bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

No evidence that androgen regulation of pulmonary TMPRSS2 explains sex-discordant COVID-19 outcomes

Mehdi Baratchian[1], Jeffrey McManus[1], Mike Berk[1], Fumihiko Nakamura[1], Serpil Erzurum[2,3], Sanjay Mukhopadhyay[4], Judy Drazba[5], John Peterson[5], Ben Gaston[6] and Nima Sharifi[1,7,8,*]

[1] Genitourinary Malignancies Research Center, Lerner Research Institute, Cleveland Clinic

[2] Department of Pathobiology, Lerner Research Institute, Cleveland Clinic

[3] Respiratory Institute, Cleveland Clinic

[4] Pathology and Laboratory Medicine Institute, Cleveland Clinic

[5] Imaging Core, Lerner Research Institute, Cleveland Clinic

[6] Herman Wells Center for Pediatric Research, Indiana University School of Medicine

[7] Department of Urology, Glickman Urological and Kidney Institute, Cleveland Clinic

[8] Department of Hematology and Oncology, Taussig Cancer Institute, Cleveland Clinic

[*]Correspondence

Nima Sharifi, M.D.

Email: sharifn@ccf.org

Phone: 216 445-9750

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The recent emergence of SARS-CoV-2 and the subsequent COVID-19 pandemic have posed a public health crisis. Higher morbidity and mortality of men with COVID-19 may be explained by androgen-driven mechanisms. One such proposed mechanism is androgen regulation of pulmonary TMPRSS2, the host co-receptor for SARS-CoV-2. We find no evidence for increased *TMPRSS2* mRNA expression in the lungs of males compared to females in humans or mice. Furthermore, in male mice, treatment with the androgen receptor antagonist enzalutamide does not decrease pulmonary TMPRSS2 expression. Nevertheless, regardless of sex, smoking significantly increases the expression of *TMPRSS2*, which reverts back to never-smoker levels in former smokers. Finally, we show that in mouse models, despite equivalent AR transcript levels, males express markedly higher amounts of AR protein. If a similar sex-specific regulation of AR protein occurs in human lung, androgens could play important roles in clinical outcome of COVID-19 through mechanisms other than TMPRSS2 regulation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The December 2019 outbreak caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China, has led to the coronavirus disease 2019 (COVID-19) pandemic (Wu et al., 2020; Zhou et al., 2020). The highly contagious nature of the disease as well as the lack of vaccines or clinically approved treatments has caused a worldwide public health emergency. Therefore, improved and timely understanding of the human susceptibilities to SARS-CoV-2 will prove invaluable in controlling the pandemic and treatment of those affected.

SARS-CoV-2 enters cells using two host cellular proteins: angiotensin converting enzyme-2 (ACE2) and transmembrane serine protease 2 (TMPRSS2). The virus first employs ACE2 as a cell entry protein, followed by TMPRSS2-mediated proteolytic processing of the SARS-2 spike protein, further facilitating viral entry (Hoffmann et al., 2020). Targeting the activity or expression of both factors by a plethora of approaches has been proposed as potential treatment (Hoffmann et al., 2020; Stopsack et al., 2020; Zhang et al., 2020). *TMPRSS2* is also a widely studied androgen-regulated gene in prostate tissue, contributing to prostate cancer pathogenesis by way of aberrantly driving oncogene expression. Approximately half of all prostate cancers harbor a fusion that juxtaposes a *TMPRSS2* transcriptional regulatory element, which is stimulated by potent androgens and the androgen receptor (AR), in front of an ERG oncogene. (Tomlins et al., 2005). The end result is AR stimulation of oncogene expression.

This raises the possibility that the physiological roles of androgens may, at least partially, account for the sex-specific clinical outcomes (Grasselli et al., 2020; Sharon Moalem, 2020). Due to its androgen-regulated nature in the prostate and its essential role in SARS-Cov-2 etiology, TMPRSS2 expression has been postulated to follow a similar pattern of regulation in pulmonary cells by the potent androgens testosterone and dihydrotestosterone (Mikkonen et al., 2010). If this link proves correct, it could pave the path to novel strategies, including re-purposing of FDA-approved potent androgen synthesis inhibitors or AR antagonists, such as enzalutamide (Enz) and apalutamide, for the treatment of COVID-19. These strategies have been the subject of clinical trial discussions (Sharifi and Ryan, 2020; Stopsack et al., 2020).

Here, we show that there is no evidence for increased TMPRSS2 expression in the lungs of males. On the other hand, cigarette smoking increases the levels of pulmonary *TMPRSS2* mRNA, which drop back to never-smoker levels after quitting smoking. We provide *in vivo* evidence that neither mRNA nor protein levels of TMPRSS2 vary by sex. In

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

addition, AR-antagonist therapy with Enz does not suppress pulmonary TMPRSS2 expression in eugonadal male mice. Interestingly, despite similar transcript levels, we found AR protein levels are substantially higher in lungs of male mice compared with females. This finding, conditional upon validation in human samples, may yet help explain sex differences in COVID-19 outcomes, but likely in a manner that is independent of TMPRSS2 regulation.

Results

TMPRSS2 and *AR* transcript expression in human lung are not higher in males than in females

Considering the poorer clinical outcome of COVID-19 in men, underlying androgen-related causes are suspected but not presently known. The SARS-Cov-2 co-receptor, TMPRSS2 harbors an AR-responsive enhancer that is induced by male gonadal hormones in prostate tissue (Wang et al., 2007), raising the possibility of a similar mode of regulation in the respiratory system. Furthermore, certain pulmonary disease process outcomes, including asthma, are sex steroid associated (Zein et al., 2020). Therefore, we sought to investigate whether male sex was associated with higher expression of *TMPRSS2* or *AR* in human lung. To this end, we acquired the publicly available expression datasets for *TMPRSS2* and *AR* in non-cancerous lung and associated respiratory tissues from the Genomic Expression Omnibus (GEO). Across all tissue-type comparisons, we found no evidence for elevated *TMPRSS2* or *AR* mRNA expression (Fig.1) in males compared with females. Together, these results suggest that worse clinical manifestations of men with COVID-19 are unlikely to be associated with *TMPRSS2*. However, this conclusion is limited to transcriptional information, and further studies are required to rule out the possibility of protein level differences in the human respiratory system.

Pulmonary TMPRSS2 expression in mice does not differ by sex and is unaffected by potent AR blockade

To further explore sex differences in TMPRSS2 expression in animal models, we harvested lungs from female and male mice treated with control diet or Enz for > 10 days and analyzed them for protein and mRNA expression. In the case of TMPRSS2, this yielded results consistent with human expression data: i.e., there were no sex-specific changes. Additionally, Enz treatment did not downregulate TMPRSS2 (Fig. 2A and B). Given the unchanged AR protein levels (despite a modest but statistically significant mRNA increase) in Enz-treated males, we infer that TMPRSS2 is not regulated by AR in the lung. Nevertheless, further

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

single-cell level experiments are needed to explore the expression pattern of TMPRSS2 in cellular subsets rather than in bulk lung and respiratory tract.

Similar to *TMPRSS2* and also consistent with human data, we found no evidence of sex-specific changes in *AR* transcript levels of mice (Fig. 2A). However, the lungs of male mice express considerably higher amounts of AR protein compared with females (Fig. 2B). This finding begs the question whether human lungs also follow a sex-discordant pattern of AR protein synthesis or if this is a mouse-specific phenomenon. If the former, androgen-AR-mediated changes may yet explain sex-specific differences in COVID-19 outcomes by TMPRSS2-independent mechanisms.

TMPRSS2 and *ACE2* mRNA expression increase in current smokers whereas in former smokers, expression returns to levels found in never smokers

In addition to male sex, smoking is a risk factor for COVID-19 susceptibility and poor clinical outcomes (Zhao et al., 2020). One recent study of 1,099 COVID-19-positive patients reported a more than two-fold increased risk for intensive care unit admission and death in smokers as compared with non-smokers (Guan et al., 2020). We identified human expression GEO datasets of bronchial/airway epithelial cells containing subject smoking status and asked whether smoking is associated with *TMPRSS2* expression. Our analysis indicated a consistent pattern where expression of both *TMPRSS2* (Fig. 3A) and the primary SARS-CoV-2 receptor *ACE2* (Fig.3B) was modestly but significantly increased in smokers compared with non-smokers. Interestingly, the levels were downregulated to never-smoker levels in former smokers. The results of our analysis are in keeping with several recent reports on ACE2 and smoking (Brake et al., 2020; Leung et al., 2020; Smith and Sheltzer, 2020). Although the p-values range widely across different data sets, almost all the data sets show changes in a consistently increased direction for current smokers including multiple data sets with small p-values (Fig. 3A and B).

For both *TMPRSS2* (Fig. 3C) and *ACE2* (Fig. 3D), there was no correlation between smoking pack years and mRNA expression in both current and former smokers, suggesting that the change does not build up over time but is instead a rapid process – akin to a switch.

Discussion

Sex-associated clinical outcomes have been long observed in a variety of infectious and inflammatory conditions. Sex steroids (i.e., androgens and estrogens) are possible mediators

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of these biologic differences. For COVID-19, potential androgen-mediated biologic differences include 1) *TMPRSS2* regulation and 2) immune modulation (Sharifi and Ryan, 2020; Stopsack et al., 2020). Our study addresses the first of these possibilities.

Although AR protein is clearly expressed in the male mouse lung at levels higher than in females, we find no evidence for androgen regulation of *TMPRSS2*. This evidence includes 1) the absence of any *TMPRSS2* increase in male compared with female human lung 2) no *TMPRSS2* increase in male compared with female mouse lung 3) no evidence for *TMPRSS2* suppression with next generation AR antagonist treatment. Therefore, given that AR protein is higher in male lungs, it is possible that AR-dependent transcription in the male lung mediates biologic differences independent of *TMPRSS2* expression that are important for driving sex-dependent differences in clinical outcomes. This has yet to be determined.

Limitations of our study include the fact that we cannot rule out cell type-specific differences in sex-associated or androgen-dependent *TMPRSS2* regulation in pulmonary tissues. Nevertheless, early autopsy studies of COVID-19 appear to show diffuse viral damage that may suggest that the expression of viral host cell receptors is likely not limited to an isolated pulmonary subset of cells (Barton et al., 2020). Although we did perform immunohistochemical staining studies for TMPRSS2 expression in the mouse lung, we encountered different patterns compared to human prostate tissue, in which TMPRSS2 is a membrane/cell surface protein. The mouse lung lacked cytoplasmic and membrane staining but expressed significant nuclear stain. This led us to conclude that much of the antibody binding in mouse was non-specific.

The possibility of suppressing androgen-regulated TMPRSS2 expression as a required SARS-CoV-2 host protein was an attractive hypothesis that may also be tested with clinical trials of next-generation anti-androgens (Sharifi and Ryan, 2020; Stopsack et al., 2020). However, our data do not support this hypothesis and we would urge caution in prioritizing COVID-19 clinical trials.

Sexually dimorphic AR expression in an organ not associated with sexual differentiation – the male lung – raises the question of function. This is reminiscent of sexually dimorphic AR protein expression in the male human kidney, in which AR function includes regulating glucocorticoid metabolism and downstream steroid receptor activity (Alyamani et al., 2020). Whether this physiology also occurs in the lung has yet to be determined.

In conclusion, we find no evidence for androgen regulation of *TMPRSS2* in the male lung. Therefore, *TMPRSS2* regulation in the lung appears to fundamentally differ from a

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

clear androgen-dependent effect in prostatic tissues. Pulmonary TMPRSS2 regulation appears not to account for the sex-discordance in COVID-19 clinical outcomes. In contrast, both *TMPRSS2* and *ACE2* appear to have increased expression in current smokers. It is not yet known whether these changes in expression have a functional impact on COVID-19 infection, but if they do, the results suggest smokers could partly mitigate their increased risk by quitting smoking. This also raises the question of whether other hazards such as high levels of air pollution could have similar effects. Nevertheless, a recent study reports that ACE2 expression is not upregulated by lung disease or exposure to carcinogens (Smith and Sheltzer, 2020). Together, these data suggest that induction of ACE2, and perhaps TMPRSS2, is mediated by specific types of tissue injury.

Materials and methods

Mice, treatments and lung harvest

Mouse studies were performed under a protocol approved by the Institutional Animal Care and Use Committee (IACUC) of the Cleveland Clinic Lerner Research Institute. A cohort of adult NSG mice (> 6 weeks old) were obtained from Cleveland Clinic Biological Resources Unit. The male mice were arbitrarily divided between two groups receiving control chow or Enz diet 62.5mg/kg. Following 11 days on diet, the mice were sacrificed using a lethal dose of Nembutal followed by cardiac puncture. Once sacrificed, the abdominal and thoracic cavities of the mice were opened, the inferior vena cava was cut, and the lungs were gently perfused with warm saline via the right ventricle. Next, the lungs were removed, and the individual lobes were either fixed with 10% formalin for paraffin embedding or snap frozen for subsequent RNA or protein analysis.

Protein and mRNA expression analysis

Approximately 40–50 mg freshly frozen lung was added to soft tissue homogenizing CK14 tubes (Betin Technologies) with 200 ml RIPA buffer containing HALT protease and phosphate inhibitor cocktail. Lung tissues were then homogenized with a homogenizer (Minilys, Betin Technologies) three times (60 s each time) at the highest speed, with 5-10 minute intervals on ice to cool lysates. The lysates were then centrifuged for 15 min at 16,000 x g and the supernatants were collected for immunoblot analysis with antibodies for TMPRSS2 (Abcam: ab92323 and Proteintech: 14437-1-AP), AR (EMD Millipore: PG-21 and Santa Cruz Biotechnology: N-20) and GAPDH (D16H11).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Total RNA was harvested by homogenizing 25 mg lung tissue in 350 µl RLT buffer (RNeasy kit, Qiagen) following the manufacturer's instructions. cDNA synthesis were then carried out with the iScript cDNA Synthesis Kit (Bio-Rad). Quantitative PCR (qPCR) analysis was conducted in triplicate in an ABI 7500 Real-Time PCR machine (Applied Biosystems) using iTaq Fast SYBR Green Supermix with ROX (Bio-Rad) with the following primer sets:

Tmprss2 forward: 5'- gtcatccacacacatcccaagtc-3'
reverse: 5'-tcccagaacctccaaagcaaga-3'

Ar forward: 5'-ggcagcagtgaagcaggtag-3'
reverse: 5'-cggacagagccagcggaa-3'

Rplp0 forward: 5'- gacctccttcttccaggctttg -3'
reverse: 5'-ctcccaccttgtctccagtcttta-3'.

Gene expression in human lung

The public genomics data repository Gene Expression Omnibus (GEO, ncbi.nlm.nih.gov/geo) was searched for data sets containing expression profiling of samples from non-cancerous human lung and bronchial/airway epithelial cells with samples identified by gender and/or smoking status of subjects. The following data sets were identified: GSE994 (airway epithelial cells from current/former/never smokers), GSE4115 (histologically normal bronchial epithelial cells from smokers with and without lung cancer), GSE7895 (airway epithelial cells from current/former/never smokers), GSE16008 (bronchial epithelial cells from healthy current and never smokers), GSE18385 (large and small airway epithelial cells from healthy current and never smokers), GSE37147 (bronchial epithelial cells from current and former smokers with and without COPD), GSE43696 (bronchial epithelial cells from asthma patients and healthy controls), GSE63127 (small airway epithelial cells from healthy current and never smokers), GSE103174 (lung tissue from smokers and nonsmokers with and without COPD), and GSE123352 (non-involved lung parenchyma from ever and never smokers with lung adenocarcinoma). TMPRSS2, AR, and ACE2 gene expression values were obtained from each data set and analyses for comparisons between groups (for each data set for which gender and/or current smoking status information was available) were performed using R.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Alyamani M, Li J, Patel M, Taylor S, Nakamura F, Berk M, Przybycin C, Posadas EM, Madan RA, Gulley JL, Rini B, Garcia JA, Klein EA, Sharifi N. 2020. Deep androgen receptor suppression in prostate cancer exploits sexually dimorphic renal expression for systemic glucocorticoid exposure. *Ann Oncol* **31**:369–376. doi:10.1016/j.annonc.2019.12.002

Barton LM, Duval EJ, Stroberg E, Ghosh S, Mukhopadhyay S. 2020. COVID-19 Autopsies, Oklahoma, USA. *Am J Clin Pathol* 1–9. doi:10.1093/ajcp/aqaa062

Brake SJ, Barnsley K, Lu W, McAlinden KD, Eapen MS, Sohal SS. 2020. Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus SARS-CoV-2 (Covid-19). *J Clin Med*. doi:10.3390/jcm9030841

Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, Cereda D, Coluccello A, Foti G, Fumagalli R, Iotti G, Latronico N, Lorini L, Merler S, Natalini G, Piatti A, Ranieri MV, Scandroglio AM, Storti E, Cecconi M, Pesenti A, COVID-19 Lombardy ICU Network. 2020. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*. doi:10.1001/jama.2020.5394

Guan W, Ni Z, Hu Yu, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui DSC, Du B, Li L, Zeng G, Yuen K-Y, Chen R, Tang C, Wang T, Chen P, Xiang J, Li S, Wang Jin-lin, Liang Z, Peng Y, Wei L, Liu Y, Hu Ya-hua, Peng P, Wang Jian-ming, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S, Zhong N. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. doi:10.1056/nejmoa2002032

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, Müller MA, Drosten C, Pöhlmann S. 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**:271-280.e8. doi:10.1016/j.cell.2020.02.052

Leung JM, Yang CX, Tam A, Shaipanich T, Hackett TL, Singhera GK, Dorscheid DR, Sin DD. 2020. ACE-2 Expression in the Small Airway Epithelia of Smokers and COPD Patients: Implications for COVID-19. *medRxiv*. doi:10.1101/2020.03.18.20038455

Mikkonen L, Pihlajamaa P, Sahu B, Zhang F-P, Jänne OA. 2010. Androgen receptor and

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

androgen-dependent gene expression in lung. *Mol Cell Endocrinol* **317**:14–24. doi:10.1016/j.mce.2009.12.022

Sharifi N, Ryan C. 2020. Androgen hazards with COVID-19. *Endocr Relat Cancer* 1–9. doi:10.1530/ERC-20-0133

Sharon Moalem. 2020. Why Are So Many More Men Dying from Coronavirus? *New York Times*.

Smith JC, Sheltzer JM. 2020. Cigarette smoke triggers the expansion of a subpopulation of respiratory epithelial cells that express the SARS-CoV-2 receptor ACE2. *bioRxiv* 2020.03.28.013672. doi:10.1101/2020.03.28.013672

Stopsack KH, Mucci LA, Antonarakis ES, Nelson PS, Kantoff PW. 2020. TMPRSS2 and COVID-19: Serendipity or opportunity for intervention? *Cancer Discov* CD-20-0451. doi:10.1158/2159-8290.CD-20-0451

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**:644–8. doi:10.1126/science.1117679

Wang Q, Li W, Liu XS, Carroll JS, Jänne OA, Keeton EK, Chinnaiyan AM, Pienta KJ, Brown M. 2007. A Hierarchical Network of Transcription Factors Governs Androgen Receptor-Dependent Prostate Cancer Growth. *Mol Cell* **27**:380–392. doi:10.1016/j.molcel.2007.05.041

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* **579**:265–269. doi:10.1038/s41586-020-2008-3

Zein J, Gaston B, Bazeley P, DeBoer MD, Igo RP, Bleecker ER, Meyers D, Comhair S, Marozkina N V., Cotton C, Patel M, Alyamani M, Xu W, Busse WW, Calhoun WJ, Ortega V, Hawkins GA, Castro M, Chung KF, Fahy J V., Fitzpatrick AM, Israel E, Jarjour NN, Levy B, Mauger DT, Moore WC, Noel P, Peters SP, Teague WG, Wenzel SE, Erzurum SC, Sharifi N. 2020. HSD3B1 genotype identifies glucocorticoid responsiveness in severe asthma. *Proc Natl Acad Sci U S A* **117**:2187–2193. doi:10.1073/pnas.1918819117

Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. 2020. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med*. doi:10.1007/s00134-020-05985-9

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, Deng Y, Lin S. 2020. The impact of COPD and smoking history on the severity of Covid-19: A systemic review and meta-analysis. *J Med Virol* 0–2. doi:10.1002/jmv.25889

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**:270–273. doi:10.1038/s41586-020-2012-7

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

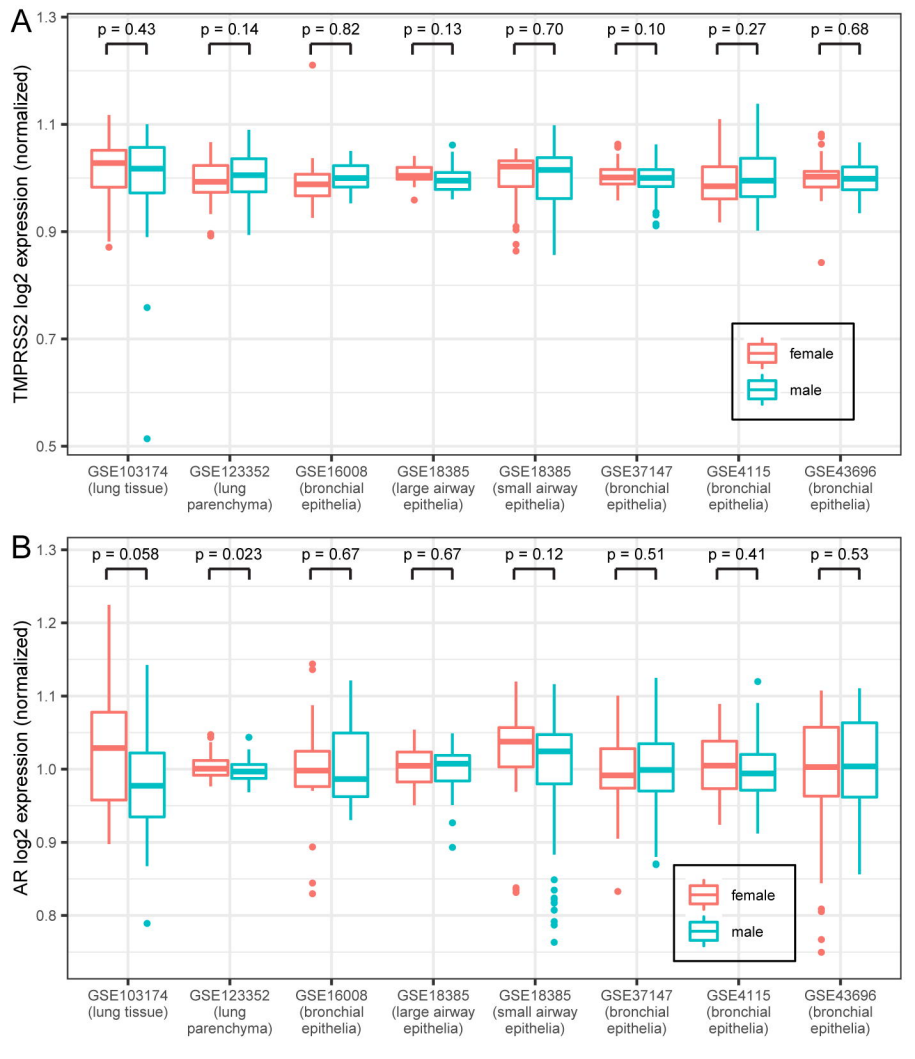
Figure 1. *TMPRSS2* and *AR* transcript expression in human lung are not higher in males than in females. (A) Box plot of *TMPRSS2* expression (normalized so that the mean within each data set equals 1) from the publicly available Gene Expression Omnibus (GEO) data sets. Center lines indicate median values, edges of boxes indicate first and third quartile values, whiskers indicate largest and smallest values extending no more than 1.5 * inter-quartile range from edges of boxes, and dots indicate outlier values. P-values from t-tests are shown for female vs. male comparison within each data set. N for each data set: GSE103174 = 22 female/31 male; GSE123352 = 81 female/95 male; GSE16008 = 15 female/11 male; GSE18385 large airway = 16 female/36 male; GSE18385 = small airway = 35 female/74 male; GSE37147 = 103 female/135 male; GSE4115 41 = female/122 male; GSE43696 = 74 female/34 male. Data sets 18385 and 4115 contained multiple TMPRSS2 reference sequences; none of the individual sequences had female vs. male expression differences with $p < 0.05$, and the expression values of the different sequences were added together for this plot. **(B)** Box plot of AR expression from the same data sets as in (A).

Figure 2. TMPRSS2 expression in mouse lung is not affected by AR-antagonist treatment and does not present a pattern of sex discordance. (A) Transcript levels of *Tmprss2* and *Ar* in NSG mice treated with control or Enz diet. Gene expression was assessed in triplicate and normalized to *Rplp0* levels. The statistical differences were calculated using one-way ANOVA with Tukey's post hoc test. Results are shown as mean ± s.d. (n = 3 technical repeats). **(B)** Immunoblots showing the expression of AR and TMPRSS2 in male mice fed with Enz or control chow for 11 days, and in male vs female mice. Results are representative of 4 technical repeats. TMPRSS2 (Ab1: ab92323; Ab2: 14437-1-AP) and AR (Ab1:PG-21 and Ab2: N-20).

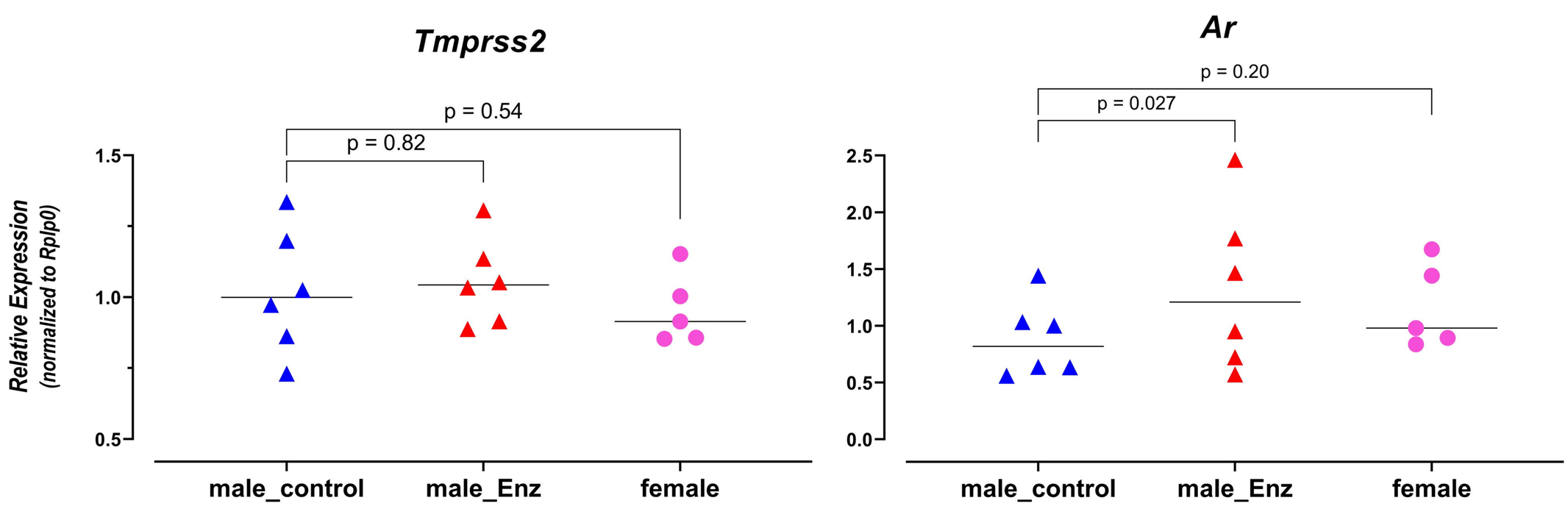
Figure 3. Expression of both *TMPRSS2* and *ACE2* transcript in human bronchial epithelia increases in current smokers compared to both former and never smokers and the increases do not depend on smoking pack years. (A) Box plot of *TMPRSS2* expression (normalized so that the mean within each data set equals 1) from publicly available GEO data sets. N for each data set: GSE16008 = 13 current smoker/13 never smoker; GSE18385 large airway = 32 current smoker/20 never smoker; GSE18385 small airway = 58 current smoker/51 never smoker; GSE37147 = 99 current smoker/139 former smoker; GSE61327 = 112 current smoker/71 never smoker; GSE7895 = 52 current smoker/31 former smoker/21

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

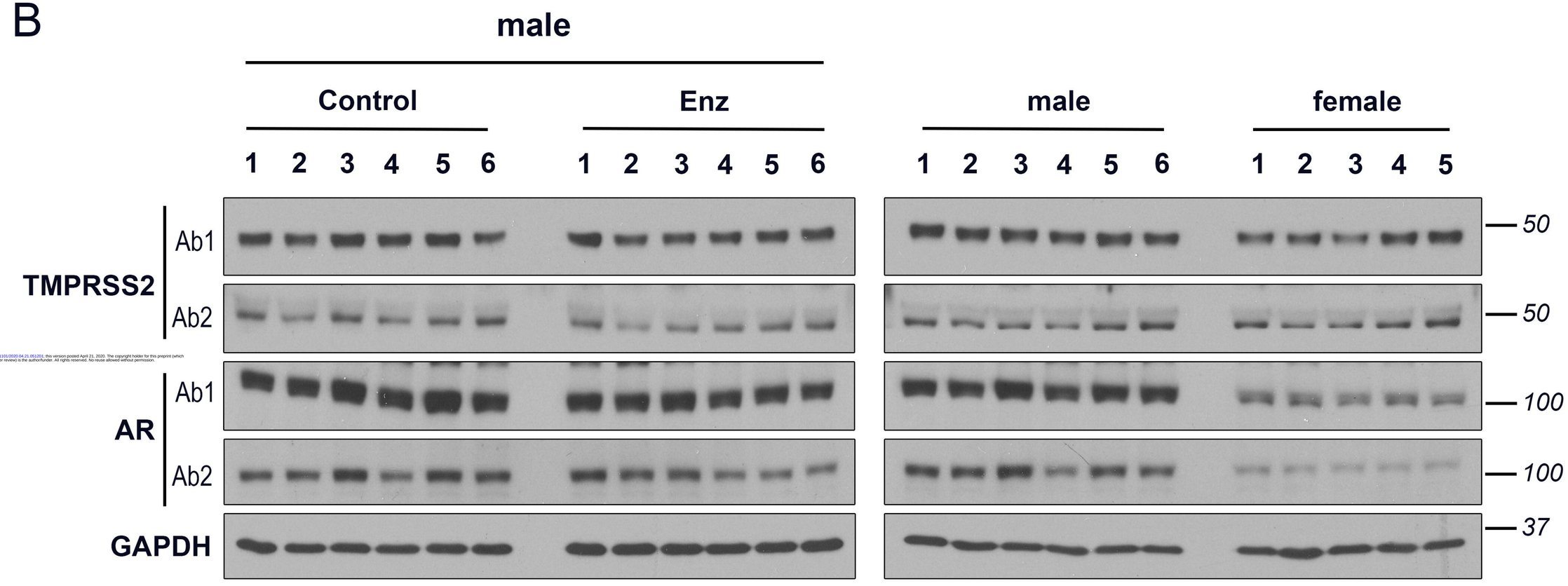
never smoker; GSE994 = 34 current smoker/18 former smoker/23 never smoker. Data sets 18385, 63127, 7895, and 994 contained multiple TMPRSS2 reference sequences and the expression values of the different sequences were added together within each data set. In all of the data sets that contained four *TMPRSS2* reference sequences (18385 both large and small airway and 63127), the difference between groups was largest for reference sequence AI660243, so expression using that reference sequence alone is also shown for those data sets. For data sets with two groups, p-values from t-tests are shown. For data sets with three groups, Tukey HSD p-values are shown that were obtained after one-way ANOVA (ANOVA p-values: GSE7895 0.009, GSE994 0.033). **(B)** Box plot of ACE2 expression from the same data sets as in (A). Data sets 18385 and 63127 contained multiple ACE2 reference sequences and the expression values of the different sequences were added together. For data sets with two groups, p-values from t-tests are shown. For data sets with three groups, Tukey HSD p-values are shown that were obtained after one-way ANOVA (ANOVA p-values: GSE7895 0.82, GSE994 0.010). **(C)** Scatter plots of normalized *TMPRSS2* expression vs. smoking pack years for current and former smokers in all the data sets that contained smoking pack year values, showing the lack of correlation between *TMPRSS2* expression and smoking pack years. Adjusted R-squared and p-values from linear fits for each data set: GSE16008 R^2 = 0.04 and p = 0.17; GSE18385 large airway R^2 = -0.03 and p = 0.63; GSE18385 small airway R^2 = -0.02 and p = 0.99; GSE37147 (current smokers) R^2 = -0.01 and p = 0.47; GSE37147 (former smokers) R^2 = 0.03 and p = 0.03; GSE7895 (current smokers) R^2 = 0.003 and p = 0.29; GSE7895 (former smokers) R^2 = -0.03 and p = 0.74. **(D)** Scatter plots of normalized ACE2 expression vs. smoking pack years for current and former smokers in all the data sets that contained smoking pack year values, showing the lack of correlation between ACE2 expression and smoking pack years. Adjusted R-squared and p-values from linear fits for each data set: GSE16008 R^2 = 0.03 and p = 0.19; GSE18385 large airway R^2 = -0.03 and p = 0.76; GSE18385 small airway R^2 = -0.01 and p = 0.41; GSE37147 (current smokers) R^2 = -0.01 and p = 0.67; GSE37147 (former smokers) R^2 = 0.001 and p = 0.30; GSE7895 (current smokers) R^2 = 0.01 and p = 0.24; GSE7895 (former smokers) R^2 = -0.02 and p = 0.60.


A
TMPRSS2 log2 expression (normalized)
1.3
1.1
0.9
0.7
0.5
p = 0.43
p = 0.14
p = 0.82
p = 0.13
p = 0.70
p = 0.10
p = 0.27
p = 0.68
female
male
GSE103174 (lung tissue)
GSE123352 (lung parenchyma)
GSE16008 (bronchial epithelia)
GSE18385 (large airway epithelia)
GSE18385 (small airway epithelia)
GSE37147 (bronchial epithelia)
GSE4115 (bronchial epithelia)
GSE43696 (bronchial epithelia)
B
AR log2 expression (normalized)
1.3
1.2
1.1
1.0
0.9
0.8
p = 0.058
p = 0.023
p = 0.67
p = 0.67
p = 0.12
p = 0.51
p = 0.41
p = 0.53
female
male
GSE103174 (lung tissue)
GSE123352 (lung parenchyma)
GSE16008 (bronchial epithelia)
GSE18385 (large airway epithelia)
GSE18385 (small airway epithelia)
GSE37147 (bronchial epithelia)
GSE4115 (bronchial epithelia)
GSE43696 (bronchial epithelia)

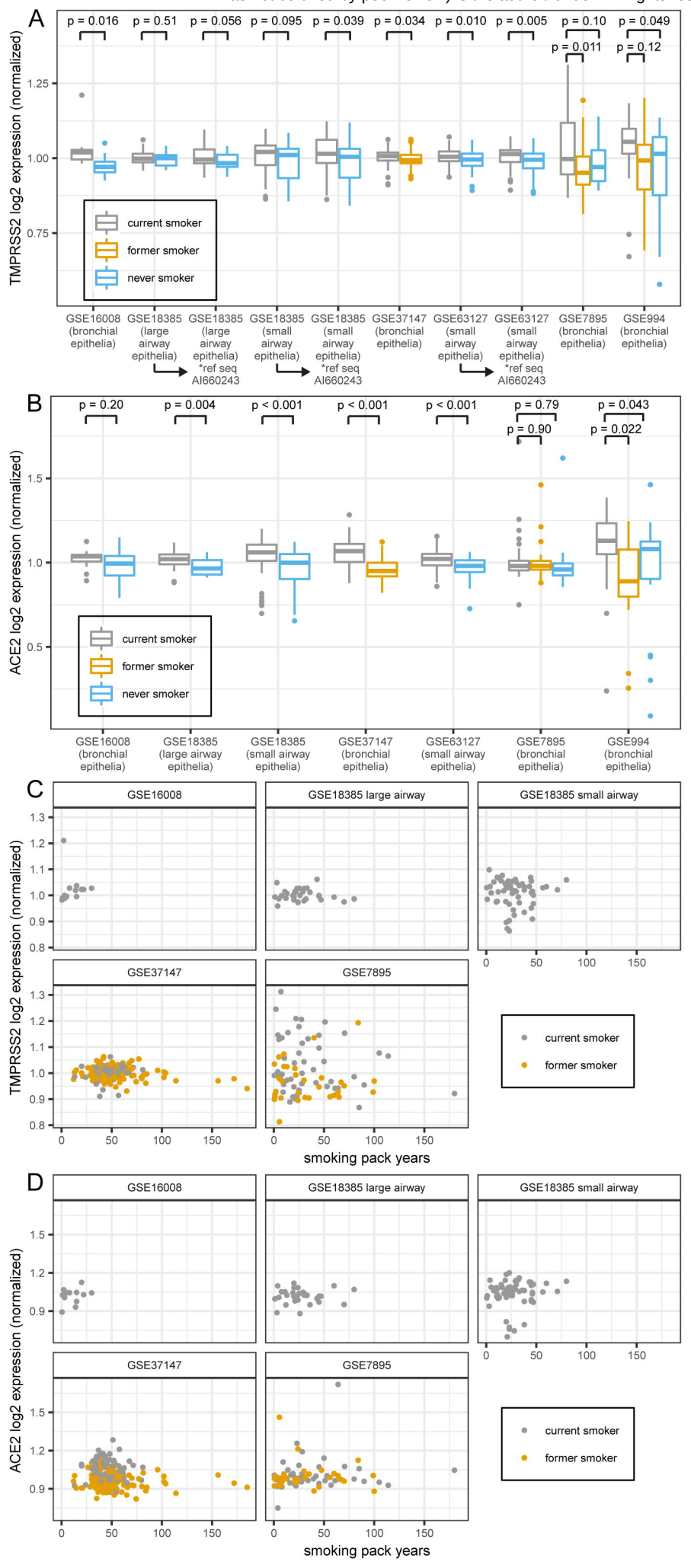
A



B



bioRxiv preprint doi: https://doi.org/10.1101/2020.04.21.051201; this version posted April 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


A
p = 0.016
p = 0.51
p = 0.056
p = 0.095
p = 0.039
p = 0.034
p = 0.010
p = 0.005
p = 0.10
p = 0.049
p = 0.011
p = 0.12
TMPRSS2 log2 expression (normalized)
current smoker
former smoker
never smoker
GSE16008 (bronchial epithelia)
GSE18385 (large airway epithelia)
GSE18385 (large airway epithelia) *ref seq AI660243
GSE18385 (small airway epithelia)
GSE18385 (small airway epithelia) *ref seq AI660243
GSE37147 (bronchial epithelia)
GSE63127 (small airway epithelia)
GSE63127 (small airway epithelia) *ref seq AI660243
GSE7895 (bronchial epithelia)
GSE994 (bronchial epithelia)
B
p = 0.20
p = 0.004
p < 0.001
p < 0.001
p < 0.001
p = 0.79
p = 0.90
p = 0.043
p = 0.022
ACE2 log2 expression (normalized)
current smoker
former smoker
never smoker
GSE16008 (bronchial epithelia)
GSE18385 (large airway epithelia)
GSE18385 (small airway epithelia)
GSE37147 (bronchial epithelia)
GSE63127 (small airway epithelia)
GSE7895 (bronchial epithelia)
GSE994 (bronchial epithelia)
C
GSE16008
GSE18385 large airway
GSE18385 small airway
GSE37147
GSE7895
TMPRSS2 log2 expression (normalized)
current smoker
former smoker
smoking pack years
D
GSE16008
GSE18385 large airway
GSE18385 small airway
GSE37147
GSE7895
ACE2 log2 expression (normalized)
current smoker
former smoker
smoking pack years