

PRDM3/16 regulate chromatin accessibility required for NKX2-1 mediated alveolar epithelial differentiation and function

Received: 8 January 2024

Accepted: 27 August 2024

Published online: 16 September 2024

Check for updates

Hua He ^{1,2,5} ✉, Sheila M. Bell^{3,5}, Ashley Kuenzi Davis ³, Shuyang Zhao ³, Anusha Sridharan³, Cheng-Lun Na³, Minzhe Guo ^{3,4}, Yan Xu ^{3,4}, John Snowball³, Daniel T. Swarr ^{3,4}, William J. Zacharias ^{3,4} & Jeffrey A. Whitsett ^{3,4} ✉

While the critical role of NKX2-1 and its transcriptional targets in lung morphogenesis and pulmonary epithelial cell differentiation is increasingly known, mechanisms by which chromatin accessibility alters the epigenetic landscape and how NKX2-1 interacts with other co-activators required for alveolar epithelial cell differentiation and function are not well understood. Combined deletion of the histone methyl transferases *Prdm3* and *Prdm16* in early lung endoderm causes perinatal lethality due to respiratory failure from loss of AT2 cells and the accumulation of partially differentiated AT1 cells. Combination of single-cell RNA-seq, bulk ATAC-seq, and CUT&RUN data demonstrate that PRDM3 and PRDM16 regulate chromatin accessibility at NKX2-1 transcriptional targets critical for perinatal AT2 cell differentiation and surfactant homeostasis. Lineage specific deletion of PRDM3/16 in AT2 cells leads to lineage infidelity, with PRDM3/16 null cells acquiring partial AT1 fate. Together, these data demonstrate that NKX2-1-dependent regulation of alveolar epithelial cell differentiation is mediated by epigenomic modulation via PRDM3/16.

Lung morphogenesis and function are orchestrated by sequential interactions among diverse progenitor cells from the foregut endoderm and splanchnic mesenchyme. Signaling and transcription programs directing cell proliferation and cell fate differentiation are required to place precise numbers of specific cell types in appropriate positions to enable pulmonary function at birth^{1–4}. Cell fate decisions for each of the many pulmonary cell types are instructed by cell-specific gene regulatory networks, which are dynamically used during morphogenesis to control transcription factor activity regulating the expression of a multitude of genes and proteins mediating cell-specific functions. Formation of the mammalian lung foregut endoderm is dependent on transcriptional programs controlled by NKX2-1 (aka thyroid transcription factor 1) and its interactions with other

transcription factors to form the conducting airways, peripheral lung saccules, and ultimately, the alveolar regions of the lung required for gas exchange after birth^{1,5–11}. NKX2-1 is expressed at variable levels in conducting and peripheral lung tubules and is required for branching morphogenesis. SOX2 defines progenitor cells forming the proximal tubules, while epithelial cells in the peripheral acinar buds are marked by expression of SOX9 and ID2^{1,4,12}. NKX2-1 plays distinct roles in regulating epithelial cell fate during early lung morphogenesis as alveolar type 2 (AT2) and alveolar type 1 (AT1) cells differentiate from SOX9 progenitors. AT1 cell differentiation is directed by the interactions of NKX2-1 with the YAP/TAZ-TEAD family of transcription factors, distinguishing AT1 cell fate choices from AT2 cell fates early in lung development^{9,11,13,14}. NKX2-1 is required for the induction of a diversity

¹Key Laboratory of Birth Defects and Related Diseases of Women and Children of MOE, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China. ²NHC Key Laboratory of Chronobiology, Sichuan University, Chengdu, Sichuan, China. ³Perinatal Institute, Division of Neonatology and Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁴Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁵These authors contributed equally: Hua He, Sheila M. Bell. ✉ e-mail: Hua.he@scu.edu.cn; Jeffrey.whitsett@cchmc.org

of transcriptional targets in AT2 cells to induce pulmonary surfactant lipid, surfactant protein, and innate host defense protein synthesis prior to birth¹⁵. Incomplete AT2 cell differentiation in preterm infants is associated with surfactant deficiency and causes respiratory distress syndrome (RDS), a major cause of infant morbidity and mortality¹⁶. NKX2-1 directly regulates the expression of genes critical for lung function at birth, including those encoding surfactant proteins A, B, C, and D, and genes controlling AT2 cell lipid homeostasis including *Abca3*, *Scd1*, and *Sreb*^{7,17}. NKX2-1 functions in gene regulatory networks with a number of transcription factors expressed in pulmonary epithelial cells, including FOXA family members (FOXA1 and FOXA2), KLF5, GATA6, TEAD, SREBF1, NF1, and CEBP α , which bind cis-active regulatory elements to activate transcriptional targets critical for surfactant homeostasis prior to and after birth^{6,9,17–20}.

Accessibility of transcription factors to their targets is regulated by the dynamic control of chromatin structure that is influenced by a diversity of histones and their post-translational modifications²¹. While the identity and roles of AT2 cell selective transcription factors required for perinatal AT2 cell differentiation and function are increasingly understood, how chromatin is remodeled to provide access to the transcription factors critical for AT2 and AT1 cell functions prior to birth is poorly understood. Remarkably, the dramatic increase in the expression of NKX2-1 transcriptional targets before birth is not accompanied by concomitant increased expression of NKX2-1, supporting the concept that NKX2-1 phosphorylation and changes in chromatin accessibility modulate the activity of NKX2-1 at its transcriptional targets^{7,11,22}.

PR(SET) Domain (PRDM) proteins are a family of 19 distinct zinc finger proteins that serve as transcription factors and co-activators known to remodel chromatin and play diverse roles in organogenesis and stem cell functions in multiple tissues. The semi-redundant proteins PRDM3 (also known as EVI1 or MECOM) and PRDM16 (also known as MEL1) share similar structures, expression patterns, and functions²³. Both contain a PR domain that has histone methyltransferase activity that may regulate the epigenetic states of chromatin^{24,25}. Multiple zinc finger domains mediate DNA binding²³. PRDM3 and PRDM16 play diverse roles in regulating gene expression, with both repressor and activator functions^{24–28}. PRDM16 is required for the normal development of hematopoietic stem cells, chondrocytes, adipocytes, maintenance of neural and intestinal stem cells, and cardiac morphogenesis^{27,29–33}. PRDM3 and PRDM16 play distinct and opposing roles during craniofacial development^{24,25,30}.

The present work evaluated the developmental function of PRDM3 and 16 in lung epithelial cells during lung morphogenesis and identified complementary and cell-type specific roles of PRDM3 and PRDM16 in alveolar epithelial cell differentiation. Deletion of PRDM3 and PRDM16 in the mouse lung endoderm enhanced AT1 cell and inhibited AT2 cell lineage decisions in the embryonic lung and blocked prenatal induction of AT2 cell selective genes critical for surfactant lipid and surfactant-associated protein production required for postnatal survival. Transcriptomic and epigenetic analyses identified gene regulatory networks implicating PRDM3, PRDM16, and NKX2-1 as critical regulators of lung epithelial formation and function, and demonstrate that PRDM3, PRDM16, and NKX2-1 bind at cis-regulatory elements to modulate the expression of AT2 and AT1 selective genes before birth. Therefore, PRDM3/16 function is required to define alveolar epithelial cell fate during embryonic lung morphogenesis and mediate AT2 cell maturation required for the transition to air breathing at birth.

Results

***Prdm3*, PRDM16, and NKX2-1 are co-expressed in lung endoderm** PRDM3 and PRDM16 share similar protein structures and are widely expressed during embryogenesis³⁴. *Prdm3* mRNA and PRDM16 immunofluorescence staining were present in fetal mouse lung epithelial cells as early as E11.5. Thereafter, *Prdm3* and PRDM16 were co-

localized with NKX2-1 in both peripheral acinar tubules and most conducting airway cells from E16.5 to E18.5 (Fig. 1a, b, Supplementary Fig. 1). At E16.5, PRDM16 expression was downregulated in maturing pulmonary neuroendocrine cells where NKX2-1 expression persisted (Supplementary Fig. 1e–h). *Prdm3*, *Prdm16*, and *Nkx2-1* mRNA levels did not change dynamically during fetal lung development, while the expression of known NKX2-1 transcriptional targets increases markedly prior to birth^{22,35} (Supplementary Fig. 2a).

***Prdm3* and *Prdm16* are required for cellular differentiation and lung function at birth**

To test the roles of PRDM3 and PRDM16 in lung morphogenesis and function, we conditionally deleted each gene individually and in combination using *Shh-Cre* to cause recombination of the *floxed* alleles in lung endoderm, producing *Prdm3*^{ShhCre Δ/Δ} , *Prdm16*^{ShhCre Δ/Δ} , and *Prdm3/16*^{ShhCre Δ/Δ} mice. *Prdm3*^{ShhCre Δ/Δ} and *Prdm16*^{ShhCre Δ/Δ} mice survived after birth without visible anatomical defects. RNA sequencing of the sorted lung epithelial cells from *Prdm3/16*^{ShhCre Δ/Δ} transgenic mice demonstrated the loss of exon 9 in *Prdm16* RNA and the loss of exon 4 in *Prdm3* RNA (Supplementary Fig. 2b), consistent with deletion of the *floxed* regions in both alleles^{36,37}. Immunofluorescence analysis demonstrated the absence of PRDM16 protein in E16.5 *Prdm3/16*^{ShhCre Δ/Δ} lung epithelial cells while staining persisted in the nontargeted endothelial cell population (Fig. 1f).

While all newborn *Prdm3/16*^{ShhCre Δ/Δ} mice died soon after birth, the only physical difference notable in *Prdm3/16*^{ShhCre Δ/Δ} fetuses was the rudimentation of the fifth digit of each paw. While similar in body weight, lung-to-body weight ratios were moderately decreased in both *Prdm3/16*^{ShhCre Δ/Δ} and *Prdm3* ^{Δ/Δ} /*16*^{ShhCre Δ/Δ} fetuses at E18.5 (Supplementary Fig. 2c, Source data are provided as a Source Data file). At E16.5 and E18.5, while the trachea and esophagus separated normally, peripheral regions of the developing lung were poorly sacculated in *Prdm3/16*^{ShhCre Δ/Δ} fetuses (Fig. 1i, m). NKX2-1 expression was normally distributed throughout the pulmonary epithelium regardless of PRDM3/16 expression (Fig. 1a, e, j, n). In *Prdm3/16*^{ShhCre Δ/Δ} E16.5 lungs, SOX2 staining was normally distributed in conducting airways surrounded by α SMA-expressing smooth muscle cells, and SOX9 expression was restricted to peripheral epithelial progenitors, demonstrating preserved cephalo-caudal patterning of the lung (Fig. 1c, g). In E18.5 lungs, immunofluorescence demonstrated normal distribution of club and ciliated cells within the mutant airways. Compared to controls, *Prdm3/16*^{ShhCre Δ/Δ} club cells expressed reduced levels of SCGB1A1 but similar levels of SCGB3A2. Basal and pulmonary neuroendocrine cells were observed in aberrant distal locations in *Prdm3/16*^{ShhCre Δ/Δ} lungs (Supplementary Fig. 3). In the alveolar region, a paucity of SFTPC⁺ AT2 cells and reduced LAMP3 levels were observed in *Prdm3/16*^{ShhCre Δ/Δ} animals (Fig. 1j, k, n, o). HOPX staining was maintained in AT1 cells in *Prdm3/16*^{ShhCre Δ/Δ} fetuses. AT2 and AT1 cells were quantified by SFTPC and HOPX staining, demonstrating a decreased proportion of AT2 cells and an increased proportion of AT1 cells in *Prdm3/16*^{ShhCre Δ/Δ} lungs (Fig. 1q, Source data are provided as a Source Data file). Ultrastructural analysis demonstrated a lack of lamellar bodies, the lipid-rich storage sites of pulmonary surfactant in *Prdm3/16*^{ShhCre Δ/Δ} AT2 cells (Fig. 1l, p). This absence of lamellar bodies and failure of postnatal survival of *Prdm3/16* deficient mice are consistent with respiratory failure caused by the lack of pulmonary surfactant and/or pulmonary hypoplasia.

Recent lineage tracing studies demonstrated that AT1 and AT2 lineage decisions from SOX9 epithelial progenitors occur early in embryonic lung morphogenesis^{38,39}; therefore, the changes in AT1 and AT2 cell allocation observed in *Prdm3/16*^{ShhCre Δ/Δ} mice are likely attributable to roles of PRDM3/16 in lineage decisions occurring early in embryonic lung formation. Consistent with this idea, immunofluorescence staining of RAGE, an AT1 cell marker, was increased in

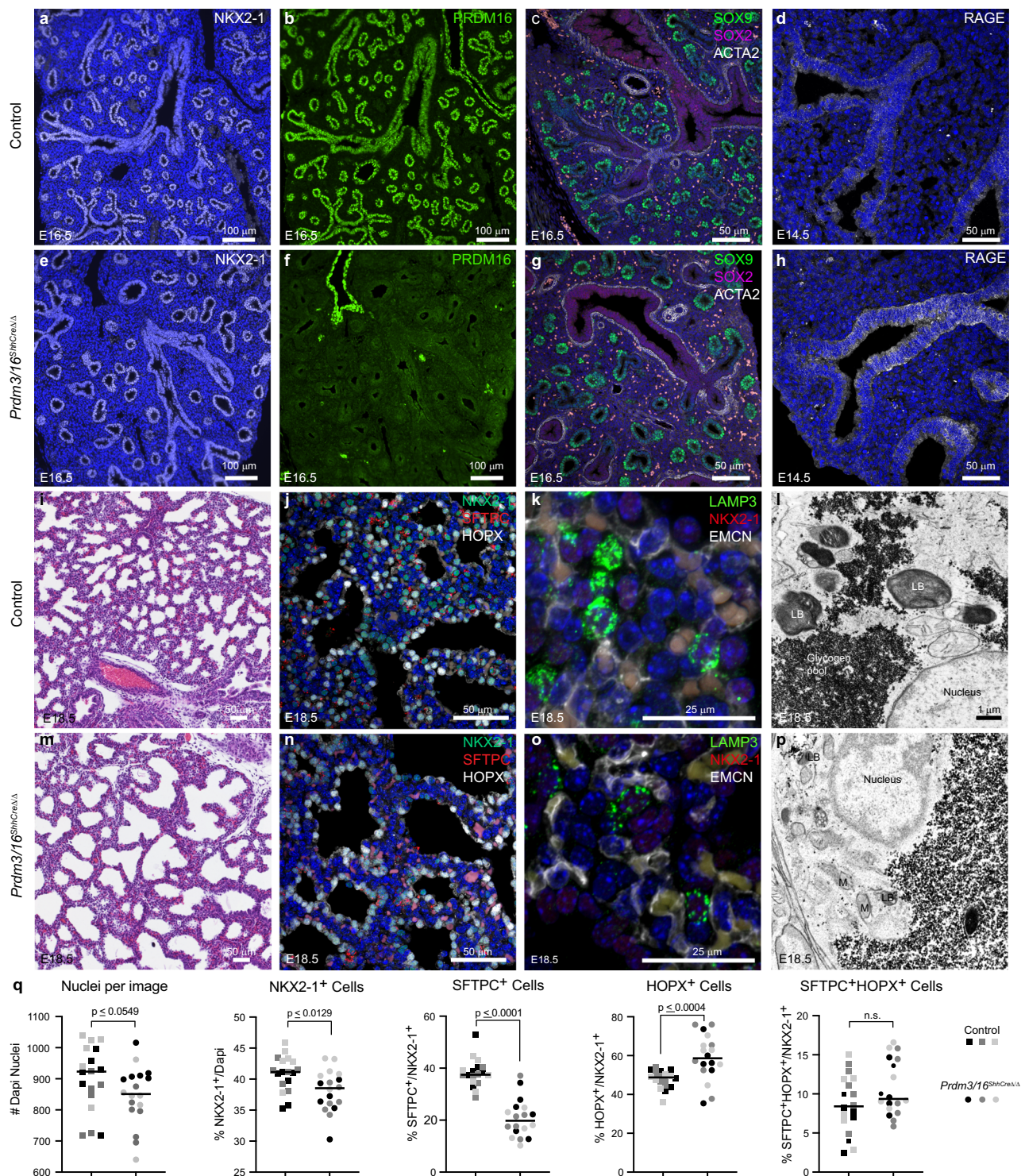


Fig. 1 | Decreased AT2 cell numbers and differentiation after deletion of *Prdm3/16*. Immunofluorescence staining of embryonic lung indicates normal expression of NKX2-1 after deletion (a, e), loss of PRDM16 staining in lung epithelium in *Prdm3/16^{ShhCreΔ/Δ}* fetuses and retention in vascular smooth muscle (b, f), $n = 4$ biological replicates. c, g Normal proximal (SOX2⁺) and distal (SOX9⁺) epithelial patterning is observed, $n = 2$ control and 4 mutant embryos. d, h At E14.5, the AT1 cell marker RAGE is increased in regions of *Prdm3/16^{ShhCreΔ/Δ}* lungs, $n = 3$ biological replicates. i, m Hematoxylin and eosin staining of E18.5 lungs demonstrating poor sacculation

in *Prdm3/16^{ShhCreΔ/Δ}* $n = 6$ biological replicates. j, k, n, o Immunofluorescence staining for SFTPC and LAMP3 identifies AT2 cells; AT1 cells are stained for HOPX demonstrating paucity of AT2 cells and reduced LAMP3 expression, $n = 3$ biological replicates. l, p Electron microscopy of E18.5 lung demonstrates absence of mature lamellar bodies in the *Prdm3/16^{ShhCreΔ/Δ}* AT2 cells, LB (lamellar body), M (mitochondria), $n = 3$ biological replicates. q Quantification of AT2, AT1, and AT2/AT1 cell numbers at E18.5 from 3 control and 3 *Prdm3/16^{ShhCreΔ/Δ}* fetuses. p -values were calculated using a 2-tailed Mann-Whitney test.

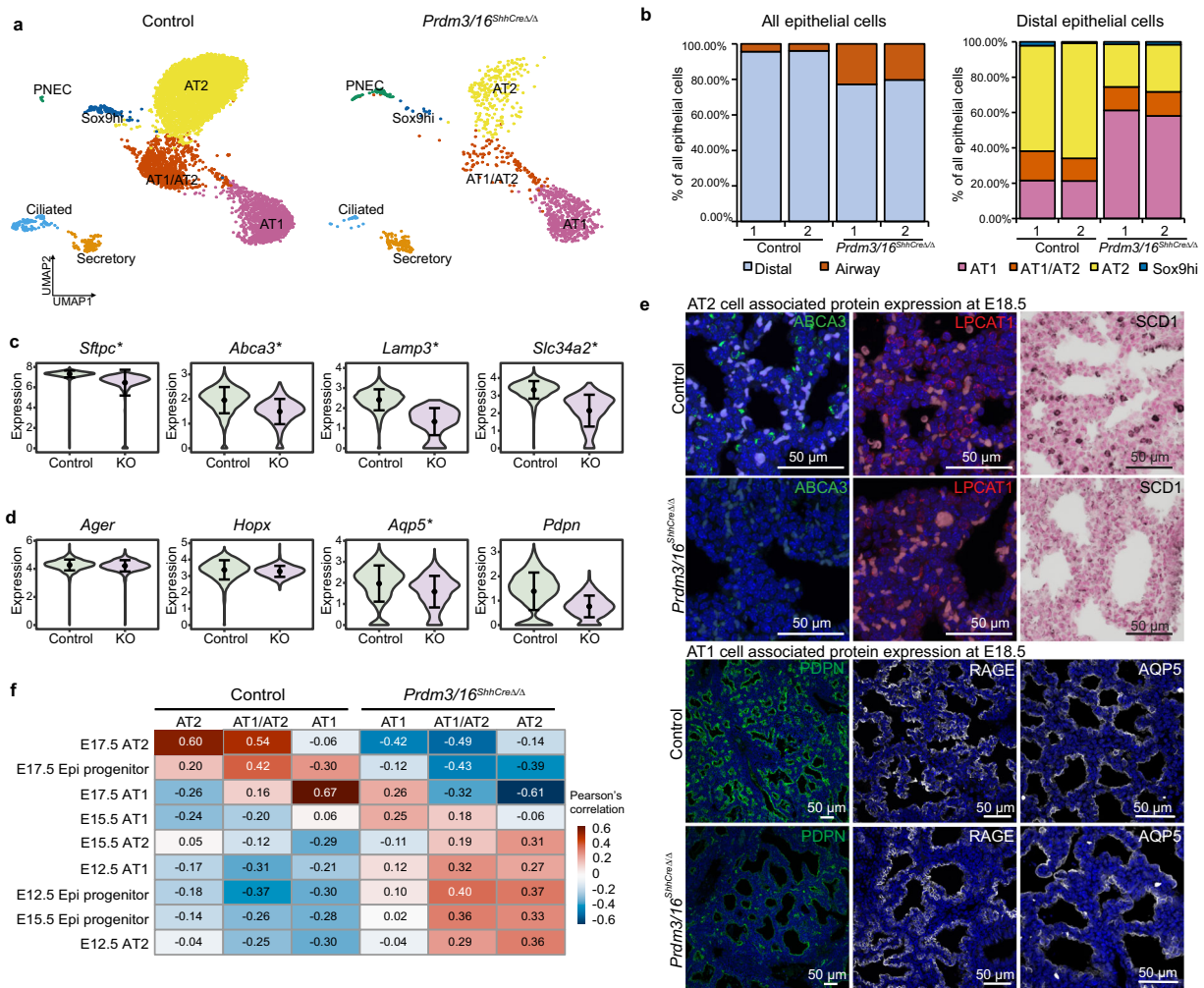


Fig. 2 | Single-cell RNA-seq (scRNA-seq) analysis of cellular and gene expression alterations in *Prdm3/16^{ShhCreΔ/Δ}* mouse lung at E18.5. **a UMAP plots of mouse epithelial cell subsets, $n = 2$ biological replicates. **b** Alterations in cell type proportions within all epithelial (left panel) and within distal epithelial (right panel) cells. scRNA-seq data of epithelial cells in **(a)** were used for the cell type proportion calculations. Violin plot visualization of representative AT2 **(c)** and AT1 **(d)** associated RNAs, black dots and error bars represent mean \pm SD; * identifies genes with statistically significant differences in expression as determined by a two-tailed**

Wilcoxon rank sum test $p \leq 0.05$, fold change ≥ 1.5 , and expression percentage $\geq 20\%$ of the cells. **e** Immunofluorescence staining of differentially expressed AT1 and AT2 genes in E18.5 lungs, $n \geq 3$ biological replicates of each genotype. **f** Pseudo-bulk correlation analysis with an independent mouse lung developmental time course scRNA-seq data (GSE149563) showing that alveolar epithelial cells from *Prdm3/16^{ShhCreΔ/Δ}* mouse lungs are most similar to cells from earlier developmental time points.

acinar regions of *Prdm3/16^{ShhCreΔ/Δ}* lung tubules as early as E14.5 (Fig. 1d, h).

PRDM3/16 regulates alveolar epithelial cell numbers and differentiation

To test the roles of PRDM3/16 in the regulation of epithelial cell numbers and gene expression, we performed single cell RNA sequencing on *Prdm3/16^{ShhCreΔ/Δ}* and control lungs at E18.5. Cell type assignments were made on the basis of interactive clustering and automated cell type annotations using the recently released LungMAP CellRef⁴⁰. We evaluated 17,584 cells comprising 29 distinct cell clusters expressing cell-type selective RNAs (Fig. 2a, Supplementary Fig. 4a). Integrated analysis demonstrated consistent cell type selective gene expression signatures in control and *Prdm3/16^{ShhCreΔ/Δ}* samples (Supplementary Fig. 4b), indicating that major cell identity genes were preserved in the absence of PRDM3/16. Consistent with the IF, scRNA-seq demonstrated a reduction in the proportion of AT2 cells and an increase in AT1 cells after the deletion of *Prdm3* and *16* (Fig. 2a, b).

While here we have prioritized the analysis of alveolar epithelial cells, these data indicated increased proportions of airway epithelial cell subtypes, including PNECs and secretory cells within the epithelial compartment of *Prdm3/16^{ShhCreΔ/Δ}* lungs (Fig. 2a, b). Despite the changes in the epithelial compartment, the distribution of cell types within the endothelial and mesenchymal compartments was unchanged (Supplementary Fig. 4c). We compared *Prdm3/16^{ShhCreΔ/Δ}* and control AT1 and AT2 cells and identified 198 (AT1) and 207 (AT2) differentially expressed genes (Supplementary Fig. 4d, Supplementary Data 1). Within the AT2 cell population, expression of genes involved in surfactant homeostasis (*Sftpc*, *Sftpd*, and *Abca3*), lipid biosynthetic processes (*Scd1*, *Fabp5*, *Pi4k2b*, *Lpcat1*, *Insig1*, *Lpin2*, *Abcd3*, and *Gpam*), lysosomal function (*Lamp3*), and transmembrane transport (*Atp6v1g*, *Slc31a1*, *Slc34a2*, *Cftr*, *Atp1b1*, and *Lcn2*) were all decreased. These data are consistent with decreased ABCA3, LPCAT1, SCD1, and LAMP3 protein detectable by IHC in PRDM3/16-deficient AT2 cells (Figs. 1o, 2c–e). Within mutant AT1 cells, we noted increased expression of genes associated with translation and regulation of cell migration,

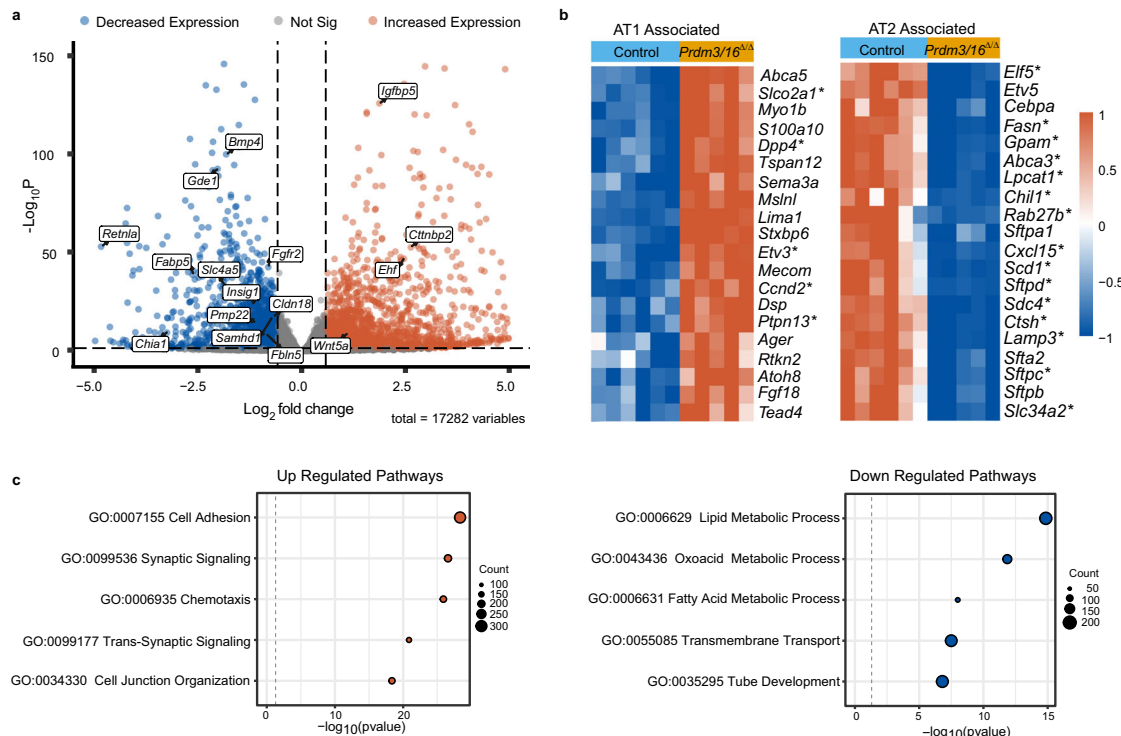


Fig. 3 | Loss of *Prdm3/16* influences cell differentiation. **a–c** Bulk RNA-Seq analysis of sorted EpCAM⁺ epithelial cells from E17.5 Control ($n = 6$) and *Prdm3/16*^{ShhCreΔ/Δ} (*Prdm3/16*^{Δ/Δ}) ($n = 5$) lungs. Wald test (default for DESeq2 r-package) was used for differential expression analysis utilizing standard cutoffs of $|\log_2$ fold change $| > 0.58$ and p -value < 0.05 . **a** Volcano plot showing 1438 genes with decreased expression and 2124 genes with increased expression, highlighting genes that are associated with epithelial cell development and mis-regulated in AT1 and AT2 cells.

b Heatmaps of normalized gene expression of AT1 and AT2 associated genes,

showing an increase in genes associated with AT1 cells and a decrease in genes associated with AT2 cells, a reflection of cell type population size. Asterisk (*) denotes statistical change in both bulk RNA-seq and single cell RNA-seq.

c Functional enrichment of gene sets with either increased expression (left panel) or decreased expression (right panel) using ToppFun and selecting highly enriched GO: Biological Processes. p -values determined by the hypergeometric probability mass function and reported as a Benjamini and Hochberg corrected FDR < 0.05 .

whereas downregulated genes were associated with cell adhesion, circulatory system development, and positive regulation of cell motility (Supplementary Fig. 4e). Expression of genes associated with AT1 cell identity and function was generally maintained, including *Ager* and *Hopx* (Fig. 2d). Expression of a subset of genes identified as NKX2-1 targets in AT1 cells was repressed in *Prdm3/16*^{ShhCreΔ/Δ} AT1 cells including *Cldn18*, *Aqp5*, *Pdpn*, *Samhd1*, *Pmp22*, *Gde1*, *Fln5*, *Slco3a1*, *Gnb4*, *Bmp4*, *Akap5*, *Lmo7*, and *Matn4* (Fig. 2d, Supplementary Data 1) suggesting incomplete terminal differentiation¹⁰. Protein expression mirrored these expression changes; immunofluorescence staining of PDPN was decreased, whereas the expression of AQP5 and RAGE was similar in control and *Prdm3/16*^{ShhCreΔ/Δ} animals at E18.5 (Fig. 2e).

Since morphological studies were consistent with pulmonary immaturity, we used correlation analysis to estimate the overall degree of immaturity within each cell lineage. We generated epithelial cell type specific gene expression profiles from pseudo-bulk analyses of single cell RNA profiles from WT and *Prdm3/16*^{ShhCreΔ/Δ} animals and compared them with pseudo-bulk analysis of single cell RNA profiles from an embryonic mouse lung developmental time course³⁹. RNA profiles from WT AT1, AT2, and AT1/AT2 cells were all highly correlated with time-matched cell-specific signatures from AT1^{ZepP}, AT2^{ZepP}, and AT1/AT2^{ZepP}. RNA profiles from *Prdm3/16*^{ShhCreΔ/Δ} AT2 cells were most correlated with those from E12.5 AT2^{ZepP}, demonstrating that lack of PRDM3/16 was sufficient to prevent progression of cell differentiation within the AT2 lineage (Fig. 2f). AT2 and AT1/AT2 cells consistently were more similar to immature rather than mature alveolar cells. RNA profiles from AT1 cells from *Prdm3/16*^{ShhCreΔ/Δ} pups were most concordant with E17.5 AT1^{ZepP}, though less correlated than in WT controls, confirming that AT1

gene expression was only mildly altered by absence of PRDM3/16 (Fig. 2f).

Strikingly, PRDM3/16 deletion caused E18.5 AT1/AT2 cells to most closely resemble the immature E12.5 and E15.5 SOX9 progenitor cells. These cells also continued to express low levels of *Sox9*. The immaturity of the *Prdm3/16*^{ShhCreΔ/Δ} AT2, AT1, and AT1/AT2 cells is also indicated by changes in gene expression shared between the cell types. We observed increased expression of genes previously reported⁴¹ to be downregulated during maturation of SOX9 progenitors including *Igf1p5*, *Pp1r14b*, *Peg3*, *Igf2*, *Cmas*, *Rbp1*, and *Tmsb10*, with concordant decreased expression of genes normally upregulated during progressive maturation including *Gde1*, *Matn4*, *Muc1*, *Nkd1*, *Mt1*, *Tspo*, *Ctsh*, *Brd7*, *Atp1b1*, *Tmbim6* and *Scd2*. Although the numbers of *Sox9* high progenitors are few at this stage of development, the *Prdm3/16*^{ShhCreΔ/Δ} population expressed higher levels of markers associated with proliferation (Supplementary Fig. 4f). Together, these observations imply a failure of distal progression within the SOX9 progenitor population during later stages of embryonic lung development, emphasizing the role of PRDM3/16 in driving cell fate acquisition and maturation.

Gene expression changes observed in the single cell analysis were reflected in bulk RNA obtained from EpCAM⁺ MACS-sorted E17.5 pulmonary epithelial cells from *Prdm3/16*^{ShhCreΔ/Δ} and control lungs (Fig. 3 and Supplementary Fig. 5a, Supplementary Data 2). Consistent with the scRNA and immunofluorescence data, genes associated with surfactant biosynthesis (*Sftpb*, *Sftpc*, and *Sftpd*), genes associated with surfactant homeostasis (*Ctsh*) and those associated with lipid metabolism (*Lpcat1*, *Abca3*, *Fasn*, *Scd1*) were significantly decreased in bulk epithelial RNA (Fig. 3b, c). Several genes not found to be differentially expressed in the single cell data but known to be required for normal

lung development and AT2 cell maturation were differentially expressed in the bulk RNA analysis including *Fgfr2* (-1.73)^{42,43}, *Cebpa* (-1.83)⁴⁴, *Etv5* (-2.82)⁴⁵, and *Wnt5a* (2.1)⁴⁶ (Fig. 3a). RNAs associated with the main AT1 cell program were increased, including *Ager*, *Rtnk2*, *Lima1*, *Tead4*, *Cttnbp2*, and *Msln1*, consistent with the increased number of AT1 cells (Fig. 3b); however, a number of genes associated with AT1 cells were downregulated in both the bulk and single cell data including *Fbln5*, *Bmp4*, *Slco3a1*, *Samhd1*, *Slc4a5*, and *Gnb4* (Fig. 3a). Taken together, analysis of both single cell and bulk RNA gene expression indicates cellular immaturity across the late embryonic lung epithelium, especially in AT2-like cells, in the absence of PRDM3 and PRDM16.

Chromatin landscape changes after loss of PRDM3/16

Because PRDM3 and 16 function as histone methyltransferases, we hypothesized that the differential gene expression and cellular immaturity found in *Prdm3/16*^{ShhCreΔ/Δ} fetuses was related to changes in chromatin organization. We used ATAC-seq to profile EpCAM⁺ sorted epithelial cells isolated from E17.5 control and *Prdm3/16*^{ShhCreΔ/Δ} lungs. Differential accessibility at 32,886 sites was identified (\log_2 fold change > 0.58 , p -value < 0.05), the majority of which annotated to intergenic or intronic regions (Fig. 4a, b, Supplementary Fig. 5b) at putative enhancer sites. Fewer differentially accessible sites were annotated to promoter regions (Fig. 4b). Homer analysis determined that peaks with increased accessibility after deletion of *Prdm3/16* were enriched in transcription factor binding sites, including FOXA1/A2, NKX2-1, and TEAD, while peaks that showed decreased accessibility were enriched in the predicted transcription factor binding sites for NKX2-1, FOS/JUN, AP-1, CEBP, and TEAD (Fig. 4c, Supplementary Data 3). The differentially accessible sites were also subjected to analysis using gkmSVM-R⁴⁷ producing AUROC scores of 0.943 indicating that the binding patterns in the open versus closed regions are different (Fig. 4d). The 500 largest positive and negative weighted kmers were identified and mapped to known transcription factor motifs (Fig. 4e). This analysis suggests that the regions with increased accessibility after deletion of *Prdm3/16* are enriched with the PRDM3 binding site indicating that PRDM3 may normally interact with those regions affecting the activity of other nearby transcription factors including FOXA1/A2 and GATA6. In contrast, the chromosomal regions closed in the absence of PRDM3/16 are more likely to be regulated by components of the FOS/Jun complex, AP-1, and NKX2-1.

Multiple changes in chromatin accessibility were adjacent to genes associated with functional enrichment annotations that were over represented in bulk and scRNA data from *Prdm3/16*^{ShhCreΔ/Δ} animals (Supplementary Fig. 5c); of the genes differentially expressed in both the bulk and single cell data, 72% were associated with nearby changes in chromatin accessibility (Supplementary Data 3). Notably, changes in accessible sites varied across a single genetic region with changes to both more open and more closed configurations when comparing WT and *Prdm3/16*^{ShhCreΔ/Δ} epithelium. Closure of the chromatin near transcriptional start sites (TSS) was observed in a subset of AT2 genes including *Cftr*, *Chia1*, *Fabp12*, *Lamp3*, *Napsa*, *Sftpd*, *Spink5*, *C3*, *Scd1*, *Atp6v1c2*, and *Retnla*. Chromatin accessibility was also altered near *Cebpa*, *Elf5*, *Fabp5*, *Slc34a2*, *Pi4k2b*, *Insig1*, *Lpin2*, *Abca3*, and *Fgfr2* (Fig. 4f, Supplementary Data 3). Increased chromatin accessibility was observed near the AT1 marker genes *Ager* and *Slco2a1*, consistent with increased expression at the RNA level (Supplementary Data 3).

To evaluate whether the observed changes in gene expression and chromatin accessibility were associated with changes in H3K4me3 methylation status, we performed CUT&RUN analysis on E17.5 EpCAM⁺ sorted epithelial cells isolated from control and *Prdm3/16*^{ShhCreΔ/Δ} lungs. Differential antibody binding was observed in 973 regions of which 614 (63%) mapped to gene promoters and 459 regions (42%) were annotated to genes identified in the bulk or single cell data as being differentially expressed. Decreases in H3K4me3 status were observed within the promoters of AT2 associated genes including *Lamp3*,

Lpcat1, and *Abca3* and AT1 associated genes (*Cldn18*, *Fbln5*, *Pdpm*) (Fig. 5a, b, Supplementary Data 4). The observed changes in H3K4me3 are in agreement with previously observed reductions in H3K4me3 levels in response to PRDM3 and 16 deficiencies²⁵. Taken together, our results suggest PRDM3 and PRDM16 activate and repress alveolar epithelial gene expression at least in part by altering the H3K4me3 methylation status modulating the accessibility of promoter and enhancer regions associated with the binding of lineage defining transcription factors including NKX2-1 and other co-factors such as CEBPA and TEAD.

PRDM16, PRDM3, and NKX2-1 share DNA binding sites

Since the AT2 signature genes whose expression was decreased in the *Prdm3/16*^{ShhCreΔ/Δ} mice were consistent with those regulated by NKX2-1⁴¹, and NKX2-1 binding sites were enriched at ATAC-seq peaks which were differentially accessible after *Prdm3* and *Prdm16* deletion, we hypothesized that PRDM3/16 functioned in alveolar epithelial cells via interaction with NKX2-1. We tested whether PRDM16, NKX2-1, and PRDM3 bound to shared genomic sites using CUT&RUN. Antibodies against PRDM16, NKX2-1, and PRDM3 were used to identify chromatin binding sites in bulk E17.5 EpCAM⁺ sorted epithelial cells. PRDM16 bound to 35,847 unique genomic regions, 22.5% near TSS (Fig. 6a, Supplementary Data 5). PRDM3 showed a similar distribution throughout the genome (Supplementary Fig. 6d, Supplementary Data 6). NKX2-1 bound to 52,716 unique genomic regions, 8,004 (~18.5%) of which were located near the promoter-TSS regions (Fig. 6b, Supplementary Data 7). Remarkably, PRDM16 and NKX2-1 bound together at ~40% of the identified peaks and PRDM16, NKX2-1, and PRDM3 coordinately bound at ~20% of the identified peaks. Motif enrichment using HOMER predicted common binding sites within the PRDM16, NKX2-1, and PRDM3 peaks for factors including TEAD, FOXA, NKX2-1, and SP1 (Fig. 6d, e; Supplementary Fig. 6e), sites also enriched within the identified ATAC-seq peaks. Binding of CTCF was also predicted within the PRDM16 and NKX2-1 peaks consistent with the idea that these factors play a role in modifying chromatin structure⁴⁸. To further verify these predictions, sequences within the PRDM16 and NKX2-1 annotated peaks were analyzed using gkmSVM-R⁴⁷ generating an AUROC score of 0.563 suggesting that NKX2-1 and PRDM16 are working through a similar mechanism. The 500 largest positive and negative weighted kmers were identified and mapped to known transcription factor motifs including TEAD, NKX2-1, FOXA, and SP1 (Supplementary Fig. 6a, b). The enrichment of TEAD binding sites within PRDM16 peaks was previously also observed in embryonic heart³³. Consistent with other PRDM16 Chip-seq data sets, the previously identified consensus binding sequence for PRDM16 was not identified within PRDM16, PRDM3, and NKX2-1 bound peaks^{27,33,49,50}.

We used CUT&RUN to identify regions bearing the chromatin mark H3K4me3 to identify regions of active transcription, and compared these regions to those bound with PRDM16, PRDM3, and NKX2-1. Analyzing only areas marked by H3K4me3, PRDM16 and NKX2-1 colocalized at 66% of peaks with 30.6% of these sites being located near the transcriptional start site (Fig. 6c). Chromatin binding sites in active regions shared by NKX2-1, PRDM16, and PRDM3 were identified near AT2 cell signature genes including *Sftpc*, *Sftpb*, *Cebpa*, *Abca3*, *Slc34a2*, *Scd1*, and *Lamp3*, genes previously shown to be bound and activated by NKX2-1 (Fig. 6f, Supplementary Fig. 6g, Supplementary Data 8)^{11,17,51}. Co-binding of PRDM16 and NKX2-1 was also observed near AT1 cell-associated genes, for example, *Rtnk2*, *Bmp4*, *Pmp22*, *Sem3b*, *Slco2a1*, *S100a10*, *Cldn18*, *Col4a2*, and *Fgfbp1* (Fig. 6g, Supplementary Fig. 6g, and Supplementary Data 8). Since PRDM16 and NKX2-1 shared occupancy at regulatory regions of target genes, we performed co-immunoprecipitation assays to evaluate direct binding. NKX2-1 and PRDM16 co-precipitated in lung tissue homogenates from E16.5 wild-type mice and after co-transfection in HEK293T cells, indicating a

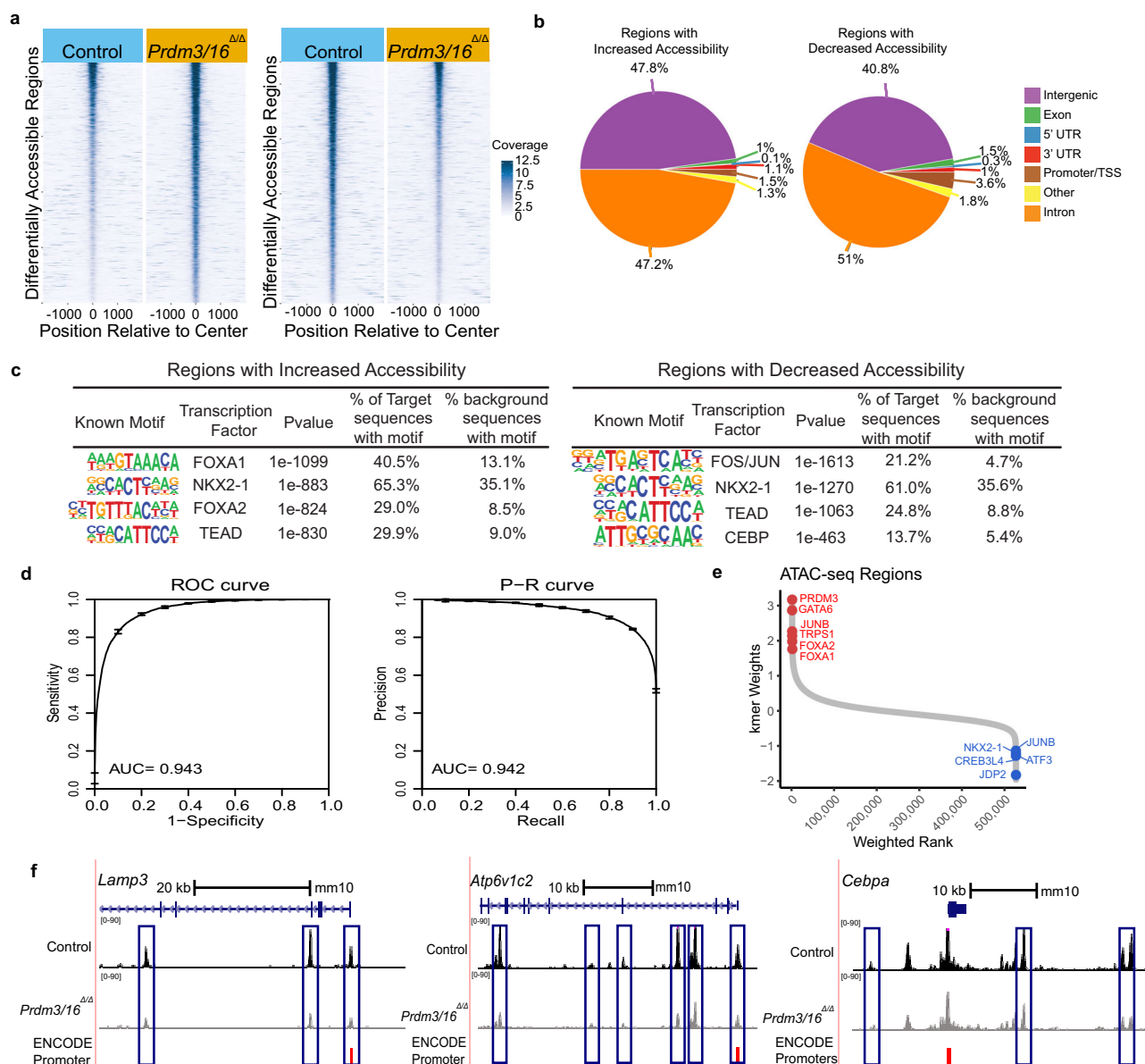


Fig. 4 | Changes in Chromatin Accessibility status following *Prdm3/16* deletion.

a Heatmaps of ATAC-seq data made with the R package tornado plot showing 5067 regions with increased chromatin accessibility (left panel) or 4577 regions with decreased chromatin accessibility (right panel) in representative individuals, $n = 6$ controls and $n = 5$ *Prdm3/16*^{ShhCreΔ/Δ} samples. Accessibility determined by differential accessibility analysis with R package DiffBind using a Wald Test with a $|\log_2$ fold change cutoff of 0.58 and a p -value < 0.05 . **b** Genomic distributions of each ATAC peakset, regions with increased or decreased accessibility, as annotated by HOMER annotatePeaks.pl. **c** Motif enrichment with HOMER searching either regions with increased or decreased accessibility, showing the putative transcription factors binding within these regions, p -value < 0.05 , determined by binomial test comparing frequency of a given motif under a set of peaks compared to a randomly

generated background. **d** ROC (Receiving Operating Characteristic) Curve and PR (Precision-Recall) Curves generated using gkmSVM to compare sequences under the open ATAC peaks trained against the closed ATAC peaks in *Prdm3/16*^{ShhCreΔ/Δ} epithelial DNA. An AUC near one indicates the model is able to distinguish differences in the sequences within the peak sets. Error bars show variability in model performance, depending on how the data set is divided into training and testing subsets. **e** The transcription factors associated with the highest 500 and lowest 500 weighted kmers identified under the peaks using gkmSVM were identified using Tomtom (FDR < 0.05) and were graphed. **f** The UCSC Genome Browser was used to visualize selected changes in promoter and enhancer chromatin accessibility observed in *Prdm3/16*^{ShhCreΔ/Δ} epithelial cells in differentially expressed genes associated with AT2 cell maturation.

potential direct interaction between NKX2-1 and PRDM16 (Fig. 7, Source data are provided as a Source Data file).

We constructed a gene regulatory network (GRN) serially integrating present data from epithelial bulk RNA-seq, scRNA-seq, PRDM3, PRDM16, NKX2-1, and H3K4me3 CUT&RUN, and NKX2-1 ChIP-seq data from Little et al.¹¹ First, we defined genes for the transcriptional regulatory network (TRN) input based on increased expression following PRDM3/16 deletion in AT1 (AT1 up regulated network) or decreased expression following PRDM3/16 deletion in AT1 (AT1 down regulated network) or AT2 cells (AT2 network). We filtered these gene sets for

those genes with neighboring chromatin regions containing PRDM3, PRDM16, and NKX2-1 binding sites by CUT&RUN or in published NKX2-1 ChIP-seq data and H3K4me3-marked active chromatin. The identified genes were used as input for IPA analysis to predict key upstream regulators of the AT1 and AT2 networks. Within the AT2 network CEBPA, ETV5, FGFR2, SREBF1, and STAT3 were predicted to have reduced activity following the deletion of *Prdm3/16*, while CBX5, ID3, and FOXA1 were predicted as activating factors in the NKX2-1/PRDM3/16 centered network (Supplementary Fig. 7a). The synergistic roles of CEBPA, ETV5, SREBF1, FOXA1/A2, and GATA6 with NKX2-1 in AT2 cell

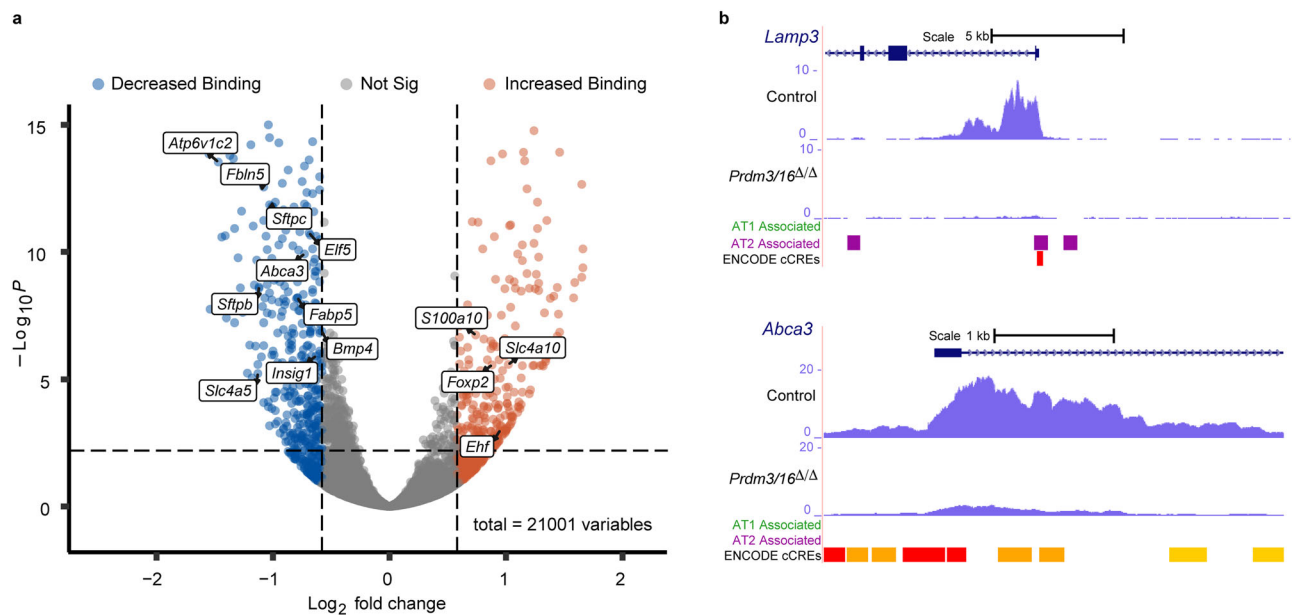


Fig. 5 | Decreased H3K4me3 binding in genes with reduced expression. **a** A Volcano plot of H3K4me3 differentially bound regions from CUT and RUN data identified 473 regions with decreased binding and 500 regions with increased binding in *Prdm3/16^{ShhCreΔ/Δ}* E17.5 epithelial cells compared to controls, ($n = 2$ biological replicates of each genotype). The indicated genes had paralleling changes in

gene expression. Differential expression was determined using a Wald Test (default for DESeq2 Rpackage) p -value < 0.05 , $|\log_2$ fold change > 0.58 . **b** The UCSC Genome Browser was used to visualize changes in promoter and enhancer binding by the H3K4me3 antibody in control and *Prdm3/16^{ShhCreΔ/Δ}* epithelial cells.

maturation, differentiation and surfactant homeostasis were previously demonstrated^{6,17,44,52,53}. Co-factors active after PRDM3/16 deletion including CBX5, ID3, and FOXA1, were predicted to regulate the “negative regulation of transcription by RNA polymerase II (GO:0000122)” and “negative regulation of cell differentiation (GO:0045596)”. The promoter sequences of two genes expressed in AT2 cells, *Sftpb* and *Abca3*, were scanned for binding motifs from the JASPAR and CIS-BP motif databases^{54,55}. Binding motifs for CEBPA, FOXA1/2, SREBF1, and STAT3 were identified in addition to NKX2-1 further confirming the network of predicted cofactors (Supplementary Fig. 7b). Within the *Sftpb* and *Abca3* promoters, the predicted cis-BP PRDM16 binding site was observed. This site was only identified in 15% of all the PRDM16 CUT&RUN peaks. Notably, within the *Abca3* promoter, the predicted cis-BP PRDM16 site was not directly associated with the identified CUT&RUN peaks suggesting that this sequence may not be the only mechanism through which PRDM16 regulates a genomic region. These analyses predict that PRDM3/16 functions with NKX2-1 to activate gene regulatory networks required for AT2 cell differentiation and surfactant homeostasis before birth.

The up-regulated AT1-associated gene network identified ARID1A, CTNBN1, HBEGF, and SMARCA4 as predicted co-factors active in the NKX2-1/PRDM centered network regulating AT1 cell gene expression (Supplementary Fig. 8a). NKX2-1, ARID1A, CTNBN1, and HBEGF are known to form complexes with SMARCA4^{56–59}. Both SMARCA4 and ARID1A are members of the SWI/SNF family involved in chromatin remodeling. ARID1A deficiency influences *Prdm16* expression in mouse embryo fibroblasts^{60,61}. The down-regulated AT1-associated network was comprised of the AT1 genes *Pdpr*, *Cldn18*, *Fbln5*, *Hs2st1*, *Phactr1*, and *Cxadr*, as well as, genes normally expressed at high levels in AT2 cells, such as *Napsa*, *Cldn18*, *Sftpd*, *Rbpjl*, *Ctsh*, and *Muc1* with an enriched function on surfactant balance (GO:0043129) and alveolar development (GO:0048286). In addition to SMARCA4 and ARID1A, NKX2-1, GATA6, KRAS, STAT6, and RXRA were identified as key upstream regulators (Supplementary Fig. 8b). Taken together, these data implicate PRDM3/16 as direct co-regulators with NKX2-1, binding cooperatively across the genome to drive chromatin organization and

define the AT1 and AT2 specification and differentiation programs during distal lung differentiation.

***Prdm3* and *Prdm16* are required for AT2 lineage fidelity during lung development**

The failure of AT2 cell differentiation in *Prdm3/16^{ShhCreΔ/Δ}* fetuses coupled with the observed early expansion of the AT1 expression domain and increase in AT1 cell numbers supported the hypothesis that PRDM3/16 loss in the AT2 lineage led to early transition of AT2-lineage cells to an AT1 cell fate. To directly test this hypothesis, we performed lineage analysis after AT2-lineage-specific PRDM3/16 deletion in *Prdm3/16^{SftpcCreERT2}; Rosa26^{lsl-TdTomato}* animals. Recombination was induced by the administration of tamoxifen to the dams at E12.5 and E13.5, and the lungs were analyzed at E18.5 (Fig. 8a). As previously reported³⁸, most control cells marked by *Sftpc^{CreERT2}* early in lung morphogenesis express only AT2 cell markers including SFTPC at E18.5 (Fig. 8b–i), with few cells expressing AT1 markers. In *Prdm3/16^{SftpcCreERT2}; Rosa26^{lsl-TdTomato}* fetuses (Fig. 8j–q), significantly more cells within the AT2 lineage (marked with TdTom) expressed HOPX alone (green arrowheads) or in combination with SFTPC (yellow arrowheads) while HOPX⁺SFTPC⁺TdTom⁺ AT2 cells (white arrowheads) were decreased. These data demonstrate that PRDM3/16 function in the AT2 lineage is required for the fate acquisition and lineage fidelity during embryonic distal lung specification (Fig. 8r–u, Source data are provided as a Source Data file).

Discussion

Respiratory function of the lung at birth requires interactions among a highly conserved network of signaling and transcription factors expressed by a diversity of cell types controlling lung morphogenesis and differentiation^{1,62}. Epithelial progenitors forming the embryonic respiratory tubules express the transcription factor NKX2-1 that distinguishes the initial foregut endoderm cells that are committed to pulmonary epithelial cell lineages^{1,4,38}. Thereafter, NKX2-1 is required for branching morphogenesis and differentiation of epithelial cells in conducting airways and alveoli, including specification of both

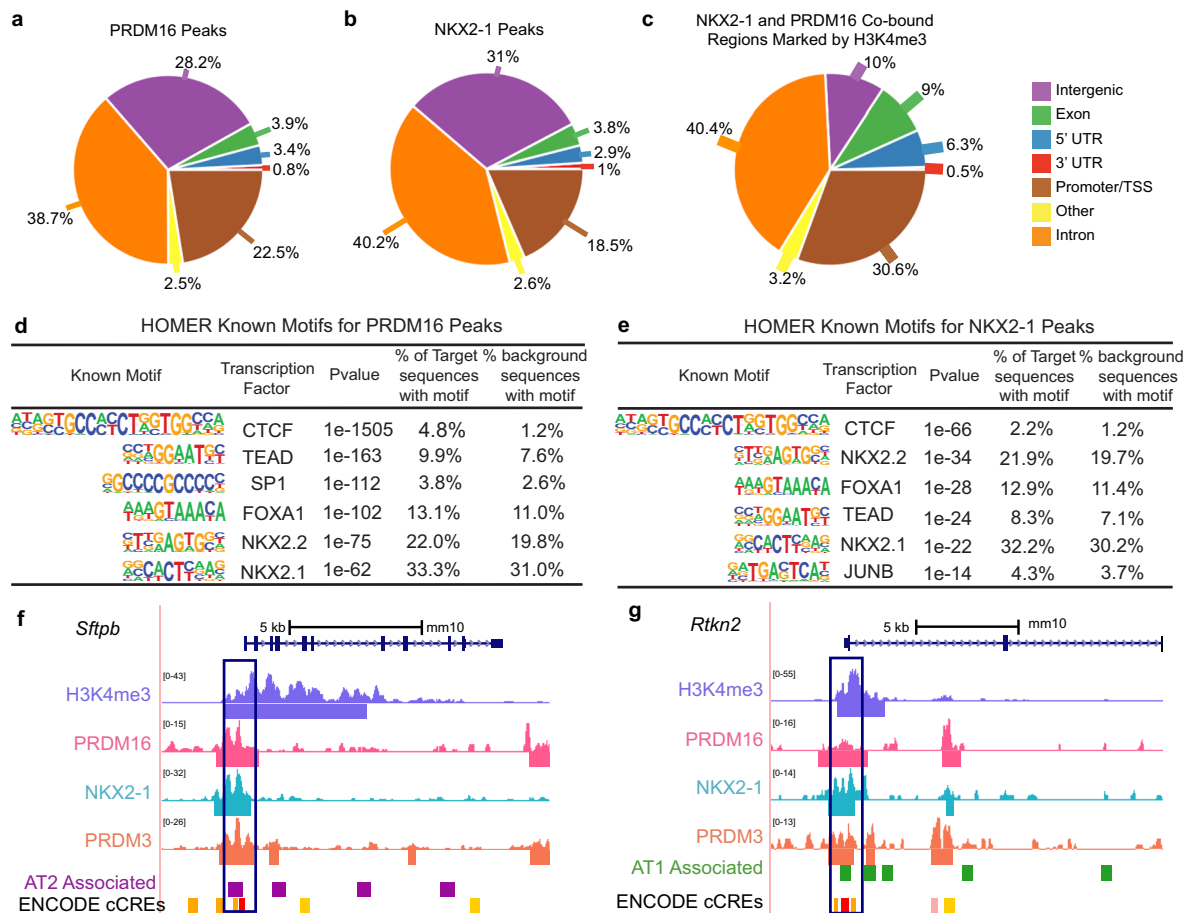


Fig. 6 | PRDM16, NKX2-1, and PRDM3 bind to shared sites throughout the genome and at promoters. a, b Genomic distributions of all called peaks from a representative PRDM16 binding experiment, ($n = 2$ replicates across two independent experiments)(a) and a representative NKX2-1 binding experiment, ($n = 3$ replicates across two independent experiments)(b) are shown. c Genomic distribution of the overlap of PRDM16 and NKX2-1 bound peaks that are marked by H3K4me3. **d, e** HOMER motif enrichment for all called peaks across the genome of a representative PRDM16 CUT&RUN (d) and a NKX2-1 CUT&RUN (e) experiment, p -

value < 0.05 , determined by binomial test comparing frequency of a given motif under a set of peaks compared to a randomly generated background. **f, g** Selected CUT&RUN gene tracks from H3K4me3, PRDM16, PRDM3, and NKX2-1 CUT&RUN experiments are visualized with the UCSC Genome Browser. AT1 and AT2 cell associated peaks were obtained from Little, et al.¹¹. ENCODE cCRE peaks are annotated from the ENCODE database of cis-regulatory elements. Binding is seen in both AT2 cell-associated genes (f) and AT1 cell-associated genes (g).

surfactant-producing AT2 cells and thin, gas exchange-promoting AT1 cells in the distal air sacs^{5,7,10,11}. Here we demonstrate that the histone lysine-methyl transferase proteins PRDM3 and 16 are required for AT1-AT2 cell fate choices. PRDM3/16 function together to promote AT2 fate acquisition and differentiation and participate in the regulation of surfactant lipid and protein expression necessary for lung function at birth. They also are required for full expression of the AT1 program, leading to accumulation of partially differentiated AT1 cells in PRDM3/16 mutants (Fig. 9). Transcriptomic and epigenomic studies demonstrated that PRDM3 and PRDM16 bind with NKX2-1 at predicted regulatory regions of their transcriptional target genes throughout the genome; without PRDM3 and PRDM16, chromatin accessibility at NKX2-1 bound promoter and enhancer sites is altered in AT2 and AT1 cell-associated genes. Thus, the present findings identify PRDM3 and 16 as epigenetic modulators of alveolar cell fate acquisition and specification required for full function of lung lineage transcription factors.

Loss of both *Prdm3* and *Prdm16* in alveolar epithelium caused surfactant deficiency at birth, consistent with the finding that both AT2 cell identity and maturation were inhibited in mutant lungs. The marked effect of PRDM3 and PRDM16 on AT2 cell lineage specification and differentiation supports the likelihood that PRDM3 and PRDM16 function in concert with other AT2-specific transcriptional co-

activators in addition to NKX2-1, a concept supported by the colocalization of predicted binding sites for other transcription factors identified by sequence analysis of the CUT & RUN experiments. For example, multi-omics and GRN inference analysis predicted that FOXA1/A2, FOS/JUN, ETV5, GATA6 and CEBPA are involved in this biological process, all of which are known to regulate a similar group of AT2 cell signature genes and to bind to regulatory regions of NKX2-1 target genes^{6,44,52,53,63}. Impaired commitment to the AT2 cell lineage was previously observed in animals deficient in epithelial FGFR2B signaling during the late pseudo glandular period^{42,43,64}. Chromatin accessibility at the *Fgfr2b* locus was altered in *Prdm3/16* deficient epithelial cells accompanied by a reduction in the expression of *Fgfr2b* and one of its known targets, *Etv5*, also known to be critical for AT2 cell specification^{53,64}. The similarity in phenotype between these studies targeting signal transduction and chromatin regulation is provocative, suggesting identification of the precise integration of intercellular signaling with activity of chromatin complexes in modulating activity of partner transcription factors is an area of high priority for future studies.

While PRDM3 and PRDM16 and NKX2-1 bound together at many AT1 cell associated genes, AT1 cell gene fate and gene expression were generally maintained after deletion of *Prdm3/16*, indicating that the gene networks required for AT1 cell lineage choice and gene

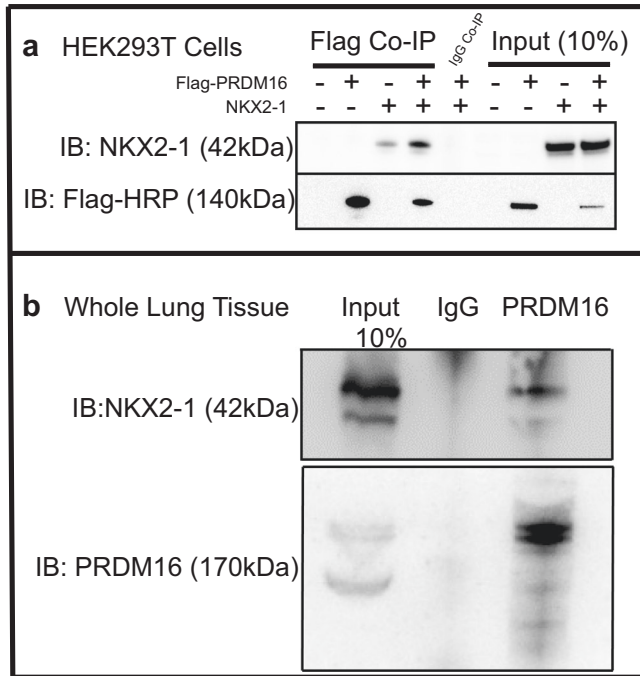


Fig. 7 | Protein interactions between PRDM16 and NKX2-1.

a Immunoprecipitation experiment demonstrates co-binding of FLAG-tagged PRDM16 and NKX2-1 after co-transfection in HEK293T cells, representative of 3 independent experiments. **b** Western blot analysis of E16.5 lung homogenates precipitated with anti-PRDM16 or IgG and immunoblotted with antibodies to PRDM16 and NKX2-1. Locations of the proteins are indicated, representative of 4 independent experiments.

expression were substantially intact. Therefore, we conclude that the AT1-centric network is likely maintained by regulators and co-activators which are distinct from that required for AT2 cell differentiation. IPA analysis based on single cell and bulk epithelial RNA from AT1 cells is distinct from that predicted for AT2 cells, suggesting important roles for PRDM3/16 and NKX2-1 and their interactions with chromatin modulators and co-activators including SMARCA4, ARID1A, CITED2, and HIF1 α . CITED2 is required for AT1 cell specification, activates SMAD2/3, directly interacts with HIF1 α , activates the *Cebpa* gene promoter in late gestation and is required for AT2 cell differentiation⁶⁵. ARID1A and SMARCA4 function together as members of the SWI/SNF chromatin regulator complex, suggesting that the interaction of PRDM/NKX2-1 and the SWI/SNF complex may be vital for sustaining the epigenetic landscape essential for the AT1 lineage. It appears that loss of PRDM attenuates this complex interplay; we note a compensatory elevation in the activity of SMARCA4 and ARID1A in PRDM3/16 knockouts, presumably as a homeostatic response to preserve the integrity of the AT1 cellular environment. In combination, this leads to increased expression of some AT1-specific genes and an expansion of the AT1 cell population, further underscoring the dynamic nature of the epigenetic regulation within AT1 cells. Notably, the ATAC and CUT&RUN data indicated that FOXA and TEAD transcription factors are prominently associated with PRDM16/NKX2-1 peaksets and have previously been implicated as primary transcription factors regulating AT1 cell maturation³⁹. We speculate that such interplay of epigenetic factors likely adds a distinct layer of regulation across the period of lung epithelial specification.

Finally, it is intriguing to note that lineage-specific deletion of PRDM3/16 led to partial lineage infidelity in the AT2 lineage during distal lung specification. Extensive recent literature has assessed the relationship between AT1 and AT2 lineages during alveolarization^{38,42,66,67}, with some data supporting a model of early

lineage specification by E14.5³⁸ and other data suggesting a population of AT1/AT2 cells retained in the late prenatal lung⁶⁷. In our study, loss of PRDM3/16 led to a decrease in the percentage of AT2 cells derived from early SFTPC-expressing epithelium at E12.5 from 80% to 60%, with concomitant increases in both HOPX⁺/SFTPC⁺ AT1 cells and HOPX⁺/SFTPC⁺ AT1/AT2 cells. Combined with data suggesting that AT1/AT2 cells in *Prdm3/16*^{ShhCre/Δ} animals were most similar to earlier progenitors, these data suggest that one key role of PRDM3/16 is to refine the lineage potential of distal lung epithelium. Therefore, functional chromatin remodeling via PRDM3/16 leads to restricted, refined lineages with appropriate quorum and fate identity, while loss of chromatin architecture prevents appropriate fate specification and leads to retention of more plastic, less mature, and less differentiated cells. Control of cellular plasticity is critical to both lung development and regeneration following lung injury, and we speculate that the definition of the epigenetic regulators, like PRDM3/16, which regulate plasticity versus differentiation may open novel therapeutic avenues for lung disease across the lifespan.

Methods

Animals

All animals were housed in specific pathogen-free facilities under the Institutional Animal Care and Use Committee protocol 2021-0053 approved by Cincinnati Children's Hospital and in the West China facility under the approval of the Institutional Animal Care and Use Committee at West China Second University Hospital, Sichuan University, protocol 2021-054. Mice were bred on a mixed background and housed with food and water ad libitum on a 14/10 (CCHMC) and a 12/12 (WCSUH) hour light/dark cycle with humidity levels between 40-60%. The *Prdm16*^{flx/flx} *tm1.1Brspl* (strain 024992)³⁷, *Shh*^{tm1(EGFP/cre/Cjt)} (strain 005622)⁶⁸, *Gt(ROSA)26Sor*^{tm9(CAG-tdTomato)/Hze} (strain 007909)⁶⁹, C57BL6/J (strain 000664) mice were acquired from Jackson Laboratories (Bar Harbor, ME). *Sftpc*^{CreERT2} mice were obtained from Dr. Harold Chapman⁷⁰. *Prdm3*^{flx/flx} mice were acquired from Dr. Kurokawa³⁶. For *ShhCre* breeding, *Prdm3*^{flx/flx} *Prdm16*^{flx/ut} *Shh*^{Cre+/ut} males were crossed with *Prdm3*^{flx/flx} *Prdm16*^{flx/flx} females to generate double knockout embryos, *Prdm3*^{flx/flx} *Prdm16*^{flx/flx} littermates were used as controls. For the lineage tracing study, *Prdm3*^{flx/ut} *Prdm16*^{flx/ut} *Sftpc*^{CreERT2+/ut} males were interbred with *Prdm3*^{flx/flx} *Prdm16*^{flx/flx} *ROSA26*^{tdTomato/Tomato} female mice; Cre positive, double heterozygous littermates (*Prdm3*^{flx/ut} *Prdm16*^{flx/ut} *Sftpc*^{CreERT2+/ut} *ROSA26*^{tdTomato/ut}) were used as controls. Animals were mated overnight, and the presence of a vaginal plug the next morning was defined as E0.5. To induce CreER activity, pregnant dams were given two doses of tamoxifen (Aladdin Scientific, T137974) dissolved in corn oil (20 mg/ml, 200 mg/kg) at E12.5 and E13.5 via oral gavage. Mice were euthanized by CO₂ followed by cervical dislocation or by the administration of a fatal dose of anesthetic followed by exsanguination. For all studies involving fetal animals, sex was not determined.

Histology and immunofluorescence analysis

Embryos were harvested and fixed in 2-4% paraformaldehyde (Electron Microscopy Sciences) in PBS at 4 °C overnight for analysis. Tissues were rinsed in PBS and either dehydrated through graded ethanols and into xylene prior to paraffin embedding and sectioning (5 μ m) or equilibrated in 30% sucrose/PBS prior to being embedded in OCT for cryosections (7 μ m). Paraffin sections were stained with Hematoxylin and Eosin (H&E). Immunofluorescence staining was performed on sections after heat mediated antigen retrieval in citrate buffer (pH 6.0). Primary antibodies used: Mouse anti-HOP (Santa Cruz Biotechnology sc-398703; 1:100), Guinea Pig anti-pro-SFTPC (In house #992; 1:500), Rabbit anti-ABCA3 (Seven Hills Bioreagents WRAB-70565; 1:100), Rabbit anti-NKX2.1 (Seven Hills Bioreagents WRAB-1231; 1:1000), Mouse anti-SOX2 clone E-4 (Santa Cruz Biotechnology SC-365823; 1:100), Rabbit anti-SOX9 (Millipore AB5535; 1:100), Guinea Pig anti-

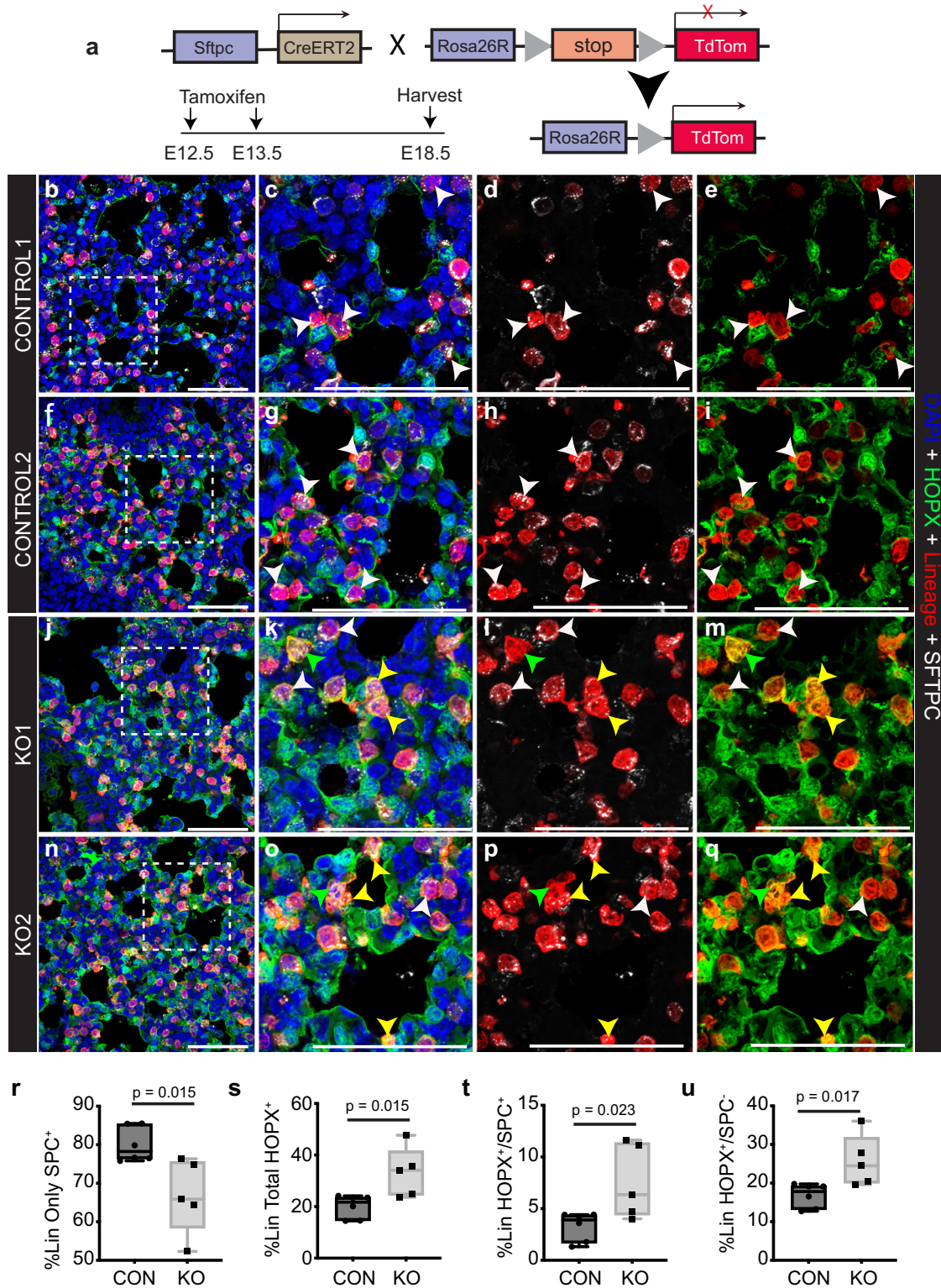


Fig. 8 | AT2-specific PRDM3/16 deletion leads to lineage infidelity during alveolar epithelial specification. **a** Dams were treated with tamoxifen at E12.5 and E13.5 to generate *Prdm3/16^{SftpcCreER+/Δ}; Rosa26^{loxP-tdTomato}* (Control) and *Prdm3/16^{SftpcCreERΔ/Δ}; Rosa26^{loxP-tdTomato}* (KO) fetuses which were harvested at E18.5 for lineage analysis. Control animals (**b–i**) demonstrated a majority of cells in the Sftpc-lineage (*Sftpc^{Lineage}*) were marked with only SFTPC protein by IHC (white arrowheads), while KO animals (**j–q**) showed significant decreases in SFTPC-only cells

(**r**), with corresponding increases in *Sftpc^{Lineage}* cells expressing HOPX (**s**), either with concomitant SFTPC expression (**t**) (yellow arrowheads in **l, m, p, q**) or without (**u**) (green arrowheads in **l, m, p, q**). *n* = 6 control and *n* = 5 KO lungs. Significance was determined by unpaired 2-tailed t-Test. The graphs are min point to max point showing all points with the center as mean and the box is 25–75%. All scale bars are 50 μm.

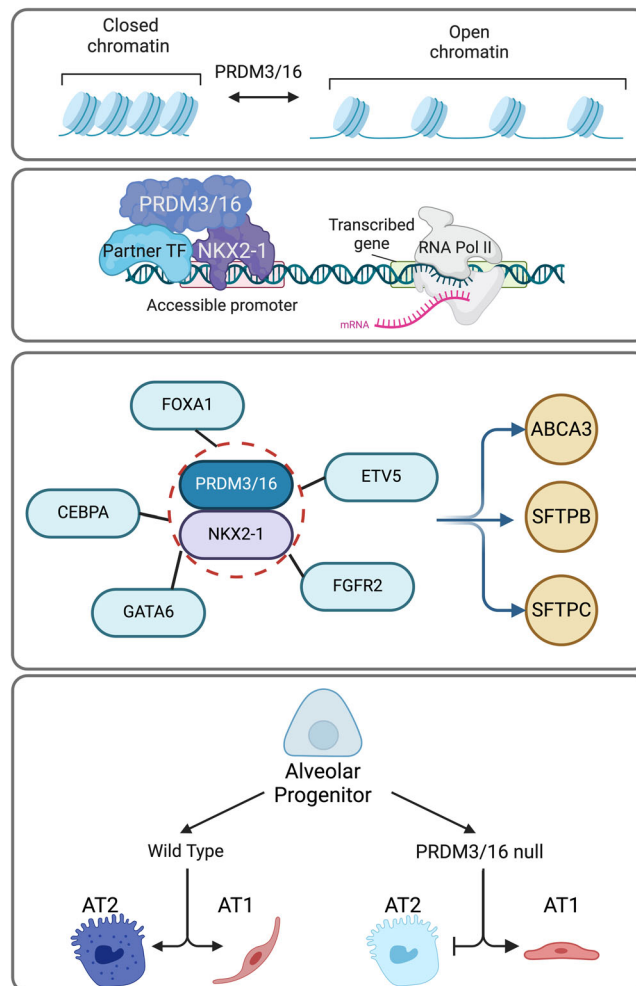


Fig. 9 | Model of the role of PRDM3/16 in alveolar development. PRDM3/16 participate in cell fate specification in the lung by modulating chromatin accessibility (top row) and by partnering with NKX2-1 and partner transcription factors to drive gene expression (second row) via a gene regulatory network required for terminal cell differentiation and surfactant expression in AT2 cells (third row).

Loss of PRDM3/16 activity in lung endoderm leads to reduced AT2 quorum, failure of AT2 surfactant function, and transition to an immature AT1 phenotype (bottom panel). Figure created with BioRender.com released under a Creative Commons Attribution-Non-commercial-NoDerivs 4.0 International license.

LPCAT1 (In house, 1:300)⁷¹, Guinea Pig anti-SCGB1A1 (in house #210, 1:500), Rabbit anti-Aquaporin5 (Abcam ab78486; 1:1200), Hamster anti-Podoplanin (DSHB 8.1.1: 1:200), Rat anti-RAGE (R&D systems mab1179; 1:100 no antigen retrieval), Guinea Pig anti-Lamp3 (Synaptic Systems 391005; 1:500), Sheep anti-PRDM16 (R&D systems AF6295; 1:100), Goat anti-endomucin (R&D systems AF4666; 1:1000), Rabbit anti-Pro SFTPC (Sigma AB3786; 1:400), Mouse anti-Calca (Sigma C7113; 1:200), Rat anti-SCGB3A2 (R&D Systems mAB3465; 1:200), Mouse anti-TRP63 (Santa Cruz sc-8431; 1:100), Rabbit anti-ARL13B (Proteintech 17711-1-AP; 1:200), Mouse anti-ASCL1 (BD Pharmingen 556604, 1:100), and Goat anti-Tdtomato (Origene AB8181; 1:500) (Supplementary Table 1). Samples were incubated with primary antibodies overnight at 4 °C, washed 3 times with PBS-TritonX100 (0.2%) prior to 1–2 hr incubation with appropriate secondary antibodies (1:200). Secondary antibodies used were the Thermo Fisher Scientific antibodies: Goat anti-mouse IgG1 Alexa Fluor 647 (A21240), Goat anti-Guinea Pig IgG Alexa Fluor 568 (A11075), Goat anti-Rabbit IgG Alexa Fluor 488 (A-11034), Goat anti-Rat IgG Alexa Fluor 647 (A21247), Goat anti-Rat IgG Alexa Fluor 555 (A-21434), Goat anti-Mouse IgG1 Alexa Fluor 568 (A21124), Donkey anti-Rabbit IgG Alexa Fluor 647 (A31573), Goat anti-Hamster IgG Alexa Fluor-488 (A21110), Donkey anti-Goat 647 (A21447), and Jackson ImmunoResearch antibodies Biotin conjugated Donkey anti-sheep (713-065-147), Alexa Fluor-488 Donkey anti-Guinea Pig IgG

(706-545-148), Alexa Fluor-647 Donkey anti-rat IgG (712-605-153), Alexa Fluor-488 Donkey anti-Rabbit IgG (711-545-152), Alexa Fluor-594 Donkey anti-Mouse IgG (715-585-151), Alexa Fluor-647 Goat anti-mouse IgG2c (115-545-208, 1:200), and Alexa Fluor-647 Donkey anti-Goat IgG (705-605-147). A Tyramide Biotin Signal Amplification (TSA) kit (Akoya/Perkin Elmer) was used for PRDM16 signal amplification in combination with Streptavidin, Alexa Fluor-488 (Life Technologies 532354; 1:200). SCD1 immunohistochemical staining was performed using Rabbit anti-SCD1 (Abcam 236868; 1:200 with TRIS EDTA pH 9.0 antigen retrieval), Goat Biotinylated anti-Rabbit IgG (Vector Labs BA-1000), the Vectastain Elite ABC Peroxidase kit (Vector Laboratories: PK-6100) and nuclear fast red post staining. Sections were imaged on Nikon AIR, Leica Stellaris 5, or Olympus IXploreSpin confocal systems with identical laser exposure between groups or a Nikon NI-E eclipse fluorescent widefield microscope. Images were minimally adjusted for LUTS. The E14.5 images for RAGE were denoised in the NIKON elements software.

For AT1 and AT2 cell quantification, six 60x confocal images taken from at least three different lobes at the distal margins were acquired from sections from 3 control and 3 *Prdm3/16*^{ShhCre/Δ} E18.5 embryos. Z-stack images comprised of 18 positions were denoised using Nikon Elements software. To reduce the background contributed by red blood cells, the signal from the TRITC channel was subtracted from the

FITC channel to create a new FITC-TRITC channel. Files were analyzed in IMARIS. Spot counting was performed with a nucleus size of 5 μm for DAPI and NKX2.1⁺ counting. A 6 μm area size was used for HOPX staining. Automatically counted DAPI⁺, NKX2.1⁺, and HOPX⁺ spots were manually evaluated, and corrections made as necessary. The presence of SFTPC was manually scored. The number of AT2 cells was determined by filtering SFTPC⁺ cells with the shortest distance to DAPI and NKX2.1 and the number of AT1 cells by filtering HOPX⁺ cells with the shortest distance to DAPI and NKX2.1. AT1/AT2 cells were DAPI⁺/NKX2.1⁺/SFTPC⁺/HOPX⁺ as defined by IMARIS and manually confirmed/corrected. For the lineage tracing study, six to eight 60x confocal images were taken from at least three different lobes from $n = 6$ control and $n = 5$ KO lungs. SFTPC, tdTomato, and nuclear HOPX signals were counted manually using the “Cell Counter” plug-in in Fiji software.

Single cell RNA sequencing

E18.5 lungs were dissected, and extrapulmonary tissue removed from embryos isolated from two litters. Samples were stored on ice in HypoThermosol FRS preservative solution (Stem Cell Technologies, 07935) until genotypes were determined. 2 control and 2 *Prdm3/16^{ShhCre/Δ}* mutant samples were generated by mincing two lungs of the same genotype together using a razor blade for 2 minutes. 25 mg of minced tissue was digested and processed as outlined in a step-by-step protocol found at <https://www.protocols.io/view/adult-mouse-lung-cell-dissociation-on-ice-ymgfu3w>. Cells were counted using a hemocytometer in the presence of trypan blue, and the concentration was adjusted to 1000 cells/ μl prior to separation using the Chromium 10X Platform. Cell separation, cDNA synthesis, and library preparations were performed by the Cincinnati Children’s Hospital Single Cell Genomics Facility (RRID:SCR_022653) and the libraries generated were sequenced by the Cincinnati Children’s Genomics Sequencing Facility (RRID:SCR_022630). The number of reads obtained were Con-1 (423,140,305), Con-2 (346,159,344), Mut-1 (411,512,209), and Mut-2 (452,966,500).

Transmission electron microscope

Lungs isolated from 3 control and 3 *Prdm3/16^{ShhCre/Δ}* E18.5 embryos from different litters were fixed in EM fixation buffer (2% glutaraldehyde, 2% paraformaldehyde (Electron Microscopy Sciences), 0.1% calcium chloride in 0.1M sodium cacodylate buffer pH 7.2) at 4 °C overnight. Lung lobes were cut into 1-2 mm blocks and processed as previously described⁷². Images were acquired by an H7650 transmission electron microscope (Hitachi High Technologies).

RNAscope

RNAscope was performed on paraffin-embedded 5 μm tissue sections using the RNAscope Multiplex Fluorescent assay v2 according to the manufacturer’s instructions (Advanced Cell Diagnostics). Antigen retrieval was performed in a boiling water bath for 15 minutes. Sections were digested with Protease Plus for 15 minutes at 40 °C prior to hybridization with a probe for *Prdm3* (Mm-Mecom 432231 lot 22067 A) and detected with Opal 570 (Akoya OP-001003 Lot 20212821, 1:500 incubated 45 minutes). Sections were post-stained with DAPI and imaged on a Nikon A1R-inverted confocal microscope. Assay was performed on multiple sections from single embryos at E11.5, E12.5, E14.5 and E16.5. Adjacent sections were stained with NKX2-1 and PRDM16.

Isolation of bulk epithelial cells with magnetic-activated cell sorting

Lungs from E17.5 mice were minced into 1 mm³ pieces and incubated in RPMI media (Gibco, 11875-093) containing Liberase TM (Roche, 540111901, 50 $\mu\text{g}/\text{ml}$) and DNase I (MilliporeSigma, 100 $\mu\text{g}/\text{ml}$) at 37 °C for 30 minutes. Tissues were transferred to C-tubes (Miltenyi) and

dissociated with the GentleMACS cell dissociator. Single cells were subjected to RBC lysis buffer for 2 minutes on ice (eBiosciences, 00-4333-57), rinsed with MACs buffer (Miltenyi), and incubated with FcR blocking reagent (Miltenyi, 130-092-575) on ice for 10 minutes, followed by incubation with CD326 anti-Mouse-EpCAM microbeads (Miltenyi, 130-105-958) at 4 °C for 15 minutes. Cells were then passed through LS columns for positive selection and collected in MACS buffer. Cells were quantified manually under a hemocytometer with trypan blue before downstream experiments.

Bulk RNA-seq, ATAC-seq, and CUT&RUN library construction

Following the EpCAM MACS sorting, 200,000 epithelial cells were aliquoted for RNA isolation with RNeasy Micro Plus kit (Qiagen, 74034) according to the manufacturer’s protocol. RNA samples were sent to Genewiz for polyA enrichment, library construction, and sequencing.

ATAC-seq libraries were constructed from 50,000 EpCAM⁺ epithelial cells isolated from 6 control and 6 *Prdm3/16^{ShhCre/Δ}* E17.5 lungs. The nuclei were isolated using lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630) and then subjected to a transposition reaction with Illumina’s Tagment DNA TDE1 Enzyme and Buffer Kit (Illumina, 20034197) at 37 °C for 30 minutes. Transposed DNA fragments were quantified by qPCR and amplified with barcoded primers using the NEBNext High-Fidelity 2X PCR master mix⁷³. The libraries were sent for sequencing by Genewiz and were sequenced on Hiseq4000 with PE150 mode for at least 50 million reads per sample.

The CUT&RUN experiments were performed using the Cut and Run Assay Kit (Cell Signaling Technology, 86652S). EpCAM⁺ epithelial cells from at least two pooled embryos were isolated from *c57Bl/6*, *Prdm3/16ShhCre* or *Prdm3/16ShhCre^{Δ/Δ}* E17.5 mouse lungs and ~100,000 cells were bound to concanavalin A beads and then incubated overnight with antibodies for PRDM16 (Sheep anti-PRDM16 R&D AF6295; 1:50 or Rabbit anti-PRDM16 1:100 kindly provided by Dr. Patrick Seale²⁹), Rabbit anti-NKX2-1 (Seven Hills Bioreagents WRAB-1231, 1:100), Rabbit anti-PRDM3 (Cell Signaling Technology 2593, 1:50), Rabbit anti-H3K4me3 (Cell Signaling Technology 9751, 1:50), or rabbit IgG or Sheep IgG for each reaction. For the PRDM16 analysis the pooled *Epcam⁺* population from *c57Bl/6* embryos was split into two samples and incubated with the different PRDM16 antibodies. The cells were incubated with pAG-MNase to facilitate antibody-guided DNA digestion. DNA was purified by phenol-chloroform-isoamyl alcohol (Thermo Fisher Scientific, 15593-031) extraction using Phasemaker tubes (Thermo Fisher Scientific, A33248) and ethanol precipitated. DNA was quantified using a Qubit and the dsDNA Quantification Assay Kit (Thermo Fisher, Q32850) or the Pico-green Assay Kit (Thermo Fisher, P7589). Libraries were constructed using the DNA Library Prep Kit for Illumina Systems (Cell Signaling Technology, 56795S) and Multiplex-oligos for Illumina Systems (Cell Signaling Technology, 47538S). Libraries were sequenced by Genewiz on a Hiseq4000 with PE150 mode for at least 15 million reads per sample.

Co-immunoprecipitation

HEK293T (2.8×10^5) (ATCC, CRL-3216) cells were plated in 6-well plates and transfected with MSCV-flag-PRDM16 (Addgene, 15504) and/or pCDNA3-NKX2-1⁷⁴ using Lipofectamine 3000 (Thermo Fisher Scientific, L3000075). Cells were lysed 48 hours post-transfection with Dynabeads Co-Immunoprecipitation kit lysis buffer (Thermo Fisher Scientific, 14321D) supplemented with 150 mM NaCl, 2 mM MgCl₂, and Complete, Mini Protease Inhibitor Cocktail (Roche, 11836153001) and PhosSTOP (Roche, 4906837001). Clarified lysates (~800 μg) were pre-cleared with Protein A/G PLUS-agarose (Santa Cruz Technology, sc-2003) and non-specific IgG for 45 minutes, before incubating overnight with EZview Red Anti-FLAG M2 Affinity Gel (Sigma-Aldrich, F2426) at 4 °C. EZview Red Protein G Affinity Gel (Sigma-Aldrich, E3403) was used for negative control samples.

Samples were washed with lysis buffer followed by 1X Tris-Buffered Saline buffer as per manufacturer's protocol. Samples were eluted by boiling in Laemmli buffer with 5% β -Mercaptoethanol and electrophoresed using 4–12% NuPAGE Bis-Tris gels (Thermo Fisher Scientific, NP0321BOX) prior to Western blotting. Blots were incubated with Rabbit mAb anti-DYKDDDK Tag (HRP Conjugated) (Cell Signaling Technology, 86861, 1:1000) or NKX2-1 antibody (Seven Hills Bioreagents, WRAB-1231, 1:1000) followed by incubation with True-Blot HRP-Conjugated secondary antibody (Rockland Immunochemicals Inc. 1:1000, 18-8816-33). Chemiluminescence detection was performed using Luminata Forte Western HRP substrate (MilliporeSigma, WBLUF0100) and a ChemiDoc Touch Imaging System (BioRad). Results shown are representative from at least three independent experiments performed.

For endogenous Co-IP experiments, 6 to 8 lungs harvested from WT E16.5 embryos were pooled and lysed in RIPA buffer (Beyotime Biotechnology, P0013D) containing protease inhibitor cocktail (Abcam, ab271306). Lysates were precleared with Protein G Dynabeads (Thermo Fisher, 10003D) for 1 h and then aliquoted and incubated with 2 μ g PRDM16 antibody (RD systems, AF6295) or the sheep IgG isotype (Sigma, I5131) overnight at 4 °C with rotation. Samples were incubated with Protein G Dynabeads at room temperature for 1 h. Beads were washed with RIPA buffer and boiled with 1X SDS-PAGE Protein Sample Loading Buffer (Beyotime Biotechnology, P0287). Protein eluents were electrophoresed in SDS-PAGE gels and blots were incubated with antibodies raised against NKX2-1 (PTMbio, PTM-5105) and PRDM16 (R&D systems, AF6295), followed by incubation with HRP-conjugated rabbit or sheep secondary antibodies (Jackson ImmunoResearch, 711-035-152 and 713-035-147). Chemiluminescence detection was performed using BeyoECL Plus reagent (Beyotime Biotechnology, P0018) and a G:BOX Chemi system (Syngene). Results shown are representative of four independent experiments.

Analysis of 10x single cell RNA-seq of control and *Prdm3/16*^{ShhCre/Δ} mouse lung

Sequencing reads from each single cell RNA-seq (scRNA-seq) sample were preprocessed, aligned, and quantified using the Cell Ranger pipeline (version 6.1.2, 10x Genomics) with mm10 mouse reference genome (refdata-gex-mm10-2020-A, 10x Genomics). For control and *Prdm3/16*^{ShhCre/Δ} samples, the following cell filtering quality control (QC) steps were performed. EmptyDrops⁷⁵ was used to identify cells (>100 unique molecular identifiers [UMIs] and false discovery rate <0.01) which deviated from empty droplets. Cells with 1000–8000 expressed (UMI > 0) genes, less than 100,000 UMIs, and less than 10% of UMIs mapped to mitochondrial genes were kept. Doublet cells in each sample were predicted and removed using Scrublet. In total, 18,084 cells from the four scRNA-seq samples passed the QC filters and were used for downstream analysis. Potential ambient RNA contamination in gene expression of each sample was corrected using SoupX⁷⁶ with automatically estimated contamination fraction rates. SoupX-corrected data from individual samples were integrated using Seurat (version 4.1.0)⁷⁷ reciprocal principal component analysis (RPCA) pipeline with SCTransform normalization. Cell clusters were identified using the Leiden clustering algorithm⁷⁸. An automated cell type annotation was performed using SingleR⁷⁹ using LungMAP mouse lung CellRef⁴⁰ as the reference. Results from the clustering analysis and the cell type annotation analysis were combined, and the final cell types were defined at the cluster level. Four clusters with co-expression of red blood cell genes (*Cypa*, *Alas2*, and *Bpgm*) and mesenchymal, epithelial, and endothelial cell markers were considered as contaminated clusters. Among the four scRNA-seq samples, the Con-1 sample has more cells (11,362 cells) than the other three samples (Con-2, 1724; Mut-1, 2257; Mut-2, 2241). To assess the robustness of our cell type identification with regard to this difference in the cell numbers, we randomly sub-sampled the Con-1 data from 11,362 cells to

2257 cells (the number of cells of the second largest sample), calculated the epithelial cell type distributions within the sub-sampled Con-1 data, re-integrated the sub-sampled Con-1 data with the other scRNA-seq, and assessed the cell type separation in the re-integrated UMAP. The sub-sampling and re-integration tests were repeated 10 times. The results showed that the epithelial cell type proportions and separations in the Con-1 sample are robust to the sub-sampling and re-integration analyses. Cell type annotations were validated using the expression of cell type marker genes from LungMAP CellCards and CellRef⁴⁰. Differential expression analysis was performed using the Seurat (v4.1.0) FindMarkers function with the Wilcoxon rank sum test. Differentially expressed genes were defined based on the following criteria, including adjusted *p*-value of Wilcoxon rank sum test <0.1, fold change \geq 1.5, and expression percentage \geq 20%.

For pseudo-bulk correlation analysis, gene expression matrix and cell type annotation of 10x single cell RNA-seq of mouse lung time course³⁹ were downloaded from Gene Expression Omnibus, accession number GSE149563. Data from alveolar epithelial cell populations (AT1, AT2, and Epi progenitor cells) from E12.5, E15.5, and E17.5 were used for pseudo-bulk correlation analysis with the present scRNA-seq data of AT1, AT1/AT2, and AT2 cells from control and mutant mouse lungs at E18.5. Gene expression was normalized using Seurat Log-Normalize function. Pseudo-bulk expression profiles for each cell population were calculated using Seurat AverageExpression function. Highly variable genes (HVGs) among pseudo-bulk profiles of each dataset were identified using Seurat FindVariableFeatures function. HVGs common in both datasets were used; then, the pseudo-bulk expression of those HVGs was z-score-scaled within each dataset. Pearson's correlations of pseudo-bulk profiles were calculated using z-score-scaled expression of HVGs in the pseudo-bulk profiles.

Bulk RNA-seq

Preprocessing of data from the EpCAM⁺ sorted cells was done using TrimGalore (0.4.2) and Cutadapt (1.8.1) to remove Illumina adapters. Quality was assessed with FastQC. Alignment was done with Bowtie (2.4.2) using the mm10 reference genome and the options '-end-to-end' and '-sensitive.' Low-quality alignments were removed with Samtools (1.10.0), and PCR duplicates were removed with Picard (2.18.22) MarkDuplicates. RNA-seq raw reads were aligned to the mm10 reference genome with Bowtie2. Raw count matrix was generated with HTSeq-count (2.0). DESeq2 (1.36.0) was used for differential expression analysis, with cutoffs of *p* < 0.05 and |log₂ fold change| > 0.58. Heatmaps were generated with the pheatmap (1.0.12) R package, and the volcano plot was generated with EnhancedVolcano (1.14.0) R package. GO term enrichment was conducted using the lists of up and down differentially expressed genes as input on the ToppFun website. Dot plots were generated in R with ggplot2 (3.4.0). Representative pathways were selected based on both *p*-value and inclusion of classic AT1 or AT2 marker genes.

Bulk ATAC-seq

Data files from EpCAM⁺ sorted cells were processed following the standard ENCODE pipeline. For data quality of the ATACseq, fastq files were evaluated with FastQC and trimmed of adapters using Cutadapt (1.9.1) and aligned with Bowtie2 (2.2.6) to the mm10 reference genome. Mitochondrial low-quality reads (Samtools (1.7) view -F 1804), and duplicates (Picard 1.126) were filtered out. All samples had a sufficient percentage of uniquely mapped reads (>70%) and PCR bottleneck coefficient of > 0.9, per ENCODE standards. Fragment inside peaks (FRiP) values were calculated for each individual, only those with FRiP > 0.1 were used in downstream analysis. The IDR conservative peakset called by ENCODE pipeline defaults (MACS, *p*-value < 0.01), along with the trimmed, sorted, deduplicated, no mitochondrial DNA bam file, were used for a DiffBind (v3) differential accessibility analysis, using full library

normalization. A $|\log_2 \text{fold change}| > 0.58$ and $p\text{-value} < 0.05$ were used to determine which regions showed increased or decreased accessibility under KO conditions. Statistically significant regions for both conditions were annotated with HOMER 'annotatePeaks.pl', and 'findMotifsGenome.pl' was used to determine enriched motifs. The gkmSVM R package⁴⁷ was utilized to generate models to evaluate the sequences under the peaks of the top 10,000 peaks based on p -value as generated by MACS2 callpeaks. Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves were generated to evaluate the performance of a given model. Ten basepair kmers were weighted by the model to find which ones were strong indicators of a transcription factor binding site. The 500 kmers with the highest weights and the 500 kmers with the lowest weights were compared to the Jaspar2022 vertebrate transcription factor binding motif database with the MEME Suite Tomtom tool to identify potential binding motifs represented by the weighted kmers. Only motifs with an FDR < 0.05 were considered.

CUT&RUN

Fastq files were evaluated with FastQC and trimmed with Trimmomatic to remove low quality reads. All samples had a sufficient percentage of uniquely mapped reads ($> 70\%$) and PCR bottleneck coefficient of > 0.9 as an indication of high library complexity, per ENCODE standards. Fragment inside peaks (FRiP) values were calculated for each individual, only those with FRiP > 0.1 were used in downstream analysis. Peaks that were present in multiple biological replicates or independent experiments were kept. The CUT&RUNTools package was used for initial fastq file processing, with default settings for alignment, most notably the '--dovetail' option in Bowtie2 (2.4.5). The MACS (2.1.4) package was used for peak calling with the following parameters: '-p 0.01 --keep-dup all', using the '-c' option with an IgG control for each antibody; a p -value of 0.001 was used to call the H3K4me3 peaks. The narrow peak setting was used for all peak calls. Homer (4.11) 'mergePeaks' function was used to generate consensus peak sets across datasets. Homer (4.11) functions 'annotatePeaks.pl' and 'findMotifs.pl' were used to find the nearest genes and enriched motifs. The 'bdgcmp' function of MACS2 was used to normalize bedgraph files with fold enrichment.

Genetic regulatory network construction

The bulk RNA-seq, scRNA-seq; PRDM16, NKX2-1, and Hek4me3 CUT&RUN data, and the NKX2-1 ChIP-Seq data sets¹¹ were used for network generation. Genes passing the following criteria were used for AT1 GRN construction: 1) gene expression increased in EpCAM+ sorted epithelial cells of *Prdm3/16^{ShhCreΔ/Δ}* vs. control ($|\log_2 \text{fold change}| > 0.58$, $p\text{-value} < 0.05$); 2) genes present in control AT1 cells ($> 15\%$) and percentage expression is higher in *Prdm3/16^{ShhCreΔ/Δ}* AT1 cells vs. control AT1 cells; and 3) genes with positive binding of PRDM3, PRDM16, NKX2-1 and H3K4me3 transcriptional activation marks based on the CUT&RUN data. In contrast, genes passing the following criteria were used for AT2 GRN construction: 1) gene expression decreased in EpCAM+ sorted epithelial cells of *Prdm3/16^{ShhCreΔ/Δ}* vs. control ($|\log_2 \text{fold change}| > 0.58$, $p\text{-value} < 0.05$); 2) genes present in control AT2 cells ($> 20\%$) and percentage expression was lower in *Prdm3/16^{ShhCreΔ/Δ}* AT2 cells vs. control AT2 cells; and 3) genes with positive binding of PRDM3 and 16 and NKX2-1 based on the Cut & Run data. The resulting gene lists were subjected to Ingenuity Pathway Analysis (IPA) upstream regulator analysis. Activation and inhibition of key factors were determined by correlations between the observed expression changes of the target genes when compared to literature knowledge. Transcriptional Regulatory Networks (TRN) were generated to reveal the potential biological interrelationships among the key upstream regulators and their target genes. IPA was used to predict inhibited or activated upstream regulators based on the gene expression changes of the input genes.

Statistics

No statistical method was used to predetermine sample sizes. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment. GraphPad Prism software was used for analysis and graph plotting. The statistical methods are supplied in figure legends and methods sections.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The ATAC-seq, CUT&RUN, bulk RNA-seq, and single cell RNA-seq data generated in this study have been deposited in the NCBI Gene Expression Omnibus with accession number [GSE250366](https://doi.org/10.1101/2024.04.30.2024). Publicly available data set [GSE149563](https://doi.org/10.1126/science.abc3172) with the <https://doi.org/10.1126/science.abc3172> and set [GSE158205](https://doi.org/10.1038/s41467-021-22817-6) with the <https://doi.org/10.1038/s41467-021-22817-6> were also used. UCSC Genome browser tracks: H3K4 comparison data in Fig. 5 can be accessed at https://genome.ucsc.edu/s/Ashley.Kuenzi%40cchmc.org/cut_run_04302024. ATAC, PRDM16, NKX2-1, and PRDM3 data in Figs. 4, 6, and Supplementary Figs. 6 and 7 can be accessed at https://genome.ucsc.edu/s/Ashley.Kuenzi%40cchmc.org/TFBS_predict_12072023. Source data are provided with this paper.

References

- Whitsett, J. A., Kalin, T. V., Xu, Y. & Kalinichenko, V. V. Building and regenerating the lung cell by cell. *Physiol. Rev.* **99**, 513–554 (2019).
- Sun, X. et al. A census of the lung: CellCards from LungMAP. *Dev. Cell* **57**, 112–145 e2 (2022).
- Zepp, J. A. & Morrissey, E. E. Cellular crosstalk in the development and regeneration of the respiratory system. *Nat. Rev. Mol. Cell Biol.* **20**, 551–566 (2019).
- Swarr, D. T. & Morrissey, E. E. Lung endoderm morphogenesis: gasping for form and function. *Annu Rev. Cell Dev. Biol.* **31**, 553–573 (2015).
- Minoo, P., Su, G., Drum, H., Bringas, P. & Kimura, S. Defects in tracheoesophageal and lung morphogenesis in *Nkx2.1(-/-)* mouse embryos. *Dev. Biol.* **209**, 60–71 (1999).
- Liu, C., Glasser, S. W., Wan, H. & Whitsett, J. A. GATA-6 and thyroid transcription factor-1 directly interact and regulate surfactant protein-C gene expression. *J. Biol. Chem.* **277**, 4519–4525 (2002).
- DeFelice, M. et al. TTF-1 phosphorylation is required for peripheral lung morphogenesis, perinatal survival, and tissue-specific gene expression. *J. Biol. Chem.* **278**, 35574–35583 (2003).
- Park, K. S. et al. TAZ interacts with TTF-1 and regulates expression of surfactant protein-C. *J. Biol. Chem.* **279**, 17384–17390 (2004).
- Gokey, J. J. et al. YAP regulates alveolar epithelial cell differentiation and AGER via NFIB/KLF5/NKX2-1. *iScience* **24**, 102967 (2021).
- Little, D. R. et al. Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proc. Natl Acad. Sci. USA* **116**, 20545–20555 (2019).
- Little, D. R. et al. Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves opposing murine alveolar cell fates in vivo. *Nat. Commun.* **12**, 2509 (2021).
- 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).
- LaCanna, R. et al. Yap/Taz regulate alveolar regeneration and resolution of lung inflammation. *J. Clin. Invest* **129**, 2107–2122 (2019).
- Nantie, L. B. et al. Lats1/2 inactivation reveals Hippo function in alveolar type I cell differentiation during lung transition to air breathing. *Development* **145**, dev163105 (2018).

15. Kolla, V. et al. Thyroid transcription factor in differentiating type II cells: regulation, isoforms, and target genes. *Am. J. Respir. Cell Mol. Biol.* **36**, 213–225 (2007).
16. Thebaud, B. et al. Bronchopulmonary dysplasia. *Nat. Rev. Dis. Prim.* **5**, 78 (2019).
17. Besnard, V., Xu, Y. & Whitsett, J. A. Sterol response element binding protein and thyroid transcription factor-1 (Nkx2.1) regulate *Abca3* gene expression. *Am. J. Physiol. Lung Cell Mol. Physiol.* **293**, L1395–L1405 (2007).
18. Ikeda, K., Shaw-White, J. R., Wert, S. E. & Whitsett, J. A. Hepatocyte nuclear factor 3 activates transcription of thyroid transcription factor 1 in respiratory epithelial cells. *Mol. Cell Biol.* **16**, 3626–3636 (1996).
19. Shaw-White, J. R., Bruno, M. D. & Whitsett, J. A. GATA-6 activates transcription of thyroid transcription factor-1. *J. Biol. Chem.* **274**, 2658–2664 (1999).
20. Bachurski, C. J., Yang, G. H., Currier, T. A., Gronostajski, R. M. & Hong, D. Nuclear factor I/thyroid transcription factor 1 interactions modulate surfactant protein C transcription. *Mol. Cell Biol.* **23**, 9014–9024 (2003).
21. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011).
22. Xu, Y. et al. Transcriptional programs controlling perinatal lung maturation. *PLoS One* **7**, e37046 (2012).
23. Hohenauer, T. & Moore, A. W. The *Prdm* family: expanding roles in stem cells and development. *Development* **139**, 2267–2282 (2012).
24. Shull, L. C. et al. PRDM paralogs antagonistically balance Wnt/ β -catenin activity during craniofacial chondrocyte differentiation. *Development* **149**, dev200082 (2022).
25. Shull, L. C. et al. The conserved and divergent roles of *Prdm3* and *Prdm16* in zebrafish and mouse craniofacial development. *Dev. Biol.* **461**, 132–144 (2020).
26. McGlynn, K. A. et al. *Prdm3* and *Prdm16* cooperatively maintain hematopoiesis and clonogenic potential. *Exp. Hematol.* **85**, 20–32 e3 (2020).
27. Baizabal, J. M. et al. The epigenetic state of PRDM16-regulated enhancers in radial glia controls cortical neuron position. *Neuron* **99**, 239–241 (2018).
28. Bard-Chapeau, E. A. et al. Mice carrying a hypomorphic *Evi1* allele are embryonic viable but exhibit severe congenital heart defects. *PLoS One* **9**, e89397 (2014).
29. Harms, M. J. et al. *Prdm16* is required for the maintenance of brown adipocyte identity and function in adult mice. *Cell Metab.* **19**, 593–604 (2014).
30. Leszczynski, P., et al. Emerging roles of PRDM factors in stem cells and neuronal system: cofactor dependent regulation of PRDM3/16 and FOG1/2 (Novel PRDM Factors). *Cells* **9**, 2603 (2020).
31. Stine, R. R. et al. PRDM16 Maintains Homeostasis of the Intestinal Epithelium by Controlling Region-Specific Metabolism. *Cell Stem Cell* **25**, 830–845 e8 (2019).
32. Shimada, I. S., Acar, M., Burgess, R. J., Zhao, Z. & Morrison, S. J. *Prdm16* is required for the maintenance of neural stem cells in the postnatal forebrain and their differentiation into ependymal cells. *Genes Dev.* **31**, 1134–1146 (2017).
33. Wu, T. et al. PRDM16 is a compact myocardium-enriched transcription factor required to maintain compact myocardial cardiomyocyte identity in left ventricle. *Circulation* **145**, 586–602 (2022).
34. Horn, K. H., Warner, D. R., Pisano, M. & Greene, R. M. PRDM16 expression in the developing mouse embryo. *Acta Histochem* **113**, 150–155 (2011).
35. Beauchemin, K. J. et al. Temporal dynamics of the developing lung transcriptome in three common inbred strains of laboratory mice reveals multiple stages of postnatal alveolar development. *PeerJ* **4**, e2318 (2016).
36. Goyama, S. et al. *Evi-1* is a critical regulator for hematopoietic stem cells and transformed leukemic cells. *Cell Stem Cell* **3**, 207–220 (2008).
37. Cohen, P. et al. Ablation of PRDM16 and beige adipose causes metabolic dysfunction and a subcutaneous to visceral fat switch. *Cell* **156**, 304–316 (2014).
38. Frank, D. B. et al. Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proc. Natl Acad. Sci. USA* **116**, 4362–4371 (2019).
39. Zepp, J. A. et al. Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* **371**, eabc3172 (2021).
40. Guo, M. et al. Guided construction of single cell reference for human and mouse lung. *Nat. Commun.* **14**, 4566 (2023).
41. Chang, D. R. et al. Lung epithelial branching program antagonizes alveolar differentiation. *Proc. Natl Acad. Sci. USA* **110**, 18042–18051 (2013).
42. Brownfield, D. G. et al. Alveolar cell fate selection and lifelong maintenance of AT2 cells by FGF signaling. *Nat. Commun.* **13**, 7137 (2022).
43. Jones, M. R. et al. FGFR2b signalling restricts lineage-flexible alveolar progenitors during mouse lung development and converges in mature alveolar type 2 cells. *Cell Mol. Life Sci.* **79**, 609 (2022).
44. Martis, P. C. et al. C/EBP α is required for lung maturation at birth. *Development* **133**, 1155–1164 (2006).
45. Herriges, J. C. et al. FGF-regulated ETV transcription factors control FGF-SHH feedback loop in lung branching. *Dev. Cell* **35**, 322–332 (2015).
46. Li, C. et al. *Wnt5a* regulates *Shh* and *Fgf10* signaling during lung development. *Dev. Biol.* **287**, 86–97 (2005).
47. Ghandi, M. et al. gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* **32**, 2205–2207 (2016).
48. Sun, X., Zhang, J. & Cao, C. CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes (Basel)* **13**, 1383 (2022).
49. Nishikata, I. et al. A novel *EVI1* gene family, *MEL1*, lacking a PR domain (*MEL1S*) is expressed mainly in t(1;3)(p36;q21)-positive AML and blocks G-CSF-induced myeloid differentiation. *Blood* **102**, 3323–3332 (2003).
50. Harms, M. J. et al. PRDM16 binds MED1 and controls chromatin architecture to determine a brown fat transcriptional program. *Genes Dev.* **29**, 298–307 (2015).
51. Bohinski, R. J., Di Lauro, R. & Whitsett, J. A. The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. *Mol. Cell Biol.* **14**, 5671–5681 (1994).
52. Wan, H. et al. *Foxa2* is required for transition to air breathing at birth. *Proc. Natl Acad. Sci. USA* **101**, 14449–14454 (2004).
53. Zhang, Z. et al. Transcription factor *Etv5* is essential for the maintenance of alveolar type II cells. *Proc. Natl Acad. Sci. USA* **114**, 3903–3908 (2017).
54. Raulusevičute, I. et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174–D182 (2024).
55. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
56. Wang, S. & Hankinson, O. Functional involvement of the Brahma/SWI2-related gene 1 protein in cytochrome P4501A1 transcription mediated by the aryl hydrocarbon receptor complex. *J. Biol. Chem.* **277**, 11821–11827 (2002).
57. Guan, B., Wang, T. L. & Shih, M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor

- suppressor in gynecologic cancers. *Cancer Res* **71**, 6718–6727 (2011).
58. Medjkane, S., Novikov, E., Versteeg, I. & Delattre, O. The tumor suppressor hSNF5/INI1 modulates cell growth and actin cytoskeleton organization. *Cancer Res* **64**, 3406–3413 (2004).
 59. Wang, W. et al. Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J.* **15**, 5370–5382 (1996).
 60. Guo, B. et al. Arid1a mutation suppresses TGF- β signaling and induces cholangiocarcinoma. *Cell Rep.* **40**, 111253 (2022).
 61. Cho, N. H. et al. OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).
 62. Guo, M. et al. Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Commun.* **10**, 37 (2019).
 63. Yang, H., Lu, M. M., Zhang, L., Whitsett, J. A. & Morrissey, E. E. GATA6 regulates differentiation of distal lung epithelium. *Development* **129**, 2233–2246 (2002).
 64. Jones, M. R. et al. Evidence for overlapping and distinct biological activities and transcriptional targets triggered by fibroblast growth factor receptor 2b signaling between mid- and early pseudoglandular stages of mouse lung development. *Cells* **9**, 1274 (2020).
 65. Xu, B. et al. Cited2 is required for fetal lung maturation. *Dev. Biol.* **317**, 95–105 (2008).
 66. Negretti, N. M. et al. A single-cell atlas of mouse lung development. *Development* **148**, dev199512 (2021).
 67. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
 68. Harfe, B. D. et al. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
 69. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
 70. Chapman, H. A. Toward lung regeneration. *N. Engl. J. Med* **364**, 1867–1868 (2011).
 71. Bridges, J. P. et al. LPCAT1 regulates surfactant phospholipid synthesis and is required for transitioning to air breathing in mice. *J. Clin. Invest* **120**, 1736–1748 (2010).
 72. Sitaraman, S. et al. Proteasome dysfunction in alveolar type 2 epithelial cells is associated with acute respiratory distress syndrome. *Sci. Rep.* **9**, 12509 (2019).
 73. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21 29 1–21 29 9 (2015).
 74. Wert, S. E., Dey, C. R., Blair, P. A., Kimura, S. & Whitsett, J. A. Increased expression of thyroid transcription factor-1 (TTF-1) in respiratory epithelial cells inhibits alveolarization and causes pulmonary inflammation. *Dev. Biol.* **242**, 75–87 (2002).
 75. Lun, A. T. L. et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).
 76. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, <https://doi.org/10.1101/303727> (2020).
 77. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
 78. Traag, V. A. & Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
 79. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).

Acknowledgements

The authors thank members of the CCHMC Single Cell Genomics Facility (RRID:SCR_022653), the CCHMC Bio-imaging and Analysis Facility (RRID:SCR_022628), and the Genomics Sequencing Facility (RRID:SCR_022630) for their advice and support for sample generation and analysis. We also thank Dr. Patrick Seale for generously providing the Rabbit anti-PRDM16 antibody. This research was supported by grants from the National Key R&D Program of China (2022YFA0806200), NSFC (82200098), and the Fundamental Research Funds for the Central Universities (SCU2022D006) to Hua He and NHLBI R01HL164414 to Jeffrey Whitsett and William Zacharias.

Author contributions

H.H. and S.M.B. designed and conducted experiments and prepared the manuscript. A.K.D. and J.S. performed ATAC, Bulk RNA, and CUT & RUN bioinformatic analyses and figure preparation, D.T.S. performed initial CUT & RUN bioinformatic analyses, C.L.N. performed EM analysis, AS immunoprecipitation studies, M.G. and S.Z. performed single cell analysis and figure preparation, Y.X. created the GRN, W.J.Z. evaluated lineage tracing data and prepared the manuscript. J.A.W. designed experiments and prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52154-3>.

Correspondence and requests for materials should be addressed to Hua He or Jeffrey A. Whitsett.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024