Smoking Modulates Different Secretory Subpopulations Expressing SARS-CoV-2 Entry Genes in the Nasal and Bronchial Airways

Ke Xu
Boston University School of Medicine
Xingyi Shi
Boston University School of Medicine
Christopher Husted
Boston University School of Medicine
Rui Hong
Boston University School of Medicine
Yichen Wang
Boston University School of Medicine
Boting Ning
Boston University School of Medicine
Travis Sullivan
Lahey Hospital and Medical Center
Kimberly Rieger-Christ
Lahey Hospital and Medical Center
Fenghai Duan
Brown University School of Public Health
Helga Marques
Brown University School of Public Health
Adam Gower
Boston University School of Medicine
Xiaohui Xiao
Boston University School of Medicine
Hanqiao Liu
Boston University School of Medicine
Gang Liu
Boston University School of Medicine
Grant Duclos
Boston University School of Medicine
Michael Platt

Boston University School of Medicine
Avrum Spira
Boston University School of Medicine
Sarah Mazzilli
Boston University School of Medicine
Ehab Billatos
Boston University School of Medicine
Marc Lenburg
Boston University School of Medicine
Joshua Campbell
Boston University School of Medicine
Jennifer Beane (jbeane@bu.edu)
Boston University School of Medicine

Research Article

Keywords: Impact of smoking on SARS-Co-V2 infection, nasal and bronchial airways, genes

Posted Date: October 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-887718/v1

Abstract

Background: SARS-CoV-2 infection and disease severity are influenced by viral entry (VE) gene expression patterns in airway epithelium. The similarities and differences of VE gene expression (ACE2, TMPRSS2, and CTSL) across nasal and bronchial compartments has not been fully characterized using matched samples from large cohorts.

Results: Gene expression data from 793 nasal and 1,673 bronchial brushes obtained from individuals participating in lung cancer screening or diagnostic workup revealed that smoking was the only clinical factor significantly and reproducibly associated with VE gene expression. ACE2 and TMPRSS2 expression were higher in smokers in the bronchus but not in the nose. scRNA-seq of nasal brushings indicated that ACE2 co-expressed genes were highly expressed in club and C15orf48^{+} secretory cells while TMPRSS2 co-expressed genes were highly expressed in keratinizing epithelial cells. In contrast, these ACE2 and TMPRSS2 modules were highly expressed in goblet cells in scRNA-seq from bronchial brushings. Cell-type deconvolution of the RNA-seq confirmed that smoking increased the abundance of several secretory cell populations in the bronchus, but only goblet cells in the nose.

Conclusions: The association of ACE2 and TMPRSS2 with smoking in the bronchus is due to their high expression in goblet cells which increase in abundance in current smoker airways. In contrast, in the nose these genes are not predominantly expressed in cell populations modulated by smoking. Smoking-induced VE gene expression changes in the nose likely has minimal impact on SARS-CoV-2 infection, but in the bronchus, smoking may lead to higher viral loads and more severe disease.

Background

As of August 14th, 2021, over 205 million confirmed cases and 4.3 million deaths have been reported globally for COVID-19 (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). SARS-CoV-2 infects cells by utilizing host cell-surface proteins for viral entry (VE). VE proteins include angiotensin-converting enzyme 2 (ACE2), which provides a binding site for the virus; proteases such as transmembrane protease serine 2 (TMPRSS2) that can cleave the viral S glycoprotein; and cathepsin L (CTSL), which can also aid with S glycoprotein priming[1–4]. Prior studies have attempted to establish associations between the expression of VE genes and clinical variables to identify factors that may influence infection incidence or COVID-19 severity.

Several studies have reported conflicting findings regarding associations between VE gene expression and clinical factors due to small sample sizes and exclusion of important covariates. Smith et al[5], Wang et al[6], and Brake et al[7] reported increase in ACE2 expression with cigarette smoke in the tracheal epithelium, while Zhang et al[8] and Aliee et al[9] reported that this observation was restricted to the small airway epithelium. Bunyavanich et al[10] reported that ACE2 expression was associated with age; however, Smith et al[5] found that ACE2 expression was equivalent between young and elderly individuals. Moreover, earlier studies with limited numbers of subjects led to the conclusion that race was associated with ACE2

expression[11], which was not found by Smith et al in a larger dataset[5]. Lung airway expression of both ACE2 and TMPRSS2 was also found to be significantly up-regulated in patients with chronic obstructive pulmonary disease (COPD) compared with healthy subjects or in smokers compared with non-smokers[12] without adjustment of important factors such as age and sex. The goal of this study is to establish associations between VE gene expression and clinical factors leveraging a large number of airway samples with extensive clinical data, matched between nasal and bronchial compartment from multiple independent cohorts, and to identify the biological processes and cell types associated with the expression of VE genes in the nasal and bronchial compartments.

Results

Gene expression profiles of 796 nasal and 1,673 bronchial brushing samples were generated using bulk RNA-seq from 4 cohorts, DECAMP[13] (Detection of Early Lung Cancer Among Military Personnel), AEGIS[14] (Airway Epithelial Gene expression in the Diagnosis of Lung Cancer), BCLHS[15] (British Columbia Lung Health Study and Pan-Canadian Lung Health Study), and PCA[16] (Pre-Cancer Atlas) (Table 1). Study participants were current or former smokers undergoing bronchoscopy as part of either a diagnostic workup or a screening process for lung cancer. Clinical data collected on samples included smoking status, sex, age, forced expiratory volume during the first second (FEV1) percent Predicted, and, in the AEGIS study, lung cancer status (**Supplemental Tables 1**). Notably, we observed different patterns of association with smoking between ACE2 and TMPRSS2 gene expression between the nasal and bronchial epithelium. ACE2 was down-regulated in the nasal epithelium of current smokers in 1 out of 2 cohorts ($p < 0.01$) but strongly up-regulated in the bronchus in all four cohorts ($p < 0.01$; Figs. 1a and 1b; **Supplemental Fig. 1a and 1b**). TMPRSS2 expression was not associated with smoking in the nose but its expression was up-regulated in current smokers in the bronchus in 3 out of 4 cohorts ($p < 0.001$). CTSL expression was down-regulated in current smokers in the nasal epithelium in 1 out of 2 cohorts ($p < 0.01$) and the bronchial epithelium in 3 out of 4 cohorts ($p < 0.05$). These genes were not strongly associated with other clinical variables (sex, age, FEV1% Predicted) across cohorts in either the nasal or bronchial samples in contrast to some previous studies[5,8,11,12]. Adjusting for lung cancer diagnosis in the AEGIS cohort did not change these observed correlations, although positive lung cancer status negatively correlated with ACE2 expression in the nasal epithelium ($p<0.05$, **Supplemental Fig. 1b**). Upon examining paired nasal and bronchial samples from the same subjects, a lack of correlation in the expression of each VE gene was observed between the two tissues ($p > 0.05$; Fig. 1c; **Supplemental Fig. 1c, Supplemental Table 3**), suggesting that the biological processes that influence the expression of these genes may differ across airway sites.

Table 1
Summary of Patient Demographics across Sample Cohorts

	DECAMP Nose	AEGIS Nose	DECAMP Bronchus	AEGIS Bronchus	BCLHS Bronchus	PCA Bronchus
Total number of samples/cells	288	505	360	938	238	137
Smoking Status (%)						
Current	125 (43.40)	186 (36.83)	153 (42.5)	442 (47.12)	99 (41.6)	77 (56.2)
Former	156 (54.17)	319 (63.17)	202 (56.11)	496 (52.88)	139 (58.4)	60 (43.8)
Never	0	0	0	0	0	0
Missing	7 (2.43)	0	5 (1.39)	0	0	0
Sex, n (%)						
Male	219 (76.04)	317 (62.78)	282 (78.33)	584 (62.26)	135 (56.72)	70 (51.09)
Female	69 (23.96)	188 (37.23)	78 (21.67)	354 (27.74)	103 (43.28)	67 (48.91)
Missing	0	0	0	0	0	0
Age						
Mean (SD)	67 (8)	60 (11)	65 (7)	63 (11)	65 (6)	58 (7)
Missing	0	0	0	0	0	0
FEV1 (% predicted)						
Mean (SD)	75.32 (19.15)	73.00 (21.28)	72.71 (20.26)	68.73 (22.18)	80.86 (20.68)	74.32 (20.72)
Missing	28	355	0	633	0	4
Having Lung Cancer (%)						
Yes	NA	309 (61.19)	NA	710 (75.69)	NA	NA
No	NA	196 (38.81)	NA	228 (24.31)	NA	NA
Missing	NA	0	NA	0	NA	NA

	DECAMP Nose	AEGIS Nose	DECAMP Bronchus	AEGIS Bronchus	BCLHS Bronchus	PCA Bronchus

Table 1. Summary of Patient Demographics. Definition of abbreviations: DECAMP = Detection of Early Lung Cancer Among Military Personnel, AEGIS = Airway Epithelial Gene expression in the Diagnosis of Lung Cancer, BCLHS = British Columbia Lung Health Study and Pan-Canadian Lung Health Study, PCA = Pre-Cancer Atlas. Values are presented as mean (SD).

To further explore the biological pathways associated with VE genes in each site of the airway, we developed co-expression networks using Weighted Gene Correlation Network Analysis (WGCNA)[17]. We identified 58 and 48 gene modules within the nasal and bronchial cohorts, respectively (Fig. 2a and 2b, **Supplemental Fig. 2; Supplemental Table 2**). The consensus module eigengenes containing the VE genes significantly correlated with their respective VE gene expression within each cohort (**Supplemental Fig. 3a**, Pearson correlation, $p < 2.2E-16$, average $R = 0.584$), suggesting that the eigengene can be used as a surrogate for each VE gene. The VE gene module eigengenes showed similar associations with the clinical covariates (**Supplemental Fig. 3b**) to those presented in Figure 1 to those presented in Figure 1. The nasal VE gene module eigengenes were not highly correlated, however, the ACE2- and TMPRSS2-associated module eigengenes in the bronchus were highly correlated across all four datasets (**Supplemental Figs. 2a and 2b**), suggesting that similar biological processes control these two gene modules in the bronchial compartment.

Comparing biological pathways enriched among VE gene modules between the nose and bronchus revealed (Fig. 2c; **Supplemental Table 3**) that genes of the nasal ACE2 module were enriched in the inflammatory response, interferon-alpha signaling, and interferon-gamma signaling pathways while the bronchial ACE2 gene module was enriched in genes associated with protein secretion and androgen response. Genes of the TMPRSS2 module in the nose were enriched in estrogen response and KRAS signaling pathways while the module in the bronchial epithelium was enriched in MTORC1 and TNFα-NFκB signaling pathways. The lack of shared biological pathways mirrors the small number of genes shared by the nasal and bronchial modules for ACE2 ($n = 5$) and TMPRSS2 ($n = 12$). In contrast, the nasal and bronchial CTSL modules shared 177 genes (Fisher's exact test, $p = 2.23E-10$) and were both enriched in genes related to inflammatory response and TNF-alpha signaling pathways. Furthermore, in the bronchus, ACE2 and TMPRSS2-associated modules share enrichment in genes involved in the p53 pathway, adipogenesis, androgen response, MTORC1 signaling, and estrogen response (**Supplemental Fig. 3c**). We also did not observe significant correlations between the eigengenes of each VE gene module in the paired nasal and bronchial samples (Fig. 2d and **Supplemental Fig. 3d**). The VE gene module and single-gene analyses are consistent and suggest that there are important differences between VE gene expression and biology between the upper and lower airways.

To determine if different cell populations contribute to the difference in VE gene expression between airway compartments, we leveraged single-cell RNA sequencing (scRNA-seq) data to characterize the expression patterns of individual VE genes and VE gene modules across major epithelial and immune cell

types. We profiled 34,833 cells from 9 nasal brushings (4 collected from two volunteers and 5 from 5 patient undergoing lung cancer screening) and 2,075 cells from 17 bronchial brushings from patients undergoing bronchoscopy for suspicion of lung cancer (**Supplemental Tables 6, 7**). A total of 17 and 15 transcriptionally distinct cell clusters were identified in the nose and bronchus, respectively, many of which expressed similar marker genes: KRT5, KRT15 (basal cells); FOXJ1, C20orf85, CDC20B (ciliated cells); SCGB1A1, SERPINB3 (club cells); MUC5AC, TFF1, TFF3 (goblet cells); FOXI1, CFTR (ionocytes); CD3D, CD8A (T cells). We also identified two bronchial goblet-like secretory populations: one characterized by high expression of CEACAM5 but not MUC5AC, previously described as peri-goblet cells[18], and the other by high expression of HLA-DQA1 and MHC class II genes (Figs. 3a-b, **Supplemental Fig. 4**). In the nose, we discovered two club-cell-like secretory cell clusters (STATH+ and C15orf48+), as well as a novel cell cluster of "keratinizing epithelial cells" expressing genes involved in cornification, epidermis development, and keratinocyte differentiation, such as SPRR3 and SPRR2A. Immune populations represent <1% and 26% of total cell populations in the nasal and bronchial epithelium, respectively. In both nasal and bronchial datasets, the cell subpopulations were consistently observed across multiple patients (**Supplemental Fig. 5**).

We found low expression of the VE genes in the single cell data, consistent with other reports (**Supplemental Fig. 6**)[19,20], so we calculated VE gene module metagene scores in each cell to understand how biological processes associated with the VE genes in the bulk RNA-seq data are distributed across nasal and bronchial cell populations (Figs. 3c-e). In the nose, the ACE2 module was moderately expressed across many cell types but was most highly expressed in C15orf48+ secretory cells and club cells. The TMPRSS2 module was predominantly expressed in the keratinizing epithelial cells, followed by C15orf48+ secretory cells. The expression of gene modules of ACE2 and TMPRSS2 is different from the gene expression pattern reported by Sungnak et al in that TMPRSS2 gene expression was limited to ACE2+ cells[21]. In the bronchus, both the ACE2 and TMPRSS2 modules were expressed at the highest levels in the goblet cell population, in agreement with previous findings[22]. The highest median metagene score for the CTSL module was in the neutrophils and macrophages in both the nose and bronchus; however, expression was also high in T cells in the nose and dendritic cells in the bronchus. We found very few cells co-expressing ACE2, CTSL, and TMPRSS2 at high levels – 114 out of 34833 cells in the nose and no cells in the bronchus. On the other hand, we found that ACE2 and TMPRSS2 modules were more likely to be highly co-expressed in nasal keratinizing epithelial cells (odds ratio = 7.56), nasal C15orf48+ secretory cells (N = 518, odds ratio = 7.36), and bronchial goblet cells (odds ratio = 16.80) (N = 806, **Supplemental Table 6**, FDR q < 0.001). Thus, ACE2 and TMPRSS2 were found to be expressed in different nasal and bronchial cell populations, which may be influenced by smoking in ways that lead to their divergent correlation with smoking in the two airway compartments.

To investigate how smoking modulates different cell populations, we computationally deconvolved cell population proportions in the bulk RNA-seq data using gene markers identified from the single-cell data (**Supplemental Table 7**). Higher proportions of goblet cells were observed in current smokers in both the nose and bronchus in all 6 cohorts, consistent with the previous studies[18]. Increased proportions of

ionocytes in current smokers were also observed in the nose (1 out of 2 cohorts) and bronchus (all 4 cohorts). The proportion of ciliated cells was significantly lower in smokers in all 6 cohorts (Figs. 4a-b, **Supplemental Figs. 7a-d**, $p < 0.05$). While goblet cell proportions were significantly higher in smokers in both the nose and bronchus, the ACE2 module was highly expressed only in bronchial, not nasal goblet cells. These results may explain the lack of association of ACE2 expression with smoking in the bulk RNA-seq nasal data. Additionally, in the AEGIS nasal data (**Supplemental Fig. 7**) the fraction of keratinizing epithelial cells is increased and STATH+ secretory cells is decreased in current smokers, suggesting that shifts in these populations may be partially responsible for the differential correlation with smoking between ACE2 and TMPRSS2 in the nose. For CTSL, both nasal datasets showed a significant decrease in the proportion of the neutrophil/macrophage population with respect to smoking. While this may partially explain the down-regulation of CTSL in smokers in the nasal bulk RNA-seq data, the overall proportions of these populations estimated in the bronchial data by the deconvolution were close to zero for most samples, which could have prevented us from confirming the decrease of this population with smoking in the bronchus.

Discussion

The current study investigated the relationship between the expression of genes encoding proteins important for SARS-CoV-2 entry (ACE2, CTSL, and TMPRSS2) and clinical factors including age, sex, COPD, and smoking in both the nose and bronchus. In our datasets, only smoking status showed a consistent association with the expression of VE genes across cohorts within each airway compartment. In the bronchus, current smoking status was associated with higher ACE2 and TMPRSS2 expression; however, in the nose, current smoking status was inversely correlated with ACE2 expression and was not associated with TMPRSS2 expression as in prior studies [5–7, 10]. The cohorts used in this study contain subjects at high risk for developing lung cancer and are thus composed of older subjects (mean age = 63, SD = 8); therefore, we did not find an association between nasal ACE expression and age as previously reported by Bunyavanich et al[8] in a study comparing children and adults. We also did not find sex to be associated with ACE2 expression as was reported by another study with a lower sample size[9] and could not confidently assess differences in expression profiles between racial subgroups as our cohorts are predominantly (> 70%) composed of White subjects. Our analyses did not find significant VE gene expression correlations between paired nasal and bronchial samples suggesting that there are differences in biological pathways or cell types between the airway compartments.

Our WGCNA analysis demonstrated that genes co-expressed with ACE2 and TMPRSS2 in the bulk RNA-seq data were more highly correlated to one another in the bronchus than in the nose and that pathway enrichment was different for these modules in the upper and lower airways. Similarly, scRNA-seq data showed that ACE2 and TMPRSS2 modules were co-expressed in bronchial goblet cells but had distinct patterns of expression across nasal cell populations. Our high-resolution nasal scRNA-seq dataset containing over 30,000 cells allowed us to differentiate between secretory cell populations to show that C15orf48+ and keratinizing epithelial cells had the highest expression of ACE2 and TMPRSS2 modules,

respectively. In contrast, we found that bronchial goblet cells had the highest expression of both ACE2 and TMPRSS2 modules, and only the goblet cell population was increased in smokers in our deconvolution analysis. Of note, ACE2 and TMPRSS2 modules were also co-expressed highly within the peri-goblet cells in the bronchus, which was also observed by Lukassen et al[20].

Interestingly, the nasal and bronchial CTSL modules showed enrichment of immune-associated biological programs and were highly expressed in similar immune cell populations. Despite these similarities, CTSL gene or module expression was not correlated between paired bronchial and nasal samples potentially due to the low abundance of immune populations in the scRNA-seq and bulk RNA-seq data.

Conclusions

Our study leveraged bulk and single-cell gene expression profiles from large cohorts to show that current cigarette smoking status is consistently associated, in a site-specific manner, with the expression of genes required for SARS-CoV-2 entry in the nose and bronchus. This difference of the association of ACE2 and TMPRSS2 expression with smoking between the nasal and bronchial compartments is likely due to the expression patterns of these genes in distinct cell subpopulations in the nose and bronchus, as well as the way the proportion of these subpopulations change with respect to smoking between the two compartments. Future work investigating other putative viral entry genes such as FURIN and ATRNL1[20–22] may build upon these findings to enhance our understanding of the effect of smoking on SARS-CoV-2 infection. The results of our study of the expression of ACE2 and TMPRSS2 suggest that smoking is unlikely to impact the likelihood of SARS-CoV-2 infection in the upper airways, but that it may play a significant role in COVID-19 disease progression and severity.

Methods

Datasets with gene expressions profiled by bulk RNA-sequencing or microarray.

Datasets of nasal and bronchial brushings from 4 published studies were analyzed, which include DECAMP (Detection of Early lung Cancer among Military Personnel)[13], AEGIS (Airway Epithelial Gene expression In the diagnosis of lung cancer)[14], BCLHS (British Columbia Lung Health Study and pan-Canadian lung health study)[15], and PCA (Pre-Cancer Atlas) (Table 1)[16]. These samples were obtained from participants with increased risks for lung cancer and were undergoing either lung cancer screening or bronchoscopy for suspicion of lung cancer. The number of samples included for each analysis was listed in Supplemental Table 1.

Datasets with gene expressions profiled by single-cell RNA-sequencing.

Cells from the inferior turbinate of the nose were collected as part of a study involving participants with indeterminate pulmonary nodules undergoing lung cancer screening at Boston Medical Center and Lahey Hospital Medical Center. Using a lab-developed protocol, cells were harvested from the nasal swabs and resuspended into single cells. A red cell lysis step was performed to eliminate most red blood cells. Samples with over 85% viability were prepared for single-cell sequencing using the 10X Genomics Platform. Cells from the main stem bronchus were brushed as part of a study involving participants undergoing diagnostic bronchoscopy for suspected lung cancer at Boston Medical Center. Cells were resuspended and sorted into 96-well PCR plates before being processed using the CEL-Seq2 RNA library preparation protocol[23]. The cDNA libraries were sequenced on an Illumina NextSeq500 in High-Output mode (75 nucleotide paired-end). The tissues were collected under the protocols approved by the Institutional Review Board (IRB) at Boston University School of Medicine and written consents were obtained from participants involved in these studies.

Derivation of consensus gene modules and their pathway enrichment.

Batch corrected and library normalized counts were used for generating the consensus gene modules using the WGCNA package[24]. Functional analysis of the genes within each of the VE gene modules was performed using MSigDB hallmark pathway gene sets[25].

Correlations between VE gene/gene modules and the clinical variables.

For the RNA-seq datasets, expression of the VE genes was correlated to clinical covariates using linear regression model via the Limma package[26] on voom-transformed data. Of note, for the PCA dataset, patient ID was included as a random effect as each participant contributed several samples. Log2 expression values of the VE genes from the microarray datasets were correlated to the clinical variables directly without voom transformation. Module eigengenes for each VE gene consensus module were used to correlate with the clinical variables using the same methodology.

Single-cell analysis of nasal and bronchial samples.

Preprocessed count matrices of the nasal and bronchial samples were pre-processed using the Scruff package[27] and analyzed using the Seurat 3.0[28] with standard settings. Quality Control of the nasal and bronchial count matrices was performed using SCTK-QC pipeline[29,30,31]. Uniform Manifold

Approximation and Projection (UMAP) was used for dimension reduction and visualizing relationships amongst cells. Canonical marker genes were utilized to identify cell types when possible. Library normalized counts were used for generating plots with single gene expression. And the meta-gene score for each VE module was calculated by averaging the module genes across normalized expressions. A high expression for a gene module is defined by an expression greater than one standard deviation above the mean.

Deconvolution of bulk RNA-seq gene expression.

Estimated proportions of cells in each of the bulk RNA samples were computed using the Autogenes Package[32] with reference gene expression profiles (GEPs) derived from the nasal and bronchial scRNA-seq datasets.

Declarations

Ethics approval and consent to participate

The DECAMP study was approved by the Human Research Protection Office (HRPO) of the Department of Defense, and the individual site IRBs for every participating site. All subjects were approved for written informed consent to participate in the study in accordance with IRB regulations. The generation of nasal scRNA-seq data was approved by the IRBs at Boston University School of Medicine and Lahey hospital. The generation of bronchial scRNA-seq data was approved by the IRB at Boston University School of Medicine. Samples were collected according to the study protocol. Written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are being prepared for submission to the Gene Expression Omnibus (GEO).

Competing interests

Avrum Spira is an employee of Johnson and Johnson Services, Inc, and has received personal fees from Veracyte Inc outside the submitted work. Marc E. Lenburg reports grants and personal fees from Johnson and Johnson and is a shareholder in Metera Pharmaceuticals; he also has a patent US PTO 9,677,138

issued. Grant Duclos is now an employee of AstraZeneca. None declared (Ke Xu, Xingyi Shi, Christopher Husted, Rui Hong, Yichen Wang, Boting Ning, Travis Sullivan, Kimberly Rieger-Christ, Fenghai Duan, Helga Marques, Adam C. Gower, Xiaohui Xiao, Hanqiao Liu, Gang Liu, Michael Platt, Sarah Mazzilli, Ehab Billatos, Joshua D. Campbell, Jennifer Beane).

Funding

The DECAMP study is supported by funds from the Department of Defense (W81XWH-11-2-0161), the National Cancer Institute (U01CA196408), and Johnson and Johnson Services, Inc (JJSI). The SU2C study is supported by a Stand Up To Cancer-LUNGevity-American Lung Association Lung Cancer Interception Dream Team Translational Cancer Research Grant (grant number: SU2C-AACR-DT23-17 to Steven M. Dubinett and Avrum E. Spira). Stand Up To Cancer is a division of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. The IDA study is supported by a Department of Defense Idea Development Award W81XWH-14-1-0234 (Jennifer Beane). Part of the work on method development was funded by the National Library of Medicine (NLM) RO1LM013154-01 (Joshua D. Campbell). This work was also funded in part by the National Cancer Institute (NCI) Human Tumor Atlas Network (HTAN) COVID NOSI 3U2CCA233238-01S1 (Avrum Spira, Jennifer Beane, Joshua D. Campbell, Sarah Mazzilli).

Author contributions

Ke Xu, Xingyi Shi, Christopher Husted, contributed equally to this work (co-first authors); Joshua Campbell, and Jennifer Beane contributed equally to this work (co-senior authors). Ke Xu is the guarantor of this paper and takes full responsibility for the integrity of the work as a whole. Avrum Spira, Marc E. Lenburg, Ehab Billatos, Joshua Campbell, Jennifer Beane, Sarah Mazzillis initiated the study design. Fenghai Duan, Helga Marques, and Adam C. Gower prepared patient characteristics for the DECAMP cohort; Ke Xu, Travis Sullivan, Kimberly Rieger-Christ, Xiaohui Xiao, Hanqiao Liu, Gang Liu, Michael Platt and Grant Duclos collected and processed samples; Ke Xu, Xingyi Shi, Christopher Husted, Rui Hong, Yichen Wang, and Boting Ning analyzed data. Ke Xu, Xingyi Shi, and Christopher Husted prepared the figures; Ke Xu, Xingyi Shi, Christopher Husted, Marc E. Lenburg, Joshua Campbell and Jennifer Beane interpreted results; Ke Xu, Xingyi Shi, and Christopher Husted drafted the manuscript; Ke Xu, Xingyi Shi, Christopher Husted, Marc E. Lenburg, Joshua Campbell, and Jennifer Beane edited and revised manuscript; all the others approved the final version of the manuscript.

Acknowledgments

We would like to thank Yuriy Alekseyev, Ashley LeClerc, and Kangning Zhang of the Single Cell Sequencing Core at Boston University School of Medicine for their help with data generation.

References

1. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor., **181**, 271–2808 (2020).
2. Jia, H. P. *et al.* ACE2 Receptor Expression and Severe Acute Respiratory Syndrome Coronavirus Infection Depend on Differentiation of Human Airway Epithelia. *J. Virol*, **79**, 14614–14621 (2005).
3. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation., **367**, 1260–1263 (2020).
4. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus.J. Virol.94, (2020).
5. Smith, J. C. *et al.* Cigarette Smoke Exposure and Inflammatory Signaling Increase the Expression of the SARS-CoV-2 Receptor ACE2 in the Respiratory Tract. *Dev. Cell*, **53**, 514–5293 (2020).
6. Wang, J., Luo, Q., Chen, R., Chen, T. & Li, J. Susceptibility Analysis of COVID-19 in Smokers Based on ACE2. (2020) doi:10.20944/preprints202003.0078.v1.
7. Brake, S. J. *et al.* Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus SARS-CoV-2 (Covid-19). *J. Clin. Med*, **9**, 841 (2020).
8. Zhang, H. *et al.* Expression of the SARS-CoV-2 ACE2 Receptor in the Human Airway Epithelium. *Am. J. Respir. Crit. Care Med*, **202**, 219–229 (2020).
9. Aliee, H. *et al.* Determinants of SARS-CoV-2 receptor gene expression in upper and lower airways. medRxiv(2020) doi:10.1101/2020.08.31.20169946.
10. Bunyavanich, S., Do, A. & Vicencio, A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. *JAMA*, **323**, 2427 (2020).
11. Zhao, Y. *et al.* Expression Profiling of ACE2, the Receptor of SARS-CoV-2. Am J Respir Crit Care Med 2020 Sep 1;202(5):756-759. doi: 10.1164/rccm.202001-0179LE.
12. Saheb Sharif-Askari, N. *et al.* Airways Expression of SARS-CoV-2 Receptor, ACE2, and TMPRSS2 Is Lower in Children Than Adults and Increases with Smoking and COPD. Mol Ther Methods Clin Dev. 2020 May 22;18:1–6. doi: 10.1016/j.omtm.2020.05.013.
13. Billatos, E. *et al.* Detection of early lung cancer among military personnel (DECAMP) consortium: study protocols. *BMC Pulm Med*, **19**, 59 (2019).
14. Silvestri, G. A. *et al.* A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N. Engl. J. Med*, **373**, 243–251 (2015).
15. Steiling, K., Lenburg, M. E. & Spira, A. Airway gene expression in chronic obstructive pulmonary disease. Proc. Am. Thorac. Soc. 6, 697–700(2009).
16. Beane, J. E. *et al.* Molecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. *Nat. Commun*, **10**, 1856 (2019).

17. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559 https://doi.org/10.1186/1471-2105-9-559 (2008).
18. Duclos, G. E. *et al.* Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Sci. Adv*, **5**, eaaw3413 (2019).
19. Qi, F., Qian, S., Zhang, S. & Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun*, **526**, 135–140 (2020).
20. Lukassen, S. *et al.* SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J*, **39**, e105114 (2020).
21. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med*, **26**, 681–687 (2020).
22. Xia, S. *et al.* The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin.Signal Transduct. Target. Ther.5, (2020).
23. Duclos, G. E. *et al.* Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Sci Adv*, **5** (12), eaaw3413 https://doi.org/10.1126/sciadv.aaw3413 (2019).
24. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559 https://doi.org/10.1186/1471-2105-9-559 (2008).
25. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinforma Oxf Engl*, **27** (12), 1739–1740 https://doi.org/10.1093/bioinformatics/btr260 (2011).
26. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43** (7), e47 https://doi.org/10.1093/nar/gkv007 (2015).
27. Wang, Z., Hu, J., Johnson, E. W. & Campbell, J. D. scruff: An R/Bioconductor package for preprocessing single-cell RNA-sequencing data. *bioRxiv. Published online January*, **16**, 522037 https://doi.org/10.1101/522037 (2019).
28. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, **36** (5), 411–420 https://doi.org/10.1038/nbt.4096 (2018).
29. Hong, R. *et al.* Comprehensive generation, visualization, and reporting of quality control metrics for single-cell RNA sequencing data. bioRxiv 2020.11.16.385328; doi: https://doi.org/10.1101/2020.11.16.385328
30. Yang, S. *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol*, **21**, 57 https://doi.org/10.1186/s13059-020-1950-6 (2020).
31. Jenkins, D. F. *et al.* Interactive single cell RNA-Seq analysis with the Single Cell Toolkit (SCTK). bioRxiv 329755; doi: https://doi.org/10.1101/329755
32. Aliee, H., Theis, F. & AutoGeneS Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *bioRxiv*. Published online February 23, 2020:2020.02.21.940650.

doi:10.1101/2020.02.21.940650

Figures

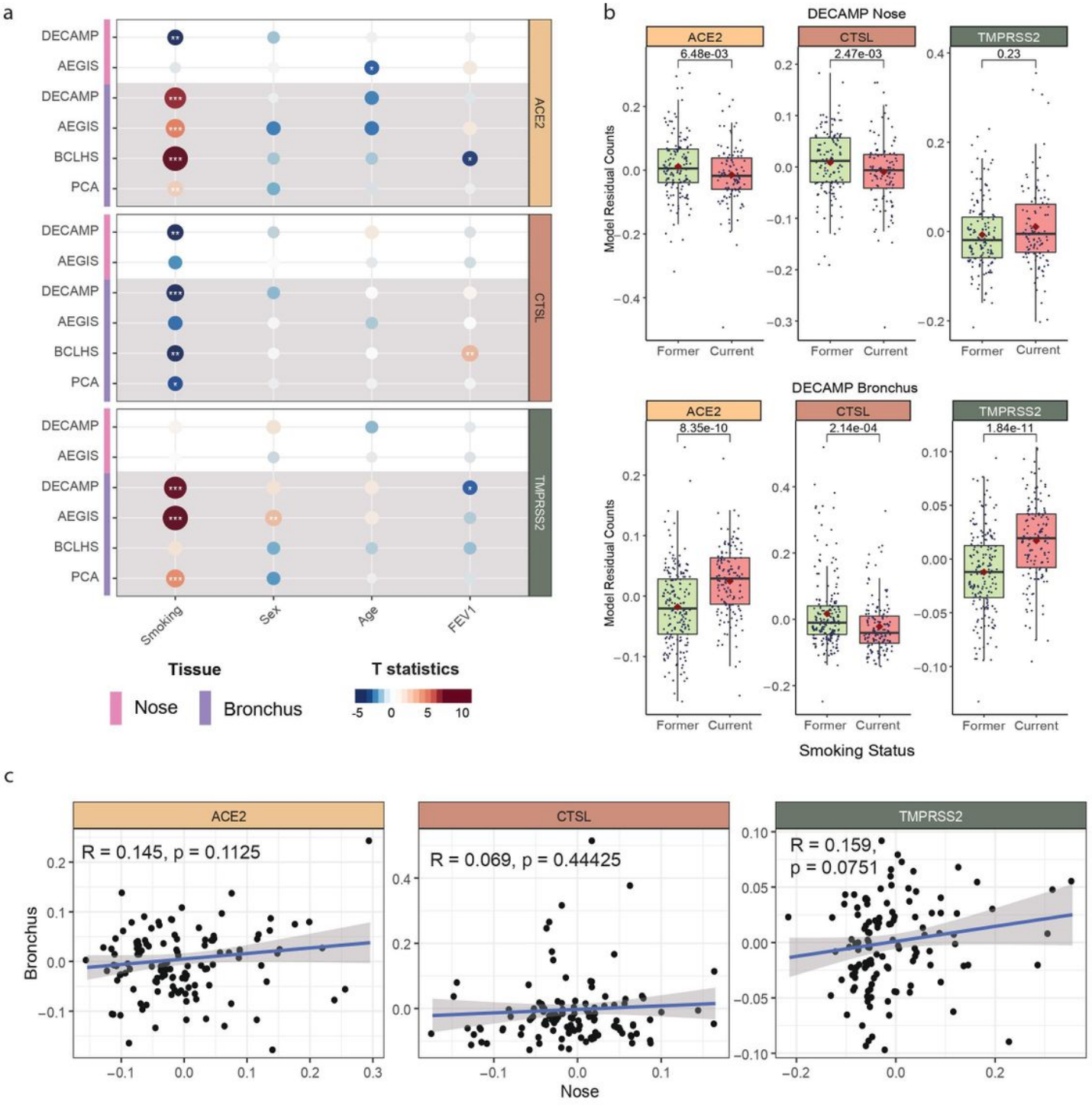

Figure 1

Differential association of VE gene expression between the nose and bronchus with respect to smoking. (a) Associations between the expression of ACE2 (tan), CTSL (brown), and TMPRSS2 (olive) and clinical covariates in the nasal and bronchial epithelium. Nasal (pink) and bronchial (purple) brushings were collected from subjects at high risk of developing lung cancer from four different studies (DECAMP bronchus: N = 341, AEGIS bronchus: N = 305, BCLHS bronchus: N = 238, PCA bronchus: N = 133, DECAMP nose: N = 253, AEGIS nose: N = 150). The size and color of the bubbles represent the significance and magnitude, respectively, of the t statistic calculated using linear modeling of VE gene expression as a function containing the four clinical variables, correcting for batch and mean Transcript Integrity Number (mTIN). Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (b) Boxplots showing the significance of associations between VE expression and smoking status in the DECAMP cohort, assessed by Student's t test on residual counts after correction for sex, age, percentage of predicted FEV1, batch, and mTIN ($p < 0.05$). (c) Expression of VE genes is not significantly correlated ($p > 0.05$, Pearson correlation) between paired DECAMP nasal (x-axis) and bronchial (y-axis) epithelial samples (N = 123). Residual counts adjusted for sex, age, percentage of predicted FEV1, batch, and mTIN were used for the comparison. The blue line is the line of best fit and the gray shading represents the 95% confidence level interval for predictions from the linear model.

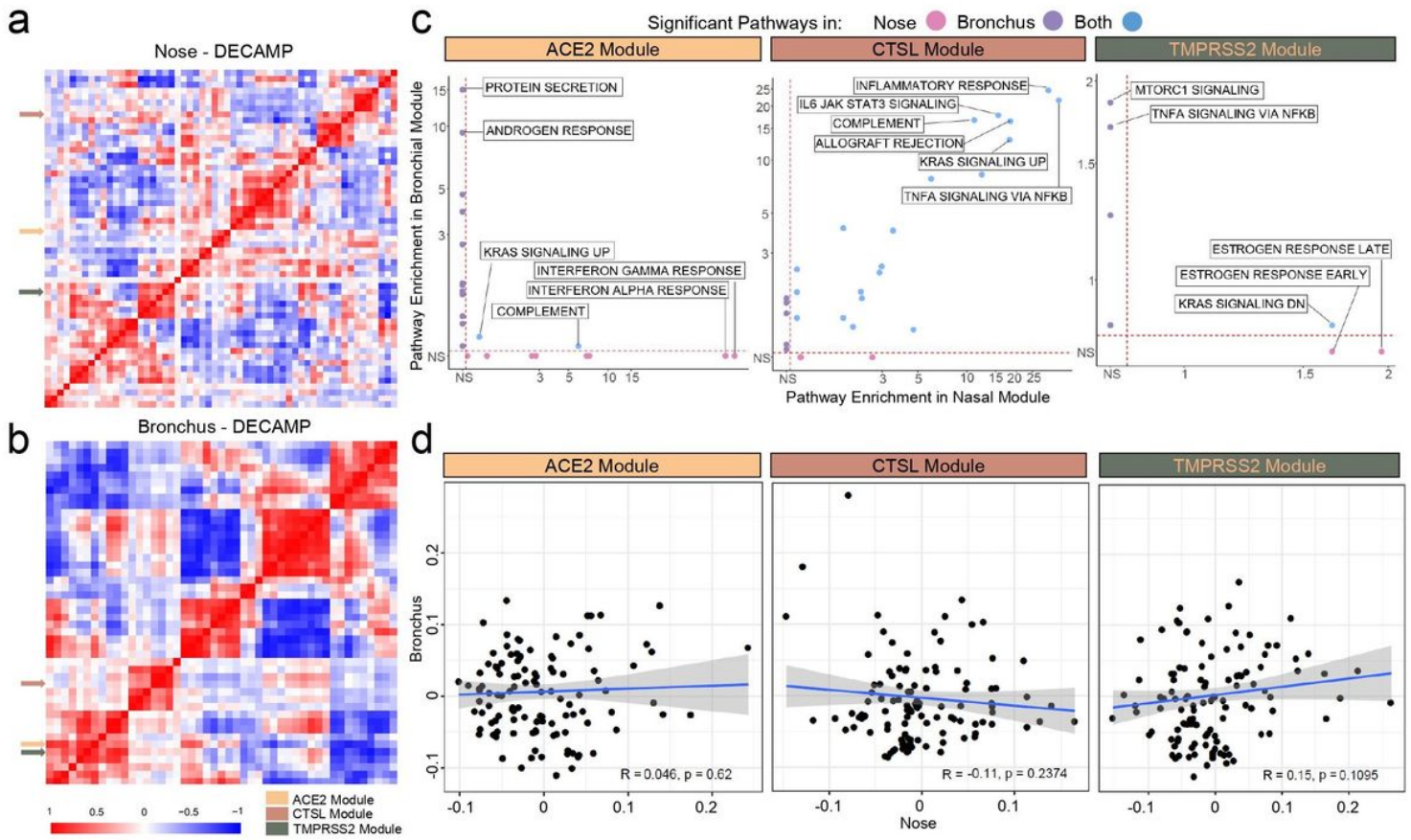

Figure 2

Consensus nasal and bronchial VE gene modules are associated with different biological pathways. (a-b) WGCNA was used to identify consensus co-expression modules in nasal (a) and bronchial (b) samples from the DECAMP cohort (n=58 and 48 modules, respectively). A heatmap of the correlation of module eigengenes computed from each module demonstrates that ACE2 and TMPRSS2 modules were more highly correlated with each other in the bronchus than in the nose. The CTSL module was not correlated with ACE2 or TMPRSS2 modules in either the nose or bronchus. (c) Scatterplots comparing the overrepresentation of MSigDB Hallmark pathway gene sets in each VE module in the nose and bronchus. The -log10(FDR q) values denoting the degree of overrepresentation of each gene set in each module in the nose and bronchus are shown on the x- and y-axes, respectively. The overrepresentation of gene sets in the ACE2 and TMPRSS2 modules is largely discordant between the nose and bronchus, whereas several immune-related pathways are overrepresented in the CTSL module in both sites. (d) VE module eigengenes were not significantly correlated (Pearson $p > 0.05$) between paired nasal (x-axis) and bronchial (y-axis) samples in DECAMP (N = 114). The blue line is the line of best fit and the gray shading represents the 95% confidence level interval for predictions from the linear model.

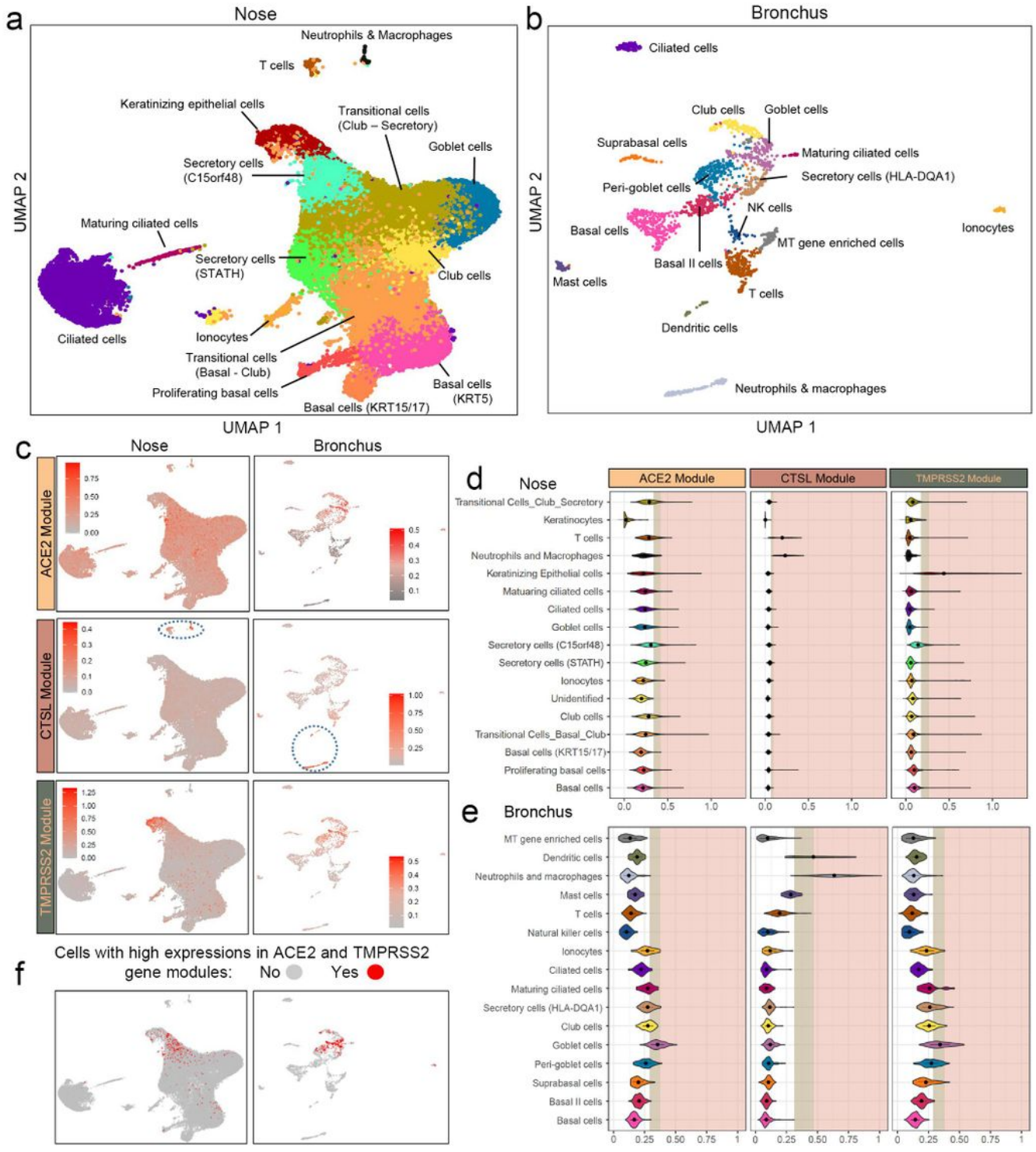

Figure 3

Comparison of expression patterns for VE gene modules across cell types from nasal and bronchial brushings. Single-cell RNA-seq was performed (a) on nasal brushings from 7 patients (n=34,833 cells) and (b) bronchial brushings from 17 patients (n=2,075 cells). Cell types were inferred from the expression of known marker genes. (c) UMAP projections showing the expression pattern of VE gene modules (ACE2, top; CTSL, middle; TMPRSS2 bottom) across different cell types in the nasal (left) and bronchial (right)

epithelium. The cells are colored gray for low expression and red for high expression of metagene scores of each VE gene module. Immune cells with high expression of the CTSL modules were encircled in blue (d-e). Violin plot showing the metagene score for each VE gene module across the cell types in (d) nasal and (e) bronchial epithelium. For each violin plot, metagene expression is designated as elevated (light brown) or highly elevated (pink) if it is greater than one or two standard deviations above the mean metagene score, respectively. (f) UMAP projections showing cells greater than 1 standard deviation above the mean for both ACE2 and TMPRSS2 modules (i.e., double-positive). Double positive cells were enriched in nasal secretory (C15orf48+) and keratinizing epithelial cells (left) and bronchial goblet cells (right).

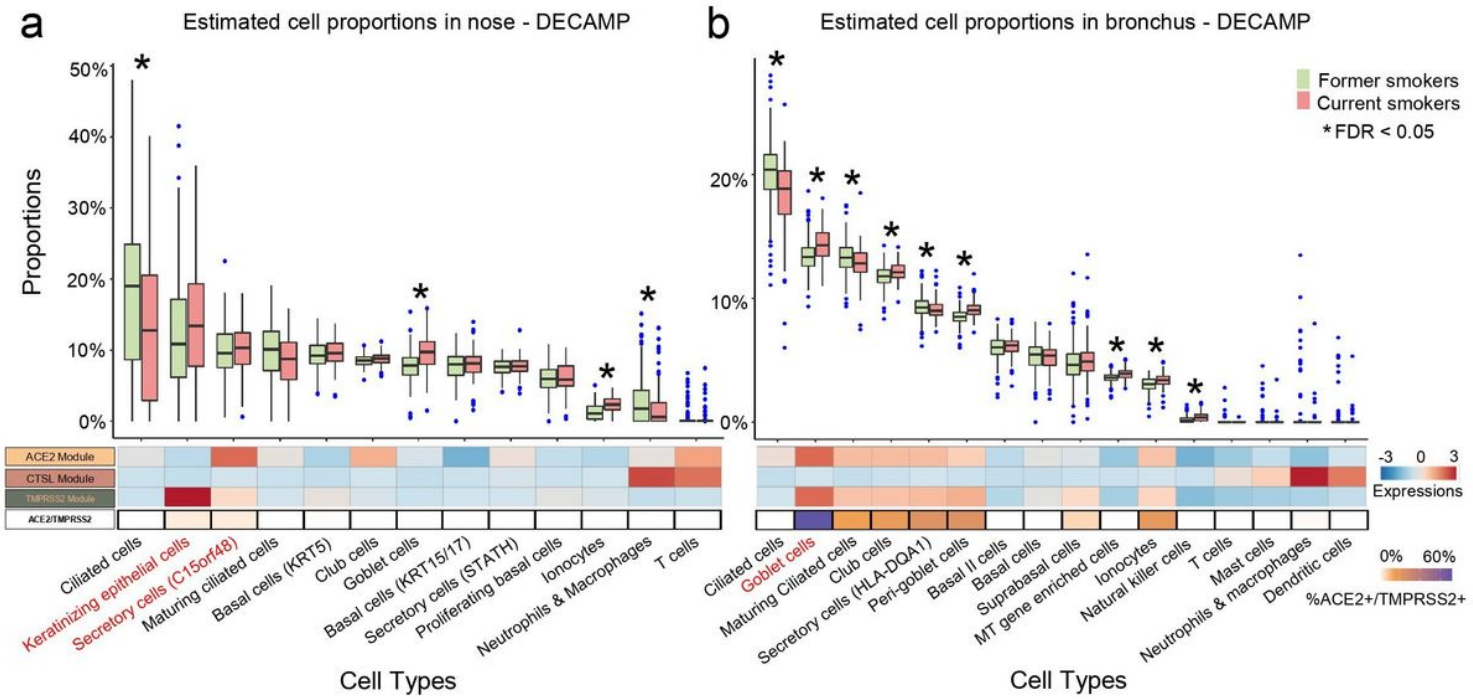

Figure 4

Deconvolution of bulk RNA-seq data shows increased proportions of goblet cells in nasal brushings from current smokers and goblet cells, peri-goblet cells, and ionocytes in bronchial brushings from current smokers. (a-b) Top: boxplots of cell type proportions estimated by AutoGeneS in bulk RNA-seq data from nasal (N = 281) and bronchial brushings (N = 355) obtained from current and former smokers in the DECAMP cohort. Significant cell proportion differences between current and former smokers were determined by the Wilcoxon test (* indicates FDR q < 0.05). Bottom: the heatmap displays the average VE module score for each cell population from the single-cell RNA-seq data (first three rows), as well as the proportion of each cell type that is ACE2+/TMPRSS2+ (i.e., the expression of both genes is one standard deviation above the average). Cell types that express both ACE2 and TMPRSS2 modules are in red. Prevalence of cell types enriched for ACE2 and TMPRSS2 expression shows positive correlation with smoking only in the bronchus, which may explain the lack of association of ACE2/TMPRSS2 expression between the bronchial and nasal bulk RNA-seq data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterials.docx
- SupplementalTables.xlsx