Gene expression and in situ protein profiling of candidate SARS-CoV-2 receptors in human airway epithelial cells and lung tissue

Jennifer A. Aguiar1, Benjamin J-M. Tremblay1, Michael J. Mansfield2, Owen Woody3, Brijallen Lobb1, Arinjay Banerjee4, Abiram Chandiramohan5, Nicholas Tiessen5, Quynh Cao5, Anna Dvorkin-Gheva4, Spencer Revill5, Matthew S. Miller4,6,7, Christopher Carlsten8, Louise Organ9, Chitra Joseph9, Alison John9, Paul Hanson10, Richard C. Austin11, Bruce M. McManus10, Gisli Jenkins9, Karen Mossman6, Kjetil Ask4,5, Andrew C. Doxey1,5,12 and Jeremy A. Hirota1,4,5,6,12

ACE2 gene and protein expression is low to absent in airway and alveolar epithelial cells in human lungs. This study suggests the presence of a mechanism dynamically regulating ACE2 expression in human lung or other receptors for SARS-CoV-2. https://bit.ly/3f85R1I


ABSTRACT In December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged, causing the coronavirus disease 2019 (COVID-19) pandemic. SARS-CoV, the agent responsible for the 2003 SARS outbreak, utilises angiotensin-converting enzyme 2 (ACE2) and transmembrane serine protease 2 (TMPRSS2) host molecules for viral entry. ACE2 and TMPRSS2 have recently been implicated in SARS-CoV-2 viral infection. Additional host molecules including ADAM17, cathepsin L, CD147 and GRP78 may also function as receptors for SARS-CoV-2.

To determine the expression and in situ localisation of candidate SARS-CoV-2 receptors in the respiratory mucosa, we analysed gene expression datasets from airway epithelial cells of 515 healthy subjects, gene promoter activity analysis using the FANTOM5 dataset containing 120 distinct sample types, single cell RNA sequencing (scRNAseq) of 10 healthy subjects, proteomic datasets, immunoblots on multiple airway epithelial cell types, and immunohistochemistry on 98 human lung samples.

We demonstrate absent to low ACE2 promoter activity in a variety of lung epithelial cell samples and low ACE2 gene expression in both microarray and scRNAseq datasets of epithelial cell populations. Consistent with gene expression, rare ACE2 protein expression was observed in the airway epithelium and alveoli of human lung, confirmed with proteomics. We present confirmatory evidence for the presence of TMPRSS2, CD147 and GRP78 protein in vitro in airway epithelial cells and confirm broad in situ protein expression of CD147 and GRP78 in the respiratory mucosa.

Collectively, our data suggest the presence of a mechanism dynamically regulating ACE2 expression in human lung, perhaps in periods of SARS-CoV-2 infection, and also suggest that alternative receptors for SARS-CoV-2 exist to facilitate initial host cell infection.
Introduction

In 2003, the severe acute respiratory syndrome (SARS) outbreak caused by the SARS coronavirus (SARS-CoV) resulted in 8096 probable cases with 774 confirmed deaths [1, 2]. In patients with SARS, deaths were attributed to acute respiratory distress associated with diffuse bilateral pneumonia and alveolar damage [3]. In December 2019, SARS-CoV-2 emerged, causing the coronavirus disease 2019 (COVID-19) pandemic. SARS-CoV-2 is spreading at a much more rapid rate than SARS-CoV [4–6]. Similar clinical reports of diffuse bilateral pneumonia and alveolar damage have been reported [7–9]. Severe cases of SARS-CoV-2 have been associated with infections of the lower respiratory tract, with detection of the virus throughout this tissue as well as the upper respiratory tract [7–9]. The biological mechanisms that may govern differences in the number of SARS and COVID-19 cases remain undefined. It is possible that SARS-CoV-2 possesses distinct molecular mechanisms that affect the virulence through viral proteins, greater susceptibility of host cells to infection, permissivity of host cells to virus replication, or some combination of these and other potentially unknown factors [10–13]. Understanding SARS-CoV and SARS-CoV-2 virus similarities and differences at the molecular level in the host may provide insights into transmission, pathogenesis and interventions.

The seminal report identifying the receptor for SARS-CoV used a HEK293 cell over-expression system to identify angiotensin-converting enzyme 2 (ACE2) as a receptor by co-immunoprecipitation with SARS-CoV spike domain 1 [14]. Subsequently, spike protein of SARS-CoV was identified as the viral interacting partner of ACE2. Host protease activity by transmembrane serine protease 2 (TMPRSS2) facilitates ACE2 ectodomain cleavage and fusion of SARS-CoV membrane with host cell membrane [15–17]. ADAM17 (ADAM metallopeptidase domain 17, a member of the ADAM (a disintegrin and metalloprotease domain) family) has also been demonstrated to cleave ACE2 ectodomain, but this was not required for SARS-CoV infection [18–20]. Mechanisms of SARS-CoV entry distinct from ACE2 have also been reported and include activation by endosomal cathepsin L and cell surface expression of CD147 (also known as basigin (BSG)) or GRP78 (78-kDa glucose-regulated protein; also known as heat shock protein family A (Hsp70) member 5 (HSPA5)) [21–23]. Each of these receptors was mechanistically interrogated and results suggested that SARS-CoV could initiate host cell entry and infection using multiple mechanisms. Recent in vitro reports have demonstrated that similar host proteins are involved in facilitating cell entry by SARS-CoV-2, such as ACE2 and TMPRSS2 [5, 24]. Biophysical and structural evidence strongly supports an interaction of ACE2 with SARS-CoV-2 spike protein, similar to SARS-CoV spike protein [12, 13]. Molecular docking studies have also suggested that SARS-CoV-2 spike protein can interact with cell surface GRP78 [25]. Indirect evidence for a role of CD147 in SARS-CoV-2 binding has been demonstrated in vitro with the use of an anti-CD147 intervention that prevented virus replication [26]. Furthermore, a clinical study with an anti-CD147 intervention reduced symptoms and duration of hospital admission for COVID-19 patients [27]. In summary, although there is evidence that SARS-CoV-2 and SARS-CoV both utilise ACE2 as a receptor to facilitate virus entry, it is possible that differences in host entry mechanisms play a role in the large epidemiological differences between the two viruses, which may include additional unidentified receptors.

ACE2 and TMPRSS2 were identified as cellular entry determinants for SARS-CoV using mechanistic studies. The original report of in situ human lung ACE2 expression described positive immunohistochemical staining for alveoli and airway epithelial cells, and immunocytochemical staining in A549 type II alveolar epithelial cells [28]. ACE2 protein expression is also present in the human lung adenocarcinoma cell line Calu-3 [29]. Similar to ACE2, the original report describing the expression of TMPRSS2 in human respiratory mucosa described expression in airway epithelium and type II alveolar epithelial cells [30]. The specificity of the ACE2 and TMPRSS2 antibodies used for analysis of expression patterns in human lung tissues remains to be addressed.
To address the uncertainties related to SARS-CoV-2 receptors in human lung, we performed gene expression and in situ protein profiling of candidate receptors in human airway epithelial cells and lung tissue. Our computational analysis used publicly available microarray gene expression datasets from airway epithelial cells of 515 unique subjects, single cell sequencing data from 10 subjects, and the FANTOM5 dataset for promoter activities of 74 lung-related cell and tissue types. Our proteomic analysis used data from the Human Proteome Map [31] and a dataset from primary human airway epithelial cells grown under air-liquid interface culture conditions [32]. For our in situ protein profiling, we performed immunohistochemical analysis of 98 human lung tissue samples. To determine antibody specificity, we performed immunoblots on protein isolated from Calu-3 cells, primary human airway epithelial cells, primary type II alveolar epithelial cells, the human bronchial epithelium cell (HBEC)-6K7 cell line, the A549 type II alveolar epithelial cell line, and HEK cells. Collectively, our data contrast with previous reports, demonstrating rare ACE2 protein expression in the airway epithelium and alveoli of human lung. Our protein expression data are consistent with low ACE2 promoter activity in a variety of lung epithelial cell samples and low ACE2 gene expression in both microarray and single cell RNA sequencing (scRNAseq) datasets. We present confirmatory evidence for the presence of TMPRSS2, CD147 and GRP78 in the respiratory mucosa. Our data suggest the presence of a mechanism dynamically regulating ACE2 expression in human lung, perhaps in periods of SARS-CoV-2 infection, and/or that alternate receptors for SARS-CoV-2 exist to facilitate initial host cell infection in lung tissue.

Methods

Human ethics
Procurement of primary human airway epithelial cells used for immunoblot and lung tissue for immunohistochemistry was approved by the Hamilton (ON, Canada) integrated Research Ethics Board (HiREB 5099T, 5305T, 11-3559 and 13-523-C). The University of British Columbia (Vancouver, BC, Canada) Research Ethics Office approved heart tissue archives and primary human airway epithelial cell collection.

Upper and lower airway gene expression analysis
Public microarray experiments using Affymetrix chips (HuGene-1.0-st-v1 and HG-U133 Plus 2) on airway epithelial cell samples collected from nasal (GSE19190) or bronchial (GSE11906) brushings of healthy nonsmokers were obtained from the NCBI Gene Expression Omnibus (GEO) database [33, 34]. This resulted in a total of 80 individual samples from the two different experiments that included 11 upper airway samples (nasal n=11) and 69 lower airway samples (trachea n=17, large airway n=17, small airway n=35). For all dataset samples, raw intensity values and annotation data were downloaded using the GEOquery R package (version 2.52.0) [35] from the Biocductor project [36]. Probe definition files were downloaded from Bioconductor and probes were annotated using Bioconductor’s “annotate” package. All gene expression data were unified into a single dataset that was then normalised by robust multiarray average (RMA) normalisation, and only genes present in both of the Affymetrix platforms (n=16013) were kept for subsequent analyses. Correction of experiment-specific batch effects was performed using the ComBat method [37] implemented using the sva R package (version 3.32.1) [38]. RMA-normalised expression levels for conventional (ACE2, TMPRSS2, ADAM17 and CTSL (cathepsin L1)) and non-conventional (CD147 and GRP78) SARS-CoV-2 receptor genes were compared across the four defined airway levels, with CDH1 (E-cadherin) expression level included as a positive control with known expression in lung tissue. Gene expression levels were tested for significant differences via pairwise Wilcoxon rank sum tests with Benjamini–Hochberg multiple testing correction using the stats R package (version 3.6.1). Gene expression box plots were generated with the ggplot2 R package (version 3.2.1).

Analysis of curated bronchial epithelial cell brushing dataset
A total of 1859 public microarray experiments using Affymetrix chips (HG-U133 Plus 2 and HuGene-1.0-st-v1) on airway epithelial cell samples were selected from the NCBI GEO database. These samples were further filtered by removing individuals with asthma or COPD, resulting in a total of 504 individual healthy samples (GSE4302, 28 samples; GSE67472, 43 samples; GSE37147, 159 samples; GSE108134, 274 samples). Within this dataset, sex and/or age information was included for 310 samples; of these, sex data were available for 86 females and 106 males. Smoking status information was also provided for 451 samples, with 260 current smokers, 82 former smokers and 109 never-smokers. For all dataset samples, raw intensity values and annotation data were downloaded as described. Probe definition files were retrieved as described. All gene expression data were unified into a single dataset that was then RMA-normalised, and only genes present in both of the Affymetrix platforms (n=16105) were kept for subsequent analyses. Correction of experiment-specific batch effects was performed as described.
**Analysis of promoter activity from the FANTOM5 dataset**

The FANTOM5 promoterome dataset [39] for the hg38 assembly [40] was used to examine promoter activity of SARS-CoV-2-related human genes, namely ACE2, TMPRSS2, ADAM17, CTSL, CD147 and GRP78. Using the ZENBU genome browser [41], the nearest cap analysis of gene expression (CAGE) peak upstream and on the same strand as each of the aforementioned genes was extracted and analysed. The dataset consists of CAGE promoter activity data for 1866 primary cells, cell lines and tissues from humans, and is quantified as normalised transcripts per million (TPM). A subset of FANTOM5 CAGE data (120 samples) is presented, considering only samples related to lung, gut, heart and prostate tissues (consisting of 74, 19, 15 and 12 samples, respectively). Normalised TPM values for each CAGE peak, an approximation for promoter activity, were \( \log_{10} \) transformed and separated according to tissue and cell type, and the radius of each point is proportional to these transformed normalised TPM values.

**Analysis of protein abundance from proteomic datasets**

Publicly available human proteomic data from the datasets of Kim et al. [31] and Foster et al. [32] were used to evaluate SARS-CoV-2 receptor-related protein expression in different human tissues and experimental conditions. Expression values were extracted from the dataset of Kim et al. [31] for ACE2, TMPRSS2, ADAM17, CTSL, CD147 and GRP78, using CDH1 as a control for airway cells. Data were created using the heatmap package in R (version 1.0.12) and expressed as \( \log_{10} \)-transformed to facilitate visualisation. Proteomic data from the dataset of Foster et al. [32] consist of bronchial epithelial cells collected from healthy nonsmokers (n=4; males) and exposed to PBS control vehicle. Intensity values for ACE2, TMPRSS2, ADAM17, CD147 and GRP78 were extracted, with CDH1 included as a positive control. Intensity values were determined by the original study authors via normalisation of all detected peptide intensities associated with a given parent protein [32]. Box plots were generated with the ggplot2 R package (version 3.2.1) with intensity values \( \log_{10} \)-transformed for visualisation purposes.

**Analysis of single cell RNA sequencing data**

Data preprocessed using the Cell Ranger pipeline (10x Genomics) were obtained from GSE135893. Samples from 10 control subjects and 12 idiopathic pulmonary fibrosis patients were downloaded and post-processed with the Seurat package in R [42]. Cell populations were defined using the markers provided in the source paper [43]. Cells belonging to the 10 control subjects were used for further analysis. Visualisations of violin plots were created using Seurat.

**Primary human airway epithelial cells**

The human lung adenocarcinoma cell line, Calu-3, was grown under culture conditions defined by the supplier (ATCC HTB-55). Primary human airway epithelial cells isolated via bronchial brushings from consenting healthy individuals were grown in PneumaCult ExPlus (Stemcell Technologies, Vancouver, BC, Canada) under submerged monolayer culture conditions and used between passages 1 and 4. The human bronchial epithelial cell line, HBEC-6KT, was grown under submerged monolayer culture conditions in keratinocyte serum-free media supplemented with epithelial growth factor (0.4 ng·mL\(^{-1}\)) and bovine pituitary extract (50 \( \mu \)g·mL\(^{-1}\)) [44–47].

**Immunoblots**

Cell protein was isolated using RIPA lysis buffer (VWR, Mississauga, ON, Canada) supplemented with protease inhibitor cocktail (Sigma, Oakville, ON, Canada) with quantification performed using Bradford assay reagents (Bio-Rad, Mississauga, ON, Canada). Immunoblots were performed using stain-free 4–20% pre-cast gradient gels and imaged on a ChemiDoc XRS+ Imaging system (Bio-Rad). For each immunoblot, 20 \( \mu \)g of protein was added per lane. ACE2 (MAB933, monoclonal, clone 171606, 2 \( \mu \)g·mL\(^{-1}\) (R&D Systems)), TMPRSS2 (HPA035787, polyclonal, 0.4 \( \mu \)g·mL\(^{-1}\) (Atlas Antibodies)), CD147 (ab666, monoclonal, clone MEM-M6/1, 1 \( \mu \)g·mL\(^{-1}\) (Abcam)) and GRP78 (610979, monoclonal, clone 40/BiP, 0.25 \( \mu \)g·mL\(^{-1}\) (BD Biosciences), and HPA038845, rabbit polyclonal (Atlas Antibodies)) primary antibodies were diluted in 5% skimmed milk/TBS with 0.1% Tween-20 and incubated overnight on a rocker at 4°C with detection performed the following day using an anti-mouse (ACE2, CD147 and GRP78 (BD Biosciences)) or anti-rabbit (TMPRSS2 and GRP78 (Atlas Antibodies)) horseradish peroxidase (HRP)-conjugated secondary antibodies at 1:3000 for 2 h at room temperature (Cell Signaling, Danvers, MA, USA). Visualisation of TMPRSS2, CD147 and GRP78 was performed using Clarity Western enhanced chemiluminescence (ECL) Substrate, while ACE2 was visualised with Clarity Max ECL Substrate (Bio-Rad). Total protein loading images were collected as a qualitative visualisation of protein loading between sample types [48]. The immunogen for ACE2 primary antibody is mouse myeloma cell line NS0-derived recombinant human ACE2 Gln18-Ser740 (predicted). The immunogen for TMPRSS2 primary antibody is the recombinant protein epitope signature tag antigen sequence, GSPPAIGP YYEHGYQPENPYPQAQPTVVTYYEVHBAQYYSPVQYAPRVTQASNPVVCTQPSPSGTVCTSKT.
The immunogen for the CD147 primary antibody is recombinant full-length protein corresponding to human CD147. The immunogen for the GRP78 BD Biosciences primary antibody is human BiP/GRP78 amino acids 525–628. The immunogen for the GRP78 Atlas Antibodies primary antibody is the recombinant protein epitope signature tag antigen sequence, EKFAEDKKLKERIDTRNEYSLKNDQII GDEKKEGLKSLSEDKETMKEEVIEWLESQDADIEDFKKKKEEIEIVQPIISSLK.

Independent immunoblot analysis (L. Organ, C. Joseph, A. John and G. Jenkins) was performed on A549, HEK and immortalised human bronchial epithelial cells. Equal amounts of protein (20 µg) were loaded on to 4–12%, Bis-Tris gradient gels (NP0326BOX; ThermoFisher) with anti-ACE2 (ab108252, rabbit monoclonal, clone EPR4435(2), 1/500 dilution of stock antibody; Abcam). A loading control of GAPDH was used to demonstrate protein loading (ab181603, rabbit monoclonal, EPR16884, 1/10 000 dilution of stock antibody; Abcam). Visualisation was performed with ECL Clarity (Bio-Rad) on a Licor C-DiGit.

Immunohistochemistry

Formalin-fixed paraffin-embedded human lung tissue from non-diseased regions was obtained from archived tissue blocks from patients who had undergone lung resection for clinical care. Human heart tissue was from the University of British Columbia Cardiovascular Tissue Registry. Sections 4 µm thick were cut and stained for ACE2 (15 µg·mL⁻¹), TMPRSS2 (10 µg·mL⁻¹), CD147 (5 µg·mL⁻¹), and GRP78 (HPA038845, 1/200 dilution) using the same antibodies used for immunoblot analysis. All staining was performed on a Leica Bond RX system with Leica Bond reagents, heat-induced antigen retrieval at pH 6 (20 min) with primary antibody incubation for 20 min. Digital slide scanning was performed using an Olympus VS120-L100 Virtual Slide System at 40× magnification with VS-ASW-L100 V2.9 software and a VC50 colour camera, followed by image visualisation with HALO image analysis software.

Results

Candidate genes important in SARS-CoV-2 infection are detectable at varying levels in human airway epithelial cells and lung tissue

We performed a targeted analysis of ACE2, TMPRSS2, ADAM17, CTSL, CD147 and GRP78 gene expression as candidates important for SARS-CoV-2 infection in human airway epithelial cells. Here and throughout the gene expression analyses, CDH1 (E-cadherin) was used as a control for lung epithelial cell phenotype. We first examined these genes in a curated dataset of upper and lower airway epithelial cell gene expression from the nasal sinus to the 12th generation of airway in the lung (figure 1).

In the upper airways, all candidates were expressed, with the highest level observed for GRP78 and the lowest level observed for ACE2. Analysis along multiple generations of the lower airways (trachea, large airway and small airway) revealed identical relative expression patterns, with ACE2 being the least expressed and GRP78 being the highest expressed. ACE2 gene expression showed the greatest variability along the upper and lower airways, with greatest expression observed in the trachea samples and the lowest expression in the small airway (figure 1).

FIGURE 1 Microarray expression profiles of candidate SARS-CoV-2 receptor genes in upper and lower airways. Normalised log2 expression levels for ACE2 (angiotensin-converting enzyme 2), TMPRSS2, ADAM17, CTSL (cathepsin L1), CD147 and GRP78 genes compared across the upper airway (nasal) and lower airways (trachea, large airway and small airway). CDH1 (E-cadherin) gene expression level is included as a positive control. Statistical values for comparisons for each gene at each airway generation were calculated; those not shown were nonsignificant. *: p<0.05; **: p<0.01; ***: p<0.001.
Following our observation of consistent expression along the upper and lower airways of candidate genes important in SARS-CoV-2 infection, we determined whether sex or age affected gene expression levels in healthy individuals using a curated dataset of bronchial brushings from 504 healthy subjects (supplementary table S1). The expression levels for the candidate genes in healthy subjects paralleled the patterns observed in the smaller survey of upper airways, trachea, large and small airways (figure 2).

FIGURE 2 Microarray expression profiles of candidate SARS-CoV-2 receptor genes in lower airway epithelial cells, analysed by age and sex. a) Clustered heatmap of log2 expression levels from NCBI Gene Expression Omnibus (GEO) samples (n=504), annotated by age, sex and microarray chip platform. Expression values reflect signal intensities, indicating lowest detected expression of ACE2 (angiotensin-converting enzyme 2) and highest expression of GRP78 and CDH1 (E-cadherin). b and c) Box plots of expression levels separated by b) sex (n=194) and c) smoking status (n=451). d and e) Plots of gene expression levels versus age, with linear regression lines of best fit, for datasets that used either d) the HG-U133 Plus 2 microarray (n=43) or e) the HuGene-1.0-st-v1 microarray (n=181). Correlations were performed separately between platforms because of differences in their age distributions. e) A weak negative correlation (r=−0.20, p=0.015) was detected for ACE2 in the dataset that used the HuGene-1.0-st-v1 microarray.
Median ACE2 gene expression was the lowest, while GRP78 gene expression was the highest (figure 2a). No gene candidate demonstrated sex dependence for expression levels (figure 2b). ACE2, TMPRSS2, CD147 and GRP78 were elevated in current smokers relative to never-smokers (figure 2c and supplementary table S2). CTSL was reduced in current smokers relative to never-smokers. No microarray chip-dependent effects were observed for relationships between sex or smoking status and gene expression. For quantitative analyses related to age and gene expression, our curated database was divided into datasets that used either the HG-U133 Plus 2 or HuGene-1.0-st-v1 microarray due to differences in age distributions. In the HuGene-1.0-st-v1 dataset (n=181), which included a greater proportion of older individuals (>50 years), we observed reduced ACE2 gene expression with age (figure 2e; p<0.05).

Promoter activity data of each of the candidate genes important in SARS-CoV-2 binding and infection were extracted and analysed from the FANTOM5 dataset, which includes 1866 primary cells, cell lines, and tissue sample types (figure 3). We selected all sample formats that included “lung”, “nasal”, “airway” or “olfactory”, to identify lung-specific sample types. Gut, heart and prostate tissue samples were analysed as controls. Consistent with our observed gene expression analysis along the upper and lower airways, normalised TPM values for each CAGE peak demonstrated that CD147 promoter activity was elevated relative to ACE2 promoter activity across airway epithelial cells and lung tissue samples. CTSL promoter activity was the lowest of all candidate genes, which contrasted with the modest expression observed at the gene level (figure 2a). Both microarray gene expression analysis and promoter activity were consistent with results of candidate gene expression in a scRNAseq dataset of 10 healthy subjects (supplementary figure S1).

Collectively, our gene expression analysis of the upper and lower airways of healthy males and females of diverse ages suggests that ACE2 gene expression is low relative to all other candidate SARS-CoV-2 receptor genes analysed in human airway epithelial cells. Furthermore, we observe no sex-dependent or age-dependent expression patterns of any candidates at the gene level, although smoking status did have an impact on gene expression levels.

In vitro and in situ protein profiling reveals distinct expression patterns for candidates important in SARS-CoV-2 infection

Analysis of transcriptional data may not be indicative of in situ protein expression levels [49]. To extend our gene expression observations, we mined publicly available proteomic data from whole lung and primary human airway epithelial cell cultures and performed in vitro immunoblots on human airway epithelial cell lysates and in situ protein immunohistochemistry on human lung tissue using the same antibodies for each method.

The Human Proteome Map is a publicly available resource that includes select adult and fetal tissues and circulating immune cell populations [31]. Using this resource, we examined protein expression of ACE2, TMPRSS2, ADAM17, CTSL, CD147 and GRP78. In human lung tissue homogenate, ACE2 was not detected, while being detected in heart, gut and testes, known positive control tissues (figure 4a). The rank order of the remaining molecules in human lung tissue homogenate was: GRP78>CD147>CTSL>ADAM17>TMPRSS2. Human lung tissue homogenate is a heterogeneous population of cells, precluding the ability to associate protein expression to a given cell type. We therefore interrogated a publicly available proteomic dataset derived from primary human airway epithelial cells grown under air–liquid interface culture conditions [32], examining the same candidates. Again, ACE2 protein expression was not detectable (figure 4b). CD147, GRP78 and CTSL were expressed with multiple peptide counts, while TMPPRSS2 and ADAM17 were only marginally expressed with low peptide counts. Collectively, two proteomic datasets from distinct lung sample formats provide complementary and consistent expression profiles of candidate molecules important in SARS-CoV-2 infection.

To localise the in situ expression of the candidate molecules of interest at the protein level, we pursued immunohistochemical analysis paired with immunoblot validation of the specificity of the selected antibodies for recognition of proteins of the predicted molecular weight. An anti-ACE2 antibody detected only a single band in Calu-3 cells at the predicted molecular weight of ACE2 protein (~110 kDa) (figure 5a, lanes 1–3). The anti-ACE2 antibody required the use of a super-sensitive ECL solution. No ACE2 protein was detected in primary airway epithelial cells or the HBEC-6KT cell line, despite confirmation of protein loading (figure 5a, lanes 4–9, protein loading shown underneath main blot). Independent immunoblotting with a distinct anti-ACE2 primary antibody was performed, with a single band observed in HEK cells, but not in immortalised human bronchial epithelial cells or A549 cells (supplementary figure S2).

An anti-TMPRSS2 antibody detected multiple bands in all airway epithelial cell samples, with a dominant band at the predicted molecular weight of ~57 kDa (figure 5b). These patterns were conserved across all cell types that were analysed.
An anti-CD147 antibody detected a single band in all airway epithelial cell samples, with a dominant band at the predicted molecular weight of $\sim$55 kDa (figure 5c). The immunoblot bands were consistent with the heavy glycosylation of CD147 [50].

An anti-GRP78 antibody (BD Biosciences 610979) detected a single band in all airway epithelial cell samples, with a dominant band at the predicted molecular weight of $\sim$78 kDa (figure 5d).

The immunoblots using anti-ACE2, anti-CD147 and anti-GRP78 demonstrated a single band of predicted molecular weight, suggesting that observed immunohistochemical staining should be specific to the protein.
of interest based on the target epitope, as both methods detect denatured proteins [51]. The same anti-ACE2 and anti-CD147 antibodies were validated for immunohistochemistry. The anti-TMPRSS2 was used for immunohistochemistry, although the multiple bands observed by immunoblot caution the specificity of any observed in situ staining. Attempts to optimise anti-GRP78 antibody application for immunohistochemistry were unsuccessful, requiring additional antibody interrogation with HPA038845.
Atlas Antibodies), which was suitable for both immunoblotting and immunohistochemistry (supplementary figure S3).

ACE2 immunohistochemistry revealed only select staining in rare cells in the airways and the alveoli of all 98 human lung samples analysed, which included healthy subjects and those with chronic lung diseases (figure 6). A single healthy human sample contained one positive airway epithelial cell with additional positive staining in the peripheral lung in cells with type II alveolar epithelial cell morphology (figure 6a).

![Immunoblot analysis of ACE2, TMPRSS2, CD147, and GRP78 protein expression in human airway epithelial cell protein lysates.](https://doi.org/10.1183/13993003.01123-2020)
A representative image of a sample from a smoker with COPD (figure 6b) shows no ACE2 protein staining in the airway epithelium and a rare positive cell in sub-basement membrane tissue. Quantification of positive pixel count for ACE2 staining normalised to total tissue pixel count revealed no differences between healthy nonsmokers and tobacco smokers (supplementary figure S4). Lung microvasculature and human heart tissue had positive staining (supplementary figures S5 and S6), consistent with previously described reports for ACE2 protein staining patterns [52, 53].

TMPRSS2 immunohistochemistry revealed diffuse staining in the airway epithelium and in immune cells in the lung periphery, with greater staining in smokers with COPD (figure 6). These observations were consistent in the majority of the 98 human samples examined.

CD147 immunohistochemistry revealed strong membrane-restricted staining in the airway epithelium and diffuse staining in immune cells in the lung periphery (figure 6). CD147 displayed greater staining in smokers with COPD. These observations were consistent in the majority of the 98 human samples examined.

GRP78 immunohistochemistry revealed diffuse staining in airway and alveolar epithelium and in immune cells in the lung periphery (figure 6; 49 samples). No qualitative differences in GRP78 staining were observed between healthy subjects and smokers with COPD.

Collectively, our in vitro and in situ protein profiling is consistent with our gene expression analysis, with CD147 and GRP78 protein expression dominant over TMPRSS2 and ACE2. Additional examples of staining in lung tissue are provided in supplementary figure S7. ACE2 protein expression is rare in human lung tissue and found in select cells in both healthy individuals and those with chronic lung diseases.

FIGURE 6 Immunohistochemical localisation of ACE2 (angiotensin-converting enzyme 2), TMPRSS2, CD147 and GRP78 protein in human lung tissue. Representative samples from a) a healthy nonsmoker with no underlying chronic airway disease, and b) a smoker with COPD. Black outlines: low magnification (12×) of conducting airways with airway epithelium; scale bars 100 μm. Green outlines: high magnification regions (60×) of conducting airway epithelium that are defined in the low magnification images by green squares; scale bars 50 μm. Red outlines: high magnification regions (50×) of lung tissue away from the airway lumen that are defined in the low magnification images by red squares; scale bars 50 μm. H&E: haematoxylin and eosin. Positive immunohistochemical staining is rust/brown. Total number of independent samples analysed was 49–98.
TMPRSS2 and CD147 protein expression are potentiated in individuals with a history of tobacco smoking and a diagnosis of COPD.

Discussion

The global COVID-19 pandemic that emerged in late 2019 is caused by SARS-CoV-2. The possible host receptor(s) for SARS-CoV-2 have not been exhaustively surveyed in human lung tissue at the gene and protein level. Understanding the expression levels and localisation of candidate SARS-CoV-2 receptors in host lung tissue may provide insights into therapeutic interventions that might reduce disease spread, viral replication or disease pathology. To address this knowledge gap, we performed gene expression, proteomic profiling at the tissue and cell level, and in situ protein profiling of candidate receptors in human airway epithelial cells and lung tissue (summarised in figure 7). Collectively, our data demonstrate rare ACE2 protein expression in human airway epithelial cells in vitro and in situ. Our protein expression data are consistent with low ACE2 promoter activity in a panel of lung epithelial cell samples and low ACE2 gene expression in bronchial epithelial cells (microarray) and lung cells (scRNAseq). We present confirmatory evidence for the presence of TMPRSS2, CD147 and GRP78 protein in vitro in airway epithelial cells and confirm broad in situ protein expression of CD147 and GRP78 in the respiratory mucosa. Our data suggest that for ACE2 to be an integral receptor for SARS-CoV-2, mechanisms are likely to exist that dynamically regulate expression in human lung, perhaps in periods of SARS-CoV-2 infection [54]. It is also possible that alternate receptors for SARS-CoV-2 are important in initial host cell infection.

Using a curated microarray gene expression dataset generated from bronchial brushings of 504 healthy subjects that considers the limitations of merging multiple datasets from distinct sources, we observed that sex did not correlate with gene expression of any candidate host molecule involved in SARS-CoV-2

![Proposed functions of host airway epithelial cell molecules for interaction with SARS-CoV-2. Proteins associated (or suggested to be associated) with host cell entry of SARS-CoV-2 and the activation of the SARS-CoV-2 spike protein (SARS-S) are displayed. Angiotensin-converting enzyme 2 (ACE2) is suggested as the primary SARS-S receptor for viral entry (interaction of ACE2 with SARS-S receptor-binding domain (RBD) leading to endosomal viral uptake), followed by activation of SARS-S via pH-dependent cleavage mediated by cathepsin L1 (CTSL). Secondary methods of viral entry and SARS-S activation are likely to involve proteases (e.g. TMPRSS2 and ADAM17) and/or secondary receptors [CD147 and GRP78]. Dashed lines indicate mechanisms that have not been fully validated. CyPA: cyclophilin A; N protein: nucleocapsid protein; ER: endoplasmic reticulum. Adapted from [20] with updates and additional information on candidate host molecules. Figure created with BioRender.com.](https://doi.org/10.1183/13993003.01123-2020)
infection and that ACE2 and TMPRSS2 were the lowest expressed genes of interest examined. In one dataset, ACE2 gene expression modestly decreased with age, although protein level confirmation was not possible. The low level of ACE2 and TMPRSS2 gene expression in bulk bronchial epithelial cell gene expression samples suggests low levels of cells expressing both of these genes within this lung tissue. We confirm that tobacco smoking is associated with elevated ACE2 gene expression levels in bronchial epithelial cell samples [55], although we were unable to confirm this with immunohistochemistry analysis of protein on human lung samples.

Advances in transcriptomics have enabled scRNAseq that has identified unique and rare cell types in human lung that may have importance in health and disease [56, 57]. scRNAseq provides an opportunity to look at transcriptional profiles in subsets of cell populations, which may isolate a cell signal from a bulk sample. We therefore utilised scRNAseq data from healthy human lung samples as a parallel approach. The resolution of scRNAseq for subpopulations of epithelial cells revealed low or absent expression of ACE2 gene in all populations examined, whereas CD147 and GRP78 were present in all populations. Our results are consistent with current publicly available data that discuss the presence of rare ACE2/TMPRSS2-positive cells [54]. Using lung samples from eight individuals (four HIV and active tuberculosis double positive, two HIV positive and tuberculosis negative, and two double negative controls), ZIEGLER et al. [54] have reported in humans that only 0.8% of type II alveolar epithelial cells expressed both ACE2 and TMPRSS2 genes. Further analysis of ciliated cells found that 5.3% of these cells expressed both ACE2 and TMPRSS2 genes. In vitro models with SARS-CoV are consistent with this finding, as ciliated cells are preferentially targeted by this coronavirus [58]. Most intriguing is that ACE2 and TMPRSS2 gene-expressing cells were only identified in the HIV and tuberculosis double positive samples. These observations were replicated in the upper airways, with only a rare population of secretory epithelial cells (0.3% of this population) co-expressing ACE2 and TMPRSS2. The reported scRNAseq results are consistent with a focused analysis looking at only ACE2 gene expression in a variety of lung cell types [59]. Importantly, these elegant transcriptomic analyses confirm our observations in bulk tissue microarray datasets.

Consortium-based publicly available datasets represent another parallel approach to confirm our data. We have used the FANTOM5 dataset containing CAGE promoter activation data for 1866 primary cells, cell lines and tissue samples from humans [39] to examine the level of promoter activity for each candidate SARS-CoV-2 receptor gene. The FANTOM5 CAGE data provide an additional and complementary approach to quantifying gene expression, since a given gene’s shared promoter can yield multiple transcripts at different expression levels, as well as being partially independent of any given transcript’s half-life in the cell. In general, the promoter activity of ACE2 in airway-related tissues is low or absent; only a single sample originating from an adult lung yielded a normalised CAGE promoter expression level >1 TPM, while expression was observed in gut cells, consistent with known patterns of ACE2 expression [60]. Consistent with the microarray data, CD147 promoter activity is elevated relative to ACE2 across airway-related cells and tissues, although the relatively low CTSL promoter activity is incongruent with modest levels of gene expression.

The expression of genes does not always correlate with protein expression [49]. With this in mind, we performed combination proteomic analyses with immunoblot analyses. For our immunoblots, we used the human Calu-3 adenocarcinoma cell line, as this cell line is susceptible and permissive to SARS-CoV-2 infection and expresses ACE2, an observation we confirm [24, 29]. We also used primary human airway epithelial cells and the bronchial epithelial cell line (HBEC-6KT). We performed immunoblots for ACE2 and TMPRSS2 as these have been highlighted as interacting with SARS-CoV-2, while we probed CD147 as recent pre-clinical and clinical studies have provided proof of concept for this as a candidate SARS-CoV-2 receptor [26, 27]. Lastly, GRP78 was dominantly expressed throughout transcriptomic studies and was selected as a positive control, as previous expression has been confirmed in human airway epithelial cells [61]. Cathepsin L was excluded from the present analysis due to low promoter activity (figure 3), while ADAM17 was excluded as the proposed function in coronavirus infections is via ACE2 [5, 24], which was included in analysis. Immunoblot with all antibodies revealed dominant bands of predicted molecular weight, with the anti-TMPRSS2 polyclonal antibody revealing additional minor bands in all cell samples examined. The identity of these other bands remains unclear and suggests downstream immunohistochemical analysis may be confounded by the specificity of this antibody. In contrast, antibodies for ACE2, CD147 and GRP78 were specific and could be used for immunohistochemistry without concerns of specificity. Interestingly, ACE2 protein could only be detected with a super-sensitive ECL solution and only in Calu-3 cells, suggesting absent protein in primary human airway epithelial cells and the HBEC-6KT cell line. Our data are consistent with previous immunoblots of primary human airway epithelial cells grown under submerged monolayer conditions using the same primary antibody, where ACE2 protein was absent, and only expressed under air–liquid interface culture conditions [62]. The
observation that CD147 and GRP78 are also expressed in Calu-3 cells encourages further interrogation into these host proteins, as they may contribute to function of ACE2 and TMPRSS2 in SARS-CoV-2 binding and fusion in this cell type. Collectively, the profiling of antibodies by immunoblot of airway epithelial cells revealed distinct band patterns demonstrative of antibody specificity for ACE2, CD147 and GRP78, and to a lesser extent for TMPRSS2.

Immunohistochemical analysis has been performed for localisation of ACE2 and TMPRSS2 in human lung [28, 30]. The observation of positive staining in human lung tissue for these proteins was not accompanied by companion immunoblot or complementary approaches to define the specificity of the antibody used [51]. In the absence of determination of antibody specificity, the historical data presented should be interpreted with caution. To address the issue of antibody specificity for immunohistochemical staining, we used the same antibodies we validated by immunoblot and confirmed findings using proteomics as an orthogonal, antibody independent, approach. We again focused on ACE2 and TMPRSS2 as these are candidate proteins important for SARS-CoV-2 infection of host cells. Our immunohistochemical staining patterns of ACE2 were consistent with transcriptional profiling and immunoblots with only one out of 98 human samples demonstrating rare staining in the airway and alveolar epithelium. Positive ACE2 staining in heart tissues and areas of lung microvasculature suggest our staining protocol was successful. These results directly contrast with those reported using antibodies that lacked validation for specificity [28, 30]. TMPRSS2 was expressed more frequently across all samples examined, with variability in the airway epithelium associated with history of smoking and/or COPD status. In contrast, CD147 expression was observed in airway epithelium of all samples. Similar to TMPRSS2, elevated CD147 expression was associated with history of smoking and/or COPD status, consistent with previous reports [50]. Our original GRP78 antibody selected for immunoblotting was not validated for immunohistochemistry. We therefore performed confirmatory GRP78 immunoblotting and immunohistochemistry with an additional antibody (HPA038845) and provided demonstration of expression in human airway epithelial cells in vitro and in situ. Importantly, it is well established that GRP78 resides in the endoplasmic reticulum (ER) under normal physiological conditions, where it acts as an ER-resident molecular chaperone to facilitate correct protein folding. However, under conditions of ER stress, including viral infection, a portion of ER-resident GRP78 relocates to the cell surface, where it may act as a viral co-receptor [63, 64]. The presence of cell surface GRP78 has been reported in atherosclerotic plaques [65], prostate cancer [66] and kidney [67]. However, there are currently no commercially available anti-GRP78 antibodies that bind specifically to cell surface GRP78 and dual immunofluorescence is used to show co-localisation of cell surface GRP78 with an established surface receptor [65]. This drawback precludes our ability to perform accurate cell surface GRP78 immunohistochemistry on lung tissue to interrogate this concept further in the context of SARS-CoV-2 receptors. Utilisation of prostate cancer patient-derived GRP78 auto-antibodies that are specific for cell surface GRP78 [65] may be suitable on lung tissue for assessment of cell susceptibility to SARS-CoV-2 infection by GRP78.

Our study has several limitations that have not already been addressed. Our observation of differences in gene expression between upper and lower airways and along the airway tree were not corroborated at the protein level. It remains possible that entirely different protein expression profiles for the candidate molecules examined exist in the upper airway, presenting a different environment for SARS-CoV-2 interaction with the respiratory mucosa. Nasal pharyngeal swabs are capable of detecting SARS-CoV-2 virus [68] and this anatomical region is probably important for subsequent infection in the lower airways [8, 69]. Related to this potential temporality of effect, it is possible that SARS-CoV-2 induces the expression of receptors on host cells following infection [54]. Our study is also limited by examining candidate molecules important in SARS-CoV-2 infection under basal conditions, in the absence of viral or environmental stimuli that may regulate gene transcription and protein translation.

SARS-CoV-2 infection and transmission has caused the global COVID-19 pandemic. An understanding of the receptors used by SARS-CoV-2 for host cell infection and the parallel characterisation in human samples is required to inform development of intervention strategies aimed at mitigating COVID-19. Our data demonstrate rare ACE2 protein expression in human airway epithelial cells in vitro and in situ, consistent with low ACE2 promoter activity and ACE2 gene expression in bronchial epithelial cells. We present confirmatory evidence for the presence of TMPRSS2, CD147 and GRP78 protein in vitro in airway epithelial cells and confirm broad in situ protein expression of CD147 and GRP78 in the respiratory mucosa. Due to the overwhelming evidence that the SARS virus interacts with ACE2, there are likely to be alternate mechanisms regulating ACE2 in the respiratory mucosa in the context of SARS-CoV-2 infection, and/or perhaps other co-receptors, beyond what is expressed under basal conditions at the protein level.

Acknowledgements: We would like to acknowledge Mary Jo Smith from the McMaster Immunology Research Centre Core Histology facility (McMaster University, Hamilton, ON, Canada) for her timely and professional expertise with antibody staining for immunohistochemistry. We would like to thank Sam Wadsworth (Aspect Biosystems, Vancouver,
BC, Canada) and John McDonough (Yale University, New Haven, CT, USA) for their intellectual discussions around CD147 and airway epithelial cell biology. We would like to thank Charles Plessy (Okinawa Institute of Science and Technology, Onna, Japan) for his suggestion to analyse the FANTOM5 data. We would like to thank all of the personal and professional support from the authors’ respective families (Candace Lee Bastedo, Julianne Meiko Hirota and Hana Lee Hirota) and research institutes and, most importantly, the frontline healthcare workers during the COVID-19 pandemic.

Conflict of interest: J.A. Aguiar has nothing to disclose. B.J-M. Tremblay has nothing to disclose. M.J. Mansfield has nothing to disclose. O. Woody has nothing to disclose. B. Lobb has nothing to disclose. A. Banerjee has nothing to disclose. A. Chandiramohan has nothing to disclose. N. Tiessen has nothing to disclose. Q. Cao has nothing to disclose. A. Dworkin-Gheva has nothing to disclose. S. Revill has nothing to disclose. M.S. Miller has nothing to disclose. C. Carlsten has nothing to disclose. L. Organ has nothing to disclose. C. Joseph has nothing to disclose. A. John has nothing to disclose. P. Hanson has nothing to disclose. R. Austin has a patent US7524826B2 issued to McMaster University and Hamilton Health Sciences Corp., McMaster University. B.M. McManus has nothing to disclose. G. Jenkins reports grants from AstraZeneca, Biogen, Galecto and GlaxoSmithKline; personal fees from Boehringer Ingelheim, Daewoong, Galapagos, Heptares, Promedior and Roche; grants and personal fees from Plantl; non-financial support from Redx and NuMedii; and is trustee for Action for Pulmonary Fibrosis, outside the submitted work. K. Mossman has nothing to disclose. K. Ask has nothing to disclose. A.C. Doxey has nothing to disclose. J.A. Hirota has nothing to disclose.

Support statement: This work was supported by start-up funds from J.A. Hirota and a CIHR grant from K. Mossman. J.A. Hirota is supported by the Canada Research Chairs program and an Ontario Early Researcher Award. A.C. Doxey is supported by NSERC and an Ontario Early Researcher Award. Funding information for this article has been deposited with the Crossref Funder Registry.

References


https://doi.org/10.1183/13993003.01123-2020


