



# Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5<sup>+</sup> basal cells

Jaymin J. Kathiriya<sup>1,5</sup>, Chaoqun Wang<sup>1,5</sup>✉, Minqi Zhou<sup>1</sup>, Alexis Brumwell<sup>1</sup>, Monica Cassandras<sup>1</sup>, Claude Jourdan Le Saux<sup>1</sup>, Max Cohen<sup>1</sup>, Kostantinos-Dionysios Alysandratos<sup>2,3</sup>, Bruce Wang<sup>1</sup>, Paul Wolters<sup>1</sup>, Michael Matthay<sup>1</sup>, Darrell N. Kotton<sup>2,3</sup>, Harold A. Chapman<sup>1,4</sup>✉ and Tien Peng<sup>1,4</sup>

**Loss of alveolar type 2 cells (AEC2s) and the ectopic appearance of basal cells in the alveoli characterize severe lung injuries such as idiopathic pulmonary fibrosis (IPF). Here we demonstrate that human alveolar type 2 cells (hAEC2s), unlike murine AEC2s, transdifferentiate into basal cells in response to fibrotic signalling in the lung mesenchyme, in vitro and in vivo. Single-cell analysis of normal hAEC2s and mesenchymal cells in organoid co-cultures revealed the emergence of pathologic fibroblasts and basaloid cells previously described in IPF. Transforming growth factor- $\beta$ 1 and anti-bone morphogenic protein signalling in the organoids promoted transdifferentiation. Trajectory and histologic analyses of both hAEC2-derived organoids and IPF epithelium indicated that hAEC2s transdifferentiate into basal cells through alveolar-basal intermediates that accumulate in proximity to pathologic *CTHRC1*<sup>hi</sup>/*TGFB1*<sup>hi</sup> fibroblasts. Our study indicates that hAEC2 loss and expansion of alveolar metaplastic basal cells in severe human lung injuries are causally connected through an hAEC2-basal cell lineage trajectory driven by aberrant mesenchyme.**

Previous studies in mice have shown that mouse alveolar type 2 cells (mAEC2s) are the resident stem cell population in the alveoli that constitute the entire gas exchange surface of the lung<sup>1,2</sup>. In idiopathic pulmonary fibrosis (IPF), the most deadly and prevalent form of diffuse parenchymal lung disease, human alveolar type 2 cells (hAEC2s) are lost from the alveoli, concurrent with the appearance of metaplastic alveolar KRT5<sup>+</sup> basal cells, which normally appear in the conducting airways<sup>3–9</sup>. Rigorous genetic lineage tracing has shown that metaplastic KRT5<sup>+</sup> cells in the murine alveoli are not derived from mAEC2s, but rather from KRT5<sup>+</sup>/SOX2<sup>+</sup> progenitors in the mouse airway after severe alveolar injury from fibrosis or viral infections<sup>5,6,10–12</sup>. However, it is not clear whether a similar population in the human airway exists that contributes to metaplastic basal cells, as the airways contain key anatomic differences across the two species<sup>13</sup>. This is a clinically relevant question, because the extent of alveolar KRT5<sup>+</sup> basal cells directly correlates with mortality in IPF<sup>14</sup>. In this study, we made a surprising finding that hAEC2s, but not mAEC2s, can readily transdifferentiate into KRT5<sup>+</sup> basal cells in organoid culture and xenotransplant. Moreover, we define pro-fibrotic mesenchymal niche-derived factors that promote hAEC2-to-basal cell transdifferentiation. Finally, quantitative spatial analysis of IPF lung tissue reveals that basal cells and advanced alveolar-basal intermediates are surrounded by aberrant, *CTHRC1*<sup>hi</sup> pro-fibrotic mesenchyme. These results identify hAEC2s as a source of metaplastic KRT5<sup>+</sup> basal cells in severe alveolar injuries and provide a potential explanation for the reported appearance of aberrant hAEC2s with basaloid features in the transcriptomes of IPF and other severe lung injuries such as COVID pneumonia<sup>8,9</sup>.

## Results

**hAEC2s transdifferentiate into basal cells in vitro.** Previously reported hAEC2 organoids have utilized MRC5, a fetal human lung fibroblast cell line, as feeders to maintain primary adult hAEC2s in vitro, at least in part through local bone morphogenic protein (BMP) and Wnt signalling<sup>1,15</sup>. We confirmed similar organoid development using low-passage MRC5 cells in our three-dimensional (3D) organoid system and hAEC2s isolated with the HTII-280 antibody (EpCAM<sup>+</sup>/HTII-280<sup>+</sup>; Fig. 1b,c and Extended Data Fig. 1a), with high purity of the hAEC2s confirmed by single-cell RNA-sequencing (scRNA-seq) and cytospin analysis (Fig. 1a and Extended Data Fig. 1b,c). We hypothesized that primary mesenchyme obtained from normal adult lungs would provide a more supportive niche. We used a flow cytometry strategy to isolate adult human lung mesenchyme (AHLM; CD45<sup>-</sup>/CD11b<sup>-</sup>/CD31<sup>-</sup>/EpCAM<sup>-</sup>) that was previously described by scRNA-seq to yield a predominantly PDGFR $\alpha$ <sup>+</sup> fibroblast population<sup>16</sup> (Extended Data Fig. 1d), and co-cultured AHLM with hAEC2s in a 3D organoid. Although there was no difference in colony-forming capacity between MRC5 cells and AHLM, we saw a dramatic loss of Surfactant Protein-C (SFTPC) accompanied by the gradual appearance of KRT5, an airway basal cell marker, in hAEC2 organoids co-cultured with AHLM (Fig. 1d,e and Extended Data Fig. 1e). By day 14 of co-culture, the majority of the organoids derived from hAEC2–AHLM co-culture contained KRT5<sup>+</sup> cells, whereas the majority of organoids in the hAEC2–MRC5 co-culture contained only SFTPC<sup>+</sup> cells (Fig. 1c,e). Further immunophenotyping of KRT5<sup>+</sup> organoids from hAEC2s co-cultured with either AHLM or MRC5 cells showed that most KRT5<sup>+</sup> cells arising from

<sup>1</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>2</sup>Center for Regenerative Medicine, Boston University School of Medicine, Boston, MA, USA. <sup>3</sup>Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA, USA. <sup>4</sup>Cardiovascular Institute and Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, University of California, San Francisco, San Francisco, CA, USA.

<sup>5</sup>These authors contributed equally: Jaymin J. Kathiriya, Chaoqun Wang. ✉e-mail: [chaoqun.wang@ucsf.edu](mailto:chaoqun.wang@ucsf.edu); [hal.chapman@ucsf.edu](mailto:hal.chapman@ucsf.edu)

MRC5 co-culture do not stain for other mature basal cell markers such as TP63 or KRT17 (Extended Data Fig. 1f,h), whereas the majority of KRT5<sup>+</sup> cells derived from AEC2/AHLM organoids are also TP63<sup>+</sup> and KRT17 (Extended Data Fig. 1g,h). Some of the organoids assume apical-basal polarity with KRT5<sup>+</sup> basal cells in the outer layer and KRT5<sup>-</sup>/KRT8<sup>+</sup> cells in the inner layer (Extended Data Fig. 1i), which have been identified as a transitional cell type derived from AEC2s<sup>17–20</sup>. To further ensure that the basal cells did not result from an outgrowth of contaminants in the HTII-280 sort, we cultured hAEC2s in previously defined mesenchymal-free, alveolar growth medium<sup>21</sup> over a month to enrich for hAEC2s (cytospin shows >99% purity) before co-culture with AHLM, which resulted in the same basal cell differentiation as seen with freshly isolated hAEC2s (Extended Data Fig. 1k). Comparison of hAEC2 organoids utilizing passage 0 (freshly sorted) versus passage 2 from the same AHLM donor did not show any significant difference in KRT5 transdifferentiation (Extended Data Fig. 1j). A parallel experiment using mAEC2s co-cultured with primary adult murine lung mesenchyme (AMLM) under identical in vitro conditions produced no KRT5<sup>+</sup> organoids (Extended Data Fig. 2a,b). Conversely, culturing hAEC2s with AMLM failed to generate organoids (Extended Data Fig. 2c).

To further study the capacity of hAEC2s to differentiate into KRT5<sup>+</sup> basal cells using a trackable differentiation system, we utilized a previously validated human induced pluripotent stem cell-derived AEC2 (iAEC2) model system<sup>22</sup>. We generated stably self-renewing iAEC2s in epithelial-only (feeder-free) cultures from the published SPC2 induced pluripotent stem cell (iPSC) line that has been previously engineered with a tdTomato reporter targeted to one allele of the endogenous *SFTPC* locus to isolate iAEC2s lacking basal cell transcripts (SPC2-ST-B2 clone, referred to as SPC2 cell line)<sup>23</sup>. As previously shown, we confirmed the purity of the iAEC2s, judged by the absence of basal cell markers and the maintenance of robust tdTomato expression (Fig. 1f). Next, we co-cultured iAEC2s with either MRC5 cells or AHLM in our previously published medium, optimized to maintain iAEC2 fate in culture<sup>22,23</sup> (CHIR99021 (CHIR), Keratinocyte growth factor (KGF), dexamethasone, cyclic adenosine monophosphate (cAMP), Isobutylmethylxanthin (IBMX), hereafter called CK/DCI). In stark contrast to iAEC2s co-cultured with MRC5 cells, iAEC2s co-cultured with AHLM dramatically lost *SFTPC* expression, as evidenced by the loss of tdTomato fluorescence (Fig. 1g). Immunophenotyping confirmed the loss of *SFTPC* in iAEC2s co-cultured with AHLM in virtually all organoids, along with the emergence of intermediate and early basal cell markers KRT8 and KRT17, respectively, in ~65% of organoids (Fig. 1i,j). Interestingly, iAEC2s co-cultured with AHLM do not completely transdifferentiate into KRT5<sup>+</sup> basal cells in CK/DCI. Further experiments revealed CHIR as the main medium component that prevents basal cell differentiation (Extended Data Fig. 2d,e) and, when CHIR was reduced by dilution, there was

complete differentiation into basal cells that are also positive for KRT5, KRT14 and p63, while retaining NKX2-1, confirming lung fate (Fig. 1k,l and Extended Data Fig. 2f).

### hAEC2s transdifferentiate to basal cells in a fibrotic host in vivo.

A previous transplant study of freshly sorted murine AEC2s showed no evidence of basal cell transdifferentiation when engrafted in injured lungs<sup>24</sup>. To determine whether the in vitro transdifferentiation capacity of hAEC2s is maintained in vivo, we transplanted freshly sorted hAEC2s into bleomycin-induced fibrotic lungs of NOD *scid* gamma (NSG) mice on day 10 (Fig. 2a). hAEC2s were transplanted either alone or with MRC5 or AHLM cells, analogous to ex vivo experiments with 3D organoids. Using the human-specific nuclear antigen (HNA), we were able to identify numerous patches of human cells predominantly in damaged alveolar regions 10 days after transplant (day 20 post bleomycin injury). Co-staining with KRT5 identified HNA<sup>+</sup>/KRT5<sup>+</sup> cells that were SFTPC<sup>-</sup>, where the metaplastic basal cells appeared to form 'bronchiolized' cysts in the murine host (Fig. 2c). Utilizing an antibody specific to human-pro-SP-C, we also noted the presence of HNA<sup>+</sup>/pro-SP-C<sup>+</sup> hAEC2 patches in the lung that were spatially distinct from dysplastic basal patches (Fig. 2b,c). Immunohistology analysis showed that, when transplanted alone or with MRC5 cells, the majority of the hAEC2s retained alveolar fate (~75%), whereas co-transplants of a similar AHLM:AEC2 ratio resulted in a majority of the patches comprising KRT5<sup>+</sup> basal cells (~80%) (Fig. 2b,c). To determine whether transplanted AHLM and hAEC2s are associated in fibrotic lung, we labelled AHLM with a lentivirus expressing red fluorescent protein (RFP) followed by co-transplantation of hAEC2s and RFP-labelled AHLM. Interestingly, we found pods of hAEC2s with basaloid feature (SFTPC<sup>+</sup>/KRT5<sup>+</sup>/HNA<sup>+</sup>, Fig. 2d, green arrows) along with basal cells (SFTPC<sup>-</sup>/KRT5<sup>+</sup>/HNA<sup>+</sup>, green arrows) in regions adjacent to engrafted RFP<sup>+</sup> AHLM (Fig. 2d, orange arrows). We also detected clusters of cells co-expressing AEC2s (SFTPC) and basal (KRT5) and transitioning (KRT8)<sup>17,19</sup> cell markers, adjacent to AHLM (Fig. 2e). The xenotransplantation experiments demonstrate that hAEC2s are capable of reconstituting a fibrotic milieu in vivo and transdifferentiate into basal cells with high efficiency in the presence of AHLM, confirming the in vivo plasticity of hAEC2s.

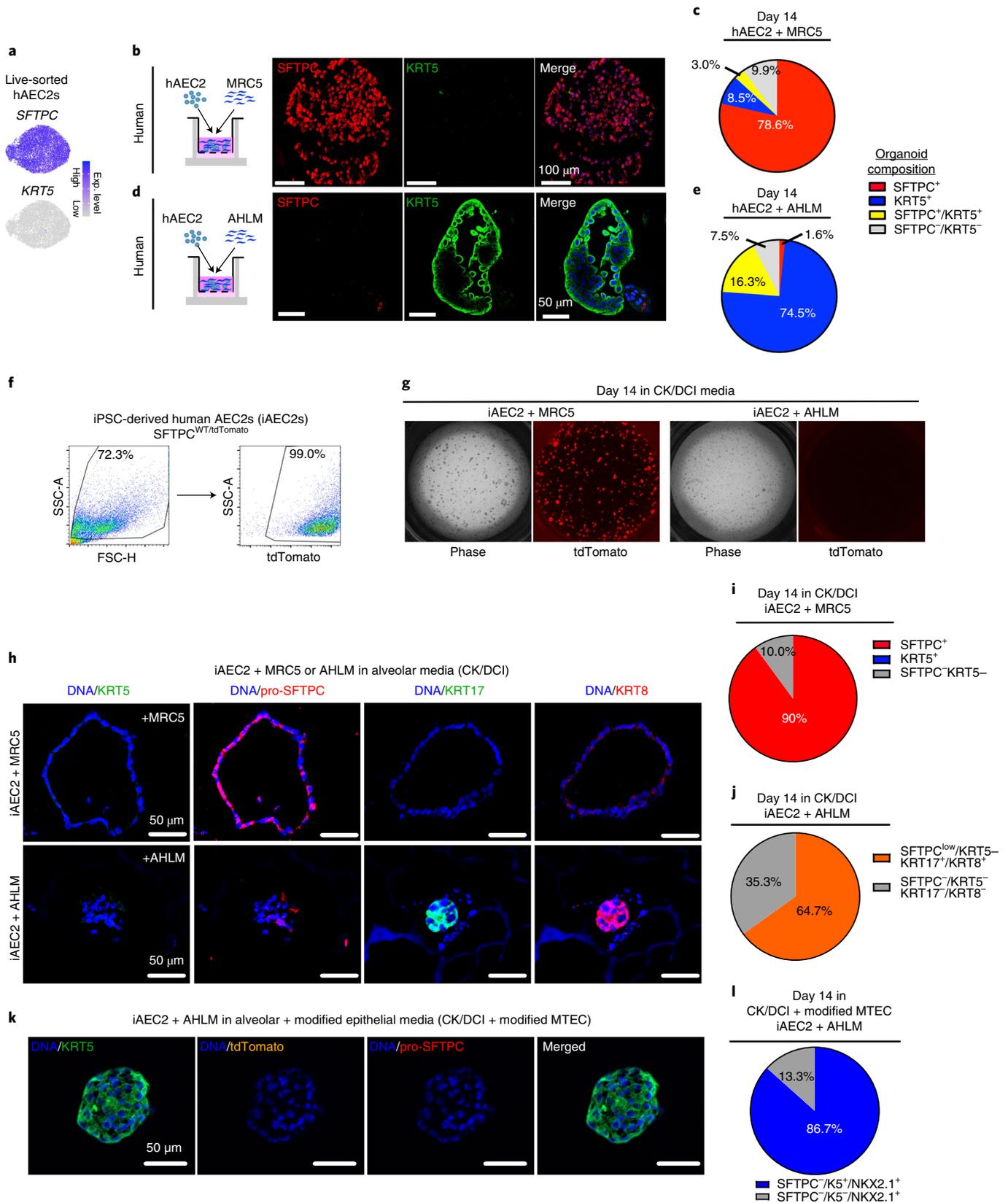
### IPF-like mesenchymal subsets emerge during AHLM culture.

hAEC2s co-cultured with IPF mesenchyme accelerated basal cell transdifferentiation compared to passage-matched AHLM from normal donors (Extended Data Fig. 3a,b). To determine whether AHLM acquires features of IPF mesenchymal subsets with culture, we analysed the scRNA-seq results for fresh (sorted from donor lung, uncultured) AHLM (normal donors, *N*=3), fresh IPF mesenchyme (*N*=3) and cultured AHLM at passage 0–1 isolated from organoids (*N*=2). IPF lungs were collected from patients undergoing lung transplantation, and AHLM was collected from cadaveric

**Fig. 1 | Primary AHLM drives hAEC2 transdifferentiation into KRT5<sup>+</sup> basal cells in vitro.** **a**, Live/EpCAM<sup>+</sup>/HTII-280<sup>+</sup> cells (hAEC2s) were sorted and sequenced, confirming a purified SFTPC<sup>+</sup> population. **b,c**, Immunofluorescence of hAEC2-derived organoids co-cultured with MRC5 cells (**b**) with quantification of epithelial lineages (**c**) found in the organoids (SFTPC<sup>+</sup> organoids, organoids with both SFTPC<sup>+</sup> cells and KRT5<sup>+</sup> cells, or KRT5<sup>+</sup> organoids). Representative of *n*=3 biologically independent replicates. Data are presented as the mean of two replicates. **d,e**, Immunofluorescence of hAEC2-derived organoids co-cultured with AHLM (**d**) with quantification of the lineages of the organoids (**e**). Data are presented as the mean of *n*=2 biological samples. Representative of *n*=2 independent experiments. **f**, iPSC-derived AEC2s express tdTomato from one allele of *SFTPC* and can be flow purified based on their tdTomato expression. **g**, iAEC2s were co-cultured with either MRC5 cells or AHLM (both passage <5) with CK + DCI medium and organoids were analysed at day 14 for tdTomato expression. Experiments were performed in biological duplicates with AHLM collected from two independent donor lungs. **h**, Organoids from iAEC2s/MRC5 or iAEC2s/AHLM were collected and stained for alveolar (SFTPC), basal (KRT5 or KRT17) or intermediate (KRT8) markers. **i,j**, Ninety percent of iAEC2s/MRC5 organoids were SFTPC<sup>+</sup>/tdTomato<sup>+</sup> (**i**), while ~65% of iAEC2s/AHLM organoids were SFTPC<sup>low</sup>, with the rest SFTPC<sup>neg</sup> (**j**). These organoids were also KRT5<sup>-</sup>/KRT17<sup>+</sup>/KRT8<sup>+</sup>, suggesting only partial differentiation. Representative of *n*=2 independent experiments. **k,l**, iAEC2/AHLM organoids, when co-cultured with a 1:1 mixture of CK/DCI and modified MTEC media show complete transdifferentiation of tdTomato<sup>+</sup> iAEC2s into tdTomato<sup>-</sup>/SFTPC<sup>-</sup>/KRT5<sup>+</sup>/NKX2-1<sup>+</sup> (86.7% of analysed colonies) lung basal cells. The remaining 13.3% of organoids were SFTPC<sup>-</sup>/K5<sup>-</sup>/NKX2-1<sup>+</sup>. Representative of *n*=2 independent experiments.

donors without prior history of lung disease. Library preparation was performed separately and mesenchymal cells were segregated based on *PDGFRA* expression and merged for uniform manifold approximation and projection (UMAP) and clustering into distinct mesenchymal subsets (Fig. 3a). UMAP analysis showed fairly

consistent clustering of subsets across AHLM and IPF donors, respectively (Extended Data Fig. 3c). Fresh AHLM consists mainly of proximal (advential), and distal (alveolar) fibroblasts<sup>16</sup>, along with a small population of smooth muscle (SM; Fig. 3c). However, cultured AHLM underwent a dramatic shift in cellular



identity, with the loss of distal alveolar fibroblast identity similar to IPF mesenchyme (Fig. 3c, blue box and Supplementary Table 1). Instead, a *CTHRC1*-high (*CTHRC1*<sup>hi</sup>) fibroblast cluster and a separate *HAS*-high (*HAS*<sup>hi</sup>) cluster emerged in cultured AHLM that were rare in fresh AHLM but enriched in IPF (Fig. 3b, red boxes in Fig. 3c and Supplementary Table 1). The merged UMAP results demonstrate a clear overlap in the identities of the *CTHRC1*<sup>hi</sup> and *HAS*<sup>hi</sup> fibroblast subsets derived from fresh IPF and cultured AHLM (Fig. 3c). The emergence of these new subsets in culture was very interesting, because both *HAS*<sup>hi</sup> (ref. 8) and *CTHRC1*<sup>hi</sup> (ref. 25) fibroblast subsets have been previously identified as fibroblast subsets in single-cell analysis of IPF lungs. The analysis of differentially expressed genes (DEGs) demonstrated highly concordant marker expressions in the pathologic mesenchymal niche subsets in cultured AHLM and IPF mesenchyme (Fig. 3d, Extended Data Fig. 3d and Supplementary Table 1). Both *CTHRC1*<sup>hi</sup> and *HAS*<sup>hi</sup> subsets arising in cultured AHLM and IPF display features of fibrotic transformation. The *CTHRC1*<sup>hi</sup> fibroblast subset is characterized by high expression of collagen along with *TGFBI* (Fig. 3b,d,e), and the *HAS*<sup>hi</sup> fibroblast subset is characterized by the elevated expression of secreted BMP antagonists that have previously been shown to be elevated in IPF lungs<sup>26</sup> (Fig. 3b,d,f). Analysis of immune subsets in IPF lungs shows that *TGFBI* is also highly expressed in the macrophage population found in fibrotic lungs (Extended Data Fig. 3e–g). This analysis shows that pathologic fibroblast subsets emerge from cultured AHLM that are present in IPF, and they express niche factors that could potentially promote metaplastic transdifferentiation of hAEC2s into basal cells.

### Pathological niche signalling modulates transdifferentiation.

To further define niche factors promoting hAEC2 transdifferentiation, we examined the transcriptome of cultured AHLM versus MRC5 cells by bulk RNA-seq. DEG analysis shows significant upregulation of multiple secreted BMP antagonists in the AHLM along with transforming growth factor- $\beta$  (TGF- $\beta$ ) ligands, whereas BMP ligands and hedgehog antagonist, *HHIP*, were preferentially upregulated in MRC5 cells (Fig. 4a and Supplementary Table 2). *HHIP* is a secreted antagonist of Sonic Hedgehog (SHH) that has previously been identified in both hAEC2s and mesenchymal niche cells<sup>27</sup>, and its expression is reduced in IPF<sup>8</sup>. Hedgehog (Hh) activation can manipulate BMP signalling to modify basal cell metaplasia in murine models<sup>26</sup>, and addition of recombinant *HHIP* to hAEC2 organoids co-cultured with AHLM significantly attenuated *KRT5* expression while increasing *SFTPC* expression in the organoids (Fig. 4b), as confirmed by immunophenotyping of organoid sections (Fig. 4c). *HHIP* treatment attenuated expression of the Hh transcriptional activator, *GLI1*, in AHLM co-cultured with hAEC2s, concurrent with an increase in the expression of BMP ligands (Fig. 4d). Similar to *HHIP*, addition of recombinant BMP4 significantly attenuated hAEC2 transdifferentiation into basal cells

(Fig. 4e,f) and other airway lineages such as club cells (Extended Data Fig. 4a). Conversely, addition of DMH1 (a small-molecule inhibitor of BMP activation) and recombinant TGF- $\beta$ 1 both augmented the number of basal cells derived from hAEC2s co-cultured with MRC5 cells (Fig. 4g,h). This demonstrates that active TGF- $\beta$  signalling promotes hAEC2 transdifferentiation into basal cells, while BMP activation in the hAEC2 niche maintains AEC2 fate, which could account for the differential effects of AHLM versus MRC5 cells in their ability to promote basal cell differentiation.

### hAEC2-derived basal cells resemble IPF metaplastic basal cells.

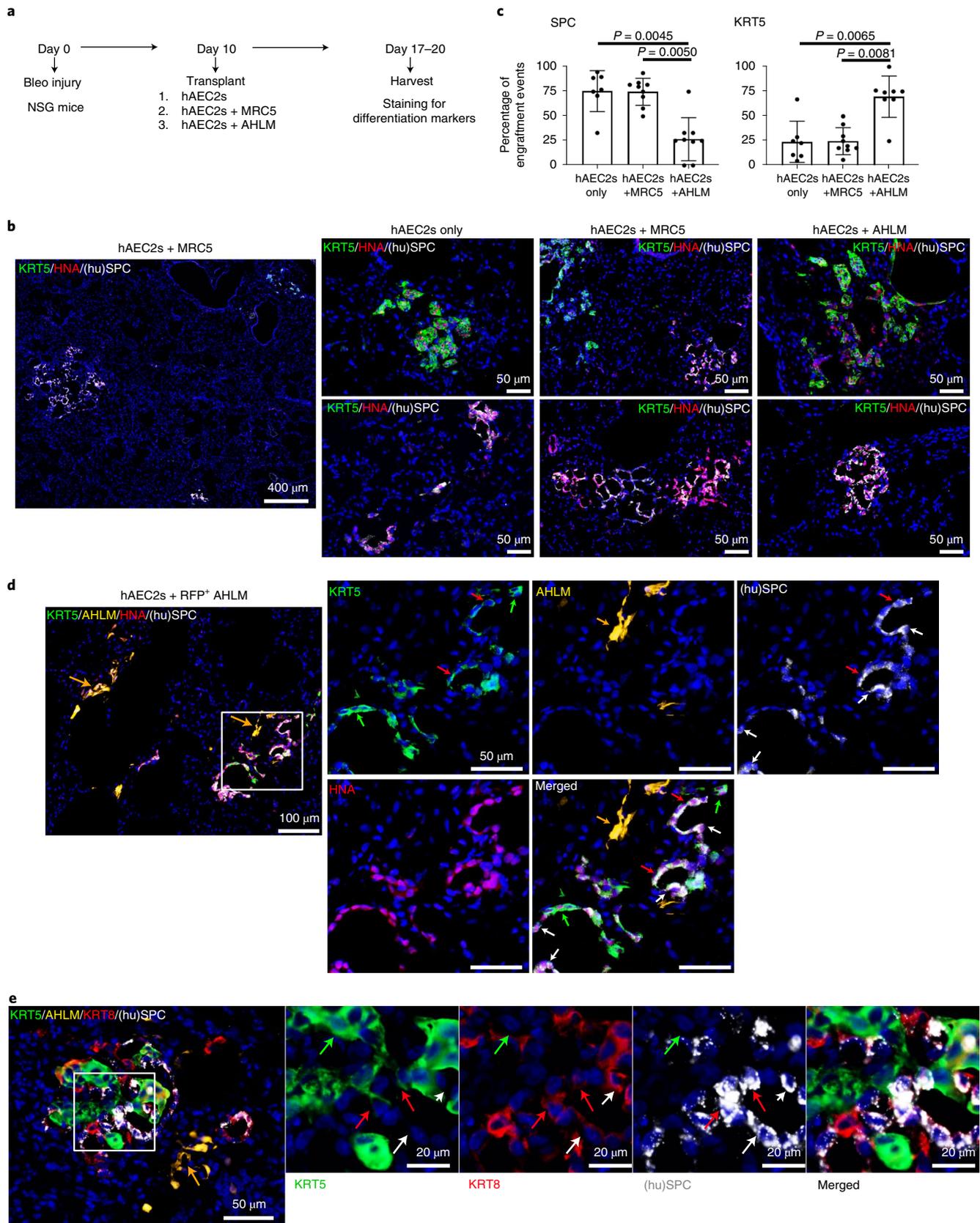
There is increased recognition of basal cell heterogeneity in normal and diseased lungs, with specific basal cell subsets increased in IPF<sup>28</sup>. To compare hAEC2-derived basal cells in organoids with basal cells isolated from intact lungs, we analysed the scRNA-seq results of freshly isolated epithelial cells isolated from a distal fragment (not from the trachea) of normal donor lungs ( $N=2$ ), IPF lungs ( $N=2$ ) and hAEC2-derived basal cells isolated from organoids ( $N=2$ ). Library preparation was performed separately, and basal cells were segregated based on *KRT5* expression and merged for UMAP and clustering into distinct basal subsets (Fig. 5a and Extended Data Fig. 5a). Clustering revealed four distinct subsets of basal cells with a high degree of gene overlap with previously identified basal cell subsets in vivo<sup>28</sup>, including secretory primed (SPB), multipotent (MPB), activated (AB) and proliferating (PB) (Fig. 5a–c and Extended Data Fig. 5b). Interestingly, the plurality of hAEC2-derived basal cells are secretory primed basal cells (Fig. 5a and Extended Data Fig. 5c, red box), a population that has been identified as the predominant basal population in the alveoli of IPF lungs<sup>28</sup>. Basal cell subset-specific marker analysis shows similarities in the expression of genes in hAEC2-derived and IPF basal cells (Fig. 5c).

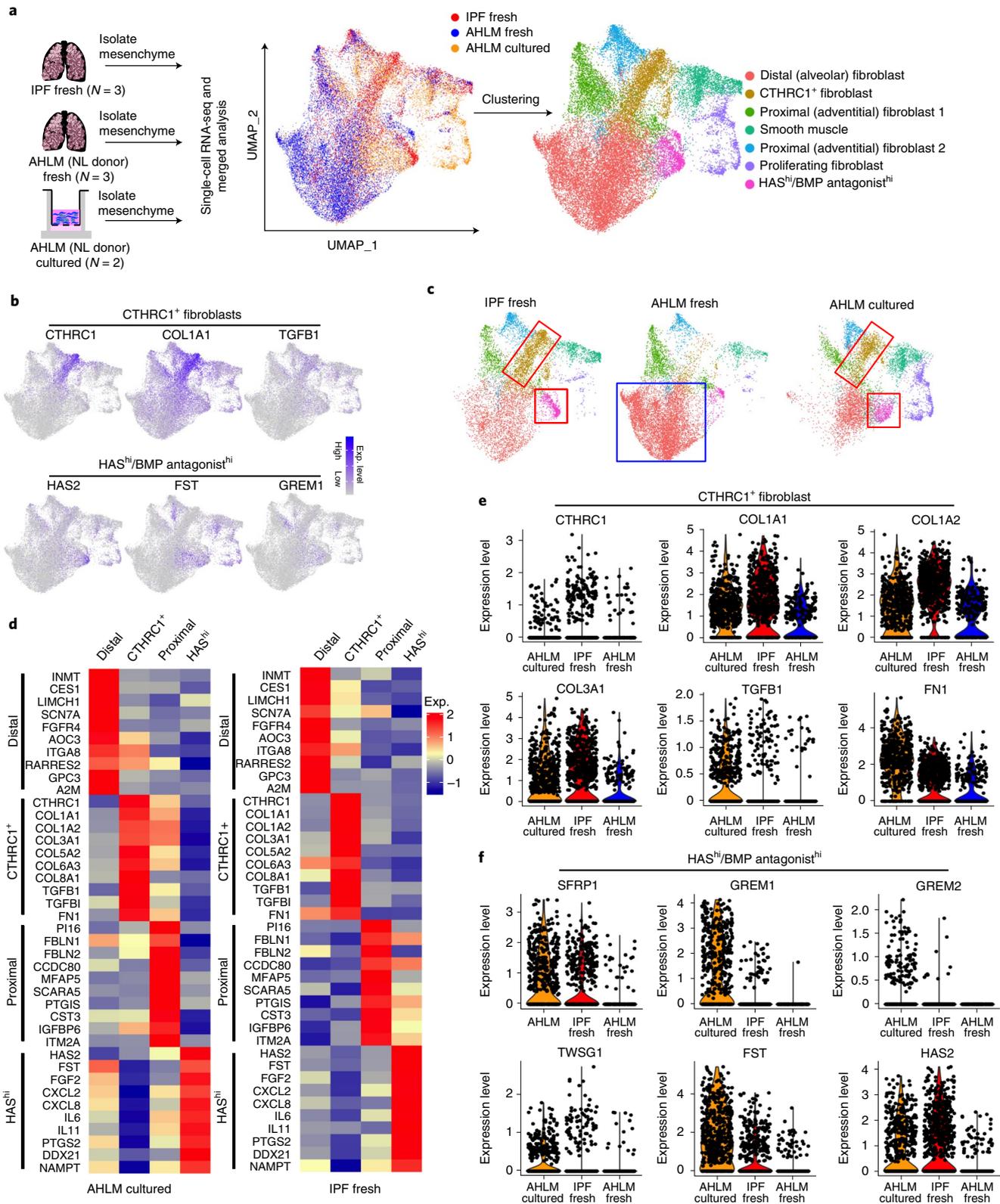
To explore further whether IPF biomarkers are upregulated in hAEC2-derived basal cells, we performed DEG analysis comparing basal cells from normal lung versus hAEC2-derived basal cells. This analysis showed that although hAEC2-derived basal cells express similar canonical basal markers such as *SOX2*, *NGFR* and *TP63*, they over-express markers previously reported to be upregulated in IPF epithelium, such as *KRT14*, *VIM* and *MMP7*<sup>7,4,8,9,29</sup> (Fig. 5d). Histologic comparison of normal lungs, IPF lungs and hAEC2-derived organoids shows that these IPF biomarkers are preferentially present in basal cells from IPF and hAEC2-derived organoids, but not normal lungs (Fig. 5e). Finally, to identify a common gene signature of IPF and hAEC2-derived basal cells, we performed an overlap analysis of DEGs of IPF basal cells and hAEC2-derived basal cells compared with normal lung basal cells. Hypergeometric probability testing shows significant enrichment of overlap genes in DEGs between hAEC2-derived and IPF basal cells (Extended Data Fig. 5d), generating a large list of DEGs that are shared between IPF and hAEC2-derived basal cells (Extended

**Fig. 2 | hAEC2s are capable of metaplastic KRT5<sup>+</sup> transdifferentiation in a fibrotic host in vivo.** **a**, Experimental set-up for xenotransplant experiment testing engraftment of hAEC2s alone or co-transplanted with MRC5 or AHLM. **b**, hAEC2s engrafted in the injured lungs and differentiated towards either KRT5<sup>+</sup> (green) basal cells or remaining as pro-SFTPC<sup>+</sup> (white) hAEC2s, as judged by co-staining with HNA (red) in the lungs of mice transplanted with hAEC2s alone or hAEC2s + MRC5 or hAEC2s + AHLM.  $n=3$  mice per group, representative of  $n=2$  independent replicates. **c**, Co-transplantation of hAEC2s with MRC5 cells does not change the relative rates of hAEC2s engraftment events when compared with hAEC2s alone. However, co-transplantation of AHLM reduced the number of hAEC2 patches observed in the lungs of mice while increasing basal cell differentiation. Each dot represents the number of engrafted regions (>5 cells) in one section. At least two sections per mouse were analysed and  $n=3$  mice per group. Data are expressed as the mean  $\pm$  s.d. The Kolmogorov-Smirnov test was used to determine normality and one-way analysis of variance (ANOVA) followed by the Kruskal-Wallis test was used for multiple comparisons. **d**, Immunofluorescence of engrafted hAEC2s actively transdifferentiating into KRT5<sup>+</sup> basal cells in proximity to engrafted AHLM in the injured lung. Orange arrows, RFP-labelled AHLM; green arrows, KRT5<sup>+</sup>/SFTPC<sup>-</sup> basal cells; white arrows, SFTPC<sup>+</sup>/KRT5<sup>-</sup> hAEC2s; red arrows, KRT5<sup>+</sup>/SFTPC<sup>+</sup> hybrid cells.  $n=3$  mice per group, representative of one independent experiment. **e**, Immunostaining of lungs transplanted with hAEC2s and tdTomato-labelled AHLM showed KRT8<sup>low</sup>/SFTPC<sup>+</sup> hAEC2s (white arrows, inset), KRT8<sup>+</sup>/SFTPC<sup>+</sup> intermediate-hAEC2s (white arrowhead, inset), KRT5<sup>+</sup> basal cells (green arrows, inset) and rare SFTPC<sup>+</sup>/KRT5<sup>+</sup> cells (red arrows, inset) adjacent to tdTomato-labelled AHLM (orange arrow) in areas of engraftment. Data are representative of  $n=3$  mice.

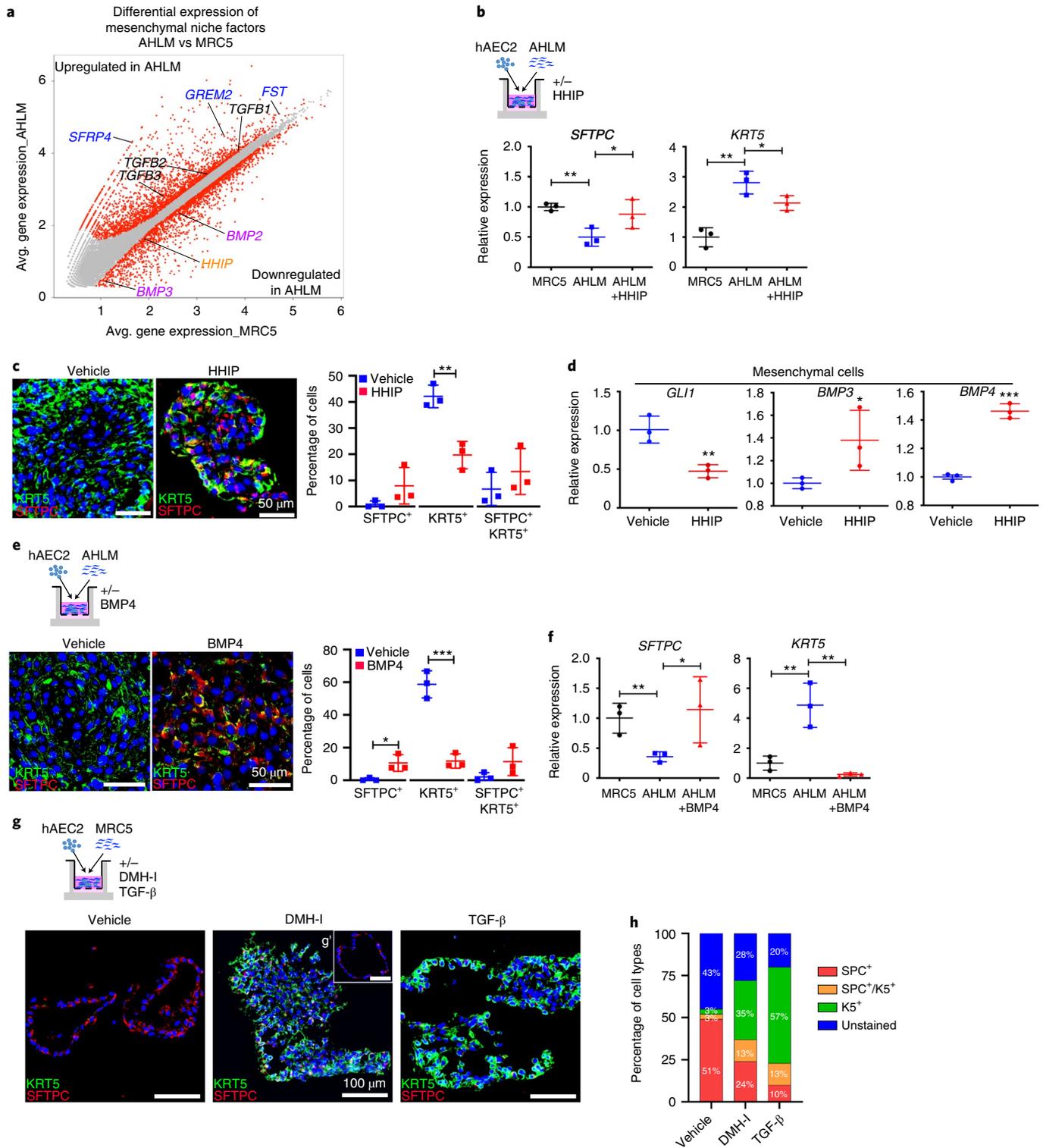
Data Fig. 5e and Supplementary Table 3). These results demonstrate that, although hAEC2-derived basal cells express canonical basal cell gene programs, they share more similarities with IPF basal cells ectopically localized in the diseased alveoli.

**hAEC2 to basal cell transdifferentiation occurs through intermediate states.** Recently, AEC2-derived intermediate cell types have been described in the alveolar space of bleomycin-challenged mice and IPF lungs<sup>8,9,17,19</sup>. To determine whether we observe





**Fig. 3 | Emergence of pathologic mesenchymal niche subsets previously seen in IPF from cultured AHLM. a**, scRNA-seq analysis of fresh (sorted from donor lung, uncultured) AHLM (normal donors,  $N = 3$ , marked as AHLM fresh), fresh IPF mesenchyme ( $N = 3$ , marked as IPF fresh) and cultured AHLM at passage 0–1 isolated from organoids ( $N = 2$ ). **b**, Signature genes in CTHRC1<sup>+</sup> and HAS<sup>hi</sup>/BMP antagonist<sup>hi</sup> clusters. **c**, scRNA-seq analysis showing a decrease of the distal (alveolar) population (blue box) and increase of CTHRC1<sup>+</sup> and HAS<sup>hi</sup>/BMP antagonist<sup>hi</sup> populations (red boxes) in both fresh IPF mesenchyme and cultured AHLM. **d**, Heatmaps showing the signature genes of each cluster in fresh IPF mesenchyme and cultured AHLM. **e**, Violin plots showing the signature genes in the CTHRC1<sup>+</sup> fibroblasts among AHLM cultured, IPF fresh and AHLM fresh. **f**, Violin plots showing the BMP antagonists in the HAS<sup>hi</sup>/BMP antagonist<sup>hi</sup> fibroblasts among AHLM cultured, IPF fresh and AHLM fresh.



similar discrete intermediates in the progressive transdifferentiation of hAEC2s toward basal cells in organoids, we analysed the time course of all epithelial cells (based on *EPCAM* expression) undergoing transdifferentiation in vitro by scRNA-seq. Epithelial cells from day-0 (live sort of HTII-280<sup>+</sup> before culture; Fig. 1a and Extended Data Fig. 1a,b), day-7 (seven days after co-culture), day-14 and day-21 organoids were sequenced separately, then merged and clustered by UMAP, whereby six main clusters emerged (Fig. 6a, Extended Data Fig. 6a and Supplementary Table 4). Four clusters were easily

identified by lineage markers of hAEC2s, basal cells, club cells and ciliated cells, respectively (Extended Data Fig. 6a,c,d). The time course shows a time-dependent loss of the hAEC2 population and gain of the *KRT5*<sup>+</sup> basal cell population, along with an early emergence of two distinct alveolar-basal intermediates (ABIs) bearing both hAEC2 and basal markers at day 7 that recede with time in culture (Fig. 6a, highlighted by a dashed line). ABI1 cells are identified by persistence of *SFTPC*, *ABCA3* and *NAPSA*, along with the emergence of *KRT17* (Fig. 6i and Extended Data Fig. 6e). ABI2 cells

**Fig. 4 | hAEC2-intrinsic and niche factors altered in IPF modulate metaplastic hAEC2 transdifferentiation.** **a**, Bulk RNA-seq analysis of AHLM and MRC5 cells with DEG analysis. Significantly differentially expressed genes ( $P_{\text{adj}} < 0.01$ ,  $\log_2\text{FC} < -0.5$  or  $> 0.5$ ; FC, fold change) are marked as red dots and are otherwise in grey. **b**, Treatment of hAEC2s + AHLM organoids with  $2.5 \mu\text{g ml}^{-1}$  HHIP increases *SFTPC* and decreases *KRT5* expression in the epithelial fractions at day 14 of co-culture. The expression levels of *SFTPC* and *KRT5* in the epithelial fractions of hAEC2s + MRC5 organoids were used as controls. Data are expressed as the mean  $\pm$  s.d. *SFTPC*,  $*P = 0.0399$  and  $**P = 0.0028$ ; *KRT5*,  $*P = 0.0294$  and  $**P = 0.0015$ . **c**, Immunofluorescence analysis of HHIP-treated hAEC2s + AHLM organoids at day 14, quantified by the percentage of *SFTPC*<sup>+</sup>, *KRT5*<sup>+</sup> and *SFTPC*<sup>+</sup>/*KRT5*<sup>+</sup> cells in total cells. Data are expressed as the mean  $\pm$  s.d.  $**P = 0.0023$ . **d**, HHIP-treated AHLM isolated from day-14 organoid co-culture demonstrates decreased *GLI1* and increased *BMP3* and *BMP4* mRNA levels. Data are expressed as the mean  $\pm$  s.d.  $*P = 0.0354$ ,  $**P = 0.0043$  and  $***P = 0.0002$ . **e**, Immunofluorescence analysis and quantification of the hAEC2s + AHLM organoids at day 14 treated with  $50 \text{ ng ml}^{-1}$  BMP4. Data are expressed as the mean  $\pm$  s.d.  $*P = 0.0142$  and  $***P = 0.0005$ . **f**, Treatment of hAEC2s + AHLM organoids with BMP4 increases *SFTPC* and decreases *KRT5* mRNA expression in the epithelial fractions at day 14 of co-culture. The expression levels of *SFTPC* and *KRT5* in the epithelial fractions of hAEC2s + MRC5 organoids were used as controls. Data are expressed as the mean  $\pm$  s.d. *SFTPC*,  $*P = 0.0358$  and  $**P = 0.0066$ ; *KRT5*,  $***P = 0.0062$  (left) and  $**P = 0.0028$  (right). **g, h**, Immunofluorescence analysis (**g**) and quantification (**h**) of day-14 hAEC2s + AHLM organoids treated with DMH-1 ( $1 \mu\text{M}$ ) or TGF- $\beta$ 1 ( $3 \text{ ng ml}^{-1}$ ). DMH-1 treatment gives rise to either *SFTPC*<sup>+</sup> alveolospheres (**g**, middle panel inset, 24% *SFTPC*<sup>+</sup> AEC2s) or *KRT5*<sup>+</sup>/*SFTPC*<sup>+</sup> basal/AEC2 mixed colonies (35% *KRT5*<sup>+</sup> basal cells). TGF- $\beta$ 1 treatment drives more efficient basal cell transdifferentiation by suppressing alveolar fate (10% *SFTPC*<sup>+</sup> AEC2s) and induces basal fate (57% *KRT5*<sup>+</sup> basal cells). The experiment was performed in a technical triplicate and data from three technical replicates are counted as one biological replicate.  $*P < 0.05$  and  $**P < 0.01$ , as determined by unpaired two-tailed Student's *t*-test. Each data point represents one technical replicate.

are identified by low levels of AEC2 markers, higher levels of *KRT17* and absence of *KRT5* (Fig. 6i and Extended Data Fig. 6e). Both these intermediates express high levels of *KRT8* (Fig. 6i and Extended Data Fig. 6i), a marker of intermediate cell types, described previously<sup>17,19</sup>. Of note, we observed a *SFTPC*<sup>low</sup>/*KRT8*<sup>+</sup>/*KRT5*<sup>-</sup> population comparable to ABIs in the epithelial cells derived from hAEC2s in vivo after engraftments in mice (Fig. 2d). Concurrent with the disappearance of the ABIs over time, there is also a progressive emergence of SCGB1A1<sup>+</sup> club cells and a FOXJ1<sup>+</sup> ciliated population in culture (Fig. 6a and Extended Data Fig. 6c,d), suggesting that transdifferentiated *KRT5*<sup>+</sup> cells are mature basal cells capable of further differentiation into club and ciliated lineages.

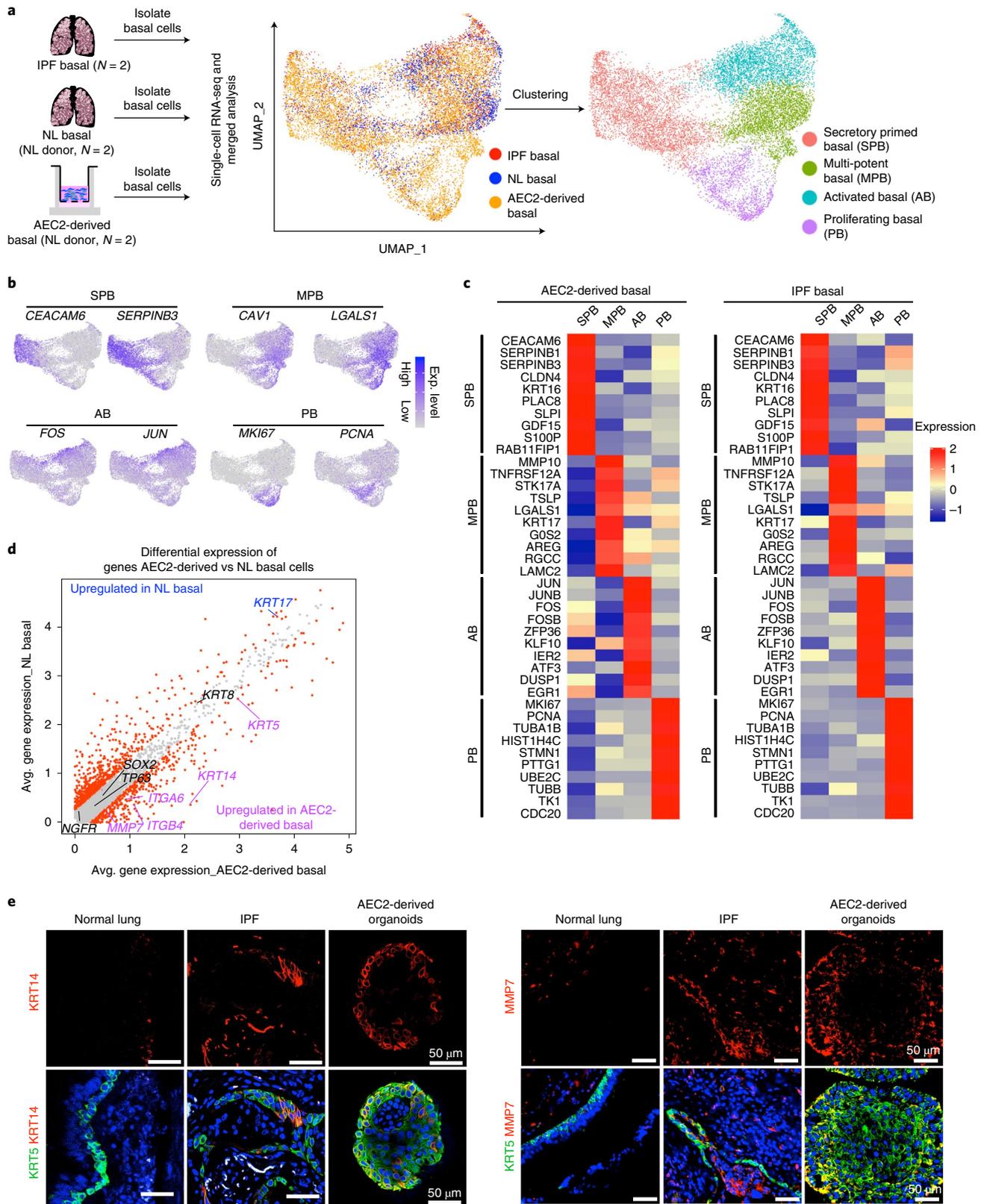
The emergence of ABI1 and ABI2 intermediate populations in the direct line of transdifferentiation to basal cells was confirmed by both Monocle and RNA trajectory analysis (Fig. 6b–d). Visualization of gene expression along the hAEC2 transdifferentiation trajectory shows that the well-established AEC2 marker, *LAMP3*, and the new hAEC2 marker, *HHIP*<sup>27</sup>, are quickly lost first as mature hAEC2s transition into ABIs (Fig. 6c and Extended Data Fig. 6a,b). The emergence of the early basal marker *KRT17* and the transitional-state marker *KRT8*<sup>17–20</sup>, concurrent with the persistence of AEC2 markers *SFTPC*, *ABCA3* and *NAPSA*, marks the ABI1 populations seen early in culture (Extended Data Fig. 6b–d). The presence of the ABI1 and ABI2 is confirmed on immunophenotyping of day-7 organoids early in the transdifferentiation, where we see *KRT17*<sup>-</sup>/*KRT8*<sup>+</sup>/*SFTPC*<sup>+</sup> and *KRT17*<sup>+</sup>/*KRT8*<sup>+</sup>/*SFTPC*<sup>low</sup> cells, in addition to a minority of organoids (9%) that contain only completely transdifferentiated basal cells (*KRT17*<sup>+</sup>/*KRT5*<sup>+</sup>; Fig. 6e). As transdifferentiation progresses, the ABIs give way to mature basal cells, as all hAEC2 markers are lost with the emergence of mature basal cell markers *KRT5* and *KRT14* (Fig. 6c,i and Extended Data Fig. 6a,e,i). An upstream IPA analysis of pathways driving hAEC2s, ABI1 and ABI2s confirmed the distinct expression profiles of these intermediates. ABI1 cells exhibited active upstream drivers typical of hAEC2s (FOXA2, GATA4, NKX2-1) relative to ABI2s (Fig. 6f). Conversely, the ABI2s are preferentially activated by basal cell pathways (TP63, HES1, SOX2) along with TGF- $\beta$ 1 signalling and expression of mesenchymal markers (Fig. 6f and Supplementary Table 5).

#### Identification of AEC2 to basal cell ABIs in fibrotic human lungs.

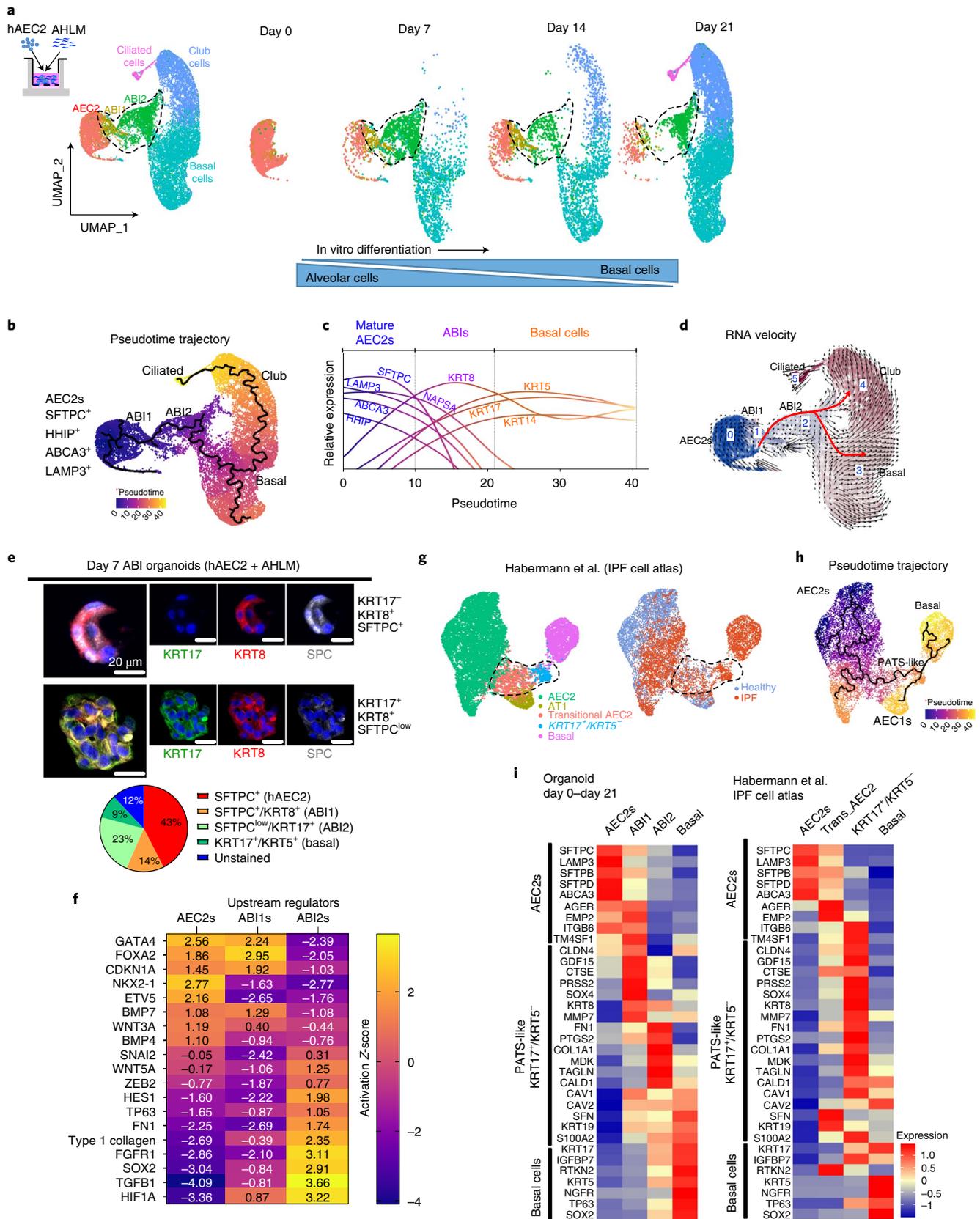
Several recent studies in mice have reported that mAEC2s undergo a transitional state when differentiating into mAEC1s in vitro and in vivo<sup>18–20</sup>. The pre-alveolar type 1 transitional cells (PATS) described in mouse<sup>20</sup> were equated with both the transitional AEC2s and *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> basaloid cells described earlier<sup>8,9</sup> and labelled as

'PATS-like cells' in human. However, basal cells were omitted from the original analysis<sup>20</sup>, even though the *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> basaloid cells express low levels of canonical basal cell transcription factor *TP63*<sup>8,20</sup>. Therefore, we hypothesized that PATS-like cells, inclusive of transitional AEC2s and *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> basaloid cells, also represent intermediate cells between AEC2s and basal cells. To this end, we re-analysed IPF epithelial transcriptomes to include basal cells from the original dataset (from Haberman et al.<sup>8</sup>, analysed by Kobayashi et al.<sup>20</sup>), and then performed Monocle trajectory analysis (Fig. 6g,h). This analysis confirmed an hAEC2–hAEC1 trajectory through transitional AEC2s. However, inclusion of basal cells in the analysis also revealed an equally prominent hAEC2-to-basal cell trajectory through the *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> basaloid cells (Fig. 6g,h). We then directly compared the genes found to be upregulated in PATS-like cells in IPF<sup>20</sup> with the organoid ABIs (ABI1 and ABI2) and observed an almost identical expression pattern (Fig. 6i and Extended Data Fig. 6f–i). The organoid ABI1 population appears similar to the transitional AEC2s, while the organoid ABI2 population appears similar to the *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> basaloid cells (Fig. 6i). Together, these findings suggest that ABIs derived from hAEC2s in our 3D organoid culture recapitulate previously identified intermediate cell populations in IPF lungs and suggest a direct trajectory from hAEC2s to basal cells through these intermediate cell types.

Consistent with our in silico prediction of a continuous trajectory from hAEC2s to ABIs to basal cells, we found evidence of all four cell types in the same cystic region (Fig. 7a and Extended Data Fig. 7a), as judged by protein staining for hAEC2s (*SFTPC*<sup>+</sup>/*KRT8*<sup>low</sup>/*KRT17*<sup>-</sup>/*KRT5*<sup>-</sup>), ABI1 (*SFTPC*<sup>+</sup>/*KRT8*<sup>hi</sup>/*KRT17*<sup>low</sup>/*KRT5*<sup>-</sup>), ABI2 (*SFTPC*<sup>low</sup>/*KRT8*<sup>hi</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup>) and basal cells (*SFTPC*<sup>-</sup>/*KRT8*<sup>-</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>+</sup>). Furthermore, ABI2s can be subdivided based on the presence of TP63, as we found several examples of both *TP63*<sup>+</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> and *TP63*<sup>-</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> ABI2s in the lungs of patients with IPF (Extended Data Fig. 7a), comparable to previously described basaloid (*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> cells that are either *TP63*<sup>+</sup> or *TP63*<sup>-</sup>) populations in IPF<sup>8,9,20</sup>. Next, we systematically inspected IPF lungs for the presence of ABIs by RNA in situ hybridization as a function of the extent of alveolar remodelling using a combination of markers for ABI1, ABI2 and basal cells, all of which express varying levels of *KRT17* messenger RNA (mRNA). As expected, ABI1 or ABI2 were extremely rare in the alveolar regions of normal lung (Extended Data Fig. 7b). Considering the highly heterogeneous histology of an IPF lung, we focused our analysis of ABIs in histologically defined regions (Fig. 7b). The relative frequencies of ABI1s and ABI2s correlated with progression of disease. In the normal-appearing and thickened alveolar regions of IPF lungs reflecting histological regions of minor injury, >80%



**Fig. 5 | hAEC2-derived basal cells resemble metaplastic basal cells seen in IPF. a**, scRNA-seq analysis of basal cells from fresh normal donors ( $N = 2$ , marked as NL basal), fresh patients with IPF ( $N = 2$ , marked as IPF basal) and hAEC2 + AHLM organoids ( $N = 2$ ). **b**, Signature genes in the SPB, MPB, AB and PB clusters. **c**, Heatmaps showing the signature genes of each cluster in hAEC2-derived basal and IPF basal cells. **d**, Differential gene expression analysis of hAEC2-derived basal versus IPF basal cells. **e**, Immunofluorescence analysis of KRT14, MMP7 and KRT5 in normal lung, IPF lung and hAEC2-derived organoids at day 14. Representative of  $n = 2$  independent experiments.



of the *KRT17*<sup>+</sup> cells were SFTPC<sup>+</sup> ABIs, and the remainder were *KRT17*<sup>+</sup>/*SFTPC*<sup>-</sup>/*KRT5*<sup>-</sup> ABIs (Fig. 7c). Conversely, in areas of moderate to severe injury reflected by the presence of microcystic regions, the majority of *KRT17*<sup>+</sup> cells were ABIs (63%) and with

emergence of a small number of *KRT17*<sup>+</sup>/*KRT5*<sup>+</sup> basal cells (10% of *KRT17*<sup>+</sup> cells) (Fig. 7c). Finally, in areas with complete bronchiolization and heavy fibrosis, we found few ABIs (2%) and ABIs (30%), with 68% of *KRT17*<sup>+</sup> cells now acquiring complete basal identity, as

**Fig. 6 | hAEC2 transdifferentiates into KRT5<sup>+</sup> basal cells through ABI cells, in vitro and in vivo.** **a**, scRNA-seq of freshly sorted HTII-280<sup>+</sup> hAEC2s (day 0) and hAEC2s co-cultured with AHLM at days 7, 14 and 21. **b–d**, Pseudotime ordering of epithelial cells from day-0, 7, 14 and 21 organoids (**b,c**) identifies an hAEC2s → ABIs → basal → club → ciliated cell trajectory based on gene expression levels over pseudotime trajectory that was confirmed via RNA velocity (**d**). **e**, Immunofluorescence staining of day-7 hAEC2 + AHLM organoids identifies the presence of KRT17<sup>-</sup>/KRT8<sup>high</sup>/SFTPC<sup>+</sup> ABI1 (14%) and KRT17<sup>+</sup>/KRT8<sup>+</sup>/SFTPC<sup>low</sup> ABI2s (23%) along with SFTPC<sup>+</sup> hAEC2s (43%) and KRT17<sup>+</sup>/KRT5<sup>+</sup> basal cells (9%). Representative of *n* = 2 biologically independent experiments, each performed in technical triplicate. Data are presented as the mean of biological replicates. **f**, IPA of the DEGs between AEC2s, ABI1 and ABI2s identifies the upregulation of alveolar signalling (NKX2-1, ETV5) in AEC2s, WNT and BMP signalling in AEC2s and ABI1 and fibrotic/airway signalling pathways (TGF-β1, HES1, extracellular matrix genes, SOX2 and TP63) in ABI2s. **g,h**, Re-analysis of IPF epithelium from Habermann et al.<sup>8</sup> (**g**) to include basal cells along with PATS-like cells (encompassing transitional AEC2s and KRT17<sup>+</sup>/KRT5<sup>-</sup> cells on UMAP) with pseudotime ordering (**h**). **i**, Heatmaps comparing lineage-specific gene expression of KRT17<sup>+</sup>/KRT5<sup>-</sup> basaloid cells in IPF (labelled as PATS-like cells by Kobayashi et al.<sup>20</sup>) and ABIs in a 3D organoid.

judged by KRT5 expression (Fig. 7c). Furthermore, by immunostaining we found evidence of ABIs in lungs with scleroderma and COVID-19, showing linearly connected AEC2s (SFTPC<sup>+</sup>), ABI1s (SFTPC<sup>+</sup>/KRT17<sup>+</sup>), ABI2s (SFTPC<sup>-</sup>/KRT17<sup>+</sup>/KRT5<sup>-</sup>) and basal cells (KRT17<sup>+</sup>/KRT5<sup>+</sup>) cells in the same area of the disease lung (Fig. 7d and Extended Data Fig. 7c). The results strongly suggest that ABI populations in vivo, similar to the population identified in our 3D organoids, appear as a function of alveolar injuries that result in transdifferentiation of hAEC2s to basal cells.

Analysis of AHLM co-cultured with hAEC2s in our organoid platform identified a *TGFβ1*<sup>hi</sup> mesenchymal population that is marked by *CTHRC1* gene expression, which has been shown previously to be associated with fibroblastic foci<sup>25</sup>. Because ABI2s have elevated TGF-β signalling (Fig. 6f), which can antagonize BMP signalling<sup>30,31</sup>, we tested the hypothesis that aberrant mesenchyme providing a pro-fibrotic *TGFβ1*<sup>hi</sup> niche is spatially associated with ABIs. We probed IPF lung specimens in situ with ABI markers (*SFTPC* and *KRT17*) and the pro-fibrotic mesenchymal marker *CTHRC1*, followed by protein staining for basal marker KRT5 (Fig. 8a). Consistent with our hypothesis, ~10% of ABI1s were adjacent to *CTHRC1*<sup>+</sup> cells, but 83% of ABI2s were found adjacent to *CTHRC1*<sup>+</sup> mesenchyme, suggesting a high degree of association between metaplastic intermediate cells and TGFβ1<sup>hi</sup> pro-fibrotic mesenchymal cells (Fig. 8b). We also found *CTHRC1*<sup>+</sup> mesenchyme to be also associated with 77% of all basal cells found in actively remodelling regions (Fig. 8b). These data indicate that hAEC2s transdifferentiate into metaplastic basal cells through a series of intermediates, at least in part in a fibrotic niche made of TGFβn<sup>hi</sup>, pro-fibrotic mesenchymal populations (Fig. 8c).

## Discussion

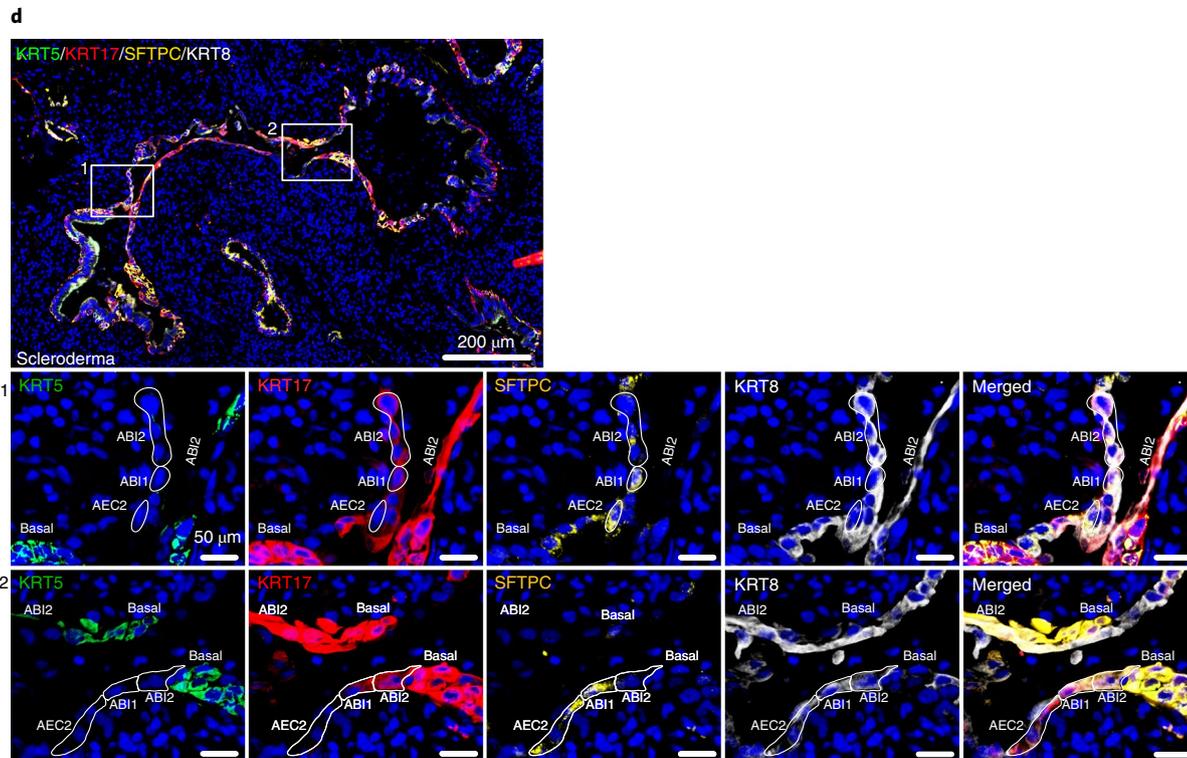
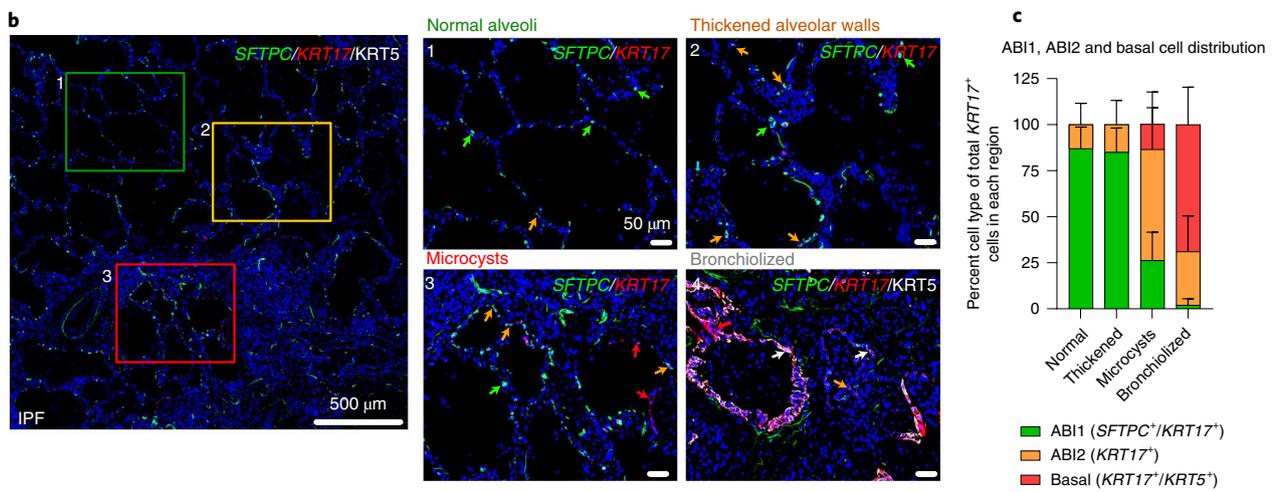
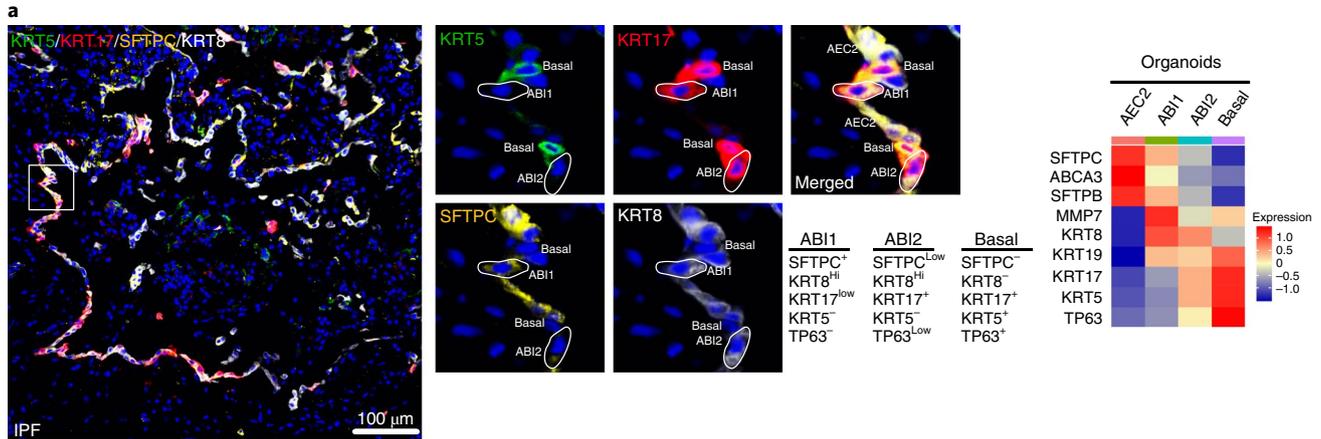
There is currently an assumption that the regenerative capacity of AEC2s, the resident stem cell of the alveolus, is highly conserved between mouse and human<sup>32</sup>. However, the unexpected finding that, unlike mAEC2s, hAEC2s robustly transdifferentiate into functional basal cells with cues from pathological mesenchymal cells arising in vitro invited the possibility that a parallel process occurs in human diseases characterized by severe alveolar injuries in vivo (for example, IPF and COVID-19). Single-cell analysis of the cultured mesenchymal niche cells responsible for the transdifferentiation (AHLM) demonstrated the emergence of two fibroblast subpopulations recently identified in freshly isolated IPF lungs that

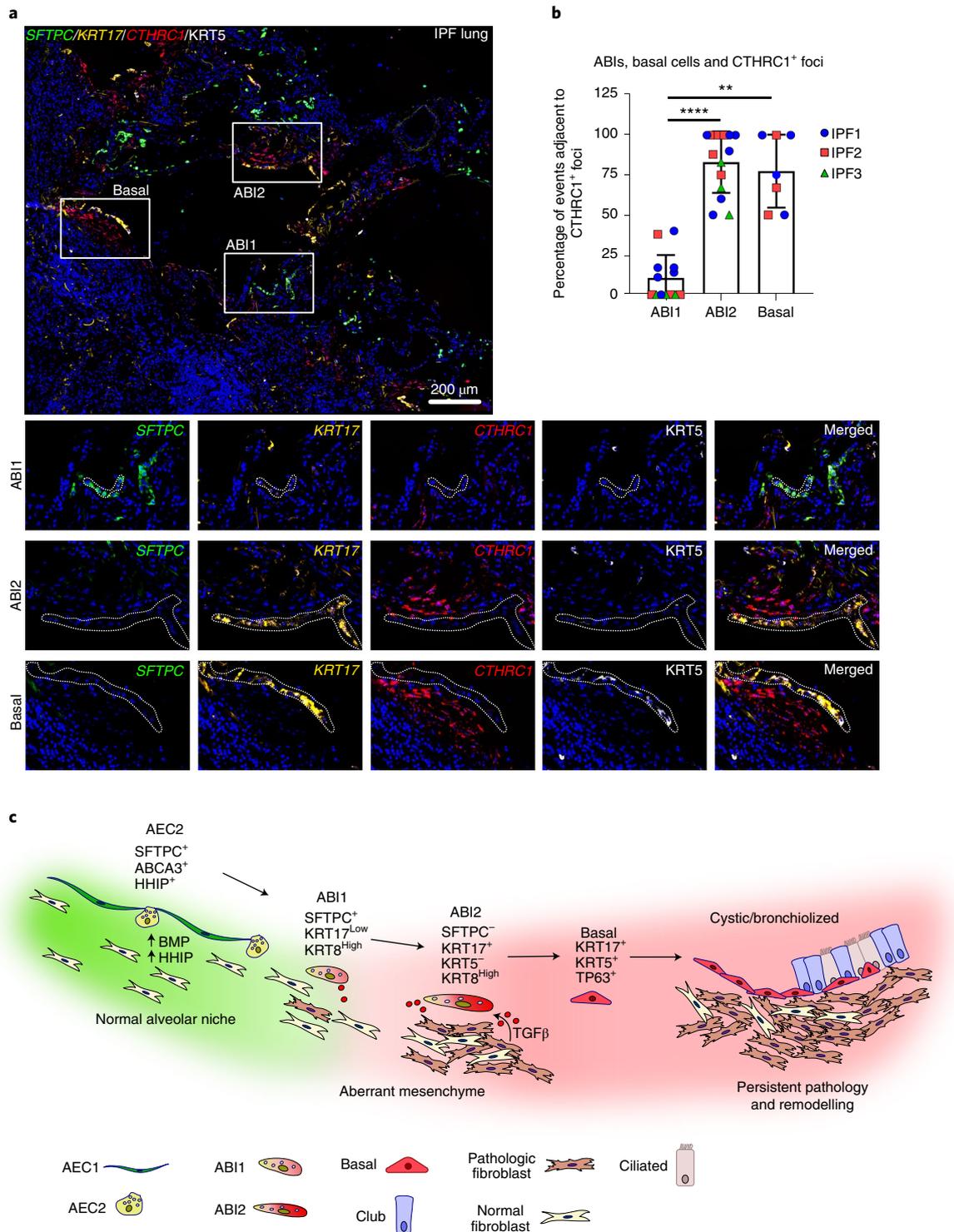
are enriched in either secreted TGF-β1 (*CTHRC1*<sup>hi</sup>) or BMP antagonists (*HAS*<sup>hi</sup>). Similarly, single-cell analysis of the hAEC2-to-basal cell trajectory in vitro revealed the presence of intermediate cell types and basal cell subsets previously identified in IPF lungs. These findings validate our AHLM/hAEC2 organoid assay as an in vitro model for hAEC2-mesenchymal crosstalk that results in stem cell metaplasia seen in severe alveolar injury, and provides experimental confirmation of a stem cell trajectory that is seen in diseased human, but not murine lung.

Another group of recent studies have also identified an intermediate cell state during AEC2 to AEC1 differentiation, first described in mice, that is marked by elevated expression of KRT8 and increased TGF-β1 signalling<sup>17,19,20</sup>. These murine intermediate cell types bear a striking resemblance to the human ABIs identified in our study, suggesting that a KRT8 high intermediate state may not be unique during AEC2-to-AEC1 transdifferentiation in mice, but also evolved to include AEC2-to-basal transdifferentiation in human. Furthermore, ABI1 and ABI2 found in vitro are remarkably similar to the discrete intermediate cell types—that is, transitional AEC2s and KRT5<sup>-</sup>/KRT17<sup>+</sup> aberrant basaloid cells—described in IPF lungs. A comparative IPA analysis of hAEC2, ABI1 and ABI2 transcriptomes (Fig. 6f) reveals the extensive transcriptional reprogramming involved in the transdifferentiation through these intermediate states, including progressive loss of AEC2-promoting signalling, such as NKX2-1, and the emergence of numerous activities promoting mesenchymal features, especially the strong activation of TGF-β signalling in ABI2s. We focused on TGF-β1, because TGF-β1 promotes hAEC2 transdifferentiation to basal cells in organoids, and *CTHRC1*<sup>hi</sup> fibroblasts known to be *TGFβ1*<sup>hi</sup> emerge both in our organoid assay and in IPF tissues. Indeed, over 80% of the ABI2 clusters and 67% of alveolar basal cells were found adjacent to *CTHRC1*<sup>hi</sup> fibroblasts (Fig. 8). These spatial relationships are consistent with the high TGF-β1 signalling observed as an upstream driver of ABI2 signalling (Fig. 6f).

That hAEC2s undergo progressive transdifferentiation to metaplastic basal cells is not unique to IPF. Alveolar metaplastic basal cells are also common in sections of scleroderma and COVID-19 lungs, and these are intermingled with ABI1 and ABI2 intermediates in areas of active remodelling (Fig. 7d). The common finding of ABIs in hAEC2-derived organoids as well as hAEC2 xenografts, as well as in histologic analyses of fibrotic lungs, suggests that hAEC2s are a major source of metaplastic basal cells in diseases with severe

**Fig. 7 | Evidence of the hAEC2-to-basal transition through ABIs in severe alveolar injury.** **a**, Immunostaining of IPF lungs finds hAEC2s (SFTPC<sup>+</sup>), ABI1 (SFTPC<sup>+</sup>/KRT17<sup>low</sup>), ABI2 (SFTPC<sup>low</sup>/KRT17<sup>+</sup>/KRT5<sup>-</sup>) and basal cells (KRT5<sup>+</sup>/KRT17<sup>+</sup>) in the same cystic regions. The heatmap shows the average expression of select candidate genes in each cell type in organoids. Representative of *n* = 3 IPF specimens. **b,c**, In situ analysis of IPF lungs identifying decreasing frequency of ABI1s (SFTPC<sup>+</sup>/KRT17<sup>+</sup>; orange arrows) and increasing frequency of ABI2s (SFTPC<sup>-</sup>/KRT17<sup>+</sup>/KRT5<sup>-</sup>; red arrows) and basal cells (SFTPC<sup>-</sup>/KRT17<sup>+</sup>/KRT5<sup>+</sup>; white arrows) as the severity of fibrosis increases. Green arrows indicate AEC2s (SFTPC<sup>+</sup>/KRT17<sup>-</sup>) (**b**). Representative of *n* = 6 independent IPF specimens used to define normal, thickened, microcystic and bronchiolized regions. Each region included at least four images taken at ×20 from at least three IPF specimens. Data in **c** are presented as the average of all the images and error bars show s.d. **d**, Immunostaining of scleroderma lung identifying linear progression from hAEC2s → ABI1 → ABI2 → basal cells in the same region. Images are representative of *n* = 1 scleroderma lung.





**Fig. 8 | ABIs and basal cells in actively remodelling regions in IPF lungs are adjacent to aberrant CTHRC1<sup>high</sup> mesenchyme. a, b,** RNA in situ hybridization was performed for *SFTPC*, *KRT17* and *CTHRC1* followed by *KRT5* protein staining (**a**). 83% of ABI2s (*SFTPC*<sup>-</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup>) and 77% of basal cells (*KRT17*<sup>+</sup>/*KRT5*<sup>+</sup>) found in actively remodelling regions were in close proximity to clusters of *CTHRC1*<sup>high</sup> fibroblasts (**b**). Representative of *n* = 3 independent IPF lungs analysed. Each dot in **b** represents an average of at least a 3 × 3 tiled region captured at ×20. In total, 160 images were captured at ×20 and counted across three IPF specimens. Data are presented as mean ± s.d. Significance is calculated by one-way ANOVA followed by the Kruskal–Wallis test for multiple comparisons: \*\**P* = 0.0014 and \*\*\*\**P* < 0.0001. **c**, Schematic illustrating a transdifferentiation pathway of hAEC2s to metaplastic basal cells through discrete ABIs and as a likely consequence the accumulation of the dysplastic alveolar structures prominent in IPF. Transdifferentiation develops as a function of the transition of surrounding mesenchyme from a BMP/Wnt high supportive niche (green) to an aberrant, pro-fibrotic niche composed of fibroblasts high in anti-BMPs, TGF-β1 and extracellular matrix (ECM) proteins (red). Although the schematic highlights active TGF-β1 deriving from fibroblasts, ABIs are also a prominent source of TGF-β1 activation and signalling (Fig. 6 and refs. <sup>20</sup> and <sup>19</sup>), probably contributing to the development of aberrant mesenchyme and a pro-fibrotic milieu.

alveolar injury. Future studies are needed to clarify whether and under what circumstances hAEC2 reprogramming towards metaplastic basal cells in the alveoli is reversible, and whether other components of the fibrotic niche such as endothelial cells and immune cells are able to drive the metaplastic phenotype.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-021-00809-4>.

Received: 12 February 2021; Accepted: 4 November 2021;

Published online: 30 December 2021

### References

- Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123**, 3025–3036 (2013).
- Rock, J. R. et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc. Natl Acad. Sci. USA* **108**, E1475–E1483 (2011).
- Smirnova, N. F. et al. Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs. *Respir. Res.* **17**, 83 (2016).
- Xu, Y. et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**, e90558 (2016).
- Xi, Y. et al. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nat. Cell Biol.* **19**, 904–914 (2017).
- Vaughan, A. E. et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621–625 (2015).
- Liang, J. et al. Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. *Nat. Med.* **22**, 1285–1293 (2016).
- Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
- Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
- Kumar, P. A. et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell* **147**, 525–538 (2011).
- Ray, S. et al. Rare SOX2<sup>+</sup> airway progenitor cells generate KRT5<sup>+</sup> cells that repopulate damaged alveolar parenchyma following influenza virus infection. *Stem Cell Rep.* **7**, 817–825 (2016).
- Yang, Y. et al. Spatial-temporal lineage restrictions of embryonic p63<sup>+</sup> progenitors establish distinct stem cell pools in adult airways. *Dev. Cell* **44**, 752–761 (2018).
- Pan, H., Deutsch, G. H., Wert, S. E. & Ontology, S. NHLBI Molecular Atlas of Lung Development Program Consortium Comprehensive anatomic ontologies for lung development: a comparison of alveolar formation and maturation within mouse and human lung. *J. Biomed. Semant.* **10**, 18 (2019).
- Prasse, A. et al. BAL cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 622–630 (2019).
- Zacharias, W. J. et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251–255 (2018).
- Wang, C. et al. Expansion of hedgehog disrupts mesenchymal identity and induces emphysema phenotype. *J. Clin. Invest.* **128**, 4343–4358 (2018).
- Jiang, P. et al. Ineffectual AEC2-to-AEC1 differentiation in IPF: persistence of KRT8<sup>hi</sup> transitional state. *Am. J. Respir. Crit. Care Med.* **201**, 1443–1447 (2020).
- Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382 (2020).
- Strunz, M. et al. Alveolar regeneration through a KRT8<sup>+</sup> transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
- Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).
- Katsura, H. et al. Human lung stem cell-based alveolospheres provide insights into SARS-CoV-2-mediated interferon responses and pneumocyte dysfunction. *Cell Stem Cell* **27**, 890–904 (2020).
- Jacob, A. et al. Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* **21**, 472–488 (2017).
- Hurley, K. et al. Reconstructed single-cell fate trajectories define lineage plasticity windows during differentiation of human PSC-derived distal lung progenitors. *Cell Stem Cell* **26**, 593–608 (2020).
- Weiner, A. I. et al. Mesenchyme-free expansion and transplantation of adult alveolar progenitor cells: steps toward cell-based regenerative therapies. *NPJ Regen. Med.* **4**, 17 (2019).
- Tsukui, T. et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).
- Cassandras, M. et al. Gli1<sup>+</sup> mesenchymal stromal cells form a pathological niche to promote airway progenitor metaplasia in the fibrotic lung. *Nat. Cell Biol.* **22**, 1295–1306 (2020).
- Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
- Carraro, G. et al. Single-cell reconstruction of human basal cell diversity in normal and idiopathic pulmonary fibrosis lungs. *Am. J. Respir. Crit. Care Med.* **202**, 1540–1550 (2020).
- Rosas, I. O. et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med.* **5**, e93 (2008).
- Ning, J., Zhao, Y., Ye, Y. & Yu, J. Opposing roles and potential antagonistic mechanism between TGF-beta and BMP pathways: implications for cancer progression. *EBioMedicine* **41**, 702–710 (2019).
- Dituri, F., Cossu, C., Mancarella, S. & Giannelli, G. The interactivity between TGFβ and BMP signaling in organogenesis, fibrosis and cancer. *Cells* **8**, 1130 (2019).
- Basil, M. C. et al. The cellular and physiological basis for lung repair and regeneration: past, present and future. *Cell Stem Cell* **26**, 482–502 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

This study complies with all relevant ethical regulations approved by the UCSF Institutional Review Board (IRB) and Institutional Animal Care and Use Committee (IACUC).

**Human lung tissue.** Studies involving human tissue were approved by the UCSF IRB. All subjects provided written informed consent. Peripheral regions of the normal lungs were obtained to select for the distal-most regions of the lung from brain-dead donors that were rejected for lung transplantation. IPF, scleroderma and acute respiratory distress syndrome (COVID-19) lung specimens were taken from the periphery of the lung at the time of lung transplant. The age and sex of tissue donors are listed in Supplementary Table 6.

**Animal studies and treatment.** Mice were housed in accordance with the UCSF IACUC protocol in humidity- and temperature-controlled rooms on a 12-h light-dark cycle with free access to food and water. Eight- to twelve-week-old animals were used for the experiments. For labelling mAEC2s, *Sftpc<sup>creERT2/+</sup>;R26R<sup>mTmG/+</sup>* mice<sup>33,34</sup> were intraperitoneally administered with tamoxifen (cat. no. T5648, Sigma) at 200 mg per kg body weight per day for three consecutive days. *NOD.Cg-Prkdc<sup>scid</sup>Il2rg<sup>tm1Wjl/SzJ</sup>* (NSG) mice have been described previously<sup>35,36</sup>. Animal studies utilized a minimum of four mice per group. Mice were injured with oral aspiration of bleomycin (2.1 U per kg body weight). Mice were weighed twice a week. Mice were euthanized between days 17 and 20 post injury for histopathological analysis.

**Histology and immunofluorescence. Paraffin embedding.** Mouse lungs were inflated with and fixed in 4% paraformaldehyde (PFA) overnight at 4°C. Human lung pieces were fixed in 4% PFA overnight at 4°C. The lungs were then washed with PBS four times for 30 min each at 4°C, then dehydrated in a series of ethanol (30%, 50%, 70%, 95% and 100%). The dehydrated lungs were incubated with xylene for 1 h at room temperature (r.t.), then embedded in paraffin. The lungs were sectioned at 8 µm on a microtome.

**Optimal Cutting Temperature (OCT) embedding.** Lungs inflated with 94%OCT/2%PFA/4%PBS were fixed with 4% PFA for 1 h at r.t., washed with PBS for 4 h at r.t. and embedded in OCT after 30% and 15% sucrose gradient washing. Organoids in 3D Matrigel were fixed with 4% PFA for 30 min at r.t. or overnight at 4°C, then washed in PBS overnight three times, followed by embedding in OCT. Sections (8-µm) were cut on a cryostat.

**Immunofluorescent staining.** Paraffin sections were twice incubated in xylene for 10 min, then rehydrated in ethanol washes (100%, 95%, 70%, 50% ethanol) for 5 min each. OCT-embedded slides were fixed in 4% PFA at r.t. for 10 min, then washed with PBS. For both paraffin- and OCT-embedded slides, antigen retrieval (cat. no. DV2004MX, Biocare) was performed for 30 min at 95°C or at 155°C followed by incubation with sodium borohydride (Sigma) in PBS. Slides were washed with 0.1% Tween-20 in PBS (PBST), blocked (3% donkey serum in PBST) for 1 h, and then incubated with primary antibodies overnight at 4°C. The following primary antibodies were used: anti-SFTPC (cat. no. AB3786, Millipore, 1:2,000), anti-SFTPC (cat. no. sc-518029, Santa Cruz, 1:100), anti-KRT5 (cat. no. 905901, BioLegend, 1:500), anti-KRT14 (cat. no. CBL197, Millipore, 1:200), anti-KRT17 (cat. no. sc-393002, Santa Cruz, 1:100), anti-p63 (cat. no. 13109, CST, 1:100), anti-KRT8 (cat. no. TROMA-I, DSHB, 0.045 µg ml<sup>-1</sup>), anti-LAMP3 (cat. no. AF0487, R&D, 1:200), anti-ABCA3 (cat. no. ab99856, Abcam, 1:500), anti-human nuclear antigen (cat. no. ab86129, Abcam, 1:500), anti-MMP7 (cat. no. MAB3315, Millipore, 1:200), anti-Ac-Tubulin (cat. no. T7451, Sigma, 1:500), anti-p-SMAD1/5/8 (cat. no. AB3848-I, Millipore, 1:100), anti-beta IV tubulin (cat. no. ab11315, Abcam, 1:200) and anti-SCGB1A1 (cat. no. MAB4218, R&D, 1:500). Slides were washed with PBST and then incubated with secondary antibodies for 1 h at r.t. The following secondary antibodies were used at 1:250: anti-chicken IgY (H + L) AF647 (cat. no. 703-605-155, Jackson ImmunoResearch), anti-rabbit immunoglobulin-G (IgG; H + L) AF555 (cat. no. A-31572, Thermo Fisher), anti-rabbit IgG (H + L) AF488 (cat. no. A-21206, Thermo Fisher), anti-mouse IgG (H + L) AF555 (cat. no. A-31570, Thermo Fisher) and anti-rat IgG (H + L) AF488 (cat. no. A-21208, Thermo Fisher). 4',6-Diamidino-2-phenylindole (DAPI) was added for 5 min, then the slides were mounted. Images were captured using a Zeiss Imager M1 and analysed using AxioVision 4.8.2 or Zeiss ZEN v3.1 software (Zeiss). Where indicated, multiple images at ×20 were captured using the 'MosaicX' function and stitched together using the 'Tile Stitch' function in AxioVision or ZEN. Final images were prepared in Affinity Designer v1.10.1.1142.

**Organoid composition calculation.** An organoid comprising SFTPC<sup>+</sup> cells and no KRT5<sup>+</sup> cells was counted as an alveolar organoid (SFTPC<sup>+</sup>), an organoid comprising both SFTPC<sup>+</sup> and KRT5<sup>+</sup> cells was counted as a hybrid organoid (SFTPC<sup>+</sup>/KRT5<sup>+</sup>) and an organoid comprising KRT5<sup>+</sup> cells and no SFTPC<sup>+</sup> cells was counted as a basal organoid (KRT5<sup>+</sup>). Organoids with no detectable staining of either SFTPC or KRT5 were counted as unstained organoids (SFTPC<sup>-</sup>/KRT5<sup>-</sup>). To characterize and compare basal cells derived from AEC2/MRC5 and AEC2/AHLM co-cultures, organoids were collected and stained at day 14 with basal cell

markers. Individual KRT5<sup>+</sup> cells were counted for protein expression of other basal markers and represented as a percentage of total KRT5<sup>+</sup> cells. In lineage analysis of day-7 organoids, and testing the effects of BMP4, HHHIP, CHIR, KGF, TGF-β1 and DMH-1 on the transdifferentiation of hAEC2s into KRT5<sup>+</sup> basal cells, the percentages of SFTPC<sup>+</sup>, KRT5<sup>+</sup> and SFTPC<sup>+</sup>/KRT5<sup>+</sup> cells and ABIs were calculated by counting all the cells across multiple organoids on at least one slide.

**Lung digestion and fluorescence-activated cell sorting.** Human lung pieces were washed in PBS (2×) and HBSS (1×) for 10 min at r.t., compressed to remove liquid, and dissected into 1-cm<sup>3</sup> pieces. Dispase II (15 U ml<sup>-1</sup>; cat. no. 17105041, Thermo Fisher), 225 U ml<sup>-1</sup> collagenase type I (cat. no. 17100017, Thermo Fisher), 100 U ml<sup>-1</sup> DNase I (cat. no. DN25, Sigma-Aldrich) and 1% Pen/Strep in 1× HBSS were used to digest the pieces for 2 h at 37°C. Fungizone (1:400) was added for the final 30 min of the digestion. The digested tissue was liquified in a blender. The suspension was serially filtered through gauze and 100-µm, 70-µm and 40-µm strainers. Red blood cells were removed using red blood cell lysis buffer (Sigma). After Fc blocking, immune and endothelial cells were depleted using biotinylated CD45 (cat. no. 368534, BioLegend, 1:200), CD31 (cat. no. 13-0319-80, eBioscience, 1:200) and CD11b (cat. no. 301304, BioLegend, 1:200) antibodies and running through streptavidin beads (cat. no. 17663, Stemcell Technologies) at 25 µl ml<sup>-1</sup>. The following antibodies were used at 1:200: anti-CD45-APC-Cy7 (cat. no. 304014, BioLegend), anti-CD11b-APC-Cy7 (cat. no. 557754, BD), anti-CD31-APC-Cy7 (cat. no. 303120, BioLegend), anti-CD326-PE (cat. no. 324206, BioLegend), anti-HTII-280 (cat. no. 303118, Terrace Biotech) and anti-mouse IgM-AF488 (cat. no. A-21042, Thermo Fisher, 1:1,000). Doublets and dead cells were excluded based on forward and side scatters and DRAQ7 (cat. no. 7406S, Cell Signaling, 1:200) or DAPI fluorescence. hAEC2s were sorted as live/EpCAM<sup>+</sup>/HTII-280<sup>+</sup> cells, and AHLM cells were sorted as live/CD45<sup>-</sup>/CD11b<sup>-</sup>/CD31<sup>-</sup>/EpCAM<sup>+</sup> cells.

Mouse lungs were incubated in the same digestion cocktail for 45 min at 37°C. The mixture was passed through a 70-µm cell strainer, resuspended in RBC lysis buffer, and filtered through a 40-µm cell strainer. Cells were stained with antibodies for 30 min at 4°C. Antibodies were used at 1:200: CD45-AF700 (cat. no. 560510, BD), CD31-APC/Fire750 (cat. no. 102528, BioLegend) and CD326-BV421 (cat. no. 563214, BD). mAEC2s were sorted using endogenous green fluorescent protein (GFP) from tamoxifen-induced *Sftpc<sup>creERT2/+</sup>;R26R<sup>mTmG/+</sup>* lungs. AMLM cells were sorted based on the selection of live EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD31<sup>-</sup>.

**Freezing/thawing primary human cells.** Cells were resuspended in F12 medium at 10<sup>7</sup> cells per ml, followed by the addition of an equal volume of 2× freeze solution (2% 1.5 M HEPES, 10% FBS, 78% F12 and 10% DMSO). The desired volume of cell aliquots was then frozen at -80°C and stored in liquid nitrogen. Frozen cells were thawed and an equal volume of warm recovery medium (DMEM with 10% FBS) was added and incubated for 1 min, followed by the addition of 12 ml of warm DMEM to the centrifuge tube. Cells were spun at 550g for 4 min and stained for fluorescence-activated cell sorting (FACS).

**Cell culture.** Mesenchymal cells were cultured in DMEM/F-12 (cat. no. 11330032, Thermo Fisher) with 10% FBS and 1% Pen/Strep. Cells were used within the first five passages of either being received from ATCC (cat. no. CCL-171) for MRC5 cells or being isolated from donor lungs for AHLM. The construct pLKO5-trFP (Addgene plasmid 57823) was used to generate RFP-expressing lentiviruses (Lenti-RFP). For labelling AHLM, AHLM cells were transduced by Lenti-RFP with 10 µg ml<sup>-1</sup> polybrene (cat. no. TR-1003-G, Millipore). iPSC-derived AEC2s with SFTPC-tdTomato reporter (iAEC2s)<sup>23</sup> were generated and maintained in CK + DCI medium as described previously<sup>22</sup>. The purity of iAEC2 culture was assessed at every passage by flow cytometry, with a cell population with >96% tdTomato<sup>+</sup> cells considered pure, as described previously<sup>22,23</sup>.

**Organoid assay.** AEC2s and mesenchymal cells were co-cultured (5,000 AEC2s: 30,000 mesenchymal cells per well) in modified MTEC medium diluted 1:1 in growth factor-reduced Matrigel (cat. no. CB-40230A, Thermo Fisher). Modified MTEC culture medium is composed of small airway basal medium (SABM, cat. no. CC-3118, Lonza) with insulin, transferrin, bovine pituitary extract, retinoic acid and Epidermal Growth Factor (EGF) as per the SAGM Bullet Kit and 0.1 µg ml<sup>-1</sup> cholera toxin (cat. no. C8052, Sigma), 5% FBS and 1% Pen/Strep. The cell suspension-Matrigel mixture was placed in a transwell and incubated with 10 µM ROCK inhibitor (cat. no. 72252, Stemcell) for the first 48 h. Each experimental condition was performed in triplicate. Where applicable, BMP4 (cat. no. 314-BP-010, R&D Systems, 50 ng ml<sup>-1</sup>), TGF-β1 (cat. no. 100-21, Peprotech, 3 ng ml<sup>-1</sup>), DMH-1 (cat. no. 73632, Stemcell, 1 µM), HHHIP (cat. no. 9280-HP-050, R&D, 2.5 µg ml<sup>-1</sup>), CHIR (cat. no. 4423, Tocris, 3 µM) and KGF (cat. no. 251KG01050, R&D, 100 ng ml<sup>-1</sup>) were added to the medium after 48 h and replenished in every medium change. iAEC2s were co-cultured with either MRC5 or AHLM cells (5,000 AEC2s: 30,000 mesenchymal cells per well; both MRC5 and AHLM with passage <5) and maintained with either CK/DCI medium or a 1:1 mixture of CK/DCI and modified MTEC medium as described above. Colonies were assayed after 7, 14 and 21 days. For the mesenchyme-free hAEC2 culture, hAEC2s were isolated from donor lungs via FACS as described above and cultured as previously described<sup>21</sup>. After two or three passages, the purity of hAEC2s was assessed by protein staining

for SFTPC and KRT5. Pure hAEC2s were then co-cultured with either MRC5 or AHLM cells as described above.

To extract RNA from organoids, the cell–Matrigel mixture in the transwell was washed with PBS and incubated in the lung digestion cocktail for 1 h at 37°C with intermittent resuspension. The mixture was removed from the transwell and resuspended in TrypLE (cat. no. 12563011, Thermo Fisher) and shaken at 37°C for 20 min. Cells were blocked with human FcR blocking reagent (cat. no. 564220, BD, 1:50) for 10 min at 4°C, then stained with biotin anti-CD326 (cat. no. 324216, BioLegend, 1:250) for 30 min at 4°C. Streptavidin beads (cat. no. 17663, STEMCELL, 1:50) were added to isolate the epithelial cells, and the rest of the cells were mesenchymal cells. For sequencing, organoids were digested as above and FACS-sorted for live EpCAM<sup>+</sup> and EpCAM<sup>-</sup> cells.

**Quantitative polymerase chain reaction.** RNA was extracted from cells using a PicoPure RNA isolation kit (cat. no. KIT0204, Applied Biosystems). Complementary DNA (cDNA) was synthesized from total RNA using a SuperScript Strand Synthesis System (cat. no. 18080044, Thermo Fisher). Quantitative PCR (qPCR) was performed using SYBR Green (cat. no. F415L, Thermo Fisher). Relative gene expression levels were defined using the  $\Delta\Delta Ct$  method. The qPCR primers (IDT) are listed in Supplementary Table 7.

**RNA in situ hybridization.** PFA-fixed OCT-embedded sections were used for RNA in situ hybridization with an RNAScope multiplex fluorescent v2 assay (ACDBio). Briefly, 7- $\mu$ m sections of normal or IPF lungs were washed, protease-dependent antigen retrieval was performed, and probes were hybridized for 2 h at 40°C, followed by step-wise amplification of each probe. RNA probes for *SFTPC* (cat. no. 452561-C1 and -C2, ACDBio), *KRT17* (cat. no. 463661-C3, ACDBio) and *CTHRC1* (cat. no. 413331, ACDBio) were used. Following completion of RNA in situ hybridization, immunostaining was performed after blocking, as already described.

**Xenotransplantation assay.** Freshly sorted hAEC2s, 700,000 alone or mixed with 200,000 cultured AHLM or MRC5 cells (passage <4) in a 40- $\mu$ l volume (1 $\times$  PBS), were transplanted at 10 days post bleomycin injury into the lungs of NSG mice via oral aspiration. Transplanted mice were euthanized 8–12 days post-transplant (a total of 18–22 days post bleomycin injury).

**Single-cell transcriptomics.** Single-cell sequencing was performed on a 10X Chromium instrument (10X Genomics) at the Institute of Human Genetics in UCSF as described before<sup>37</sup>. Briefly, cells were isolated via FACS and then loaded on a Chromium Controller instrument to generate single-cell Gel Bead-In-EMulsions (GEMs). Libraries were prepared by performing reverse transcription on a Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad), then GEMs were collected to amplify cDNA using a Bio-Rad C1000 Touch Thermal Cycler. SPRiselect (Beckman Coulter) was used to select for amplified cDNA. Indexed sequencing libraries were constructed using the Chromium Single-Cell 3' library kit (10X Genomics) and sequenced on a NovaSeq 6000 instrument (Illumina) with the following parameters: Read 1 (26 cycles), Read 2 (98 cycles) and i7 index (eight cycles) to obtain a sequencing depth of ~100,000 reads per cell. Reads were aligned to the appropriate mouse or human genome and quantified using the Cell Ranger Single-Cell Software Suite.

**Quantification and statistical analysis.** *Statistical analysis.* All statistical analyses were performed in GraphPad Prism. Unpaired one-tailed *t*-tests were used to determine the *P* values, and the data in the graphs are presented as mean  $\pm$  s.d. For quantification of xenotransplants, the Kolmogorov–Smirnov test was used to determine normality and one-way ANOVA followed by the Kruskal–Wallis test was used for multiple comparisons. For normally distributed data, ordinary one-way ANOVA followed by Tukey's multiple comparisons test was performed. For all tests, *P* values < 0.05 were considered significant. The representation factor was calculated to represent the number of overlapping genes divided by the number of expected overlapping genes drawn from two independent groups, as calculated on [nemates.org](http://nemates.org) with a base value of 30,000 genes in the human genome. A representation factor of >1 indicates more overlapping genes than expected of two independent groups.

**Immunofluorescence image quantification.** Sections were imaged for quantification on a Zeiss Lumar V12 or Zeiss AxioImager.M1 microscope. Cell counts for stained cells were performed on Fiji using the 'Cell Counter' plug-in. The results were averaged between each specimen and s.d. values were calculated per condition. For quantification of xenografts, at least three mice in each group were analysed with at least two sections taken at 200  $\mu$ m apart to capture different regions of the lung. Any HNA<sup>+</sup> region of more than five cells was counted as one engrafted region. At least three lobes were analysed at each section for three mice per condition, and a minimum of two sections per mouse were analysed. Data are presented as mean  $\pm$  s.d., \*\**P* < 0.01 (one-way ANOVA followed by Kruskal–Wallis for multiple comparison). For quantification of ABIs, at least six distinct areas per IPF specimen (*n* = 3), covering at least three different zones of disease severity, were counted. Zones of disease severity were defined as follows: (1) normal-looking alveoli,

alveoli with a single layer of DAPI<sup>+</sup> cells; (2) thickened alveolar walls, alveoli with multiple layers of DAPI<sup>+</sup> cells in the alveolar wall; (3) microcysts, cyst-like structures that line and overlay the fibrotic regions as judged by accumulated DAPI<sup>+</sup> cells in alveolar space; (4) heavily fibrotic/bronchiolized zone, fibrotic lesions with autofluorescent collagen deposition around bronchiolized epithelium. *SFTPC*<sup>+</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> cells were detected by mRNA in situ for *SFTPC* and *KRT17* and protein staining for *KRT5* and were counted as ABI1s, *SFTPC*<sup>-</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> cells were counted as ABI2s and *SFTPC*<sup>-</sup>/*KRT17*<sup>+</sup>/*KRT5*<sup>+</sup> cells were counted as mature basal cells. IPF lung specimens (*n* = 3) were probed with probes against *SFTPC*, *KRT17* and *CTHRC1* mRNA and *KRT5* protein staining to identify ABIs and adjacent *CTHRC1*<sup>+</sup> mesenchyme. At least a 3  $\times$  3 mosaic image was captured at  $\times$ 20 (nine images covering an area of 3.28 mm<sup>2</sup>), stitched together and counted as one data point. In total, 160 images were captured at  $\times$ 20 across three IPF lungs and every instance of ABI1, ABI2 or basal cell in the actively remodelled region was counted. A cell cluster of at least five cells was counted as an ABI or a *CTHRC1*<sup>+</sup> mesenchymal region. A mesenchymal region was marked as adjacent to ABIs if it was within a three-cell distance of an ABI cluster.

**Analysis of single-cell RNA-sequencing.** FASTQ files were run through Cell Ranger v2.1.1 software with default settings for de-multiplexing, aligning reads with STAR software to Hg19 or GRCh38, and counting unique molecular identifiers (UMIs). Seurat package v4.0<sup>38</sup> in R v3.5.1 was used for downstream analysis. Low-quality cells were filtered (expressing fewer than 200 genes, >10% mitochondrial reads and >6,000 unique gene counts). Principal component analysis was performed on log-normalized and scaled data using 2,000 variable genes. The top 10 principal component analyses were used for clustering and visualized using the UMAP algorithm in the Seurat package. Monocle trajectory analysis was performed using Monocle 3<sup>39</sup> by importing the counts from the Seurat object. RNA velocity was calculated using the scVelo v0.1.16 package<sup>40</sup> in Python v3.6 and velocity calculations were overlaid on UMAP projections calculated in Seurat. Single-cell transcriptomes of healthy and IPF lungs were obtained from GSE135893 and GSE132771 and processed using Seurat. AEC2s, AEC1s, transitional AEC2s, *KRT17*<sup>+</sup>/*KRT5*<sup>-</sup> cells and basal cells were extracted and analysed. RNA values were averaged for each cluster using the AverageExpression function followed by the DoHeatmap function in Seurat to generate heatmaps. Genes specific to the PATS-like cells reported in Kobayashi et al.<sup>20</sup> were picked for heatmaps. The lists of DEGs were identified with a Model-based Analysis of Single-cell Transcriptomics (MAST) test. Upstream regulator analysis of gene lists containing significantly differentially expressed genes (*P*<sub>adj</sub> < 0.05, logFC < -0.15 or > 0.15) was done with Ingenuity Pathway Analysis (Qiagen).

**Bulk RNA-sequencing analysis.** Total RNA was extracted from AHLM and MRC5 cells using the RNeasy Kit (cat. no. 74004, Qiagen). Sequencing was performed using HiSeq 4000. Quality control of reads was conducted using FastQC (Babraham Bioinformatics). Ligation adaptors were removed using Cutadapt and Sickle 1.33. Sequencing reads were aligned using STAR 2.7.9a, and UCSC human GRCh38/hg38 was used as the reference genome. The differential gene expression list was generated using DESeq2.

**Statistics and reproducibility.** No statistical method was used to predetermine sample size. No data were excluded from analyses. Sample ID was blinded to the investigator for image quantification.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

## Data availability

The RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under the following accession codes: GSE148992 (human bulk RNA-seq), GSE150068 (organoids and IPF) and GSE150247 (human scRNA-seq). Previously published scRNA-seq data that are re-analysed here are available at GSE135893 and GSE132771. All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

## Code availability

No custom codes were developed and used in this manuscript. All codes are available by request to the corresponding author.

## References

- Chapman, H. A. et al. Integrin  $\alpha 6 \beta 4$  identifies an adult distal lung epithelial population with regenerative potential in mice. *J. Clin. Invest.* **121**, 2855–2862 (2011).
- Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
- Coughlan, A. M. et al. Myeloid engraftment in humanized mice: impact of granulocyte-colony stimulating factor treatment and transgenic mouse strain. *Stem. Cells Dev.* **25**, 530–541 (2016).

36. Shultz, L. D. et al. Human lymphoid and myeloid cell development in NOD/LtSz-*scid* IL2R $\gamma^{null}$  mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* **174**, 6477–6489 (2005).
37. Kathiriya, J. J., Brumwell, A. N., Jackson, J. R., Tang, X. & Chapman, H. A. Distinct airway epithelial stem cells hide among club cells but mobilize to promote alveolar regeneration. *Cell Stem Cell* **26**, 346–358 (2020).
38. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
39. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
40. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

### Acknowledgements

We thank A. Baldwin for providing technical assistance, Parnassus Flow Cytometry Core for assistance with cell sorting for bulk and single-cell RNA analysis (P30DK063720), E. Wan and the Institute for Human Genetics Core for processing of single-cell RNA samples and high-throughput sequencing. The GEO accession number for raw RNA-sequencing data is given in the Methods. This work is supported by NIH grants nos. DP2AG056034, R01HL142552 and R01HL155622 to T.P., R01HL128484, R35HL150767 and U01HL134766 to H.A.C. and F32HL143931-01A1 and

K99HL155785-01 to J.J.K., a tobacco-related disease research programme postdoctoral award to C.W. and a Nina Ireland Program Award to M.M. for human lung collection.

### Author contributions

J.J.K., C.W., H.A.C. and T.P. conceived the experiments and wrote the manuscript. C.W., J.J.K., M.Z., A.B., C.L.S., M.C. assandras, M. Cohen, B.W. and K.-D.A. performed the experiments, collected samples and analysed data. D.N.K., P.W., B.W. and M.M. provided materials and input to the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

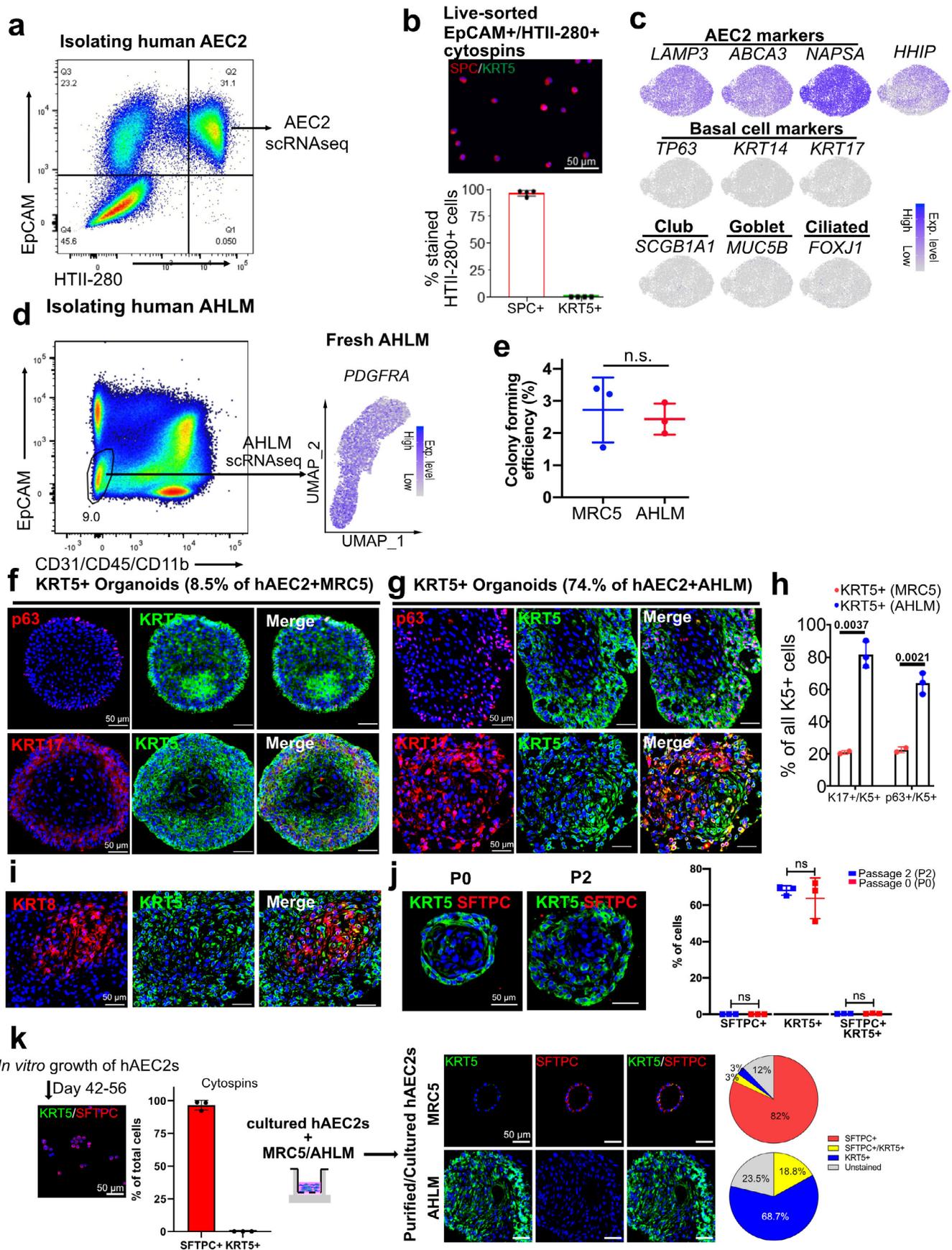
**Extended data** is available for this paper at <https://doi.org/10.1038/s41556-021-00809-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41556-021-00809-4>.

**Correspondence and requests for materials** should be addressed to Chaoqun Wang or Harold A. Chapman.

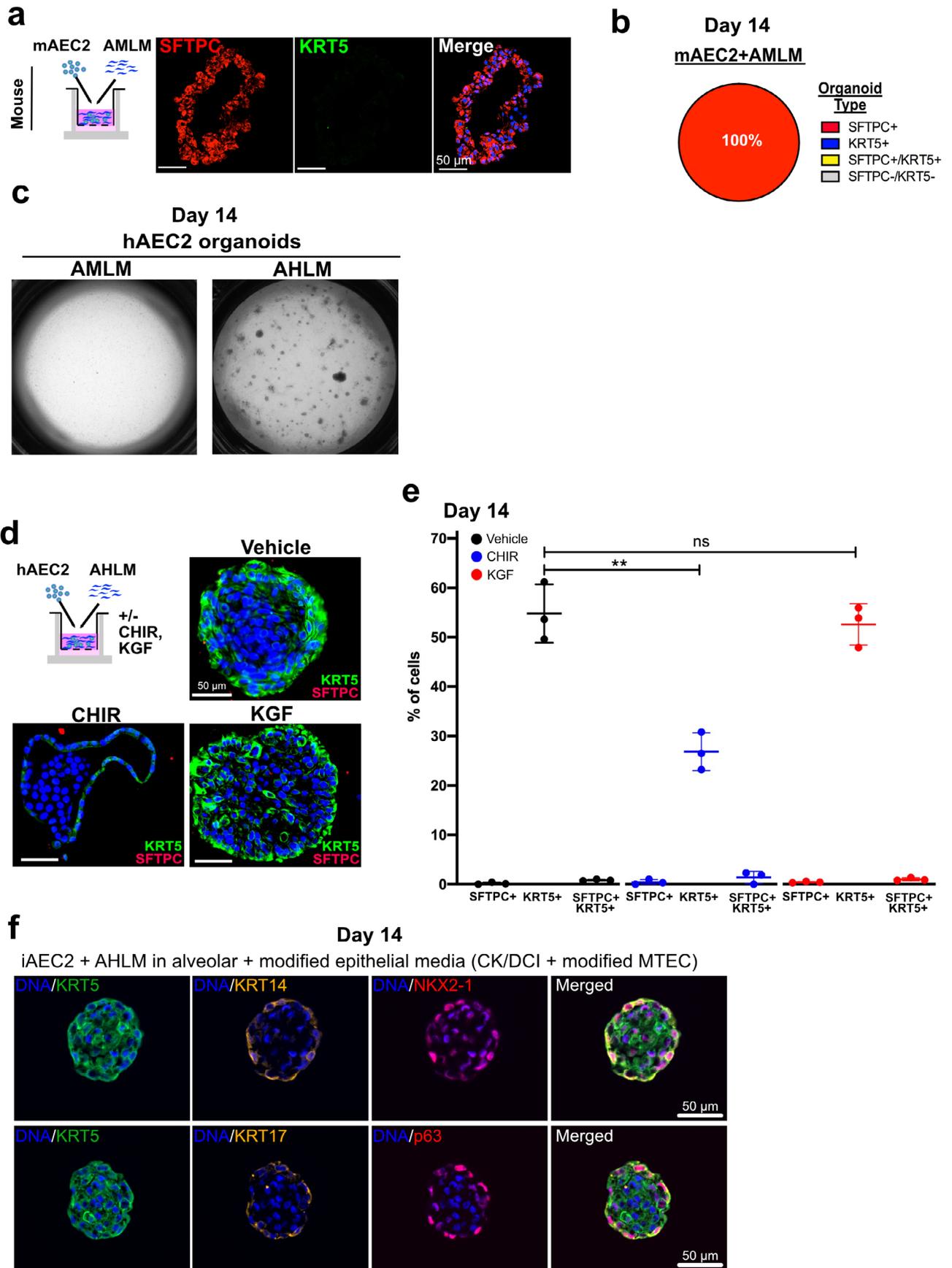
**Peer review information** *Nature Cell Biology* thanks Bi-Sen Ding and the other, anonymous, reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



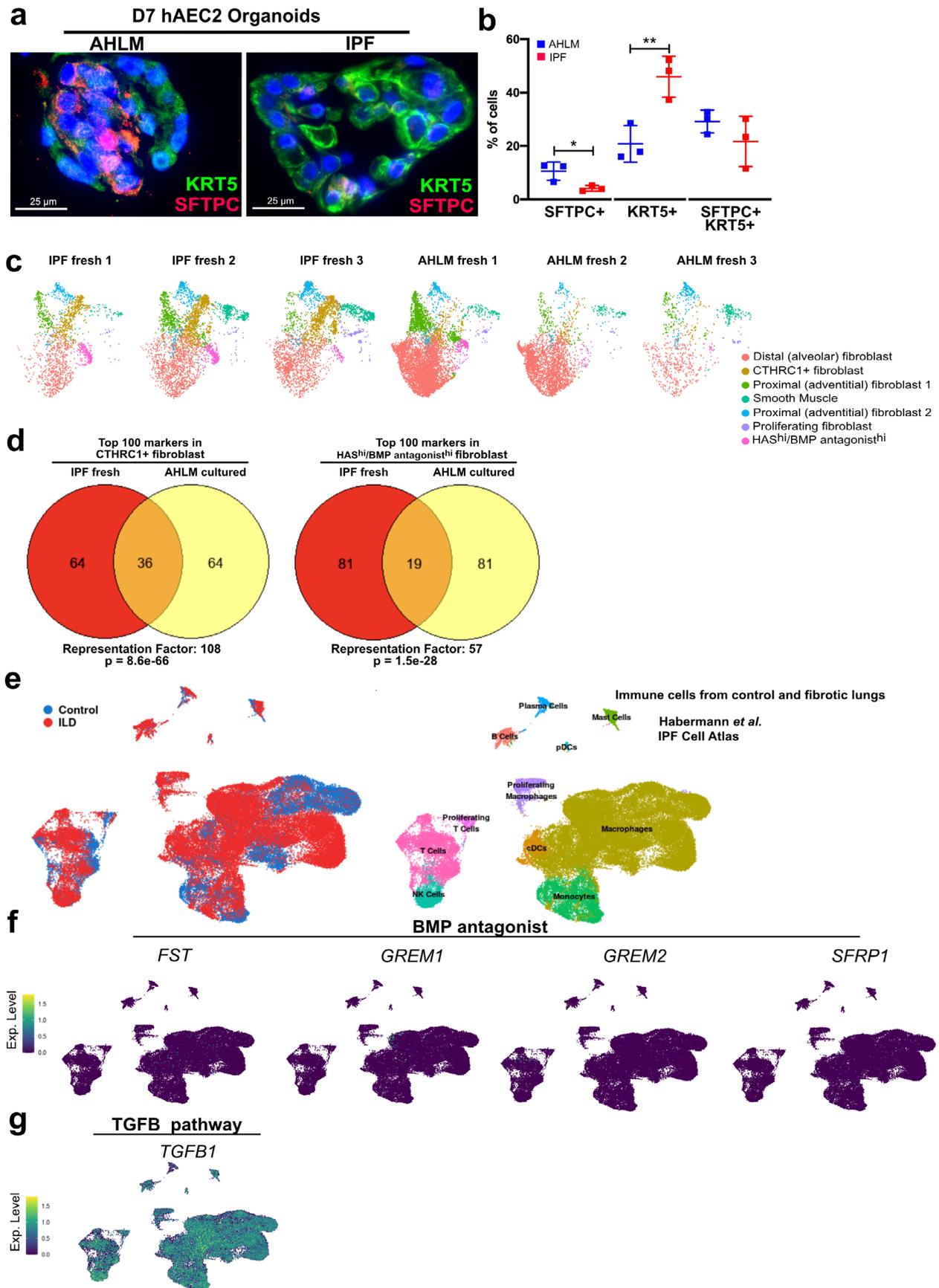
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Purified hAEC2s transdifferentiate into KRT5+ basal cells when co-cultured with AHLM.** **(a)** FACS plot shows the strategy for isolating hAEC2s. After selecting live/CD45-/CD31-/CD11b- cells, hAEC2s were further isolated by gating EpCAM+/HTII-280+ cells. **(b)** Quantification of SFTPC+ and KRT5+ cells in freshly sorted HTII-280+ cells by cytopins. Data are mean  $\pm$  SD, representative of  $n=4$  biologically independent human lung cytopins analyzed. **(c)** confirmation of the purity of HTII-280+ cells by scRNA-seq showed no appreciable detection of non-AEC2 markers. **(d)** scRNA-seq of AHLM isolated as live/CD31-/CD45-/CD11b-/EpCAM- cells that express *PDGFRA*. **(e)** Colony forming efficiency of hAEC2/MRC5 and hAEC2/AHLM co-culture at D14. Data are mean  $\pm$  SD, representative of  $n=3$  biologically independent for AHLM and hAEC2 and representative of  $n=3$  independent experiments for MRC5. **(f)** and **(g)** IF analysis of p63 and KRT17, and KRT5 in the hAEC2s + MRC5 and hAEC2s+AHLM organoids at D14. Images are representative of  $n=2$  (in **f**) and  $n=3$  (in **g**) biological replicates. **(h)** Quantification of KRT17+ and p63+ cells in total KRT5+ basal cells in the hAEC2s + MRC5 and hAEC2s+AHLM organoids at D14.  $n=2$  (for MRC5) and  $n=3$  (for AHLM) biologically independent samples were analyzed. Data are mean  $\pm$  SD. **(i)** IF analysis of KRT8 in the hAEC2s+AHLM organoids at D14. Representative of  $n=2$  independent experiments. **(j)** IF analysis and quantification of KRT5+ and SFTPC+ cells in the hAEC2-derived organoids co-cultured with freshly sorted AHLM (Passage 0) and cultured AHLM (Passage 2) at D14. Each dot represents a technical replicate. Representative of  $n=2$  independent experiments. **(k)** Enrichment of hAEC2s using a mesenchyme-free system, confirmed by IF analysis and quantification of SFTPC+ and KRT5+ cells (left), and co-culture of the enriched hAEC2s with AHLM or MRC5, the organoids are analyzed at D14 (right). Data are expressed as the mean  $\pm$  SD, representative of  $n=3$  biologically independent experiments.



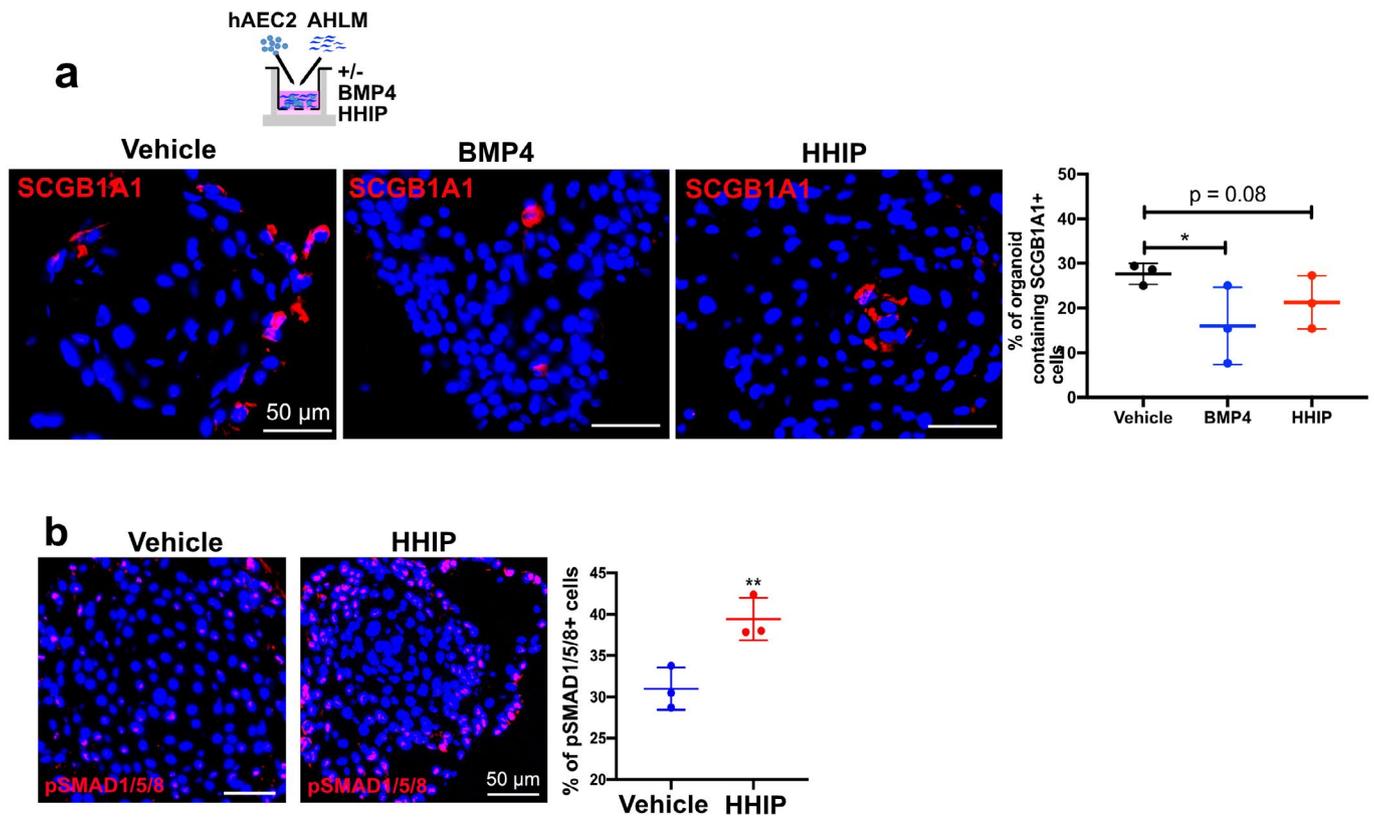
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Difference in niche provided by mouse and human lung mesenchyme and components of the CK/DCI media on hAEC2s-to-basal transdifferentiation.** **(a)** IF analysis of mouse AEC2s (mAEC2s)-derived organoids co-cultured with the adult mouse lung mesenchyme (A MLM) at D14. mAEC2s (GFP+) were sorted from the lung of adult *SftpcCreERT2/+;R26RmTmG/+* mouse. Representative of  $n = 3$  independent experiments. **(b)** Quantification of lineages of organoids (SFTPC+ organoids, organoids with both SFTPC+ cells and KRT5+ cells, or KRT5+ organoids) in mAEC2+ A MLM organoids. **(c)** hAEC2s co-cultured with either adult murine lung mesenchyme (A MLM) or AHLM. **(d,e)** IF analysis and quantification of hAEC2+ AHLM organoids treated with  $3\ \mu\text{M}$  CHIR or  $100\ \text{ng/ml}$  KGF at D14. Representative of  $n = 2$  independent experiments. Data are expressed as the mean  $\pm$  SD, representative of  $n = 3$  technical triplicates. \*\*  $p = 0.0012$ . **(f)** IF analysis of human iAEC2+ AHLM organoids cultured in 1:1 dilution of CK-DCI:SABM/MTEC standard media show canonical basal markers (KRT14, KRT17, P63) and lung lineage (NKX2-1). Data are expressed as the mean  $\pm$  SD, representative of  $n = 2$  independent experiments. \*\* indicates  $p < 0.01$  as determined by unpaired one-tailed student's t-test.

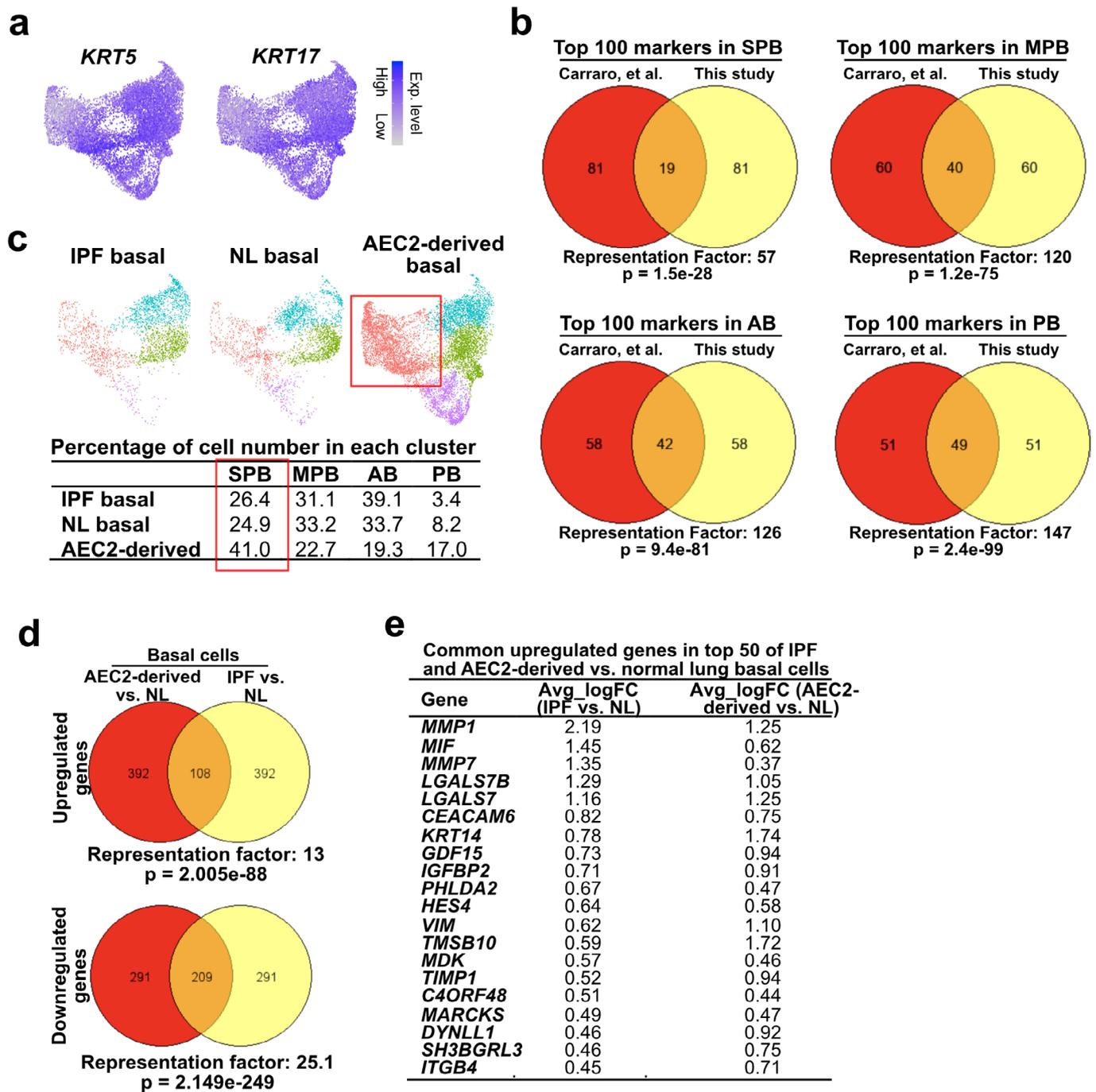


Extended Data Fig. 3 | See next page for caption.

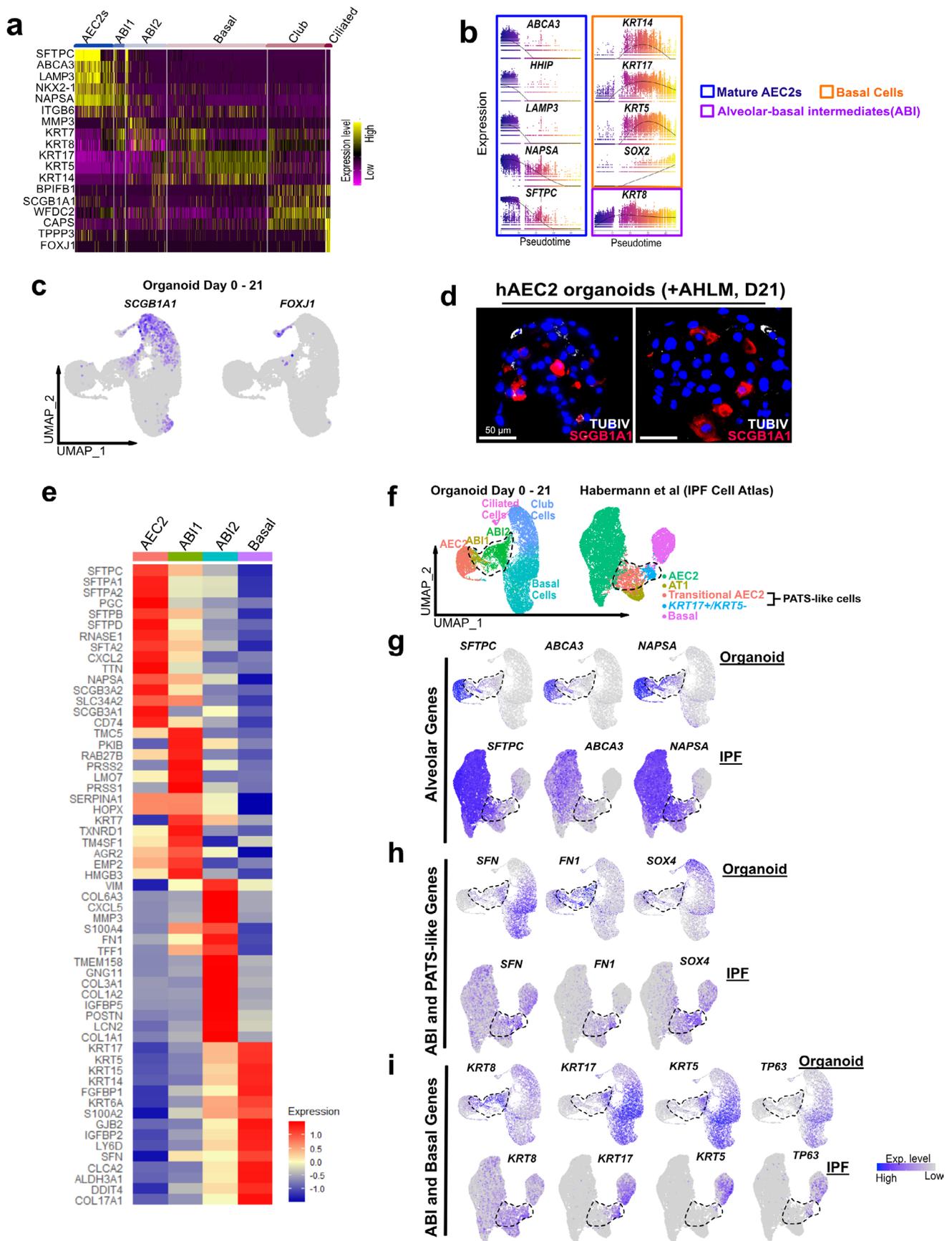
**Extended Data Fig. 3 | Comparison of pathologic mesenchymal niche subsets arising from IPF and cultured AHLM.** (a, b) IF analysis of hAEC2s-derived organoids co-cultured with human adult normal lung (AHLM) and IPF mesenchyme at D7, and quantified by the percentage of SFTPC+, KRT5+, and SFTPC+/KRT5+ cells in total cells. Data are expressed as the mean  $\pm$  SD, representative of n=3 technical replicates. \* p=0.0182 and \*\* p=0.0066. Representative of n=2 independent experiments. (c) UMAP plots of mesenchyme from each individual IPF or normal donor that was used in scRNA-seq analysis. (d) Comparison of top 100 markers in *CTHRC1*+ (left) and *HAS*<sup>high</sup>/BMP antagonist<sup>high</sup> (right) clusters of uncultured IPF (IPF fresh) and cultured AHLM with statistical comparison for degree of gene overlap of two independent gene sets denoted by representation factor (see methods). (e, f, and g) UAMP plots show BMP antagonists and *TGFB1* expression in IPF lungs (from IPF Cell Atlas, Habermann et al.). Data are expressed as the mean  $\pm$  SD. \* indicates p<0.05 and \*\* indicates p<0.01 as determined by unpaired one-tailed student's t-test. Each data point on the graph represents a technical replicate.



**Extended Data Fig. 4 | Effect of HHIP and BMP4 in modulating metaplastic hAEC2 transdifferentiation. (a)** IF analysis of SCGB1A1 in the hAEC2+ AHLM organoids treated with 2.5  $\mu$ g/ml HHIP and 50 ng/ml BMP4, quantified by percentage of organoid containing SCGB1A1+ cells. Data are expressed as the mean  $\pm$  SD. \*  $p = 0.0441$ . **(b)** IF analysis of pSMAD1/5/8 in the hAEC2+ AHLM organoids treated with 2.5  $\mu$ g/ml HHIP, quantified by percentage of pSMAD1/5/8+ cells in total cells. Data are expressed as the mean  $\pm$  SD. \*\*  $p = 0.0079$ . Experiment was performed in a technical triplicate. \* indicates  $p < 0.05$  and \*\* indicates  $p < 0.01$  as determined by unpaired two-tailed student's t-test.

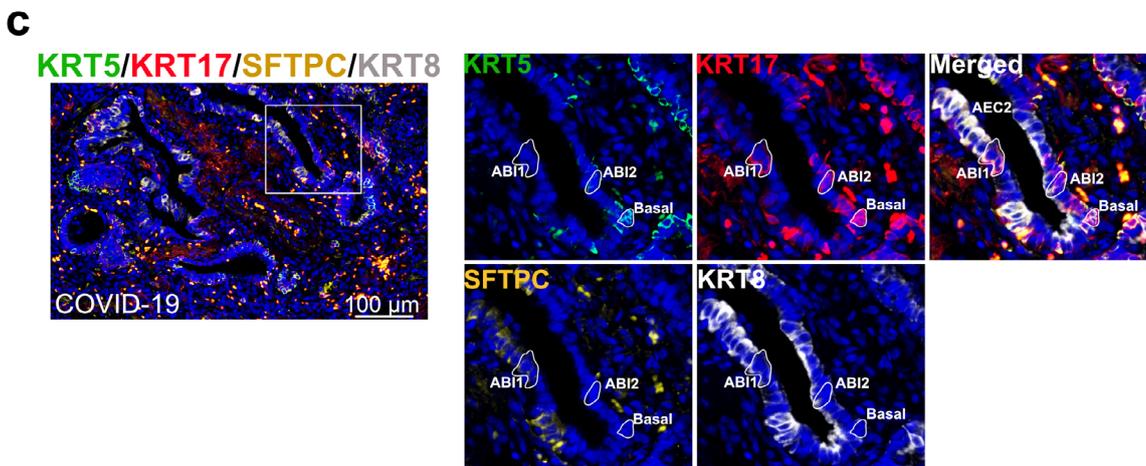
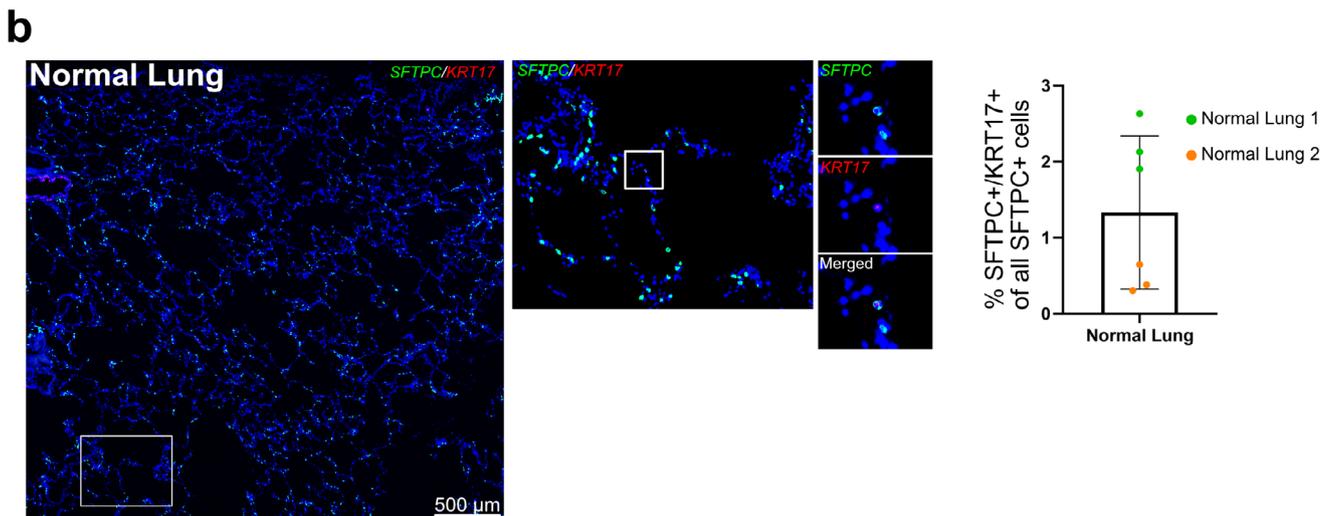
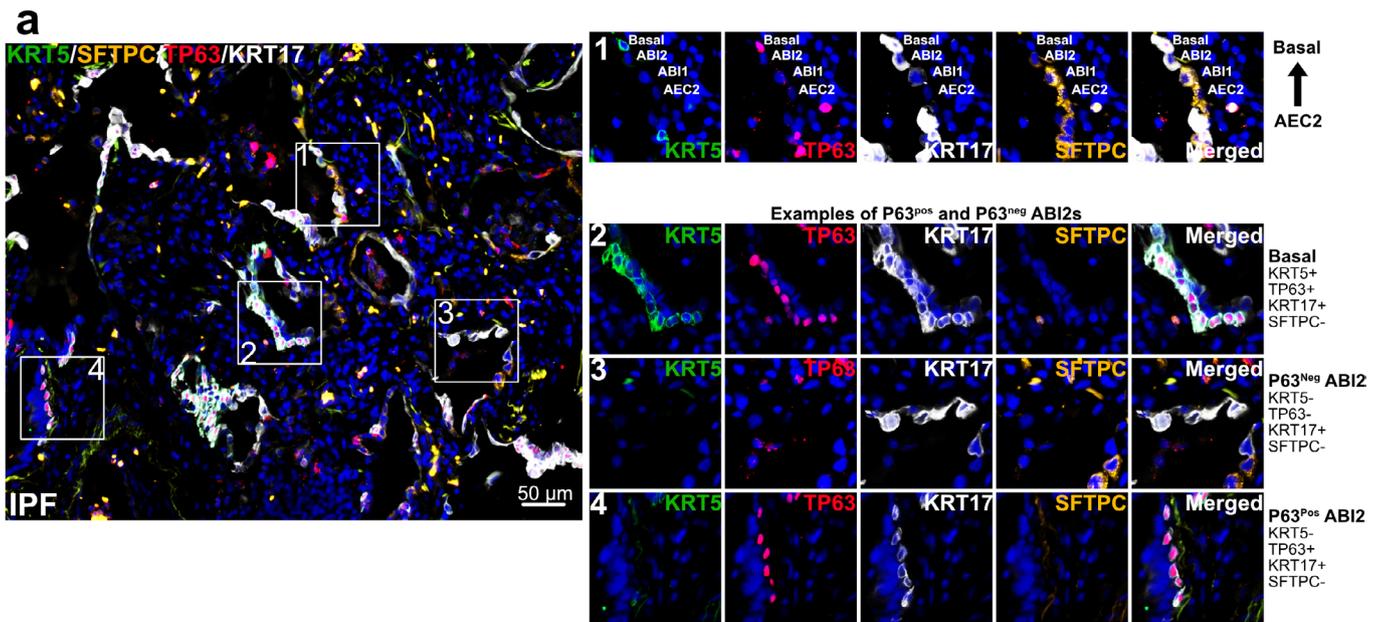


**Extended Data Fig. 5 | Comparison of hAEC2-derived and IPF basal cells.** (a) Feature plots show KRT5 and KRT17 expression in the merged basal cell data from fresh normal donors (N=2, marked as NL basal), fresh IPF patients (N=2, marked as IPF basal), and hAEC2+AHLM organoids (N=2). (b) Comparison of top 100 markers in SPB, MPB, AB, and PB clusters of this study and Carraro *et al.*<sup>3</sup> with statistical comparison for degree of gene overlap of two independent gene sets denoted by representation factor was done via <http://nemates.org> (see methods) (see methods). (c) UMAP plots of basal cells from IPF, normal donors, and hAEC2-derived organoids with contributing cell number in each cluster. (d) Comparison of top 500 differentially expressed genes in hAEC2-derived and IPF basal cells against normal lung basal cells with statistical comparison for degree of gene overlap of two independent gene sets denoted by representation factor was done via <http://nemates.org> (see methods). (e) Representative upregulated genes in top 50 of IPF and hAEC2-derived vs. normal lung basal cells. SPB=secretory primed basal, MPB=multipotent basal, AB=activated basal, PB=proliferating basal.



Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Expression of ABI markers in organoid and comparison with PATS-like cells in IPF.** **(a)** Heatmap of lineage marker expressions in different cellular subsets arising from hAEC2-derived organoids. **(b)** Expression along pseudotime trajectory of AEC2, ABI, and basal cell markers. **(c)** Featureplots of organoid scRNA-seq (D0-D21) show SCGB1A1+ club cells and FOXJ1+ ciliated cells. **(d)** IF staining of hAEC2+AHLM organoids (Day 21) show SCGB1A1+ club and rare TUBB4+ ciliated cells. Representative of n = 2 independent experiments. **(e)** Top 15 significantly differentially expressed genes between AEC2s, ABI1s, ABI2s, and basal cells show AEC2-like profile of ABI1s and basal-like profile of ABI2s in addition to high ECM related gene signature of ABI2s. **(f)** UMAP of hAEC2-derived organoids co-cultured with AHLM and IPF epithelium where transitional AEC2s and KRT17+/KRT5- cells as originally described by Habermann et al<sup>4</sup>, appear. The combination of transitional AEC2s and KRT17+/KRT5- cells were labelled as PATS-like cells by Kobayashi et al<sup>1</sup>. Feature plots comparing **(g)** hAEC2 lineage markers expression, **(h)** PATS-like markers expression, and **(i)** basal lineage marker expression in the UMAPs of organoid and IPF.



Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Identification of ABIs in normal, IPF, and COVID-19 lungs.** (a) Immunostaining of IPF lung with AEC2, ABIs, and basal markers show (1) linearly connected AEC2, ABI1, ABI2, and basal cell, (2) a cluster of basal cells (KRT17+/KRT5+/TP63+), and two types of ABI2s (KRT17+/KRT5-): (3) TP63<sup>neg</sup> and (4) TP63<sup>pos</sup>. Representative of n=3 independent IPF lungs. (b) In situ analysis hybridization of normal lung with *SFTPC* and *KRT17* identifies rare ABI1s (*SFTPC*+/*KRT17*+ ) as a fraction of total *SFTPC*+ cells. Data are represented as mean  $\pm$  SD, representative of n=2 biologically independent samples. (c) Immunostaining of COVID-19 lung shows presence of AEC2s, ABI1 and ABI2s, and basal cells in the same bronchiolized region. Representative of n=1 sample.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

STAR; CellRanger v2.1.1; BD FACSDiva v8.01; Zeiss Lumar V12; Invitrogen EVOS M5000; Zeiss Zen 3.1; Affinity Designer v1.10.1; R v3.5.1; Python v3.6; DEseq2, Sickel v1.33, STAR v2.7.9a

Data analysis

No custom codes were developed in this study.

For single cell RNA-seq data analysis the CellRanger v2.1.1 software with default settings was used for de-multiplexing, aligning reads with STAR software to mouse genome GRCh38, and counting unique molecular identifiers (UMIs). Downstream data analysis was conducted with SeuratV3.0 in Rv3.5.1. RNA velocity was calculated on integrated objects using scVelo packageV0.1.16 and RNA velocity in SeuratWrappers package in R. Differential gene expression was conducted using DEseq in R. Gene ontology enrichment analysis was performed using the PANTHER Overrepresentation test on the top differentially expressed genes. Image analysis and quantification was performed using Fiji 1.0 and Imaris 8.3.1. FlowJo v10 was used for flow cytometry analysis. GraphPad Prism 8.0 was used for all statistical testing.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The human bulk RNA-seq data reported in this paper is deposited in NCBI Gene Expression Omnibus (GEO) under the accession number GSE148992. The human

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For animal studies, sample size was determined based on the number of biological replicates available, which were strictly controlled for sex and age between experimental groups to minimize confounding factors in our analysis. No statistical test was performed to determine sample size. For other experiments, no prior calculations were performed to determine sample size. Samples size were determined by availability of tissue specimens (in case of disease lungs). Significance was calculated based on pairwise comparisons which required a minimum of three samples. We had at least three samples in all experiments.
Data exclusions	No data was excluded from the analysis presented in this paper.
Replication	All animal studies were independently performed twice with successful replication. Three or more biological replicates were used for every study. All in vitro experiments were performed at least two times to confirm reproducibility and all attempts at replications were successful.
Randomization	For animal studies, age/sex/litter-matched mice were randomized into groups based on the presence of specific genetic alleles. For studies using animals of identical genotypes, age/sex/litter-matched mice were randomly assigned to experimental groups using publicly available online "list randomizer" tool. Animals of both experimental groups were housed in the same cages when possible to minimize confounding factors. Samples receiving in vitro treatments were completely randomized.
Blinding	For RNA and histologic quantification, performer was blinded to specimen/genotype during data collection and analysis. For organoid size/efficiency quantification, the performer was blinded. For in silico analyses of scRNA- or bulkRNA-seq, the performer was not blinded because the coding required explicit labeling of different specimens.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used

The following primary antibodies were used for immunofluorescence: rabbit anti-SFTPC (Cat#AB3786; Millipore Sigma; used at 1:2000), mouse anti-SFTPC (Cat#sc-518029; Santa Cruz; used at 1:100), chicken anti-KRT5 (Cat#905901; BioLegend; used at 1:500), mouse anti-KRT14 (Cat#CBL197; Millipore Sigma; used at 1:200), mouse anti-human KRT17 (Cat#sc-393002; Santa Cruz; used at 1:100), rabbit anti-p63 (cat#13109; Cell Signaling Technology; used at 1:100), rat anti-KRT8 (cat#TROMA-I; DSHB; used at 0.045 µg/mL), goat anti-human LAMP3 (cat#AF0487; R&D; used at 1:200), rabbit anti-ABCA3 (Cat#ab99856; Abcam; used at 1:500), rabbit anti-human nuclear antigen (Cat#ab86129; Abcam; used at 1:500), mouse anti-MMP7 (Cat#MAB3315; Millipore Sigma; used at 1:200), mouse anti-beta IV Tubulin (Abcam; Cat# ab11315; used at 1:200), mouse anti-Ac-Tuba1a (Sigma; Cat# T7451, used at 1:500),

rabbit anti-p-SMAD1/5/8 (Cat#AB3848-I; Millipore Sigma; used 1:100),  
rat anti-SCGB1A1 (R&D; Cat#MAB4218; used at 1:500).

The following secondary antibodies were used at 1:250 for immunofluorescence:  
Donkey Anti-chicken IgY (IgG) (H+L) Alexa Fluor 647 (Jackson ImmunoResearch; Cat# 703-605-155; used at 1:250),  
Donkey anti-rabbit IgG (H+L) Alexa Fluor 555 (Thermo Fisher; Cat# A-31572; used at 1:250),  
Donkey anti-rabbit IgG (H+L) Alexa Fluor 488 (Thermo Fisher; Cat# A-21206; used at 1:250),  
Donkey anti-mouse IgG (H+L) Alexa Fluor 555 (Thermo Fisher; Cat# A-31570; used at 1:250),  
Donkey Anti-rat IgG (H+L) Alexa Fluor 488 (Thermo Fisher; Cat# A-21208; used at 1:250)  
Goat anti-mouse IgM Antibody Alexa Fluor 488 (Thermo Fisher; Cat# A-21042; used at 1:250)

For flow cytometry on mouse cells, the following antibodies were used:  
Rat anti-Mouse CD45-Alexa Fluor 700 (BD; Cat#560510; Lot#8338884)  
Rat anti-Mouse CD31-APC/Fire 750 (BioLegend; Cat#102528; Lot#B259777)  
Rat anti-Mouse CD326-BV421 (BD; Cat#563214; Lot#8228560)

For flow cytometry on human cells, the following antibodies were used:  
Biotinylated CD45 (clone 2D1; Cat#368534; BioLegend, used at 1:200),  
CD31 (clone WM-59; Cat#13-0319-80; eBioscience; used at 1:200),  
CD11b (clone ICRF44; Cat#301304; BioLegend; used at 1:200)  
mouse anti-human CD45 APC-Cy7 (BioLegend; Cat#304014; used at 1:200),  
mouse anti-human anti-CD11b-APC-Cy7 (BD; Cat#557754; used at 1:200),  
mouse anti-human CD31 APC-Cy7 (BioLegend; Cat#303120; used at 1:200),  
mouse anti-human CD326 PE (BioLegend; Cat#324206; used at 1:200)  
mouse IgM anti-human HTII-280 (Terrace Biotech, Cat#303118; used at 1:200)  
goat anti-mouse IgM Alexa Fluor 488 (Thermo Fisher; Cat# A-21042; used at 1:1000)  
rat anti-mouse CD45 (Alexa Fluor 700; Cat#560510; BD; used 1:200),  
rat anti-mouse CD31 (APC/Fire750; Cat#102528; BioLegend; used 1:200)  
DRAQ7 (Cat#7406S; Cell Signaling Technologies) was used to stain dead cells for flow cytometry

#### Validation

All antibodies have been validated by the manufacturer using specific assay for mouse or human cells and have multiple citations for the relevant cell type. Investigators independently validated antibodies by reproducing findings in published work. Primary and secondary antibody controls were used for immunofluorescence. FMO or IgG controls were used for flow cytometry.

## Eukaryotic cell lines

### Policy information about [cell lines](#)

#### Cell line source(s)

MRC5 cells Cat#ATCC-CCL-171 were directly obtained from ATCC.

#### Authentication

Cells were immediately used (within the first five passages) after purchasing from ATCC. They were not authenticated in the research laboratory.

#### Mycoplasma contamination

All cell lines when obtained from ATCC were screened for mycoplasma contamination by ATCC. No subsequent tests were performed after receiving cells from ATCC.

#### Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified lines were used in this manuscript.

## Animals and other organisms

### Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

#### Laboratory animals

Adult transgenic mice in C57BL/6 background were used at 6-12 weeks of age. The laboratory mouse strains used in this study are: R26-mTmG (The Jackson Laboratory; Stock No: 007576), SFTPC-CreERT2 (Chapman et al, 2011), and NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NOD scid gamma; NSG; The Jackson Laboratory, Stock No: 005557; Shultz et al 2005). For all strains, equal number of males and females were utilized in all experiments where applicable.

#### Wild animals

No wild animals were involved in this study.

#### Field-collected samples

No field-collected samples were involved in this study.

#### Ethics oversight

All animals were housed and treated in accordance with the Institutional Animal Care and Use Committee (IACUC) protocol approved at the University of California, San Francisco.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Scleroderma or IPF lungs were collected from patients with established relevant diagnoses undergoing lung transplantation, and the control healthy lungs were from sex and age-matched cadaveric donors without prior history of lung disease obtained from brain-dead donors that were rejected for lung transplantation. For histologic analysis, four independent IPF samples were used. Across the entire manuscript, a total of 16 IPF specimens with age range of 49-69 and sex distribution of 5 females/11 males were used. Detailed information is available in supplementary table 6.
Recruitment	We did not recruit donors specifically for this study. Participants did not receive compensation from the authors. Healthy human lungs were provided by the Northern California Donor Network from deceased donors in which the lung was independently rejected for transplantation. Disease lungs were obtained from end-stage patients receiving lung transplantation, and therefore exhibited severe to end-stage disease phenotype. Samples were taken from areas of severe injury. Investigators were blind to patient information besides relevant history of lung disease.
Ethics oversight	Studies involving human tissue were approved by the UCSF Institutional Review Board. Informed consent was received from each donor.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Human lung tissue was obtained and dissected into 5 cm <sup>3</sup> pieces. Tissue was washed two times in 500 ml sterile PBS and one time in HBSS for 10 min at room temperature. Using autoclaved Kim Wipes, tissue was compressed to remove as much liquid as possible and further dissected into 1 cm <sup>3</sup> pieces. Sterile HBSS buffer containing 15 U/ml Dispase II (Cat#17105041; Thermo fisher), 225 U/mL Collagenase Type I (Cat#17100017; Thermo Fisher), 100 U/mL Dnase I (Cat#DN25; Sigma-Aldrich), and 1% penicillin/streptomycin was added to the tissue pieces. Tissue was digested for 2 h at 37 °C and Fungizone (1:400) was added for the final 30 min of the digest. After digest the enzyme solution was diluted with an equal volume of sort buffer (DMEM without phenol red, 2% FBS and penicillin/streptomycin) and liquefied using an Osterizer 12 Speed Blender as follows: 5 s pulse, 5 s grate, and 5 s pulse. The suspension was poured through a glass funnel lined with sterile 4x4 gauze, applying some compression to recover as much of the solution as possible. The cell suspension was sequentially filtered through 100um, 70um and 40um strainers. Red blood cells were removed using Red Blood Cell Lysis Buffer (Sigma-Aldrich). Immune cells were depleted by resuspending 2.0 X 10 <sup>7</sup> cells/ml and FC blocked for 10 minutes rocking at 4 °C. Biotinylated CD45 (clone 2D1; Cat#368534; BioLegend, used at 1:200), CD31 (clone WM-59; Cat#13-0319-80; eBioscience; used at 1:200), and CD11b (clone ICRF44; Cat#301304; BioLegend; used at 1:200) were spiked into the cell suspension at 1:200 and rocked an additional 20 minutes. Streptavidin beads (Cat#17663; STEMCELL Technologies; used 1:50) were added at 25ul/ml to exclude immune and endothelial cells using MagRack 6 (GE Healthcare 89129-096). Cells were either frozen for future use or stained for flow cytometry. The following FACS antibodies were used at 1:200: mouse anti-human CD45 APC-Cy7 (BioLegend; Cat#304014; used at 1:200), mouse anti-human anti-CD11b-APC-Cy7 (BD; Cat#557754; used at 1:200), mouse anti-human CD31 APC-Cy7 (BioLegend; Cat#303120; used at 1:200), mouse anti-human CD326 PE (BioLegend; Cat#324206; used at 1:200), and mouse IgM anti-human HTII-280 (Terrace Biotech, Cat#303118; used at 1:200). After 30 min incubation, goat anti-mouse IgM Alexa Fluor 488 (Thermo Fisher; Cat# A-21042; used at 1:1000) was added to label HTII-280+ cells. Immune and endothelial cells were further excluded on FACS by gating CD45-/CD11b-/CD31-. Human AEC2s were sorted for EpCAM+/HTII-280+ cells, and adult human lung mesenchymal cells (AHLM) were sorted for CD45-/CD11b-/CD31-/EpCAM- cells. Mouse lungs were harvested from adult animals and digested in a cocktail of Collagenase Type I (Cat#17100017; Thermo Fisher; used 225 U/mL), Dispase II (Cat#17105041; Thermo fisher; used 15 U/mL) and Dnase I (Cat#DN25; Sigma-Aldrich; used 50 U/mL). Single cell suspension were prepared by sequential filtering the homogenate through 100um, 70um, and 40um strainers before proceeding with antibody staining and flow cytometry.
Instrument	All data was collected on a BD FACSAria II.
Software	Data was collected using BD FACSDiva software. Data was analyzed using FlowJo v10.
Cell population abundance	The sorted HTII-280+ cells were approximately 85% of total live/EpCAM+ cells in normal human lung and ~10-50% of total live/EpCAM+ cells in IPF lungs. Human mesenchymal cells were approximately 10% of all live cells. Small portions of sorted cells were run through the machine to confirm purity of over 85%. Immunofluorescent staining of sorted cells were also performed to confirm expression of expected markers.
Gating strategy	Cells were selected using FSC/SSC scatter profile to remove debris and doublets were excluded using FSC-A/FSC-H to select for

Gating strategy

single cells. Dead cells were removed using DRAQ7 or DAPI exclusion. Mesenchymal cells from normal donor human lungs were isolated using the gating strategy: live/CD45-/CD31-/EpCAM-. HTII-280+ AEC2s from normal donor human lungs were isolated using the gating strategy: live/CD45-/CD31-/EpCAM+/HTII-280+. These strategies are included in the methods section.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.