SARS-CoV-2 receptor ACE2 and TMPRSS2 are predominantly expressed in a transient secretory cell type in subsegmental bronchial branches

AUTHORS

Soeren Lukassen[1,2#], Robert Lorenz Chua[1,2#], Timo Trefzer[1,2#], Nicolas C. Kahn[3,4#], Marc A. Schneider[4,5#], Thomas Muley[4,5], Hauke Winter[4,6], Michael Meister[4,5], Carmen Veith[7], Agnes W. Boots[8], Bianca P. Hennig[1,2], Michael Kreuter[3,4*], Christian Conrad[1,2*] & Roland Eils[1,2,9*]

[#] These authors contributed equally to this work

* corresponding authors, shared senior authorship

Affiliation

[1] Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany

[2] Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany

[3] Thoraxklinik, Heidelberg University Hospital, Department of Pneumology and Critical Care Medicine, Roentgenstrasse 1, 69126 Heidelberg, Germany

[4] Translational Lung Research Center Heidelberg (TLRC), Member of the German Center for Lung Research (DZL), Im Neuenheimer Feld 156, 69120 Heidelberg, Germany

[5] Thoraxklinik, Heidelberg University Hospital, Translational Research Unit, Roentgenstrasse 1, 69126 Heidelberg, Germany

[6] Thoraxklinik, Heidelberg University Hospital, Department of Thoracic Surgery, Roentgenstrasse 1, 69126 Heidelberg, Germany

[7] Division of Redox Regulation, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

[8] Department of Pharmacology and Toxicology, NUTRIM School of Nutrition, Translational Research and Metabolism. Faculty of Health, Medicine and Life Sciences, Maastricht University, Universiteitssingel 50, Maastricht, the Netherlands

[9] Health Data Science Unit, Heidelberg University Hospital and BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

Correspondences:

roland.eils@bihealth.de

christian.conrad@bihealth.de

kreuter@uni-heidelberg.de

SUMMARY

The SARS-CoV-2 pandemic affecting the human respiratory system severely challenges public health and urgently demands for increasing our understanding of COVID-19 pathogenesis, especially host factors facilitating virus infection and replication. SARS-CoV-2 was reported to enter cells via binding to ACE2, followed by its priming by TMPRSS2. Here, we investigate *ACE2* and *TMPRSS2* expression levels and their distribution across cell types in lung tissue (twelve donors, 39,778 cells) and in cells derived from subsegmental bronchial branches (four donors, 17,521 cells) by single nuclei and single cell RNA sequencing, respectively. While *TMPRSS2* is expressed in both tissues, in the subsegmental bronchial branches *ACE2* is predominantly expressed in a transient secretory cell type. Interestingly, these transiently differentiating cells show an enrichment for pathways related to RHO GTPase function and viral processes suggesting increased vulnerability for SARS-CoV-2 infection. Our data provide a rich resource for future investigations of COVID-19 infection and pathogenesis.

KEYWORDS

SARS-CoV-2, COVID-19, ACE2, TMPRSS2, FURIN, human cell atlas, healthy lung tissue, healthy bronchial cells, epithelial differentiation, single cell RNA sequencing

INTRODUCTION

In December 2019, a disease affecting predominantly the respiratory system emerged in Wuhan, province Hubei, China, with its outbreak being linked to the Huanan seafood market as about 50% of the first reported cases either worked at or lived close to this market [Chen et al. (2020), Huang et al. (2020)]. Among the first 99 reported patients with an average age of 55.5 years, $^{2}/_{3}$ were male and 50% suffered from chronic diseases [Chen et al. (2020)]. The disease rapidly spread to other provinces in China, neighboring countries and eventually worldwide [Wang et al. (2020a), Wu and McGoogan (2020), Zhu et al. (2020)]. The World Health Organization (WHO) named the disease **co**rona**vi**rus **d**isease 20**19***,* COVID-19 (formerly known as 2019-nCov), and the virus causing the infection was designated as **s**evere **a**cute **r**espiratory **s**yndrome **co**rona**v**irus **2**, SARS-CoV-2 [Coronaviridae Study Group of the International Committee on Taxonomy of (2020)], belonging to the *Coronaviridae* family.

The two coronavirus infections affecting global public health in the 21st century were caused by SARS-CoV and MERS-CoV (**M**iddle **E**ast **r**espiratory **s**yndrome **co**rona**v**irus) [de Wit et al. (2016)]. As suggested for SARS-CoV-2, these two coronaviruses had a zoonotic origin with human-to-human transmission after their outbreak [de Wit et al. (2016)]. SARS-CoV and MERS-CoV infected 8,096 and 2,494 humans with 10% and 37% deaths, respectively [WHO (2004), WHO (2016), de Wit et al. (2016)]. As of 13 March 2020, 128,392 SARS-CoV-2 infections were reported worldwide with 4,728 deaths. The mean incubation time is hard to be precisely defined during an ongoing outbreak, but currently predicted to rank form 4 - 6.4 days [Guan et al. (2020), Backer et al. (2020)] up to 14 days as estimated by the WHO for COVID-19 and thus very similar to those of SARS and MERS [Backer et al. (2020)]. Until 11 February, most of the COVID-19 patients showed only mild symptoms (80%) or none at all (1% confirmed cases) and 14% had severe symptoms, while 5% of the COVID-19 patients were critically affected [Wu and McGoogan (2020)]. These data suggest that the estimated number of undetected SARS-CoV-2 infections is much higher and even patients with mild or no symptoms are infectious [(Lai et al., 2020); Rothe et al. (2020)]. Deaths were solely reported for the critical cases (49%), mainly for patients with preexisting comorbidities such as cardiovascular disease, diabetes, chronic respiratory disease, hypertension or cancer [Wu and McGoogan (2020)]. Although COVID-19 has a milder clinical impairment compared to SARS and MERS for the vast majority of patients, SARS-CoV-2 infection shows dramatically increased human-to-human transmission rate with the total number of deaths significantly exceeding those of SARS and MERS patients already within the first three months of the COVID-19 outbreak. The emergent global spread of SARS-CoV-2 and its strong impact on public health immediately demands for joint efforts in bio-medical research increasing our understanding of the virus' pathogenesis, its entry into the host's cells and host factors facilitating its fast replication that explains the high human-to-human transmission rates.

SARS-CoV-2 is an enveloped virion containing one positive-strand RNA genome and its sequence has already been reported [Chan et al. (2020), Lu et al. (2020), Wu et al. (2020a), Wu et al. (2020c), Zhou et al. (2020)]. The genome of SARS-CoV-2 comprises 29.9 kb and shares 79.5% and 96% identity with SARS-CoV and bat coronavirus, respectively [Zhou et al. (2020)]. Coronaviruses were reported to use

different cellular entry mechanisms in terms of membrane fusion activities after receptor binding [White and Whittaker (2016)]. SARS-CoV was previously shown to bind to angiotensin-converting enzyme 2 (ACE2) for cell entry, mediated via the viral surface spike glycoprotein (S protein) [Gallagher and Buchmeier (2001), Li et al. (2003), Simmons et al. (2013)]. Comparison of the SARS-CoV and SARS-CoV-2 S protein sequence revealed 76% protein identity [Wu et al. (2020b)] and recent studies reported that SARS-CoV-2 is also binding to ACE2 *in vitro* [Hoffmann et al. (2020), Walls et al. (2020), Yan et al. (2020), Zhou et al. (2020)]. Subsequently, the S protein is cleaved by the transmembrane protease serine 2 TMPRSS2 [Hoffmann et al. (2020)]. Simultaneously blocking TMPRSS2 and the cysteine proteases CATHEPSIN B/L activity inhibits entry of SARS-CoV-2 *in vitro* [Kawase et al. (2012)], while the SARS-CoV-2 entry was not completely prohibited *in vitro* [Hoffmann et al. (2020)]. It remains to be determined whether additional proteases are involved in priming of the SARS-Cov-2 but not SARS-CoV S protein. One candidate is FURIN as the SARS-CoV-2 S protein contains four redundant FURIN cut sites (PRRA motif) that are absent in SARS-CoV. Indeed, prediction studies suggest efficient cleavage of the SARS-CoV-2 but not SARS-CoV S protein by FURIN [Coutard et al. (2020), Wu et al. (2020b)]. Additionally, FURIN was shown to facilitate virus entry into the cell after receptor binding for several coronaviruses, e.g. MERS-CoV [Burkard et al. (2014)] but not SARS-CoV [Millet and Whittaker (2014)]. In addition, although *ACE2* was previously described to be expressed in the respiratory tract [Jia et al. (2005), Ren et al. (2006), Xu et al. (2020), Zhao et al. (2020), Zou et al. (2020)], still little is known about the exact cell types expressing *ACE2* and *TMPRSS2* serving as entry point for SARS-CoV and the currently emerging SARS-CoV-2. Therefore, there is an urgent need for investigations of tissues in the upper and lower airways in COVID-19 patients but also healthy individuals to increase our understanding of the host factors facilitating the virus entry and its replication, ultimately leading to treatment strategies of SARS-CoV-2 infections.

As pointed out recently [Zhang et al. (2020b)], our understanding of host genetic factors involved in COVID-19 disease outcome is still poor.

RESULTS

ACE2 and its co-factor *TMPRSS2* are expressed in the lung and bronchial branches

Here, we established a rich reference data set that describes the transcriptional landscape at the single cell level of the lung and subsegmental bronchial branches of in total 16 individuals (Fig. 1A). Based on this resource, we set out to identify potential key mechanisms likely involved in the SARS-CoV-2 pathway. First, we investigated the expression patterns of the SARS-CoV-2 receptor *ACE2* and the serine protease priming its S-protein, *TMPRSS2*, in individual cells in the lung and in subsegmental bronchial branches (Fig. 1A).

To quantify gene expression in the lung, single nuclei RNA sequencing was performed on surgical specimens of healthy, non-affected lung tissue from twelve lung adenocarcinoma (LADC) patients, resulting in 39,778 sequenced cell nuclei. All major cell types known to occur in the lung were identified (Fig. 1B, C). Independent of the cell types present in the lung, the median *ACE2* levels were below five counts per million (CPM) (Fig. 1D), which given a typical mRNA content of 500,000 mRNA molecules per cell indicate that only about half of all cells were statistically expected to contain even a single *ACE2* transcript. The reads per patient and cell type were therefore aggregated into pseudo-bulks and analysis was continued. As expected from prior literature, the AT2 cells showed highest *ACE2* expression in the lung both in terms of their CPM values (Fig. 1D, further referred to as *ACE2*$^+$ cells) and the percentage of positive cells (Supp. Fig. 3). The expression of *TMPRSS2* across cell types of the lung (further referred to as *TMPRSS2*$^+$ cells) was much stronger with a certain specificity for AT2 cells (Fig. 1E), which is in agreement with previous studies.

For the subsegmental bronchial branches, air-liquid-interface (ALI) cultures were grown from primary human bronchial epithelial cells (HBECs) and subjected to single cell RNA sequencing, resulting in 17,451 cells across four healthy donors (further referred to as HBECs). In this dataset, we identified all expected cell types, including recently described and rare cell types such as *FOXN4*-positive cells and ionocytes (Fig. 1F, Supp. Fig. 1, 2). Expression levels of *ACE2* were comparable but slightly elevated compared to the lung tissue samples, with the strongest expression being

observed in a subset of secretory cells ('secretory3'; Fig. 1G). While *TMPRSS2* was less strongly expressed in HBECs than in lung tissue cells, its expression was still markedly higher than that of *ACE2* (Fig. 1H). Interestingly, *ACE2*$^+$ cells were also most strongly expressing *TMPRSS2* (Fig. 1F, H). In the HBECs, a significant enrichment of the number of *ACE2*$^+$/*TMPRSS2*$^+$ double-positive cells was observed (1.6 fold enrichment, p=0.002, Supp. Fig. 4).

Transient secretory cells display a transient cell state between goblet and ciliated cells.

As the 'secretory3' cells in the HBECs showed comparably strong expression of both genes implicated in SARS-CoV-2 entry, with *ACE2* expression levels in a similar to higher range than those in AT2 cells, we decided to further investigate this cell type. As airway epithelial cells are known to continuously renew in a series of differentiation events involving nearly all major HBE cell types in a single trajectory, we mapped the 'secretory3' cells according to their position in this hierarchy using pseudo-time mapping (Fig. 2A). 'Secretory3' cells were found at an intermediate position between club or goblet cells and ciliated cells (Fig. 2A), which could be confirmed by shared marker gene expression (Fig. 2B). This placed 'secretory3' cells at a point shortly before terminal differentiation is reached (Fig. 2C, D). Hence, we will refer to these cells as 'transient secretory cells'.

In agreement with ongoing differentiation towards an epithelial cell lineage, the cell-specific markers of the transient secretory cells showed an enrichment of RHO GTPases and their related pathways (Fig. 2E). Note that we find this enrichment of pathways exclusively in the transient secretory cells but in none of the other cell types. As RHO GTPases have been implicated in membrane remodeling and the viral replication cycle, especially entry, replication, and spread [Van den Broeke et al. (2014)], transient secretory cells may be more permissive to SARS-CoV-2 infection adding to a potential vulnerability caused by considerably high co-expression levels of *ACE2* and *TMPRSS2*. Interestingly, these cell marker genes were not enriched in any of the cell types in the whole lung indicating the absence of this transient secretory cell type in the lung samples. This may indicate a different entry route for the SARS-CoV-2 virus in the more central bronchial branches compared to the peripheral lung tissue.

FURIN expressing cells overlap with *ACE2*$^+$/*TMPRSS2*$^+$ and *ACE2*$^+$/*TMPRSS2*$^-$ cells in lung and in bronchial cells.

We next sought to explore whether we could derive additional evidence for other factors recently suggested to be involved in SARS-CoV-2 host cell entry. Therefore, we investigated the expression of *FURIN*, as SARS-CoV-2 but not SARS-CoV was reported to have a FURIN cleavage site in its S protein, potentially increasing its priming upon ACE2 receptor binding. In lung tissue, AT2, endothelial, and immune cells were most strongly positive for *FURIN* expression (Fig. 3A). Across cell types, we observed a marked enrichment of the number of double and triple positive cells for any combination of *ACE2*, *TMPRSS2*, and/or *FURIN* expression, indicating a preference for co-expression (Fig. 3B). Interestingly, also transient secretory cells in the HBECs showed an intermediate expression of *FURIN* (Fig. 3C). While *TMPRSS2* and *FURIN* were less strongly expressed in HBECs than in distant cells from surgical lung tissue, their expression was still markedly higher compared to *ACE2*. In the HBECs, a significant enrichment of the number of *ACE2*$^+$/*TMPRSS2*$^+$ double-positive cells was observed, while *ACE2*$^+$/*TMPRSS2*$^+$/*FURIN*$^+$ enrichment did not reach significance (Fig. 3D). The latter finding, however, may be due to differences in co-expression or the generally lower number of positive cells reducing statistical power. Interestingly, our data showed that the additional possibility of priming the SARS-CoV-2 S protein by FURIN would potentially render 25% more cells vulnerable for infection as compared to by exclusively TMPRSS2 S protein priming. Thus, our data suggest that FURIN might increase overall permissiveness of cells in the respiratory tract by potentially equipping more cells with proteolytic activity for SARS-CoV-2 S protein priming after ACE2 receptor binding.

Sex, age and history of smoking in correlation with *ACE2* expression

Early epidemiological data on SARS-CoV-2 transmission and spread have suggested age, sex and history of smoking among others as potential confounding factors impacting SARS-CoV-2 infection [Brussow (2020), Chen et al. (2020), Huang et al. (2020), Zhang et al. (2020a)]. We thus investigated these possible risk factors for COVID-19 and their influence on receptor gene expression (Fig. 4A). No sex-related difference could be observed in the *ACE2* expression in individual cell types neither in the lung tissue nor in the HBECs (Fig. 4C, D). This observation is in full agreement

with recent reports based on a large number of COVID-19 patients [Guan et al. (2020)]. Although the first reported cases suggested higher infection rates for males [Chen et al. (2020), Huang et al. (2020)], no significant differences in the infection rate of males and females were found with increasing numbers of COVID-19 patients [Wang et al. (2020b), Brussow (2020), Zhang et al. (2020a)].

It has to be noted, though, that the power of our datasets to detect such changes with lung cells from surgical lung tissue samples from twelve patients (three male, nine female) and HBECs from four patients (three male, one female) is limited. On the individual cell type level, no *ACE2* expression differences correlating with age could be observed (Fig. 4E, F), and neither did the cell type composition change notably (Supp. Fig. 6). Although no dependency of *ACE2* expression on age, gender and sex was found in single cell populations, we observed interesting trends in the lung for age and gender dependencies when aggregating all reads per sample into a single *ACE2* expression value (Figure 4G, H). Here, we see a trend for increasing overall expression of *ACE2* with age for all female lung samples (R^2=0.40; p=0.065). Since we only have samples from males of younger age available, we cannot study age dependencies for males here. However, when comparing *ACE2* expression of all five female and three male samples in the young age group between 40 and 50 years of age, we see a clearly higher overall expression of *ACE2* in males compared to females (p=0.048). Note that smoking did not seem to affect *ACE2* expression levels in the lung cells within our dataset as all smokers in our study were of about the same age and *ACE2* expression correlated with age (Fig. 4B).

Altogether, we here demonstrate how single cell sequencing studies of samples from anatomically precisely defined parts of the respiratory system can provide new insights into potential host cell vulnerability for SARS-CoV-2 infection.

DISCUSSION

The here presented transcriptome data on single cell level of healthy human lung tissues, including surgical lung specimen and subsegmental bronchial branches, contains valuable information about human host factors for SARS-CoV-2 infection and other related infectious diseases affecting the respiratory tract. This resource certainly comes with limitations (see below). However, we believe the unprecedented

depth of our analysis on a single cell level will provide a valuable resource for future mechanistic studies and target mining for pulmonary host factors that are involved in facilitating virus entry and replication, ultimately leading to defining genes that are of urgent interest for studying transcriptional changes in COVID-19 patients or the SARS-CoV-2 pathogenicity in general.

This data set comprises 16 individuals and a total of 57,229 cells. Thus, this large single-cell and nuclei expression resource enables the analysis of weakly expressed genes such as *ACE2* and identification of rare cell types and the transient secretory cell type, for which our data showed a particularly high level of potential vulnerability for SARS-CoV-2 infection assuming that ACE2 is the receptor that is likely to be crucial for its cell entry. Although we strongly believe that the here presented data is well suited as reference data set to study SARS-CoV-2 infection on the single cell transcriptional level, we are also well aware of potential limitations of our data.

First, the here presented data is derived from individuals that have no infection history with SARS-CoV-2. However, we deem our data clinically meaningful as our patient cohort is representative for those being affected by SARS-CoV-2 associated disease. The patient cohort analyzed here is in line with recently published data on characteristics of COVID-19 patients with regards to age (this study: median age 48 years - lung cohort or 55 years - HBEC cohort; Chen *et al.*, 2020: 55 years; Guan *et al.*, 2020: 47 years) [Chen et al. (2020), Guan et al. (2020)] (Guan NEJM, Chen Lancet) and especially comorbidity burden [Chen et al. (2020)] with 50% of the patients being affected by chronic cardiovascular or pulmonary disease or diabetes (this study: 50%; Chen *et al.*, 2020: 51%; Guan *et al.*, 2020: no data available). However, there was a different prevalence of smoking or history of smoking in our cohort (this study: 25% - lung cohort or 100% HBEC cohort; Guan *et al.*, 2020: 85%)

Second, we are combining data of different sections of the lung and use two different RNA sequencing methods on single cell level. While single cell RNA sequencing typically delivers higher quality data, the cells must be intact for processing, as damaged cells present a loss of RNA molecules with mitochondrial reads preferentially retained, introducing a skewed expression profile [Luecken and Theis (2019)]. Adequate sampling of lung specimen for preparing suspensions of intact

cells is often challenging. Single nuclei RNA sequencing also reveals transcriptome data on single cell level and can be used on frozen tissue. This method typically provides a more faithful estimate of cell type proportions due to lessened dissociation bias. However, it typically comes with less reads per cell and less genes detected [Bakken et al. (2018)], although this bias seems low in our dataset (Supp. Fig. 7). The identification of closely related cell populations is mostly concordant in both methods, with single cell RNA sequencing being inferior for detection of lowly expressed genes on a single-cell level [Bakken et al. (2018)]. The gene expression shown here for cells from surgical lung tissue samples and ALI-cultured cells derived from the subsegmental bronchial tree (HBECs) should not be directly compared, as the different techniques are likely to have an influence on absolute expression values. However, qualitative comparisons of relative abundances as performed here are still appropriate.

Third, the number of samples included here is very much limiting the scope of understanding the pathogenicity of SARS-CoV-2 in the context of different confounding factors such as age, gender and history of smoking. As a consequence of small sample numbers, the significance level for dependency of expression of single genes, most predominantly *ACE2*, on such confounding factors are at best weak. However, we believe that it is a strength of our data that our clinical samples are well annotated w.r.t. anatomical position and patient characteristics. Thus, our data is in principle suited for testing hypothesis derived from larger epidemiological studies. Clearly, measuring expression levels for specific genes in larger patient cohorts is required to further substantiate the hypothesis derived from our and others' data.

Bearing the above limitations in mind, we investigated *ACE2* expression levels in all cell types in the lung and bronchial branches revealing very low expression levels that are nevertheless significantly enriched in AT2 cells in the lung as described previously [Hamming et al. (2004)] and are comparable with the levels in specific cell types in the bronchial branches. As ACE2 is the only currently known SARS-CoV-2 receptor on the host cell surface, we propose that higher expression levels of *ACE2* facilitate infection by SARS-CoV-2.

While we did not find any dependency of *ACE2* expression on sex, gender and age on the single cell level, we observed a trend for age dependency on *ACE2* expression aggregated over all cell types in lung samples from females. Recent data suggest that infection rates are similar in children as in adults, but remains undiagnosed since the clinical picture is often subclinical [Bi et al. (2020)]. Hence, it would be of utmost interest to study both healthy and infected children to understand the transcriptional basis accounting for differences in development of clinical symptoms. Furthermore, analyzing further samples from both males and females, both healthy and infected, of different age groups on the single cell level will lay the foundation for understanding vulnerability of different cell types for SARS-CoV-2 infection in different parts of the respiratory system.

One emergent question is why the human-to-human transmission of SARS-CoV-2 is much higher compared to SARS-CoV or MERS-CoV. Potential explanations comprise i) the binding of SARS-CoV-2 to another, yet unknown receptor on the host cell surface, ii) enhanced cleavage of the SARS-CoV-2 S protein resulting in higher efficiency of the virus' entry into the cell, and iii) additional host factors increasing the virus entry into the cell, e.g. by facilitating membrane fusion. We used our reference map and our above described findings, seeking for additional host factors that might be involved in any of those mechanisms.

Coronaviruses were shown to be able to enter into the host cell via several pathways, including endosomal and non-endosomal entry in the presence of proteases [Simmons et al. (2004), Matsuyama et al. (2005), Wang et al. (2008), White and Whittaker (2016)]. Different proteases were previously shown to facilitate entry into the cells for different coronaviruses [Hamming et al. (2004)]. Both SARS-CoV and SARS-CoV-2 were shown to share the ACE2 as cell surface receptor and TMPRSS2 as the major protease facilitating their entry into the host cell [Hoffmann et al. (2020), Walls et al. (2020), Yan et al. (2020), Zhou et al. (2020)]. Nevertheless, SARS-CoV-2 human-to-human transmission rates appear to be significantly higher as compared to of SARS-CoV, given the massively higher transmission rate three months after its first detection, with about 130,000 confirmed infections and more than 4,700 deaths (as compared to 8,096 SARS patients with 2,494 deaths from 2002-2003).

SARS-CoV-2 was shown to comprise a FURIN cleavage site that is absent in SARS-CoV [Coutard et al. (2020), Wu et al. (2020b)]. In addition to the presence of this FURIN cleavage site and experimental evidence that MERS-CoV is primed by FURIN, also recent data published by Hoffmann *et al.* (2020) suggest an additional protease besides TMPRSS2 and CATHEPSIN B/L as blockage of both completely prohibited the entry of SARS-CoV but not SARS-CoV-2. However, the involvement of FURIN in the virus infection pathway remains to be confirmed by future experimental approaches. Further evidence for an involvement of FURIN in the entry of SARS-CoV-2 into the host cell comes from very recent studies predicting increased binding affinity between the SARS-CoV-2 S protein and the human ACE2 receptor as compared to the SARS-CoV S protein [Wrapp et al. (2020), Wu et al. (2020b)]. Wu *et al.* postulate that the cleavage of the S protein at the FURIN cut site might cause the increased binding affinity of SARS-CoV-2 to its receptor, e.g. triggered by structural rearrangements of the cleaved S protein as it was previously shown for other coronavirus S proteins [Kirchdoerfer et al. (2016), Walls et al. (2016), Wrapp et al. (2020)].

As the involvement of an additional protease during S protein priming might enhance entry of SARS-CoV-2 compared to SARS-CoV, we also investigated co-expression of *ACE2*, *TMPRSS2*, and *FURIN*. Using the here presented reference data set, we were able to show co-expression of *ACE2* with *TMPRSS2* and/or *FURIN* in both healthy human lung tissue and HBECs. The potential additional priming of 25% of cells for SARS-CoV-2 infection by FURIN may support the hypothesis for increased human-to-human transmission rates. This hypothesis, however, must be followed up by future studies.

Besides higher SARS-CoV-2 infection rate of the cells due to FURIN cleavage, also the severity of COVID-19 in some of the clinical cases might be explained by this additional FURIN motif. Recent prediction studies suggest efficient cleavage of the SARS-CoV-2 S protein by FURIN that might result in higher pathogenicity of the virus [Coutard et al. (2020), Wu et al. (2020b)], potentially due to an increased affinity to the ACE2 receptor [Wrapp et al. (2020), Wu et al. (2020b)]. Further evidence for correlation of increased virus pathogenicity with FURIN priming arises from the

observation that the presence of a FURIN-like cleavage site positively correlated with pathogenicity for different influenza strains [Kido et al. (2012)].

MERS-CoV is an additional coronavirus that comprises a FURIN-like motif [Burkard et al. (2014)]. Although this virus did not infect as many individuals as SARS-CoV or the currently emerging SARS-CoV-2, it was highly lethal (37% of MERS patients) in 2016 [WHO (2016), de Wit et al. (2016)]. However, although both SARS-CoV-2 and MERS-CoV are likely to be cleaved by FURIN during S protein priming, we speculate that the MERS-CoV receptor DPP4 does not have any or high relevance for COVID-19 as DPP4 is solely expressed in lung cells but not in the subsegmental bronchial branches (Supp. Fig. 5). Taken together, these data offer the interesting hypothesis that both high human-to-human transmission rate of SARS-CoV-2 and the severe COVID-19 cases might be caused by additional cleavage sites, resulting in higher ACE2 binding affinity and/or more efficient membrane fusion.

The most striking finding of this study was the detection of a specific cell type in the bronchial branches expressing *ACE2*, i.e. the transient secretory cells. Subsequent cell type marker analysis revealed that these cells reflect differentiating cells on the transition from club or goblet to terminally differentiated ciliated cells, the latter having an important role in facilitating the clearance of viral particles [Dumm et al. (2019), Sims et al. (2008)]. RHO GTPase activating CIT and viral processes regulating pathways are strongly enriched in the transient secretory cells. Interestingly, about 40% of the *ACE2*$^{+}$ transient secretory cells are co-expressing the protease *TMPRSS2* that is known to be involved in S protein priming, making this cell type potentially highly vulnerable for SARS-CoV-2 infection. If we include FURIN as another protease priming in SARS-CoV-2 entry of the host cell, the percentage of transient secretory cells expressing the receptor *ACE2* and either both or one of the proteases *TMPRSS2* and *FURIN*, increases to 50%. Hence, transient secretory cells would be potentially highly vulnerable for SARS-CoV-2 infection.

Interestingly, SARS-CoV replication was reported to be 100- to 1,000-fold higher via the non-endosomal pathway as compared to the endosomal pathway, with the non-endosomal pathway being dependent on the presence of proteases [Matsuyama et al. (2005)]. Therefore, our data suggest to experimentally investigate whether SARS-

CoV-2 is able to enter into the cell also via the non-endosomal pathway, either by the involvement of several proteases or by the ongoing membrane remodeling in the transient secretory cells, which might include RHO GTPases as one prominent pathway specifically enriched in this cell type.

Taken together, we present a rich resource for studying transcriptional regulation of SARS-CoV-2 infection, which will serve as reference dataset for future studies of primary samples of COVID-19 patients and *in vitro* studies addressing the viral replication cycle. With all limitations in mind, which we discussed above, we demonstrate the potential of this resource for deriving novel and testing existing hypotheses that need to be followed up by independent studies.

ACKNOWLEDGEMENTS

Cryopreserved surgical lung tissues from patients were kindly provided from the Lung Biobank Heidelberg, a member of the accredited Tissue Bank of the National Center for Tumor Diseases (NCT) Heidelberg, the BioMaterialBank Heidelberg and the Biobank platform of the German Center for Lung Research (DZL). We thank Martin Fallenbüchel and Christa Stolp for collection of tissue samples and Elizabeth Chang Xu for establishment of ALI cultures. We thank Leif Erik Sander, Irina Lehmann, and Saskia Trump for advice in this study. This study was supported by the European Commission (ESPACE, 874710, Horizon 2020) and the German Center for Lung Research (DZL, 82DZL00402).

AUTHOR CONTRIBUTIONS

N.C.K, M.A.S., M.M., C.C., and R.E. conceived and designed the project. N.C.K, M.A.S., M.M., A.W.B., B.P.H., M.K., C.C., and R.E. supervised the project. R.L.C, T.T., M.A.S., and C.V. performed experiments. S.L., R.L.C., T.T., and B.P.H. analyzed data. N.C.K., T.M., H.W., C.V., A.W.B. provided tissue and cells. S.K., M.A.S, T. M., B.P.H., M.K. C.C., and R.E. wrote the manuscript, all authors read and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1 – *ACE2* and *TMPRSS2* are expressed in specific cell types in lungs and HBECs. A) Sampling location of the surgical lung specimens and human bronchial epithelial cells (HBECs) used in this study. Blue rectangle is zoomed in B) Overview of the major cell types in the lung and airways. C) Uniform manifold approximation and projection (UMAP) of primary lung samples single nuclei RNA sequencing (cell types are color-coded). D) Expression values of *ACE2* in the cell types of primary lung samples. E) Expression values of *TMPRSS2* in the cell types of primary lung samples. F) UMAP projections of HBE single cell RNA sequencing data (cell types are color-coded), G) Expression values of *ACE2* in the cell types of HBECs. H) Expression values of *TMPRSS2* in the cell types of HBECs. Boxes in box plots indicate the first and third quartile, with the median shown as horizontal lines. Whiskers extend to 1.5 times the inter-quartile-range. All individual data points are indicated on the plot.

Figure 2 – Characterization of transient secretory (secretory3) cells. A) Pseudo-time trajectory projected onto a UMAP embedding of HBECs. The location of the secretory3 cell type is marked by a red outline (cell types are color-coded). B) Expression of goblet (top row) and ciliated cell markers (bottom row) in club, goblet, secretory3, and ciliated cells. C) Pseudo-time trajectory projected onto a UMAP embedding of HBECs (pseudo-time values are color-coded). D) Cell type composition along a binned pseudo-time axis. The earliest stages are located on the left. E) Pathway enrichment values for secretory3-specific marker genes.

Figure 3: *FURIN* is expressed in *ACE2*$^+$ and *ACE2*$^+$/*TMPRSS2*$^+$ cells. A) Expression values of *FURIN* in the cell types of primary lung. B) Overlap of *ACE2*$^+$, *TMPRSS2*$^+$, and *FURIN*$^+$ cells in the lung dataset. RF: representation factor, enrichment. P: hypergeometric tail probability. Total number of cells: 39,778. C) Expression values of *FURIN* in HBECs. D) Overlap of *ACE2*$^+$, *TMPRSS2*$^+$, and *FURIN*$^+$ cells in the HBEC dataset. RF: representation factor, enrichment. P:

hypergeometric tail probability. Total number of cells: 17,451. Boxes in box plots indicate the first and third quartile, with the median shown as horizontal lines. Whiskers extend to 1.5 times the inter-quartile-range. All individual data points are indicated on the plot.

Figure 4 – Age, sex and smoking behavior as possible factors influencing *ACE2* expression. A) Metadata of the sample donors. B) Expression of *ACE2* in the cell types of the primary lung (history of smoking color-coded). C) and D) Expression of *ACE2* in C) HBECs and D) primary lung (sex is color-coded). E) and F) Expression of *ACE2* in E) HBECs and F) primary lung (age is color-coded). G) and H) *ACE2* expression over all cell types versus age in G) female primary lung cells and H) primary lung cells of patients aged 40 to 50 (sex is color-coded). Boxes in box plots indicate the first and third quartile, with the median shown as horizontal lines. Whiskers extend to 1.5 times the inter-quartile-range. All individual data points are indicated on the plot.

METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Roland Eils (roland.eils@charite.de).

Primary lung and subsegmental bronchial samples are available upon request and upon installment of a material transfer agreement (MTA).

There are restrictions to the availability of the dataset due to potential risk of de-identification of pseudonymized RNA sequencing data. Hence, the raw data will be available under controlled access. Accession numbers, DOIs, or unique identifiers for raw data will be listed here upon submission to the community-endorsed public repository.

Processed data in the form of count tables and metadata tables containing patient ID, sex, age, smoking status, cell type and QC metrics for each cell are available on FigShare (https://doi.org/10.6084/m9.figshare.11981034.v1) and Mendeley Data (https://data.mendeley.com/datasets/7r2cwbw44m/1).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human lung tissues and bronchial branches

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki. The use of biomaterial and data for this study was approved by the local ethics committee of the Medical Faculty Heidelberg (S-270/2001 and S-538/2012).

Cryoconserved surgical healthy lung tissue from twelve patients was provided by the Lung Biobank Heidelberg.

Sex, age and smoking behavior for every individual is provided in figure 4a and additional information can be found in table 1.

METHOD DETAILS

Lung tissues

Tissue was assembled during routine surgical intervention from lung cancer patients. A representative part of normal lung tissue distant from the tumor was cut into pieces of about 0.5-1 cm^3 immediately after resection. Up to 4 pieces were distributed in 2 ml cyro vials (Greiner Bio-One GmbH). Thereafter, the vials were snap frozen in liquid nitrogen (within 0.5 h after resection). The tissue samples had no direct contact to the liquid liquid nitrogen. After snap-freezing, the vials were stored at –80°C until use [Muley et al. (2012)]. Patients' characteristics are shown in table 1 (lung tissues).

Air liquid interface (ALI) cultures of HBECs

Human bronchial epithelial cells (HBECs) were obtained from endobronchial lining fluid (ELF) by minimally invasive bronchoscopic microsampling (BMS) from subsegmental airways for further investigation of indeterminate pulmonary nodules as described previously [Kahn et al. (2009)]. ELF was obtained from a noninvolved segment from the contralateral lung. Patients' characteristics are shown in table 1 (ELF patients).

Primary HBECs were maintained in DMEM/F12 media (Gibco, Carlsbad, CA) supplemented with bovine pituitary extract (0.004 mL/mL), epidermal growth factor (10 ng/mL), insulin (5 µg/mL), hydrocortisone (0.5 µg/mL), triiodo-L-thyronine (6.7 ng/mL) and transferrin (10 µg/mL) (PromoCell), 30 nM sodium selenite (Sigma), 10 µM ethanolamine (Sigma), 10 µM phosphorylethanolamine (Sigma), 0.5 µM sodium pyruvate (Gibco), 0.18 mM adenine (Sigma), 15 mM Hepes (Gibco), 1x GlutaMAX (Gibco) and in the presence of 10 µM Rock-inhibitor (StemCell). After reaching confluency, cells were transferred to 12-well plates and plated at 90.000 cells/insert using the ThinCert™ with 0.4 µm pores (#665641, Greiner Bio-One). After 2-3 days, cells were airlifted by removing the culture from the apical chamber and PneumaCult-ALI media (StemCell) was added to the basal chamber only. Differentiation into a pseudostratified mucociliary epithelium was achieved after approximately 24-27 days. Afterwards, ALI cultures were characterized for their proper differentiation using Uteroglobin/CC10 (#RD181022220-01, BioVendor), Mucin 5AC (#ab3649, abcam), Keratin 5 (#HPA059479, Merck) and Tubulin beta 4 (#T7941, Merck). To do so, filters were fixed with 4% paraformaldehyde (PFA, Merck) and permeabilized with 0.1% Triton X-100 (Merck). First antibodies were applied over night at 4°C. Filters were incubated with Alexa Fluor 488 and 549 secondary antibodies (Dianova) for 40 minutes at 37°C. DNA was stained using Hoechst 33342 (#B2261, Merck). Pictures were taken using Leica SP5 confocal microscope and Leica Application Suite software (Leica). Pictures were assembled using Photoshop CS6 (Adobe).

Table 1. Patient characteristics. ELF: Epithelial lining fluid; ECOG: Eastern Cooperative Oncology Group Performance Status. Scale Data are expressed as range (mean ± SD). (* $P < 0.05$, *** $P < 0.001$).

	Surgically lung patients	**ELF sampled patients**
Number and gender (male/female)	12 (3/9)	4 (3 / 1)
Age (years)	44-79 (48 ± 13)	53-67 (55 ± 6.4)
ECOG		
0	12	3
1	0	1
Smoking status		
current smokers	3	2
ex-smokers	0	2

never smokers	9	0

SINGLE NUCLEI AND SINGLE CELL RNA SEQUENCING

Single nuclei isolation (lung tissues)

Essentially, single nuclei were isolated as described elsewhere [Tosti et al. (2019)]. Briefly, snap frozen healthy lung tissue from lung adenocarcinoma patients was cut into pieces with less than 0.3 cm diameter and single nuclei were isolated at low pH by homogenizing the cells in 1 ml of citric acid-based buffer (Sucrose 0.25 M, Citric Acid 25 mM, Hoechst 33342 1 g/ml) at 4°C using a glass dounce tissue grinder. After one stroke with the “loose” pestle, the tissue was incubated for 5 minutes on ice, further homogenized with 3-5 strokes of the “loose” pestle, followed by another incubation at 4°C for 5 minutes. Nuclei were released by 5 additional strokes with the “tight” pestle and the nuclei solution was filtered through a 35 µm cell strainer. Cell debris was removed via centrifugation at 4°C for 5 minutes at 500x g, the supernatant was removed, followed by nuclei cell pellet resuspension in 700 µl citric-acid based buffer and centrifugation at 4°C for 5 minutes at 500x g. After carefully removing the supernatant, the nuclei cell pellet was resuspended in 100 µl cold resuspension buffer (25 mM KCl, 3 mM $MgCl_2$, 50 mM Tris-buffer, 400 U RNaseIn, 1 mM DTT, 400 U SuperaseIn (AM2694, Thermo Fisher Scientific), 1g/ml Hoechst (H33342, Thermo Fisher Scientific). Nuclei concentration was determined using the Countess II FL Automated Cell Counter and optimal nuclei concentration was obtained by adding additional cold resuspension buffer, if needed. Subsequently, samples were processed using the 10x Chromium device with the 10x Genomics scRNA-Seq protocol v2 to generate cell and gel bead emulsions, followed by reverse transcription, cDNA amplification, and sequencing library preparation following the

manufacturers' instructions. Libraries have subsequently been sequenced one sample per lane on HiSeq4000 (Illumina; paired-end 26 x 74 bp)

Dissociation and Cryopreservation of ALI culture cells

Single cell suspensions were generated by adding 1x DPBS (Thermo, 14190094) to the apical chamber, followed by an incubation at 37°C for 15 minutes and subsequent washing of the cells with 1x DPBS (three times) to remove excess mucus. Adherent cells were dissociated via 0.25% Trypsin-EDTA (ThermoFisher Scientific, 25200056) treatment for 5 minutes at 37°C, followed by Dispase treatment (Corning, 354235) for 10 minutes at 37°C. Trypsinization was inactivated by adding cell culture medium with soybean trypsin inhibitor. Cells were spun down at 300x g for 5 minutes, resuspended in 1x DPBS, and passed through a 20 µm cell strainer (pluriSelect, 41-50000-03) to obtain a single cell solution and to remove cell clumps. Cells were spun at 300x g for 5 minutes and resuspended in cryopreservation medium (10% DMSO, 10% FBS, and 80% Culture Medium (PneumaCult-ALI) and frozen down gradually before shipping on dry-ice.

Single cell RNA sequencing library preparation (ALI cultures)

Cryopreserved cells were thawed at 37°C, spun down at 300xG for 5 minutes, the cell pellet was resuspended in 1x PBS with 0.05% BSA (Sigma-Aldrich, 05479), and passed through a 35 µm filter (Corning, 352235) to remove cell debris. Single cell suspensions were loaded onto the 10x Chromium device using the 10x Genomics Single Cell 3' Library Kit v2 (10x Genomics; PN-120237, PN-120236, PN-120262) to generate cell and gel bead emulsions, followed by reverse transcription, cDNA amplification, and sequencing library preparation following the manufacturers' instructions. The resulting libraries were sequenced with one sample per lane using the NextSeq500 (Illumina; high-output mode, paired-end 26 x 49 bp) or with two samples per lane using the HiSeq4000 (Illumina; paired-end 26 x 74 bp).

Pre-processing and data analysis

CellRanger software version 2.1.1 (10x Genomics) was used for processing of the raw sequencing data and the transcripts were aligned to the 10x reference human genome hg19 1.2.0. Low-quality cells were removed during pre-processing using Seurat version 3.0.0 (https://github.com/satijalab/seurat) based on the following

criteria: (a) >200 or, depending on the sample, <6000 – 9000 genes (surgical lung tissues)/ <3000 – 5000 genes (ALI cultures), (b) <15% mitochondrial reads (Supp. Fig. 7). The remaining data were further processed using Seurat for log-normalization, scaling, merging, clustering, and gene expression analysis. Afterwards, all control samples were merged using the “FindIntegrationAnchors” and “IntegrateData” functions with their results being merged again and used for downstream analysis. Monocle3 [Cao et al. (2019)] was used infer cellular trajectories and dynamics. Dimensionality reduction, cell clustering, trajectory graph learning, and pseudotime measurement were performed with this tool.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses

Statistical analyses were performed using R and Python 3.7.1 with scipy 0.14.1 and statsmodels 0.9.0. For comparisons between CPM values, the Wilcoxon rank sum test was used. To calculate the p-value for the overlap between sets, a hypergeometric test was employed. Expected molecule counts per cell were modelled using a binomial distribution and the observed detection probability. Gene set enrichment analysis were performed using Metascape Zhou et al. (2019)] on the KEGG, Canonical pathways, GO, Reactome, and Corum databases. P-values were adjusted for multiple testing using the Benjamini-Hochberg method Benjamini and Hochberg (1995)]. Boxes in box plots indicate the first and third quartile, with the median shown as horizontal lines. Whiskers extend to 1.5 times the inter-quartile-range. All individual data points are indicated on the plot.

SUPPLEMENTAL INFORMATION

Supp. Figure 1 – Exemplary characterization of an air liquid interface (ALI) culture derived from HBECs. Filters of an ALI culture were stained for the indicated bronchial epithelial markers and analyzed using confocal microscopy.

Supp. Figure 2 – Cell type identification by marker genes. DotPlots indicate the percent expressing cells (size) and expression level (color) for selected marker genes for each cell type in HBECs.

Supp. Figure 3 – Percentage of *ACE2*$^{+}$/*TMPRSS2*$^{+}$/*FURIN*$^{+}$ positive cells. Percentage of positive cells (defined here as having at least one read of the respective gene) for *ACE2*, *TMPRSS2*, and *FURIN* in primary lung and HBECs.

Supp. Figure 4 – Quantification of ACE2+/TMPRSS2+ double positive cells. Overlaps and enrichment statistics for all ACE2 and TMPRSS2 single and double positive cells in the primary lung and HBEC dataset. RF: representation factor, enrichment. P: hypergeometric tail probability.

Supp. Figure 5 – Expression of DPP4 in primary lung. Expression levels and percentage of positive cells for *DPP4* in the primary lung dataset.

Supp. Figure 6 – Age-dependent cell type composition. Cell type composition in the primary lung and HBEC dataset (age is color-coded).

Supp. Figure 7 - single-cell sequencing quality control. Numbers of unique molecular identifiers (UMIs, top row) and genes per cell are shown for each individual sample from primary lung tissue and HBECs.

REFERENCES

Backer, J.A., Klinkenberg, D., and Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. Euro Surveill *25*.

Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N.*, et al.* (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One *13*, e0209648.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society *57*, 289-300.

Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T.*, et al.* (2020). Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv *https://www.medrxiv.org/content/10.1101/2020.03.03.20028423v1*.

Brussow, H. (2020). The Novel Coronavirus - A Snapshot of Current Knowledge. Microb Biotechnol.

Burkard, C., Verheije, M.H., Wicht, O., van Kasteren, S.I., van Kuppeveld, F.J., Haagmans, B.L., Pelkmans, L., Rottier, P.J., Bosch, B.J., and de Haan, C.A. (2014). Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog *10*, e1004502.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J.*, et al.* (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496-502.

Chan, J.F., Yuan, S., Kok, K.H., To, K.K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C., Poon, R.W.*, et al.* (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet *395*, 514-523.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y.*, et al.* (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet *395*, 507-513.

Coronaviridae Study Group of the International Committee on Taxonomy of, V. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., and Decroly, E. (2020). 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.

de Wit, E., van Doremalen, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol *14*, 523-534.

Dumm, R.E., Fiege, J.K., Waring, B.M., Kuo, C.T., Langlois, R.A., and Heaton, N.S. (2019). Non-lytic clearance of influenza B virus from infected cells preserves epithelial barrier function. Nat Commun *10*, 779.

Gallagher, T.M., and Buchmeier, M.J. (2001). Coronavirus spike proteins in viral entry and pathogenesis. Virology *279*, 371-374.

Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C.*, et al.* (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med.

Hamming, I., Timens, W., Bulthuis, M.L., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol *203*, 631-637.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A.*, et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X.*, et al.* (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497-506.

Jia, H.P., Look, D.C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., and McCray, P.B., Jr. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. J Virol *79*, 14614-14621.

Kahn, N., Kuner, R., Eberhardt, R., Meister, M., Muley, T., Winteroll, S., Schnabel, P.A., Ishizaka, A., Herth, F.J., Poustka, A.*, et al.* (2009). Gene expression analysis of endobronchial

epithelial lining fluid in the evaluation of indeterminate pulmonary nodules. J Thorac Cardiovasc Surg *138*, 474-479.

Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., and Matsuyama, S. (2012). Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. J Virol *86*, 6537-6545.

Kido, H., Okumura, Y., Takahashi, E., Pan, H.Y., Wang, S., Yao, D., Yao, M., Chida, J., and Yano, M. (2012). Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure. Biochim Biophys Acta *1824*, 186-194.

Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M., Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., and Ward, A.B. (2016). Pre-fusion structure of a human coronavirus spike protein. Nature *531*, 118-121.

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., and Hsueh, P.R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents, 105924.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C.*, et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450-454.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N.*, et al.* (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565-574.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol *15*, e8746.

Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M., and Taguchi, F. (2005). Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. Proc Natl Acad Sci U S A *102*, 12543-12547.

Millet, J.K., and Whittaker, G.R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci U S A *111*, 15214-15219.

Muley, T.R., Herth, F.J., Schnabel, P.A., Dienemann, H., and Meister, M. (2012). From tissue to molecular phenotyping: pre-analytical requirements heidelberg experience. Transl Lung Cancer Res *1*, 111-121.

Ren, X., Glende, J., Al-Falah, M., de Vries, V., Schwegmann-Wessels, C., Qu, X., Tan, L., Tschernig, T., Deng, H., Naim, H.Y.*, et al.* (2006). Analysis of ACE2 in polarized epithelial cells: surface expression and function as receptor for severe acute respiratory syndrome-associated coronavirus. J Gen Virol *87*, 1691-1695.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W.*, et al.* (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med *382*, 970-971.

Simmons, G., Reeves, J.D., Rennekamp, A.J., Amberg, S.M., Piefer, A.J., and Bates, P. (2004). Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. Proc Natl Acad Sci U S A *101*, 4240-4245.

Simmons, G., Zmora, P., Gierer, S., Heurich, A., and Pohlmann, S. (2013). Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. Antiviral Res *100*, 605-614.

Sims, A.C., Burkett, S.E., Yount, B., and Pickles, R.J. (2008). SARS-CoV replication and pathogenesis in an in vitro model of the human conducting airway epithelium. Virus Res *133*, 33-44.

Tosti, L., Hang, Y., Trefzer, T., Steiger, K., Ten, F.W., Lukassen, S., Ballke, S., Kuehl, A.A., Spieckermann, S., Bottino, R.*, et al.* (2019). Single nucleus RNA sequencing maps acinar cell states in a human pancreas cell atlas. bioRxiv.

Van den Broeke, C., Jacob, T., and Favoreel, H.W. (2014). Rho'ing in and out of cells: viral interactions with Rho GTPase signaling. Small GTPases *5*, e28318.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell *journal pre-proof*.

Walls, A.C., Tortorici, M.A., Bosch, B.J., Frenz, B., Rottier, P.J.M., DiMaio, F., Rey, F.A., and Veesler, D. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature *531*, 114-117.

Wang, C., Horby, P.W., Hayden, F.G., and Gao, G.F. (2020a). A novel coronavirus outbreak of global health concern. Lancet *395*, 470-473.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y.*, et al.* (2020b). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA.

Wang, H., Yang, P., Liu, K., Guo, F., Zhang, Y., Zhang, G., and Jiang, C. (2008). SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. Cell Res *18*, 290-301.

White, J.M., and Whittaker, G.R. (2016). Fusion of Enveloped Viruses in Endosomes. Traffic *17*, 593-614.

WHO (2004). Summary of probably SARS cases with onset of illness from 1 November 2002 to 31 July 2003. . http://wwwwhoint/csr/sars/country/table2004_04_21/en/

WHO (2016). Middle East respiratory syndrome coronavirus (MERS-CoV) – Saudi Arabia. http://wwwwhoint/csr/don/26-april-2016-mers-saudi-arabia/en/

Wrapp, D., Wang, N.h.s.g.o.t.n.c.-n.c.a.f.-l.c.s.a.i.C.o.t.s.c., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., *et al.* (2020a). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe.

Wu, C., Yang, Y., Liu, Y., Zhang, P., Wang, Y., Wang, Q., Xu, Y., Li, M., Zheng, M., Chen, L., *et al.* (2020b). Furin, a potential therapeutic target for COVID-19. chinaRxiv.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., *et al.* (2020c). A new coronavirus associated with human respiratory disease in China. Nature.

Wu, Z., and McGoogan, J.M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA.

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., Li, T., and Chen, Q. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci *12*, 8.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science.

Zhang, J.J., Dong, X., Cao, Y.Y., Yuan, Y.D., Yang, Y.B., Yan, Y.Q., Akdis, C.A., and Gao, Y.D. (2020a). Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy.

Zhang, Y.-Z., Koopmans, M., and Yuen, K.Y. (2020b). The Novel Coronavirus Outbreak: What We Know and What We Don't. Cell *https://doi.org/10.1016/j.cell.2020.02.027*.

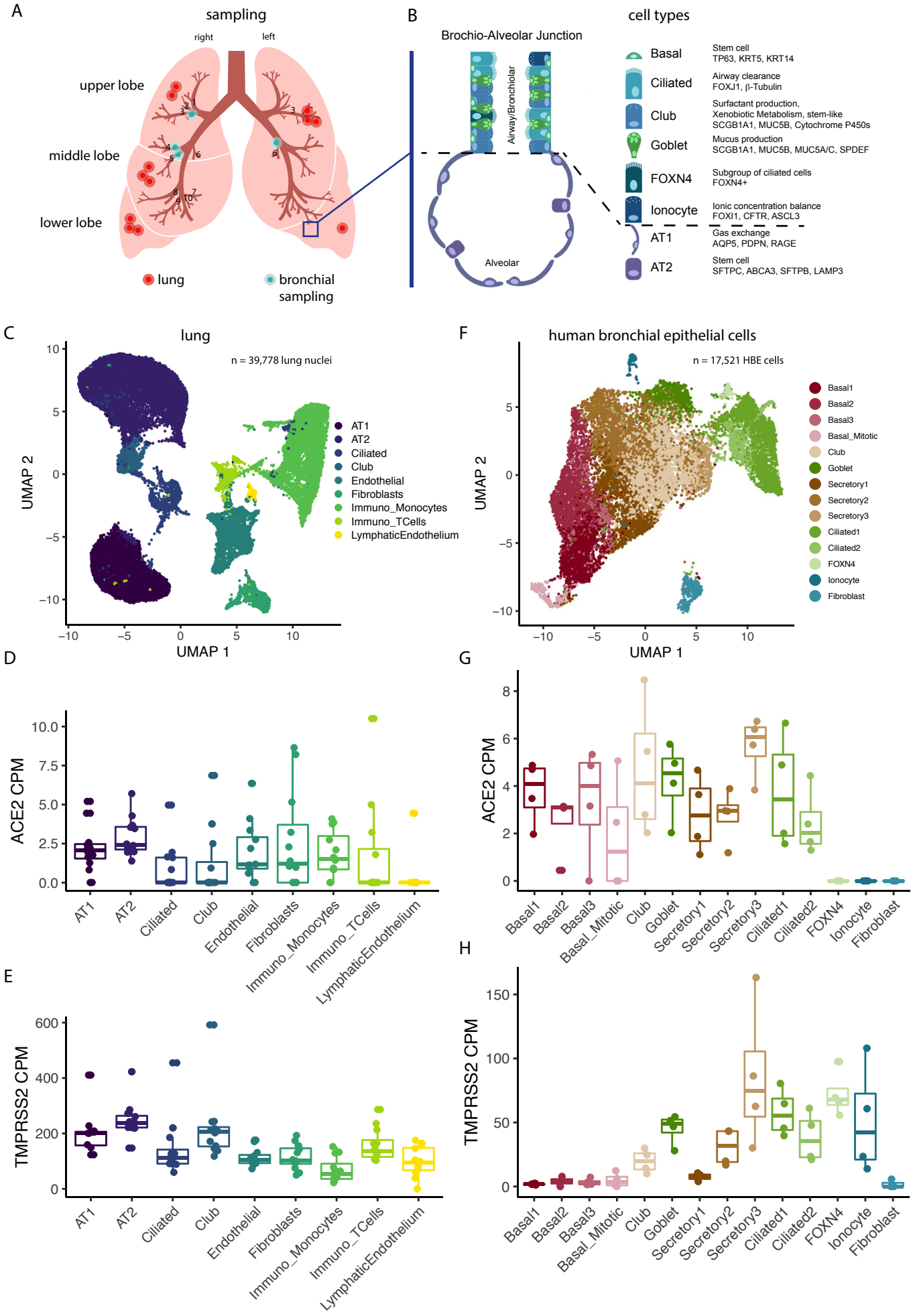
Zhao, Y., Zaho, Z., Wand, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. BioRxiv.

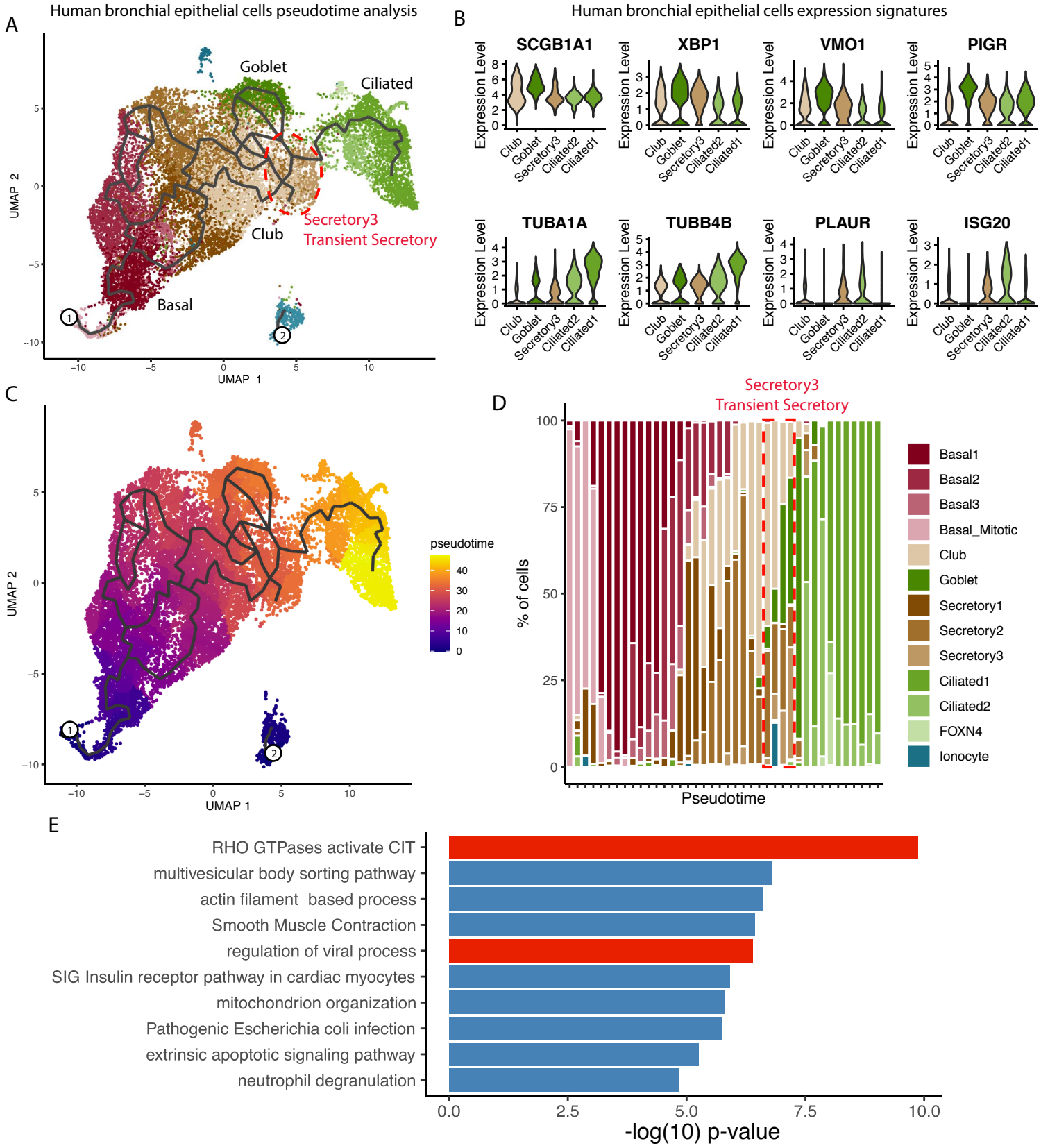
Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L.*, et al.* (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature.

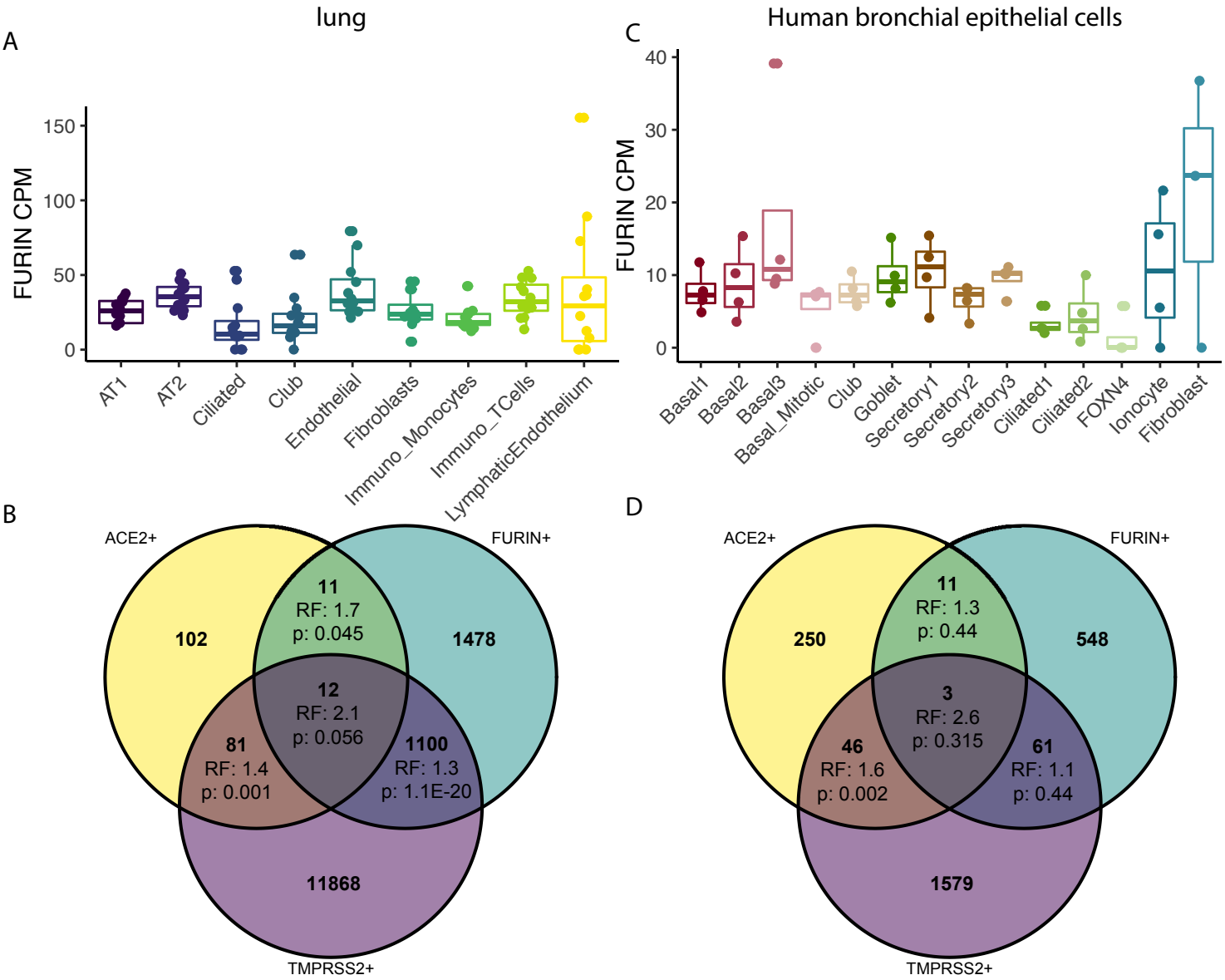
Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun *10*, 1523.

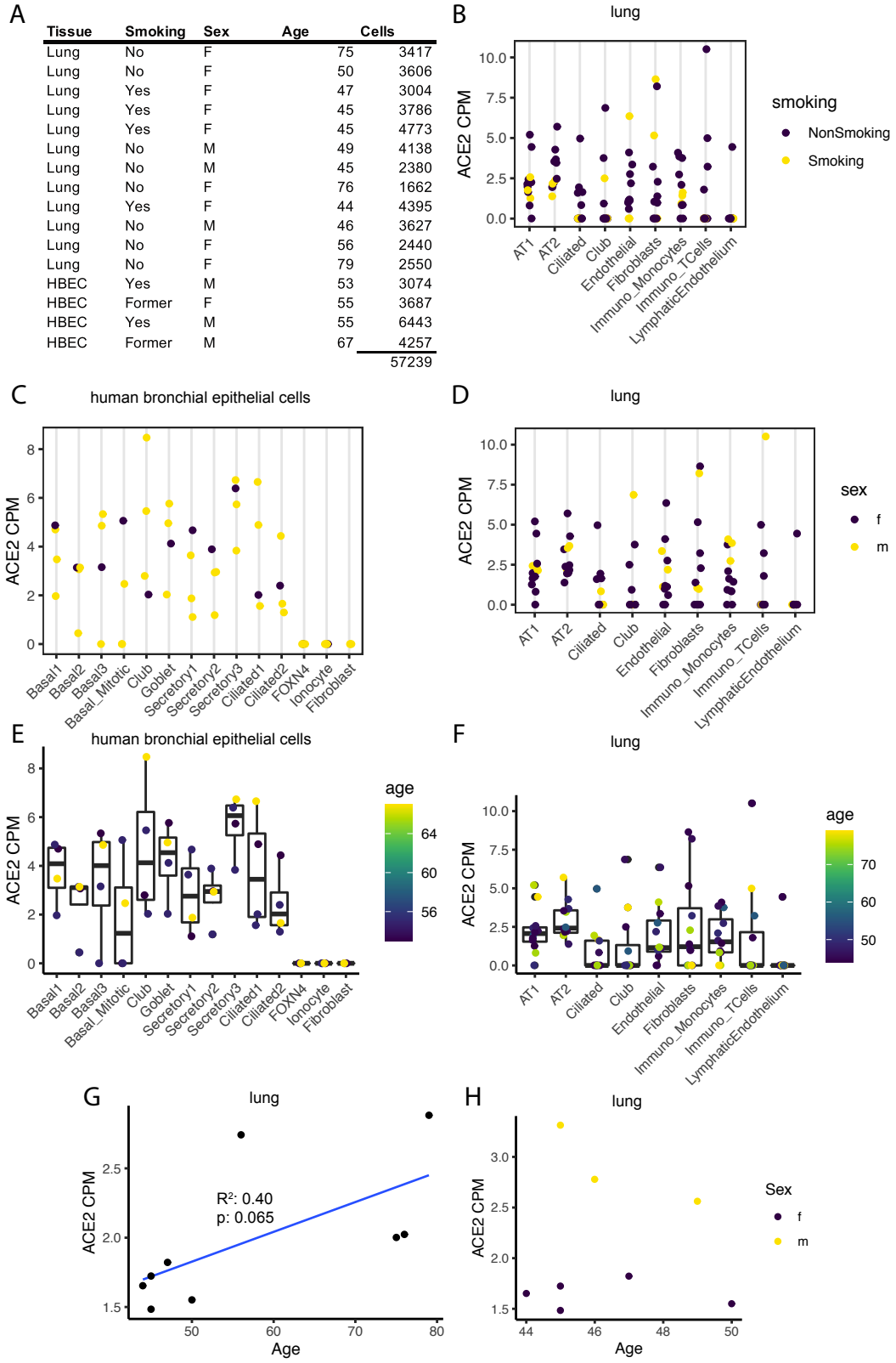
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R.*, et al.* (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med *382*, 727-733.

Zou, X., Chen, K., Zou, J., Han, P., Hao, J., and Han, Z. (2020). The single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to Wuhan 2019-nCoV infection. Frontiers of Medicine *Epub ahead of print*.


A
sampling
right
left
upper lobe
middle lobe
lower lobe
lung
bronchial sampling
B
cell types
Brochio-Alveolar Junction
Airway/Bronchiolar
Alveolar
Basal
Stem cell
TP63, KRT5, KRT14
Ciliated
Airway clearance
FOXJ1, β-Tubulin
Club
Surfactant production,
Xenobiotic Metabolism, stem-like
SCGB1A1, MUC5B, Cytochrome P450s
Goblet
Mucus production
SCGB1A1, MUC5B, MUC5A/C, SPDEF
FOXN4
Subgroup of ciliated cells
FOXN4+
Ionocyte
Ionic concentration balance
FOXI1, CFTR, ASCL3
AT1
Gas exchange
AQP5, PDPN, RAGE
AT2
Stem cell
SFTPC, ABCA3, SFTPB, LAMP3
C
lung
n = 39,778 lung nuclei
UMAP 2
UMAP 1
AT1
AT2
Ciliated
Club
Endothelial
Fibroblasts
Immuno_Monocytes
Immuno_TCells
LymphaticEndothelium
F
human bronchial epithelial cells
n = 17,521 HBE cells
UMAP 2
UMAP 1
Basal1
Basal2
Basal3
Basal_Mitotic
Club
Goblet
Secretory1
Secretory2
Secretory3
Ciliated1
Ciliated2
FOXN4
Ionocyte
Fibroblast
D
ACE2 CPM
AT1
AT2
Ciliated
Club
Endothelial
Fibroblasts
Immuno_Monocytes
Immuno_TCells
LymphaticEndothelium
G
ACE2 CPM
Basal1
Basal2
Basal3
Basal_Mitotic
Club
Goblet
Secretory1
Secretory2
Secretory3
Ciliated1
Ciliated2
FOXN4
Ionocyte
Fibroblast
E
TMPRSS2 CPM
AT1
AT2
Ciliated
Club
Endothelial
Fibroblasts
Immuno_Monocytes
Immuno_TCells
LymphaticEndothelium
H
TMPRSS2 CPM
Basal1
Basal2
Basal3
Basal_Mitotic
Club
Goblet
Secretory1
Secretory2
Secretory3
Ciliated1
Ciliated2
FOXN4
Ionocyte
Fibroblast


A
Human bronchial epithelial cells pseudotime analysis
Goblet
Ciliated
Secretory3
Transient Secretory
Club
Basal
UMAP 2
UMAP 1
B
Human bronchial epithelial cells expression signatures
SCGB1A1
XBP1
VMO1
PIGR
TUBA1A
TUBB4B
PLAUR
ISG20
Expression Level
Club
Goblet
Secretory3
Ciliated2
Ciliated1
C
pseudotime
D
Secretory3
Transient Secretory
% of cells
Pseudotime
Basal1
Basal2
Basal3
Basal_Mitotic
Club
Goblet
Secretory1
Secretory2
Secretory3
Ciliated1
Ciliated2
FOXN4
Ionocyte
E
RHO GTPases activate CIT
multivesicular body sorting pathway
actin filament based process
Smooth Muscle Contraction
regulation of viral process
SIG Insulin receptor pathway in cardiac myocytes
mitochondrion organization
Pathogenic Escherichia coli infection
extrinsic apoptotic signaling pathway
neutrophil degranulation
-log(10) p-value


A
lung
FURIN CPM
150
100
50
0
AT1
AT2
Ciliated
Club
Endothelial
Fibroblasts
Immuno_Monocytes
Immuno_TCells
LymphaticEndothelium
C
Human bronchial epithelial cells
FURIN CPM
40
30
20
10
0
Basal1
Basal2
Basal3
Basal_Mitotic
Club
Goblet
Secretory1
Secretory2
Secretory3
Ciliated1
Ciliated2
FOXN4
Ionocyte
Fibroblast
B
ACE2+
FURIN+
TMPRSS2+
102
11
RF: 1.7
p: 0.045
1478
12
RF: 2.1
p: 0.056
81
RF: 1.4
p: 0.001
1100
RF: 1.3
p: 1.1E-20
11868
D
ACE2+
FURIN+
TMPRSS2+
250
11
RF: 1.3
p: 0.44
548
3
RF: 2.6
p: 0.315
46
RF: 1.6
p: 0.002
61
RF: 1.1
p: 0.44
1579

A

Tissue	Smoking	Sex	Age	Cells
Lung	No	F	75	3417
Lung	No	F	50	3606
Lung	Yes	F	47	3004
Lung	Yes	F	45	3786
Lung	Yes	F	45	4773
Lung	No	M	49	4138
Lung	No	M	45	2380
Lung	No	F	76	1662
Lung	Yes	F	44	4395
Lung	No	M	46	3627
Lung	No	F	56	2440
Lung	No	F	79	2550
HBEC	Yes	M	53	3074
HBEC	Former	F	55	3687
HBEC	Yes	M	55	6443
HBEC	Former	M	67	4257
				57239