Published in final edited form as:

Nat Med. 2021 March ; 27(3): 546–559. doi:10.1038/s41591-020-01227-z.

Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics

A full list of authors and affiliations appears at the end of the article.

Abstract

ACE2 and accessory proteases (TMPRSS2, CTSL) are needed for SARS-CoV-2 cellular entry, and their expression may shed light on viral tropism and impact across the body. We assess the cell type-specific expression of *ACE2*, *TMPRSS2*, and *CTSL* across 107 single-cell RNA-Seq studies from different tissues. *ACE2, TMPRSS2*, and *CTSL* are co-expressed in specific subsets of respiratory epithelial cells in the nasal passages, airways, and alveoli, and in cells from other organs associated with COVID-19 transmission or pathology. We performed a meta-analysis of 31 lung scRNA-seq studies with 1,320,896 cells from 377 nasal, airway, and lung parenchyma samples from 228 individuals. This revealed cell type specific associations of age, sex, and smoking with expression levels of *ACE2*, *TMPRSS2*, and *CTSL.* Expression of entry factors increased with age and in males, including in airway secretory cells and alveolar AT2 cells. Expression programs shared by *ACE2*[+]*TMPRSS2*[+] cells in nasal, lung and gut tissues included

^Corresponding author: lung@humancellatlas.org, hca@humancellatlas.org.

*These authors contributed equally

Author Contributions

principal investigators listed in The Human Cell Atlas Lung Biological Network authors provided resources.
sample collection was performed by A.W., B.L., D.T.M., S.L., I.R., J.C.S., M.S., K.A.V. and HCA Lung Biological Network
C.M., A.W., Y.K., E.T.D., B.L., D.T.M., T.S.A., S.J.A., Z.A., I.A., K.B., J.B., Lu.B., Li.B., E.B., L.T.B., S.C., E.C., T.M.C., M.S.C., S.G., A.G., X.H., L.H., Y.H., T.S.K., M.L., L.S.L., W.L., H.M., E.M., L.M., K.M., I.M.M., A.M.M., A.N.N., S.K.N., L.P., C.Q., D.R., R.V.S., M.S., N.P.S., M.S., D.S., J.T., K.J.T., M.H.W., J.W., W.Y., C.G.K.Z. performed experiments
C.M., M.D.L., G.E., L.S., A.W., Y.K., G.H., E.D.V., A.S., C.S.S., K.A.J., E.T.D., E.F., E.T.T., M.A., P.C., B.L., J.B., S.C., J.S., A.L.H., H.C., D.T.M. analyzed and interpreted data with input from T.S.A., H.A., S.J.A., O.A., C.B., I.B., M.C., J.C., T.M.C., A.S.C., M.D., G.D., D.S.F., S.G., A.G., B.G., J.G., M.G., A.C.H., T.H., P.H., A.J., L.J., P.J., T.S.K., L.L., M.A.L., S.L., H.M., E.M., C.H.M., I.M.M., A.M.M., L.P., A.N.N., O.B.P., S.P., S.K.N., R.Q., D.R., C.V.S., X.S., R.S., K.S., T.B.S., C.T., P.T., K.J.T., N.R.T., M.H.W., X.W., W.Y., W.Z., C.G.Z.
A.R., C.M., M.D.L., G.E., L.S., A.W., G.H., E.D.V., A.S., C.S.S., K.A.J., E.F., E.T.T., B.L., S.C., J.S., A.L.H., D.T.M. wrote the manuscript with input from all authors and guidance from A.R., J.R., F.J.T., M.C.N.

Conflict of interest statement

N.K. was a consultant to Biogen Idec, Boehringer Ingelheim, Third Rock, Pliant, Samumed, NuMedii, Indaloo, Theravance, LifeMax, Three Lake Partners, Optikira and received non-financial support from MiRagen. All of these outside the work reported. J.L. is a scientific consultant for 10X Genomics Inc. A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Asimov, and Neogene Therapeutics. O.R.R. and A.R. are co-inventors on patent applications filed by the Broad Institute to inventions relating to single cell genomics applications, such as in PCT/US2018/060860 and US Provisional Application No. 62/745,259. A.K.S. compensation for consulting and SAB membership from Honeycomb Biotechnologies, Cellarity, Cogen Therapeutics, Orche Bio, and Dahlia Biosciences. S.A.T. was a consultant at Genentech, Biogen and Roche in the last three years. F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH, and ownership interest in Cellarity Inc. L.V. is founder of Definigen and Bilitech two biotech companies using hPSCs and organoid for disease modelling and cell based therapy. J.A.K. has received advisory board fees from Boehringer Ingelheim, Inc, and has research contracts with Genentech. Eric S. Lander serves on the Board of Directors for Codiak BioSciences and serves on the Scientific Advisory Board of F-Prime Capital Partners and Third Rock Ventures; he is also affiliated with several non-profit organizations including serving on the Board of Directors of the Innocence Project, Count Me In, and Biden Cancer Initiative, and the Board of Trustees for the Parker Institute for Cancer Immunotherapy. He has served and continues to serve on various federal advisory committees. Joakim Lundeberg is a scientific consultant for 10X Genomics Inc. Jennifer Beane, Joshua Campbell, Mary Reid and Sarah Mazzilli are funded in part by a sponsored research agreement from Janssen Pharmaceuticals, Inc. Avrum Spira is an employee of Johnson & Johnson. Ramnik J. Xavier is a co-founder Celsius Therapeutics and Jnana Therapeutics, and a consultant at Novartis. All other authors declare no conflicts of interest.

genes that may mediate viral entry, key immune functions and epithelial-macrophage cross-talk, such as genes involved in the IL6, IL1, TNF and complement pathways. Cell type-specific expression patterns may contribute to COVID-19 pathogenesis , and our work highlights putative molecular pathways for therapeutic intervention.

INTRODUCTION

COVID-19, caused by SARS-CoV-2 infection, can manifest with pathologies in multiple systems, including the lungs and airways, gastrointestinal tract, kidney, liver, and heart, and multiorgan failure[1-3]. SARS-CoV-2 RNA has been found in nasal and throat secretions, saliva and stool specimens[4].

Virion infection of host cells is initiated by the viral spike (S)-protein binding to ACE2. *ACE2* expression has been correlated with increased viral load in human cell lines[5,6] and in mice[7]. Viral infection further requires proteolytic cleavage of the S-protein, and TMPRSS2 or Cathepsin L, encoded by the *CTSL* gene, can provide this role for cellular entry[8].

There is substantial variation in the clinical consequences of infection across individuals, from asymptomatic to death. Disease severity and mortality rise with age[9,10], with a slightly higher incidence and mortality in men[2]. Children are significantly less likely to develop severe acute disease[11]. Smoking may be associated with more severe disease[12]. Finally, adults with pre-existing cardiovascular disease may have higher rates of disease acuity and death[2].

Identifying specific cell types that can be infected by SARS-CoV-2 and relating SARS-CoV-2 entry factors to key co-variates, like age or sex could inform our understanding of COVID-19 tropism and heterogeneity in disease outcomes. The Human Cell Atlas (HCA) community has generated single-cell cell atlases of diverse tissues in healthy individuals, which can now be leveraged to enable such studies. Early analyses of Human Cell Atlas data revealed that some of the cells of the nasal passages, airways, lung parenchyma, and gut express *ACE2* and *TMPRSS2*[13,14], most notably nasal goblet cells and multiciliated cells[13] in the airways and AT2 cells in the distal lung[13,15,16], and identified *ACE2* and *TMPRSS2* expression in colonic enterocytes[13,17].

Here, we chart the cell-type-specific expression patterns of *ACE2* and accessory proteases by integrated analysis of 116 single-cell and single-nucleus RNA-Seq studies, including 31 studies of the lung and airways, and 85 studies of other diverse tissues. With the lung and airway studies, we performed the first single-cell meta-analysis of atlas datasets associating cell type specific changes in expression level with age, sex and smoking status. We identify cross-tissue and tissue-specific gene programs enriched in immune-associated genes in *ACE2*$^{+}$ *TMPRSS2*$^{+}$ cells and highlight other proteases that are significantly co-expressed with *ACE2* and could play a role in infection.

RESULTS

Double positive *ACE2*$^+$*TMPRSS2*$^+$ cells across the lung, airways and other organs associated with COVID-19

We enumerated the proportion of double positive *ACE2*$^+$*TMPRSS2*$^+$ cells and *ACE2*$^+$*CTSL*$^+$ cells across 92 human single-cell or single-nucleus RNA-seq (sc/snRNA-seq) studies (including seven of the lung and airways) (Fig. 1, Methods, Supplementary Table 1 and 2). We surveyed published datasets, assigning cells to five broad categories (Fig. 1a,b, Extended Data Fig. 1, Extended Data Fig. 2, Supplementary Table 1), and analyzed more finely annotated published and unpublished datasets (Methods, Fig. 1c,d, Supplementary Table 1,3).

ACE2$^+$*TMPRSS2*$^+$ epithelial cells were most prevalent (in order) within the ileum, liver, lung, nasal mucosa, bladder, testis, prostate, and kidney (Fig. 1a). Consistent with previous reports[33], double positive *ACE2*$^+$*TMPRSS2*$^+$ cells in the nose and airways were largely secretory goblet and multiciliated cells, and double positive cells in the distal lung were largely AT2 cells (Fig. 1c, Extended Data Fig. 3a). *ACE2* and *TMPRSS2* expression in secretory and AT2 cells is also supported by scATAC-seq from the primary carina and subpleural parenchyma of one adult individual, respectively, as well as secretory and multiciliated cells, and to a lesser extent some basal and tuft cells (Supplementary Fig. 1a-d, n=3 samples per location, n=1 patient, Methods). In a larger aggregation of lung and nasal datasets (Methods), we observed *ACE2*$^+$ *TMPRSS2*$^+$ cells in various lung epithelial cells in pediatric samples (Extended Data Fig. 3b,c), also supported by single-cell chromatin accessibility by transposome hypersensitive sites sequencing (scTHS-Seq)[18] (Extended Data Fig. 4, Methods). Significant double positive *ACE2*$^+$ *TMPRSS2*$^+$ cells in other tissues included enterocytes, pancreatic ductal cells, prostate luminal epithelial cells, brain oligodendrocytes, kidney proximal tubular cells and principal cells of the collecting duct, inhibitory enteric neurons, heart fibroblasts/pericytes, and fibroblasts and pericytes in multiple tissues (Fig. 1a-c). Notably, some of the cell types in which there are double positive cells (including brain oligodendrocytes, multiciliated cells of the upper respiratory tract, and sustentacular cells in olfactory epithelium) are cell types that also express *MYRF* (albeit not always significant triple expressors; Supplementary Fig. 2). MYRF is a transcription factor that induces expression of the myelin proteins MBP (myelin basic protein) and MOG (myelin oligodendrocyte glycoprotein)[19] Autoimmune reactions against these proteins are known to potentially induce neurological symptoms (Discussion).

ACE2$^+$ *CTSL*$^+$ co-expressing cells were enriched among AT1 and AT2 cells, enterocytes, ventricular cardiomyocytes and heart macrophages, as well as fibroblasts and pericytes in multiple tissues, including the placenta, heart, lung, kidney and ENS (Fig. 1d). We did not observe substantial *ACE2* mRNA expression in scRNA-seq profiles in the bone marrow or cord blood (Fig. 1a,b), although there was *ACE2* expression in alveolar and heart macrophages (Extended Data Fig. 5). Notably, in human placenta[20-22], *ACE2* was expressed (1.4%) in maternal decidual/stromal cells, maternal pericytes, and fetal extravillous trophoblasts, cytotrophoblasts, and syncytiotrophoblast in both first-trimester and term placenta (Fig. 1d). While there was little expression of *TMPRSS2* (0.2%), *CTSL* was expressed in most cells (56%), and there were *ACE2*$^{+}$*CTSL*$^{+}$ double positive cells (1.3%).

Cell type specific expression of additional proteases that may be relevant to infection

SARS-CoV-2 infects cells in the absence of *TMPRSS2*[8], so additional proteases likely play roles in proteolytic cleavage of viral proteins for entry and egress. To predict such proteases, we tested the co-expression of *ACE2* with each of 625 annotated human protease genes[23] in a declined donor transplant dataset ("Regev/Rajagopal", Supplementary Table 1). *TMPRSS2* was significantly co-expressed in multiple lung epithelial cell types (Fig. 2a, Supplementary Table 4, 5), as were multiple members of the proprotein convertase subtilisin kexin (*PCSK*) family (Fig. 2a,b), including *FURIN*, *PCSK2*, *PCSK5*, *PCSK6* and *PCSK7* in AT2 cells. Proprotein convertases have known roles in coronavirus S-protein priming. We obtained similar results in an independent dataset from 40 samples (Extended Data Fig. 6a,b, Supplementary Table 1, datasets "Barbry", "Kropski", "Lafyatis/Rojas", "Misharin_new", "Nawijn/Teichmann", "Northwestern_Misharin_ 2018Reyfman", "Sanger_Meyer_2019Madissoon"). As previously reported[24], the SARS-CoV-2 S-protein has a polybasic motif in the S1/S2 region (Extended Data Fig. 6c) that corresponds to cleavage motifs of PCSK family proteases (Extended Data Fig. 6d)[24] and an additional site at the S2' position (Extended Data Fig. 6e)[25].

FURIN, *PCSK5* and *PCSK7* were co-expressed with *ACE2* across multiple lung cell types (Fig. 2c, Extended Data Fig. 6f), *PCSK1* and *PCSK2* were mostly restricted to neuroendocrine cells[26], and *PCSK2* also detected in some AT2 cells (Fig. 2d, Extended Data Fig. 6g). In AT2 cells, proximal multiciliated cells, and basal cells, dual expression of PCSKs with *ACE2* was at fractions comparable to or higher than *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells (Fig. 2e, Extended Data Fig. 6h). Co-expression is significant across other tissues (Extended Data Fig. 6i,j), including liver, ileum, kidney, and nasal airways.

Because different host proteases may contribute to different stages of the viral life cycle[25], we examined the prevalence of *ACE2*$^{+}$*TMPRSS2*$^{+}$*PCSK*$^{+}$ triple-positive cells (TPs) in the lung. *ACE2*$^{+}$*TMPRSS2*$^{+}$*PCSK7*$^{+}$ were the main TPs in multiciliated (0.75%) and secretory (0.72%) cells of proximal airways, and *ACE2*$^{+}$*TMPRSS2*$^{+}$*FURIN*$^{+}$ TPs were the most common within AT2 cells (0.36%) (Extended Data Fig. 6k). Among all known human proteases (Fig. 2f, Supplementary Fig. 3), cathepsins (*CTSB*, *CTSC*, *CTSD*, *CTSL*, *CTSS*), proteasome subunits (*PSMB2, PSMB4, PSMB5*), and complement proteases (*C1R*, *C2*, *CFI*), were the most commonly co-expressed with *ACE2* in lung epithelial cell types.

Orthogonal validation of *ACE2*, *TMPRSS2* and *CTSL* expression in the lungs

As *ACE2* expression is quite low, we next validated some of these patterns by fluorescence *in situ* hybridization and immunofluorescence in tissue sections of airways and alveoli from three healthy donor lungs that were rejected for lung transplantation. *ACE2*, *CTSL* and *TMPRSS2* were co-expressed by fluorescence in situ hybridization in alveolar cells, albeit at low levels (Fig. 1e,f). Co-staining with cell type-specific markers, showed *ACE2* expression and *TMPRSS2* expression in some HTII-280^{+} AT2 cells (Fig. 1g,h); we confirmed the latter by TMPRSS2 protein immunostaining (Extended Data Fig. 7d).

TMPRSS2 protein was expressed at low levels in some AT1 cells (identified by AGER, Extended Data Fig. 7d). Some non-epithelial cells also expressed these three genes. We further validated *ACE2* expression by bulk mRNA-seq of sorted AT2 cells (Extended Data Fig. 7e). Immunohistochemistry with antibodies used previously to block cellular viral entry specifically labeled adult pro-$SFTPC^{+}$ AT2 cells (Extended Data Fig. 7c, Supplementary Table 6, Methods).

Previous studies revealed that ACE2 is highly enriched in nasal and intestinal mucous cells[13,14]. While mucous cells are relatively rare in healthy surface airway epithelium, they are abundant in submucosal glands (SMGs). scRNA-seq of microdissected SMGs of healthy donors showed enrichment of *ACE2*, *TMPRSS2* and *CTSL* in mucous cells (Extended Data Fig. 7f). *In situ* analysis confirmed the presence of *ACE2* transcripts in acinar epithelial cells of the SMGs (Extended Data Fig. 7g), and cells expressing *ACE2* in the large airway epithelium (Extended Data Fig. 7).

Association of *ACE2*, *TMPRSS2*, and *CTSL* expression in lung and airway cells with age, sex and smoking

We next asked how the expression of *ACE2*, *TMPRSS2*, and *CTSL* in specific cell subsets relates to three key covariates associated with disease severity: age (older individuals are more severely affected), sex (males are more severely affected), and smoking (smokers are more severely affected)[27]. As no single dataset to date was sufficiently large, we aggregated samples across 31 sc/snRNA-seq studies (Supplementary Table 2; 14 published[16,28-38]; 17 not yet published[39,40]). This analysis spanned 1,320,896 cells from 228 individuals without known lung disease or from histologically normal-appearing lung adjacent to the site of disease, across 377 nasal, lung, and airway samples from either brushes, scrapings, biopsies, bronchoalveolar lavages, resections, entire lungs that could not be used for transplant or post mortem examinations (Fig. 3a). From unpublished data, we only obtained single-cell expression counts for the three genes (pre-processed by each data generator), total UMI counts per cell, cell identity annotations (which we harmonized to three resolution levels across studies; Fig. 3a,b, Supplementary Table 2, Extended Data Fig. 8, Methods), and age, sex, and smoking status (when ascertained). We modeled the association between the expression counts of each gene and age, sex, and smoking status using a linear model, accounting for technical variation arising from dataset-related factors and covariate interactions (Methods). We fitted this model within each cell type to non-fetal lung data of donors for whom smoking history was known (985,420 cells, 286 samples, 164 donors, 21 datasets), and fitted a model without smoking status covariates to the full non-fetal lung data (1,096,604 cells, 309 samples, 185 donors, 24 datasets).

For simplicity, we treated each cell as an independent observation. This implicitly combines variability in both donors and cells, and, because cells from the same donor are not truly independent observations, can result in inflated p-values, especially when there are few donors for a particular cell type. To address this, account for covariate interactions, and ensure robustness, we: (**1**) used a simple noise model (Poisson) to reduce overfitting of donor variability; (**2**) confirmed that effect directions of significant associations are consistent in a pseudo-bulk analysis (modeling only donor variation;

Methods, Supplementary Data 1-4); (**3**) confirmed summarized age, sex, and smoking associations with a model including interaction terms (Methods, Supplementary Data 1-4); and (**4**) separated significant associations that passed all above confirmations into *robust trends* and *indications* depending on their robustness to holding out individual datasets (Methods, Supplementary Data 1-4). We focused on trends or indications in cell types where *ACE2* and *TMPRSS2* are co-expressed (Fig. 3c): airway epithelial cells (basal, multiciliated, and secretory cells), alveolar AT1 and AT2 cells, and submucosal gland secretory cells.

We find robust trends of *ACE2* expression with age, sex, and smoking status in these cell types (Fig. 3d, Extended Data Fig. 9, Supplementary Fig. 4-6; non-smoking model results in Supplementary Fig. 7-10): *ACE2* expression increases with age in AT2 cells, and is elevated in males in airway secretory cells and alveolar AT1 and AT2 cells. *ACE2* levels are higher in past or current smokers in basal and submucosal secretory cells, and lower in AT2 cells (Fig. 3d). Analysis of bulk RNA-Seq data from bronchial brushings[41] indicated an upregulation of both *ACE2* and *TMPRSS2* in current *vs*. former smokers (Extended Data Fig. 10). Furthermore, we find indications of increased *ACE2* expression with age and in males in multiciliated cells, but those rely on inclusion of the dataset with the most cells and samples ("Regev/Rajagopal"; Extended Data Fig. 9, Methods). All above trends and indications for sex and age were validated in a simplified model without smoking status on the full non-fetal lung dataset (Supplementary Fig. 7, Supplementary Data 5-8, Methods).

Examining joint trends of *ACE2* and the protease genes within the same cell type, we found robust trends of *ACE2* and *TMPRSS2* co-expression increasing with age in AT2 cells, in males in AT1 cells, and an indication of the two genes being elevated in males in multiciliated cells (*ACE2* indication dependent on "Regev/Rajagopal" dataset; Fig. 3d, Extended Data Fig. 9). *ACE2* and *CTSL* show robust trends of joint up-regulation in males in AT2 cells, and in smokers in submucosal secretory cells. Indications of joint up-regulation of these genes were found in males in AT1 cells, and in smokers in basal cells (Fig. 3d, Extended Data Fig. 9, Methods). All joint trends for age and sex covariates were confirmed on the full non-fetal lung data using the simple model without smoking covariates (Supplementary Fig. 7).

An immune gene program in *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in airway, lung and gut

Our previous analyses revealed immune signaling genes that co-vary with *ACE2* and *TMPRSS2* in airway and lung cells[13,14] . To explore these in a broader context, we identified tissue and cell programs related to double positive *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in the nasal epithelium, lung, and gut (Supplementary Tables 7-10). Tissue programs are shared across double positive cells from different cell types in one tissue; cell programs distinguish double positive cells from the rest of the cells of the same type (Methods).

Tissue programs were enriched in pathways related to viral infection and immune response, including phagosome structure, antigen processing and presentation, and apoptosis (Fig. 4a,b, Supplementary Fig. 11a,b for selected genes, Supplementary Tables 7-10). These include *CEACAM5* (lung, nasal, gut programs) and *CEACAM6*[42] (lung), surface attachment factors for coronavirus spike protein; *SLPI* (lung, nasal)[43]; *PIGR* (lung, gut; may promote antibody-dependent enhancement via IgA[44]); and *CXCL17* (lung, nasal)[45]. Tissue programs also had genes associated with cholesterol and lipid metabolic pathways and endocytosis (*DHCR24*, *LCN2*, *FASN*); MHC I and MHC II pathways[46]; preparation against cellular injury (interferons, extracellular RNAse: *PLAC8*, *TXNIP*); complement (*C3*, *C4BPA*); immune modulation (*BTG1*) and tight junctions (*DST*, *CLDN3*, *CLDN4*).

Cell programs (Fig. 4c,d, Supplementary Fig. 12a-c, Supplementary Tables 7-10) were enriched in many of the same genes and pathways (e.g., *CEACAM5*, *CXCL17*, *SLPI*), and further captured unique functions, including TNF signaling in lung secretory cells (*e.g.*, *RIPK3*[47]), lysosomal functions in lung secretory and multiciliated cells[48], the immunoproteasome (AT1 cells, Fig. 4c), cytokines, chemokines and their receptors (nasal goblet cells: *CSF3*, *CXCL1*, *CXCL3*, *IL19*, CCL20; AT1 cells: *IL1R1*), and genes that encode surfactant proteins (AT2 cells, *SFTPA*, *SFTPA2*). Cell programs from multiple tissues (Fig. 4c,d) included genes related to TNF signaling, raising the possibility that anti-TNF therapy may impact the expression of *ACE2* and/or *TMPRSS2*. Some of the genes encode proteins that are targets of known drugs[49] (*e.g.*, in lung secretory cells: *C3*, *HDAC9*, *IL23A*, *PIK3CA*, *RAMP1*, and *SLC7A11*), other gene products have been shown to interact with SARS-CoV-2 proteins[50] (*e.g.*, GDF15[68], a central regulator of inflammation[51]), and yet others may be related to COVID-19 pathological features, including *MUC1*[52] (in tissue and specific cell programs), *IL6ST* (lung tissue and gut enterocyte programs), and *IL6* (AT2 program, Supplementary Fig. 12d). Other cell types, such as heart pericytes, are enriched for cells co-expressing *ACE2* with *IL6R* or *IL6ST* (Supplementary Fig. 13). The immune-like programs of $ACE2^+$ epithelial cells are also reflected in the regulatory features of the *ACE2* locus by scATAC-Seq (Fig. 4f). Cell-cell interaction analysis[53] (Methods) predicted interactions (Supplementary Table 11) between AT2 cells (overall or $ACE2^+TMPRSS2^+$) and myeloid cells through oncostatin, complement, IL1 receptor and CSF signaling.

Conserved expression patterns in mouse models

Preclinical studies of SARS-CoV-2 infection and treatment require model systems that approximate human physiology. Transgenic hACE2 mouse models have been identified as a valuable resource to evaluate diverse therapeutics for COVID-19[54]. We thus asked whether expression patterns of SARS-CoV-2 entry factors were similar in human and mouse model cell types of interest.

$Ace2^+Tmprss2^+$ and $Ace2^+Ctsl^+$ double positive cells were present primarily in club and multiciliated cells in the airway epithelia of healthy mice[55] (Fig. 5a), consistent with human airways (Extended Data Fig. 3a), and increased from 2 to 4 months old (Fig. 5a,b). Moreover, the expression patterns observed in scRNA-seq data of whole lungs from mice exposed daily to cigarette smoke for two months (Fig. 5c-k, Methods) are consistent with our observations in human airway epithelial cells (Fig. 3d, Extended Data Fig. 9a). Upon smoke exposure, there was a significant increase in the number $Ace2^+$ cells and *Ace2* expression in airway secretory cell numbers, but not AT2 cells (Fig. 5f-i). There was also agreement in expression patterns between the human placenta and mouse placenta development (Fig 1c,d, Fig. 5l, Supplementary Fig. 14).

DISCUSSION

To the best of our knowledge, this study represents the first single-cell meta-analysis. Our meta-analysis provided the required power to uncover age, sex and smoking associations at single-cell resolution. The contrasting smoking associations of *ACE2* across epithelial cell types show the importance of single-cell resolution, as down-regulation in AT2 cells would have been otherwise masked by increases in airway epithelial signal in bulk RNA-Seq[56]. Although we have aggregated over 200 donors in our dataset, effects such as race, ethnicity, genetic ancestry, cumulative smoking, or healthy tissue with a distal disease site may still confound the associations we have obtained.

Our models included tested covariates, technical covariates, and interaction terms, which allowed us to uncover complex associations (*e.g.*, sex and smoking associations are typically stronger for younger individuals; Supplementary Fig. 5). Modeling the smoking status of a donor was important to reduce background variation and account for the unbalanced distribution of covariates. Fitting this model required aggregating many datasets, harmonized by a consistent cell type annotation. However, the annotation remains coarse in some cases, where cell labels still aggregate over considerable diversity, and can be further refined in the future. As the HCA grows and further datasets become available, our model could be extended to allow nonlinear associations with the tested covariates. Such associations may uncover e.g. distinct effects in the particularly affected geriatric population. While there is a trend of increased proportion of $ACE2^{+}TMPRSS2^{+}$ cells with age (Extended Data Fig. 3b,c), this cannot be modeled reliably given the compositional diversity (Fig. 3a, Supplementary Fig. 15), potential confounders, and limited sample numbers. Further metadata can help address this.

Our findings in human and mouse models are consistent with respect to smoking and age associations. In line with our human data, we find an increase in Ace2 expression in maturing mice (2-4 months). Others have reported lower expression of entry factors in aged mice (24 months), showing potential limitations of mice as a model system.

Our comprehensive cross-tissue analysis expands on our[13,14,16,57] and others'[58-60] earlier efforts, identifying cell subsets across tissues that may be implicated in transmission or pathogenesis. For example, double positive cells in the submucosal glands may be a reservoir for viruses that escape from expulsion associated with severe cough in the airway luminal surface. Another intriguing hypothesis is that neurologic symptoms[61-63] and Guillain-Barré Syndrome[64] may arise as an autoimmune response to myelin antigens expressed by infected $ACE2^{+}TMPRSS2^{+}$ and $ACE2^{+}$ cells that express myelin-producing genes (Supplementary Fig. 2, Supplementary Table 7).

ACE2 and *TMPRSS2* expression in lung, nasal and gut epithelial cells is associated with programs involving key immunological genes and genes related to viral infection. Expression of *IL6*, *IL6R* and *IL6ST* in lung epithelial cells raises the hypothesis that infection may trigger uncontrolled cytokine expression, as elevated IL-6 levels were reported in more severe COVID-19 patients[65] . The prediction of TNF, complement, and IL1 pathways may suggest a benefit for therapies that target these axes. The accessibility of

STAT and IRF binding sites in scATAC-Seq data is consistent with interferon regulation of *ACE2* expression in epithelial cells[14] and with high activity of STAT1/2 and IRF1/2/5/7/8/9 in macrophage states increased in severe COVID-19 patients[66]. Future lines of inquiry could include investigating the impact of lysosomal genes in lung secretory and multiciliated cells on viral infection and of *RIPK3* expression in airway cells on necroptosis.

Finally, the expression of other potential accessory proteases may help pursue therapeutic hypotheses related to disruption of viral processing via protease inhibition. *FURIN*, *PCSK5* and *PCSK7* are more broadly expressed than *TMPRSS2* across lung cell types (Fig. 2d) and across tissues (Extended Data Fig. 6i). Viral proteins may physically interact with PCSK6[50], which is significantly co-expressed with *ACE2* in AT2 cells (Fig. 2b, Extended Data Fig. 6b). Because PCSKs are localized in different membrane compartments[26], they might process SARS-CoV-2 S-proteins at different viral stages. Altogether, this could provide SARS-CoV-2 with immense flexibility in entry and egress.

Our meta-analysis provides a detailed molecular and cellular map to aid in our understanding of SARS-CoV-2 transmission, pathogenesis and clinical associations. We have demonstrated here how this can be done despite restrictions on data sharing. As the HCA progresses, we envision such meta-analyses in the context of other diseases, for example by combining large healthy reference atlases with both epidemiological and genetic risk factors. In parallel, as new atlases are generated from COVID-19 tissues and models, their integration will further advance our understanding of this disease.

METHODS

Patient samples

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)

Integrated analysis of published datasets

Publicly available (Supplementary Table 1) single-cell RNA-seq datasets were downloaded from Gene Expression Omnibus (GEO). We searched GEO for datasets that met all of the following criteria: (1) provided unnormalized count data; (2) was generated using the 10X Genomics’s Chromium platform; and (3) profiled human samples. These samples spanned a wide range of tissues, including primary tissues, cultured cell lines, and chemically or genetically perturbed samples. Applying these filters increases standardization of sample as the vast majority were prepared using the same 10X Chromium instrument and Cell Ranger pipelines.

Datasets comprise of one or more samples (individual gene expression matrices), which often correspond to individual experiments or patient samples. In total, this yielded 2,333,199 cells from 469 samples from 64 distinct datasets (Supplementary Table 1). To allow comparison across samples and datasets, we mapped through a common dictionary of gene symbols and excluded unrecognized symbols. If a gene from an aggregated master list was not found in a sample, the expression was considered to be zero for every cell in that sample.

After all datasets were collected, we quantified the percentage of cells with >0 UMIs for both *ACE2* and *TMPRSS2* or *ACE2* and *CTSL*. For further analyses with broad cell classes, we only used datasets with more than 15 double positive cells yielding 252,871 cells from 40 samples.

For integration across datasets, we used two levels of annotations. When possible, every sample was annotated with its tissue of origin based on the available metadata from GEO. We excluded any sample for which tissue was not specified. For the smaller subset of 252,871 cells we then manually annotated cell clusters with broad cell type classes using marker genes. These clusters were generated using the harmony-pytorch Python implementation (version 0.1.1 (https://github.com/lilab-bcb/harmony-pytorch) of the Harmony scRNA-seq integration method[67] for batch correction and leiden clustering from the Scanpy package (version 1.4.5). Clusters without clear markers distinguishing types were excluded from further analysis.

Data was processed using Scanpy. Individual datasets were normalized log (UMIs/10,000 +1) by column sum and the log1p function ($\ln(10{,}000 * g_{ij} + 1)$ where a gene's expression profile, g, is the result of the UMI count for each gene, *i*, for cell *j*, normalized by the sum of all UMI counts for cell *j*. This data normalization step was only used for generating the clusters and cell type annotations.

All other statistical tests for the integrated analysis were performed on the cell's binary classification as a double positive or not. For example, for a cell to be considered ACE2+, it has >0 ACE2 transcripts. Double positive cells have >0 transcripts for both genes of interest. We used Fisher's exact test to test for statistical dependence between the expression of *ACE2* and *TMPRSS2* or *CTSL* and corrected for multiple testing via Benjamini-Hochberg over all tests for each gene pair.

Bronchial brushings from current and former smokers

Bronchial brushings were obtained from high-risk subjects undergoing lung cancer screening at ~1-year intervals by white light and autofluorescence bronchoscopy and computed tomography (n=137 brushings from n=50 patients, GSE109743) and profiled via RNA-seq as described previously[41]. Differential expression analysis of entry factors in former and current smokers was performed via voom-limma[68] using the model:

$$Y_i \sim smoking + batch + TIN + (1 \mid patient),$$

where smoking denotes the encoded smoking status ("current" or "former"), batch refers to the experimental batch effect derived from the sequencing run, *TIN* represents the RNA integrity score, and *(1 / patient)* is a random intercept per patient. Multiple testing correction was performed via Benjamini-Hochberg to obtain an FDR-corrected p-value.

Integrated co-expression analysis of high resolution cell annotations across tissues

We compiled a compendium of published and unpublished datasets consisting of 2,433,890 cells from 21 tissues and/or organs including adipose, bone marrow, brain, breast, colon, cord blood, enteric nervous system, esophagus mucosa, esophagus muscularis, anterior eye, heart, kidney, liver, lung, nasal, olfactory epithelium, pancreas, placenta, prostate, skeletal muscle and skin. After the harmonization of cell type annotations, *ACE2*-*TMPRSS2* and *ACE2*-*CTSL* coexpression were assessed using a logistic mixed effect model:

$$Y_i \sim ACE2 + (1 \mid sample\_id) \quad (1)$$

where Y_i was the binarized expression level of either *TMPRSS2* or *CTSL*, and covariates were binarized *ACE2* expression in cell i and a sample-level random intercept.

Models were fit separately for each cell type in each dataset. In order to avoid spurious associations in cell types with very few *ACE2*$^+$ cells and due to very low expression of *ACE2*, we subsampled *ACE2*$^-$ cells to the number of *ACE2*$^+$ cells within each cell type and discarded cell types containing fewer than 5 cells expressing either *ACE2* or fewer than 5 cells expressing the other gene being tested after the subsampling procedure. The significance of the association between *ACE2* and *TMPRSS2*/*CTSL* is controlled for 10% FDR using the statsmodels Python package (version 0.11.1)[69]. Data processing was performed using Scanpy Python package (version 1.4.6)[70] and logistic models were fit using lme4 R package (version 1.1.21)[71].

Single-cell ATAC-Seq analysis

Library Generation and Sequencing.—We performed single-cell ATAC-seq from primary carina and subpleural parenchyma of one individual (n=3 samples per location). Libraries were generated using the 10x Chromium Controller and the Chromium Single Cell ATAC Library & Gel Bead Kit (#1000111) according to the manufacturer's instructions (CG000169-Rev C; CG000168-Rev B) with unpublished modifications relating to cell handling and processing. Briefly, human lung derived primary cells were processed in 1.5ml DNA LoBind tubes (Eppendorf), washed in PBS via centrifugation at 400g, 5 min, 4C, lysed for 3 min on ice before washing via centrifugation at 500g, 5 min, 4C. The supernatant was discarded and lysed cells were diluted in 1x Diluted Nuclei buffer (10x Genomics) before counting using Trypan Blue and a Countess II FL Automated Cell Counter to validate lysis. If large cell clumps were observed, a 40µm Flowmi cell strainer was used prior to the tagmentation reaction, followed by Gel Bead-In-Emulsions (GEMs) generation and linear PCR as described in the protocol. After breaking the GEMs, the barcoded tagmented DNA was purified and further amplified to enable sample indexing and enrichment of scATAC-seq libraries. The final libraries were quantified using a Qubit dsDNA HS Assay kit (Invitrogen) and a High Sensitivity DNA chip run on a Bioanalyzer 2100 system (Agilent).

All libraries were sequenced using Nextseq High Output Cartridge kits and a Nextseq 500 sequencer (Illumina). 10x scATAC-seq libraries were sequenced paired-end (2 x 72 cycles).

Initial data processing and QC.: Fastq files were demultiplexed using 10x Genomics CellRanger ATAC mkfastq (version 1.1.0). We obtained peak-barcode matrices by aligning reads to GRCh38 (CR v1.2.0 pre-built reference) using CellRanger ATAC count. Peak-barcode matrices from six channels were normalized per sequencing depth and pooled using CellRanger ATAC aggr.

The aggregated, depth-normalized, filtered dataset was analyzed with Signac (v0.1.6, https://github.com/timoast/signac), a Seurat[72] extension developed for the analysis of scATAC-seq data. All the analyses in Signac were run with a random number generator seed set as 1234. Cells that appeared as outliers in QC metrics (peak_region_fragments ≤ 750 or peak_region_fragments ≥ 20,000 or blacklist_ratio ≥ 0.025 or nucleosome_signal ≥ 10 or TSS.enrichment ≤ 2) were excluded from the analysis.

Normalization and dimensionality reduction.: The aggregated dataset was processed with Latent Semantic Indexing[73], i.e. datasets were normalized using term frequency-inverse document frequency (TF-IDF), then singular value decomposition (SVD), ran on all binary features, was used to embed cells in low-dimensional space. Uniform Manifold Approximation and Projection (UMAP)[74] was then applied for visualization, using the first 30 dimensions of the SVD space.

Gene activity matrix and differential motif activity analysis.: A gene activity matrix was calculated as the chromatin accessibility associated with each gene locus (extended to include 2kb upstream of the transcription start site, as described in the vignette 'Analyzing PBMC scATAC-seq' (version: March 13, 2020, https://satijalab.org/signac/articles/pbmc_vignette.html), using as gene annotation the genes.gtf file provided together with Cellranger's atac GRCh38-1.2.0 reference genome. For the motif analysis, we note that because epithelial cells with an accessible *ACE2* locus tend to have a higher number of fragments in peaks than cells with inaccessible *ACE2* (Supplementary Fig. 1e), consistent also with higher UMIs in scRNA-seq, some of the cells with inaccessible *ACE2* could be false negatives, reducing our power.

Clusters were annotated using label transfer from matching scRNA samples or by literature / expert search of marker "active" (i.e. accessible) genes. Differential motif activity analysis was performed using Signac's implementation of ChromVAR[75], with motif position frequency matrices from JASPAR2020[76] (http://jaspar.genereg.net/) selecting transcription factors motifs from human (species=9606), broadly following the vignette 'Motif analysis with Signac' (https://satijalab.org/signac/articles/motif_vignette.html). Cells were identified as positive for *ACE2* and/or *TMPRSS2* (i.e. with the loci accessible) if at least one fragment was overlapping with the gene locus or 2kb upstream. Differential activity scores between epithelial cells positive for *ACE2* (with the above-mentioned definition of 'positive') and non-expressing *ACE2* was performed with the *FindMarkers* function of Seurat (version 3.1.1), using as test 'LR' (i.e. logistic regression) and as latent variable the number of counts in peak. The function constructs a logistic regression model predicting group membership based on each motif score individually and compares this to a null model with a likelihood ratio test. Adjusted p-value is the result of Bonferroni correction.

Immunohistochemistry and Proximity ligation *in situ* hybridization (PLISH)

Proximity ligation in situ hybridization (PLISH) was performed as described previously76. Briefly, frozen human trachea and distal lung sections were fixed with 4.0% paraformaldehyde for 20 min, treated with protease (20 μg/mL proteinase K for lung or Pepsin for trachea for 9 min) at 37°C, and dehydrated with up-series of ethanol. The sections were incubated with gene-specific oligos (Supplementary Table 6) in hybridization buffer (1 M sodium trichloroacetate, 50 mM Tris [pH 7.4], 5 mM EDTA, 0.2 mg/mL heparin) for 2 h at 37°C. Common bridge and circle probes were added to the section and incubated for 1 h followed by T4 ligase reaction for 2 h. Rolling circle amplification was performed by using phi29 polymerase (#30221, Lucigen) for 12 hours at 37°C. Fluorophore-conjugated detection probe was applied and incubated for 30 min at 37°C. For combination of PLISH and Immunostaining, sections were incubated with primary antibody for HTII-280 (Terrace Biotech, TB-27AHT2-280), pro-SFTPC (Millipore, ab3786) or ACTA2 (Sigma, F3777) for 1 h at room temperature. Sections were incubated with secondary antibody (goat anti-mouse IgM secondary antibody (Thermo Scientific, A21044) or donkey anti-rabbit IgG secondary antibody (Thermo Scientific, A32795) for 45 min at room temperature, then sections were mounted in medium containing DAPI. We imaged three representative areas per patient for three patients total for images and quantification shown in Fig. 1 and imaged one representative area for a single patient for Extended Data Fig. 7a,c,d,g. Images were captured using Olympus Confocal Microscope FV3000 with Olympus FLUOVIEW FV31S-SW v2.1.1.98 using 20× or 60× objective.

THS-Seq on human pediatric samples

THS-Seq was performed as previously reported[18] on human pediatric samples (full gestation, with no known lung disease) collected at day 1 of life, 14 months, 3 years, and 9 years (n=1 at each time point).

Integrated analysis for associating *ACE2, TMPRSS2, and CTSL* expression with age, sex and smoking status in nasal, airway and lung cells

To assess the association of age, sex, and smoking status with the expression of *ACE2*, *TMPRSS2*, and *CTSL*, we aggregated 31 scRNA-seq datasets of healthy human nasal and lung cells, as well as fetal samples containing the expression counts of only the 3 genes. Aggregation of these datasets was enabled by harmonizing the cell type labels of individual datasets and dataset concatenation within Scanpy[70] (version 1.4.5.1). We harmonized annotations manually on the basis of provided cell type labels together with data contributors using a preliminary ontology generated on the basis of 5 published datasets [30-32,36,38] with 3 levels of annotations. Level 1 has the lowest resolution and distinguishes epithelial from stromal/mesenchymal, endothelial and immune cells. Level 2 breaks up each of the level 1 categories in the coarsest available further observed annotations. Level 3 in turn splits up the observed level 2 annotations where finer annotations were available. (Supplementary Table 2, consent to publish was obtained from all contributors). To compare AT2 cells and their fetal progenitors possible, we mapped progenitor cells labeled “AT2-like” and “SpC+ progenitors” to the AT2 label. We further harmonized metadata by collapsing the smoking covariate into “has smoked” and “has never smoked” and by taking mean ages where only age ranges were given. This resulted in a dataset of 1,320,896 cells and 3 genes in 377 samples from 228 donors (the cell by three-gene count matrix with annotations is available on the Single Cell Portal (SCP1257)). We divided the data into fetal (136,450 cells, 41 samples, 34 donors), adult nasal (57,548 cells, 20 samples, 18 donors), and adult lung (1,126,898 cells, 316 samples, 187 donors) datasets based on metadata provided.

To get an overview of sample diversity, we clustered the samples using the proportion of cells in level 2 cell types as features. Clustering was performed using louvain clustering (resolution 0.3; louvain package version 0.6.1) on a knn-graph (k=15) computed on Euclidean distances over the top 5 principal components of the cell type proportion data within Scanpy. This produced four clusters. Sample cluster labels were assigned based on cell type compositions and metadata for anatomical location that was obtained from the published datasets and via input from the data generators.

Within non-fetal datasets we modeled the association of age, sex, and smoking status with gene expression for *ACE2*, *TMPRSS2*, and *CTSL* within each cell type using a generalized linear model with the log total counts per cell as offset and Poisson noise as implemented in statsmodels[69] (version 0.11.1) and using a Wald test from Diffxpy (www.github.com/theislab/diffxpy; version 0.7.3, batchglm version 0.7.4). Specifically, we fit the model:

$$Y_{ij} \sim age + sex + age{:}sex + smoking + sex{:}smoking + age{:}smoking + dataset, \quad (2)$$

which models effects of age, sex and smoking while accounting for potential interactions between covariates and the uneven distribution of covariates across the dataset. Here, Y_{ij} denotes the raw count expression of gene i in cell j, age, sex, and smoking denote the modeled covariates, and *age:sex*, *sex:smoking*, and *age:smoking* represent the interaction terms between these covariates. The interaction terms model whether there is a difference in the smoking effect in men and women, and likewise whether the age effect is different for smokers and non-smokers. We included the *dataset* term to model the technical variation (e.g., sampling and processing differences) between the diverse datasets, and the log total counts per cell was used as an offset. Here, the total counts were scaled to have a mean of 1 across all cells before the log was taken. Due to the inclusion of interaction terms, the complex interaction model (2) fits the overall effects of age (k_{age}), sex (k_{sex}), and smoking ($k_{smoking}$) as linear functions of the other two covariates respectively, given by the equations:

$$\begin{aligned} k_{age}(sex, smoking) &= \beta_{age} + sex\beta_{age:sex} + smoking\beta_{age:smoking}, \\ k_{sex}(age, smoking) &= \beta_{sex} + age\beta_{age:sex} + smoking\beta_{sex:smoking}, \\ k_{smoking}(age, sex) &= \beta_{smoking} + age\beta_{age:smoking} + sex\beta_{sex:smoking}. \end{aligned}$$

Here, β_{age} and $\beta_{age:sex}$ represent the model coefficients for age and the interaction of age and sex in model (2) respectively, and *age* denotes the age covariate. Sex and smoking covariates were converted into a one-hot encoded format such that *sex=0* denoted females and *smoking=0* denoted non-smokers. As linear dependencies on covariates can be summarized by showing 2 values per covariate, we displayed effect sizes for the overall age, sex, and smoking associations by computing k_{age}, k_{sex}, and $k_{smoking}$ for sex∈{0,1}, smoking∈{0,1}, and age∈{31,62} (the first and third quartiles of the age distribution). Standard errors for these effects were computed using the variance-covariance matrix Σ via $SE = \sqrt{C^T \Sigma C}$, where SE is the standard error and C is the vector of covariate values used to compute the respective overall effect (e.g., k_{age}). P-values were obtained using a Wald test, and multiple testing correction was performed over all tests on the same cell type data via Benjamini-Hochberg. In order to fit this model we pruned the data to contain only datasets that have at least 2 donors and for which smoking status metadata was provided. This resulted in a dataset of 985,420 cells and 286 samples from 164 donors for adult lung data. Only 15 donors remained for adult nasal data after this filtering, which we deemed too few to obtain robust results. To obtain cell-type specific associations the above model was fit within each cell type for all cell types with at least 1,000 cells.

While cells from different donors are not truly independent observations, model (2) treats them as such and thus models cellular and donor variation jointly. As donor variation tends to be larger than single-cell variation, when most cells come from few donors (either there are few donors, or few donors contribute most of the cells), this can lead to an inflation of p-values. To counteract this effect, we verified that significant associations are consistent when modeling only donor variation via pseudo-bulk analysis (Supplementary Data 1-4). Furthermore, we tested whether effects are dependent on few donors by holding out datasets.

Pseudo-bulk data was generated by computing the mean for each gene expression value and nUMI covariate for cells in the same cell type and donor. After filtering as described above, model (2) was fit to the data (Supplementary Data 1-4). In contrast to the single-cell model, pseudo-bulk analysis underestimates certainty in modeled effects as uncertainty in the pseudo-bulk means are not taken into account when estimating background variance. Thus, we used only effect directions from pseudo-bulk analysis to validate single-cell associations. In further analysis, we regarded only those associations as confirmed by pseudo-bulk analysis, where the FDR-corrected p-value in the single-cell model is below 0.05, and the sign of the estimated effect is consistent in both the single-cell and the pseudo-bulk analysis.

We further separated significant associations into robust trends and indications depending on the holdout analysis. A significant association was regarded as a *robust trend* if the effect direction is consistent when holding out any dataset when fitting the model (without considering the p-value). In the case that holding out one dataset caused the maximum likelihood estimate of the coefficient to be reversed, we denote this as the effect *no longer being present*, which characterized the association as an *indication*. Two dataset holdouts led to *indications* in our analysis: the largest declined donor transplant dataset (Supplementary Table 2, "Regev-Rajagopal", most cells and most samples; indication in *ACE2* multiciliated lineage age and sex associations, and *CTSL* AT1 sex association), and a declined donor tracheal epithelium dataset (“Seibold”, Supplementary Table 2, most donors in the smoking analysis; *CTSL* basal smoking association).

At least 4 values per covariate are required to describe a single association in model (2) (e.g., male non-smoker, female non-smoker, male smoker, and female smoker for the k_{age} effect). To summarize these effects and present a single association per covariate, we also fit the simplified model:

$$Y_{ij} \sim age + sex + smoking + dataset\,. \qquad (3)$$

As in model (2), the logarithmized, scaled total counts per cell were used as an offset, data were filtered as described, and multiple testing correction was performed via Benjamini-Hochberg. To increase the robustness of our reported associations, we again performed pseudo-bulk and holdout analysis. Additionally, to still account for covariate interactions, we discarded associations where the complex model (2) and the simplified model (3) results were inconsistent. Here, consistency was defined by two criteria: at least one model (2) indication or robust trend in the same direction as the model (3) effect, and no model (2) indication or robust trend in the opposite direction to the model (3) effect.

As metadata on smoking status was only available for a subset of the data, we also fitted a reduced version of models (2) and (3) without the smoking covariate on a larger dataset to confirm sex and age associations (Supplementary Data 5-8). The non-smoking model was fit on 1,096,604 cells in 309 samples from 185 donors of adult lung data. Again, log total counts (scaled) was used as an offset, pseudo-bulk and holdout analysis was performed, and associations from the simple model were tested for consistency with the complex model.

Normalizing *ACE2*$^+$*TMPRSS2*$^+$ double positive fractions of human lung samples

Proportions of *ACE2*$^+$*TMPRSS2*$^+$ cells (Extended Data Fig. 3a, Supplementary Fig. 15) were normalized to account for differences in total UMI counts. Normalization was done per donor, per cell type by calculating $\frac{X_{i,j}}{N_{i,j}} * 10,000$, where $X_{i,j}$ is the DP fraction of cell type i in donor j, and $N_{i,j}$ represents the median total UMI count of cells of type i in donor j.

Identification of gene programs using feature importance for a random forest trained to classify *ACE2*$^+$*TMPRSS2*$^+$ *vs ACE2-TMPRSS2*$^-$ cells

To infer tissue programs, we trained a random forest classifier to discriminate between double positive and double negative cells (excluding *ACE2* and *TMPRSS2*; 75:25 class balanced test-train split), generalizing across multiple cell types in one tissue, and ranked genes according to their importance scores in the classifier. To infer cell programs, we performed differential expression analysis between double positive and double negative cells within each cell subset.

Importantly, these methods do not assume that *ACE2*$^+$*TMPRSS2*$^+$ cells form a distinct subset within each cell type. Rather, our goal is to leverage the variation among single cells within a single type to identify gene programs that are co-regulated with *ACE2* and *TMPRSS2* within each expressing cell subset.

For each of the lung, nasal, and gut datasets, we labeled the cells with non-zero counts for both *ACE2* and *TMPRSS2* as double positive cells (DPs), and the cells with zero counts for both *ACE2* and *TMPRSS2* as double negative cells (DNs). Within each tissue, we identified cell types with greater than 10 DPs, and for each of these cell types, we selected the genes with increased expression (log fold change greater than 0) in DPs vs DNs (so that we focus on important ”positive” features). We trained a classifier with 75:25 train:test split to classify the DPs from DNs within each of these cell types using the *sklearn* (version 0.21.3) [77] *RandomForestClassifier* function with the following parameters: *n_estimators* set to *100*, the *criterion* as *gini*, and the *class_weight* parameter set to *balanced_subsample*. We first trained individual classifiers separately for each of the cell types, and pooled genes with positive feature importance values (using the *feature_importance*[78] field in the trained *RandomForestClassifier* object) to train a final DP vs DN classifier across each tissue. We used the top 500 genes, as ranked by their feature importance scores, to define the signature for the gene expression program of DPs for the tissue. This procedure was carried out in lung, nasal, and gut datasets, yielding tissue-specific signatures for gene expression programs of DPs from each tissue.

For visualization purposes only, we generated network diagrams using the *networkx* (version 2.2) tool with the *ForceAtlas2* (version 0.3.5) graph layout algorithm [79]. We scored genes that appeared in signatures for multiple tissues by their aggregated feature importance (using a plotting heuristic that used the sum of importance ranks for genes in individual tissues and by assigning a large valued rank (10000) to a gene that did not appear in a particular tissue) and selected the top 10 genes that were shared by each pair of tissues or shared by all tissues along with additional genes that included the ones unique to each tissue’s signature to plot in the network visualization. The GO terms enriched in the gene expression programs shared by DPs across tissues were found using *gprofiler* (version 1.0.0) [80] using the *scanpy.queries.enrich* tool.

This analysis was performed in two ways: on the original data, as well as after accounting for differences in distribution of the number of UMIs (nUMI) per cell between DPs and DNs. This was done by binning the nUMI distribution in the DPs for each tissue into a 100 bins and then randomly sampling from the nUMI distribution for the DNs in each bin to match the distribution of the DPs in that bin. The nUMI distributions before and after the matching are shown in Supplementary Fig. 11b.

Identification of gene programs enriched in DP *vs*. DN cells using regression

In parallel, we used a regression framework to recover gene modules enriched in DP *vs.* DN cells (Fig. 4c,d, Supplementary Fig. 12a,b) in the nasal, lung, and gut datasets. We first restricted our analysis to cell subsets derived from at least two donor individuals that each contained a mixture of DN and DP cells (Nawijn Nasal: multiciliated, Goblet; Regev/Rajagopal Lung: AT1, AT2, Basal, multiciliated, Secretory; Aggregated Lung: AT2, multiciliated, Secretory; Regev/Xavier Colon: *BEST4*$^{+}$ Enterocytes, Cycling TA (Transit Amplifying), Enterocytes, Immature Enterocytes 2, TA-2). For each of these cell subsets, we then used MAST (version 1.8.2) [81] to fit the following regression model to every gene with cells as observations:

$$Y_i \sim X + (1 \mid S),$$

where Y_i is the expression level of gene i in cells, measured in units of log2(TP10K+1), X is the binary co-expression state of each cell (*i.e.* DP *vs.* DN), and S is the donor that each cell was isolated from. To control for donor-specific effects (*i.e.* batch effects), we used a mixed model with a random intercept that varies for each donor. To fit this model, we subsampled cells from DP and DN groups to ensure that both the donor distribution and the cell complexity (*i.e.* the number of genes per cell) were evenly matched between the two groups, as follows. First, for each subset, we restricted our analysis to donors containing at least two DN and two DP cells. Using these samples, we partitioned the cells into 10 equally-sized bins based on cell complexity and subsampled DN cells from each bin to match the cell complexity distribution of the DP cells. Finally, we fit the mixed model (above), controlling for both donor and cell complexity.

To build gene modules for DP cells, we prioritized genes by requiring that they be expressed in at least 10% of DP cells, and to have a model coefficient greater than 0 with an FDR-adjusted p-value less than 0.05 (for the combined coefficient in the hurdle model). After this filtering step, genes were ranked by their model coefficient (*i.e.* estimated effect size). The top 12 genes were selected for network visualization within each cell type (Fig. 4c,d, Supplementary Fig. 12a,b). In three cases (gut Cycling TA, TA-2 and *BEST4*$^+$ cells), *RP11-** antisense genes were flagged and excluded from visualizations. To visualize overlap across each network, we indicated whether each gene was among the top 250 genes from each of the other cell types. Putative drug targets were identified by querying the Drugbank database[49]. Gene set enrichment analysis was performed using the R package EnrichR (version 1.0)[82], selecting the top 25 genes from each cell type for the pan-tissue analysis ("All" category; Fig. 4e), and the top 50 genes from each cell type for the tissue-specific analyses ("Nose", and "Lung" categories; Fig. 4e). We note a few caveats/challenges/limitations that may influence our results, including non-uniform sampling across donors; variation in cell compositions across regions (e.g., distal lung vs carina), and additional cellular heterogeneity that the current level of broad subset annotation may not have been captured.

Cell-Cell interaction analysis

CellphoneDB [53] v.2.0.0 was run with default parameters on the 10 human lung samples of the Regev/Rajagopal dataset (41 samples, 10 patients, 2-6 locations each), analyzing the cells from each dissected region separately. For each sample (patient/location combination), for each cell type we distinguished double positive cells (*ACE2* > 0 and *TMPRSS2* > 0) from all others. Only interactions highlighted as significant, i.e. present in the "significant means" output (p <0.05) from CellphoneDB were considered. AT2 cells and myeloid cells were present in lung lobes samples from all 10 patients, whereas samples from 5 patients contained both *ACE2*$^+$*TMPRSS2*$^+$ double positive AT2 cells and myeloid cells.

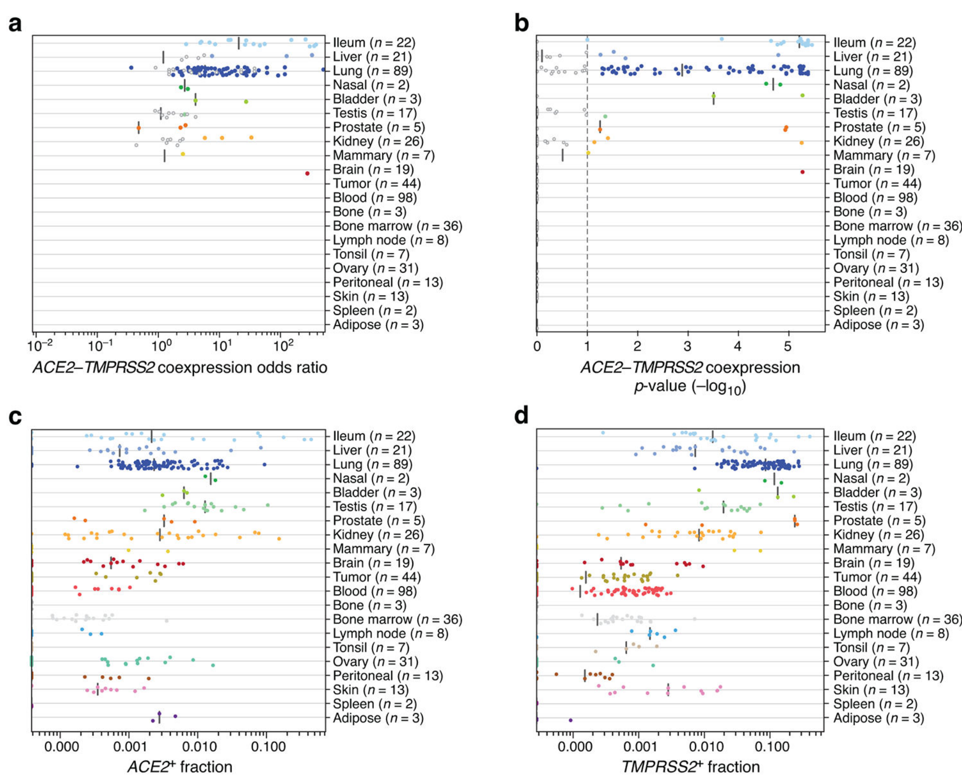
Co-expression patterns of additional proteases and *IL6*/*IL6R*/*IL6ST*

ACE2-protease co-expression (Fig. 2, Extended Data Fig. 5) and *ACE2*-*IL6*/*IL6R*/*IL6ST* co-expression (Supplementary Fig. 13) were tested via the logistic mixed-effects model described in "Integrated co-expression analysis of high resolution cell annotations across tissues" (Equation 1, above).

Mouse smoke exposure experiments

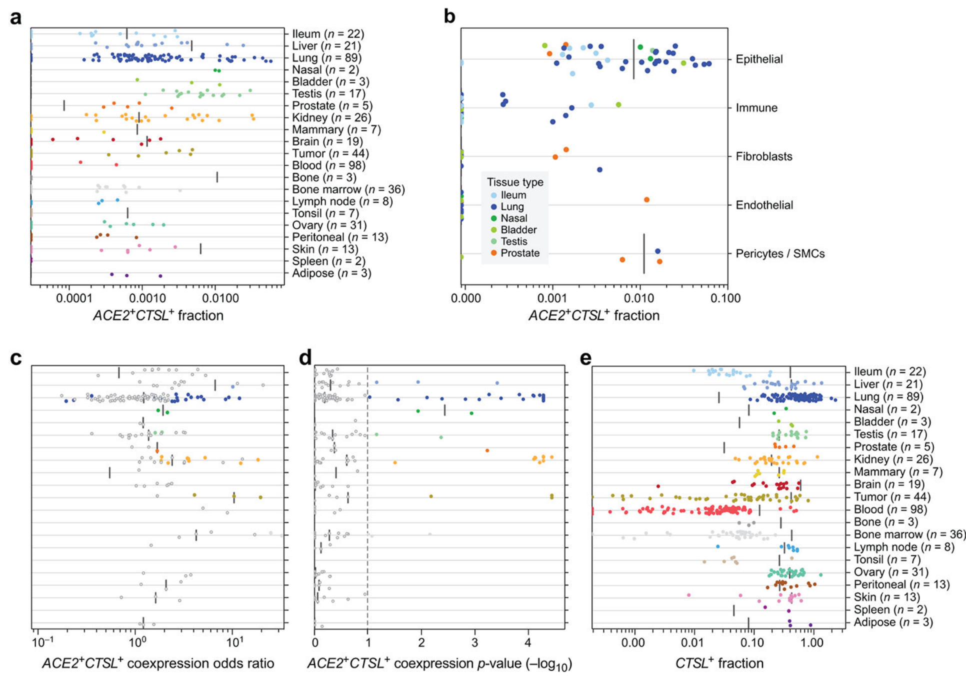
For these experiments, 8 to 10 week old pathogen-free female wild-type C57BL/6 mice were obtained from Charles River (Sulzfeld, Germany) and housed in rooms maintained at constant temperature and humidity with a 12 hour light cycle. Animals were allowed food and water ad libitum. All animal experiments 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. The female C57BL/6 mice (n=5) were whole body exposed to 100% mainstream cigarette smoke at a particle concentration of 500 mg/m3, generated from 3R4F research cigarettes (Filter removed, Tobacco Research Institute, University of Kentucky), for 50 min twice/day, 5 days/week for 2 months to mimic human smoking habits [83]. Control mice (n=3) were exposed to filtered air, but exposed to the same stress as mice exposed to cigarette smokè.

Extended Data



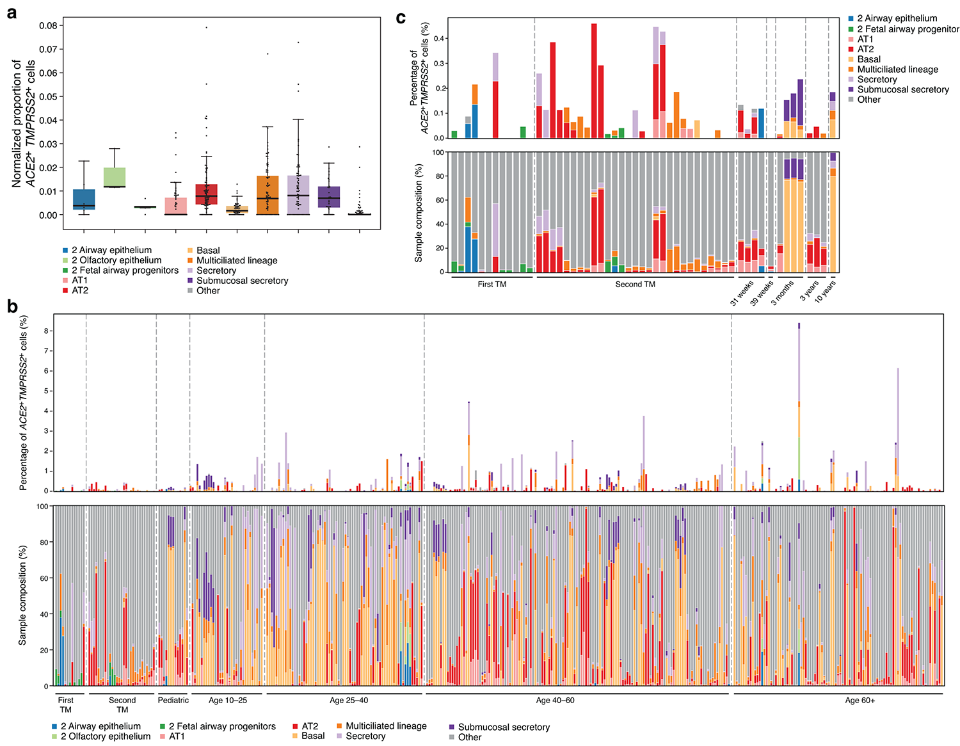
Extended Data Fig. 1. A cross-tissue survey of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in published single-cell datasets.

(**a**) Odds ratio (x axis) of *ACE2*$^{+}$*TMPRSS2*$^{+}$ co-expression in single-cell datasets (dots) from different tissues (y axis). (**b**) Significance ($-\log_{10}$(p-value) using two-sided Fisher's exact test, x axis) of co-expression of *ACE2*$^{+}$*TMPRSS2*$^{+}$ in single-cell datasets (dots) from different tissues (y axis). (**c,d**) Proportion (x axis) of *ACE2*$^{+}$ cells per dataset (c) and *TMPRSS2*$^{+}$ cells per dataset (d) across different tissues (y axis).



Extended Data Fig. 2. A cross-tissue survey of *ACE2*$^{+}$*CTSL*$^{+}$ cells in published single-cell datasets.

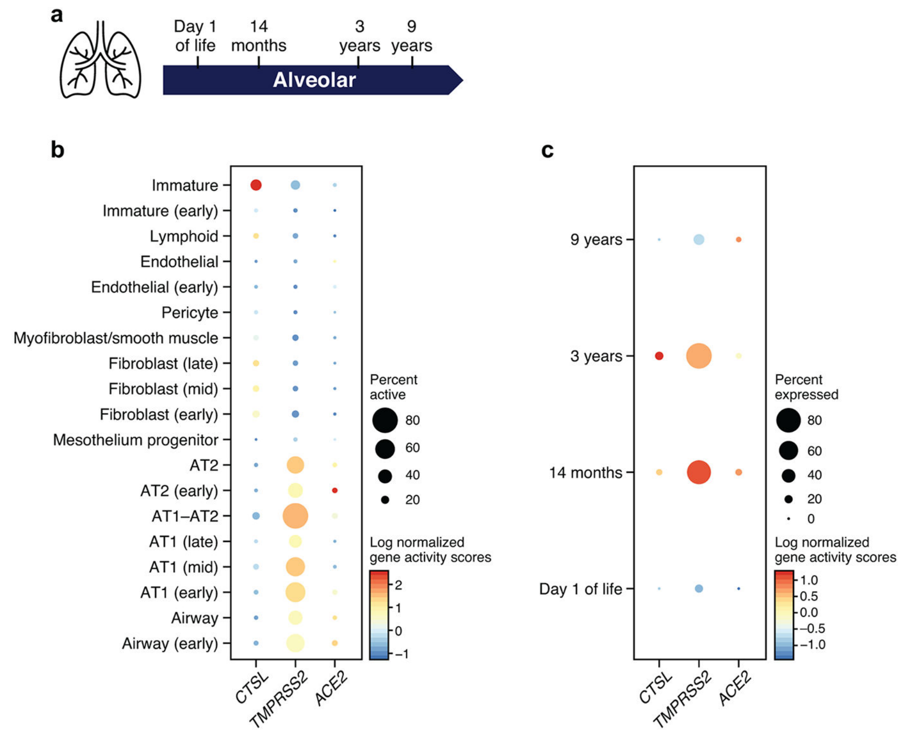
(**a**) Proportion (x axis) of *ACE2*$^{+}$*CTSL*$^{+}$ cells per dataset (dots) across different tissues (y axis). (**b**) Proportion (x axis) of *ACE2*$^{+}$*CTSL*$^{+}$ cells within clusters annotated by broad cell-type categories (dots) in each of the top 7 enriched datasets (y axis; color legend, inset). (**c**) Odds ratio (x axis) of *ACE2*$^{+}$*CTSL*$^{+}$ co-expression in single-cell datasets (dots) from different tissues (y axis). (**d**) Significance ($-\log_{10}$(p-value) using two-sided Fisher's exact test, x axis) of co-expression of *ACE2* and *CTSL* in single-cell datasets (dots) from different tissues (y axis). (**e**) Proportion (x axis) of *CTSL*$^{+}$ cells per dataset across different tissues (y axis).



Extended Data Fig. 3. Cellular composition and fraction of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells across the aggregated lung dataset

(**a**) Boxplot of normalized donor fractions of *ACE2*$^{+}$ *TMPRSS2*$^{+}$ (double positive - DP) cells per cell type. The box indicates the median and first and third quartile, whiskers extend to points within 1.5 times the interquartile range. For each cell type, only donors that have at least 100 cells of the cell type were included. Cell types with at least 10 *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in the entire dataset were labeled, the remaining cell types were grouped under 'Other'. Cell type labels preceded by a "2" consist of cells that had no annotation available at level 3 and therefore kept their level 2 annotation. Cells with only level 1 annotations were grouped under "Other". (2_Airway epithelium: n=6, 2_Olfactory epithelium: n=3, 2_fetal airway progenitors: n=5, AT1: n=60, AT2: n=92, Basal: n=56, Multiciliated lineage: n=88, Secretory: n=79, Submucosal Secretory: n=35, Other: n=180 donors.)

(**b**) Percentage of *ACE2*$^{+}$ *TMPRSS2*$^{+}$ cells across 377 samples and with sample composition. Top: Percentage *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in each sample, categorized by level 3 annotations. Bottom: Sample compositions. Samples are ordered by age, with 31-week pre-term births and 39-week full-term births both set to age 0. (**c**) Zoom in on fetal and pediatric samples of plot (b). Samples are ordered and labeled by age. Fetal samples are partitioned into first and second trimester (TM) and pediatric samples are divided into 31-week pre-term births, 39-week full term births, 3 month, 3 year, and 10 year old children. AT1, 2: alveolar type 1, 2. AT2 progenitor cells were grouped under AT2.



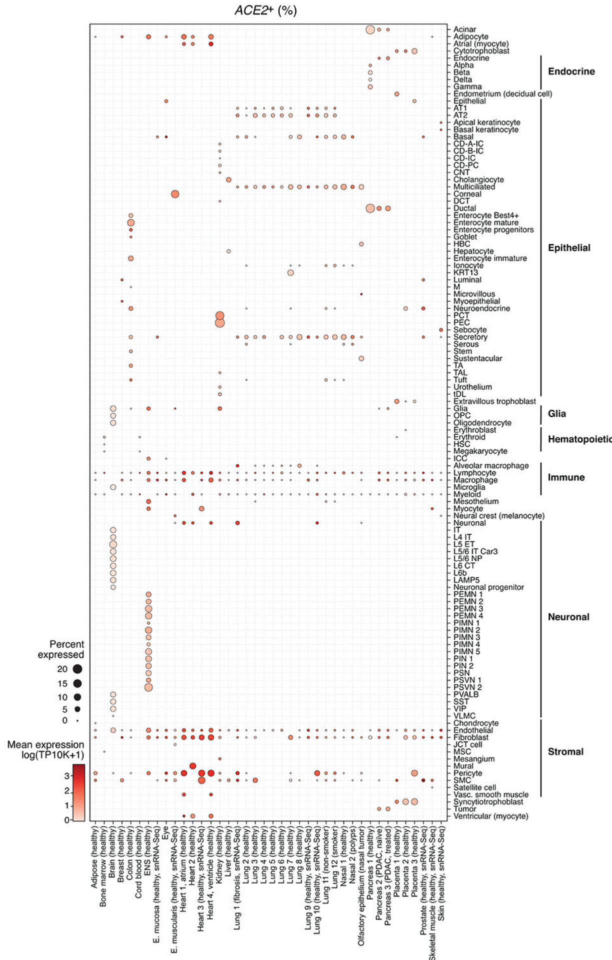
d scTHS-Seq accessibility in all cell types by age: *ACE2*, *TMPRSS2* and *CTSL* (human)

Time point	Day 1	14 mos.	3 yr	9 yr
ACE2+*CTSL*+	81	72	113	4
ACE2+*TMPRSS2*+	349	243	384	54
Total cells	11,444	9077	10,510	4232

Extended Data Fig. 4. Chromatin accessibility at the *ACE2, TMPRSS* and *CTSL* loci across lung cells in early life

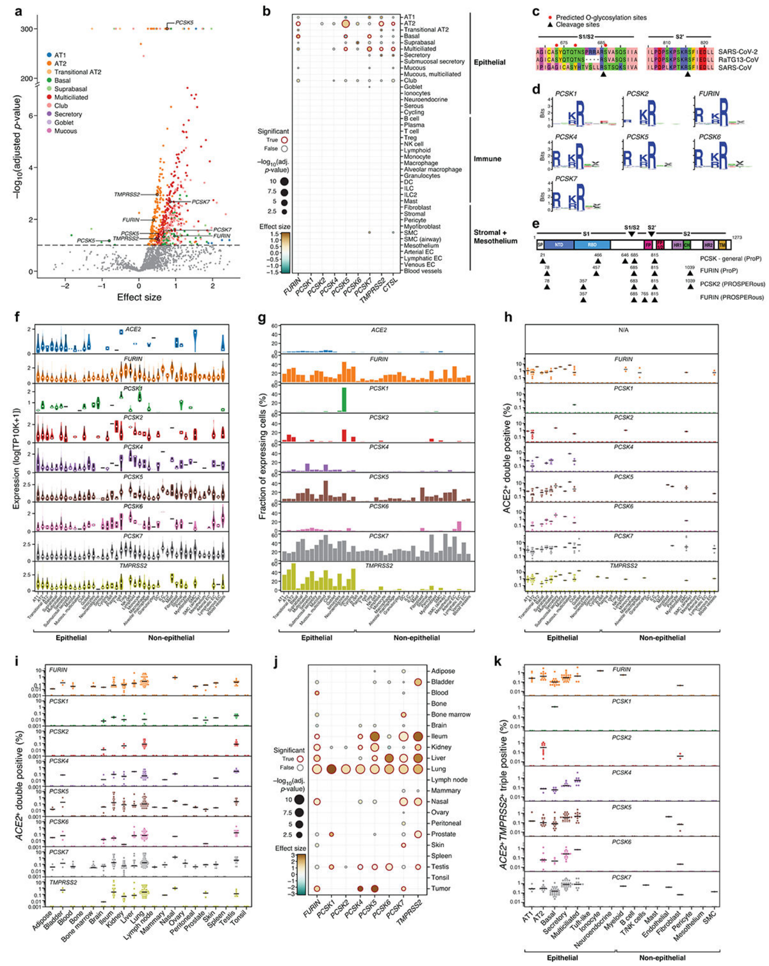
(**a**) Schematic: single-cell chromatin accessibility by transposome hypersensitive sites sequencing (THS-Seq) from human pediatric samples (full gestation, no known lung disease) collected at day 1 of life, 14 months, 3 years, and 9 years (n=1 at each time point). (**b**) Accessibility (dot color log normalized gene activity scores), and % of cells with accessible loci (dot size) for the *ACE2, TMPRSS,* and *CTSL* loci (columns) across different cell types (rows) in scTHS-Seq with all time points aggregated. (**c**) Accessibility (dot color log normalized gene activity scores), and % of cells with accessible loci (dot size) of *ACE2, TMPRSS* and *CTSL* in AT1--AT2 cells in scTHS-Seq at day 1 of life, 14 months, 3 years, and 9 years (rows). (**d**) Number of *ACE2*+ *CTSL*+ and *ACE2*+ *TMPRSS2*+ cells per time point.

Author Manuscript
Author Manuscript
Author Manuscript



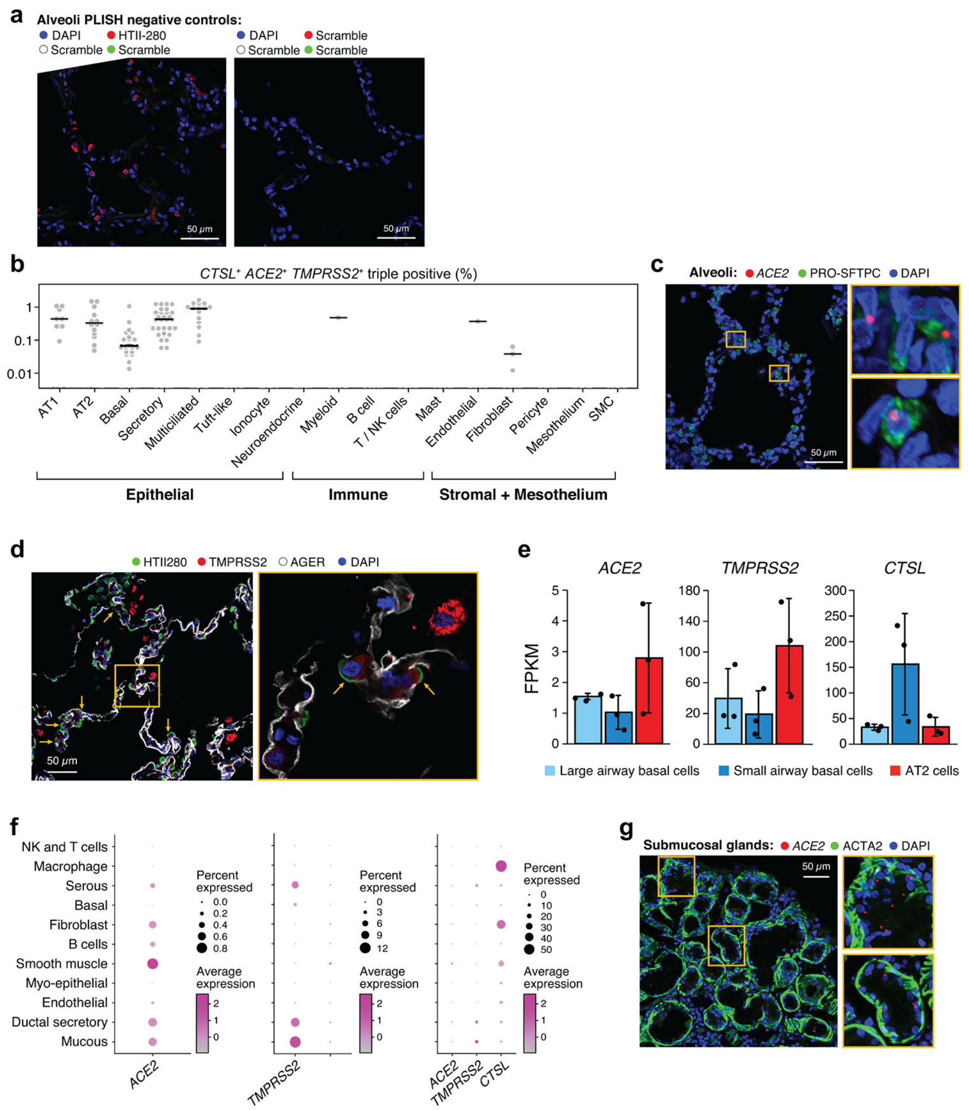
Extended Data Fig. 5. *ACE2* expression across tissues and cell types.

Shown are fractions of *ACE2* expressing cells (dot size) and mean *ACE2* expression level in expressing cells (dot color) across datasets (rows) and cell types (columns).



Extended Data Fig. 6. Additional analyses to identify other proteases that may have a role in infection.

(**a**) Multiple proteases are co-expressed with *ACE2* in another human lung scRNA-seq (“aggregated lung”). Scatter plot of significance (y axis, −log10(adjusted p value) by two-sided Wald test. (Methods)) and effect size (x axis) of co-expression of each protease gene (dot) with *ACE2* within each indicated epithelial cell type (color). Dashed line: significance threshold. *TMPRSS2* and PCSKs that significantly co-expressed with *ACE2* are marked. (**b**) *ACE2*-protease co-expression with PCSKs, *TMPRSS2* and *CTSL* across lung cell types (“aggregated lung”). Significance (dot size, −log10(adjusted p value) by two-sided Wald test. (Methods)) and effect size (color) for co-expression of *ACE2* with selected proteases (columns) across cell types (rows). (c-d) Predicted cleavage sites in the SARS-CoV-2 S-protein S1/S2 region. (**c**) Multiple amino acid sequence alignment of SARS-CoV-2 S-protein S1/S2 region with orthologous sequences from other betacoronaviruses (top) and polybasic cleavage sites of other human pathogenic viruses (bottom). (**d**) Sequence logo plot showing cleavage site preference derived from MEROPS database for PCSK1, PCSK2, FURIN, PCSK4, PCSK5, PCSK6 and PCSK7. (**e**) Protease cleavage sites (triangles) predicted by ProP and PROSPERous in the SARS-CoV-2 spike protein. Top: Full-length SARS-CoV-2 S-protein sequence schematic with predicted functional protein domains and motifs. Numbers: amino acid residues after which cleavage occurs; SP: signal peptide; NTD: N-terminal domain; RBD: Receptor-binding domain; FP: Fusion peptide; FP1/2: Fusion peptide 1/2; HR1: Heptad repeat 1; CH: connecting helix; HR2: Heptad repeat 2; TM: Transmembrane domain. (**f,g**) Multiple proteases are expressed across lung cell types ("aggregated lung"). (f) Distribution of non-zero expression (y axis) for *ACE2*, PCSKs and *TMPRSS2* across lung cell types (x axis). White dot: median non-zero expression. (g) Proportion of cells (y axis) expressing *ACE2*, *PCSK* family or *TMPRSS2* across lung cell types (x axis), ordered by compartment. (**h**) *ACE2*$^{+}$ *PCSK*$^{+}$ double positive cells across lung cell types. Fraction (y axis) of different *ACE2*$^{+}$ *PCSK*$^{+}$ or *ACE2*$^{+}$ *TMPRSS2*$^{+}$ double positive cells across lung cell types, ordered by compartment (x axis). Dots: different samples, line: median of non-zero fractions. (**i,j**) *ACE2*$^{+}$ *PCSK*$^{+}$ co-expression across human tissues (collection of published scRNA seq datasets). (i) Percent (y axis) of different *ACE2*$^{+}$ *PCSK*$^{+}$ or *ACE2*$^{+}$ *TMPRSS2*$^{+}$ double positive cells across human tissues (x axis). Dots: different single-cell datasets, line: median of non-zero fractions. (j) *ACE2* co-expression with PCSKs or *TMPRSS2* across human tissues. Significance (dot size, −log10(adjusted p value) by two-sided Wald test. (Methods)) and effect size (dot color) of co-expression. (**k**) Fraction of *ACE2*$^{+}$ *TMPRSS2*$^{+}$ *PCSK*$^{+}$ cells across lung cell types ("Regev/Rajagopal dataset"). Dots: samples, line: median of non-zero fractions.



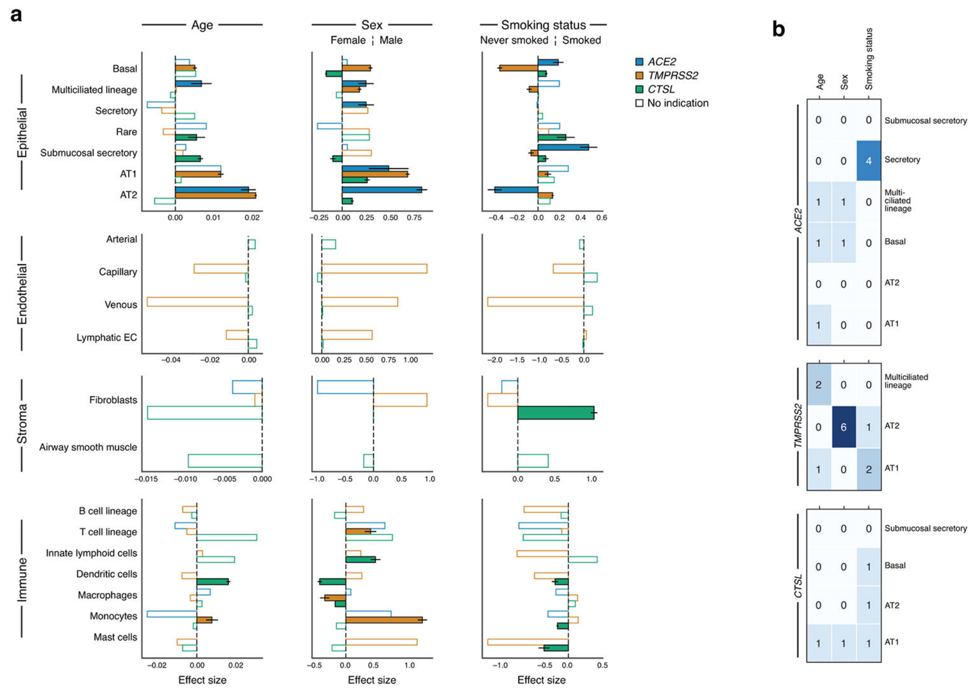
Extended Data Fig. 7. *ACE2, TMPRSS2, CTSL* Immunofluorescence and RNA profiling

(**a**) Negative control of PLISH in human lung alveoli. Left shows scrambled probe detection in three indicated colors. Right shows HTII-280 antibody staining (red) with 2 color scramble probe detection. DAPI (blue) indicates nuclei. (**b**) Frequency of *ACE2, CTLS* and *TMPRSS2* triple positive cells in each sample (n = 60) (dots) in the Regev/Rajagopal dataset. (c) PLISH and immunostaining in human adult lung alveoli for *ACE2* (red), PRO-SFTPC (green), DAPI (blue).

(d) Immunostaining in human adult lung alveoli. HTII-280 (green) , TMPRSS2 (red) and AGER (white). Blue shows DAPI in nuclei. (e) Mean expression (y axis, FPKM, from bulk RNA-seq, error bars: standard error) of *ACE2, CTSL, TMPRSS2* in sorted cells from 3 different human explant donors using the following markers: large and small airway basal cells (NGFR+), AT2 cells (HT-II 280+) and alveolar organoids (HT-II 280+). (**f**) Expression in the submucosal gland. Mean expression (color) and proportion of expressing cells (dot size) of *ACE2*, *TMPRSS2* and *CTSL* in key cell types (rows), from scRNA-seq of human large airway submucosal glands. (g) PLISH and immunostaining in human large airway submucosal glands. *ACE2* (red), *ACTA2* (green) and DAPI (blue). We imaged one representative area for a single patient for a,c,d,g (Methods).

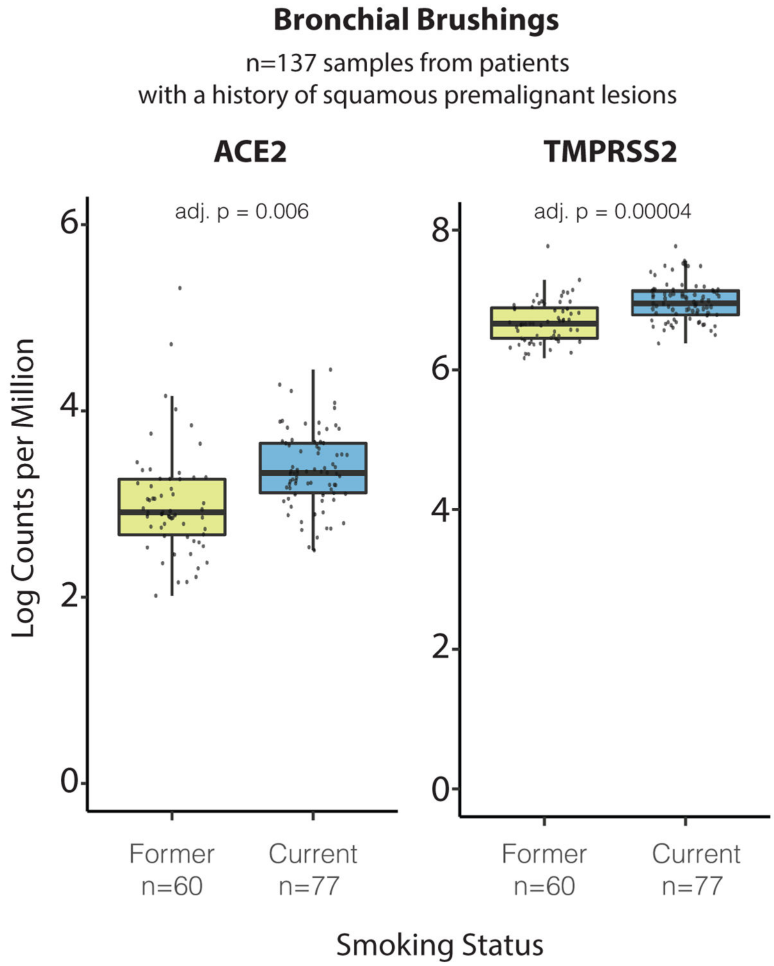
Level 1	Level 2	Level 3
Epithelial	Airway epithelium	Basal Multiciliated Secretory Rare (PNEC, Tuft-like, Ionocyte) proliferat. Epithelial cells KRT5-/KRT17+
	Submucosal gland	Submucosal secretory Acinar
	Alveolar epithelium	AT1 AT2
	Olfactory epithelium	Olfactory epithelium Basaloid
Endothelial	Blood vessels	Arterial Capillary Venous Bronchial Vessel 1 Bronchial Vessel 2 Capillary Intermediate 1 Capillary Intermediate 2
	Lymphatic	Lymphatic EC
Stroma	Fibroblast	Fibroblast Myofibroblast
	Smooth muscle	Airway smooth muscle Venous smooth muscle Fibromyocyte
	Mesothelium	Mesothelium
Immune	Lymphoid	B cell lineage T cell lineage Innate lymphoid
	Myeloid	Dendritic Macrophage Monocyte Mast
	Granulocytes	Basophilic Neutrophilic Eosinophilic
	Megakariocytic	Megakaryocytes Erythrocytes
Cycling cells	Cycling	

Extended Data Fig. 8. An overview of the three-level lung cell ontology used for cell annotation harmonization.



Extended Data Fig. 9. Age, sex, and smoking status associations with expression of *ACE2*, *TMPRSS2*, and *CTSL* across level 3 cell type annotations modeled without interaction terms.

(**a**) Age, sex, and smoking assocations with expression of *ACE2* (blue), *TMPRSS2* (yellow), and *CTSL* (green) modeled without interaction terms on 985,420 cells from 164 donors. Level 3 cell types are shown on the y-axes, and are subdivided by level 1 cell type annotations (top to bottom: epithelial, endothelial, stromal and immune cells). The effect size (x axis) is given as a log fold change (sex, smoking status) or the slope of log expression per year (age). Positive effect sizes indicate increases with age, in males, and in smokers. As the age effect size is given per year, it is not directly comparable to the sex and smoking status effect sizes. Colored bars: associations with an FDR-corrected p-value<0.05 (one-sided Wald test on regression model coefficients), consistent effect direction in pseudo-bulk analysis, and consistent results using the model with interaction terms (Methods). White bars: associations that do not pass all of the three above-mentioned evaluation criteria. Error bars: standard errors around coefficient estimates. Error bars are only shown for colored bars (indications or robust trends) to limit figure size. Only cell types with at least 1000 cells across donors are included. Number of cells and donors per cell type: Basal: 155877, 105, Multiciliated lineage: 37530, 157, Secretory: 22306, 140, Rare: 2676, 71, Submucosal secretory: 33661, 45, AT1: 29973, 101, AT2: 155512, 104, Arterial: 3497, 37, Capillary: 15745, 34, Venous: 7173, 33, Lymphatic EC: 5055, 76, Fibroblasts: 9112, 51, Airway smooth muscle: 1077, 13, B cell lineage: 11761, 90, T cell lineage: 52139, 97, Innate lymphoid cells: 29836, 56, Dendritic cells: 9017, 90, Macrophages: 156964, 89, Monocytes: 42703, 96, Mast cells: 13581 cells, 88 donors. (**b**) Robustness of associations to holding out a dataset. The values show the number of held-out datasets that result in loss of association between a given covariate (rows) and *ACE2, TMPRSS2,* or *CTSL* expression in a given cell type (columns). Robust trends are determined by significant effects that are robust to holding out any dataset (0 values). From left to right: results for *ACE2, TMPRSS2,* and *CTSL.* AT1, 2: alveolar type 1, 2. EC: endothelial cell.



Extended Data Fig. 10. *ACE2* and *TMPRSS2* are up-regulated in bronchial brushings from current versus former smokers.

Boxplots of log counts per million normalized gene expression for *ACE2* and *TMPRSS2* are plotted across current (red, n=70 samples) versus former (green, n=60 samples) smokers. Both genes are significantly up-regulated in current versus former/never (*ACE2*, FDR=0.006; and *TMPRSS2*, FDR=0.00004) based on a linear model using voom-transformed data that included genomic smoking status, batch, and RNA quality (TIN) as covariates and patient as a random effect. Multiple testing correction was performed via Benjamini-Hochberg to obtain an FDR-corrected p-value. (Methods)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Christoph Muus*,^,

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; John A. Paulson School of Engineering and Applied Sciences, Harvard, University, Cambridge, MA 02138

Malte D. Luecken[*,^],
Institute of Computational Biology, Helmholtz Zentrum München, , Neuherberg, Germany

Gokcen Eraslan[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Lisa Sikkema[*],
Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany

Avinash Waghray[*],
Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA; Departments of Internal Medicine and Pediatrics, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA, USA; Harvard Stem Cell Institute, Cambridge, MA, USA

Graham Heimberg[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Yoshihiko Kobayashi[*],
Department of Cell Biology, Duke University Medical School, Durham, NC 27710, USA

Eeshit Dhaval Vaishnav[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02140, USA

Ayshwarya Subramanian[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Christopher Smillie[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Karthik A. Jagadeesh[*],
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Elizabeth Thu Duong[*],
University of California San Diego, Department of Pediatrics, Division of Respiratory Medicine

Evgenij Fiskin[*],

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Elena Torlai Triglia*,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Meshal Ansari*,
Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany; Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

Peiwen Cai*,
Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Brian Lin*,
Center for Regenerative Medicine, Massachusetts General Hospital,Boston, MA, USA; Departments of Internal Medicine and Pediatrics, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA, USA; Harvard Stem Cell Institute, Cambridge, MA, USA

Justin Buchanan*,
Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Sijia Chen*,
Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Jian Shu*,
Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

Adam L. Haber*,
Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Hattie Chung*,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Daniel T. Montoro*,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Taylor S. Adams,
Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

Hananeh Aliee,
Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

Samuel J. Allon,
Institute for Medical Engineering and Science & Department of Chemistry, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

Zaneta Andrusivova,
SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Ilias Angelidis,
Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Orr Ashenberg,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Kevin Bassler,
Department for Genomics & Immunoregulation, LIMES-Institute, University of Bonn, 53115 Bonn, Germany

Christophe Bécavin,
Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France

Inbal Benhar,
Klarman Cell Observatory, Broad Institute of MIT and Harvard,Cambridge, MA, 02142, USA

Joseph Bergenstråhle,
SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Ludvig Bergenstråhle,
SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Liam Bolt,
Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Emelie Braun,
Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute

Linh T. Bui,
Translational Genomics Research Institute, Phoenix, AZ

Steven Callori,
Department of Medicine, Boston University School of Medicine; Bioinformatic Program, Boston University

Mark Chaffin,
Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142

Evgeny Chichelnitskiy,

Institute of Transplant Immunology, Hannover Medical School, MHH, Carl-Neuberg Str. 1, 30625 Hannover, Germany, phone +40 511 532 9745; fax +40 511 532 8090; German Center for Infectious Diseases DZIF, TTU-IICH 07.801

Joshua Chiou,
Biomedical Sciences Graduate Program, University of California-San Diego, La Jolla, CA, 92093.

Thomas M. Conlon,
Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Michael S. Cuoco,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Anna S.E. Cuomo,
European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK

Marie Deprez,
Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France

Grant Duclos,
Boston University School of Medicine, Boston, MA 02118, USA

Denise Fine,
Boston University Medical Center

David S. Fischer,
Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

Shila Ghazanfar,
Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

Astrid Gillich,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Bruno Giotti,
Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Joshua Gould,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Minzhe Guo,

Divisions of Pulmonary Biology; Perinatal Institute, Cincinnati Children's Hospital Medical Center

Austin J. Gutierrez,
Translational Genomics Research Institute, Phoenix, AZ

Arun C. Habermann,
Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

Tyler Harvey,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Peng He,
Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Xiaomeng Hou,
Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Lijuan Hu,
Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute

Yan Hu,
Division of Pulmonary Sciences and Critical Care Medicine, School of Medicine, University of Colorado, Aurora, CO, USA 80045

Alok Jaiswal,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Lu Ji,
Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

Peiyong Jiang,
Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

Theodoro S. Kapellos,
Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, 53115 Bonn, Germany

Christin S. Kuo,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Ludvig Larsson,
SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Michael A. Leney-Greene,

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Kyungtae Lim,
Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

Monika Litviňuková,
Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom.; Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

Leif S. Ludwig,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA Division of Hematology / Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

Soeren Lukassen,
Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany; Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany

Wendy Luo,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Henrike Maatz,
Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

Elo Madissoon,
European Molecular Biology Laboratory - European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK; Wellcome Sanger Institute, Cellular Genetics Programme Wellcome Genome Campus, Hinxton, Cambridge, CB10 1HH, UK

Lira Mamanova,
Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Kasidet Manakongtreecheep,
Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Sylvie Leroy,
Université Côte d'Azur, Pulmonology Department, CHU Nice, NICE, France; Institut de Pharmacologie Moléculaire et Cellulaire, Sophia-Antipolis, France.

Christoph H. Mayr,
Helmholtz Zentrum München, Institute of Lung Biology and Disease, Group Systems Medicine of Chronic Lung Disease, Member of the German Center for Lung Research (DZL), Munich, Germany

Ian M. Mbano,
Africa Health Research Institute,Durban, South Africa. School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of Kwazulu Natal, Durban, South Africa.

Alexi M. McAdams,
Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear, Boston, MA 02114

Ahmad N. Nabhan,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Sarah K. Nyquist,
Computational and Systems Biology, CSAIL, Institute for Medical Engineering and Science & Department of Chemistry, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

Lolita Penland,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Olivier B. Poirion,
Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Sergio Poli,
Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

CanCan Qi,
Dept. of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; GRIAC Research Institute, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Rachel Queen,
Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Bioscience West Building, Newcastle upon Tyne NE1 3 BZ, UK

Daniel Reichart,
Department of Genetics, Harvard Medical School, Boston, MA, United States.; Department of Cardiology, University Heart & Vascular Center, University of Hamburg, Hamburg, Germany

Ivan Rosas,
Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

Jonas C. Schupp,
Section of Pulmonary, Critical Care, and Sleep Medicine, Yale University School of Medicine, New Haven, CT, USA

Conor V. Shea,
Boston University School of Medicine, Boston, MA 02118, USA

Xingyi Shi,
Department of Medicine, Boston University School of Medicine; Bioinformatic Program, Boston University

Rahul Sinha,
Institute for Stem Cell Biology and Regenerative Medicine, Stanford Medicine, Stanford, CA 94305, USA

Rene V. Sit,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Kamil Slowikowski,
Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Michal Slyper,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Neal P. Smith,
Massachusetts General Hospital Center for Immunology and Inflammatory Diseases

Alex Sountoulidis,
Stockholm University, Department of Molecular Biosciences, The Wenner-Gren Institute.

Maximilian Strunz,
Comprehensive Pneumology Center (CPC) and Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Travis B. Sullivan,
Lahey Hospital & Medical Center

Dawei Sun,
Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

Carlos Talavera-López,
Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom.

Peng Tan,

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Jessica Tantivit,
Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Kyle J. Travaglini,
Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Nathan R. Tucker,
Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142; Masonic Medical Research Institute, Utica, NY, USA 13501

Katherine A. Vernon,
Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA

Marc H. Wadsworth,
Institute for Medical Engineering and Science, Department of Chemistry & Koch Institute for Integrative Cancer Research, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

Julia Waldman,
Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Xiuting Wang,
Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Ke Xu,
Boston University School of Medicine, Boston, MA 02118, USA

Wenjun Yan,
Center for Brain Science, Harvard University, Cambridge, MA 02138; Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

William Zhao,
Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Carly G.K. Ziegler[^],
Harvard-MIT Health Sciences and Technology, Institute for Medical Engineering and Science, Koch Institute for Integrative Cancer Research, MIT; Broad Institute of MIT and Harvard; Ragon Institute of MGH, MIT and Harvard

The NHLBI LungMAP Consortium, The Human Cell Atlas Lung Biological Network

Affiliations

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; John A. Paulson School of Engineering and Applied Sciences, Harvard, University, Cambridge, MA 02138

Institute of Computational Biology, Helmholtz Zentrum München, , Neuherberg, Germany

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany

Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA; Departments of Internal Medicine and Pediatrics, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA, USA; Harvard Stem Cell Institute, Cambridge, MA, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Department of Cell Biology, Duke University Medical School, Durham, NC 27710, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02140, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

University of California San Diego, Department of Pediatrics, Division of Respiratory Medicine

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany; Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Center for Regenerative Medicine, Massachusetts General Hospital,Boston, MA, USA; Departments of Internal Medicine and Pediatrics, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA, USA; Harvard Stem Cell Institute, Cambridge, MA, USA

Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

Institute for Medical Engineering and Science & Department of Chemistry, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Department for Genomics & Immunoregulation, LIMES-Institute, University of Bonn, 53115 Bonn, Germany

Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France

Klarman Cell Observatory, Broad Institute of MIT and Harvard,Cambridge, MA, 02142, USA

SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute

Translational Genomics Research Institute, Phoenix, AZ

Department of Medicine, Boston University School of Medicine; Bioinformatic Program, Boston University

Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142

Institute of Transplant Immunology, Hannover Medical School, MHH, Carl-Neuberg Str. 1, 30625 Hannover, Germany, phone +40 511 532 9745; fax +40 511 532 8090; German Center for Infectious Diseases DZIF, TTU-IICH 07.801

Biomedical Sciences Graduate Program, University of California-San Diego, La Jolla, CA, 92093.

Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK

Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France

Boston University School of Medicine, Boston, MA 02118, USA

Boston University Medical Center

Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Divisions of Pulmonary Biology; Perinatal Institute, Cincinnati Children's Hospital Medical Center

Translational Genomics Research Institute, Phoenix, AZ

Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute

Division of Pulmonary Sciences and Critical Care Medicine, School of Medicine, University of Colorado, Aurora, CO, USA 80045

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, 53115 Bonn, Germany

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom.; Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA Division of Hematology / Oncology, Boston Children’s Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany; Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

European Molecular Biology Laboratory - European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK; Wellcome Sanger Institute, Cellular Genetics Programme Wellcome Genome Campus, Hinxton, Cambridge, CB10 1HH, UK

Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Université Côte d'Azur, Pulmonology Department, CHU Nice, NICE, France; Institut de Pharmacologie Moléculaire et Cellulaire, Sophia-Antipolis, France.

Helmholtz Zentrum München, Institute of Lung Biology and Disease, Group Systems Medicine of Chronic Lung Disease, Member of the German Center for Lung Research (DZL), Munich, Germany

Africa Health Research Institute,Durban, South Africa. School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of Kwazulu Natal, Durban, South Africa.

Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear, Boston, MA 02114

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Computational and Systems Biology, CSAIL, Institute for Medical Engineering and Science & Department of Chemistry, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093. Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

Dept. of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; GRIAC Research Institute, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Bioscience West Building, Newcastle upon Tyne NE1 3 BZ, UK

Department of Genetics, Harvard Medical School, Boston, MA, United States.; Department of Cardiology, University Heart & Vascular Center, University of Hamburg, Hamburg, Germany

Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine

Section of Pulmonary, Critical Care, and Sleep Medicine, Yale University School of Medicine, New Haven, CT, USA

Boston University School of Medicine, Boston, MA 02118, USA

Department of Medicine, Boston University School of Medicine; Bioinformatic Program, Boston University

Institute for Stem Cell Biology and Regenerative Medicine, Stanford Medicine, Stanford, CA 94305, USA

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Massachusetts General Hospital Center for Immunology and Inflammatory Diseases

Stockholm University, Department of Molecular Biosciences, The Wenner-Gren Institute.

Comprehensive Pneumology Center (CPC) and Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

Lahey Hospital & Medical Center

Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom.

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Broad Institute of MIT and Harvard, Cambridge, MA, USA; Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

Department of Biochemistry and Wall Center for Pulmonary Vascular Disease

Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142; Masonic Medical Research Institute, Utica, NY, USA 13501

Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA

Institute for Medical Engineering and Science, Department of Chemistry & Koch Institute for Integrative Cancer Research, MIT; Ragon Institute of MGH, MIT and Harvard; Broad Institute of MIT and Harvard

Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Boston University School of Medicine, Boston, MA 02118, USA

Center for Brain Science, Harvard University, Cambridge, MA 02138; Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

Department of Genetics and Genomic Sciences, Icahn School of Medicineat Mount Sinai, New York, NY 10029, USA

Harvard-MIT Health Sciences and Technology, Institute for Medical Engineering and Science, Koch Institute for Integrative Cancer Research, MIT; Broad Institute of MIT and Harvard; Ragon Institute of MGH, MIT and Harvard

Acknowledgements

We thank all donors, patients and their families for their contributions to the studies that are part of our integrated analysis. We thank Leslie Gaffney and Anna Hupalowska for help with figure preparation, Carl de Boer for critical reading of the manuscript, and Dr. Elmar Spiegel from the statistical consulting core facility at the Institute of Computational Biology, Helmholtz Center Munich for advice on statistical modeling.

N.E.B. is supported by NIH/NHLBI R01HL145372, Department of Defense W81XWH1910416.

Joseph Collin is supported by grants MRC (#MR/S035826/1) and ERC (#614620). Roland Eils and Christian Cond are supported by the European Commission - ESPACE, 874710 and Horizon 2020, Tushar Desai is supported by HubMap consortium and Stanford Child Health Research Institute- Woods Family Faculty Scholarship. Oliver Eickelberg is supported by CZI Seed network and NIH 1R01HL146519 (OE). Christine Falk is supported by DFG, SFB 738 project B3; DFG FA-483/1-1. Ian A. Glass and the University of Washington Laboratory of Developmental Biology was supported by NIH award number 5R24HD000836 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Anna Greka is supported by Seed Network Grant from the Chan Zuckerberg Initiative. PH acknowledges support from the LENDULET-BIOMAG Grant (2018-342) and the Chan Zuckerberg Initiative (CZF2019-002448). Norbert Hubner acknowledges support by BHF/DZHK grant, ERC Advanced Grant under the European Union Horizon 2020 Research and Innovation Program and the Federal Ministry of Education and Research of Germany in the framework of CaRNAtion. William Janssen received funding from the NIH (R35HL140039, R01HL130938). Naftali Kaminski received funding from NIH grants R01HL127349, U01HL145567 and an unrestricted grant from Three Lakes Foundation. Melanie Koenigshoff received funding from National Institute of Health Grant R01HL141380. Gerard Koppelman and Malte Kuhnemund received funding from the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Mark Krasnow received funding from Howard Hughes Medical Institute, Chan Zuckerberg Initiative, Wall Center for Pulmonary Vascular Disease. Jonathan Kropski received funding from NIH R01HL145372(JAK/NEB), K08HL130595(JAK), Doris Duke Charitable Foundation (JAK). Majlinda Lako received funding from ERC (#614620). Haeock Lee acknowledges funding from the National Research Foundation of Korea. Sarah Mazzilli, Joshua Campbell, Avrum Spira, Marc E. Lenburg, and Jennifer Beane acknowledge support from a Stand Up to Cancer-LUNGevity-American Lung Association Lung Cancer Interception Dream

Team Translational Cancer Research Grant (grant number: SU2C-AACR-DT23-17 to S.M. Dubinett and A.E. Spira). Stand Up to Cancer is a division of the Entertainment Industry Foundation. Sarah Mazzilli, Joshua Campbell, Marc E. Lenburg and Jennifer Beane acknowledge funding from Sponsored Research Agreements with Janssen Pharmaceuticals, Inc. Jennifer Beane and Joshua Campbell acknowledge funding from Department of Defense W81XWH1410234. Sylvie Leroy acknowledges funding from the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Sten Linnarson acknowledges funding from Knut and Alice Wallenberg Foundation (2015.0041, 2018.0172), Erling-Persson Family Foundation (HDCA) and Swedish Foundation for Strategic Research (SB16-0065, RIF14-0057). Joakim Lundeberg acknowledges funding from the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR), Knut and Alice Wallenberg Foundation (2018.0172) and Erling-Persson Family Foundation (HDCA). B.D.M. is supported by National Institutes of Health grant R01 HL133153. Kerstin Meyer acknowledges funding from Chan Zuckerberg Initiative grant 2017-174169 (5022), Wellcome Trust grants 206194/Z/17/Z and 211276/Z/18/Z, Medical Research Council grant MR/S035907/1, the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Alexander Misharin acknowledges funding from NIH grants HL135124, AG049665 and AI135964, and grant number CZF2019-002438 from the Chan Zuckerberg Initiative Foundation awarded to the HCA Lung Seed Network. Martijn Nawijn acknowledges funding from grant number CZF2019-002438 from the Chan Zuckerberg Initiative Foundation awarded to the HCA Lung Seed Network, GSK Ltd, Netherlands Lung Foundation project no. 5.1.14.020 and 4.1.18.226 and the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Marko Z. Nikolić acknowledges funding from Rutherford Fund Fellowship allocated by the Medical Research Council and the UK Regenerative Medicine Platform (MR/5005579/1); Rosetrees Trust (Grant number M899 to Marko Z Nikolic). Michela Noseda acknowledges funding from a BHF/DZHK grant and British Heart Foundation (PG/16/47/32156), Chan Zuckerberg Initiative RFA CZF2019-002431e for Research Excellence and Centre for Regenerative Medicine, Imperial College London, London, UK. Jose Ordovas Montanes acknowledges funding from Richard and Susan Smith Family Foundation. Gavin Y. Oudit acknowledges support from Canada Research Chair (CRC), Canadian Institute of Health Research (CIHR) and the Heart and Stroke Foundation (HSF). Dana Pe'er acknowledges funding from Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center. Stephen R Quake acknowledges funding from the CZI Biohub. Jayaraj Rajagopal acknowledges funding from LungMAP, CZI Seed Network. Purushothama Rao Tata acknowledges funding from R01HL146557 from NHLBI/NIH and CZI-HCA seed projects . Emma L Rawlins acknowledges funding from MRC: MR/S035907/1 and MR/P009581/1. Wellcome:109146/Z/15/Z. Core support from the Wellcome Trust: 203144/Z/16/Z and Cancer Research UK: C6946/A24843. AR and O.R.-R. work was supported by HHMI, the Klarman Cell Observatory, the Manton Foundation and the Chan Zuckerberg Initiative. Paul Reyfman acknowledges funding from NIH K08HL146943; Parker B. Francis Fellowship, ATS Foundation/Boehringer Ingelheim Pharmaceuticals Inc. Research Fellowship in IPF. Mauricio Rojas acknowledges funding from 1 U01 HL14555-01. Kourosh Saeb-Parsy acknowledges funding from NIHR Cambridge Biomedical Research Centre. Christos Samakovlis acknowledges funding from the Swedish research Council, Swedish Cancer Society, CPI, the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Herbert B. Schiller's work was supported by grant number CZF2019-002438 from the Chan Zuckerberg Initiative Foundation awarded to the HCA Lung Seed Network, the German Center for Lung Research and Helmholtz Association, the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR). Joachim Schultze's work was supported by JLS funded in part by Boehringer Ingelheim, by the German Research Foundation (DFG; EXC2151/1, ImmunoSensation2 - the immune sensory system, project number 390873048), project numbers 329123747, 347286815) and by the HGF grant sparse2big. Christine Seidman was supported by Howard Hughes Medical Institute, NIH (NHLBI): 2R01HL080494. Jon Seidman was supported by NIH (NHLBI): 2R01HL080494. Alex K. Shalek was supported by the Beckman Young Investigator Program, a Sloan Fellowship in Chemistry, the NIH (5U24AI118672), and the Bill and Melinda Gates Foundation. Douglas Shepherd was supported by the CZI Seed Network grant. JRS is supported by the National Heart, Lung, and Blood Institute (NHLBI - R01HL119215), by the NIAID Novel Alternative Model Systems for Enteric Diseases (NAMSED) consortium (U19AI116482) and by grant number CZF2019-002440 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. Fabian J. Theis was supported by grant number CZF2019-002438 from the Chan Zuckerberg Initiative Foundation awarded to the HCA Lung Seed Network, the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant # ZT-I-PF-5-01), the European Union's H2020 Research and Innovation Program under grant agreement no. 874656 (discovAIR) and the German Center for Lung Research. Alexander Tsankov was supported by CZI Lung Atlas and NSF award IOS-2028295. Ludovic Vallier was supported by the ERC advanced grant New-Chol, the Cambridge University Hospitals National Institute for Health Research Biomedical Research Centre and the core support grant from the Wellcome Trust and Medical Research Council of the Wellcome–Medical Research Council Cambridge Stem Cell Institute. Maarten van den Berge was supported by the ministry of Economic Affairs and Climate Policy by means of the PPP. Ramnik J Xavier was supported by DK 043351, DK114784, AI142784, DK117263. Laure Emmanuelle Zaragosi was supported by the Agence Nationale de la Recherche (UCAJEDI, ANR-15-IDEX-01; SAHARRA, ANR-19-CE14-0027; France Génomique, ANR-10-INBS-09-03); Fondation pour la Recherche Médicale (DEQ20180339158); Chan Zuckerberg Initiative (Silicon Valley Foundation, 2017-175159-5022); Conseil Départemental des Alpes Maritimes (2016-294DGADSH-CV; 2019-390DGADSH-CV). Darin Zerti was supported by MRC (#MR/S035826/1) and ERC (#614620). H.Z. is supported by the National Key R&D Program (no. 2019YFA0801703) and National Natural Science Foundation of China (no. 31871370). This study was supported by NHLBI Molecular Atlas of

Lung Development Program Human Tissue Core grant U01HL122700 and HL148861. Jeffrey Whitsett, Gail H. Deutsch and Yan Xu acknowledge support from National Institutes of Health, U01 HL148856 LungMap Phase II – Building a multidimensional map of developing human lung. Xin Sun, Allen Wang, Sebastian Preissl, Thomas J. Mariani

Code Availability Statement

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 (requires Google account). The analysis can also be accessed at https://github.com/theislab/Covid_meta_analysis/.

Data Availability Statement

Availability of published datasets is summarized in Supplementary Table 1 and 2. Interactive visualization and download of select (as indicated in Supplementary Table 1 and 2) gene expression data can be accessed on the Single Cell Portal at http://broad.io/hcacovid19

NHLBI LungMap Consortium author

Gail H. Deutsch	Department of Pathology, Seattle Children's Hospital, University of Washington, Seattle, Washington
Jennifer Dutra	University of Rochester Biocomputational Center, Research Data Integration & Analytics Group, University of Rochester Medical Center, Rochester, New York; Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, New York;
Kyle J Gaulton	Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.
Jeanne Holden-Wiltse	University of Rochester Biocomputational Center, Research Data Integration & Analytics Group, University of Rochester Medical Center, Rochester, New York; Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, New York;
Heidie L. Huyck	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York
Thomas J. Mariani	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York Program in Pediatric Molecular and Personalized Medicine, Department of Pediatrics, University of Rochester Medical Center, Rochester, New York
Ravi S. Misra	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York
Cory Poole	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York;
Sebastian Preissl	Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093.Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.
Gloria S. Pryhuber	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York
Lisa Rogers	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York;
Xin Sun	Department of Pediatrics, University of California-San Diego School of Medicine, La Jolla, CA 92093. Department of Biological Sciences, University of California-San Diego, La Jolla, CA 92093.
Allen Wang	Center for Epigenomics, University of California-San Diego School of Medicine, La Jolla, CA, 92093.

	Department of Cellular and Molecular Medicine, University of California-San Diego School of Medicine, La Jolla, CA, 92093.
Jeffrey A Whitsett	Cincinnati Children’s Hospital Medical Center, Cincinnati, OHIO
Yan Xu	Divisions of Pulmonary Biology and Biomedical Informatics; Perinatal Institute, Cincinnati Children's Hospital Medical Center; University of Cincinnati College of Medicine

HCA Lung Biological Network author

Jehan Alladina	Division of Pulmonary and Critical Care Medicine, Department of Medicine, Massachusetts General Hospital, Boston, USA
Nicholas E Banovich	Translational Genomics Research Institute, Phoenix, AZ.
Pascal Barbry	Université Côte d’Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France
Jennifer E. Beane	Department of Medicine, Boston University School of Medicine, Boston, MA, USA
Roby P. Bhattacharyya	Infectious Disease and Microbiome Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA Infectious Diseases Division, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
Katharine E. Black	Division of Pulmonary and Critical Care Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
Alvis Brazma	European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
Joshua D. Campbell	Division of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA.
Josalyn L. Cho	Department of Medicine, Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; Center for Immunology and Inflammatory Diseases, Division of Rheumatology, Allergy and Immunology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA;
Joseph Collin	Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Bioscience West Building, Newcastle upon Tyne NE1 3 BZ, UK
Christian Conrad	Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany
Kitty de Jong	Roswell Park Comprehensive Cancer Center, Buffalo, NY 14203, USA
Tushar Desai	Department of Medicine and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94116
Diane Z. Ding	Boston University School of Medicine, Boston, MA 02118, USA
Oliver Eickelberg	Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO, US
Roland Eils	Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany Health Data Science Unit, Heidelberg University Hospital and BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany
Patrick T. Ellinor	Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA 02114
Alen Faiz	Respiratory Bioinformatics and Molecular Biology, University of Technology Sydney, Sydney, New South Wales, Australia.
Christine S. Falk	Institute of Transplant Immunology, Hannover Medical School, MHH, Germany

Michael Farzan	Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter, Florida, USA (33458)
Andrew Gellman	Department of Statistics, Columbia University
Gad Getz	Broad Institute of MIT and Harvard, Cambridge, MA, USA Department of Pathology, Harvard Medical School, Boston, MA, USA Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
Ian A Glass	Department of Pediatrics, Genetic Medicine, University of Washington, Seattle, Washington;
Anna Greka	Brigham and Women's Hospital, Harvard Medical School, and Broad Institute of MIT and Harvard
Muzlifah Haniffa	Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK; Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK; Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK.
Lida P Hariri	Division of Pulmonary and Critical Care Medicine and Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA
Mark W. Hennon	Roswell Park Comprehensive Cancer Center, Buffalo, NY 14203, USA
Peter Horvath	Synthetic and Systems Biology Unit, Hungarian Academy of Sciences, Biological Research Center (BRC), Temesvári körút 62, 6726 Szeged, Hungary. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Tukholmankatu 8, 00014 Helsinki, Finland.
Norbert Hübner	Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany; DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, 13347 Berlin, Germany; Berlin Institute of Health (BIH), 10178 Berlin, Germany; Charité-Universitätsmedizin, 10117 Berlin, Germany
Deborah T. Hung	Professor of Genetics, Department of Genetics at Harvard Medical School and Department of Molecular Biology at Massachusetts General Hospital; Co-Director, Infectious Disease and Microbiome Program and Core Faculty Member, Broad Institute of MIT & Harvard
Heidie L. Huyck	Department of Pediatrics, Division of Neonatology, University of Rochester Medical Center, Rochester, New York
William J. Janssen	Division of Pulmonary, Critical Care and Sleep Medicine National Jewish Health Division of Pulmonary Medicine and Critical Care Sciences University of Colorado Denver
Dejan Juric	Department of Medicine, Harvard Medical School and Massachusetts General Hospital Cancer Center, Boston, MA 02114
Naftali Kaminski	Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine
Melanie Koenigshoff	Division of Pulmonary Sciences and Critical Care Medicine, School of Medicine, University of Colorado, Aurora, CO, USA 80045 Lung Repair and Regeneration Unit, Helmholtz-Zentrum Munich, Ludwig-Maximilians-University, University Hospital Grosshadern, Member of the German Center of Lung Research (DZL), Munich, Germany 81377
Gerard H. Koppelman	Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children’s Hospital, University of Groningen, University Medical Center Groningen (UMCG), Groningen Research Institute for Asthma and COPD, Groningen, Netherlands
Mark A. Krasnow	Department of Biochemistry and Wall Center for Pulmonary Vascular Disease
Jonathan A Kropski	Division of Allergy, Pulmonary and Critical Care Medicine,Department of Medicine, Vanderbilt University Medical Center, Nashville, TN Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN Department of Veterans Affairs Medical Center, Nashville, TN
Malte Kuhnemund	Cartana AB, Nobels vag 16, 17165 Stockholm, Sweden
Robert Lafyatis	Division of Rheumatology, Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, USA.

Majlinda Lako	Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Bioscience West Building, Newcastle upon Tyne NE1 3 BZ, UK
Eric S. Lander	Broad Institute of Harvard and MIT, Cambridge, MA, USA Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA Department of Systems Biology, Harvard Medical School, Boston, MA
Haeock Lee	Department of Biomedicine and Health Sciences, The Catholic University of Korea, Seoul, Korea
Marc E Lenburg	Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine
Sten Linnarsson	Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute
Gang Liu	Boston University School of Medicine, Boston, MA 02118, USA
Yuk Ming Dennis Lo	Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China
Joakim Lundeberg	SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology
John C. Marioni	Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, UK Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK
Charles-Hugo Marquette	Université Côte d'Azur, CHU de Nice, FHU OncoAge, CNRS, Inserm, IRCAN team 3, Pulmonology Department, Nice, 06000, France
Sarah A. Mazzilli	Boston University School of Medicine, Boston, MA 02118, USA
Benjamin D. Medoff	Division of Pulmonary and Critical Care Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
Ross J. Metzger	Department of Biochemistry and Wall Center for Pulmonary Vascular Disease
Kerstin B Meyer	Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
Zhichao Miao	Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
Alexander V Misharin	Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, Illinois
Martijn C Nawijn	Department of Pathology and Medical Biology, University of Groningen, GRIAC Research institue, University Medical Center Groningen, the Netherlands
Marko Z Nikolić	UCL Respiratory, Division of Medicine, University College London, London, UK.
Michela Noseda	National Heart and Lung Institute, Imperial College London, London, UK; British Heart Foundation Centre for Research Excellence and Centre for Regenerative Medicine, Imperial College London, London, UK
Jose Ordovas-Montanes	Division of Gastroenterology Boston Children's Hospital, Boston, MA, USA; Program in Immunology, Harvard Medical School, Boston, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA; Harvard Stem Cell Institute, Cambridge, MA, USA.
Gavin Y. Oudit	Division of Cardiology, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada Mazankowski Alberta Heart Institute, Edmonton, Alberta, Canada
Dana Pe'er	Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA
Joseph E Powell	Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia; UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia
Stephen R Quake	Depts of Bioengineering and Applied Physics, Stanford University, and the Chan Zuckerberg Biohub.
Jayaraj Rajagopal	Harvard Stem Cell Institute, Cambridge, Massachusetts; Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts
Purushothama Rao Tata	Department of Cell Biology, Regeneration Next Initiative, Duke University School of Medicine, Durham, NC, USA, 27710

Emma L. Rawlins	Wellcome Trust/ CRUK Gurdon Institute and Department Physiology, Development and Neuroscience, University of Cambridge
Aviv Regev	Klarman Cell Observatory, Broad Institute of MIT and Harvard, Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02142
Mary E. Reid	Roswell Park Comprehensive Cancer Center, Buffalo, NY 14203, USA
Paul A. Reyfman	Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, Illinois
Kimberly M. Rieger-Christ	Lahey Hospital & Medical Center, Burlington, MA 01805
Mauricio Rojas	Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh
Orit Rozenblatt-Rosen	Orit Rozenblatt-Rosen, Affiliation: Klarman Cell Observatory, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
Kourosh Saeb-Parsy	Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, UK
Christos Samakovlis	SciLifeLab, Department of Molecular Biosciences, Stockholm University, Stockholm Sweden and Cardiopulmonary Institute, Justus Liebig University; Giessen Germany
Joshua R. Sanes	Center for Brain Science, Harvard University, Cambridge, MA 02138 Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138
Herbert Schiller	Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany
Joachim L Schultze	Department for Genomics & Immunoregulation, LIMES-Institute, University of Bonn, 53115 Bonn, Germany PRECISE Platform for Single Cell Genomics & Epigenomics, Germany Center for Neurodegenerative Diseases and University of Bonn, Bonn, Germany
Roland F. Schwarz	Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany
Ayellet V. Segre	Broad Institute of MIT and Harvard, Cambridge, MA, USA. Harvard Medical School, Boston, MA, USA. Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, USA.
Max A. Seibold	Department of Pediatrics; Center for Genes, Environment, and Health; National Jewish Health; Denver, CO 80206
Christine E. Seidman	Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; Cardiovascular Division, Brigham & Women’s Hospital, Boston, MA 02115, USA; Howard Hughes Medical Institute
Jon G. Seidman	Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
Alex K. Shalek	Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA; Institute for Medical Engineering and Science (IMES), Koch Institute for Integrative Cancer Research, and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA
Douglas P Shepherd	Center for Biological Physics and Department of Physics, Arizona State University, Tempe, AZ USA
Rahul Sinha	Institute for Stem Cell Biology and Regenerative Medicine, Stanford Medicine, Stanford, CA 94305, USA
Jason R. Spence	Department of Internal Medicine, Gastroenterology, University of Michigan Medical School, Ann Arbor, MI 48109, USA; Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA; Department of Biomedical Engineering, University of Michigan College of Engineering, Ann Arbor, MI 48109, USA.
Avrum Spira	Boston University School of Medicine, Boston, MA 02118, USA & Johnson and Johnson Innovation, Cambridge, MA 02142, USA.
Xin Sun	Department of Pediatrics, Department of Biological Sciences, University of California SD, 9500 Gilman Dr. MC0766, San Diego, CA 92093-0766

Erik Sundström	Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
Sarah A. Teichmann	Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom. Dept Physics/Cavendish Laboratory, University of Cambridge, JJ Thompson Ave, Cambridge CB3 0EH, United Kingdom.
Fabian J. Theis	Institute of Computational Biology, Helmholtz Zentrum München and Departments of Mathematics and Life Sciences, Technical University Munich, Germany
Alexander M. Tsankov	Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA
Ludovic Vallier	Wellcome and MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Biomedical Campus, Puddicombe Way, Cambridge CB2 0AW, UK; Department of Surgery, Cambridge Biomedical Campus, Hills Rd, Cambridge, CB2 0QQ, UK
Maarten van den Berge	Department of Pulmonary diseases and tuberculosis, University of Groningen, GRIAC Research institue, University Medical Center Groningen, the Netherlands
Tave A. Van Zyl	Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear, Boston, MA 02114
Alexandra-Chloé Villani	Broad Institute of MIT and Harvard, Cambridge, MA, USA Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA
Astrid Weins	Department of Pathology, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA 02115, USA.
Ramnik J Xavier	Broad Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital
Ali Önder Yildirim	Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany
Laure Emmanuelle Zaragosi	Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, 06560, France
Darin Zerti	Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Bioscience West Building, Newcastle upon Tyne NE1 3 BZ, UK; Microscopy Centre and Department of Applied Clinical Sciences and Biotechnology, University of L'Aquila, via Vetoio, 67100 Coppito, L'Aquila, Italy
Hongbo Zhang	Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education and Department of Histology and Embryology of Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China
Kun Zhang	UCSD Department of Bioengineering, 9500 Gilman Drive, MC0412, PFBH402, La Jolla, CA 92093-0412
Xiaohui Zhang	Boston University School of Medicine, Boston, MA 02118, USA

REFERENCES

1. Wang D et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA 323, 1061–1069 (2020). [PubMed: 32031570]
2. Huang C et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506 (2020). [PubMed: 31986264]
3. Chen N et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513 (2020). [PubMed: 32007143]
4. Wang W et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA (2020) doi:10.1001/jama.2020.3786.
5. Jia HP et al. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. J. Virol 79, 14614–14621 (2005). [PubMed: 16282461]

6. Hou YJ et al. SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. Cell 182, 429–446.e14 (2020). [PubMed: 32526206]
7. McCray PB Jr et al. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. J. Virol 81, 813–821 (2007). [PubMed: 17079315]
8. Walls AC et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281–292.e6 (2020). [PubMed: 32155444]
9. Perez-Saez J et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. Lancet Infect. Dis (2020) doi:10.1016/S1473-3099(20)30584-3.
10. Zhou F et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395, 1054–1062 (2020). [PubMed: 32171076]
11. Ludvigsson JF Systematic review of COVID- 19 in children shows milder cases and a better prognosis than adults. Acta Paediatr. 109, 1088–1095 (2020). [PubMed: 32202343]
12. Guo FR Smoking links to the severity of COVID- 19: An update of a meta- analysis. J. Med. Virol 1 (2020).
13. Sungnak W et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med 26, 681–687 (2020). [PubMed: 32327758]
14. Ziegler CGK et al. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell 181, 1016–1035.e19 (2020). [PubMed: 32413319]
15. Qi F, Qian S, Zhang S & Zhang Z Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. Biochem. Biophys. Res. Commun 526, 135–140 (2020). [PubMed: 32199615]
16. Lukassen S et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. (2020) doi:10.15252/embj.20105114.
17. Zhang H et al. Specific ACE2 expression in small intestinal enterocytes may cause gastrointestinal symptoms and injury after 2019-nCoV infection. Int. J. Infect. Dis 96, 19–24 (2020). [PubMed: 32311451]
18. Sos BC et al. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. Genome Biol. 17, 20 (2016). [PubMed: 26846207]
19. Emery B et al. Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination. Cell 138, 172–185 (2009). [PubMed: 19596243]
20. Vento-Tormo R et al. Single-cell reconstruction of the early maternal-fetal interface in humans. Nature 563, 347–353 (2018). [PubMed: 30429548]
21. Suryawanshi H et al. A single-cell survey of the human first-trimester placenta and decidua. Sci Adv 4, eaau4788 (2018). [PubMed: 30402542]
22. Tsang JCH et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. Proc. Natl. Acad. Sci. U. S. A 114, E7786–E7795 (2017). [PubMed: 28830992]
23. Pérez-Silva JG, Español Y, Velasco G & Quesada V The Degradome database: expanding roles of mammalian proteases in life and disease. Nucleic Acids Res. 44, D351–5 (2016). [PubMed: 26553809]
24. Coutard B et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 176, 104742 (2020). [PubMed: 32057769]
25. Millet JK & Whittaker GR Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. Virology 517, 3–8 (2018). [PubMed: 29275820]
26. Seidah NG & Prat A The biology and therapeutic targeting of the proprotein convertases. Nat. Rev. Drug Discov 11, 367–383 (2012). [PubMed: 22679642]
27. Cai H Sex difference and smoking predisposition in patients with COVID-19. The Lancet Respiratory Medicine vol. 8 e20 (2020). [PubMed: 32171067]
28. Goldfarbmuren KC et al. Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. Nat. Commun 11, 2485 (2020). [PubMed: 32427931]

29. Duclos GE et al. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. Sci Adv 5, eaaw3413 (2019). [PubMed: 31844660]
30. Vieira Braga FA et al. A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med 25, 1153–1163 (2019). [PubMed: 31209336]
31. Reyfman PA et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med 199, 1517–1536 (2019). [PubMed: 30554520]
32. Madissoon E et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 21, 1 (2019). [PubMed: 31892341]
33. Ordovas-Montanes J et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. Nature 560, 649–654 (2018). [PubMed: 30135581]
34. Miller AJ et al. In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. Dev. Cell 53, 117–128.e6 (2020). [PubMed: 32109386]
35. Adams TS et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci Adv 6, eaba1983 (2020). [PubMed: 32832599]
36. Habermann AC et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. Science Advances vol. 6 eaba1972 (2020). [PubMed: 32832598]
37. Deprez M et al. A single-cell atlas of the human healthy airways. Am. J. Respir. Crit. Care Med (2020) doi:10.1164/rccm.201911-2199OC.
38. Morse C et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur. Respir. J 54, (2019).
39. Travaglini KJ, Nabhan AN, Penland L & Sinha R A molecular cell atlas of the human lung from single cell RNA sequencing. BioRxiv (2020).
40. Mayr CH et al. Integrated Single Cell Analysis of Human Lung Fibrosis Resolves Cellular Origins of Predictive Protein Signatures in Body Fluids. (2020) doi:10.2139/ssrn.3538700.
41. Beane JE et al. Molecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. Nat. Commun 10, 1856 (2019). [PubMed: 31015447]
42. Chan C-M et al. Carcinoembryonic Antigen-Related Cell Adhesion Molecule 5 Is an Important Surface Attachment Factor That Facilitates Entry of Middle East Respiratory Syndrome Coronavirus. Journal of Virology vol. 90 9114–9127 (2016). [PubMed: 27489282]
43. Wahl SM et al. Secretory leukocyte protease inhibitor (SLPI) in mucosal fluids inhibits HIV-1. Oral Dis. 3, S64–S69 (1997). [PubMed: 9456660]
44. Turula H & Wobus C The Role of the Polymeric Immunoglobulin Receptor and Secretory Immunoglobulins during Mucosal Infection and Immunity. Viruses vol. 10 237 (2018).
45. Burkhardt AM et al. CXCL17 is a mucosal chemokine elevated in idiopathic pulmonary fibrosis that exhibits broad antimicrobial activity. J. Immunol 188, 6399–6406 (2012). [PubMed: 22611239]
46. Debbabi H et al. Primary type II alveolar epithelial cells present microbial antigens to antigen-specific CD4+ T cells. Am. J. Physiol. Lung Cell. Mol. Physiol 289, L274–9 (2005). [PubMed: 15833765]
47. Yue Y et al. SARS-Coronavirus Open Reading Frame-3a drives multimodal necrotic cell death. Cell Death Dis. 9, 904 (2018). [PubMed: 30185776]
48. Burkard C et al. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog. 10, e1004502 (2014). [PubMed: 25375324]
49. Wishart DS et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018). [PubMed: 29126136]
50. Gordon DE et al. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. bioRxiv 2020.03.22.002386 (2020) doi:10.1101/2020.03.22.002386.
51. Luan HH et al. GDF15 Is an Inflammation-Induced Central Mediator of Tissue Tolerance. Cell 178, 1231–1244.e11 (2019). [PubMed: 31402172]

52. Dhar P & McAuley J The Role of the Cell Surface Mucin MUC1 as a Barrier to Infection and Regulator of Inflammation. Front. Cell. Infect. Microbiol 9, 117 (2019). [PubMed: 31069176]
53. Efremova M, Vento-Tormo M, Teichmann SA & Vento-Tormo R CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc 15, 1484–1506 (2020). [PubMed: 32103204]
54. Bao L et al. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. Nature 583, 830–833 (2020). [PubMed: 32380511]
55. Montoro DT et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature 560, 319–324 (2018). [PubMed: 30069044]
56. Smith JC et al. Cigarette smoke exposure and inflammatory signaling increase the expression of the SARS-CoV-2 receptor ACE2 in the respiratory tract. doi:10.1101/2020.03.28.013672.
57. Tucker Nathan R et al. Myocyte-Specific Upregulation of ACE2 in Cardiovascular Disease. Circulation 142, 708–710 (2020). [PubMed: 32795091]
58. Hamming I et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J. Pathol 203, 631–637 (2004). [PubMed: 15141377]
59. Zhao Y et al. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv 2020.01.26.919985 (2020) doi:10.1101/2020.01.26.919985.
60. Venkatakrishnan AJ et al. Knowledge synthesis from 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. arXiv [q-bio.GN] (2020).
61. Mao L et al. Neurological manifestations of hospitalized patients with COVID-19 in Wuhan, China: a retrospective case series study. (2020).
62. Poyiadji N et al. COVID-19--associated acute hemorrhagic necrotizing encephalopathy: CT and MRI features. Radiology 201187 (2020).
63. Helms J, Kremer S & Meziani F More on Neurologic Features in Severe SARS-CoV-2 Infection. Reply. The New England journal of medicine vol. 382 e110 (2020).
64. Toscano G et al. Guillain–Barré Syndrome Associated with SARS-CoV-2. N. Engl. J. Med 382, 2574–2576 (2020). [PubMed: 32302082]
65. Del Valle DM et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat. Med (2020) doi:10.1038/s41591-020-1051-9.
66. Liao M et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Med 26, 842–844 (2020). [PubMed: 32398875]

Methods references

67. Korsunsky I et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019). [PubMed: 31740819]
68. Law CW, Chen Y, Shi W & Smyth GK voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29 (2014). [PubMed: 24485249]
69. Seabold S & Perktold J Statsmodels: Econometric and statistical modeling with python. in Proceedings of the 9th Python in Science Conference vol. 57 61 (Austin, TX, 2010).
70. Wolf FA, Angerer P & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018). [PubMed: 29409532]
71. West BT, Welch KB & Galecki AT Linear Mixed Models: A Practical Guide Using Statistical Software, Second *Edition*. (CRC Press, 2014).
72. Stuart T et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019). [PubMed: 31178118]
73. Cusanovich DA et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914 (2015). [PubMed: 25953818]
74. McInnes L & Healy J UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML] (2018).

75. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978 (2017). [PubMed: 28825706]

76. Fornes O et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 48, D87–D92 (2020). [PubMed: 31701148]

77. Pedregosa F et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12, 2825–2830 (2011).

78. Breiman L, Friedman J, Stone CJ & Olshen RA Classification and regression trees. (CRC press, 1984).

79. Jacomy M, Venturini T, Heymann S & Bastian M ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 9, e98679 (2014). [PubMed: 24914678]

80. Raudvere U et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198 (2019). [PubMed: 31066453]

81. Finak G et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16, 278 (2015). [PubMed: 26653891]

82. Kuleshov MV et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–7 (2016). [PubMed: 27141961]

83. Jia J et al. Cholesterol metabolism promotes B-cell positioning during immune pathogenesis of chronic obstructive pulmonary disease. EMBO Mol. Med 10, (2018).

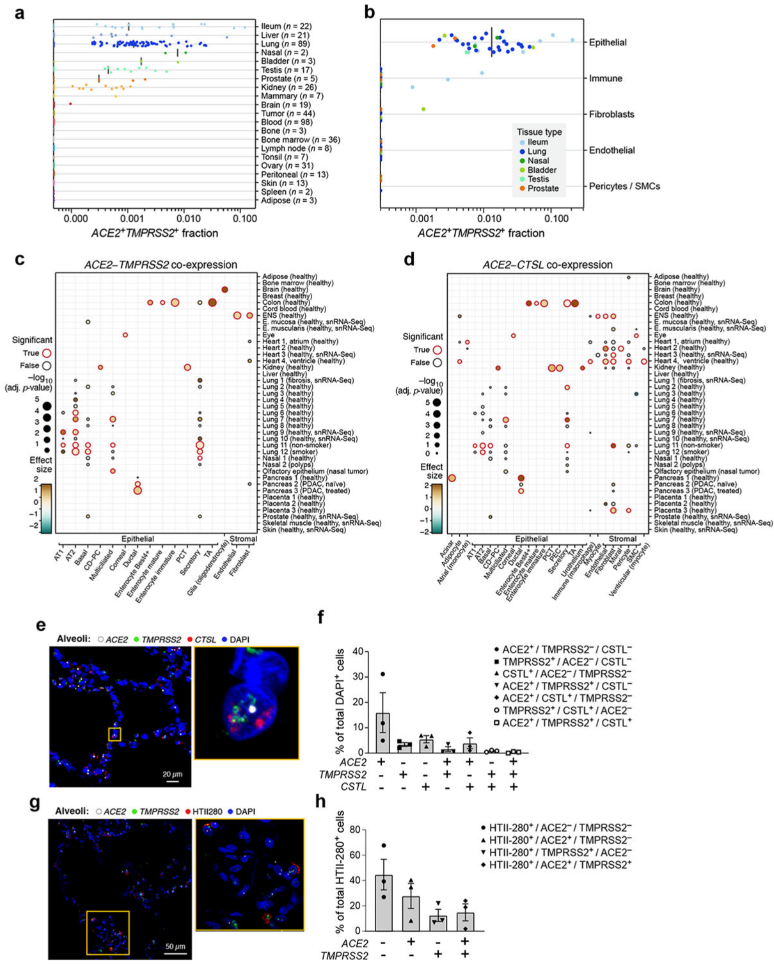


Figure 1. A cross-tissue survey of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells shows enrichment in cells at reported sites of disease transmission or pathogenesis.

(a,b) Double positive cells are more prevalent in epithelial organs and cells. (a) Proportion of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells (y axis) per dataset (dots) from 21 tissues and organs (rows). (b) Proportion of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells (y axis) within cell clusters (dots) annotated by broad cell-type categories (rows) within each of the top 7 enriched datasets (color legend, inset). (c,d) Significant co-expression of *ACE2*$^{+}$*TMPRSS2*$^{+}$ or *ACE2*$^{+}$*CTSL*$^{+}$ highlights cells from tissues implicated in transmission or pathogenesis. Significance of co-expression (dot size −log10(adjusted P-value), by two-sided Wald test (Methods); red border: FDR<0.1) of *ACE2*$^{+}$*TMPRSS2*$^{+}$ (c) or *ACE2*$^{+}$*CTSL*$^{+}$ (d) and effect size (dot color, color bar) for finely annotated cell classes (columns) from diverse tissues (rows). Only tissues and cells in at least one significant co-expression relationship are shown (Methods). (e-h) *In situ* validation of double positive cells in the lung, airways, and submucosal gland (n = 3 donors per experiment, imaged three randomly chosen areas per donor). PLISH and immunostaining (e,g) and quantification (error bars: standard error) (f,h) in human adult lung alveoli for (e) *ACE2* (white), *TMPRSS2* (green) and *CTSL* (red) (total of 1487 DAPI positive cells examined for quantification (f)) and (g) *ACE2* (white), *TMPRSS2* (green) and HTII-280 (red) (total of HTII-280 positive 482 cells examined for qualitification (h)).

Author Manuscript
Author Manuscript
Author Manuscript

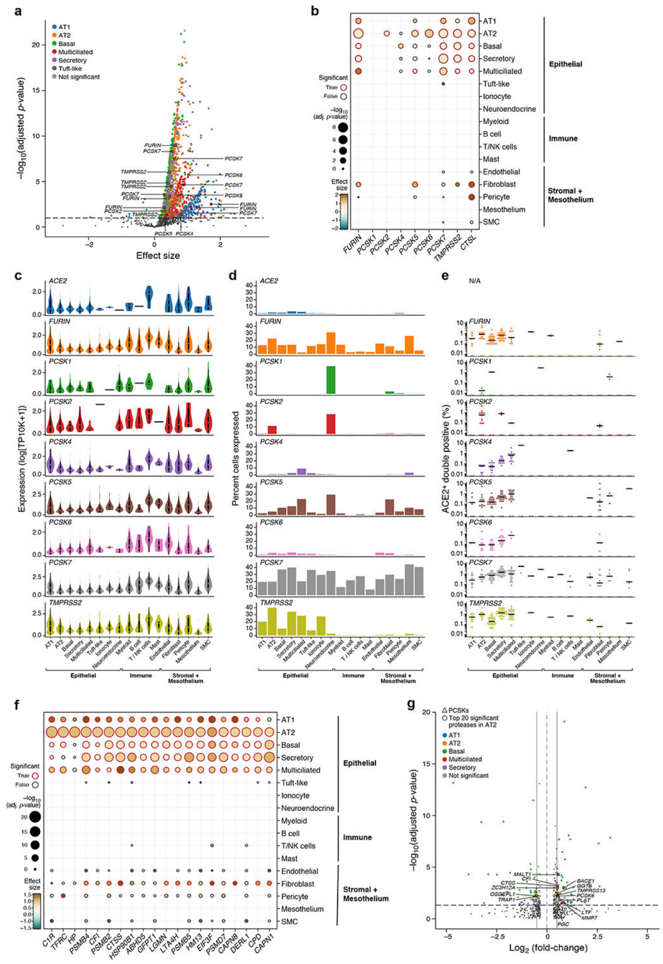


Figure 2. *ACE2*-protease co-expression and SARS-CoV-2 S-protein cleavage sites suggest a possible role for additional proteases in infection.

(**a**) Multiple proteases are co-expressed with *ACE2* in human lung scRNA-seq. Scatter plot of significance (y axis, −log10(adjusted P value)), by two-sided Wald test. (Methods) and effect size (x axis) of co-expression of each protease gene (dot) with *ACE* within each indicated epithelial cell type (color). Dashed line: significance threshold. *TMPRSS2* and PCSKs that significantly co-expressed with *ACE2* are marked. (**b**) *ACE2*-protease co-expression with PCSKs, *TMPRSS2* and *CTSL* across lung cell types. Significance (dot size, −log10(adjusted P value), by two-sided Wald test. (Methods)) and effect size (color) for co-expression of *ACE2* with selected proteases (columns) across cell types (rows). (**c,d**) Multiple proteases are expressed across lung cell types. (**c**) Distribution of non-zero expression (y axis) for *ACE2*, *PCSKs* and *TMPRSS2* across lung cell types (x axis). White dot: median non-zero expression. **(d)** Proportion of cells (y axis) expressing *ACE2*, *PCSK* family or *TMPRSS2* across lung cell types (x axis), ordered by compartment. **(e)** *ACE2*$^{+}$ *PCSK*$^{+}$ double positive cells across lung cell types. Fraction (y axis) of different *ACE2*$^{+}$ *PCSK*$^{+}$ or *ACE2*$^{+}$ *TMPRSS2*$^{+}$ double positive cells across lung cell types (x axis). Dots: different samples, line: median of non-zero fractions. **(f)** *ACE2*-protease co-expression analysis for the 20 most significant human proteases in AT2 cells. Significance (dot size, −log10(adjusted P value), by two-sided Wald test. (Methods)) and effect size (color) for co-expression of *ACE2* with different proteases (columns) across cell types (rows). **(g)** Additional protease expression in $ACE2^{+}$ $TMPRSS2^{+}$ double positive cells. Significance (y axis, −log10(adjusted P value), by two-sided Wald test. (Methods)) and fold change (x axis) of differential expression for each human protease between $ACE2^{+}$ $TMPRSS2^{+}$ double positive vs double negative cells within each indicated epithelial cell types (color). Significantly differentially expressed proteases within AT2 cells and PCSKs across all epithelial cell types are highlighted.

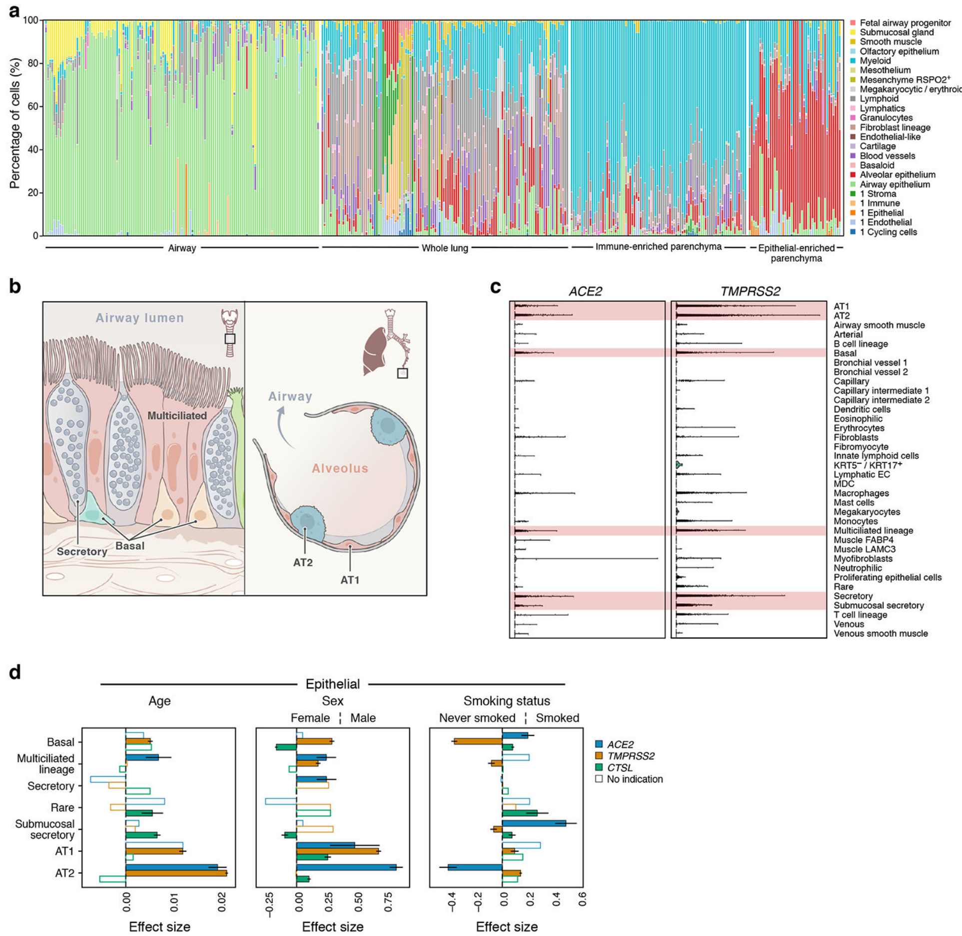


Figure 3. *ACE2, TMPRSS2*, and *CTSL* expression increases with age and in men, and shows cell type specific associations with smoking

(a) Samples in the aggregated lung and airway dataset partition to several classes by their cell composition. Percentage of cells (y axis) by level 2 cell annotations (Annotations with a preceding "1" indicate coarse annotations of cells that had no annotation at level 2) across samples (x axis). The 377 samples are ordered by sample composition clusters (Methods). (b) Schematic of key lung and airway epithelial cell types highlighted in the study. (c) Distribution of normalized *ACE2* and *TMPRSS2* expression across level 3 lung cell types in 1,031,254 cells from 228 donors. Red shading indicates the main cell types that express both *ACE2* and *TMPRSS2*. (d) Age, sex, and smoking status associations with expression of *ACE2* (blue), *TMPRSS2* (orange), and *CTSL* (green) in level 3 epithelial cells. The effect size (x axis) of the association is given as a log fold change (sex, smoking status) or the slope of log expression per year with age. As the age effect size is given per year, it is not directly comparable to the sex and smoking status effect sizes. Positive effect sizes indicate increases with age, in males, and in smokers. Colored bars: associations with an FDR-corrected p-value<0.05 (one-sided Wald test on regression model coefficients), consistent effect direction in pseudo-bulk analysis, and consistent results using the model with interaction terms (Methods). White bars: associations that do not pass all of the three above-mentioned evaluation criteria. Error bars: standard errors around coefficient estimates. Error bars are only shown for colored bars (indications or robust trends) to limit figure size.

Number of donors and cells per cell type: Basal: 155877, 105, Multiciliated lineage: 37530, 157, Secretory: 22306, 140, Rare: 2676, 71, Submucosal secretory: 33661, 45, AT1: 29973, 101, AT2: 155512 cells, 104 donors. AT1, AT2: alveolar type 1, 2; EC: endothelial cell; MDC: monocyte derived cell.

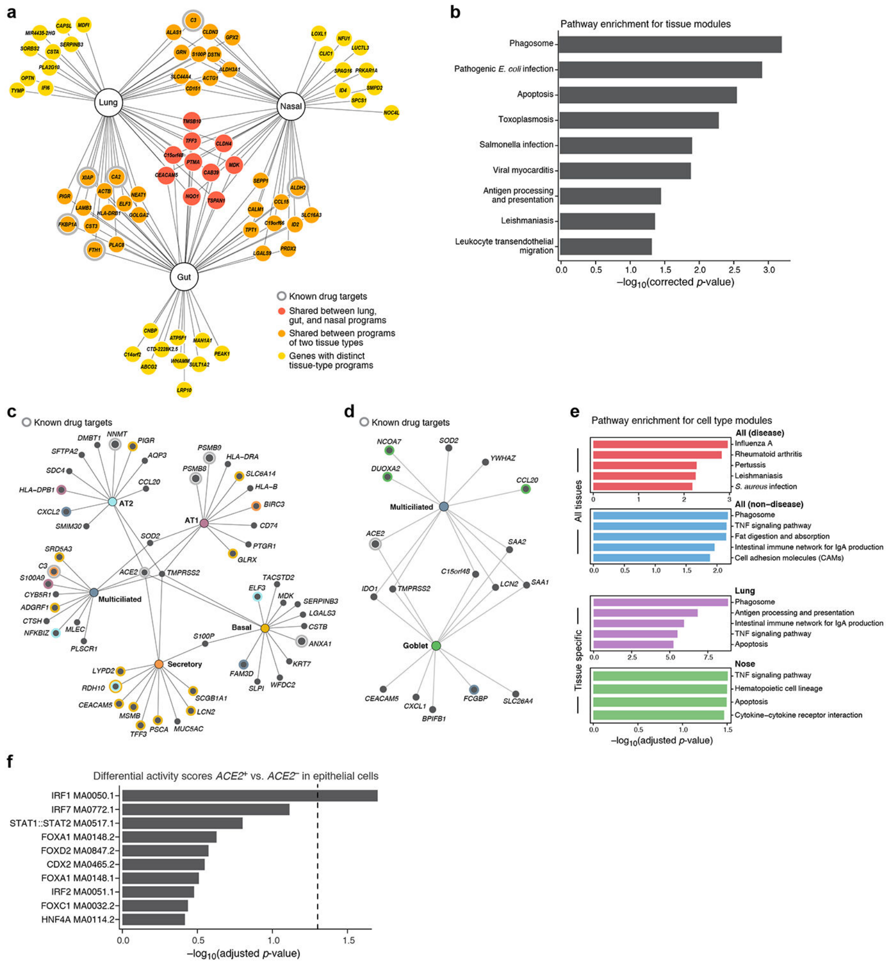


Figure 4: Tissue and cell-type-specific gene modules in *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells highlight immune and inflammatory features

(**a,b**) Tissue programs of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in lung, gut, and nasal samples. (**a**) Selected tissue program genes. Node: gene; Edge: program membership. Genes are selected heuristically for visualization (Methods). (b) Enrichment was tested using a hypergeometric test exactly as performed by gprofiler in scanpy.queries.enrich (−log10(adj P-value), x axis) of KEGG pathway gene sets (y axis) in the full tissue programs. (**c-e**) Cell programs of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells. (c,d) Top 12 genes from each cell program recovered for different lung (**c**) or (**d**) nasal epithelial cell-type (nodes, colors). Colored concentric circles: overlap with a gene in the top 250 significant genes in other cell types. *ACE2* and *TMPRSS2* are included even if not among the top 12. (**e**) Enrichment (−log10(adj P-value), x axis) of KEGG disease and non-disease pathway gene sets in either highly significant genes across all tissues (top) or in specific tissues (lung, nose, bottom). (**f**) Motif activity in immune TFs in *ACE2*$^{+}$ cells. Significance (−log10(adjusted p-value), x axis) of the top 10 differential “motif activity scores” (Methods) between epithelial *ACE2*$^{+}$ cells or *ACE2*$^{−}$ cells (y axis). (Epithelial cells are: AT1, AT2, secretory, ciliated, ionocytes, and neuroendocrine cells, highlighted in the gray shaded area in Supplementary Fig. 1a). (n=2 locations: primary carina and lung lobes, n=3 samples per location, n=1 patient). Motifs are extracted from the JASPAR2020 database, motif code is shown in each row. Dashed line: threshold for significance (adjusted p-value of 0.05). P-values were calculated by logistic regression and likelihood ratio test, adjusted through Bonferroni correction (see Methods).

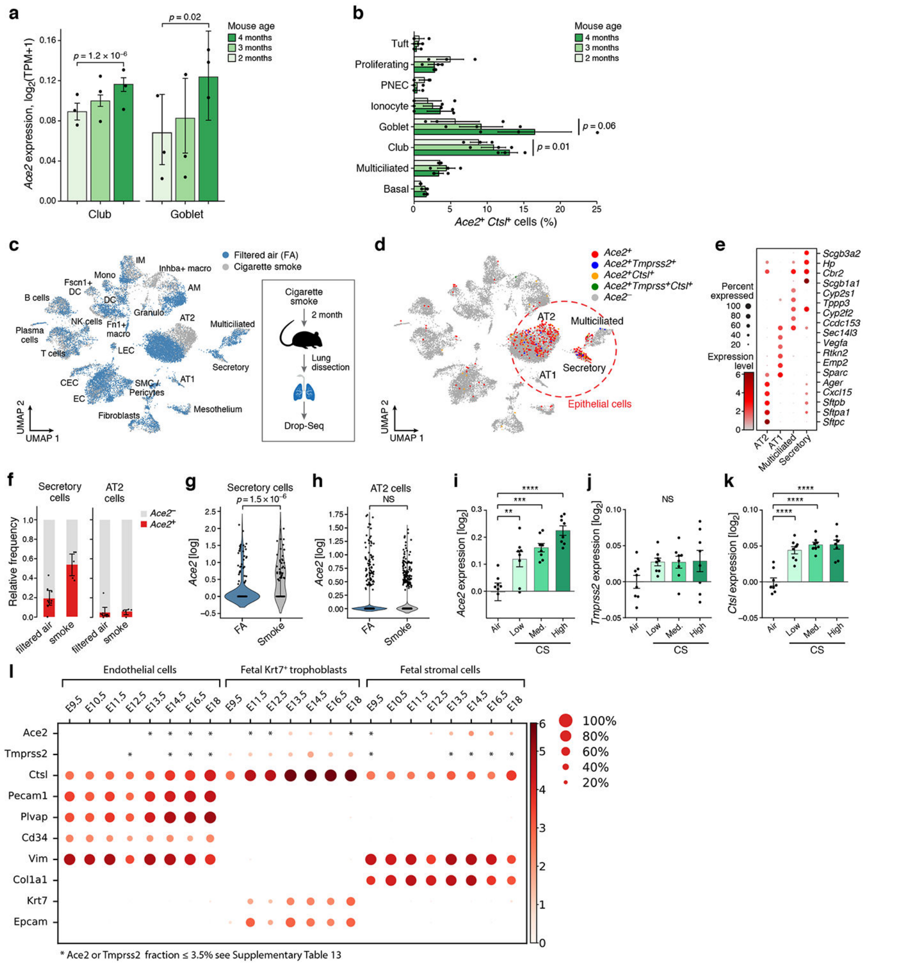


Figure 5: *Ace2*, *Tmprss2* and *Ctsl* expression in mouse in similar cell types and follows similar patterns with age and smoking.

(**a**) Gradual increase in *Ace2* expression by airway epithelial cell type with age. Mean expression (y axis) of *Ace2* in different airway epithelial cells (*x* axis) of mice of three consecutive ages (color legend, upper right). Shown are replicate mice (dots, n=3 for each age), mean (bar), and error bars (standard error of the mean (SEM)). The effect of mouse age was tested using a two-sided Wald test (p-values). (**b**) Increase in proportion of *Ace2*$^+$*Ctsl*$^+$ goblet and club cells with age. Percent of *Ace2*$^+$*Ctsl*$^+$ cells (x axis) in different airway epithelial cell types (y axis) of mice of three consecutive ages (color legend, upper right). Shown are replicate mice (dots), mean (bar), and error bars (SEM). The effect of mouse age was tested using Wald test (p-values). (**c-k**) Increase in *Ace2* expression in secretory cells with smoking. Mice were daily exposed to cigarette smoke or filtered air (FA) as control for two months after which cells from whole lung suspensions were analyzed by scRNA-seq (Drop-Seq). (**c,d**) UMAP of scRNA-seq profiles (dots) colored by experimental group (c) or by *Ace2*$^+$ cells and indicated double positive cells (d). Alveolar epithelial cells (AT1 and AT2) and airway epithelial secretory and ciliated cells are marked. (**f**) The relative frequency of *Ace2*$^+$ cells is increased by smoking in airway secretory cells but not AT2 cells. Relative proportion (y axis) of *Ace2*$^+$ (red) and *Ace2*$^-$ (grey) cells in smoking and control mice of different cell types (x axis) (filtered air (FA): n = 9 mice, smoke exposed:

n=5 mice, error bars represent 95% confidence intervals). (**g, h**) Expression of *Ace2* is increased in airway secretory cells (filtered air: 187 cells, smoke exposure: 62 cells) , but not in AT2 cells (filtered air: 3808, smoke exposure: 1882). Distribution of *Ace2* expression (y axis) in secretory (f) and AT2 (g) cells from control and smoking mice (x axis), (p-value = 1.5 10^{-6} by Wilcoxon rank-sum test). (**i-k**) Re-analysis of published bulk mRNA-Seq74 of lungs exposed to different daily doses of cigarette smoke show increased expression of (**i**) *Ace2*, (**j**) *Tmprss2*, and (**k**) *Ctsl* after five months of chronic exposure. n=8 mice per condition. Bars show mean, error bars show standard error. (** p=0.0046, *** p=0.0002, **** p<0.0001, one-way ANOVA with Dunnett's multiple comparisons test, compared to Air group.) (**l**) Expression in placenta. Mean expression (color) and proportion of expressing cells (dot size) of *Ace2, Tmprss2* and *Ctsl* along with marker genes (see Supplementary Fig. 14) in single and double positive cells from embryonic days 9.5 to 18 of mouse placenta development.