nature research

Corresponding author(s): Muus, Luecken, Regev, HCA Lung Biological Network

Last updated by author(s): Dec 3, 2020

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	R packages: lme4 (1.1.21), Signac (0.1.6), JASPAR (2020), EnrichR (1.0), Harmony (0.1.1), Seurat (3.1.1), chromVAR (1.6.0), TFBSTools (1.25.1), MAST (1.8.2), gprofiler (1.0.0), GenomeInfoDb (1.22.0), python packages: harmony-pytorch (version 0.1.1), scanpy (versions 1.4.5, 1.4.5.1, 1.4.6), statsmodels (version 0.11.1), sklearn (version 0.21.3), ForceAtlas2 (version 0.3.5), louvain (version 0.6.1), networkx (version 2.2), diffxpy (version 0.7.3), batchglm (version 0.7.4) 10X Genomics CellRanger ATAC mkfastq (version 1.1.0), CellphoneDB (version 2.0.0), Star (Version 20201), Cufflinks (Version 2.2.1.3), Cuffdiff (Version 2.2.1.6)

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

Data

Policy information about availability of data

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

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

Data and an interactive analysis examining the co-expression of genes across datasets can be accessed via the open-source data platform, Terra at https://app.terra.bio/#workspaces/kco-incubator/COVID-19_cross_tissue_analysis.

Interactive visualization and download of gene expression data can be accessed on the Single Cell Portal at http://broad.io/hcacovid19

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Meta-analysis of single-cell atlases, see main text and supplementary tables for included studies.
Data exclusions	Meta-analysis of single-cell atlases, see main text and supplementary tables for included studies.
Replication	N/A (the data was not specifically collected for this meta-analysis)
Randomization	N/A (the data was not specifically collected for this meta-analysis)
Blinding	N/A (the data was not specifically collected for this meta-analysis)

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

Primary antibodies
ACTA2 Mouse IgG2a, FITC-conjugated (Sigma, F3777, Clone: 1A4, lot: 038M4865V, 1:500)
AGER Goat IgG (R&D systems, AF1145, lot: HCE0719011, 1:200)
HTII-280 Mouse IgM (Terrace Biotech, TB-27AHT2-280, lot: not available, 1:100)
Pro-SFTPC Rabbit IgG (Sigma, ab3786, lot: 3267937, 1:500)
TMPRSS2 Rabbit IgG (Abcam, ab109131, Clone: EPR3862, lot: GR3248440-1, 1:200)

Secondary antibodies
Alexa Fluor 488 Donkey anti-Mouse IgM (Thermo fisher scientific, A21042, lot: 2160416, 1:400)
Alexa Fluor 488 Donkey anti-Rabbit IgG (Thermo fisher scientific, A32795, lot: 1981155, 1:400)
Alexa Fluor 594 Donkey anti-Rabbit IgG (Thermo fisher scientific, A21207, lot: 1987293, 1:400)
Alexa Fluor 594 Goat anti-Mouse IgM (Thermo fisher scientific, A21044, lot: 1806144, 1:400)
Alexa Fluor 647 Donkey anti-Goat IgG (Thermo fisher scientific, A21447, lot:2175459, 1:400)

Validation

Validation statements and relevant citations of the listed antibodies are available in the manufacturer's websites:

Primary antibodies:
ACTA2 Mouse IgG2a, FITC-conjugated(https://www.sigmaaldrich.com/catalog/product/sigma/f3777?lang=en&region=US)
AGER Goat IgG (https://www.rndsystems.com/products/human-mouse-rat-rage-ager-antibody_af1145)
HTII-280 Mouse IgM (https://www.terracebiotech.com/product-page/anti-ht2-280-1ml)
Pro-SFTPC Rabbit IgG (http://www.emdmillipore.com/US/en/product/Anti-Prosurfactant-Protein-C-proSP-C-Antibody,MM_NF-AB3786?ReferrerURL=https%3A%2F%2Fwww.google.com%2F&bd=1)
TMPRSS2 Rabbit IgG (https://www.abcam.com/tmprss2-antibody-epr3862-ab109131.html)

Secondary antibodies:
Alexa Fluor 488 Donkey anti-Mouse IgM (https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgM-Heavy-chain-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21042)
Alexa Fluor 488 Donkey anti-Rabbit IgG (https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32795)
Alexa Fluor 594 Donkey anti-Rabbit IgG (https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21207)
Alexa Fluor 594 Goat anti-Mouse IgM (https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgM-Heavy-chain-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21044)
Alexa Fluor 647 Donkey anti-Goat IgG (https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21447)

Animals and other organisms

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

Laboratory animals	Wild-type Mus musculus, C57BL/6J and Cast/EiJ, both Females and Males were ordered from the Jackson Lab. Mice were housed under standard barrier conditions at the Whitehead Institute for Biomedical Research. Experimental mice were obtained by crossing Cast/EiJ mice with C57BL/6J mice. For the smoke exposure experiments, 8 to 10 week old pathogen-free female wild-type C57BL/6 mice were obtained from Charles River (Sulzfeld, Germany).
Wild animals	none
Field-collected samples	none
Ethics oversight	Mice were housed under standard barrier conditions at the Whitehead Institute for Biomedical Research. All experiments performed in this study were in accordance with the relevant animal husbandry standards of the Committee on Animal Care. All mouse exposure experiments were approved by the ethics committee for animal welfare of the local government for the administrative region of Upper Bavaria (Regierungspräsidium Oberbayern) and were conducted under strict governmental and international guidelines in accordance with EU Directive 2010/63/EU.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Provided in supplementary table 2 for lung single-cell datasets, N/A or not specifically collected for the remaining studies.
Recruitment	N/A (the data was not specifically collected for this meta-analysis)
Ethics oversight	Sample collection underwent IRB review and approval at the institutions where the samples were originally collected. "Adipose_Healthy_Manton_unpublished" was collected under IRB 2007P002165/1(ORSP-3877). Tissue samples from breast, esophagus muscularis, esophagus mucosa, heart, lung, prostate, skeletal muscle and skin referred to as "Tissue_Healthy_Regev_snRNA-seq_unpublished" were collected under ORSP-3635. Samples referred to as "Eye_Sanes_unpublished" were collected under Dana Farber / Harvard Cancer Center Protocol Number 13-416 and Massachusetts Eye and Ear Protocol Number 18-034H. Samples referred to as "Kidney_Healthy_Greka_unpublished" were collected under Massachusetts General Hospital IRB number 2011P002692. Samples referred to as "Liver_Healthy_Manton_unpublished" were collected under IRB 02-240; ORSP 1702 as well as and ORSP-2630 under ORSP-2169. Lung samples from smokers and non-smokers (41 samples, 10 patients, 2-6 locations each) with suffix "Regev/ Rajagopal_unpublished" were collected under Massachusetts General Hospital IRB 2012P001079 / (ORSP-3900) under ORSP-3490. Healthy and fibrotic lung samples with suffix "Xavier_snRNA-seq_unpublished" were collected under Massachusetts General Hospital IRB number 2003P000555 (CG-5242) under ORSP-3490, Medoff, 2015P000319 (CG-5145) under ORSP-3490. Pancreas PDAC samples were collected under Fernandez-del Castillo, 2003P001289 (CG-4692) under ORSP-3490 Massachusetts General Hospital IRB number Fernandez-del Castillo, 2003P001289 (CG-4692) under ORSP-3490. Samples in the dataset "Barbry" were derived from a study that was approved by the Comité de Protection des Personnes Sud Est IV (approval number: 17/081) and informed written consent was obtained from all participants involved. All experiments were performed during 8 months, in accordance with relevant guidelines and French and European regulations. No deviations were made from our approved protocol named 3Asc (An Atlas of Airways at a single cell level - ClinicalTrials.gov identifier: NCT03437122). IPF and COPD lungs in the "Kaminski" dataset were obtained from patients undergoing transplant while healthy lungs were from rejected donor lung organs that underwent lung transplantation at the Brigham and Women's Hospital or donor organs provided by the National Disease Research Interchange (NDRI). Patient tissues relating to the dataset "Krasnow" were obtained under a protocol

approved by Stanford University's Human Subjects Research Compliance Office (IRB 15166) and informed consent was obtained from each patient prior to surgery. The study protocol was approved by the Partners Healthcare Institutional Board Review (IRB Protocol # 2011P002419). Samples in the dataset "Kropski_Banovich" were collected under Vanderbilt IRB # 060165, 171657, and Western IRB#20181836. Ethics approval number 2018/769-31. "Meyer_b" were collected under CBTM (Cambridge Biorepository for Translational Medicine), research ethics approval number: UK NHS REC approval reference number 15/EE/0152. Samples in the dataset "Linnarsson" are covered by (2018/769-31) approved by the Swedish Ethical Review Authority. Samples in the "Misharin" dataset were collected under (STU00056197, STU00201137, and STU00202458) approved by the Northwestern University Institutional Review Board. Samples in the "Rawlins" dataset were obtained from terminations of pregnancy from Cambridge University Hospitals NHS Foundation Trust under permission from NHS Research Ethical Committee (96/085) and the Joint MRC/Wellcome Trust Human Developmental Biology Resource (grant R/R006237/1, www.hdbr.org, HDBR London: REC approval 18/LO/0822; HDBR Newcastle: REC approval 18/NE/0290). The studies relating to datasets "Schultze" and "Schultze_Falk" were approved by the ethics committees of the University of Bonn and University hospital Bonn (local ethics vote 076/16) and the Medizinische Hochschule Hannover (local ethics vote 7414/2017). Fifteen human tracheal airway epithelia in the "Schultze" dataset were isolated from de-identified donors whose lungs were not suitable for transplantation. Lung specimens were obtained from the International Institute for the Advancement of Medicine (Edison, NJ) and the Donor Alliance of Colorado. The National Jewish Health Institutional Review Board (IRB) approved the research under IRB protocols HS-3209 and HS-2240. Samples in the "Xu/Whitsett" dataset were provided through the federal United Network of Organ Sharing via the National Disease Research Interchange (NDRI) and International Institute for Advancement of Medicine (IIAM) and entered into the NHLBI LungMAP Biorepository for Investigations of Diseases of the Lung (BRINDL) at the University of Rochester Medical Center, overseen by the IRB as RSRB00047606. (Supplementary Table 1, 2)

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