nature research

Corresponding author(s): Aviv Regev and Jayaraj Rajagopal

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Initial processing of single-cell RNA-sequencing data was performed using the commercial CellRanger pipelines (10X Genomics, versions 1.0.1 and 1.3.1, described in the Methods section of the paper). Subsequent analysis was performed using the open-source R programming language. FACS Diva software was used for cell sorting on BD FACSAria II equipment. Isc readings were recorded using Acquire & Analyze software.
Data analysis	ImageJ2 was used to analyze immunofluorescent images. All statistical calculations and computational analyses in the paper were performed using the open-source R programming language. De-multiplexing was performed using Illumina Bcl2fastq (v2.17.1.14). Read alignment to transcriptome for SMART-Seq2 data was performed using Bowtie (v2.1.0) and RSEM (v1.2.3), and batch correction was performed using ComBat from the R package 'sva' (v3.26.0). FACS analysis was performed using FlowJo (v10.1). To establish ciliary beat frequency (CBF) custom code in Matlab (Mathworks, Natick, MA) was used to quantify Fourier analysis of the reflectance of beating cilia.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

All RNA-sequencing data for this study is deposited in GEO (accession GSE103354) and can be browsed using the Broad Institute 'Single Cell Portal' (password-protected until publication) at https://portals.broadinstitute.org/single_cell/study/trachea-epithelium. This is described in the manuscript on page 62.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	No sample size calculations were performed for mouse or ferret experiments. Experiments which involved the analysis of many individual cells from each biological replicate (Pulse-Seq and immunostains) we used 3 mice per time point. For analyses in which a single data point was collected per biological sample (qRT-PCR, Ussing chamber), 3-6 mice or ferrets were used per time point, with the limitation of low frequency genotypes. The points on each graph represent independent biological replicates throughout the manuscript.
Data exclusions	In sequencing experiments, cells with fewer than 1,000 genes (3' droplet) or 2,000 genes (full-length) detected were excluded as low-quality cells. In the single case of Foxi1 KO (Fig. 5e), two heterozygous samples were identified as outliers and removed using a standard implementation of DBscan clustering using the full dataset of all genes assayed using qRT-PCR.
Replication	Key findings were reproduced throughout the manuscript. For example, Pulse-Seq recapitulated the unique expression profiles of all cell types and subtypes described in the initial droplet experiment. Key functional studies were also independently replicated. This includes the proximal-distal mucous metaplasia study, which was verified in a second independent cohort, and the equivalence current readings on the Foxi1-KO samples were replicated using the same stem cell-derived samples with a more extensive panel of ion channel inhibitors, confirming our initial finding.
Randomization	Littermate mice were assigned into groups on the basis of genotype.
Blinding	Blinding was used for data analysis throughout the manuscript - including the genotype of mouse samples for expression studies, electrophysiology studies, and characterization of physiologic parameters at the epithelial surface (pH, ASL, mucus, CBF, viscosity).

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	rabbit anti-Atp6v0d2 (1/300; pa5-44359, Thermo), goat anti-CC10 (aka Scgb1a1, 1:500; SC-9772, Santa Cruz), anti-mouse CD45-PE (1/500; #12-0451-83, eBioscience), hamster anti-CD81(1/500; MA1-70091, Thermo), rabbit anti-CFTR (1:100; ACL-006, Alomone), mouse anti-Chromogranin A (1/500; sc-393941, Santa Cruz), rat anti-Cochlin (1/500; MABF267, Millipore), anti-mouse

EpCAM-PECy7 (1/500; 324221, Biolegend), goat anti-FLAP (aka Alox5ap, 1:500; NB300-891, Novus), goat anti-Foxi1 (1:250; ab20454, Abcam), chicken anti-GFP (1:500; GFP-1020, Aves Labs), rabbit anti-Gnat3 (1/300; sc-395, Santa Cruz), rabbit anti-Gng13 (1:500; ab126562, Abcam), rabbit anti-Krt13 (1/500; ab92551, Abcam), goat anti-Krt13 (1/500; ab79279, Abcam), goat anti-Lipf (1:100; MBS421137, mybiosource.com), mouse anti-Muc5ac (1/500; ma1-38223, Thermo), mouse anti-Muc5ac (1/500; ma1-38223, Thermo), mouse anti-p63 (1:250; gtx102425, GeneTex), rabbit anti-Tff2 (1/500; 13681-1-AP, ProteinTech), rabbit anti-Trpm5 (1:500; ACC-045, Alamone), mouse anti-tubulin, acetylated (1:100; T6793, Sigma). All secondary antibodies were Alexa Fluor conjugates (488, 594 and 647) and used at 1:500 dilution (Life Technologies): dk anti-chicken 488 A-11039, dk anti-goat 488 A-11055, dk anti-mouse 488 A-21202, dk anti-rabbit 488 A-21206, dk anti-rat 488 A-21208, dk anti-goat 594 A-11058, dk anti-mouse 594 R37115, dk anti-rabbit 594 R37119, dk anti-hamster 647 A-21451, dk anti-goat 647 A-21447, dk anti-mouse 647 A-31571, dk anti-rabbit 647 A-31573.

Validation

The specificity of each primary antibody was validated by staining directly against species-matched IgG controls.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

HEK 293T cells were used for lentiviral production. Primary mouse and ferret cells were used for all other cell based experiments. Cells were dissociated from mouse or ferret airways and subsequently assayed in differentiation experiments.

Authentication

No authentication was performed on HEK293T cells used to make lentivirus.

Mycoplasma contamination

No mycoplasma testing was performed on HEK293T cells used to make lentivirus. No mycoplasma testing was performed on primary mouse or ferret cells.

Commonly misidentified lines
(See ICLAC register)

HEK293T cells were used to make lentivirus.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

The MGH Subcommittee on Research Animal Care approved animal protocols in accordance with NIH guidelines. Krt5-creER and Scgb1a1-creER mice were described previously. Foxi1-eGFP mice were purchased from GENSAT. C57BL/6J mice (stock no. 000664), LSL-mT/mG mice (mouse stock no. 007676), and LSL-tdTomato (stock no. 007914), Ascl3-EGFP-Cre mice (stock no. 021794), and Foxi1-KO mice (stock no. 024173) were purchased from the Jackson Laboratory. Male C57BL/6 mice were used for the full length and initial 3' scRNA-seq experiments. Both male and female mice were used for lineage tracing, expression analysis on enriched populations, knockout expression and phenotyping studies, and 'Pulse-Seq' experiments. We used three mice for each lineage time point. All mice were between 6-8 weeks old when used for experiments,. In the case of lineage tracing experiments, lineage labeling was initiated at 6-8 weeks of age and mice were collected at specified subsequent time points (30 days) and (60 days).

Wild animals

This study did not involve wild animals.

Field-collected samples

This study did not involve animals collected from the field.

Flow Cytometry

Plots

Confirm that:

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

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

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

Methodology

Sample preparation

Primary mouse cells were isolated as detailed in the methods, "Cell dissociation and FACS".

Instrument

BD FACSAria II was used for cell sorting.

Software

FACS Diva Version 6.1.3 software was used for cell sorting and FlowJo10.1 for data analysis.

Cell population abundance

In the included representative FACS data (Supplemental Figure X), we gated for size (Scatter, 22% of all events), Forward scatter

Cell population abundance

singlets (80.5% of Scatter), Side scatter singlets (83.6% of Forward scatter singlets), Live cells by the exclusion of 7AAD (86.6% of Side Scatter Singlets), EpCAM+ (89.4% of Live cells), membrane-bound tdTomato+ (Tom+, 49.2% of EpCAM+), and membrane-bound eGFP (GFP+, 50.7% of EpCAM+). Subsequent scRNA-seq analysis filtered 0.798% of these cells as contaminating immune or mesenchymal cells, denoting the high purity of epithelial cells (99.202%) isolated by our EpCAM+ sorting protocol.

Gating strategy

The initial gating stategy used SSC-A vs FSC-A for discarding debris. Singlets were gated for by FSC-W vs FSC-A and SSC-W vs SSC-A. Live/dead cells were gated using PerCP-A vs FSC-A. EpCAM cells were gated from SSC-A vs Pe-Cy7 Green-A. GFP+ and tdTomato+ cells were selected on the basis of PE-A vs FITC-A. Fluorescence minus one controls were used to determine positive and negative boundaries.

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