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

Haijun Zhang, MD, PhD, Mahboubeh R. Rostami, Phd, Philip L. Leopold, PhD,

Jason G. Mezey, PhD, Sarah L. O'Beirne, MD, PhD, Yael Strulovici-Barel, PhD,

and Ronald G. Crystal, MD

ONLINE DATA SUPPLEMENT

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-

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

Healthy nonsmokers and phenotypic healthy 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-

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 microRNA (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

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 smoker males and females

were also matched for pack-yr. 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 smoker males and females were also matched for pack-yr. 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 10th to 12th airway fiberoptic bronchoscopy (3). Brushed cells were treated with ACK lysing buffer to re-

move red blood cells and treated with 0.05% trypsin-ethylenediaminetetraacetic acid (GIBCO, 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 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*. Mucus 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 mucus 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* nonsmokers.

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 fil-

ters 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.

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 pseudo-stratified airway epithelium in PneumaCult-ALI medium (Stemcell Technologies) as previously described (14). Briefly, 24 hr after seeding BC onto the filter, the medium was replace with fresh PneumaCult Ex Plus growth medium; after additional 24 hr, medium was removed from the apical chamber to create ALI conditions, and the basal medium was replaced with to PneumaCult-ALI differentiation medium (15). 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 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 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 influ-

enced by brief fluctuations in $PM_{2.5}$ concentrations and daily variations in individuals' exposure to outdoor air.

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.

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.

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	Unpublished	(6)
	SAE	Smoker	73	HG-U133	Unpublished	(6)
Figure 2C, D	SAE	Nonsmoker	20	Illumina HiSeq 2500	Unpublished	Unpublished
	SAE	Smoker	23	Illumina HiSeq 2500	Unpublished	Unpublished
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	Unpublished	Unpublished
	SAE	Smoker	23	Illumina HiSeq 2500	Unpublished	Unpublished
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	Unpublished	Unpublished
	LAE	Smoker	10	Illumina HiSeq 2500	Unpublished	Unpublished
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 #.

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.

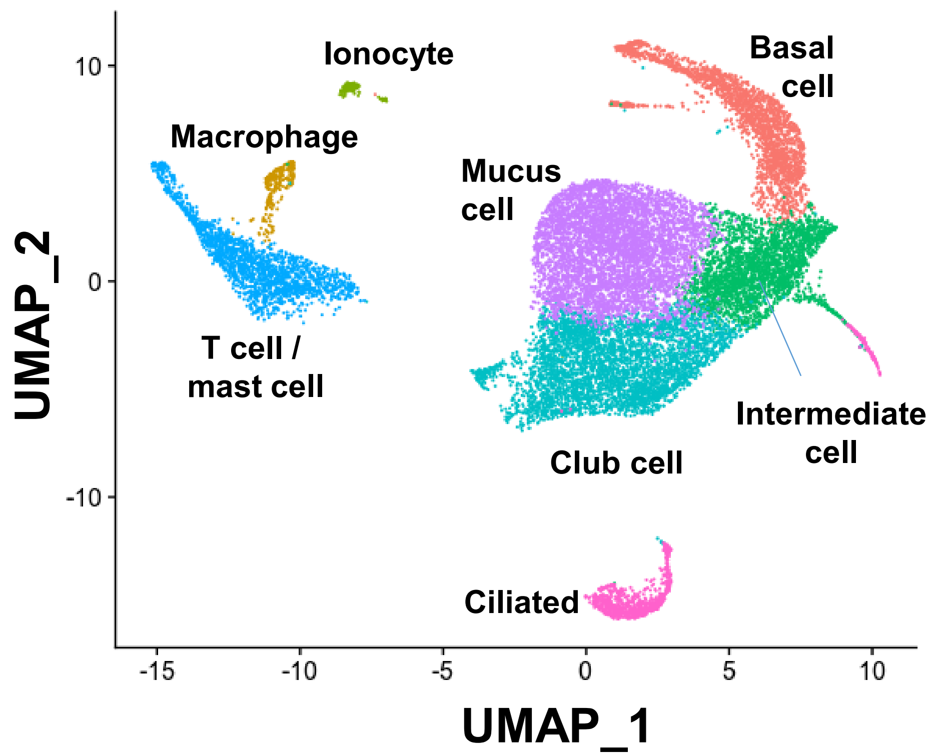
Figure E2. Expression of *ACE2* in alveolar epithelium type 2 (AT2) cells. **A.** Analysis of the single cell 10x transcription data of Reyfman *et al* (18), GEO accession number 122960. The data was generated from samples of lung parenchyma from 8 control lung biopsies from lung donors with no pulmonary 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

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

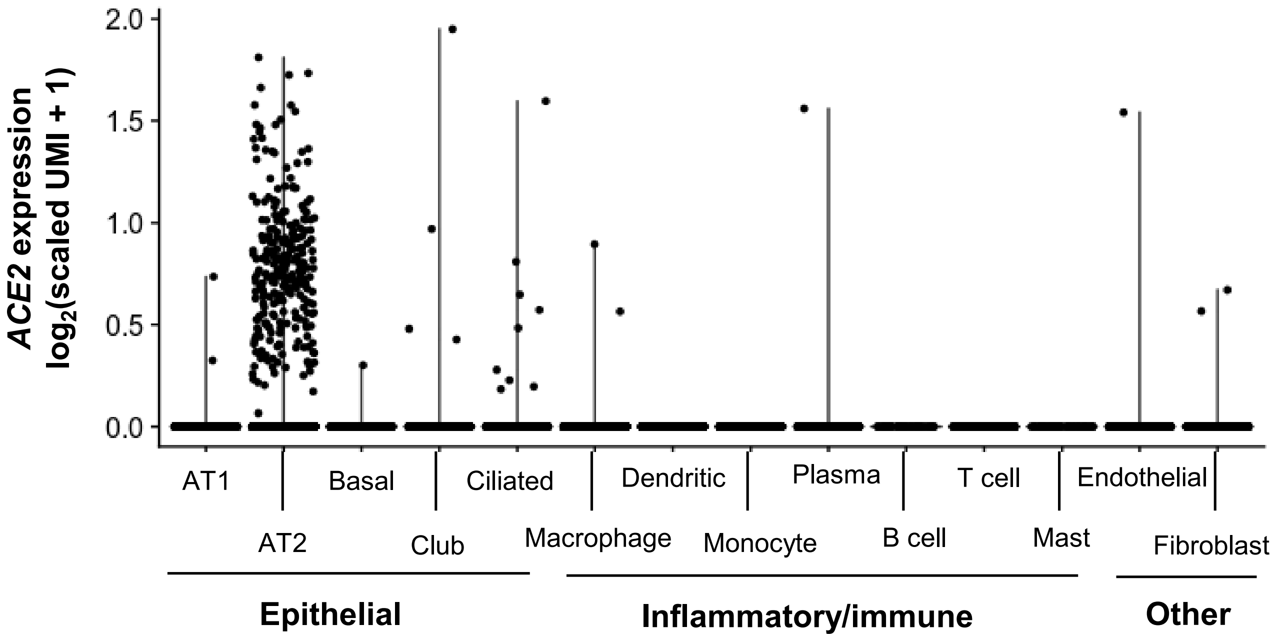
Figure E3. Expression of *ACE2* in the large airway epithelium (LAE) of nonsmokers compared to smokers. **A.** LAE, nonsmokers (n=21) *vs* smokers (n=31), Affymetrix HG-U133 Plus 2.0 microarray, analysis based on the data of Vanni *et al* (4) GEO accession number 10135. **B.** LAE, effect of sex and 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; *** = $p<0.001$. Open symbols – nonsmokers, closed symbols – smokers. FPKM = fragments per kilobase of exon per million fragments sequenced.

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. **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 symbols – nonsmokers, closed symbols – smokers.

Figure E1



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



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

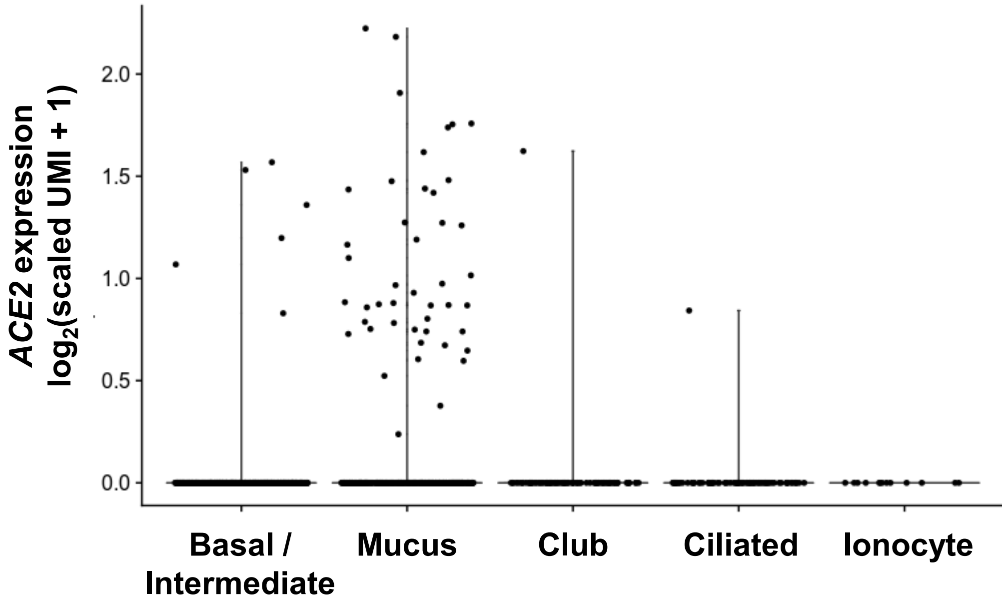


Figure E3

A. *ACE2* expression, large airway epithelium, microarray

Relative genes expression

NS

Nonsmokers (n=21)

Smokers (n=31)

B. *ACE2* expression, large airway epithelium, microarray

Relative genes expression

NS NS NS NS NS ***

Male (n=15) Female (n=6)

Nonsmokers

Male (n=21) Female (n=10)

Smokers

C. *ACE2* expression, large airway epithelium, RNA-seq

FPKM

NS

Nonsmokers (n=10)

Smokers (n=10)

D. *ACE2* expression, large airway epithelium, RNA-seq

FPKM

NS

M (n=8) F (n=2)

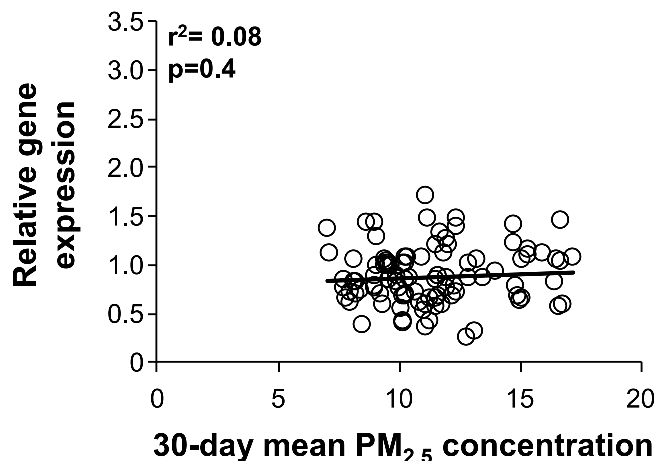
Nonsmokers

M (n=8) F (n=2)

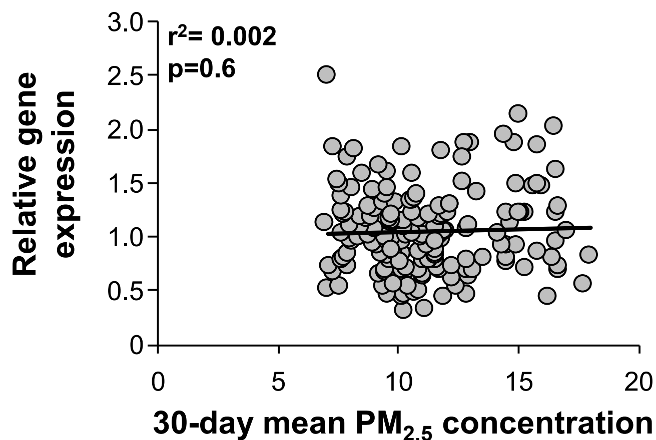
Smoers

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