Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor

William J. Zacharias1,2, David B. Frank2,3,4,†, Jarod A. Zepp1,2, Michael P. Morley1,2,4, Farrah A. Alkhaleel1,2, Jun Kong1,2, Su Zhou1,4, Edward Cantu5 & Edward E. Morrissey1,2,4,6,7

Functional tissue regeneration is required for the restoration of normal organ homeostasis after severe injury. Some organs, such as the intestine, harbour active stem cells throughout homeostasis and regeneration1; more quiescent organs, such as the lung, often contain facultative progenitor cells that are recruited after injury to participate in regeneration2,3. Here we show that a Wnt-responsive alveolar epithelial progenitor (AEP) lineage within the alveolar type 2 cell population acts as a major facultative progenitor cell in the distal lung. AEPs are a stable lineage during alveolar homeostasis but expand rapidly to regenerate a large proportion of the alveolar epithelium after acute lung injury. AEPs exhibit a distinct transcriptome, epigenome and functional phenotype and respond specifically to Wnt and Fgf signalling. In contrast to other proposed lung progenitor cells, human AEPs can be directly isolated by expression of the conserved cell surface marker TM4SF1, and act as functional human alveolar epithelial progenitor cells in 3D organoids. Our results identify the AEP lineage as an evolutionarily conserved alveolar progenitor that represents a new target for human lung regeneration strategies.

Wnt signalling, which is revealed by expression of the direct target gene Axin2, has previously been shown to have an important role in the development of both the surfactant-producing alveolar type 2 (AT2) cells and the alveolar type 1 (AT1) cells that form the gas-exchange surface of the lung alveolus4. In the lungs of adult Axin2creERT2-tdT;R26ReYFP (R26RɛYFP (R26RɛYFP is also known as Gt(ROSA)26Sortm1(ɛYFP)Cgn), mice, Axin2+ Wnt-responsive epithelial cells are restricted to the alveolar region and express the AT2 cell marker Sftpce (Fig. 1a–d and Extended Data Fig. 1a–c). Few Axin2+ cells express AT1 markers, including Hopx (Fig. 1e and Extended Data Fig. 1k, l). These Axin2+ AT2 cells represent an alveolar epithelial progenitor lineage—hereafter referred to as AEPs—that comprises approximately 20% of adult AT2 cells (Fig. 1f). AEPs express the same level of AT2 marker genes as other AT2 cells (Extended Data Fig. 1f), and also show enriched expression of Wnt targets (Extended Data Fig. 1g). We performed one-, three- and nine-month lineage tracing using Axin2creERT2-tdT;R26ReYFP mice to define AEP dynamics during adult homeostasis (Fig. 1a). AEPs are notably stable, and we observed only a small increase in the number of AEP-marked cells after nine months (Fig. 1g and Extended Data Fig. 2a–c). In contrast to developmental alveologenesis4 (Extended Data Fig. 3), few Axin2+ AT2 cells become AEPs during homeostasis (Fig. 1h).

To assess dynamics of AEPs in lung injury, we used the H1N1 influenza virus to injure the lungs of adult Axin2creERT2-tdT;R26ReYFP mice, causing spatially heterogeneous injury that is similar to human influenza infection5. We defined four regions of injury severity: (i) zone 1, no morphological changes; (ii) zone 2, a minor injury with mild interstitial thickening; (iii), zone 3, substantial injury; and (iv) zone 4, total alveolar destruction (Fig. 1i). We used this spatially specific response to analyse the contribution of AEPs to lung regeneration.

Recent studies have shown that Sox2-derived Krt5+ epithelial cells migrate to damaged distal lung regions to re-create an epithelial barrier6–10. We observe Krt5+ epithelium specifically in zone 4 after influenza infection (Extended Data Fig. 4a–d, f), but lineage tracing demonstrates that no Krt5+ cells are derived from AEPs (Extended Data Fig. 4g). Furthermore, AEPs express minimal levels of Krt5 or Sox2 RNA and no detectable protein (Extended Data Figs 1f, 4e), further indicating that AEPs and Krt5+ cells derive from distinct lineages. In zone 4, Sftpce+ and Krt5+Sftpce+ cells are very rare (Extended Data Fig. 4i), which confirms previous reports5 that the Krt5+ lineage cells do not efficiently regenerate Sftpce+ cells except after forced Wnt activation6.

One month after influenza injury, AEPs and their progeny are present at homeostatic levels in zone 1. However, in zones 2 and 3 the number of AT2 cells expands significantly (Extended Data Fig. 4b)11,12, with a large increase in the percentage of AT1 and AT2 cells arising from the AEP lineage (Fig. 1j–l and Extended Data Figs 2d–i, 4j–l). This robust labelling is independent of the timing of tamoxifen injection before influenza infection (Extended Data Fig. 5g–h). Notably, in zone 2 and zone 3 the AEP lineage shows a marked and specific increase in proliferation (Fig. 1m and Extended Data Fig. 5k–o). Three months after injury, within 300 microns of a persistent Krt5+ pod, a majority of AT2 cells and many AT1 cells in regenerated alveoli are derived from the AEP lineage (Fig. 1o–q). Immunohistochemistry and fluorescence-activated cell sorting (FACS) analysis after influenza injury demonstrate that AEPs self-renew to maintain the AEP lineage and generate a large number of new lineage-traced alveolar epithelial progeny (Fig. 1n and Extended Data Figs 2j, 5a–e). Notably, few non-AEP AT2 cells acquire the AEP phenotype even in the setting of considerable lung injury (Fig. 1n and Extended Data Fig. 5e).

AEPs exhibit a distinct gene expression profile enriched in lung developmental genes (Fig. 2a–d), including the key genes Fgffr2, Nkx2-1, Id2, Etv4, Etv5 and Foxa1 (Extended Data Fig. 6 and Supplementary Table 1). Furthermore, analysis by assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Extended Data Fig. 7) revealed a marked difference between AEPs and AT2 cells, with more than 40% of the genome containing differential open chromatin (Fig. 2a). Although many regions of common open chromatin are found near housekeeping genes, regions of AEP-enriched open chromatin are found near lung development genes (Extended Data Fig. 7c). DNA binding-site motif analysis shows that AEP-enriched chromatin contains binding sites for AEP-enriched transcription factors of the Klf, Six, Sox, Nkx2 and E11/Ets families (Extended Data Fig. 7d, e), all of which are known to be regulators of progenitor cell behaviour13–17. Moreover, a group of primed cell-cycle regulators near AEP-enriched open chromatin were

*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Figure 1 | Identification of an Axin2\(^+\) AEP in the adult lung that regenerates a substantial percentage of the alveolar epithelium.

a, Schematic of Axin2\(^{creERT2-tdT,R26R:eYFP}\) mice. Lineage tracing experimental design is as indicated. D, day; mo, month. b–d, Axin2 marks a subset of AT2 cells. eYFP is detected by an anti-GFP antibody. Unmarked, white arrowheads. AEP-marked, yellow arrowheads; marks a subset of AT2 cells. eYFP is detected by an anti-GFP antibody. Experimental design is as indicated. D, day; mo, month.

j, k, l, m, n, Experimental design is as indicated. D, day; mo, month.

To isolate human AEPs, we identified cell surface markers that were enriched in mouse AEPs (Fig. 3a). These studies identified the epithelial cancer stem cell membrane protein Tm4sf1\(^{22,23}\) as a marker for mouse AEPs (Fig. 3b and Extended Data Fig. 8a–c). Immunohistochemistry and FACS analysis demonstrates that Tm4sf1 marks approximately 20% of labelled mouse AT2 cells and dynamically regulated in AEPs two weeks after influenza infection\(^{18–21}\) (Fig. 2b–d and Extended Data Fig. 6e–g).

To isolate human AEPs, we identified cell surface markers that were enriched in mouse AEPs (Fig. 3a). These studies identified the epithelial cancer stem cell membrane protein Tm4sf1\(^{22,23}\) as a marker for mouse AEPs (Fig. 3b and Extended Data Fig. 8a–c). Immunohistochemistry and FACS analysis demonstrates that Tm4sf1 marks approximately 20% of labelled mouse AT2 cells and
more than 90% of mouse AEPs (Fig. 3c, d and Extended Data Fig. 8a). Using a combination of a human TM4SF1 antibody (Extended Data Fig. 8d) and human-AT2-specific HTII-280 antibody24 (Extended Data Fig. 8b, e–h), we were able to identify a distinct subset of HTII-280®TM4SF1®EPICM® putative human AEPs in normal human lung. These human AEPs comprise approximately 29% of the human AT2 population (Fig. 3e) and express SFTPc, but not KRT5 or SOX2, mRNA (Supplementary Table 2).

Using clonal alveolar organoid assays25, both mouse AEPs and human AEPs form a greater number of, and larger, organoids than both AT1 and AT2 cells but no SOX2⁺ or KRT5⁺ cells (Extended Data Fig. 8i, j). AEPs also demonstrate increased responsiveness to Wnt modulation when compared to AT2 cells (Fig. 3f–n and Extended Data Fig. 9). Notably, depletion of TM4SF1⁺ cells from the human AT2 population leads to a dramatic loss of organoid formation (Fig. 3o–s). Wnt inhibition promoted AT1 cell differentiation and Wnt activation promoted AT2 cell formation in both mouse and human organoids, but not in human AEP-depleted organoids (Fig. 3o–r, u, v and Extended Data Fig. 9o, p). These data suggest that TM4SF1⁺HTII-280⁺ human AEPs are the functional equivalent of mouse AEPs.

RNA sequencing analysis (RNA-seq) demonstrated that a large proportion of human AEP-enriched genes (35.6%)—including key progenitor cell regulators—were evolutionarily conserved with mouse AEPs (Fig. 4a, b and Extended Data Fig. 10a, b). In particular, mouse and human AEPs are both enriched for Wnt pathway targets, including AXIN2 and FGFR2; FGFR2 is the primary receptor for FGF7 and FGF1026–32 (Fig. 4c, Extended Data Fig. 10k and Supplementary Table 2). DNA binding motif analysis shows that a majority of these primed genes are associated with cell-cycle activation. Reg., regulation. For full details of the experimental design and statistical methods for these analyses, see Methods.
Importantly, the conservation and accessibility of both mouse and human AEPs provides an opportunity for mechanistic studies to shed light on human lung progenitor cell biology, and assist in the development of new treatments for acute and chronic lung diseases.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 April 2017; accepted 24 January 2018.
Published online 28 February 2018.
2. Stanger, B. Z. Probing hepatocyte heterogeneity. 
7. Vaughan, A. E. 
6. Kumar, P. A. 
12. Rock, J. R. 
11. Barkauskas, C. E. 
4. Frank, D. B. 
3. Afelik, S. & Rovira, M. Pancreatic 
Extended Data Fig. 10. 
treatment. Dep., AEP-depleted AT2 cells. Additional data are shown in 
organoid assays show that mouse AEPs ( human only, 
and Fgf signalling. a 
Figure 4 

Figure 4 | AEPs display an evolutionarily conserved response to Wnt and Fgf signalling. a, Human AEPs exhibit a distinct transcriptome enriched for Wnt responsiveness. b, More than one third of human AEP-enriched genes are shared with mouse AEPs. c, Volcano plot of 15,628 genes tested using limma shows extensive overlap between upregulated genes in mouse and human AEPs. FGF2R2 is indicated. d–w, Alveolar organoid assays show that mouse AEPs (d–m) and human AEPs (n–w) display a significant increase in colony formation and size upon FGF7 treatment. Dep., AEP-depleted AT2 cells. Additional data are shown in Extended Data Fig. 10. a–n, 4 individual organoid experiments. Statistics are inclusive of all biological replicates. ***P < 0.001 and **P < 0.01 by ANOVA with adjustment for multiple comparison testing. NS, not significant. Plots are centred on mean with bars indicating standard deviation. Scale bars: 25 μm.

1. Beumer, J. & Clevers, H. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. Development 143, 3639–3649 (2016).
2. Stanger, B. Z. Probing hepatocyte heterogeneity. Cell Res. 25, 1181–1182 (2015).
3. Afelik, S. & Rovira, M. Pancreatic islet-cell regeneration: facultative or dedicated progenitors? Mol. Cell. Endocrinol. 445, 85–94 (2017).
4. Frank, D. B. et al. Emergence of a wave of Wnt signaling that regulates lung alveologenesis by controlling epithelial self-renewal and differentiation. Cell Reports 17, 2312–2325 (2016).
5. Töpfer, L. et al. Influenza A (H1N1) vs non-H1N1 ARDS: analysis of clinical course. J Crit. Care 29, 340–346 (2014).
6. 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).
7. Vaughan, A. E. et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. Nature 517, 621–625 (2015).
8. Zuo, W. et al. p63+Krt5+ distal airway stem cells are essential for lung regeneration. Nature 517, 616–620 (2015).
9. Xi, Y. et al. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. Nat. Cell Biol. 19, 904–914 (2017).
10. Ray, S. et al. Rare Sox2+ airway progenitor cells generate KRT5+ cells that repopulate damaged alveolar parenchyma following influenza virus infection. Stem Cell Reports 7, 817–825 (2016).
11. Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. J. Clin. Investig. 123, 3025–3036 (2013).
12. 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).
13. El-Hashash, A. H. et al. Six1 transcription factor is critical for coordination of epithelial, mesenchymal and vascular morphogenesis in the mammalian lung. Dev. Biol. 353, 242–258 (2011).
14. Herriges, J. C. et al. FGF-regulated ETV transcription factors control FGF–SHH feedback loop in lung branching. Dev. Cell 35, 322–332 (2015).
15. Kherrouche, Z. et al. PEA3 transcription factors are downstream effectors of Met signaling involving in migration and invasiveness of Met-addicted tumor cells. Mol. Oncol. 9, 1852–1867 (2015).
16. Rockich, B. E. et al. Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. Proc. Natl Acad. Sci. USA 110, E4456–E4464 (2013).
17. Wan, H. et al. Kruppel-like factor 5 is required for perinatal lung morphogenesis and function. Development 135, 2563–2572 (2008).
18. Bogunovic, M. et al. Origin of the lamina propria dendritic cell network. Immunity 31, 513–525 (2009).
19. Wu, L. et al. Mat1-Gld modulated Cak activity regulates cell cycle G1: exit. Mol. Cell. Biol. 21, 260–270 (2001).
20. Lin, D. et al. Constitutive expression of B-myb can bypass p53-induced Waf1/Cip1-mediated G1 arrest. Proc. Natl Acad. Sci. USA 91, 10079–10083 (1994).
21. Schmidt-Eckardt, U., Daniel, G., Hoffmann, A. & Spengler, D. Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. Mol. Cell. Biol. 34, 1020–1030 (2014).
22. Wan, H., Schievchen, G. L., Hellström, I., Hellström, K. E. & Aruffo, A. Cloning and expression of the tumor-associated antigen L6. Proc. Natl Acad. Sci. USA 89, 3503–3507 (1992).
23. Gao, H. et al. Multi-organ site metastatic reactivation mediated by non-canonical discoidin domain receptor 1 signaling. Cell 166, 47–62 (2016).
24. Gonzalez, R. F., Allen, L., Gonzales, L., Ballard, P. L. & Dobbs, L. G. HTII-280, a bimarker specific to the apical plasma membrane of human lung alveolar type II cells. J. Histochem. Cytochem. 56, 891–901 (2010).
25. Barkauskas, C. E. et al. Lung organoids: current uses and future promise. Development 144, 986–997 (2017).
26. Shu, W. et al. Wnt/Li-catenin signaling acts upstream of N-myc, BMP4, and FGF signaling to regulate proximal-distal patterning in the lung. Dev. Biol. 283, 226–239 (2005).
27. Zhang, X. et al. Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. J. Biol. Chem. 281, 15694–15700 (2006).
28. Yano, T. et al. KGF regulates pulmonary epithelial proliferation and surfactant protein gene expression in adult rat lung. Am. J. Physiol. Lung Cell. Mol. Physiol. 279, L1146–L1158 (2000).
29. Yano, T., Deterting, R. R., Simonet, W. S., Shannon, J. M. & Mason, R. J. Keratinocyte growth factor reduces lung damage due to acid instillation in rats. Am. J. Respir. Cell Mol. Biol. 15, 433–442 (1996).
30. Panos, R. J., Rubin, J. S., Caskey, G. K., Aaronson, S. A. & Mason, R. J. Keratinocyte growth factor and hepatocyte growth factor/scatter factor are heparin-binding growth factors for alveolar type II cells in fibrillar-conditioned medium. J. Clin. Investig. 92, 969–977 (1993).
31. Quin qui, J. et al. Influenza virus infects epithelial stem/progenitor cells of the distal lung: impact on Fgf2b-driven epithelial repair. PLoS Pathog. 12, e1005544 (2016).
32. Nikolaidis, N. M. et al. Mitogenic stimulation accelerates influenza-induced mortality by increasing susceptibility of alveolar type II cells to infection. Proc. Natl Acad. Sci. USA 114, E6613–E6622 (2017).
33. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature 515, 355–364 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants from the National Institutes of Health (T32-HL007586 to W.J.Z.; T32-HL007915, K12-HD043245 to D.B.F., T32-HL007843 to J.A.Z. and HL110942, HL087825, HL132999, HL129478, HL134745 to E.E.M.). We thank the Flow Cytometry Core Laboratory of Children’s Hospital of Philadelphia and the CIVI Histology Core, Next Generation Sequencing Core and CDB Microscopy Core at the University of Pennsylvania for technical assistance.

Author Contributions W.J.Z., D.B.F., J.A.Z., F.A.A., S.Z. and J.K. performed the experiments. W.J.Z., D.B.F., J.A.Z., M.P.M. and E.E.M. analysed the data. E.C. provided access to human samples and assisted W.J.Z. with all human experiments. E.E.M. supervised the project. W.J.Z. wrote the first draft of the manuscript. All authors contributed to the writing of the final manuscript.

Author Information Reprints and permissions information is available at http://www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.E.M. (emorrise@pennmedicine.upenn.edu).

Reviewer Information Nature thanks C. Dean and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
METHODS
Ethical compliance. All mouse studies were performed under guidance of the University of Pennsylvania Institutional Animal Care and Use Committee in accordance with institutional and regulatory guidelines. This study used cells derived from de-identified non-used lungs donated for organ transplantation via an established protocol (PROPEL, approved by University of Pennsylvania Institutional Review Board) with informed consent in accordance with institutional and NIH procedures. All patient information was removed before use. This use does not meet the current NIH definition of human subject research, but all institutional and regulatory guidelines for human subject research were followed throughout the reported experiments.

Mice and Cre recombinase induction. The generation and genotyping of the Axin2creERT2flrt mouse line generated in our laboratory has previously been described. The SfpcreERT2 mouse line was a gift of H. Chapman; their genotyping and generation have previously been described. HoxpTgFPGFP mice were a gift of R. Jain and J. Epstein, and are available from Jackson Laboratories. The R26R-TdTomato mice are available from Jackson Laboratories. All mouse studies were performed under guidance of the University of Pennsylvania Institutional Animal Care and Use Committee. Mice were maintained on a mixed CD-1 and C57BL/6 background. For induction of all Cre recombinase models, tamoxifen (Sigma) was dissolved in 100% ethanol and diluted with corn oil (Sigma) to produce a 10% ethanol:tamoxifen:corn oil mixture at 20 mg/ml. Six-to-eight-week-old mice were injected intraperitoneally with 200 μg per g body weight on between three and five consecutive days to induce recombination. All lineage-tracing experiments represent a minimum of n = 6 mice in all groups to allow for effective statistical evaluation. qPCR experiments represent a minimum of n = 3 mice in all groups. Mouse experiments were performed on both male and female mice in all conditions, and mice were chosen at random from the cohort but not formally randomized. Randomization between experimental conditions was not possible owing to the nature of the injury experiments.

Influenza lung injury. PR8 H1N1 influenza was a gift of J. Wherry. Reformation for lineage tracing was performed using 3 daily tamoxifen injections, 7 or 28 days before viral infection. For infection, the virus was diluted in PBS and a dose of 0.3 LD50 was administered via intranasal instillation. After infection, mice were weighed and monitored daily for 14–28 days, and mice that lost >30% of their starting weight or were moribund were euthanized humanely. Post-influenza RNA was obtained at 14 days after infection, and lung regeneration was analysed from tissue collected from mice between 28 days and 3 months after infection. FACS data were generated from influenza-infected and uninfected mice using the same protocols, as described later. Regionalized lung injury was assessed via histology, and adjacent sections were used for all immunostaining and quantification.

Histology. At the time of tissue collection, mice were euthanized by CO2 inhalation. The chest cavity was exposed and the lungs cleared of blood by perfusion with cold PBS via the right ventricle. Lungs were inflated with 2% paraformaldehyde under constant pressure of 30 cm water, and allowed to fix overnight. The lungs were removed and deemed to be low-expressed. These data were transformed using the negative binomial distribution in the R package edgeR, and genes with a CPM < 1 in 25% of samples were adjusted for background. For induction of all Cre recombinase models, tamoxifen (Sigma) and CHIR99021 1 μM (Fisher) were used. DMSO was used as a control. The ROCK inhibitor Y27632 (Sigma) was used at 1 μM.

Gene expression. Using the following antibodies on paraffin sections: GFP (chicken, BioGenex, MU178-UC, 1:20), Sftpc (rabbit, Millipore, ABC99, 1:250), Sftpc (goat, Santa Cruz, sc-37972, 1:200), and TM4SF1 (rabbit, LSBiosciences, LS-C109221, 1:25) and rabbit IgG isotype control (LS Biosciences, LS-C109221, 1:25) with donkey anti-rabbit 488 secondary (Life Technologies, A212016, 1:200). B7077, 1:25) and rabbit IgG isotype control (LS Biosciences, LS-C109221, 1:25) with donkey anti-rabbit 488 secondary (Life Technologies, A212016, 1:200). Human samples of normal, de-identified human lungs were obtained as non-used lungs donated for organ transplantation via an established protocol (PROPEL, approved by University of Pennsylvania Institutional Review Board) with informed consent in accordance with institutional procedures. A 2 × 2 cm piece of distal lung tissue was obtained, pleura and large airways were carefully dissected away and tissue was processed into a single-cell suspension using the same combination of dispase, collagenase I and DNase, as previously described. The Axin2creERT2fMc mice were isolated from lungs of 6–8-week-old SfpcreERT2R26R-YFP mice 5 days after induction with 200 μg per g body weight tamoxifen. eYFP+ cells were then isolated via FACS sorting, as previously described. For sorting and quantification, the following antibodies were used: Pdpn-eFluor660 (BioSource, Clone G8.8, 1:100), EpCAM-APC (BioSource, Clone G8.8, 1:200), EpCAM-eFluor488 (BioSource, Clone G8.8, 1:200), CD31-PECy7 (BioSource, Clone 390, 1:200) and CD45-PECy7 (BioSource, Clone 30-F11, 1:200). Two anti-mouse Tim4 antibodies were used to ensure specificity: sheep anti-mouse Tim4 (R&D systems, SF61, 1:100) and sheep IgG isotype control (R&D systems, 5-001-A, 1:100) with anti-sheep 488 secondary (Jackson ImmunoResearch). Library premixes, B7077, 1:25) and rabbit IgG isotype control (LS Biosciences, LS-C109221, 1:25)

RNAseq analysis. Cells were sorted using the protocols described earlier into Trizol LS (Life Technologies). For mouse, six individual mice were sorted and pooled into two individual pools for Axin2creERT2fMc and three individual pools for SfpcreERT2flrt. For human, cells from three individual patients were sorted separately and prepared for sequencing individually. RNA was then extracted using a combination of the Trizol protocol and MinElute RNA Cleanup Kit (Qiagen). RNA integrity was confirmed via Bioanalyzer evaluation and samples with RNA integrity number > 6 were used. High-quality nanoRNA-Seq libraries were prepared using the Illumina TruSeq Stranded mRNA library preparation protocol. Library quality was monitored using the Agilent Bioanalyzer. High-quality libraries were sequenced on the Illumina HiSeq 2500 using 75-bp paired-end reads. Sequenced reads were aligned to the mm9 or human reference (hg19/hGRC37) genome using the STAR aligner. Gene expression data were generated from influenza-infected and uninfected mice using the same protocols, as described earlier. For induction of all Cre recombinase models, tamoxifen (Sigma) and CHIR99021 1 μM (Fisher). DMSO was used as a control. The ROCK inhibitor Y27632 (Sigma) was used at 1 μM.

Data availability. Ex vivo alveolar organoids. Clonal alveolar organoid assays were performed as previously described, with some modifications from the original protocol. In brief, 5 × 103 epithelial cells (AT2 or AEP for mouse) were seeded at 2800-28000, HT2-280-7TM4SF1+ and HT2-280-7TM4SF1+ for human) were isolated as described earlier, and mixed with 5 × 104 lung fibroblasts (isolated from adult wild-type mice as previously described) for mouse; MRC5 cells (ATCC CCL-171, tested negative for mycobacterial contamination, at no greater than passage 10) for human. Cells were then suspended in a 1:1 mixture of SAGM medium (Lonza, with all additives except epinephrine) and growth factor-reduced, phenol-free Matrigel (Corning). Ninety microliters of the cell–medium–Matrigel mixture was then aliquoted into individual 24-well cell culture inserts and allowed to solidify at 37 °C. SAGM was then placed into each well of the 24-well plate. The Rock inhibitor Y27632 (Sigma) was included in the medium for the first two days. After two days of culture, Y27632 was removed and ligand treatments of organoids were performed using the following reagents at the indicated concentrations: Wnt3a 200 ng/ml (R&D systems), Fgf7 50 ng/ml (R&D Systems), XAV939 10 μM (Sigma) and CHIR99021 1 μM (Fisher). DMSO was used a control. The medium was changed every 48 h, and fresh ligands were included at each medium change. Fgf7, XAV939 and CHIR99021 were fixed in 2% paraformaldehyde, embedded in Histogel (Richard-Allan), dehydrated, paraffin-embedded, and sectioned and immunostained as described earlier.

References: 1. Scadden DT, et al. Nat Rev Mol Cell Biol 2013; 14:11–22. 2. Walmsley JS, et al. Stem Cells Dev 2017; 26:235–45. 3. Tuszynski MH, et al. Nat Neurosci 2012; 15:710–718. 4. Cui H, et al. Nat Commun 2017; 8:1363. 5. Cui H, et al. Nat Methods 2015; 12:1009–1012. 6. Chen Y, et al. Cell Stem Cell 2015; 16:179–190. 7. Liu X, et al. Cell Stem Cell 2015; 16:161–178. 8. Zhang P, et al. Cell Stem Cell 2015; 16:191–201. 9. Majors B, et al. Cell Stem Cell 2015; 16:132–145.
multiple comparisons using Benjamini–Hochberg procedure. For the human data, a paired analysis was employed using the patient as a blocking variable. Heat maps and principal component analysis plots were generated in R. Gene Ontology (GO) enrichment analysis was performed using the ToppGene Suite (http://toppgene.cchmc.org/)40.

**ATAC-seq analysis.** Individual ATAC-seq libraries were generated from sorted Axin2+ and Sftpc− AT2 cells as described earlier, using previously published methods41. In brief, 5 × 10^5 cells were sorted into PBS, washed and lysed to obtain nuclei. Nuclei were exposed to TN5 transposase (Illumina), and fractionated DNA was used for amplification and library preparation. Libraries were then purified and paired-end sequenced. Following sequencing, FASTQ files were aligned against the mouse reference genome (mm9) using the STAR aligner38, with default parameters plus options to suppress the matching of spliced reads (–outFilterMatchNminOverLread 0.4–outFilterScoreMinOverLread 0.4). Duplicate reads were flagged using the MarkDuplicates program from Picard tools and removed using samtools. MACS242 was used to call peaks from the following options –nomodel–shift -100–etextsize 200. Differential ATAC-seq peaks were determined using the beddiff command from MACS2 and default options. Peaks were filtered to have a MACS2 log10 likelihood ratio score > 10 and to be within −50 kb and 10 kb of the transcription start site of Ensembl 67 protein-coding genes. ATAC-seq enrichment heat maps were created using deepTools243.

**Motif analysis.** The intersection of gene promoter regions (−5 kb, 600 bp, Ensembl v.67) with identified ATAC-seq peaks was performed using bedtools. The FASTA file of genome sequence (mm9) of promoter ATAC-seq peaks was created using bedtools and scanned for TCF/LEF motifs using FIMO44. Motif enrichment analysis was performed using the findMotifsGenome.pl program in the HOMER software suite45, with the peak search size option set to 50 bp.

**Chromatin immunoprecipitation.** Chromatin immunoprecipitation was performed using the High Sensitivity ChIP Kit (Abcam) with 3 μg of anti-β-catenin (Santa Cruz sc-7963) or anti-IgG1 isotype control (Santa Cruz sc-3877). In brief, 1 × 10^5 Axin2+ or Sftpc− AT2 cells were sorted into SAGM (Lonza), whole chromatin was prepared, chromatin was cross-linked and sonicated using a Covaris sonicator to an optimal size of 300 bp, and chromatin was immunoprecipitated using the antibodies listed earlier, following the Abcam protocol. Library quality was confirmed using Bioanalyzer, and the enrichment of genomic DNA was assessed using qPCR to compare β-catenin immunoprecipitate to immunoprecipitate from an IgG control for each cell type. qPCR data represents n = 2 individual immunoprecipitation experiments, and was performed in triplicate.

**Statistical analysis.** No statistical methods were used to predetermine sample size. Statistical analysis was performed in Prism for Mac, and R. A two-tailed Student's t-test was used for the comparison between two experimental groups. For experiments with more than two groups, an ANOVA was performed followed by planned contrasts; pairwise comparisons and P-value adjustments for multiple comparisons were performed using Dunnett's procedure. The generation of odds ratios for the distribution of ATAC regions near genes was evaluated using Fisher's exact test and contingency table analysis. Statistical data were considered significant when P < 0.05. Centre values of all plots represent means and error bars represent standard deviations, with the exception of error bars for odds ratios (which represent confidence intervals).

**Data availability.** ATAC-seq and RNA-seq sequencing data generated during this study have been deposited in the Gene Expression Omnibus database with the primary accession GSE97055. All upregulated and downregulated genes identified during the RNA-seq experiments described in this paper are found in Supplementary Table 1 (mouse data) or Supplementary Table 2 (human data). Source Data for all plots in all figures are available online. The detailed protocol for the cell isolation and propagation of human AEPs is available on the Protocol Exchange46. All other datasets generated during and/or analysed in the current study are available from the corresponding author on request.
Extended Data Figure 1 | Location of Axin2\(^{+}\) epithelial cells within the adult mouse lung. a, Low-power view of the lung showing that E-cadherin\(^{+}\) Axin2\(^{+}\) epithelial cells are found only in the alveolar region, and not in the airway of the lung. b, c, Immunohistochemistry for ciliated (b) and secretory (c) markers shows no evidence of Axin2-lineage labelled cells co-expressing either of these markers. d, e, Quantification of the location of Axin2\(^{+}\) epithelial cell distribution in the lung. f, qPCR showing that Axin2\(^{+}\) AEPs and AT2 cells express similar levels of AT2 markers and other lung epithelial cell markers. AEPs express slightly higher levels of Abca3. g, AEPs express increased levels of Wnt signalling pathway components and targets by qPCR. h–j, Cytopsins and quantification demonstrating that the majority of sorted Axin2\(^{+}\) epithelial cells are Sftpc\(^{+}\). k, l, FACS analysis of Axin2\(_{tdT}\) positive, Hopx\(_{EYFP}\) mice demonstrating that few Axin2\(^{+}\) epithelial cells express Hopx, consistent with the immunohistochemistry data shown in Fig. 1. Data in this figure represent \(n = 3\) (k, l), 4 (d–j) or 10 (all other panels) mice from three individual experiments. Statistics are representative of all biological replicates. All data are shown as centred on mean with bars indicating standard deviation. *\(P < 0.05\), **\(P < 0.01\) by two-tailed \(t\)-test (f, g) or ANOVA with preplanned pairwise comparisons and adjustment for multiple comparison testing (d). Scale bars: a–c, 100\(\mu\)m; h, i, 25\(\mu\)m.
Extended Data Figure 2 | Characterization of Axin2++ Wnt responsive cells in the adult lung. a, Lineage tracing for three months shows a stable population of AEPs and progeny in the alveolar epithelium. Yellow arrow, labelled cell; white arrow, unlabelled cell. b, c, Quantification of AT1 and AT2 cells labelled by the AEP lineage mark at homeostasis. Lower power (d–f) and higher power (g–i) images showing expansion of AEPs in a regional fashion, one month after influenza injury. Dotted white line in f shows the edge of a Krt5++ pod, with a dearth of AEP-lineage-labelled cells. Panels g–i show additional channels of the same fields as shown in Fig. 1i, j. j, Representative FACS plot showing expansion of AEP-lineage-labelled epithelial cells after influenza. The quantification of these FACS plots can be found in Fig. 1n. k–o, Comparison of Ki67++ expression in AT2 cells and AEPs after influenza. In areas of regeneration, Ki67++ AEPs constitute the majority of cells entering the cell cycle, when compared to AT2 cells. Data shown represent n = 5 (j–o), 6 (a–c) or 10 (d–i) independent mice from three individual experiments, except for the nine-month lineage tracing which was performed in two separate experiments. Statistics are representative of all biological replicates. All data are shown as centred on mean with bars indicating standard deviation. *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001 by ANOVA with preplanned pairwise comparisons and adjustment for multiple comparison testing. Scale bar, 50 μm.
Extended Data Figure 3 | In contrast to adult lung homeostasis, the Wnt response in the alveolar epithelium during alveologenesis is dynamic.

a, Schematic of lineage labelling procedure to assess Wnt-responsive epithelium during alveologenesis. b, Epithelial cells were identified by FACS as Epcam$^+$ CD45$^-$ CD31$^-$. Cells were then gated for tdTomato and eYFP expression as shown. c, Quantification of Wnt responsiveness in the alveolar epithelium over a 1-day or 3-week lineage trace. d, Model of directionality and magnitude of AT2 and AEP transitions. During alveologenesis, AT2 and AEP fates are somewhat fluid, though the AEP population decreases during this period of lung development. During adult homeostasis, few if any AT2 cells take on the AEP fate (see Fig. 2). After injury, AEPs expand to create AT2 cells, but even after injury very few AT2 cells adopt the AEP fate. Data shown represent $n=3$ mice. Statistics are representative of all biological replicates. Data in c are centred on mean with bars indicating standard error of the mean.
Extended Data Figure 4 | AEPs are a distinct lineage compared to Sox2-derived Krt5+ cells and are capable of generating AT1 cells. 
a–d. AEPs and Krt5+ cells inhabit distinct regions of the regenerating mouse lung. 
a. Overview of a region surrounding a Krt5+ pod. b. In regions of mild injury, AEPs and AEP-lineage-marked AT2 cells predominate and no Krt5+ cells are seen. Yellow arrow, AEP-labelled cell. 
c. At the border of zone 4 areas of alveolar destruction, AEPs are observed regenerating AT2 cells. d. Krt5+ cells are distinct from AEPs and never bear the AEP lineage mark. Red arrow, Krt5+ cell. e. AEP-lineage cells do not express Krt5 or Sox2 protein at baseline, in contrast to previously reported lineages.5,4 Arrows represent probable AEPs by morphology. f. Krt5+ cells predominate in zone 4 regions, where AEPs are not present. 
g. Quantification demonstrating that Krt5+ cells are never marked with the AEP lineage mark. h. AT2 populations expand markedly after influenza injury, except in zone 4. i. Krt5+ cells rarely express SftpC in zone 4 regions. j–l. One month after influenza injury, AEPs give rise to a small number of Hopx+ AT1 cells, predominantly in zone 2 of mild injury. Yellow arrow, AEP-labelled cells; white arrow, unlabelled cells. Zone 3 (l) has very few AEP-derived Hopx+ cells, which may be due to a lag in AT1 regeneration from AEPs in this more severely affected region. Data shown represent n = 6 (a–g, i) or 10 (h, j, k) independent mice across three individual experiments. Statistics are representative of all biological replicates. All data are shown as centred on mean with bars indicating standard deviation. **P < 0.01 and ***P < 0.001, by ANOVA with preplanned pairwise comparisons and adjustment for multiple comparison testing. Scale bars: a, 200 μm; b–d, j–l, 50 μm.
Extended Data Figure 5 | Wnt signalling in the alveolar epithelium is largely stable after influenza infection, and AEP lineage labelling is not affected by tamoxifen perdurance. a, FACS gating strategy used for all post-influenza FACS experiments in Fig. 1, Extended Data Fig. 2 and b, c. SSC-A, side-scatter area, SSC-H, side-scatter height, FSC-H, forward-scatter height. b, c, FACS analysis demonstrates that Axin2tdT intensity is mildly decreased in the epithelium at 7 and 14 days after influenza infection. d, In regions of milder lung injury, most lineage-labelled AT2 cells are eYFP+ and tdTomato−, which suggests that these cells are the progeny of AEPs. e, In zone 3, we detect a mix of eYFP+tdTomato+AEPs (red arrowheads) and eYFP−tdTomato−AEP progeny (yellow arrowheads) among the AT2 cell population. f, Experimental design of lineage tracing experiment in g–i, with a longer incubation time after tamoxifen treatment than in the experiments that generated the data presented in a–e, and Fig. 1 and Extended Data Figs 4, 6. g, h, Confocal imaging demonstrating lineage labelling of AT2 cells with the AEP lineage mark 28 days after influenza-mediated injury. White arrows, unlabelled AT2 cells; yellow arrows, AEP-labelled cells. i, Quantification of lineage-labelled AT2 cells in multiple regions of lung injury. Representative seven-day lineage data is reproduced from Fig. 1 for comparison. Data shown represent n = 4 (a–c) or 5 (d–i) independent mice across two different experiments. Statistics are representative of all biological replicates. All data were analysed with ANOVA followed by preplanned pairwise comparisons and adjustment for multiple comparison testing, and are shown centred on mean with bars indicating standard deviation. **P < 0.01. Scale bars, 50 μm.
Extended Data Figure 6 | Transcriptome analysis of AEPs versus AT2 cells, and activation of cell-cycle genes in AEPs after influenza injury. a, Volcano plot of 14,618 genes tested using a linear model in the R package limma, showing the distinct differences in gene expression in AEPs and AT2 cells. Notable lung-progenitor developmental signalling and transcription factors are indicated. b, GO analysis of the top 500 most-differentially expressed genes, showing the enrichment of categories related to lung development and morphogenesis in AEPs. c, Heat maps of two of the AEP-enriched GO categories. Important regulators of lung-progenitor-cell biology are indicated. d, qPCR confirms upregulation of a subset of important regulators of lung progenitor biology in AEPs. e, AT2 and AEP open chromatin is found near distinct sets of genes involved in the cell cycle. f, Schematic of analysis of changes in expression of AEP-primed genes after influenza infection. g, A subset of primed cell-cycle regulators in AEPs show expression changes after influenza infection. qPCR data are from $n = 4$ mice from two separate infections. All data are shown as centred on mean with bars indicating standard deviation. Statistics are representative of all biological replicates. *$P < 0.05$ and **$P < 0.01$ by two-tailed t-test.
Extended Data Figure 7 | ATAC-seq reveals distinct differences in open chromatin architecture in AEPs versus AT2 cells. a, ATAC-seq peaks in both AT2 cells and AEPs are similar to previously described mouse lung genome-wide DNase hypersensitivity profiling. b, AT2 and AEP ATAC peaks are distributed in a similar fashion, predominantly within intergenic regions and introns. c, GO enrichment analysis of the nearest neighbour genes in the vicinity of AT2 peaks, AEP peaks and peaks common to both AEPs and AT2 cells shows that common peaks are enriched for general cellular housekeeping roles, whereas AT2 open chromatin is enriched near genes associated with exocytosis and cell differentiation. By contrast, AEP peaks are enriched near genes associated with lung development processes. d, e, Examination of the genes associated with open chromatin in AEPs reveals a strong enrichment for transcription factors associated with lung endoderm progenitor cells, including members of Klf, Six, Sox, Nkx2 and Elf/Ets families. By contrast, AT2 cell open chromatin is associated with a unique set of transcriptional regulators that includes members of the NfI and Cebp families, which are known to regulate AT2 cell surfactant genes. For details of ATAC analysis, see Methods.
Extended Data Figure 8 | The combination of HT2-280 and TM4SF1 antibodies are capable of identifying AEPs in human lung. a, Top panels show isotype and active antibody gates for sheep anti-mouse Tm4sf1 FACS. The bottom panels show that the Tm4sf1 antibody detects approximately 20% of SftpccreERT2eYFP labelled AT2 cells. b, Isotype and active antibody gates for human HT2-280 (AT2 marker) antibody and TM4SF1 antibody. c, d, An example of the FACS gating strategy used to generate the data shown in Fig. 3. e, f, Selection for HT2-280 strongly enriches for human AT2 cells. g, h, The majority of isolated HT2-280 cells express SFTPC protein by cytopsin. i, j, Human AEPs in organoid culture do not express KRT5 or SOX2 protein at detectable levels. Each FACS panel shown in a–f shows gates from cells of one individual mouse or patient and is representative of n = 6 independent mice across two individual experiments or n = 4 human patients. Isotype staining was performed three times to confirm specificity. Statistics are representative of all biological replicates. Statistics in h are calculated with two-tailed t-test, displayed as mean with bars showing standard deviation. Scale bars: g, 25 μm; i, j, 50 μm (i, j).
Extended Data Figure 9 | Mouse AEPs generate more alveolar organoids compared to AT2 cells, and cells in these organoids are restricted from AT1 cell differentiation by Wnt signalling. a, Schematic of mouse alveolar organoid culture method. b–m, Sftpc<sup>+</sup> mouse AT2 cells (b–d, h–j) and mouse AEPs (e–g, k–m) were isolated from the indicated mouse lines and cultured in alveolar organoid assays. AT2 cells (b) and AEPs (e) both form alveolar organoids. AEPs generate more numerous and larger organoids than do AT2 cells. Activation of Wnt signalling using CHIR99021 does not increase the organoid-forming efficiency of either AT2 cells (c) or AEPs (f) but does increase the number of Sftpc<sup>+</sup> cells in treated organoids (i, l, o). Inhibition of Wnt signalling using XAV939 increases the number and size of alveolar organoids (d, g, n, q), decreases the number of Sftpc<sup>+</sup> AT2 cells and increases the number of Aqp5<sup>+</sup> AT1 cells (j, m, p). For tests of all parameters, AEPs exhibited a more marked response to Wnt modulation than did AT2 cells. Data shown represent \( n = 12 \) wells from \( n = 4 \) individual mice in each group, across 3 individual experiments. Quantitative counting shown for cell differentiation (o, p) represents counting of \( n > 400 \) organoids from \( n = 4 \) mice. All data were analysed with ANOVA followed by preplanned pairwise comparisons and adjustment for multiple comparison testing, and are shown centred on mean with bars indicating standard deviation. * \( P < 0.05 \), ** \( P < 0.01 \), *** \( P < 0.001 \) and **** \( P < 0.0001 \). Statistics are representative of all biological replicates. Scale bars: 50 µm.
Extended Data Figure 10 | Combination of ATAC-seq and RNA-seq emphasizes the Wnt- and FGF-responsive nature of AEPs and identifies several novel AEP-enriched direct Wnt target genes. a, Schematic of human RNA-seq experiments. b, GO term analysis of the top 300 human AEP-enriched genes shows enrichment of several categories associated with lung progenitor cell function, similar to observations made of mouse AEPs. c, Evaluation of chromatin accessibility in the mouse genome near common AEP-enriched genes demonstrates a significant overrepresentation of Tcf binding sites, particularly in putative regulatory regions 5 kb immediately upstream of the transcriptional start site. For details of enrichment analysis, see Methods. d, Schematic of areas of AEP-enriched open chromatin near selected AEP-enriched genes. Peak height represents coverage of the indicated genomic region in the ATAC library, and the number indicates the fold enrichment in the indicated peak. e, Chromatin immunoprecipitation qPCR on AEP versus AT2 chromatin demonstrates Ctnnb1 antibody binding at the differentially accessible genomic regions near Etv4, Sftpα, Lamp3 and Gpr116 in AEP cells, indicating that these genes are direct Wnt targets. Data are shown as mean with individual data points showing summary data from two independent chromatin immunoprecipitation experiments with multiple technical replicates. f–j, Fgfr2 activation in mouse AEPs drives increased proliferation and the formation of larger organoids; quantification shown in j. See Fig. 4 for additional data. k, RNAscope showing enriched expression of Fgfr2 (red) in lineage-labelled AEPs. l–q, Similar to treatment with Fgf7, Fgf10 treatment drives increased colony-forming efficiency in both mouse AEPs (l–p) and human AEPs (q). Data shown in f–j, l–q represent a minimum of \( n = 12 \) wells across two individual experiments. Statistics are representative of all biological replicates. Data were analysed with ANOVA followed by preplanned pairwise comparisons and adjustment for multiple comparison testing, and are shown centred on mean with bars indicating standard deviation. *\( P < 0.05 \), **\( P < 0.01 \), ***\( P < 0.001 \) and ****\( P < 0.0001 \).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

1. Sample size
   Describe how sample size was determined.

Sample sizes for animal studies were made as large as possible based on the complex genetics and long term nature of these experiments. As a baseline, N=5 was chosen empirically given the observed variation in response to injury from a series of preliminary studies. Many experiments in the manuscript have N=10 to ensure robustness. For all animal experiments, multiple animals were tested in 2-4 individual experiments to control for inter-cohort variation. Statistics were calculated across all biological and technical replicates.

2. Data exclusions
   Describe any data exclusions.

No data were excluded

3. Replication
   Describe whether the experimental findings were reliably reproduced.

There were no difficulties with reproduction of any of the reported findings. All experiments for this study are from multiple biological and technical replicates, with N and replication numbers reported in the legend of each Figure.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.

For all animal experiments, experimental and control animals were chosen at random from among littermates with the appropriate genotype. All injury experiments were performed with uninjured littermates as controls. All animals in these studies were maintained on a similar mixed C57Bl/6 x CD1 genetic background. Both males and females were used in this study, and results are representative of data obtained from animals of both sexes.

For human studies, untreated or control samples from each individual were used as comparator for experimental manipulations on parallel samples. Larger numbers of technical replicates were used given the range of human genetic variation. No randomization was possible for these samples.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The genotype of all animals was known to the investigators at the outset of the studies as all animals in each experiment and treatment group were littermates with the same genotype. Blinding to experimental condition was challenging as the reported influenza model caused obvious illness in animals and injury in examined tissues. Blinding to condition for quantitative counting was attempted but is not possible given the severity of injury and the morphological differences in injured tissues.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | confirmed |
|-----|-----------|
| ☑   |           |

- The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. \( P \) values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Policy information about availability of computer code

Software

Describe the software used to analyze the data in this study.

- Imaging and Analysis - Adobe Photoshop CS6, FUJI 2.0.0-rc-43/1.51f, Nikon Elements v4.60.00, Leica LAS X
- FACS analysis - FlowJo 10.0.8r1
- Statistical Analysis - Graphpad Prism (version 7.0a), R v3.3
- Bioinformatics - No custom code was used.
  - Fastq v0.11.5
  - Star v2.5.2b
  - Rsubread, EdgeR, and Limma (bioconductor v3.4)
  - Deeptools v2.4
  - Samtools v1.3.1
  - R v3.3

For RNA-Seq: Fastq files were assessed for quality control using the FastQC program. Fastq files were aligned against the mouse reference genome (mm9) or human reference (hg19/hGRC37) genome using the STAR aligner 40. Duplicate reads were flagged using the MarkDuplicates program from Picard tools. Per gene read counts for Ensembl (v67) gene annotations for the mouse samples or Ensembl (v75) for the human were computed using the R package Rsubread with duplicate reads removed. Gene counts represented as counts per million (CPM) were first normalized using trimmed mean of M-values method in the R package edgeR and genes with 25% of samples with a CPM < 1 were removed and deemed low expressed. This data was transformed using the VOOM function from the limma R package. Differential gene expression was performed using a linear model with the limma package. Given the small sample size of the experiment, we employed the empirical Bayes procedure as implemented in limma to adjust the linear fit and calculate \( P \) values. \( P \) values were adjusted for multiple comparisons using Benjamini-Hochberg procedure. For the human data a paired analysis was employed using the patient as a blocking variable. Heatmaps and PCA plots were generated in R. Gene Ontology enrichment analysis was performed using the ToppGene Suite (http://toppgene.cchmc.org/).

For ATAC-Seq - Following sequencing, Fastq files were aligned against the mouse reference genome (mm9) using the STAR aligner 40, with default parameters plus options to suppress the matching of spliced reads (’--outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4’). Duplicate reads were flagged using the MarkDuplicates program from Picard tools and removed using samtools. MACS2 was used to call peaks with the following options ‘--nomodel --shift -100 --extsize 200’ 44. Differential ATAC-seq peaks were
determined using the bgdDiff command from MACS2 and default options. Peaks were filtered to have a MACS2 log10 likelihood ratio score > 10 and within -50kb and +10kb of the transcription start site of Ensembl 67 protein coding genes. ATAC-seq enrichment heatmaps were created using deepTools2.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Axin2-CreERT2-tdTomato animals are available by request to Dr. Morrisey.

The SftpC-CreERT2 line used in this study was developed by Dr. Harold Chapman, UCSF. These animals were a gift of Dr. Chapman and are available on request directly from Dr. Chapman.

All other animals and materials used in this manuscript are publicly available. Other mouse models used are available at the Jackson Laboratories.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Extensive details of all antibodies used is available in the Materials and Methods

For Tm4sf1 antibodies:

1) Mouse - multiple Tm4sf1 antibodies were obtained and tested at multiple concentrations in IHC and FACS. Staining was confirmed as specific by FACS in comparison to an isotype antibody.

2) Human Tm4sf1-APC concentration was titrated by FACS, compared to isotype control antibody, and used at the optimal FACS concentration for cell sorting. Data is presented on the quality controls used for the human experiments.

Additional antibody information:

For sorting and quantification, the following antibodies were used: Pdpn-eFluor660 (eBioscience, Clone 8.1.1, 1:100) EpCAM-APC (eBioscience, Clone G8.8, 1:200), EpCAM-eFluor488 (eBioscience, Clone G8.8, 1:200), CD31-PeCy7 (eBioscience, Clone 390, 1:200), CD45-PeCy7 (eBioscience, Clone 30-F11, 1:200). Two anti-mouse Tm4sf1 antibodies were used to ensure specificity: Sheep anti-mouse Tm4sf1 (R&D systems, AF7514, 1:10) and Sheep IgG isotype control (R&D systems, 5-001-A, 1:10) with anti-Sheep 488 secondary (Abcam, ab150177, 1:50) or Rabbit anti-mouse Tm4sf1 (LS Biosciences, B7077, 1:25) and Rabbit IgG isotype control (LS Biosciences, LS-C109221, 1:25) with Donkey anti-rabbit 488 secondary (Life Technologies, A212016, 1:200).

Immunohistochemistry was used to detect protein expression using the following antibodies on paraffin sections: GFP (chicken, Aves, GFP-1020, 1:500), GFP (goat, Abcam, ab5450, 1:100), RFP (rabbit, Rockland, 600-901-379, 1:250), Scgb1a1 (goat, Santa Cruz, sc-9772, 1:20), Tubbig (mouse, BioGenex, MU178-UC, 1:20), SftpC (rabbit, Millipore, ABC99, 1:250), SftpC (goat, Santa Cruz, sc-7750, 1:50), Pdpn (mouse, HybriDNA Bank, Clone 8.11.1, 1:50), Aqp5 (rabbit, Abcam, ab92320, 1:100), and Ki67 (rabbit, Abcam, clone SP6, ab16667, 1:50), anti-mouse Tm4sf1 (rabbit, LS Biosciences, B7077, 1:500).

For human experiments: EPCAM-PE (BD, mouse, Clone 1B7, 1:50), HT2-280 (mouse IgM, a gift of Leland Dobbs, UCSF, 1:50), TM4SF1-APC (mouse, R&D Systems, Clone 877621, 1:100), Mouse IgG1-APC isotype control (R&D systems, 1C002A, 1:100), anti-APC microbeads (Miltenyi, 130-090-855, 1:20), anti-mouse IgM microbeads (Miltenyi, 130-047-302, 1:20).

For CHIP: anti-β-catenin (Santa Cruz sc-7963) or anti-IgG1 isotype control (Santa Cruz sc-3877).
10. Eukaryotic cell lines
   a. State the source of each eukaryotic cell line used. MRC-5 cells were obtained from ATCC and used as supportive mesenchymal cells in human lung organoid experiments.
   b. Describe the method of cell line authentication used. Cells were used at low passage number and discarded after P10. No specific validation was used.
   c. Report whether the cell lines were tested for mycoplasma contamination. All cells were tested negative for mycobacterial infection.
   d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. No commonly misidentified lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   All animal studies were performed under guidance of the University of Pennsylvania Institutional Animal Care and Use Committee. For all animal experiments, experimental and control animals were chosen at random from among littermates with the appropriate genotype. All injury experiments were performed with uninjured littermates as controls. All animals in these studies were maintained on a similar mixed C57Bl/6 x CD1 genetic background. Both males and females were used in this study, and results are representative of data obtained from animals of both sexes. Animals were housed in a barrier facility, with animals for influenza infection housed in a BSL-2 barrier facility.

   For induction of all Cre recombinase models, tamoxifen (Sigma) was dissolved in 100% ethanol and diluted with corn oil to produce a 10% ethanol:tamoxifen:corn oil mixture at 20 mg/mL. 6-8 week old mice were injected intraperitoneally (IP) with 200mg/gm on 3-5 consecutive days to induce recombination. Number and timing of injection for all experiments is reported in the associated Figure.

   For influenza infection, we utilized PR8 influenza virus, which was a gift of Dr. John Wherry at the University of Pennsylvania. Virus was produced in their laboratory and titered based on LD50, and for experimental infection, virus was diluted in PBS and a dose of 0.3 LD50 was administered via intranasal instillation. Animals were monitored on a daily basis, with determination of weight and scoring of body condition on the BSC scale. In the event of rapid weight loss or dehydration, animals were provided with moistened chow and diet gel for support. Animals which lost >30% of weight or developed a BSC score of 1 (indicative of clinically relevant weight loss or dehydration) were sacrificed humanely prior to death. At the doses utilized in this study, less than 10% of animals met these IACUC-approved end-points. All other animals survived infection, and were analyzed at the planned regeneration end-point.

Policy information about studies involving human research participants

12. Description of human research participants
   This study utilized cells derived from de-identified non-utilized lungs donated for organ transplantation via an established protocol (PROPEL, approved by University of Pennsylvania Institutional Review Board) with informed consent in accordance with institutional procedures. All patient information was removed prior to use. This use does not meet the current NIH definition of human subject research, but all institutional procedures required for human subject research were followed throughout the reported experiments.
Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

▶ Data presentation

For all flow cytometry data, confirm that:

☑ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☑ 2. 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).

☑ 3. All plots are contour plots with outliers or pseudocolor plots.

☑ 4. A numerical value for number of cells or percentage (with statistics) is provided.

▶ Methodological details

5. Describe the sample preparation. All samples in this manuscript are whole single cell suspensions derived from human or mouse lung. Briefly, whole lung was digested with a mix of dispase, collagenase, and DNAse, macerated, and agitated at 37°C. Samples were then passed over multiple filters, followed by a final 40 micron filtration. Red blood cells were lysed, staining was performed with the antibodies and concentrations noted in the materials and methods, and samples were then taken for analysis or cell sorting without fixation. DAPI was utilized to identify dead cells.

6. Identify the instrument used for data collection. FACS quantification was performed on BD Accuri and BD LSRFortessa. FACS sorting was performed on MoFlo Astrios and BD FACSJazz.

7. Describe the software used to collect and analyze the flow cytometry data. Raw FCS files were obtained from each experiment and analyzed in FlowJo. Compensation was performed per experiment with single stained controls using the FlowJo compensation feature.

8. Describe the abundance of the relevant cell populations within post-sort fractions. Typically, EPCAM+ lung epithelial cells comprise 10-15% of single cell suspensions prepared as above and gated as below. Percentages for fractions of the epithelial population are noted in each figure and legend.

9. Describe the gating strategy used. 1) All captured events were gated for singlets using SSC-A v SSC-H gating 2) Intact cells were separated from debris using SSC-A vs FSC-A 3) For most experiments, live cells were identified by absence of DAPI 4) Epithelial cells were identified as CD45- CD31- EPCAM+ cells 5) Relevant experimental comparisons were made as noted in each figure.

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