Expression of the SARS-CoV-2 *ACE2* Receptor in the Human Airway Epithelium

Haijun Zhang, MD, PhD[1]*, Mahboubeh R. Rostami, Phd[1]*, Philip L. Leopold, PhD[1], Jason G. Mezey, PhD[1,2], Sarah L. O'Beirne, MD, PhD[1], Yael Strulovici-Barel, PhD[1‡], and Ronald G. Crystal, MD[1‡]

[1]Department of Genetic Medicine,

Weill Cornell Medical College

New York, NY

and

[2]Department of Biological Statistics and Computational Biology

Cornell University

Ithaca, NY

* These authors contributed equally

‡ YSB and RGC contributed equally as senior investigators for this study

Author contributions: HZ – data analysis, manuscript writing; MRR - data analysis, manuscript writing; PLL - data analysis, manuscript writing; JGM – statistical analysis; SLO'B – subject recruitment and phenotyping, bronchoscopy, manuscript writing; YSB - data analysis, manuscript writing; RGC - data analysis, manuscript writing

Descriptor: 10.9 Pathogen/Host Cell Interactions

Online Data Supplement: This article has an online data supplement, which is accessible from this issue's table of content online at **www.atsjournals.org**"

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Running head: *ACE2* expression in airway epithelium

Correspondence: Ronald G. Crystal, MD

Department of Genetic Medicine

Weill Cornell Medical College

1300 York Avenue, Box 164

New York, New York 10065

Phone: (646) 962-4363

Fax: (646) 962-0220

E-mail: geneticmedicine@med.cornell.edu

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

At a Glance Commentary

Scientific Knowledge on the Subject

COVID-19, a viral disease with severe respiratory morbidity, is caused by the SARS-CoV-2 coronavirus. The major tropism determinant for SARS-CoV-2 virus is availability of angiontensin converting enzyme 2 (ACE2), the primary viral receptor expressed on the surface of cells. The high level of contagion argues that inhalation of airborne virus-containing droplets is a major route of exposure, so an understanding of COVID-19 may benefit from characterization of ACE2 expression in the airway.

What This Study Adds to the Field

We report ACE2 gene expression in the small airway, large airway, and trachea using microarray, bulk RNA seq, miRNA, and single cell RNA sequencing data sets. Broad expression of ACE2 was found throughout the airway with higher expression in proximal segments. In addition, all major epithelial cell types expressed ACE2. Smoking was associated with higher ACE2 mRNA expression in the small airway. Male smokers had the highest ACE2 expression levels, potentially providing a partial explanation for elevated COVID-19 incidence among men compared with women. ACE2 expression might be influenced by low miR-1246 expression in smokers. These data may provide insight into the pathogenesis of COVID-19 and risk factors in the population.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Abstract

Rationale: Infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus disease (COVID-19), a predominantly respiratory illness. The first step in SARS-CoV-2 infection is binding of the virus to angiotensin converting enzyme 2 (*ACE2*) on the airway epithelium.

Objectives: The objective was to gain insight into the expression of *ACE2* in the human airway epithelium.

Methods: Airway epithelium sampled by fiberoptic bronchoscopy of trachea, large airway epithelium (LAE) and small airway epithelium (SAE) of nonsmokers and smokers was analyzed for expression of *ACE2* and other coronavirus infection-related genes using microarray, RNA-seq and 10x single cell transcriptome analysis, with associated examination of *ACE2*-related miRNA.

Measurements and Main Results: (1) *ACE2* is expressed similarly in the trachea and LAE with lower expression in the SAE; (2) in the SAE, *ACE2* is expressed in basal, intermediate, club, mucus and ciliated cells; (3) *ACE2* is up-regulated in the SAE by smoking, significantly in males; (4) levels of miR-1246 expression could play a role in *ACE2* up-regulation in the SAE of smokers; and (5) *ACE2* is expressed in airway epithelium differentiated *in vitro* on air-liquid interface cultures from primary airway basal stem/progenitor cells; this can be replicated using LAE and SAE immortalized basal cell lines derived from healthy nonsmokers.

Conclusions: *ACE2*, the gene encoding the receptor for SARS-CoV-2, is expressed in the human airway epithelium, with variations in expression relevant to the biology of initial steps in SARS-

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

CoV-2 infection.

Abstract word count: 236

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for coronavirus disease 2019 (COVID-19), a global pandemic characterized by fever, dry cough, dyspnea, lymphopenia and a significant mortality rate, primarily due to respiratory complications (1-6). The disease is spread primarily through person-to-person transmission *via* respiratory droplets and by contact with contaminated surfaces (4, 5, 7, 8). Approximately 50% of hospitalized COVID-19 patients have pre-existing medical conditions, including diabetes, cardiovascular disease, chronic obstructive pulmonary disease (COPD) and malignancy (1, 2, 5, 9, 10) and males have increased susceptibility to infection, more severe disease and higher mortality (1-3, 5, 9).

The SARS-CoV-2 virus is a novel coronavirus distinct from the coronaviruses causing human severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) (7, 11-13). Like other closely related coronaviruses, SARS-CoV-2 interacts with cells through the virus spike protein, an envelope glycoprotein which binds to the host cell receptor, angiotensin converting enzyme 2 (*ACE2*), and mediates viral entry (14-17). Following binding, lineage B coronavirus spike proteins are modified by one or more cellular proteases, either at the cell surface or following endocytosis. At the cell surface, the transmembrane serine protease, *TMPRSS2*, can cleave the SARS-CoV-2 spike protein, leading to exposure of a fusion peptide that can guide direct fusion of the coronavirus envelope with the plasma membrane of the target cell, as observed in other coronaviruses (14, 16). Alternatively, either cathepsin L or furin, intracellular proteases, can activate spike-mediated coronavirus envelope fusion with internal cellular membranes (14, 18). Secondary activation of the SARS-CoV-2 by either cathepsin L or furin, both widely expressed in airway epithelium, has not yet been demonstrated, but the

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

sequence of the SARS-CoV-2 spike protein contains two furin cleavage sites (15). Coronavirus fusion is also affected by the activity of an enzyme, phosphatidylinositol 4-kinase IIIβ (*PI4KB*), a gene expressed in the airway epithelium (19). *PI4KB* phosphorylates phosphatidylinositol in a pathway leading to generation of inositol triphosphate, an intracellular signaling molecule. Both pharmacological inhibition of *PI4KB* activity and siRNA-mediated knockdown of *PI4KB* inhibited infection of cells *in vitro* with a SARS-CoV spike protein pseudotyped virus (19).

Based on the knowledge that SARS-CoV-2 infection is primarily a respiratory illness, SARS-CoV-2 has been isolated from respiratory epithelial lining fluid, and SARS-CoV-2 infects human airway epithelium (4, 7, 12), it is highly likely that the cells mediating entry of SARS-CoV-2 in the majority of cases can be found in the respiratory epithelium. In this context, we searched our extensive airway epithelial transcriptome data of healthy nonsmokers and smokers for evidence of expression of *ACE2*, with a focus on the extent of expression, which cell types express the receptor, whether sex and/or cigarette smoking influence *ACE2* expression in airway epithelium, and biologic processes in human airway epithelium that may be linked to *ACE2* expression. Finally, we observed that the immortalized BCi-NS1.1 cell line, an immortalized airway basal cell (BC) line derived from BC collected from the large airway epithelium (LAE) of a healthy nonsmoker (20) and the hSABCi-NS1.1 cell line, an immortalized airway BC line derived from BC collected from the small airway epithelium (SAE) of a healthy nonsmoker (21), both express *ACE2* and, when cultured on air-liquid interface (ALI), the differentiated progeny express *ACE2*. In the context that the airway epithelium is a likely entry site for SARS-CoV-2, these cell lines should be useful investigative tools for studying SARS-CoV-2 interaction with the human airway epithelium and assessing therapeutic agents to treat the infection.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Methods

The assessment of expression of *ACE2* and associated genes in the airway epithelium of nonsmokers and smokers was derived from multiple databases of our laboratory's assessment of the transcriptome of nonsmokers and smokers representing 744 independent samples of airway epithelium derived from 267 subjects. We obtained airway epithelium by fiberoptic bronchoscopy and brushing of trachea, large airway (1-5 generations, brushing typically at 3-4 generations) and small airway (6-23 generations, brushing typically at 10-12 generations) (22-24) from phenotypically normal nonsmokers and smokers. Details regarding inclusion/exclusion criteria for nonsmokers and smokers are presented in Supplemental Methods, as are the details regarding sample processing and analysis. Expression was assessed by microarray, RNA-seq or 10x single cell analysis (23-27) and comparisons were made between nonsmokers and smokers; p value <0.05 was considered significant. The majority of the *ACE2* transcriptome data is from our previously published datasets. The study population, samples, and transcriptome quantification methodology for each table and figure are detailed in Table E1, along with references and GEO accession numbers in the public repository of data in the Gene Expression Omnibus website, https://www.ncbi.nlm.nih.gov/geo/, if relevant. If the dataset is new, it is listed in Table E1 as "unpublished." The quantification of SAE microRNAs that could bind to *ACE2* mRNA is based on our publication by Wang *et al* (27). In addition to the published databases we used: (1) microarrays (Affymetrix HG-U133 Plus 2.0, Affymetrix, Santa Clara, CA) to assess the expression of *ACE2* on a well-differentiated airway epithelium cultured at ALI using primary trachea BC from healthy nonsmokers; (2) RNA-seq (Illumina HiSeq 2500, San Diego, CA) to assess *ACE2* expression in 2 immortalized BCi-NS1.1 (large airway epithelium) and hSABCi-NS1.1 (small airway epithelium) cell lines (20, 21); and (3) single cell RNA sequencing (10x

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Genomics, Pleasanton, CA) to analyze the expression of *ACE2* in SAE from 5 nonsmokers and 5 smokers. In addition to expression of *ACE2*, we assessed the expression of genes that have been identified to participate in the initial steps of other similar coronaviruses, with the likelihood that some of these genes participate in the early events of SARS-CoV-2 infection. Finally, we assessed the data of O'Beirne *et al* (25) for effects, if any, of New York City pollution levels over time on small airway epithelium *ACE2* levels.

Results

Expression of *ACE2* in Normal Airway Epithelium

Analysis of trachea, LAE and SAE demonstrated *ACE2* is expressed in all regions of the tracheobronchial tree of healthy nonsmokers, with higher expression in the trachea and LAE than in the SAE (Affymetrix HG-U133 Plus 2.0 microarray; Figure 1A). Assessment of primary trachea basal stem/progenitor cells of healthy nonsmokers differentiating on ALI showed that BC express *ACE2*, as do, to a greater extent, the differentiated progeny of the BC (Affymetrix HG-U133 Plus 2.0 microarray; Figure 1B). Single cell transcriptome analysis identified all the major cell types of the SAE of healthy nonsmokers, including basal, intermediate, club, mucus, and ciliated cells, as well as ionocytes, macrophages, T cells, and mast cells (10x, Figure E1). *ACE2* expression was noted mainly in epithelial cells including basal, intermediate, club, mucus, and ciliated cells (Figure 1C). The low % of positive cells is partially a consequence of the technology that samples a fraction of the transcripts in a given cell (28). As a result, low abundance transcripts are not detected in every cell that expresses the gene. Of interest, the single cell data of Reyfman *et al* (29) derived from the lung parenchyma of individuals without lung disease demonstrated that alveolar type 2 cells and other epithelial cells express *ACE2*

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

(Figure E2A). The similar percentages of epithelial cells that were observed to express *ACE2* in the Reyfman study (29) compared with this study support the observation that *ACE2* is expressed in a broad array of epithelial cells. Similarly, an analysis of data from Duclos et al. (30), who obtained bronchial epithelium via bronchoscopic brushing in a similar manner to this study, reveals the presence of *ACE2* in cells in a variety of epithelial cells (Figure E2B). Of note, *ACE2* was detected in a higher percentage of mucus cells than other epithelial cells in the Duclos study (30), possibly indicating a difference between the large airway bronchial epithelia analyzed in that study compared with the small airway epithelia analyzed in this study. From this data, we conclude that the *ACE2* receptor for SARS-CoV-2 is distributed throughout the lung epithelial surface, and the site of infection would be dictated by the size of the inhaled droplet and respiration parameters (31).

We also analyzed cells recovered by bronchoalveolar lavage (BAL) for *ACE2* expression. Consistent with our analysis of the data from Reyfman *et al* (29) (Figure E2A), data obtained using microarray, RNA-seq and 10x single cell transcriptome analysis indicated that *ACE2* expression was undetectable or rarely detected in alveolar macrophages, T cells, B cells or dendritic cells (data not shown).

Smoking and Sex Influence on *ACE2* Expression

Based on the clinical data that SARS-CoV-2 lung infection is characterized in chest imaging as distal infiltrates (32), it is likely that the SAE is an important site of SARS-CoV-2 binding. *ACE2* expression was higher in the SAE of smokers than nonsmokers (Affymetrix HG-U133 Plus 2.0 microarray, $p<10^{-5}$, Figure 2A). When nonsmokers and smokers were divided by sex, male smokers exhibited significantly higher *ACE2* expression levels than female smokers or

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

nonsmokers of either sex ($p<0.005$, all comparisons, Figure 2B). RNA-seq analysis of a different dataset confirmed that SAE *ACE2* levels were higher in smokers than nonsmokers (Illumina HiSeq 2500, Figure 2C), with the *ACE2* levels in SAE of male smokers higher than male nonsmokers (Figure 2D; a low n=3 in females obviated comparison of smokers *vs* nonsmokers *ACE2* levels in females). Analysis of *ACE2* levels in the LAE showed no differences relevant to sex or smoking, except that *ACE2* levels were significantly higher in male *vs* female nonsmokers when assessed by Affymetrix HG-U133 Plus 2.0 microarray (Figure E3).

Pollution Effects on *ACE2* Expression

Correlations between ambient pollution levels and COVID-19 cases have been reported (33). To assess if low levels of air pollution might affect SAE *ACE2* expression, we analyzed the dataset of O'Beirne *et al* (25), a study carried out to evaluate the relationship between gene expression in the SAE of healthy nonsmokers and smokers in New York City with average monthly pollution levels ($PM_{2.5}$) as reported by the United States Environmental Protection Agency (EPA). While no differences were observed between SAE *ACE2* expression and $PM_{2.5}$ levels (nonsmokers and smokers combined or separately; Figure E4), the peak 30-day mean $PM_{2.5}$ levels in New York City during this study was 18 $\mu g/m^3$, a level considered "safe" by the EPA (25).

Possible Influences of microRNA Expression on *ACE2*

Our database was assessed for microRNAs homologous to the 3' end of the *ACE2* gene that are significantly modulated by smoking. Of interest, miR-1246 has homology to *ACE2*, and miR-1246 is down-regulated in SAE of smokers compared to nonsmokers (27) (Affymetrix miRNA 2.0 arrays, Figure 3). We have insufficient data to determine if this smoking-related

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

decrease in miR-1246 plays a role in *ACE2* expression in the SAE, but this is a possible mechanism to be assessed in future studies.

Expression of Genes Related to the Initial Steps of Coronavirus Infection

In addition to evaluating the expression of *ACE2*, the gene encoding the primary SARS-CoV-2 receptor, we assessed the SAE mRNA expression levels of other cellular proteins reported to be related to the early steps in infection pathway based on lineage B coronaviruses, and thus possibly relevant to airway epithelial infection by SARS-CoV-2.

Cell surface *ACE2* levels can be regulated by *ADAM10* and *ADAM17*, cell surface disintegrins that mediate shedding of *ACE2* from the cell surface (34). Both *ADAM10* and *ADAM17* are expressed in the SAE (Figures 4A, B). *TMPRSS2*, *TMPRSS11A* and *TMPRSS11D* are proteases that cleave the SARS-CoV spike protein at the cell surface to facilitate fusion of the coronavirus envelope with the cell membrane, a critical step in transfer of the nucleocapsid to the cytosol (16, 35). All 3 enzymes are expressed in SAE (Figures 4C-E). Interestingly, expression of *TMPRSS2* is up-regulated in the SAE of smokers *vs* nonsmokers (Figure 4C). Furin and cathepsin L, two proteases encoded by the genes *FURIN* and *CTSL* are found in the endo-lysosomal pathway, and can cleave the coronavirus spike protein leading to intracellular fusion of the envelope with the organelle membrane relevant to infection by SARS and other coronaviruses (18, 36, 37). Both proteins are expressed in the SAE (Figures4F, G). Finally, inhibition of *PIK4B*, an enzyme that phosphorylates phosphatidylinositol, results in inhibition of SARS-CoV infection (19). *PIK4B* is also expressed in SAE (Figure 4H). Other than the up-regulation of *TMPRSS2* in smokers, there were no other smoking-related changes in SAE expression among the genes related to coronavirus infection. Separately, an analysis of single

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

cell RNA sequencing data showed that coronavirus infection-related genes were broadly expressed in airway epithelium, including basal, intermediate, club, mucus, and ciliated cells in both healthy nonsmokers and smokers (Figure 5). As reported above, *ACE2* expression was observed in all of the epithelial cell types; however, in contrast to data from RNA sequencing and microarrays, *ACE2* did not exhibit a smoking-dependent increase in gene expression in the single cell transcriptome data (Figure 5A). The absence of an observed difference in *ACE2* expression among nonsmokers and smokers in the single cell RNA-seq dataset compared with the bulk RNA dataset and microarray dataset likely reflects the technical details inherent in the three types of analysis. Single cell RNA-seq data is derived from a smaller number of individuals and a smaller number of cells per individual compared with bulk RNA-seq or microarray analysis. The preparation time and, therefore, potential changes due to RNA degradation are greater for single cell RNA-seq than bulk RNA-seq or microarray analysis, which may lead leading to a disproportionate increase in variability for low abundance transcripts like *ACE2* in single cell RNA-seq compared with bulk RNA-seq or microarray analysis. Finally, single cell RNA-seq is also influenced by the relative size and shape of cells with larger, fragile, differentiated cells exhibiting higher losses during recovery from the airway epithelium. For example, ciliated cells are known to comprise 60-70% in cell differentials of epithelium taken from small airway but make up less than 20% of live single cells that survive to be included in single cell analysis. Despite these caveats, single cell RNA-seq provides an opportunity to gain expression data on single cell types in a mixed population which is not possible using bulk RNA-seq or microarray analysis. Among the other host factors related to coronavirus infection, all except *TMPRSS11A* were detected by single cell RNA sequencing, all were widely expressed in small airway epithelial cells, and none were significantly different in nonsmokers *vs* smokers

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

(Figure 5B-H). *CLEC4M*-mediated expression of CD209L, a cell surface protein reported to have binding activity for the SARS-CoV (38), was not detected in airway epithelial samples by single cell RNA sequencing. For transcriptomic analyses including single cell RNA sequencing, microarray analysis and bulk RNA sequencing, it is important to remember that mRNA levels do not always precisely predict protein levels in tissues and that correlative studies to assess protein levels need to be performed.

Finally, we assessed *ACE2* expression in 2 cell lines generated from basal cells of healthy nonsmokers, including BCi-NS1.1 [derived from a single BC from the LAE of a healthy nonsmoker (20)] and hSABCi-NS1.1 [derived from a single BC from the SAE of a healthy nonsmoker (21)]. With the caveat of small "n," these data were derived from RNA-seq data on a single sample of each cell line at each stage of differentiation and are subject to validation, both cell lines expressed low levels of *ACE2* prior to differentiation. When allowed to differentiate on an ALI, both cell lines expressed elevated levels of *ACE2* (RNA-seq, Illumina HiSeq 4000, Table 1). When queried for expression levels of other genes encoding proteins important to coronavirus infection, relatively high levels of *ADAM10*, *ADAM17*, *FURIN*, *CTSL*, and *PI4KB* were all detected and were maintained with differentiation. *TMPRSS2* expression was low in undifferentiated large airway basal cells but became much more pronounced in differentiated large airway epithelium. In contrast, the level of *TMPRSS2* was higher in the small airway immortalized basal cell line, and expression was maintained at approximately the same expression level following differentiation. Protease family members *TMPRSS11A* and *TMPRSS11D* showed low expression at both stages of differentiation. These cell lines can be grown indefinitely and should be useful for investigate the biology of SARS-CoV-2 infection, including screening of therapies designed to inhibit the early stages of infection.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Discussion

Biology of SARS-CoV-2 Infection

Respiratory disease is the dominant manifestation of SARS-CoV-2 which is both acquired and transmitted by inhalation of airborne droplets and by contact routes (4, 5, 7, 8). SARS-CoV-2 expressed the spike protein that binds to angiotensin converting enzyme 2 (*ACE2*) on the surface of airway epithelial cells (7, 13, 15-17). In order to provide additional insights into the biology of this initial interaction of SARS-CoV-2 with the airway epithelium, we assessed our published and unpublished transcriptome databases of human airway epithelium to assess the extent of expression of *ACE2* in healthy nonsmokers and smokers. The data demonstrates widespread expression of *ACE2* throughout the respiratory epithelial surface. In the SAE, a likely site of entry of SARS-CoV-2, all major epithelial cell types express the *ACE2* gene, as indicated by analysis of our single cell RNA data as well as re-analysis of single cell RNA-seq data from Reyfman et al. and Duclos et al. presented in the Supplementary data (29, 30). It is important to note that while the absolute number of cells identified with *ACE2* is very low in this study and in the Reyfman and Duclos studies, the percentage is necessarily an underestimate of the proportion of cells in the airway epithelium that are actually expressing ACE2 due to technical details of the single cell RNA-seq method (28). In support of our data, Harmer et al. (39) found *ACE2* mRNA by qRT-PCR at several levels of the airway. Using data from Duclos et al. (30), following supervised clustering, *ACE2* was apparent in large airway basal/intermediate cells, club cells, mucus cells, and ciliated cells. Finally, several recently submitted manuscripts also indicate that *ACE2* is expressed in a broad array of epithelial cells at various positions along the airway (40, 41). At the protein level, Hamming et al. (42) observed *ACE2* staining in alveolar epithelial cells and in the basal layer of airways. A previous report by Jia el al. localized the

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

majority of *ACE2* protein to ciliated cells in an *in vitro* differentiated large airway cell culture (43), contrasting with the broader expression in epithelial cells we identified. The level of *ACE2* expression in this differentiated large airway epithelial culture was sufficient to measure *ADAM10*- and *ADAM17*-dependent *ACE2* shedding (34). Data from Jia et al. (43) and from our analysis of undifferentiated and differentiated immortalized airway basal cells suggest that *in vitro* cultures of airway epithelium may be useful in studying SARS-CoV-2 infection.

Of interest, expression of the *TMPRSS2* gene is up-regulated in the SAE of smokers. Other members of the transmembrane serine protease family, including *TMPRSS11A* and *TMPRSS11D* (also known as human airway tryptase, HAT), share the ability to activate the fusion peptide in the spike protein via proteolysis (35). While *TMPRSS11A* was not detected in the single cell transcriptome data, both *TMPRSS2* and *TMPRSS11D* were detected in basal, intermediate, club, mucus, and ciliated cells. *TMPRSS2* exhibited strong expression through the airway epithelium, while *TMPRSS11D* was less prevalent in fully differentiated airway epithelial cells. Other than *TMPRSS2*, smoking did not affect the expression of other proteases of this class. Furthermore, intracellular proteases furin and cathepsin L were also widely expressed, suggesting that SARS-CoV-2 could be induced to escape from an endosome following entry, and *PI4KB* was also expressed, completing the complement of cellular proteins needed to support a productive infection of most airway epithelial cell types. Additional host proteins will surely be demonstrated to play modifying roles in SARS-CoV-2 infection. For example, the IFITM family of interferon-inducible anti-viral proteins has previously been shown to have anti-SARS activity although the mechanism of action has not yet been determined (44). Of interest, the fact that cigarette smoke blocks interferon signaling might provide yet another link between cigarette smoking and SARS-CoV-2 infection.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Control of Airway Epithelium *ACE2* Gene Expression

The relative contribution of smoking to the acquisition and course of a SARS-CoV-2 infection has been a source of controversy as the COVID-19 pandemic has developed (45, 46). The finding of the smoking-specific difference in *ACE2* expression in the small airway, but not in the large airway or trachea, is of interest, as it suggests potential differences in airway epithelial susceptibility. Depending on droplet size, inhalation of droplets can lead to deposition of materials throughout the airway (31). At present, the precise avenues of infection by SARS-CoV-2 are not well understood, a point highlighted by the revelation that many contagious individuals are likely asymptomatic, suggesting that they may have active infection in the upper airway without involvement of the lower airway (47). A study of pulmonary infections in mice, caused by a closely related coronavirus, showed that airway infection preceded alveolar involvement, was worse in aged mice, and that an *ACE2* knockout model was protected from infection in the lung, all of which implicate the airway epithelium as a critical element of pathogenesis (48).

We observed that SAE *ACE2* expression is higher in male smokers compared with female smokers, all nonsmokers and male nonsmokers, separately. Combined with the observation that expression of the gene encoding the infectivity activating protease, *TMPRSS2*, was elevated in smokers, there are reasons that smokers, and, in particular, male smokers, would be at greater risk of propagating a SARS-CoV-2 infection. However, whether smoking is a significant risk factor for COVID-19 infection and/or the intensity of the infection is not clearly defined at the clinical level. Cai et al. (49) noted a disproportionate number of men reported with COVID19 across several epidemiological studies among Asian populations early during the pandemic and postulated that the high incidence in men was due to a higher incidence of

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

smoking in that population. Whether the disproportionate incidence among men can be attributed to smoking-induced changes in gene expression, smoking-associated co-morbidities, or some other factor, remains to be determined. A mechanistic explanation for explaining the variations in *ACE2* gene expression in this study has not yet been established.

Our data also showed that miR-1246, a microRNA with homology to *ACE2*, is down-regulated in the SAE of smokers, providing a potential mechanism for smoking-related upregulation of *ACE2*. Other notable aspects of *ACE2* expression imply that sex-specific gene expression would be anticipated. Despite the fact that *ACE2* is located on the X-chromosome and is known to escape X-inactivation, the gene exhibits a variable sex- and tissue-specific bias with lower expression observed in the female lungs compared with male lungs (50). *ACE2* gene expression may be one of several factors contributing to prevalence of COVID-19 in men.

BCi-NS1.1 and hSABCi-NS1.1 Cell Lines

One of the challenges in studying the early steps in virulent coronaviruses like SARS-CoV-2 is establishing an *in vitro* cell culture system that reflects, as close as possible, the interaction of the virus with the human respiratory epithelium. In the context that the primary human airway basal cells express *ACE2*, and then when differentiated on ALI, the differentiated progeny express *ACE2*, basal cell differentiation provides an *in vitro* culture model to assess SARS-CoV-2-airway epithelium interaction. While primary normal human airway epithelium cannot be maintained more than 3-4 passages *in vitro*, we have immortalized 2 cell lines, BCi-NS1.1 and hSABCi-NS1.1, each derived from a single basal cell of a healthy nonsmoker from large or small airway epithelium, respectively (20, 21). Genes encoding *ACE2*, *TMPRSS2*, and the other cellular factors that collaborate to create a successful SARS-CoV-2 infection were

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

found to be expressed *in vitro* in airway epithelium differentiated from the immortalized the LAE BCi-NS1.1 and SAE hSABCi-NS1.1 cell lines (20, 21); these cells lines should be useful in studying the early events of SARS-CoV-2 infection of SAE and assessment of potential therapies to prevent the progression of COVID-19. Both lines can be genetically manipulated, and both can be passaged indefinitely. Both cell lines are available to the coronavirus community by contacting the senior author.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Acknowledgments. We thank the Epigenomics, Genomics, and Flow Cytometry Core Facilities, Weill Cornell Medical College, for assistance in generating the single cell RNA sequencing data. and N Mohamed for editorial assistance. These studies were supported by HL134549, HL107882, HL107882-S, HL118857 and HL118541 and the Department of Genetic Medicine, Weill Cornell Medical College. MR Rostami was supported, in part, by T32HL094284; SLO'B was supported, in part, by Pulmonary Fibrosis Foundation and the American Thoracic Society.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; 395: 507-513.

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497-506.

3. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Li M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JTK, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020; 382: 1199-1207.

4. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 2020; 323: 1061-1069.

5. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY,

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020; 382: 1708.

6. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020; 323: 1239-1242.

7. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; 395: 514-523.

8. World Health Organization, Depart of Communications. Modes of transmission of virus causing COVID-19:implications for IPC precaution recommendations. 2020. https://www.who.int/publications-detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations. [last accesesd 4/30/20]

9. Novel Coronavirus Pneumonia Emergency Response Epidemiology T. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020; 41: 145-151.

10. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020; 395: 1054-1062.

11. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; 579: 270-273.

12. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. *Nature* 2020; 579: 265-269.

13. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; 382: 727-733.

14. Millet JK, Whittaker GR. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 2015; 202: 120-134.

15. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 2020; 176: 104742.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

16. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020; 181: 271-280.

17. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 2020; 5: 562-569.

18. Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* 2005; 102: 11876-11881.

19. Yang N, Ma P, Lang J, Zhang Y, Deng J, Ju X, Zhang G, Jiang C. Phosphatidylinositol 4-kinase IIIbeta is required for severe acute respiratory syndrome coronavirus spike-mediated cell entry. *J Biol Chem* 2012; 287: 8457-8467.

20. Walters MS, Gomi K, Ashbridge B, Moore MA, Arbelaez V, Heldrich J, Ding BS, Rafii S, Staudt MR, Crystal RG. Generation of a human airway epithelium derived basal cell line with multipotent differentiation capacity. *Respir Res* 2013; 14: 135.

21. Wang G, Lou HH, Salit J, Leopold PL, Driscoll S, Schymeinsky J, Quast K, Visvanathan S, Fine JS, Thomas MJ, Crystal RG. Characterization of an immortalized human small airway basal stem/progenitor cell line with airway region-specific differentiation capacity. *Respir Res* 2019; 20: 196.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

22. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG. Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med (Berl)* 2007; 85: 39-53.

23. Turetz ML, O'Connor TP, Tilley AE, Strulovici-Barel Y, Salit J, Dang D, Teater M, Mezey J, Clark AG, Crystal RG. Trachea epithelium as a "canary" for cigarette smoking-induced biologic phenotype of the small airway epithelium. *Clin Transl Sci* 2009; 2: 260-272.

24. Vanni H, Kazeros A, Wang R, Harvey BG, Ferris B, De BP, Carolan BJ, Hubner RH, O'Connor TP, Crystal RG. Cigarette smoking induces overexpression of a fat-depleting gene AZGP1 in the human. *Chest* 2009; 135: 1197-1208.

25. O'Beirne SL, Shenoy SA, Salit J, Strulovici-Barel Y, Kaner RJ, Visvanathan S, Fine JS, Mezey JG, Crystal RG. Ambient Pollution-related Reprogramming of the Human Small Airway Epithelial Transcriptome. *Am J Respir Crit Care Med* 2018; 198: 1413-1422.

26. Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, Zhou XK, Raman T, Omberg L, Clark A, Mezey J, Crystal RG. Biologic phenotyping of the human small airway epithelial response to cigarette smoking. *PLoS One* 2011; 6: e22798.

27. Wang G, Wang R, Strulovici-Barel Y, Salit J, Staudt MR, Ahmed J, Tilley AE, Yee-Levin J, Hollmann C, Harvey BG, Kaner RJ, Mezey JG, Sridhar S, Pillai SG, Hilton H, Wolff G, Bitter H, Visvanathan S, Fine JS, Stevenson CS, Crystal RG. Persistence of smoking-induced dysregulation of miRNA expression in the small airway epithelium despite smoking cessation. *PLoS One* 2015; 10: e0120824.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

28. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 2014; 11: 740-742.

29. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen CI, Ren Z, Verma R, Abdala-Valencia H, Nam K, Chi M, Han S, Gonzalez-Gonzalez FJ, Soberanes S, Watanabe S, Williams KJN, Flozak AS, Nicholson TT, Morgan VK, Winter DR, Hinchcliff M, Hrusch CL, Guzy RD, Bonham CA, Sperling AI, Bag R, Hamanaka RB, Mutlu GM, Yeldandi AV, Marshall SA, Shilatifard A, Amaral LAN, Perlman H, Sznajder JI, Argento AC, Gillespie CT, Dematte J, Jain M, Singer BD, Ridge KM, Lam AP, Bharat A, Bhorade SM, Gottardi CJ, Budinger GRS, Misharin AV. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1517-1536.

30. Duclos GE, Teixeira VH, Autissier P, Gesthalter YB, Reinders-Luinge MA, Terrano R, Dumas YM, Liu G, Mazzilli SA, Brandsma C-A, van den Berge M, Janes SM, Timens W, Lenburg ME, Spira A, Campbell JD, Beane J. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Science Advances* 2019; 5: eaaw3413.

31. Gonda I. Particle deposition in the human respiratory tract. In: Crystal RG, West JB, Weibel ERB, P.J., editors. The Lung: Scientific Foundations. Philadelphia: Lippincott-Raven; 1997. p. 2289-2294.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

32. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 2020; 20: 425-434.

33. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Sci Total Environ* 2020; 727: 138704.

34. Jia HP, Look DC, Tan P, Shi L, Hickey M, Gakhar L, Chappell MC, Wohlford-Lenane C, McCray PB, Jr. Ectodomain shedding of angiotensin converting enzyme 2 in human airway epithelia. *Am J Physiol Lung Cell Mol Physiol* 2009; 297: L84-96.

35. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* 2011; 85: 873-882.

36. Bosch BJ, Bartelink W, Rottier PJ. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol* 2008; 82: 8887-8890.

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

38. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD, Jr., Thackray LB, Young MD, Mason RJ, Ambrosino DM, Wentworth DE, Demartini JC, Holmes KV. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* 2004; 101: 15748-15753.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

39. Harmer D, Gilbert M, Borman R, Clark KL. Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. *FEBS Lett* 2002; 532: 107-110.

40. Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Winter H, Meister M, Veith C, Boots AW, Hennig BP, Kreuter M, Conrad C, Eils R. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *Embo j* 2020: e105114.

41. Ziegler C, Allon SJ, Nyquist SK, Mbano I, Miao VN, Cao Y, Yousif AS, Bals J, Hauser BM, Feldman J, Muus C, Wadsworth II MH, Kazer S, Hughes TK, Doran B, Gatter GJ, Vukovic M, Tzouanas CN, Taliaferro F, Guo Z, Wang JP, Dwyer DF, Buchheit KM, Boyce J, Barrett NA, Laidlaw TM, Carroll SL, Colonna L, Tkachev V, Yu A, Zheng HB, Gideon HP, Winchell CG, Lin P, L. , Berger B, Leslie A, Flynn JL, Fortune SM, Finberg RW, Kean L, Garber M, Schmidt A, Lingwood D, Shalek AK, Ordovas-Montanes J, Lung Biological Network H. Title (In press).

42. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* 2004; 203: 631-637.

43. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, Farzan M, Wohlford-Lenane C, Perlman S, McCray PB, Jr. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol* 2005; 79: 14614-14621.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

44. Bailey CC, Zhong G, Huang IC, Farzan M. IFITM-Family Proteins: The Cell's First Line of Antiviral Defense. *Annu Rev Virol* 2014; 1: 261-283.

45. Lippi G, Henry BM. Active smoking is not associated with severity of coronavirus disease 2019 (COVID-19). *Eur J Intern Med* 2020.

46. Vardavas CI, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. *Tob Induc Dis* 2020; 18: 20-20.

47. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, Bardossy AC, Oakley LP, Tanwar S, Chisty Z, Bell JM, Methner M, Harney J, Jacobs JR, Carlson CM, McLaughlin HP, Stone N, Clark S, Brostrom-Smith C, Page LC, Kay M, Lewis J, Russell D, Hiatt B, Gant J, Duchin JS, Clark TA, Honein MA, Reddy SC, Jernigan JA, County PHSK, Team CC-I. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility — King County, Washington, March 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69: 377–381.

48. Menachery VD, Yount BL, Jr., Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco WA, Shi ZL, Baric RS. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 2015; 21: 1508-1513.

49. Cai H. Sex difference and smoking predisposition in patients with COVID-19. *Lancet Respir Med* 2020; 8: e20.

50. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A,

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Lappalainen T, Regev A, Ardlie KG, Hacohen N, MacArthur DG. Landscape of X chromosome inactivation across human tissues. *Nature* 2017; 550: 244-248.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure Legends

Figure 1. Expression of *ACE2* in the human airway epithelium of healthy nonsmokers. **A, B.** Expression level is presented as relative gene expression compared to all other genes on the array. See Supplemental Methods for details on normalization. **A.** Comparison of *ACE2* expression in trachea epithelium, large airway epithelium (LAE) and small airway epithelium (SAE). Quantification by Affymetrix HG-U133 Plus 2.0 microarrays. The data was generated from the datasets of GEO accession numbers 13933, 10135 and 11784 (23, 24, 26) and compared using a 2-way ANOVA (gender was identified as a source of variation). **B.** *ACE2* expression during *in vitro* differentiation of airway epithelium derived from primary tracheal basal cells (BC). RNA was collected by brushing from freshly isolated, purified tracheal BC and from cells derived from the BC on an air-liquid interface (ALI) culture at initiation of the culture (day 0) and at days 7 to 28 of culture. *ACE2* levels (determined by Affymetrix HG-U133 Plus 2.0 microarrays) increased as BC differentiated into airway epithelial cells (*ACE2* levels at day 28 compared to day 0, $p<10^{-5}$). **C.** Single cell 10x analysis of *ACE2* expression in the different cell populations comprising the normal SAE of healthy nonsmokers. All the major cell types express *ACE2*, including basal, intermediate, club, mucus and ciliated cells. Each data point represents a single cell. *ACE2* was detected in a minority of epithelial cells from each cluster (1.2% of basal cells, 2.6% of intermediate cells, 1.7% of club cells, 2.4% of mucus cells, and 1.0% of ciliated cells). These values are useful for comparison among the epithelial cell types, but underestimate the actual percentage of cells expressing the gene (28). See Supplemental Methods for markers used to define each cell type and for details on calculation of scaled UMI and transformation for data presentation. NS = non-significant; * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure 2. Effect of smoking and sex on *ACE2* expression in the small airway epithelium (SAE). **A, B.** Expression level is presented as relative gene expression compared to all other genes on the array. See Supplemental Methods for details on normalization. **A.** Healthy smokers *vs* nonsmokers, male and female combined, Affymetrix HG-U133 Plus 2.0 microarrays. The data was generated from the dataset of Tilley et al (26), GEO accession number 11784. **B.** Male *vs* female for smokers *vs* nonsmokers, Affymetrix HG-U133 Plus 2.0 microarrays, same dataset as in panel A. **C.** Smokers *vs* nonsmoker, male and female combined, RNA-seq (Illumina HiSeq 2500). **D.** Male *vs* female for smokers *vs* nonsmokers, RNA-seq, same dataset as in panel C. A 2-way ANOVA (sex was identified as a source of variation) was used for analysis. NS = non-significant; * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$. Open circles – nonsmokers, gray circles – smokers.

Figure 3. Possible relationship of miR-1246 levels to modulate the levels of *ACE2* in the small airway epithelium (SAE). Assessment of the dataset of Wang et al (27) (GEO accession number 53519) of healthy nonsmokers (n=9) and healthy smokers (n=10) for smoking-related significant changes in levels of miRNA in the SAE for miRNA-1246 with sequences that complement the sequence of the 3' untranslated region (3'UTR) of *ACE2* mRNA. **A.** Predicted pairing of target region in *ACE2* 3'UTR (top) and human miR-1246 (bottom) analyzed by TargetScanHuman 7.2 (http://www.targetscan.org/vert_72/). **B, C.** miR-1246 levels are decreased in the SAE of smokers compared to nonsmokers. A 2-way ANOVA (age was identified as a source of variation) was used for analysis. NS = non-significant; * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$. **B.** Assessment by Affymetirx miRNA 2.0 arrays. Expression of miRNA is presented as relative miRNA expression compared to all other human mature miRNA. See Supplemental

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Methods for details). **C.** Assessment by TaqMan PCR. Data is from Wang et al (27). Open circles – nonsmokers, gray circles – smokers.

Figure 4. Assessment of the small airway epithelium (SAE) of healthy nonsmokers (n=20) and smokers (n=23) for expression of genes that may be relevant to SARS-CoV-2 infection. See text for details regarding the possible relevance of these genes to SAE infection by SARS-CoV-2. Quantification by RNA-seq (Illumina HiSeq 2500). A 1-way ANOVA was used for analysis. NS = non-significant; * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$. Open circles – nonsmokers, gray circles – smokers.

Figure 5. Single cell 10x transcriptome analysis of the small airway epithelium (SAE) of healthy nonsmokers (n=5) and smokers (n=5) for expression of genes that may be relevant to SARS-CoV-2 infection of the SAE. A total of 18,263 cells from nonsmokers and 16,678 cells from smokers were analyzed. *ACE2* cells were detected in a minority of epithelial cells from both non-smokers and smokers (NS%/S%: 1.2%/0.6% of basal cells, 2.6%/2.1% of intermediate cells, 1.7%/1.3% of club cells, 2.4%/1.9% of mucus cells, and 1.0%/1.2% of ciliated cells). These values are useful for comparison among the epithelial cell types, but underestimate the actual percentage of cells expressing the gene (28). Expression of *TMPRSS11A* was detected by bulk RNA-seq (Illumina HiSeq 2500, Figure 4D) but not detected by single cell RNA sequencing (10x). For all comparisons of gene expression in all cell types, the differences in expression levels in nonsmokers (NS) and smokers (S) were <10%; no significant differences in gene expression were observed. Statistical comparisons were performed using the Wilcoxon rank sum test with p values adjusted using the Bonferroni correction. See Supplemental Methods for markers used to define each cell type and for details on calculation of scaled UMI and transformation for data presentation.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

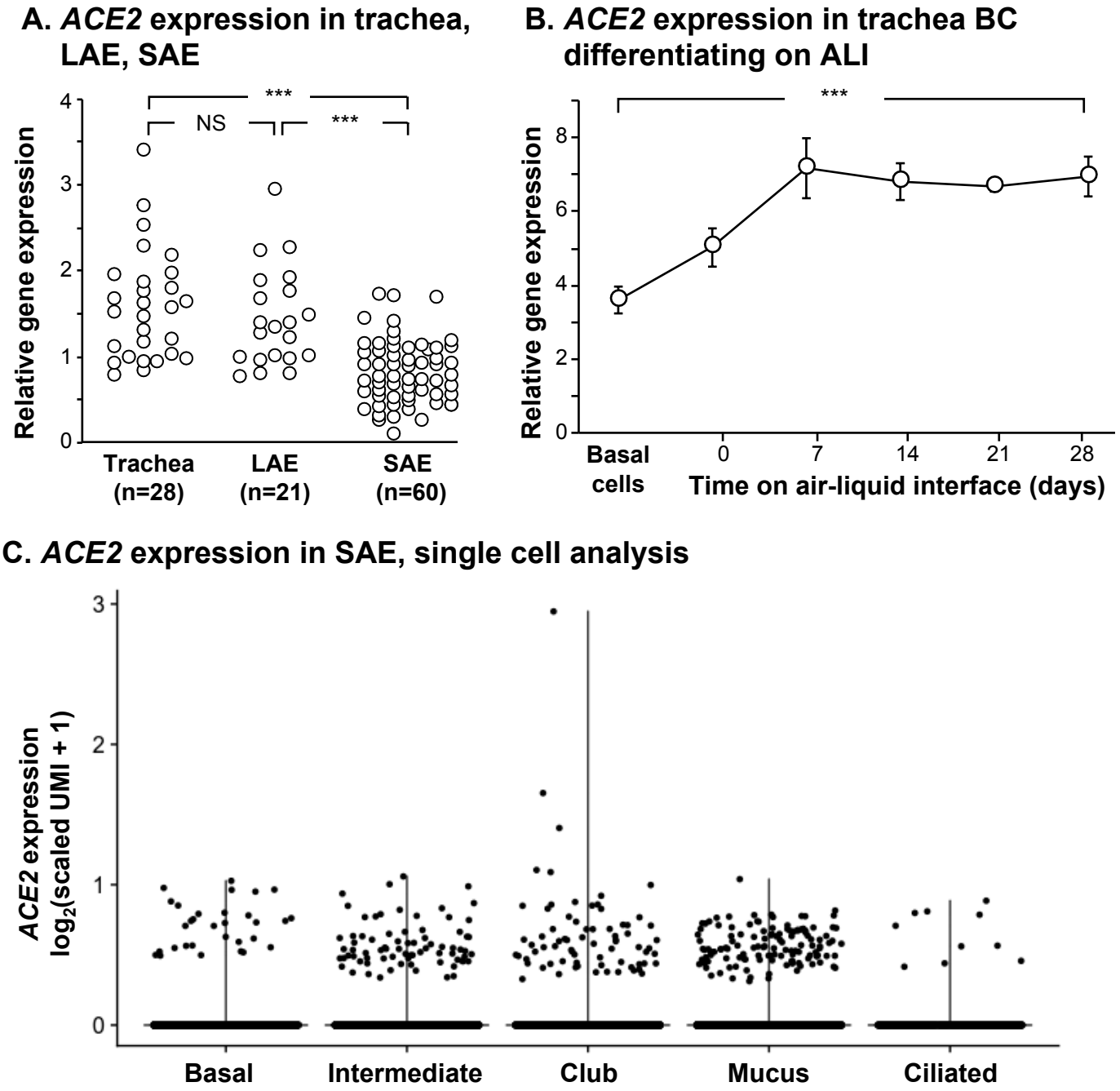
Table 1. Expression of *ACE2* and other Genes that may be Relevant to SARS-CoV-2 Infection in the BCi-NS1.1 and hSABCi-NS1.1 Immortalized Large and Small Airway Epithelium Basal Cell Lines at Baseline and after Differentiation on Air-Liquid Interface[1]

	BCi-NS1.1[2]		hSABCi-NS1.1[3]	
Gene	**Basal cell baseline**	**Air-liquid interface day 28**	**Basal cell baseline**	**Air-liquid interface day 28**
ACE2	0.4	1.7	0.4	2.8
ADAM10	23.8	13.6	23.4	21.8
ADAM17	9.6	6.9	14.0	11.8
TMPRSS2	3.3	25.4	12.2	15.2
TMPRSS11A	0	0.5	0.5	0
TMPRSS11D	0.1	0.7	0.4	0.2
FURIN	14.9	12.5	46.8	29.5
CTSL	37.5	33.1	74.7	67.0
PI4KB	18.0	28.6	19.2	22.6

[1] Expression assessed by RNA-seq (Illumina HiSeq 4000); data is presented in FPKM.
[2] The BCi-NS1.1 line was derived from large airway epithelium basal cell of a healthy nonsmoker and differentiated on air-liquid interface for 28 days (20).
[3] The hSABCi-NS1.1 line was derived from small airway epithelium basal cell of a healthy nonsmoker and differentiated on air-liquid interface for 28 days (21).

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

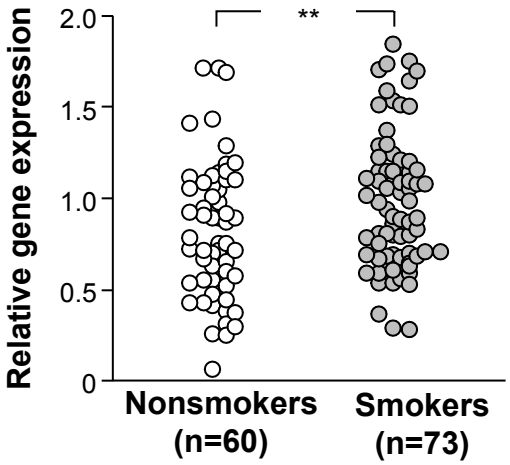
Figure 1



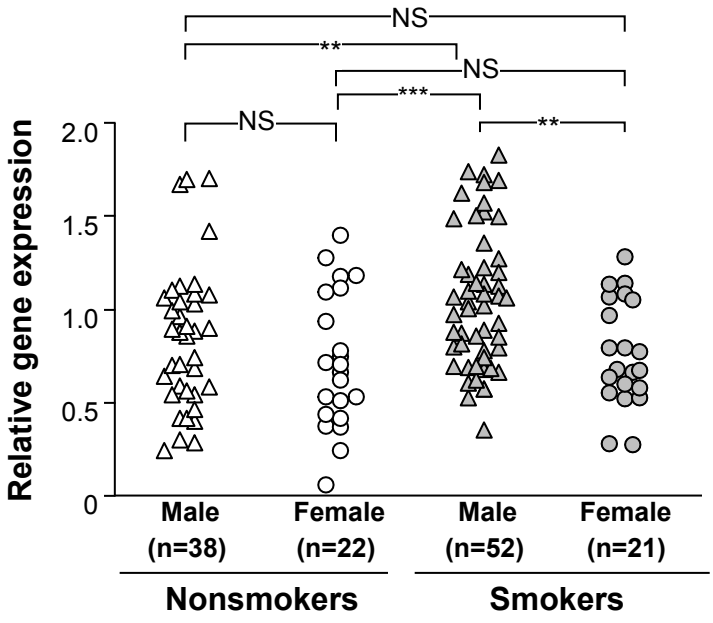
AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure 2

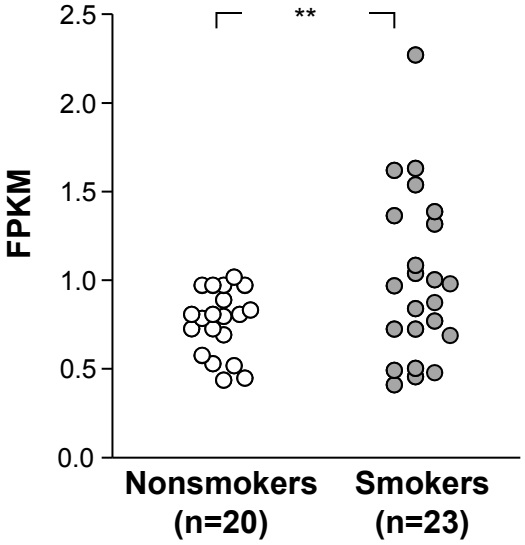
A. *ACE2* expression, SAE, smokers *vs* nonsmokers, microarray



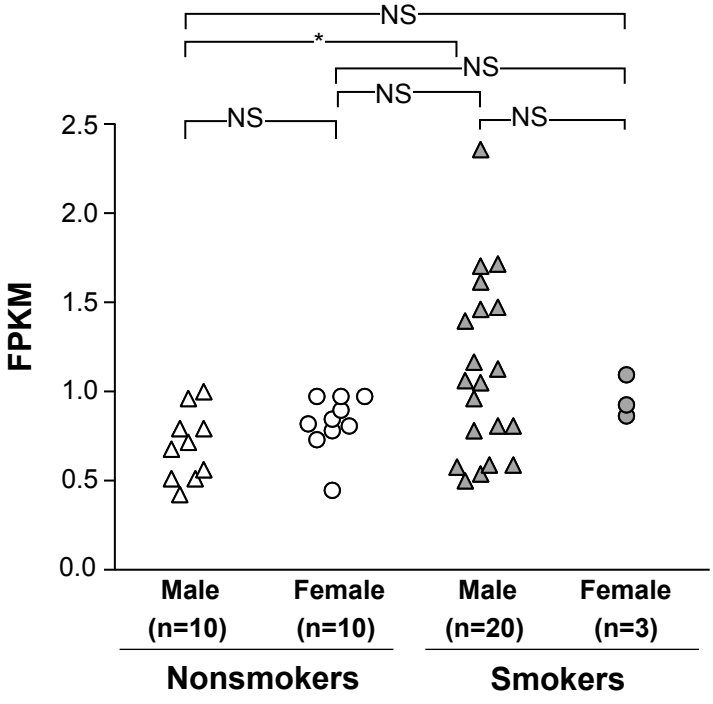
B. *ACE2* expression, SAE, smokers *vs* nonsmokers, microarray



C. *ACE2* expression, SAE, smokers *vs* nonsmokers, RNA-seq



D. *ACE2* expression, SAE, smokers *vs* nonsmokers, RNA-seq



AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure 3

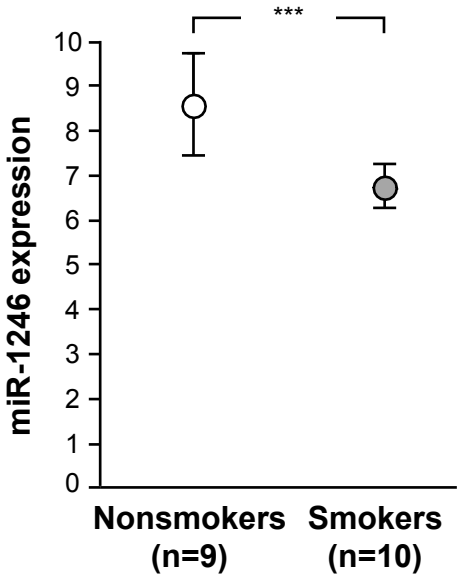
A. miR-1246 and *ACE2*

Position 461-467 of ACE2 3' UTR `5' ...AGCUCACUUUCAUUUAAUCCAUU...`

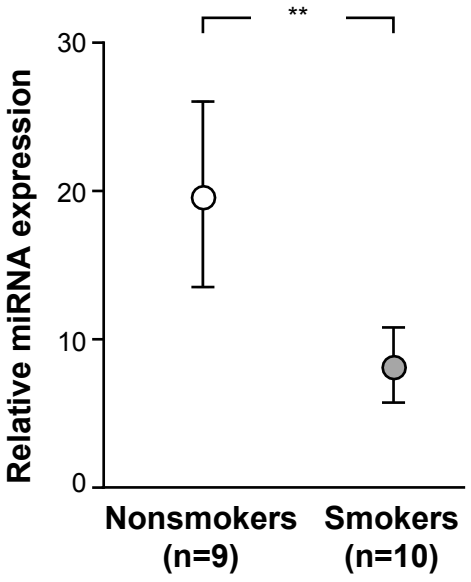
`||||||| `

hsa-miR-1246 `3' GGACGAGGUUUUUAGGUAA`

B. miR-1246, microarray

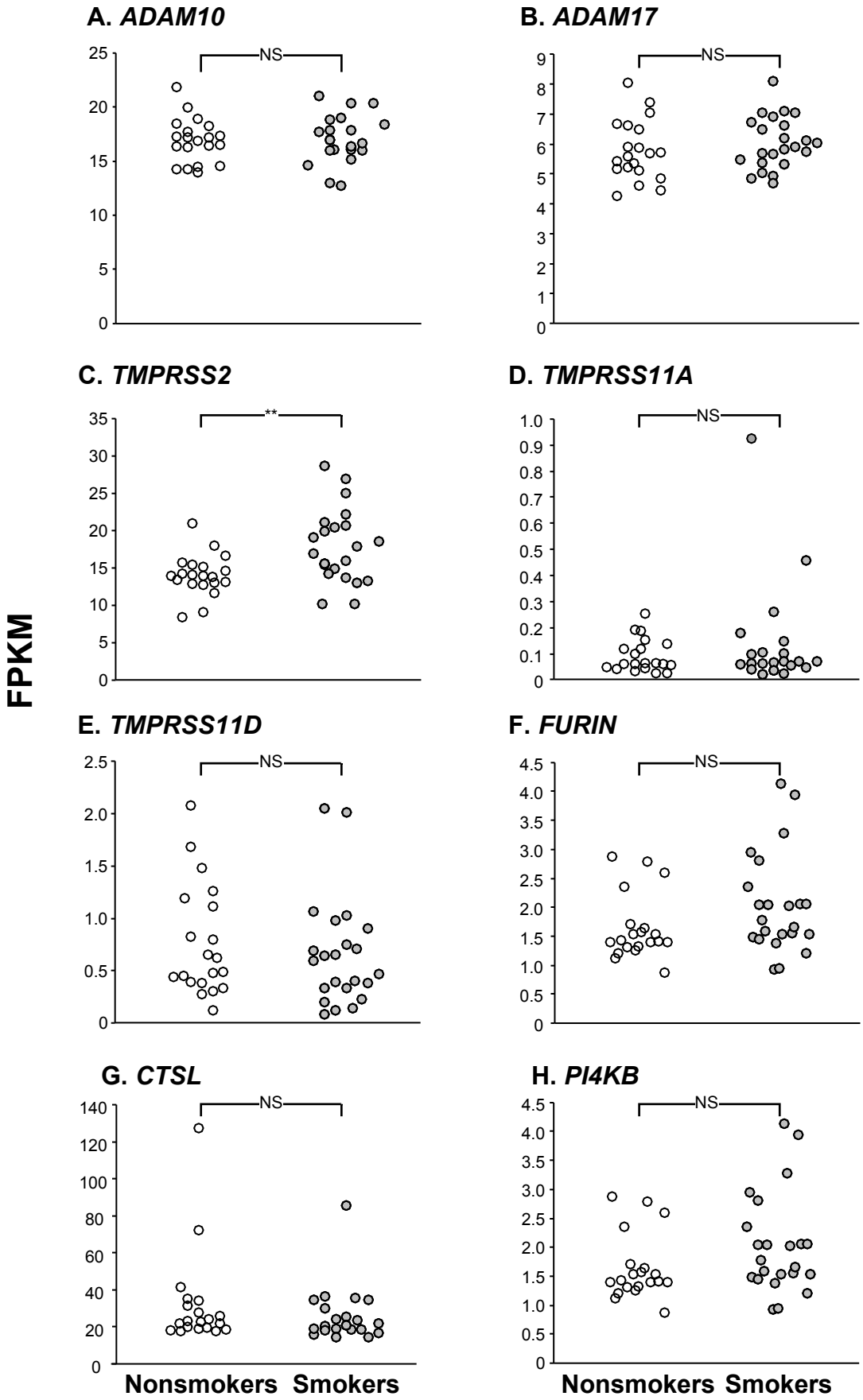


C. miR-1246, TaqMan



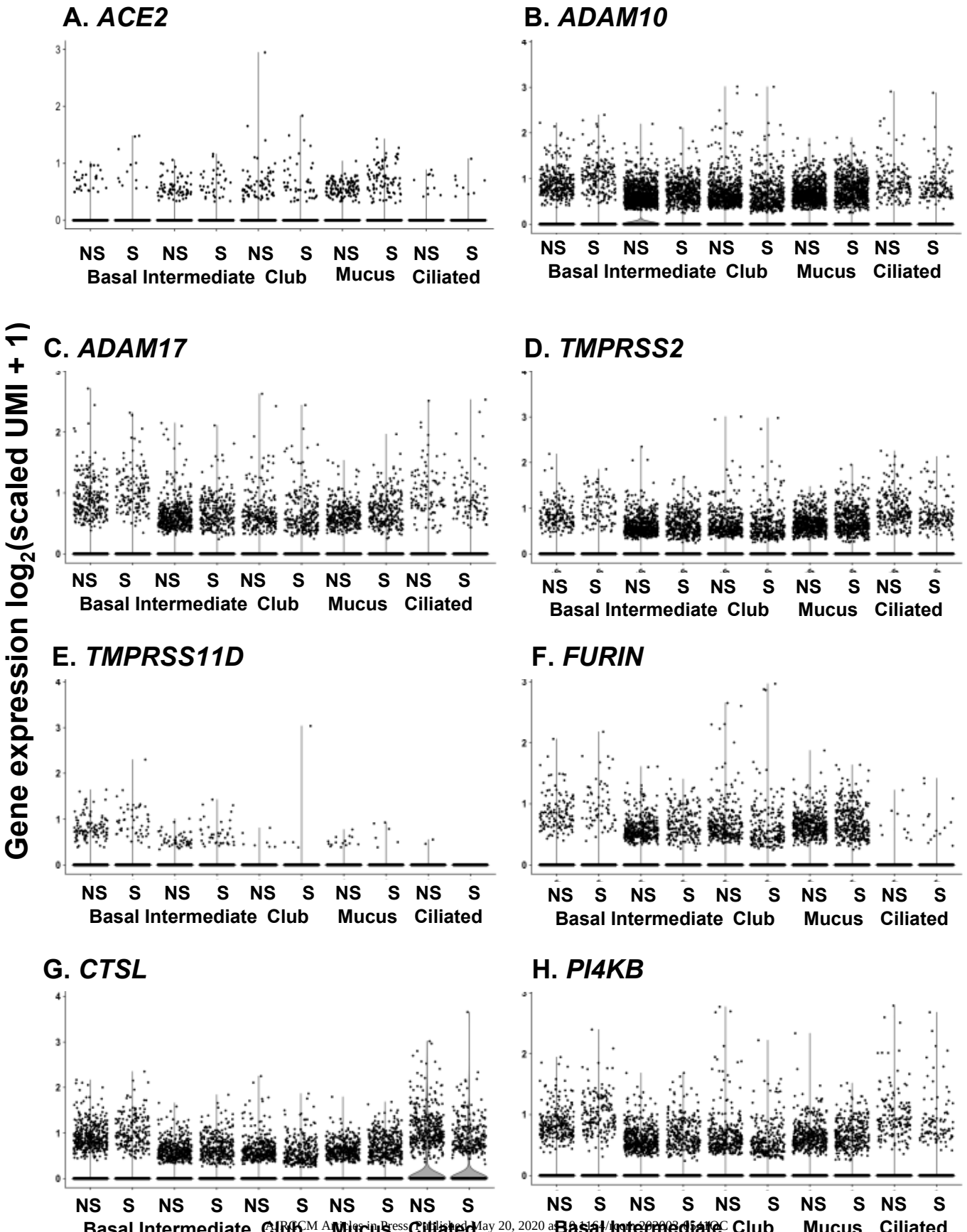
AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure 4



AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure 5



Online Supplemental Methods

Study Population Inclusion / Exclusion Criteria Inclusion/Exclusion Criteria

Healthy nonsmokers

Inclusion criteria

- Capable of providing informed consent
- Willingness to participate in the study
- Men and women, age 18 or older
- Negative HIV
- Not pregnant (women)
- Good overall health without history of chronic lung disease, including asthma, and without recurrent or recent (within 3 months) acute pulmonary disease
- No history of allergies to medications to be used in the bronchoscopy procedure
- Not taking any medications relevant to lung disease or having an effect on the airway epithelium
- Normal physical examination
- Normal routine laboratory evaluation, including general hematologic studies, general serologic/immunologic studies, general biochemical analyses, and urine analysis
- Normal electrocardiogram
- Normal chest X-ray (PA and lateral)
- Normal serum α1-antitrypsin levels
- Self-reported never smokers, with smoking status validated by the absence of nicotine and cotinine in urine (nicotine<2 ng/ml, cotinine<5 ng/ml) (1)
- Normal lung function, including forced expiratory volume in 1 second (FEV1) ≥80% predicted, forced vital capacity (FVC) ≥80% predicted, FEV1/FVC ≥0.7 based on pre-bronchodilator spirometry, total lung capacity (TLC) ≥90% predicted and DLCO ≥80% predicted

Healthy smokers

Inclusion criteria

- Capable of providing informed consent
- Willingness to participate in the study
- Men and women, age 18 or older
- Negative HIV
- Not pregnant (women)
- Good overall health without history of chronic lung disease, including asthma, and without recurrent or recent (within 3 months) acute pulmonary disease
- No history of allergies to medications to be used in the bronchoscopy procedure
- Not taking any medications relevant to lung disease or having an effect on the airway epithelium
- Normal physical examination
- Normal routine laboratory evaluation, including general hematologic studies, general sero-

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

logic/immunologic studies, general biochemical analyses, and urine analysis
- Normal electrocardiogram
- Normal chest X-ray (PA and lateral)
- Normal serum α1-antitrypsin levels
- Self-reported current daily smokers with pack-yr >5, validated by urine nicotine > 30 ng/ml and cotinine > 50 ng/ml (1)
- Normal lung function, including FEV1 ≥80% predicted, FVC ≥80% predicted, FEV1/FVC ≥0.7 based on pre-bronchodilator spirometry, TLC ≥90% predicted and DLCO ≥80% predicted

All Subjects

Exclusion criteria
- Unable to meet the inclusion criteria
- Current active infection or acute illness of any kind
- Evidence of malignancy within the past 5 years
- Alcohol or drug abuse within the past 6 months

Study Population and Biologic Samples

Normal nonsmokers and phenotypic normal smokers were recruited using local print and online media. Research subjects were evaluated at the Weill Cornell Medical College Clinical Translational and Science Center and the Department of Genetic Medicine Clinical Research Facility under IRB-approved protocols. After providing written consent, all subjects underwent a detailed screening visit and assessment of medical history, physical exam, complete blood count, coagulation studies, liver function tests, urine analysis, chest X-ray, high resolution chest CT scan, EKG and pulmonary function tests and determined to be phenotypically normal (see Inclusion/Exclusion criteria).

Sampling trachea, large and small airway epithelium

After evaluation, all individuals who met inclusion/exclusion criteria underwent either tracheal brushing without conscious sedation to obtain trachea epithelial cells as previously described (2) or sampling of the small or large airway epithelium using brushing under mild sedation and anesthesia of the vocal cords as previously described (3, 4). Smokers were asked not to smoke the evening prior to the procedure. For trachea sampling, a fiberoptic bronchoscope (Pen-

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

tax, Tokyo, Japan, EB-1530T3) was used to collect tracheal epithelial cells. A 2 mm cytology brush (Kimberly Clark, Roswell, GA) was advanced through the working channel of the bronchoscope and tracheal epithelial cells were obtained by gently gliding the brush back and forth 20 times on tracheal epithelium in at least five different locations (2).

To sample the large or small airway epithelium, the fiberoptic bronchoscope was positioned proximal to the opening of a desired lobar bronchus. A 2 mm diameter brush was used for gentle brushing of the 3rd to 4th order bronchi [large airway epithelium (LAE) (4)] or 10th to 12th generation branch [small airway epithelium (SAE) (3)] of the right lower lobe. Cells were collected by gently gliding the brush back and forth on the epithelium 5 to 10 times in 8 to 10 different locations in the same general area.

There was an overlap of a subset of the individuals used for analysis by the different methods as detailed per figure: Figure 1A, nonsmoker microarray analysis of n=60 nonsmokers with SAE samples, n=8 had LAE samples, additional n=12 had trachea samples, and additional n= 9 had both LAE and trachea samples. Figure 1B, ALI culture of trachea samples from nonsmokers: there is no overlap of the nonsmokers with any other data sets used for analysis in the manuscript. Figure 2A-D and Figure 4, SAE microarray and RNA-seq: there is no overlap of any of the subjects analyzed using microarray and RNA-Seq. Figure 3, SAE miRNA: of n=9 nonsmokers, n=2 were also analyzed for SAE RNA-Seq (Figure 2C-D). None of the n=10 smokers were used in any other analyses in this manuscript. Figure 1C, 6 and E2 – there is no overlap of any subjects with single cell samples with any other data sets used analyzed in this manuscript.

Trachea, Large and Small Airway Sample Processing

Collected cells were processed as previously described (2-4). Briefly, cells were dislodged from the cytology brush by flicking into 5 ml of ice-cold Bronchial Epithelium Basal

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Medium (BEBM, Lonza, Basel, Switzerland) and kept on ice until processed. A 0.5 ml aliquot was used for differential cell count and the remaining 4.5 ml were immediately processed for RNA extraction. Total cell number was determined by counting on a hemocytometer and cell morphology and differential cell count (percentage of inflammatory and epithelial cells as well as proportions of ciliated, basal, secretory, and undifferentiated epithelial cells) were assessed on sedimented cells prepared by centrifugation (Cytospin 11, Shandon instruments, Pittsburgh, PA) and stained with Diff -Quik (Dade Behring, Newark, NJ).

Microarray Assessment of Expression of *ACE2*

Total RNA was prepared for microarray transcriptome analysis using the 3'IVT Express kit (Affymetrix, Santa Clara, CA) and assessed using Affymetrix HG-U133 Plus 2.0 microarrays (Affymetrix), as previously described (5, 6). Briefly, RNA quantity was assessed by Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE) and RNA quality by Bioanalyzer (Agilent Technologies, Santa Clara, CA) (7, 8). Total RNA (1 to 2 μg) was used to synthesize double stranded cDNA and Affymetrix kits were used to quantify the biotin-labeled cDNA yield (5). RNA was hybridized on the arrays with probes for >54,000 genome-wide transcripts, using Affymetrix protocols, hardware and software (8). Microarray quality was verified by signal intensity ratio of *GAPDH* 3' to 5' probe sets ≤ 3.0 and multi-chip normalization scaling factor ≤ 10.0 (7). The MAS5 algorithm (GeneSpring version 7.3, Affymetrix Microarray Suite Version 5) was used to normalize the data per array to the median expression value of each sample. *ACE2* gene was represented by one probe (219962_at) selected based on highest specificity and sensitivity scores (Affymetrix) and expression level is presented as relative gene expression compared to all other genes on the array (n=14,465, each represented by one probe set/gene). The image files from the microarrays were processed in Partek Genomics Suite software version 6.6, 2012 (Partek, St. Louis, MO). All compared groups were matched for ethnicity and pack-yr (among smoker males

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

and females). A 1-way ANOVA was used to compare the LAE (age showed a trend towards being a source of variation) and a 2-way ANOVA was used to compare the trachea *vs* large *vs* SAE of nonsmokers and to compare the SAE of smokers and nonsmokers as sex was identified as a source of variation (Partek). When compared on a genome-wide basis, *ACE2* was found to be differentially expressed in trachea *vs* SAE and LAE *vs* SAE of nonsmokers, and in SAE o*f* smokers *vs* nonsmokers [$p<0.05$, with Benjamini-Hochberg correction for multiple testing (9)].

RNA-seq Assessment of Expression of *ACE2* and Related Genes

RNA was purified, amplified and loaded onto an Illumina flowcell for paired-end sequencing reactions using the Illumina HiSeq 2500 or Illumina HiSeq 4000 (Illumina, San Diego, CA), as previously described (10). The library was prepared using (0.5 µg total RNA) TruSeq RNA Library Prep Kit v2. Illumina HiSeq paired-end reads were aligned to GRCh37/hg19 human reference genome and RefSeq gene definitions (2014-06-02) using STAR (2.3.1z13_r470). Cufflinks (2.2) was used to convert aligned reads into fragments per kilobase of exon per million fragments sequenced (FPKM) using RefSeq gene definitions. All compared groups were matched for ethnicity and pack-yr (among smoker males and females). A 1-way ANOVA was used to compare the LAE (age showed a trend towards being a source of variation) and a 2-way ANOVA was used to compare the SAE as sex was identified as a source of variation (Partek). When compared on a genome-wide basis, *ACE2* was found to be differentially expressed in the SAE of smokers *vs* nonsmokers [$p<0.05$, with Benjamini-Hochberg correction for multiple testing (9)].

10x Single Cell Assessment of Expression of *ACE2* and Related Genes

Small airway epithelial cells were obtained via brushing at the level of the 10^{th} to 12^{th} airway fiberoptic bronchoscopy (3). Brushed cells were treated with ACK lysing buffer to remove red blood cells and treated with 0.05% trypsin-ethylenediaminetetraacetic acid (GIBCO,

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

ThermoFisher, Waltham, MA) (5 min) to release single cells from clumps. Trypsin activity was stopped by addition of 15% fetal bovine serum (GIBCO) in HEPES buffered saline solution (Lonza, Morristown, NJ). Cells were washed by centrifugation at 500x*g* in PBS supplemented with 0.04% IgG free, protease free bovine serum albumin (Jackson ImmunoResearch, West Grove, PA), filtered through a 100 μm cell strainer filtered through a 100 μm Falcon cell strainer (Fisher Scientific, Waltham, MA) and a 35 μm Falcon cell strainer into a flow cytometry staining tube (Fisher Scientific), and stained with DAPI (Sigma Chemical, St. Louis, MO) to identify dead cells. Fluorescence activated cell sorting was used to collect live, single cells for analysis by single cell RNA sequencing (10x Genomics, Pleasanton, CA).

Downstream single cell analysis of small airway epithelial cells was performed using Seurat package V3 and R 3.5 (11). Genes were eliminated from the analysis if they were detected in less than 10 cells. Cells were eliminated from the data set if they contained less than 200 detected genes or if mitochondrial genes accounted for more than 25 of expressed genes. A total of 38,173 cells was recovered from the 10 samples. The IntegrateData function was used to integrate the datasets from individual subjects, and anchors between data sets were identified using FindIntegrationAnchors. Normalization of 18,990 identified genes was performed by the total number of unique molecular identifiers (UMI) per cell, multiplying by a scale factor (10,000) and then applying a log transformation. Principal Component Analysis (PCA) was carried out for the top 2,000 variable genes and used first 15 principal components (PCs) to project cells onto a two-dimensional map using Uniform Manifold Approximation and Projection (UMAP) dimensional reduction method using the RunUMAP function. To identify cell types from small airway epithelium, all the cells from the 10 subjects were clustered using K-nearest neighbor (KNN) graph-based clustering algorithm and FindNeighbors function. Finally, FindClusters function (resolution parameter=0.2) was used to establish cell clusters. Following unsupervised clustering

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

of cells, cluster signature genes were identified as genes present in at least 10% of the cells in the population and showing significantly elevated gene expression (0.25 log scale) in one cluster compared with expression of the marker in the total population all other clusters using Wilcoxon rank sum test with Bonferroni correction (p adj <0.05) (12). Clusters were characterized by their top 20 signature genes based on a ranking of most significant p values. The basal cell cluster was identified by the expression of cytoskeletal signature genes *KRT5*, *KRT15*, and *TP63*. The club cell cluster was characterized by expression of host defense genes *SCGB1A1*, *SCGB3A1*, and *C3*, and the absence of *MUC5AC*. Mucous cells signature genes include host defense genes *MUC5AC*, *MUC5B*, and *TFF3*. The ciliated cell cluster was characterized by signature genes relating to cilia structure and function, including *FOXJ1*, *DNAI1*, and *TUBB4B*. The intermediate cell cluster was identified by expression of signature genes shared with other clusters albeit at lower levels, including the basal cell markers (*KRT5*, *KRT15*, and *KRT19*), club cell markers (*KRT19* and *SLPI*) and the mucous cell marker (*CEACM6*). Other smaller clusters, identified as ionocytes, proliferating basal cells, and precursor ciliated cells, did not contain a sufficient number of cells to make an analysis of *ACE2*-positive cells within the clusters to be informative. Statistical comparisons were performed using the Wilcoxon rank sum test with p values adjusted using the Bonferroni correction. No parameters were identified a source a variation.

A genome-wide analysis did not identify *ACE2* as differentially expressed in smokers *vs* non-smokers.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Air-liquid Interface Cultures

Primary tracheal basal cells (BC) isolated from brushed trachea epithelium cell were maintained in culture Bronchial Epithelium Growth Medium (BEGM, Lonza, Walkersville, MD) as previously described (13). BC were seeded onto human type IV collagen-coated transwell filters and induced to differentiate into airway epithelial cells on air-liquid interface (ALI) in a 1:1 mixture of DMEM and Ham's F12 medium with supplements including 2% Ultroser G (BioSerpa S.A., Cergy-Saint-Christophe, France) (13). An ALI culture was introduced two days after establishing the culture by removal of medium from the apical chamber. The medium was changed every 2 to 3 days and the apical surface was rinsed with Search Results
Web results phosphate-buffered saline (PBS) once per week to clear accumulating mucus. Cultures were maintained up to 28 days.

Immortalized cell lines derived from the large airway epithelium BC (BCi-NS1.1) and small airway epithelium BC (hSABCi-NS1.1) were propagated on collagen IV-coated plastic in PneumaCult Ex Plus medium (Stemcell Technologies, Inc., Vancouver, BC), transferred to collagen IV-coated tissue culture filter inserts, and differentiated at ALI to develop a pseudostratified airway epithelium in PneumaCult-ALI medium (Stemcell Technologies) as previously described (14). Briefly, 24 hr after seeding BC onto the filter, the medium was changed from PneumaCult Ex Plus growth medium to PneumaCult-ALI differentiation medium (15). After additional 24 hr, medium was removed from the apical chamber to create ALI conditions. Medium was replaced in the basal chamber every 2 to 3 days, and the apical surface was washed with PBS once per week to remove mucus. In all cases, the establishment of an epithelial barrier was confirmed by measuring trans-epithelial electrical resistance (TEER), which achieved >200 Ohms per cm^2 by ALI day 7 and was maintained at that level through ALI day 28. RNAseq (Illumina HiSeq 4000) was performed on RNA extracted from ALI cultures using Trizol (13) on

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

the day indicated in the text.

miRNA

Following RNA extraction and sample quality assessment, miRNA microarray analyses were performed using Affymetrix miRNA 2.0 arrays (Affymetrix) as previously described (16). Briefly, total RNA was extracted using miRNeasy mini kit (Qiagen, Valencia, CA) and RNA integrity was assessed on an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA) and determined as >5.5 for all samples. Total RNA was then hybridized to Affymetrix miRNA 2.0 arrays (Affymetrix) and miRNA QC tool (Affymetrix) was used for quality control. The image files from Affymetrix miRNA 2.0 arrays were processed in Partek Genomics Suite software version 6.6, 2012 (Partek). Affymetrix Robust Multi-array Average (RMA) algorithm was used to normalize the data per array to the median expression value of each sample. There were a total of 1,100 human mature miRNA and the expression of miRNA is presented as relative miRNA expression compared to all other human mature miRNA. A two-way ANOVA (age was identified as a source of variation) was used to compare smokers and nonsmokers. Our previous analysis of 1100 mature human miRNA in smokers compared to nonsmokers (16) identified 34 miRNA differentially expressed ($p<0.05$, fold-change >1.5). Of those, miR-1246 is the only one that regulates *ACE2* expression. It was the most down-regulated miRNA in smokers compared to nonsmokers ($p<10^{-3}$, fold change = 3.8). Confirmation of levels of miR-1246 was carried out by TaqMan PCR as previously described (16). MicroRNA let-7a was used as endogenous control.

Correlation of *ACE2* Expression with New York City (NYC) Pollution Levels

ACE2 expression levels were analyzed for correlation with 30-day mean levels (particulate matter ≤2.5 μm; $PM_{2.5}$) in nonsmoker and smoker SAE samples obtained by bronchoscopy from the same subject several times over 1 yr (total n=98 nonsmoker samples and n=176 smoker samples). NYC air quality was assessed using daily mean $PM_{2.5}$ levels recorded in the United

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

States Environmental Protection Agency (EPA) Air Quality System's Database (https://www.epa.gov/outdoor-air-quality-data) for sites around the city and averaged per month as previously described (17). In all cases, parameter code 88502 (acceptable $PM_{2.5}$ AQI and speciation mass) data was utilized. The mean $PM_{2.5}$ concentration for the 30 days prior to each bronchoscopy was calculated and used for analysis as it reflects the effects of subacute $PM_{2.5}$ exposure on the SAE transcriptome rather than daily $PM_{2.5}$ levels where analysis could be influenced by brief fluctuations in $PM_{2.5}$ concentrations and daily variations in individuals' exposure to outdoor air.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Table E1. Databases Used for Quantification of Expression of *ACE2* and Associated Genes in the Airway Epithelium of Nonsmokers and Smokers

Table / Figure	Cell source	Clinical pheno-type	n	Methodology[1]	GEO accession #[2]	Reference
Table I	BCi-NS1.1	Nonsmoker	1	Illumina HiSeq 4000	unpublished	(15)
	hSABCi-NS1.1	Nonsmoker	1	Illumina HiSeq 4000	unpublished	(14)
Figure 1A	Trachea	Nonsmoker	28	HG-U133	13933 (published)	(2)
	LAE	Nonsmoker	21	HG-U133	10135 (published)	(4)
	SAE	Nonsmoker	60	HG-U133	11784 (published)	(6)
Figure 1B	Trachea ALI	Nonsmoker	4	HG-U133	unpublished	unpublished
Figure 1C	SAE	Nonsmoker	5	10x single cell	unpublished	unpublished
Figure 2A, B	SAE	Nonsmoker	60	HG-U133	11784 (published)	(6)
	SAE	Smoker	73	HG-U133	11784 (published)	(6)
Figure 2C, D	SAE	Nonsmoker	20	Illumina HiSeq 2500	133924 (pending)	pending
	SAE	Smoker	23	Illumina HiSeq 2500	133924 (pending)	pending
Figure 3	SAE	Nonsmoker	9	miRNA 2.0	53519 (published)	(16)
	SAE	Smoker	10	miRNA 2.0	53519 (published)	(16)
Figure 4	SAE	Nonsmoker	20	Illumina HiSeq 2500	133924 (pending)	pending
	SAE	Smoker	23	Illumina HiSeq 2500	133924 (pending)	pending
Figure 5	SAE	Nonsmoker	5	10x single cell	unpublished	unpublished
	SAE	Smoker	5	10x single cell	unpublished	unpublished
Figure E1	SAE	Nonsmoker	5	10x single cell	unpublished	unpublished
	SAE	Nonsmoker	5	10x single cell	unpublished	unpublished
Figure E2A	Lung	Control	8	10x single cell	122960 (published)	(18)
Figure E2B	LAE	Nonsmoker	6	Illumina HiSeq 2500	131391 (published)	(19)
	LAE	Smoker	6	Illumina HiSeq 2500	131391 (published)	(19)
Figure E3A, B	LAE	Nonsmoker	21	HG-U133	10135 (published)	(4)
	LAE	Smoker	31	HG-U133	10135 (published)	(4)
Figure E3C, D	LAE	Nonsmoker	10	Illumina HiSeq 2500	101353 (pending)	pending
	LAE	Smoker	10	Illumina HiSeq 2500	101353 (pending)	pending
Figure E4	SAE	Nonsmoker	98	Illumina HiSeq 2500	108134 (published)	(17)
	SAE	Smoker	176	Illumina HiSeq 2500	108134 (published)	(17)

[1] HG-U133=Affymetrix HG-U133 Plus 2.0 microarrays; miRNA 2.0 = Affymetrix miRNA 2.0 arrays.
[2] All detailed GEO # are GSE accession #.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Supplemental References

1. Moyer TP, Charlson JR, Enger RJ, Dale LC, Ebbert JO, Schroeder DR, Hurt RD. Simultaneous analysis of nicotine, nicotine metabolites, and tobacco alkaloids in serum or urine by tandem mass spectrometry, with clinically relevant metabolic profiles. *Clin Chem* 2002; 48: 1460-1471.
2. Turetz ML, O'Connor TP, Tilley AE, Strulovici-Barel Y, Salit J, Dang D, Teater M, Mezey J, Clark AG, Crystal RG. Trachea epithelium as a "canary" for cigarette smoking-induced biologic phenotype of the small airway epithelium. *Clin Transl Sci* 2009; 2: 260-272.
3. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG. Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med (Berl)* 2007; 85: 39-53.
4. Vanni H, Kazeros A, Wang R, Harvey BG, Ferris B, De BP, Carolan BJ, Hubner RH, O'Connor TP, Crystal RG. Cigarette smoking induces overexpression of a fat-depleting gene AZGP1 in the human. *Chest* 2009; 135: 1197-1208.
5. Strulovici-Barel Y, Omberg L, O'Mahony M, Gordon C, Hollmann C, Tilley AE, Salit J, Mezey J, Harvey BG, Crystal RG. Threshold of biologic responses of the small airway epithelium to low levels of tobacco smoke. *Am J Respir Crit Care Med* 2010; 182: 1524-1532.
6. Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, Zhou XK, Raman T, Omberg L, Clark A, Mezey J, Crystal RG. Biologic phenotyping of the human small airway epithelial response to cigarette smoking. *PLoS One* 2011; 6: e22798.
7. Raman T, O'Connor TP, Hackett NR, Wang W, Harvey BG, Attiyeh MA, Dang DT, Teater M, Crystal RG. Quality control in microarray assessment of gene expression in human airway epithelium. *BMC Genomics* 2009; 10: 493.
8. Tumor Analysis Best Practices Working G. Expression profiling--best practices for data generation and interpretation in clinical trials. *Nat Rev Genet* 2004; 5: 229-237.
9. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995; 57: 289-300.
10. Ryan DM, Vincent TL, Salit J, Walters MS, Agosto-Perez F, Shaykhiev R, Strulovici-Barel Y, Downey RJ, Buro-Auriemma LJ, Staudt MR, Hackett NR, Mezey JG, Crystal RG. Smoking dysregulates the human airway basal cell transcriptome at COPD risk locus 19q13.2. *PLoS One* 2014; 9: e88051.
11. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell* 2019; 177: 1888-1902 e1821.
12. Jafari M, Ansari-Pour N. Why, When and How to Adjust Your P Values? *Cell J* 2019; 20: 604-607.
13. Hackett NR, Shaykhiev R, Walters MS, Wang R, Zwick RK, Ferris B, Witover B, Salit J, Crystal RG. The human airway epithelial basal cell transcriptome. *PLoS One* 2011; 6: e18378.
14. Wang G, Lou HH, Salit J, Leopold PL, Driscoll S, Schymeinsky J, Quast K, Visvanathan S, Fine JS, Thomas MJ, Crystal RG. Characterization of an immortalized human small airway basal stem/progenitor cell line with airway region-specific differentiation capacity. *Respir Res* 2019; 20: 196.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

15. Walters MS, Gomi K, Ashbridge B, Moore MA, Arbelaez V, Heldrich J, Ding BS, Rafii S, Staudt MR, Crystal RG. Generation of a human airway epithelium derived basal cell line with multipotent differentiation capacity. *Respir Res* 2013; 14: 135.
16. Wang G, Wang R, Strulovici-Barel Y, Salit J, Staudt MR, Ahmed J, Tilley AE, Yee-Levin J, Hollmann C, Harvey BG, Kaner RJ, Mezey JG, Sridhar S, Pillai SG, Hilton H, Wolff G, Bitter H, Visvanathan S, Fine JS, Stevenson CS, Crystal RG. Persistence of smoking-induced dysregulation of miRNA expression in the small airway epithelium despite smoking cessation. *PLoS One* 2015; 10: e0120824.
17. O'Beirne SL, Shenoy SA, Salit J, Strulovici-Barel Y, Kaner RJ, Visvanathan S, Fine JS, Mezey JG, Crystal RG. Ambient Pollution-related Reprogramming of the Human Small Airway Epithelial Transcriptome. *Am J Respir Crit Care Med* 2018; 198: 1413-1422.
18. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen CI, Ren Z, Verma R, Abdala-Valencia H, Nam K, Chi M, Han S, Gonzalez-Gonzalez FJ, Soberanes S, Watanabe S, Williams KJN, Flozak AS, Nicholson TT, Morgan VK, Winter DR, Hinchcliff M, Hrusch CL, Guzy RD, Bonham CA, Sperling AI, Bag R, Hamanaka RB, Mutlu GM, Yeldandi AV, Marshall SA, Shilatifard A, Amaral LAN, Perlman H, Sznajder JI, Argento AC, Gillespie CT, Dematte J, Jain M, Singer BD, Ridge KM, Lam AP, Bharat A, Bhorade SM, Gottardi CJ, Budinger GRS, Misharin AV. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1517-1536.
19. Duclos GE, Teixeira VH, Autissier P, Gesthalter YB, Reinders-Luinge MA, Terrano R, Dumas YM, Liu G, Mazzilli SA, Brandsma C-A, van den Berge M, Janes SM, Timens W, Lenburg ME, Spira A, Campbell JD, Beane J. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Science Advances* 2019; 5: eaaw3413.
20. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 2014; 11: 740-742.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Supplemental Figure Legends

Figure E1. Representation of cell clustering in single cell RNA sequencing data. UMAP projection of unsupervised clustering with clusters identified based on marker gene expression. See Supplemental Methods for marker genes for each cluster.

Figure E2. Expression of *ACE2* is alveolar epithelium type 2 (AT2) cells. **A.** Analysis of the single cell 10x transcription data of Reyfman *et al* (18), GEO accession number 122960. The data was generated from samples of lung parenchyma from 8 control lung biopsies from lung donors with no pulmonary pathology (5 nonsmokers, 2 smokers, and 1 ex-smoker). Data was processed using the pipeline described in the Supplemental Methods. AT2 cells are the dominant cell type expressing *ACE2*, likely due to the high representation of this cell type in the analysis. The level of detection of *ACE2* was similar for all epithelial cell types including AT1, AT2, basal, club and ciliated cells. A total of 24,957 epithelial cells were analyzed of which 91% were AT2 cells. *ACE2* was detected in a minority of epithelial cells from each cluster (0.3% of AT1 cells, 1.2% of AT2 cells, 0.7% of club cells, and 1.2% of ciliated cells). Inflammatory/immune cells did not express *ACE2*. **B.** Analysis of the single cell 10x transcription data of Duclos *et al*. (19), GEO accession number 131391. The data were generated from samples of bronchial airway epithelium obtained by fiberoptic bronchoscopy from 6 never smokers and 6 current smokers. Data was processed using the pipeline described in the Supplemental Methods. Cell clusters were identified as basal/intermediate, mucus, club, ciliated, and ionocyte, although each cluster had some intermediate cell character. The total number of epithelial cells analyzed was 1,105. *ACE2* cells were detected in a minority of epithelial cells from each cluster except ionocytes for which only 13 cells were detected (1.1% of basal/intermediate cells, 1.1% of club cells, 11.5% of mucus cells, and 1.2% of ciliated cells). Values indicating percentage of positive cells above are

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

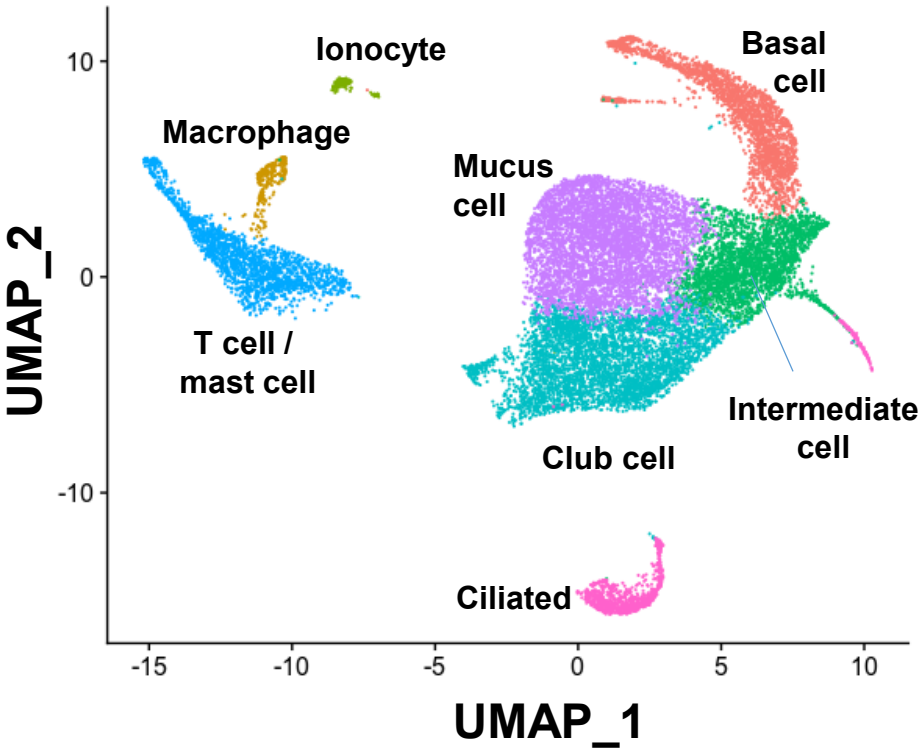
useful for comparison among the epithelial cell types in the data set but underestimate the actual percentage of cells expressing the gene (20).

Figure E3. Expression of *ACE2* in the large airway epithelium (LAE) of nonsmokers compared to smokers. **A.** LAE, nonsmokers (n=21) *vs* smokers (n=31), Affymetrix HG-U133 Plus 2.0 microarray, analysis based on the data of Vanni *et al* (4) GEO accession number 10135. **B.** LAE, effect of sex and smoking. Affymetrix HG-U133 Plus 2.0 microarrays; same data as panel A. **A-B.** Expression level is presented as relative gene expression compared to all other genes on the array. See Supplemental Methods for details on normalization. **C.** LAE, nonsmokers (n=10) *vs* smokers (n=10), RNA-seq (Illumina HiSeq 2500). **D.** LAE, effect of sex. RNAseq; same data as in panel C. There was insufficient female data to determine significance. A 1-way ANOVA was used to compare the groups. NS = non-significant; * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$. Open circles – nonsmokers, gray circles – smokers.

Figure E4. Effect of pollution in New York City on *ACE2* expression in the small airway epithelium (SAE) of nonsmokers and smokers. Fiberoptic bronchoscopy was used to sample the SAE from the same subject at different times over one year. *ACE2* levels in SAE samples were compared to the air pollution levels in New York City in the month the SAE was sampled. The analysis of *ACE2* expression is based on the Affymetrix HG-U133 Plus 2.0 microarray data of O'Beirne *et al*., (17), GEO accession number GSE108134. Open circles - nonsmokers, gray circles - smokers. **A.** healthy nonsmokers; **B.** healthy smokers; and **C.** combined healthy nonsmokers and smokers. There is no correlation of SAE *ACE2* levels and the levels of air pollution in New York City. Particulate matter ≤2.5 μm ($PM_{2.5}$) levels did not exceed 18 $\mu g/m^3$ (units) during the observation period. Open circles – nonsmokers, gray circles – smokers.

AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

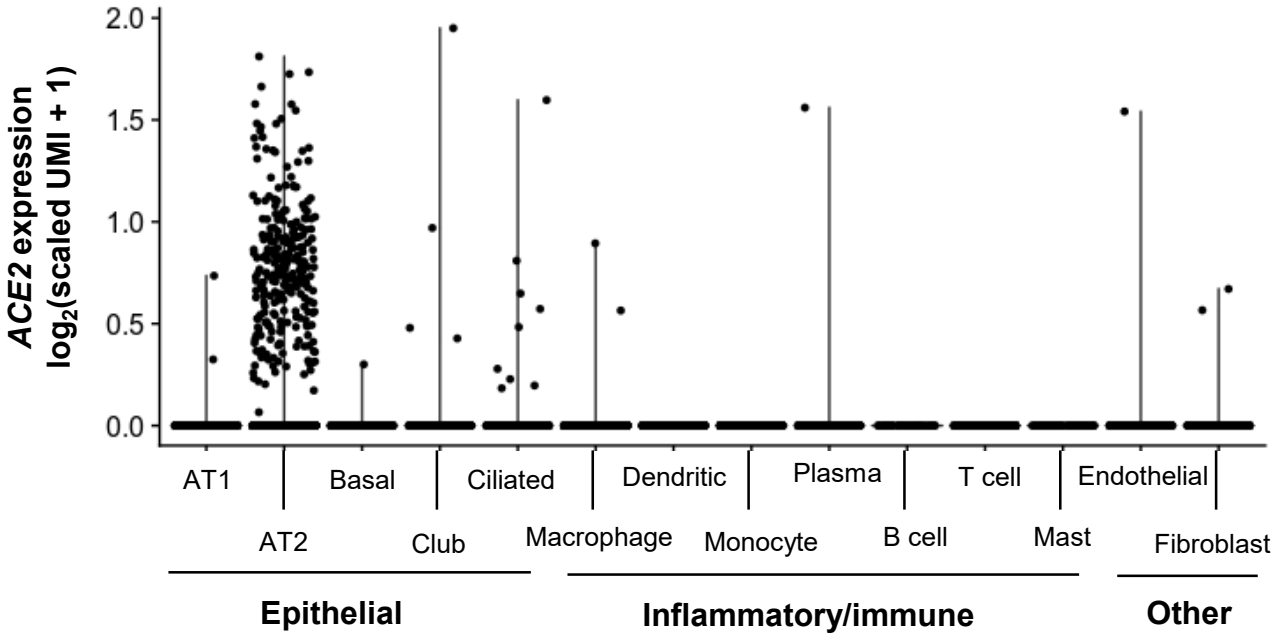
Figure E1



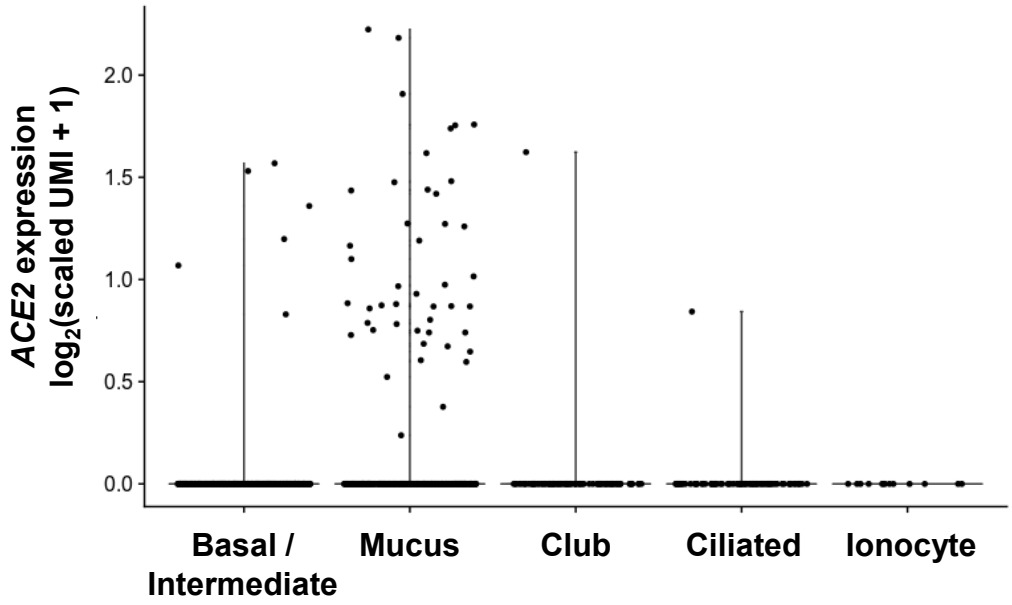
AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure E2

A. ACE2 expression in lung (Reyfman et al.)

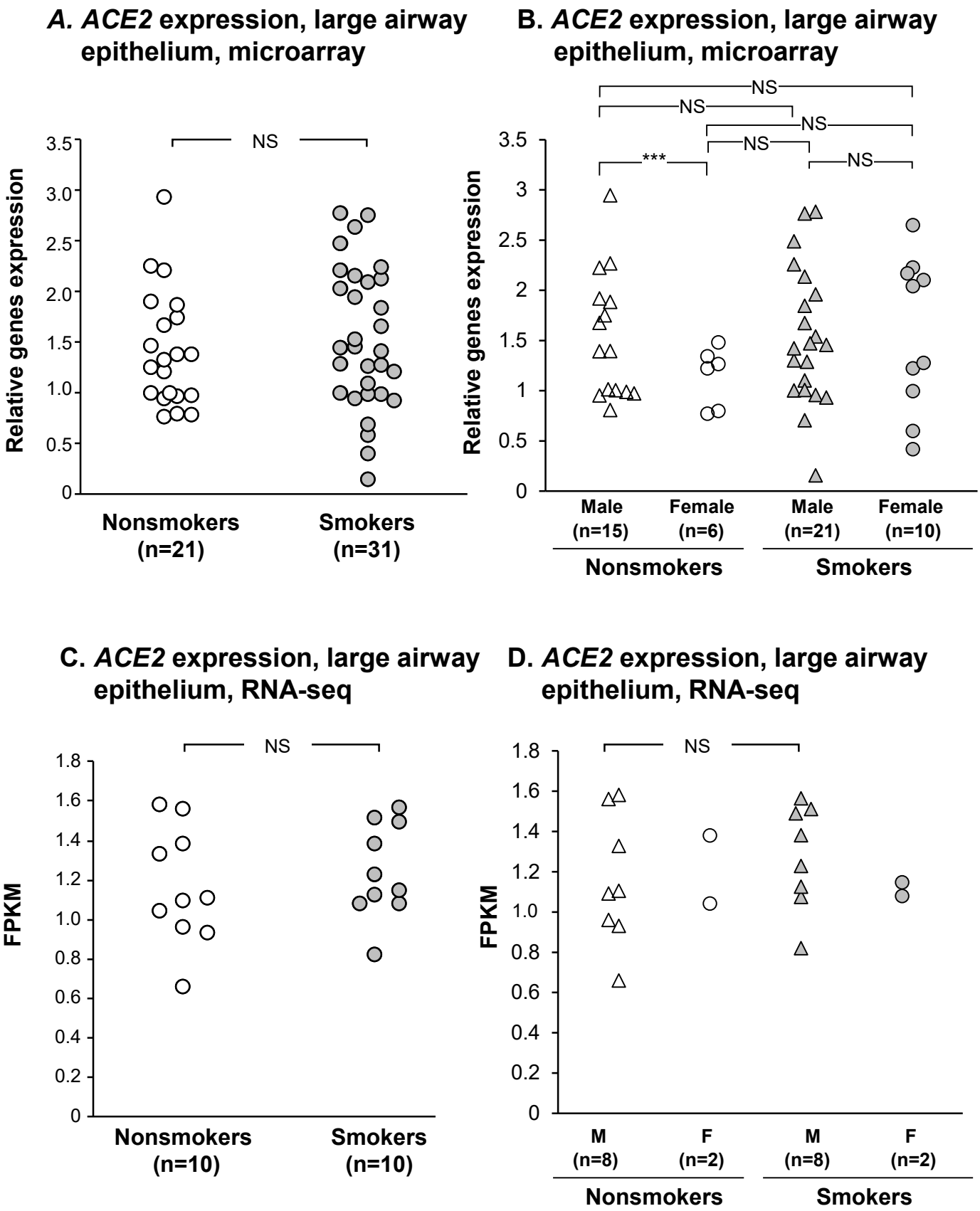


B. *ACE2* expression in large airway (Duclos et al.)



AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

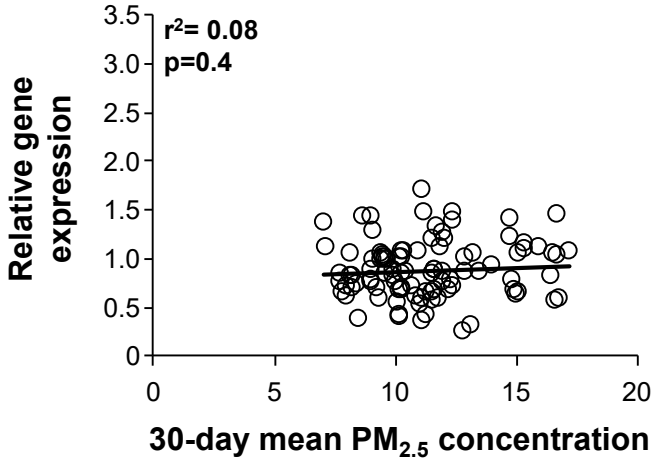
Figure E3



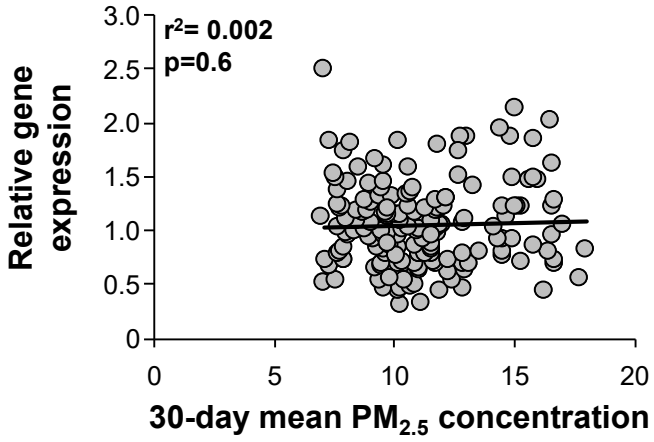
AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC

Figure E4

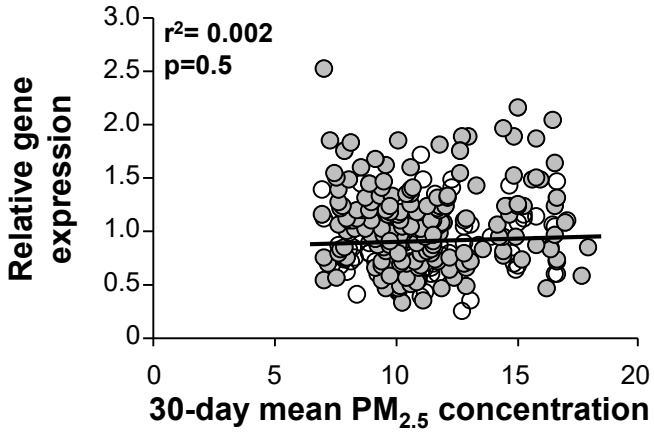
A. Correlation of monthly pollution concentration and *ACE2* expression in healthy nonsmokers



B. Correlation of monthly pollution concentration with *ACE2* expression in healthy smokers



C. Correlation of monthly pollution concentration and *ACE2* expression in healthy nonsmokers and smokers



AJRCCM Articles in Press. Published May 20, 2020 as 10.1164/rccm.202003-0541OC