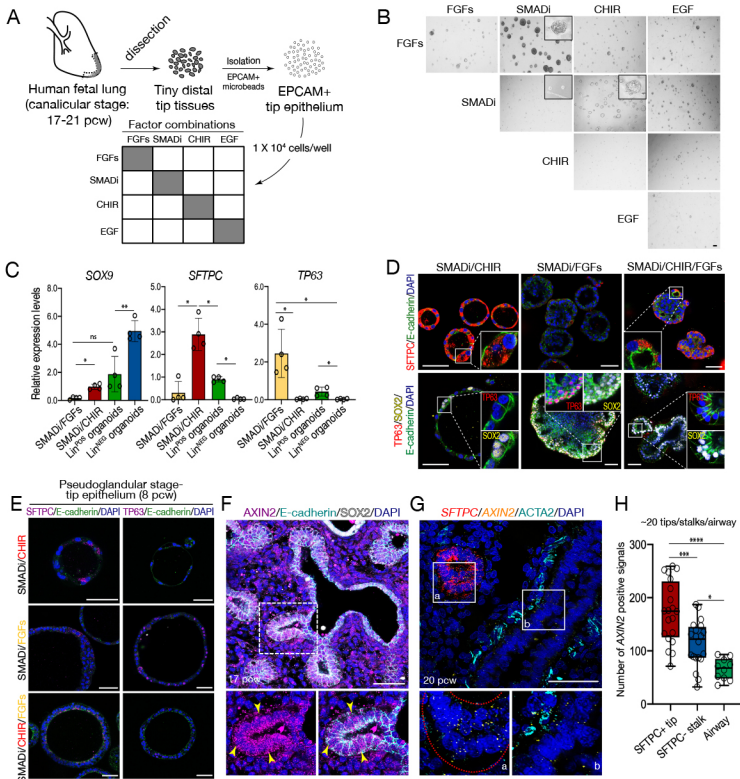


Fig. 4. Spatial patterning of the differentiating epithelium by NOTUM+ myofibroblasts

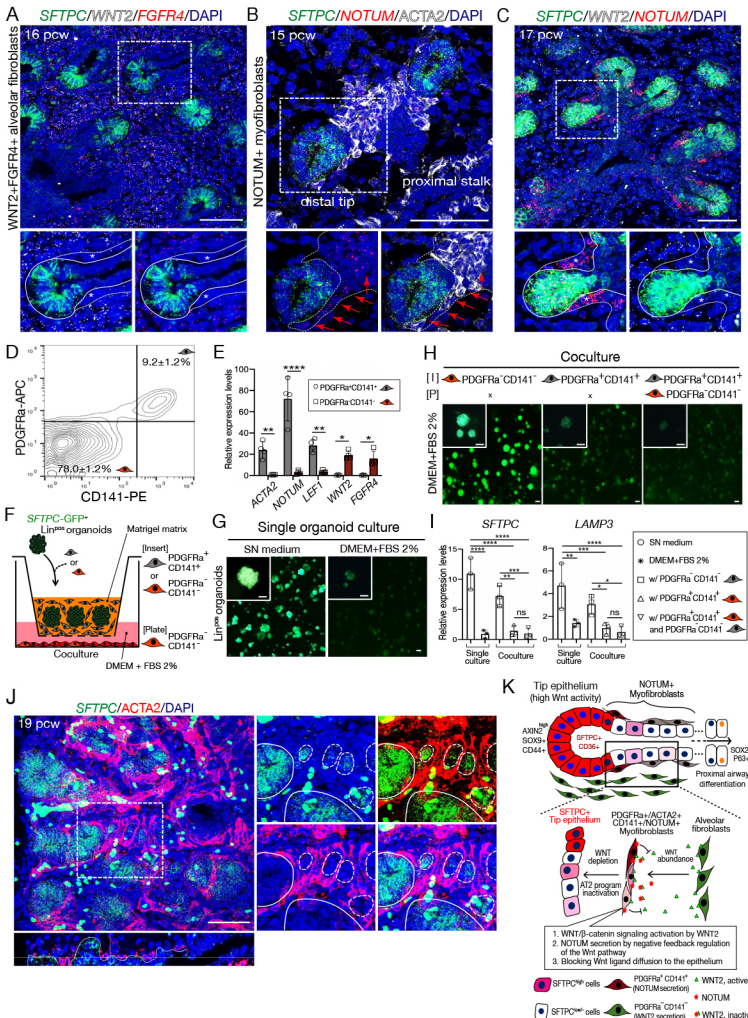


Fig. 5. Identification of key transcription factors controlling airway and alveolar lineage differentiation using the organoid system.

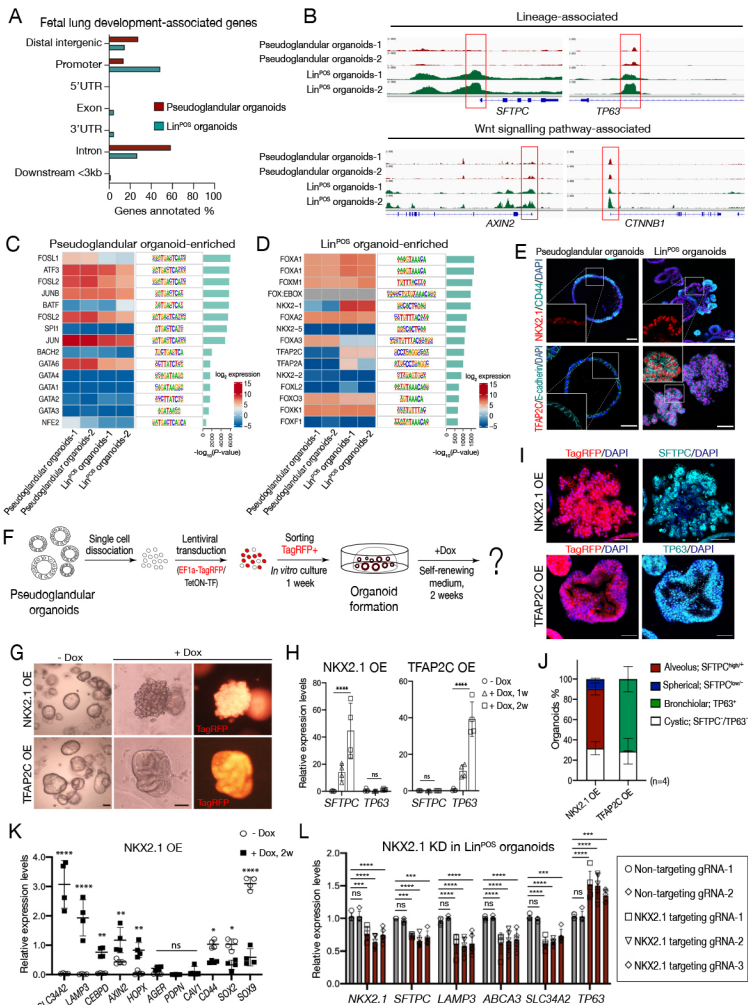


Fig. 6. Analysis of naturally occurring human genetic variation using organoid assays.

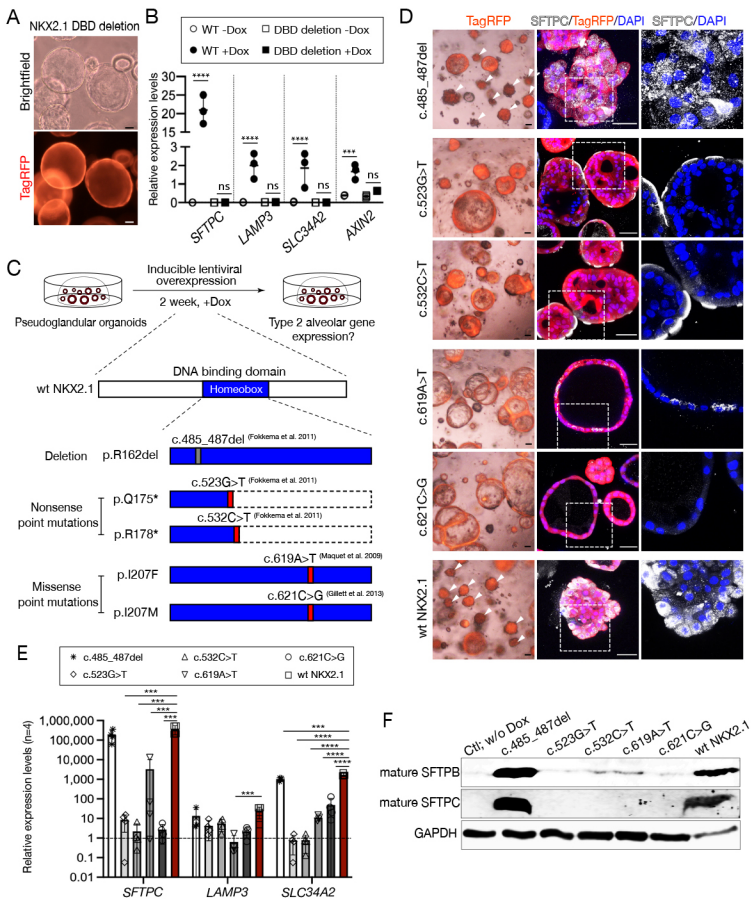
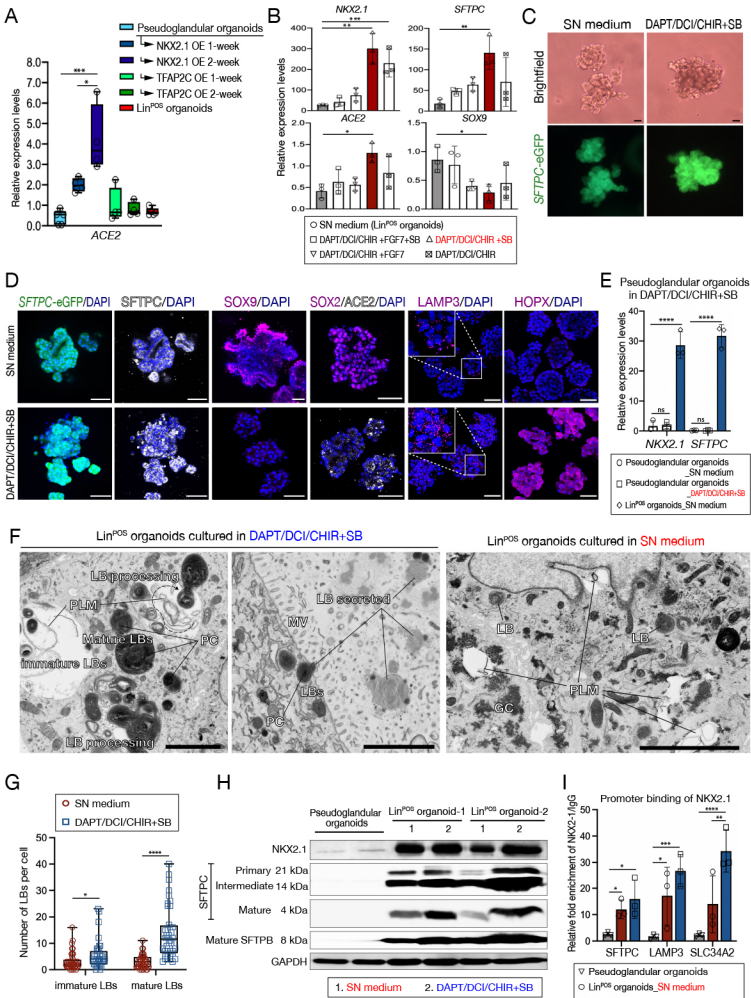
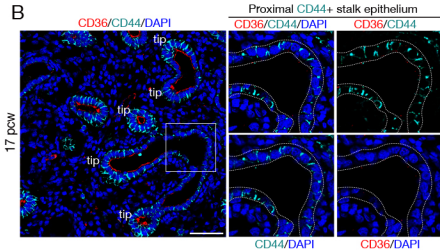
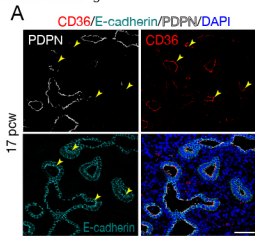
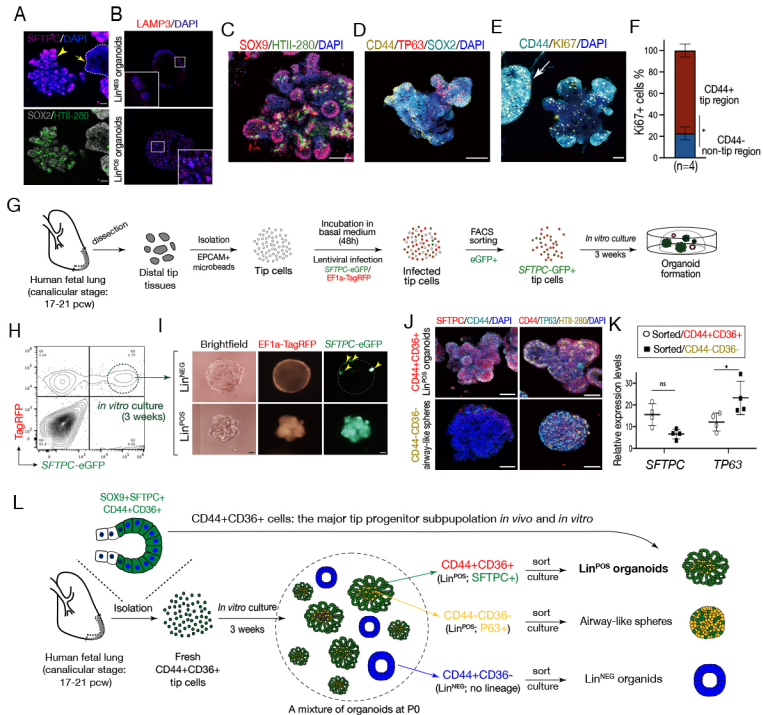


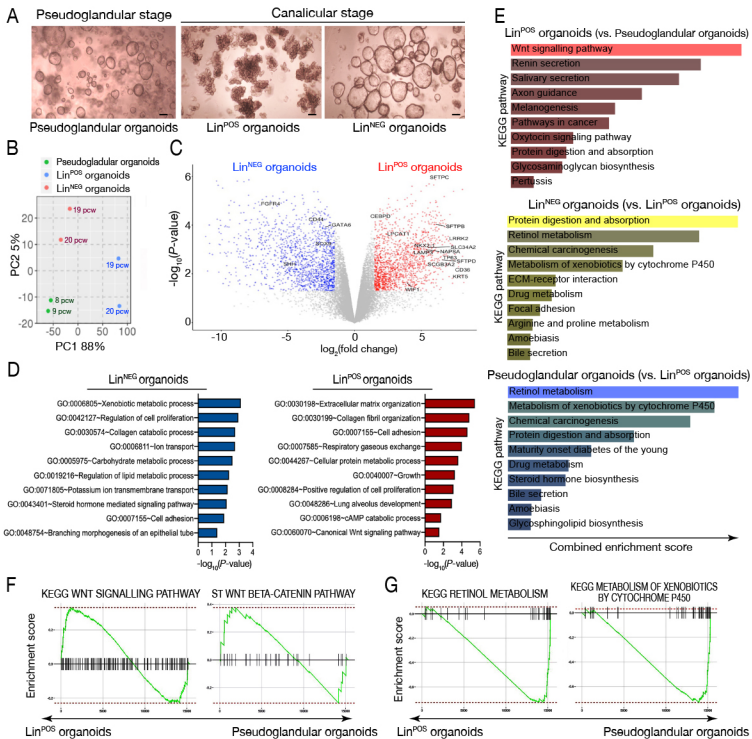
Fig.7. Efficient *in vitro* differentiation of Lin^{POS} organoids to alveolar type 2 cell fate

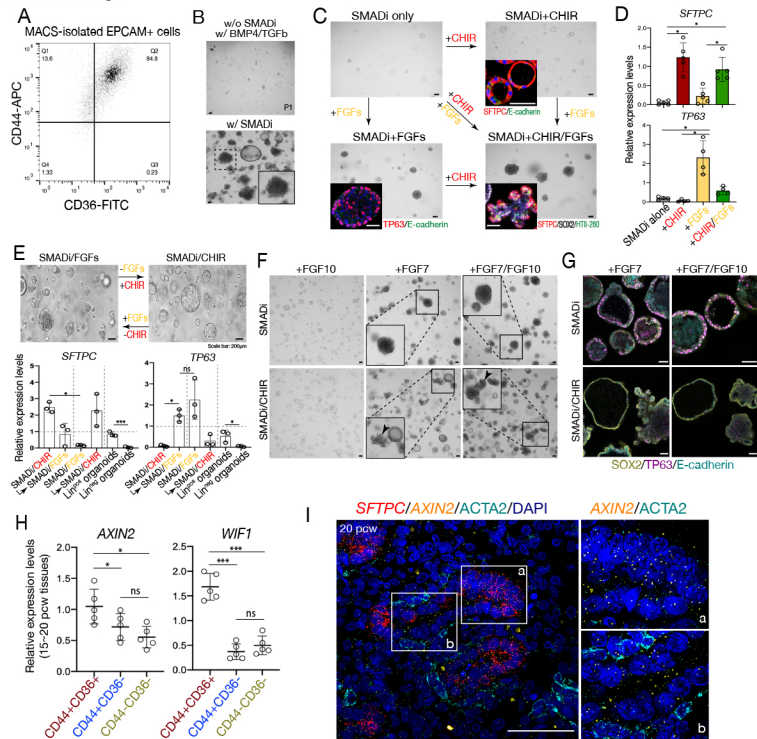


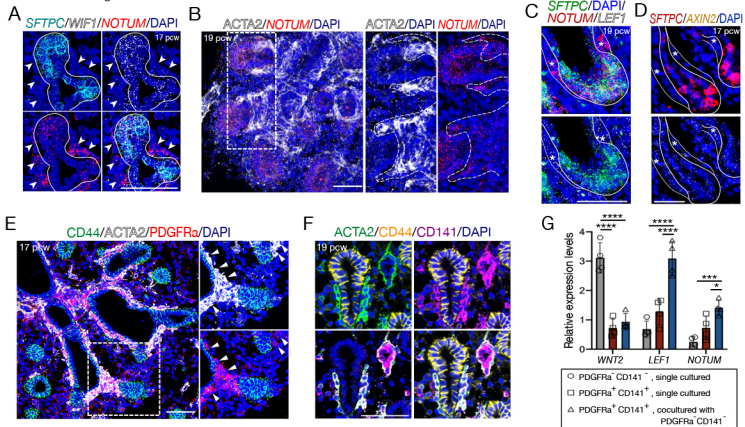
Extended Data Fig. 1

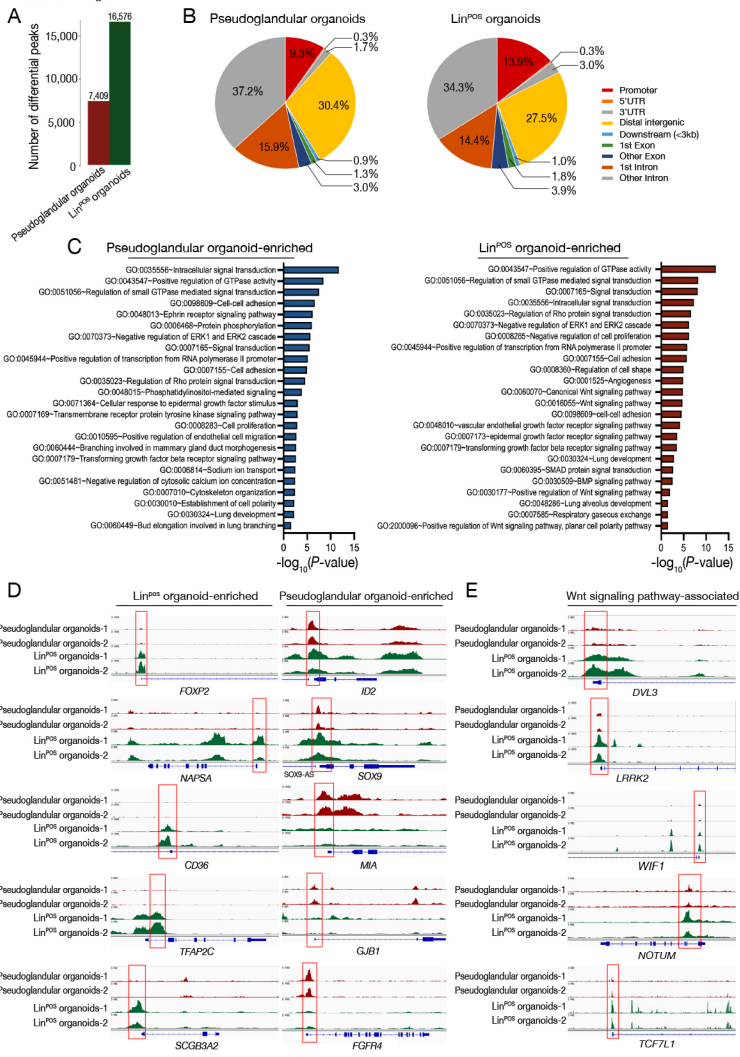


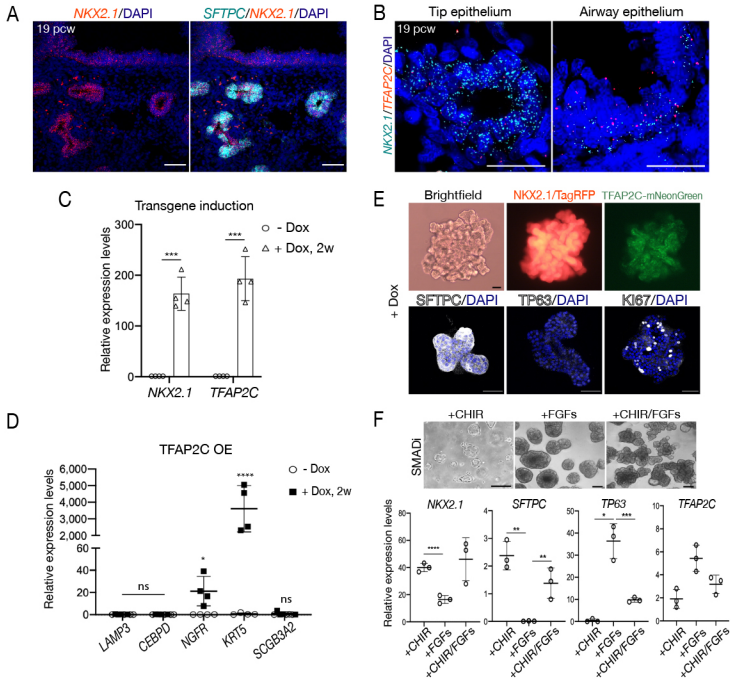






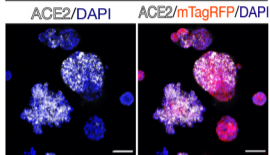




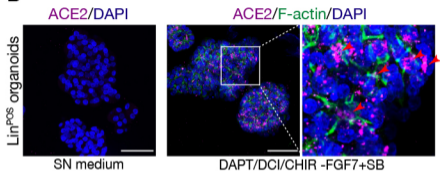


Extended Data Fig. 8

A Pseudoglandular organoids NKX2.1 OE 2-week



B



C

