

Systemic analysis of putative SARS-CoV-2 entry and processing genes in cardiovascular tissues identifies a positive correlation of BSG with age in endothelial cells

Blerina Ahmetaj-Shala, Ricky Vaja, Santosh S Atanur, Peter M. George, Nicholas S. Kirkby and Jane A. Mitchell

*BAS, RV and SA contributed equally to this study.

Affiliations

BA-S, RV, NSK and JAM: Cardiorespiratory Interface, National Heart and Lung Institute, Imperial College London, SW7 2AZ

SA: Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London

Institute of Translational Medicine and Therapeutics ([ITMAT](#)) Data Science Group, NIHR, BRC, Imperial College London

PMG: Interstitial Lung Disease Unit, National Heart and Lung Institute, Imperial College London, Royal Brompton and Harefield NHS Foundation Trust, Sydney Street, London, SW3 6NP, UK.

Correspondence to Dr Blerina Ahmetaj-Shala: b.ahmetaj@imperial.ac.uk or

Professor Jane A. Mitchell: j.a.mitchell@ic.ac.uk

Abstract:

COVID-19, caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has rapidly spread throughout the world with unprecedented global healthcare and socio-economic consequences. There is now an established secondary syndrome of COVID-19 characterised by thrombosis, vascular dysfunction and hypertension, seen in those most severely affected. Advancing age in adults is the single most significant risk factor for hospitalisation and death with COVID-19. In light of the cardiovascular/thrombotic sequelae associated with severe COVID-19 disease and the overwhelming risk that increased age carries, in this study, our aim was to obtain mechanistic insight by interrogating gene expression profiles in cardiovascular tissues and cells. Our focus was on the two putative receptors for SARS-CoV-2, *ACE2* and *BSG* along with a selected range of genes thought to be involved in virus binding/processing. In this study we have made four important observations: (i) Cardiovascular tissues and/or endothelial cells express the required genes for SARS-CoV-2 infection, (ii) SARS-CoV-2 receptor pathways, *ACE2/TMPRSS2* and *BSG/PP1B(A)* polarise to lung/epithelium and vessel/endothelium respectively, (iii) expression of SARS-CoV-2 host genes are, on the whole, relatively stable with age and (iv) notable exceptions were *ACE2* which decreases with age in some tissues and *BSG* which increases with age in endothelial cells. Our data support the idea that that *BSG* is the dominate pathway utilised by SARS-CoV-2 in endothelial cells and are the first to demonstrate a positive correlation with age. We suggest *BSG* expression in the vasculature is a critical driver which explains the heightened risk of severe disease and death observed in those >40 years of age. Since *BSG* is utilised by other pathogens our findings have implications beyond the current pandemic. Finally, because *BSG* functions in a range of cardiovascular diseases and fibrosis, our observations may have relevance to our understanding of the diseases associated with aging.

Introduction:

COVID-19, caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), was first reported in Wuhan, China in December 2019. SARS-CoV-2 is relatively contagious and whilst producing mild symptoms in the majority of people, can progress to severe or fatal disease in susceptible individuals. As such the virus has rapidly spread throughout the world with unprecedented global healthcare and socio-economic consequences.

SARS-CoV-2 is related to SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) which caused respiratory epidemics in 2003 and 2012 respectively. Based on what was known about human host interactions with SARS-CoV and MERS-CoV along with recent research using SARS-CoV-2 tools, a list of key entry and processing genes utilised by the virus to infect host cells has been defined. SARS-CoV-2 enters host cells by binding of the spike protein with two putative receptors; ACE2¹ and BSG (also known as Basigin, CD147 or EMMPRIN)^{2,3}. For viral entry by ACE2, it is thought that the SARS-CoV-2 spike protein is primed and ACE2 cleaved, by the cellular serine proteases TMPRSS2¹ and ADAM17. FURIN cleaves viral enveloping proteins providing another putative priming step for the spike protein of SARS-CoV-2⁴. For viral entry via BSG, less is known regarding specific receptor/viral processing partners for SARS-CoV-2. However, for SARS-CoV⁵, HIV⁶ and the measles virus⁷, respectively, peptidylprolyl isomerase A (PPIA; also known as cyclophilin A) and peptidylprolyl isomerase B (PPIB; also known as cyclophilin B), which are natural ligands for BSG, incorporate into virus and facilitate binding to BSG. Similarly, PPIB forms a complex with the malaria pathogen (*Plasmodium falciparum* merozoites) and BSG to facilitate infection of red blood cells⁸. Intracellular processing of SARS-CoV-2 spike protein is thought to involve the lysosomal cysteine proteases cathepsin B/L (CTSL, CTSB) which, can also substitute for TMPRSS2 in some cells¹.

Initial infection with SARS-CoV-2 occurs via the respiratory epithelium; high gene expression of *ACE2* and *TMPRSS2* in nasal epithelium^{9,10} have been taken to imply that the nose is a primary entry point for the virus⁹. ACE2 and TMPRSS2 are also co-expressed in bronchial epithelium⁹⁻¹¹. However, where COVID-19 progresses to severe disease the lung and other organs are also affected. The emerging pattern of severe and fatal COVID-19 disease includes pneumonia with acute respiratory distress syndrome, cytokine storm, widespread

vasculopathy, thrombosis, renal failure, hypertension and endothelial dysregulation seen across multiple vascular beds and organ systems^{12,13}. Furthermore, COVID-19 is associated with an increased risk of arterial thrombosis^{14,15} and venous thromboembolism with pulmonary embolism^{16,17}, both likely to represent an important source of acute and post-COVID-19 morbidity and mortality. While hypertension and thrombosis are common features after COVID-19^{18,19}, the important question as to whether COVID-19 as an independent risk factor for cardiovascular disease in the acute setting and during the recovery period, is a concern and remains to be established. Similarly the rates of post-COVID-19 pulmonary fibrosis remain unknown but at discharge the vast majority of patients with COVID-19 have evidence of persisting pulmonary infiltrates on computerised tomography scans²⁰ and almost half have physiological impairment²¹. This secondary thrombotic/vascular clinical syndrome of severe COVID-19 suggests that SARS-CoV-2 infects not only respiratory epithelium but also the endothelium disrupting barrier function and allowing access to cardiovascular tissues and other organs of the body²². This idea is supported by reports showing that SARS-CoV-2 can infect endothelial cells in vitro²³ and that coronaviruses including SARS-CoV-2 can progress to a systemic infection^{24,25} with some patients showing detectable viral RNA in blood samples²⁶⁻²⁸.

The reasons that underpin progression of mild to severe or fatal COVID-19 disease remain incompletely understood but risk factors have been defined²⁹; these include established cardiovascular disease, diabetes, obesity and black and minority ethnicity [BAME]. However, the dominant risk factor for severe COVID-19 across all datasets is age, with the vast majority of those in hospital with COVID-19 disease being over 40 years. Indeed, a recent disparities in outcomes report by Public Health England found age to be the largest disparity with likelihood of death in adults increasing in an age-dependent manner from around 40 years³⁰. Importantly, while positive tests for SARS-CoV-2 infection increase with age the relative rate of infection between adult age groups profoundly underpredicts the effect of advancing age on the risk of death from COVID-19^{29,30}. Therefore, understanding how SARS-CoV-2 causes severe disease with pulmonary, thrombotic, cardiorenal and vascular complications is critically important in managing the pandemic and identifying therapeutic strategies.

While some studies report expression profiles of *ACE2* and *TMPRSS2* in epithelial cells^{9,11} and immune cells^{10,11}, expression patterns of a wider range of host SARS-CoV-2 entry and processing genes in these cells has been recently reported¹¹. However, the relative expression levels of SARS-CoV-2 entry and processing genes in vessels and in endothelial cells has not been fully established. Finally, the impact of age on the expression of these genes in a cardiovascular setting is incompletely understood.

Here we have used publicly available gene expression data to determine the relative expression of key SARS-CoV-2 host entry/ processing genes in human cardiovascular tissues including aorta, coronary artery, heart (atria and left ventricle), whole blood and the kidney and for comparison the colon, spleen and lung. We went on to investigate gene expression in endothelial cells and, for comparison, airway (nasal and bronchial) epithelium and leukocytes (peripheral blood mononuclear cells; PBMCs). We used blood outgrowth endothelial cells as a model because, since they are obtained from blood samples of living donors, data sets across age ranges have been created. Furthermore, blood outgrowth endothelial cells are an accepted model for application in personalised medicine since they retain elements of disease phenotype across a number of cardiovascular and other conditions³¹⁻³³. After mapping gene expression across our target tissues and cells, our primary objective was to determine how age, as the single most dominant risk factor for severe COVID-19, impacts on expression of SARS-CoV-2 entry and processing genes in human cardiovascular and other tissues.

Methods:

Genotype-Tissue Expression (GTEx) analysis

The Genotype-Tissue Expression (GTEx) project³⁴ is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and. We downloaded gene expression data from GTEx version 8 (<https://www.gtexportal.org/home/datasets>) which contain expression data from 54 tissues from 948 donors. We identified tissues of interest based on organ systems affected by severe COVID-19 disease and extracted expression data specifically from those tissues. Tissues were split into two categories; (i) cardiovascular tissues including aorta, coronary artery, heart (atrial and appendage), left ventricle, kidney (cortex) and whole blood and (ii) 'other tissues' including lung, colon and spleen. We performed principle component analysis (PCA) on gene expression data from each tissue of interest. We observed that the major variation in gene expression was due to type of death (Hardy Score; Supplementary Figure 1) and so corrected for this. We normalised the gene expression data for each tissue separately using COMBAT-seq³⁵ with Hardy score as a batch. After normalisation expression data was extracted for our target genes (*ACE*, *ACE2*, *ADAM17*, *BSG*, *CTSB*, *CTSL*, *FURIN*, *PPIA*, *PPIB* and *TMPRSS2*). The following number of donors were identified for each tissues; aorta (432), coronary artery (240), atrial appendage (429), left ventricle (432), kidney cortex (85), whole blood (755), lung (578), colon (779) and spleen (241). Age identifiers in GTEx are grouped by decade, as such results were analysed based on samples that associated with 20-29 to 70-79 years of age. Principle component analysis (PCA) plots of raw and processed GTEx data is presented in Supplementary Figure 1.

Gene expression dataset systematic review analysis

Using ArrayExpress and NCBI GEO we identified the raw datasets (.CEL files) of transcriptomic gene expression profiling by microarray of healthy adult donors for blood outgrowth endothelial cells, peripheral blood mononuclear cells (PBMCs) and bronchial airway (obtained from bronchial brushing) and nasal (obtained nasal brushing) epithelium. We applied strict inclusion criteria; (i) only datasets that used Affymetrix Gene Chips (.CEL files) were included, (ii) only datasets where individual ages are defined were included and (iii) only datasets for 'untreated' cells were included. The following studies and number of donors were identified (see Supplementary Table 1); blood outgrowth endothelial cells; 3 studies with 63 donors,

PBMCs; 6 studies with 84 donors, airway (bronchial) epithelium; 2 studies with 74 donors and nasal epithelium; 3 studies with 111 donors.

Transcriptomic expression profiling of cell data

Raw (.CEL) files were imported into Partek Flow® software and aligned with STAR to the human assembly (hg19) whole genome. The data was quantified to an annotation model using 'Ensembl Transcripts release 75' and normalised to 'Counts Per Million' and filtered to remove genes below the reliable quantitation threshold. The gene expression values from different studies were merged based on gene names. To correct for the batch effects, the data was normalised using the empirical Bayes model ComBat³⁶. The normalised expression values of our target genes (*ACE2*, *CYPA*, *CYPB*, *BSG*, *ADAM17*, *TMPRSS2*, *FURIN*, *CTSB* and *CTSL*) were then extracted for further downstream analysis. PCA plots of raw and processed cell data is presented in Supplementary Figure 1.

Statistical analysis:

All data were analysed on GraphPad Prism v8 and are shown as mean +/- SEM for samples from 'n' = individual donors. Data were grouped into two groups; samples from adults below the age of 40 and above the age of 40. Data were tested for normality of distribution and analysed using parametric (Student T-test) or non-parametric (Mann Whitney U-test) tests as appropriate as a discovery exercise. Where $p < 0.05$ follow on correlation tests were performed using Pearson's, for continuous variables (cell data), or Spearman's, for ordinal variables (tissue data), tests. Details of tests used are given in individual figure legends.

Results:

We quantified nine SARS-CoV-2 entry and processing genes (*ACE2*, *BSG*, *ADAM17*, *TMTRSS2*, *CYPA*, *CYPB*, *CTSB*, *CTSL* and *FURIN*) (Figure 1 and Figure 2) along with *ACE* in cardiovascular tissues (including kidney and whole blood) and other organs (including lung spleen and colon), (Figure 1) and in endothelial cells, respiratory epithelial cells and in PBMCs (Figure 2). *ACE* is not directly related to cellular processing of the virus but represents a pharmacological link with *ACE2*.

Relative expression of SARS-CoV-2 entry genes across organs (Figure 1 and Figure 3)

As expected, *ACE* was highly expressed in the lung³⁷ with lower but consistent levels expressed across other tissues and with very low levels present in blood. Of the two putative SARS-CoV-2 receptors, *BSG* was highly expressed across all tissues with higher levels seen in most cardiovascular tissues than in the lung or spleen. *ACE2*, across all tissues, was expressed in relatively low levels. However, cardiovascular tissues including kidney, heart and blood vessel (coronary artery) expressed higher levels of *ACE2* than the lung or spleen. Relatively low levels of *ACE2* were seen in whole blood. The colon was positioned mid group for both *BSG* and *ACE2* expression. Of the putative processing genes, required for spike protein conditioning and/or cleavage of *ACE2* allowing viral entry, *ADAM17* and *FURIN* were each enriched in the lung with relatively stable levels of expression across cardiovascular and other target tissues. *TMTRSS2* was also enriched in the lung, colon and kidney cortex with very low levels present in arteries, heart, spleen and blood. For the putative viral partner ligands of *BSG*, *PPIA* and *PPIB*, both were expressed throughout our selected tissues with higher levels expressed in arteries than kidney, blood or heart tissues. Lung, spleen and colon expressed high or midranking levels of *PPIA* and *PPIB*. The endosomal proteases, *CTSB* and *CTSL* showed similar expression patterns across our tissues of interest. Both *CTSB* and *CTSL* were enriched in the lung and spleen (*CTSB*) with relatively high levels across all tissues. *CTSB* and *CTSL* were more highly expressed in arteries than kidney, heart or blood.

Relative expression of SARS-CoV-2 entry genes in endothelial cells versus airway epithelium and PBMCs (Figure 2 and Figure 3)

To complement and extend the above organ-level approach we focussed on endothelial cells versus respiratory epithelial cell types and immune cells (PBMCs), in line with COVID-19

disease pathology. As expected, *ACE* was highly enriched in endothelial cells³⁸ with lower levels present in PBMCs, bronchial and nasal epithelium. *BSG* was enriched in endothelial cells and PBMCs with lower levels expressed in nasal and bronchial airway epithelium. *ACE2* was enriched in nasal epithelium followed by endothelial cells and lower levels in PBMCs and bronchial epithelium. *ADAM17* was highly enriched in PBMCs followed by endothelial cells and nasal and bronchial epithelial cells. *FURIN* was enriched in endothelial cells with mid-ranking levels expressed in PBMCs and lower levels in nasal and bronchial epithelium; *TMPRSS2* was highly enriched in nasal epithelial cells followed by bronchial epithelial cells, endothelial cells and PBMCs. *PPIA* and *PPIB* were highly expressed in endothelial cells with lower levels in PBMCs, nasal and bronchial epithelial cells. Intracellular proteases *CTSB* and *CTSL* were both also enriched in endothelial cells followed by airway epithelium (*CTSB*, nasal>bronchial; *CTSL*, bronchial>nasal) and low levels in PBMCs.

Next, in line with recent Public Health England's review of disparities in risks and outcomes for COVID-19¹⁷, we grouped data into two age categories, <40 and >40 years to determine differences in gene expression. Where differences were found and based on clinical evidence showing that the risk of death from COVID-19 directly correlates with age¹⁹, we performed follow-on correlation analysis.

Effect of age on expression of SARS-CoV-2 entry genes across cardiovascular and COVID-19 target tissues

Arteries:

In the aorta, *FURIN* and *PPIB* were increased while *ACE2* was decreased in samples from adults >40 years of age (Figure 3). Reductions of *ACE2* or increases of *PPIB* linearly correlated with age (Figure 4; Supplementary Figure 2). *FURIN* expression did not linearly correlate with age (Supplementary Figure 2). None of our selected genes were affected by age in the coronary artery (Figure 3).

Heart and Kidney:

In the heart (atria and left ventricle) or kidney none of our selected genes were affected by age (Figure 3).

Whole Blood:

In whole blood *ACE* expression increased while *ACE2* and *ADAM17* were reduced in samples from individuals >40 years of age (Figure 3; Figure 4; Supplementary Figure 4). Of these genes reduced expression of *ACE2* (Figure 4) and *ADAM17* and increased expression of *ACE* linearly correlated with age (Supplementary Figure 4).

Colon:

In the colon *ACE2*, *ADAM17* and *TMPRSS2* were decreased in samples from individuals >40 years of age (Figure 3; Figure 4; Supplementary Figure 5). Of these genes, reduced expression levels of *ACE2* and *TMPRSS2* but not *ADAM17* linearly correlated with age (Supplementary Figure 5).

Lung:

In the lung expression levels of *CTSL* and *PPIA* were decreased in samples from adults >40 years of age (Figure 3). Reduced expression linearly correlated with age for *PPIA* but not *CTSL* (Supplementary Figure 6).

Spleen:

None of our studied genes were altered with age in the spleen (Figure 3).

Effect of age on expression of SARS-CoV-2 entry genes in endothelial cells, airway cells and leukocytes

In endothelial cells *BSG*, but not other genes, was increased in samples from adults >40 years (Figure 3 and Figure 5; Supplementary Figure 7) and levels showed a positive linear correlation with age (Figure 5). By contrast, only *ACE* was fractionally (but statistically significantly) reduced in nasal epithelium between age categories (Figure 3; Supplementary Figure 8) and this did not linearly correlate with age (Supplementary Figure 9). No genes were altered in bronchial epithelium (Supplementary Figure 10) and PBMCs (Supplementary Figure 11) with age (Figure 3).

Summary:

In light of the cardiovascular/thrombotic sequelae associated with severe COVID-19 disease and the overwhelming risk that increased age carries, our aim was to obtain mechanistic insight by interrogating gene expression profiles in cardiovascular tissues and cells. Our focus was on the two putative receptors for SARS-CoV-2, *ACE2* and *BSG* along with a selected range of genes thought to be involved in virus binding/processing. In this study we have made four important observations: (i) Cardiovascular tissues and/or endothelial cells express the required genes for SARS-CoV-2 infection, (ii) SARS-CoV-2 receptor pathways, *ACE2/TMPRSS2* and *BSG/PP1B(A)* polarise to lung/epithelium and vessel/endothelium respectively, (iii) expression of SARS-CoV-2 host genes are, on the whole, relatively stable with age and (iv) notable exceptions were *ACE2* which decreases with age in some tissues and *BSG* which increases with age in endothelial cells.

Discussion:

Initial SARS-CoV-2 infection occurs in the airways. For most people infection is either asymptomatic or associated with mild symptoms consistent with localised viral infection in respiratory tissues. Naturally, therefore, most research to date has focused on investigations in the lung or airway cells and understanding how to manage the complications of pneumonia and ventilation failure. However, in some individuals SARS-CoV-2 infection progresses to severe COVID-19 disease, which is a systemic illness with complications specifically associated with the cardiorenal system, endothelium and thrombosis. Now, understanding the vascular component of severe COVID-19 disease associated with SARS-CoV-2 infection is emerging as an urgent unmet clinical need.

Here, we report that SARS-CoV-2 receptors and processing genes are expressed across all cardiovascular target tissues and/or in endothelial cells, supporting the idea that systemic organs contain the required machinery to be infected by the virus. In regard to the two SARS-CoV-2 receptor pathways; expression levels of *ACE2* were higher in cardiovascular tissues than the lung. By contrast, *TMPRSS2*, thought to be required for SARS-CoV-2 infection in epithelial cells, was present in lung colon and kidney but essentially absent in other cardiovascular tissues (heart, vessels and whole blood). Our findings describing the relative

levels of these genes in human tissues are in line with others using similar approaches for *ACE2*³⁹⁻⁴³ and *TMPRSS2*^{41,44}. Moreover, we found that both *ACE2* and *TMPRSS2* were enriched in nasal epithelium with low levels in bronchial epithelium and PBMCs which is in agreement with recent work from others (nasal versus bronchial epithelium^{9,10} and versus PBMCs¹⁰). Radzikowska et al., profiled a wider range of SARS-CoV-2 entry genes in immune cells and differentiated primary bronchial epithelial cells and also reported relatively high levels of expression of *PPIA*, *BSG*, *PPIB* with much lower levels of *TMPRSS2* followed by *ACE2* in airway cells¹¹. However, our focus was on the cardiovascular system. Importantly we found endothelial cells to express much lower levels of *ACE2* and *TMPRSS2* than nasal epithelium but higher levels than those expressed in PBMCs. Nevertheless, in contrast to the *ACE2/TMPRSS2* pathway genes, *BSG* and *PPIB* were expressed in higher levels in vessels and endothelial cells than in the lung and airway epithelial cells. Together our findings suggest that SARS-Cov-2 and other relevant viruses exploit *BSG* as a receptor pathway in the vasculature. Our findings align with recent work by Ganier and colleagues⁴⁵ and add evidence to the recent ‘proposed mechanism’ explaining how SARS-CoV-2 accesses endothelial cells, presented by Acosta Saltos and Acosta Saltos⁴⁶.

Severe COVID-19 is exceptionally rare in children. In adults the strongest risk factor for severe disease and death is age, with those under 40 years being at very low risk; the risk of severe COVID-19 disease and death increases proportionally after the age of 40³⁰. Of the genes that we studied, several candidates, including *ACE2*, were affected by age but with the exception of *BSG* in endothelial cells and *PPIB* and *FURIN* in aorta, expression was reduced in those >40. We found consistent age-related reductions in *ACE2* in whole blood, aorta and in the colon. Our findings are in line with those published by Chen and co-workers who also reported a negative correlation between *ACE2* and age in a range of tissues including colon and blood³⁹. Moreover, our work corroborates earlier studies showing that *ACE2* (protein) declines with age in mouse aorta⁴⁷. Other studies in rats also showed that *ACE2* declines with age in the lung and kidney^{48,49}. It should be noted, however, that Li and colleagues found no effect of age on *ACE2* expression across a similar selection of tissues⁴³ and that Santesmasses and colleagues found that *ACE2* expression increased with age in the lung⁵⁰. We also found a trend for *ACE2* to increase in the lung but this did not reach statistical significance in our study. Key

differences between the studies include the analytical approaches applied, the number of tissues selected, and the age groups used.

Since ACE2 is a receptor for SARS-CoV-2, which declines with age in some settings (this study)^{43,47-49} and because age is the strongest predictor for fatal COVID-19 disease a paradox has emerged⁵¹: *SARS-CoV-2 receptor expression does not positively correlate with high risk groups of severe COVID-19 disease*. One explanation of the paradox has been that because ACE2 is a cardioprotective enzyme, while ACE2 is the receptor for airway infection of SARS-CoV-2, low levels of ACE2 in the circulation of the elderly and those with cardiovascular disease increases the risk of cardiovascular complications associated with severe COVID-19 disease⁵¹. Additionally, it has been hypothesised that BSG acts as a receptor for SARS-CoV-2 in endothelial cells⁴⁶. BSG expression is increased in a range of cardiovascular diseases which could compensate for any age/disease associated reductions in ACE2 in regard to viral infection. In line with this in our study, *BSG* positively correlated with age and this association was only seen in endothelial cells. Others have found that BSG increases with age in the skin⁵². It is not clear why we did not see age related increases in *BSG* in the aorta or coronary artery or in organ samples. One explanation could be that in vessels the delicate lining of the endothelium may have been lost during tissue dissection and/or that the age effects on *BSG* expression in endothelial cells is diluted out in complex tissues by expression levels in other cells which make up the bulk of the samples. It should also be considered that whilst blood outgrowth endothelial cells display key ubiquitous features of endothelial cells and retain disease phenotypes³¹, heterogeneity exists within endothelial cells as they age with passage and in different vascular beds.

Based on our findings together the cardiovascular, renal and thrombotic complications associated with severe COVID-19 disease, we suggest that BSG expression in the pulmonary and systemic vasculature maybe an important driver which explains the heightened risk of severe disease and death observed in those >40 years of age. These observations add to the growing evidence and provide additional mechanistic insight supporting the targeting the BSG: PPIA/PPIB axis in severe COVID-19 disease.

BSG is an attractive lead to have emerged from this work. BSG is upregulated in a range of diseases including those co-morbidities/morbidities associated with increased risk of severe COVID-19 disease and poorer outcomes including thrombosis⁵³, pulmonary hypertension⁵⁴, renal disease⁵⁵, obesity¹¹ and diabetes⁵⁶. Moreover BSG, as both a receptor for SARS-Cov-2 and as an inducer of extracellular matrix metalloproteinases, may be relevant when considering the potential role of the vasculature and, specifically endothelial dysfunction, in propagating and driving pulmonary fibrosis^{57,58} – one of the most feared long term complications of COVID-19⁵⁹.

Clearly, a full understanding of BSG interactions with SARS-CoV-2 will provide valuable mechanistic insight and could identify new therapeutic targets and/or provide additional insight for experimental drugs currently in trials for the prevention and treatment of severe COVID-19 disease. For example, cyclosporin, an immunosuppressive drug used in the prevention of organ rejection after transplant, has been suggested as a therapy for cytokine storm associated with severe COVID-19 disease. To this end clinical trials to assess cyclosporin drugs in COVID-19 have been initiated^{42,43}. Cyclosporin also binds PPIA and therefore may additionally inhibit viral binding to BSG. Our findings that BSG expression is increased with age may provide additional rational for the use of drugs that interfere with cyclophilin binding. In light of our findings and the overwhelming clinical information indicating vascular inflammation in severe COVID-19 disease, the relative role of SARS-CoV-2 receptors and processing proteins in endothelial cells should be investigated. Our data indicates that blood outgrowth endothelial cells will be a useful tool in future work exploring mechanisms of viral infection and inflammation in COVID-19. In addition, since blood outgrowth endothelial cells can be obtained from blood samples of living donors, functional assays using these cells from protected and at-risk populations may provide a means of identifying personalised therapies and those at risk of severe disease.

Since BSG is utilised by other pathogens our findings have implication beyond the current pandemic. Finally, because BSG is implicated in a range of cardiovascular diseases and fibrosis our observations may have relevance to our understanding of the diseases associated with aging.

Limitations and future studies:

The details of the cellular events involved in SARS-CoV-2 infection in different cell types has not yet been fully established which means that the weighted relevance of genes that we have investigated in COVID-19 remains to be elucidated. However, each of the genes that we studied are established in various aspects of human physiology and pathology and so our analysis has relevance to the understanding of aging in a setting wider than infection. We used publicly available data sets which, while were analysed in a systematic and unbiased manner, require replication and validation in a prospective follow up clinical and mechanistic studies. Furthermore, gene expression data invariably requires biological validation in functional assays.

Figure 1: Expression of SARS-CoV-2 entrance/processing genes in different cardiovascular and non-cardiovascular tissues

Standardised expression levels for the genes *ACE*(A), *BSG* (B), *ACE2* (C), *ADAM17* (D), *FURIN* (E), *TMPRSS2* (F), *PPIA* (G), *PPIB* (H), *CTSB* (I) and *CTSL* (J) were obtained from GTEx from cardiovascular tissues (aorta, coronary artery, atrial appendage, left ventricle, kidney cortex, and whole blood; red columns) and other tissues (lung, blue columns; spleen, grey columns and colon, green columns). Data for each tissue was corrected for batch effects COMBAT-seq and expressed as mean \pm S.E.M. Tissues were ranked in order of expression for each gene. A heat map showing expression of ACE2 and BSG pathways and viral processing proteases in each tissue was generated (K). Data were coloured by gene, whereby black is the lowest expressing tissue and red is highest expressing tissue.

Figure 2: Expression of SARS-CoV-2 entrance/processing genes in blood outgrowth endothelial cells, PBMCs and epithelial cells (nasal and bronchial)

Standardised expression levels for the genes *ACE*(A), *BSG* (B), *ACE2* (C), *ADAM17* (D), *FURIN* (E), *TMPRSS2* (F), *PPIA* (G), *PPIB* (H), *CTSB* (I) and *CTSL* (J) were obtained from online databases from human blood outgrowth endothelial cells (Endothelial cells, red columns), peripheral blood mononuclear cells (PBMCs, grey columns) and epithelial cells (nasal and bronchial, blue columns). The data was aligned and analysed using PartekFlow® and corrected for batch effects using COMBAT-seq and expressed as mean \pm S.E.M. Cells were ranked in order of expression each gene. A heat map showing expression of ACE2 and BSG pathways and viral processing proteases in each cell type was generated (K). Data were coloured by gene, whereby black is the lowest expressing cell type and red is highest expressing cell type.

Figure 3: Heat map representing expression of SARS-CoV-2 entrance/processing genes significantly altered by age

Heatmaps were generated for the expression of *ACE*, *ACE2*, *BSG*, *ADAM17*, *CTSB*, *CTSL*, *FURIN*, *PPIA*, *PPIB* and *TMPRSS2* in cells (A) and organs (B). Data were analysed based on two adult age groups; under or over 40 years. Data were analysed using an unpaired Mann-Whitney t-test or unpaired t-test depending on normality distributions. Significant data ($p < 0.05$) are shown as either increased (red) or decreased (green) expression; black corresponds to no significant change.

Figure 4: The effect of age on *ACE2* expression in aorta (A,B), whole blood (C,D) and colon (E,F)

ACE2 levels in aorta (A,B), whole blood (C,D) and colon (E,F) were analysed based on adults under vs over 40-year-olds (A, C, E). Data is expressed as mean \pm S.E.M. and analysed using unpaired Mann-Whitney T-test (A,C,E) or Spearman's correlation test (B,D,F); significance was accepted when $*p < 0.05$.

Figure 5: The effect of age on *BSG* expression in blood outgrowth endothelial cells

BSG levels in blood outgrowth endothelial cells (Endothelial cells) were analysed in adults under 40 years (<40) versus over 40 years (>40) using an unpaired Mann-Whitney T-test (A) and correlations with age determined using Pearson's correlation analysis (B); significance was accepted when $*p < 0.05$.

Supplementary Table 1: Identified cell studies. Raw transcriptomic gene expression profiling datasets were obtained from ArrayExpress and NCBI GEO for blood outgrowth endothelial cells (BOECs; also known as 'late outgrowth endothelial cells'), peripheral blood mononuclear cells (PBMCs) and bronchial airway (obtained from bronchial brushing) and nasal (obtained nasal brushing) epithelium.

Supplementary Figure 1: Principle component analysis (PCA) of genes of interest pre and post COMBAT-Seq normalisation. PCA plots for all organs (A) and cells (B) pre and post COMBAT-Seq normalisation were generated.

Supplementary Figure 2: The effect of age on *FURIN* (A,B) and *PPIPB* (C,D) in aorta

Gene expression levels in aorta are shown as mean \pm S.E.M and were analysed based on adults under vs over 40-year-olds using an unpaired Mann-Whitney T-test (A,C) and correlations with age determined using Spearman's correlation test analysis (B,D); significance was accepted when $*p < 0.05$.

Supplementary Figure 3: The effect of age *CTSL* heart left ventricle (heartLV)

Gene expression levels in heartLV are shown as mean \pm S.E.M and were analysed in adults under 40 years (<40) versus over 40 years (>40) using an unpaired Mann-Whitney T-test (A) and correlations with age determined using Spearman's correlation test analysis (B); significance was accepted when $*p < 0.05$.

Supplementary Figure 4: The effect of age on ACE (A,B) and ADAM17 (C,D) in whole blood

Gene expression levels in whole blood are shown as mean \pm S.E.M and were analysed in adults under 40 years (<40) versus over 40 years (>40) using an unpaired Mann-Whitney T-test (A,C) and correlations with age determined using Spearman's correlation test analysis (B,D); significance was accepted when $*p < 0.05$.

Supplementary Figure 5: The effect of age on ADAM17 (A,B) and TMPRSS2 (C,D) in colon

Gene expression levels in colon were analysed in adults under 40 years (<40) versus over 40 years (>40) using an unpaired Mann-Whitney T-test (A,C) and correlations with age determined using Spearman's correlation test analysis (B,D); significance was accepted when $*p < 0.05$.

Supplementary Figure 6: The effect of age on CTSL (A,B) and PPIA (C,D) in lung

Gene expression levels in lung were analysed in adults under 40 years (<40) versus over 40 years (>40) using an unpaired Mann-Whitney T-test (A,C) and correlations with age determined using Spearman's correlation test analysis (B,D); significance was accepted when $*p < 0.05$.

Supplementary Figure 7: The effect of age (adult under 40 years; <40 versus over 40 years; >40) on of SARS-CoV-2 entrance/processing genes in blood outgrowth endothelial cells

Data is presented as individual points for each gene and analysed using Students T-Test or Mann-Whitney T-test as appropriated; significance was accepted when $*p < 0.05$.

Supplementary Figure 8: The effect of age (adult under 40 years; <40 versus over 40 years; >40) on of SARS-CoV-2 entrance/processing genes in nasal epithelial cells

Data is shown as individual points for each gene and analysed using Students T-Test or Mann-Whitney T-test as appropriated; significance was accepted when $*p < 0.05$.

Supplementary Figure 9: Correlation analysis of ACE expression in nasal epithelial cells with age

Data is shown as individual points for *ACE* expression levels and age (years). Data was analysed using Pearson's correlation analysis.

Supplementary Figure 10: The effect of age (adult under 40 years; <40 versus over 40 years; >40) on of SARS-CoV-2 entrance/processing genes in bronchial epithelial cells

Data is shown as individual points for each gene and analysed using Students T-Test or Mann-Whitney T-test as appropriated; significance was accepted when $*p < 0.05$.

Supplementary Figure 11: The effect of age (adult under 40 years; <40 versus over 40 years; >40) on of SARS-CoV-2 entrance/processing genes in peripheral blood mononuclear cells (PBMCs).

Data is shown as individual points for each gene and analysed using Students T-Test or Mann-Whitney T-test as appropriated; significance was accepted when $*p < 0.05$.

Figure 1

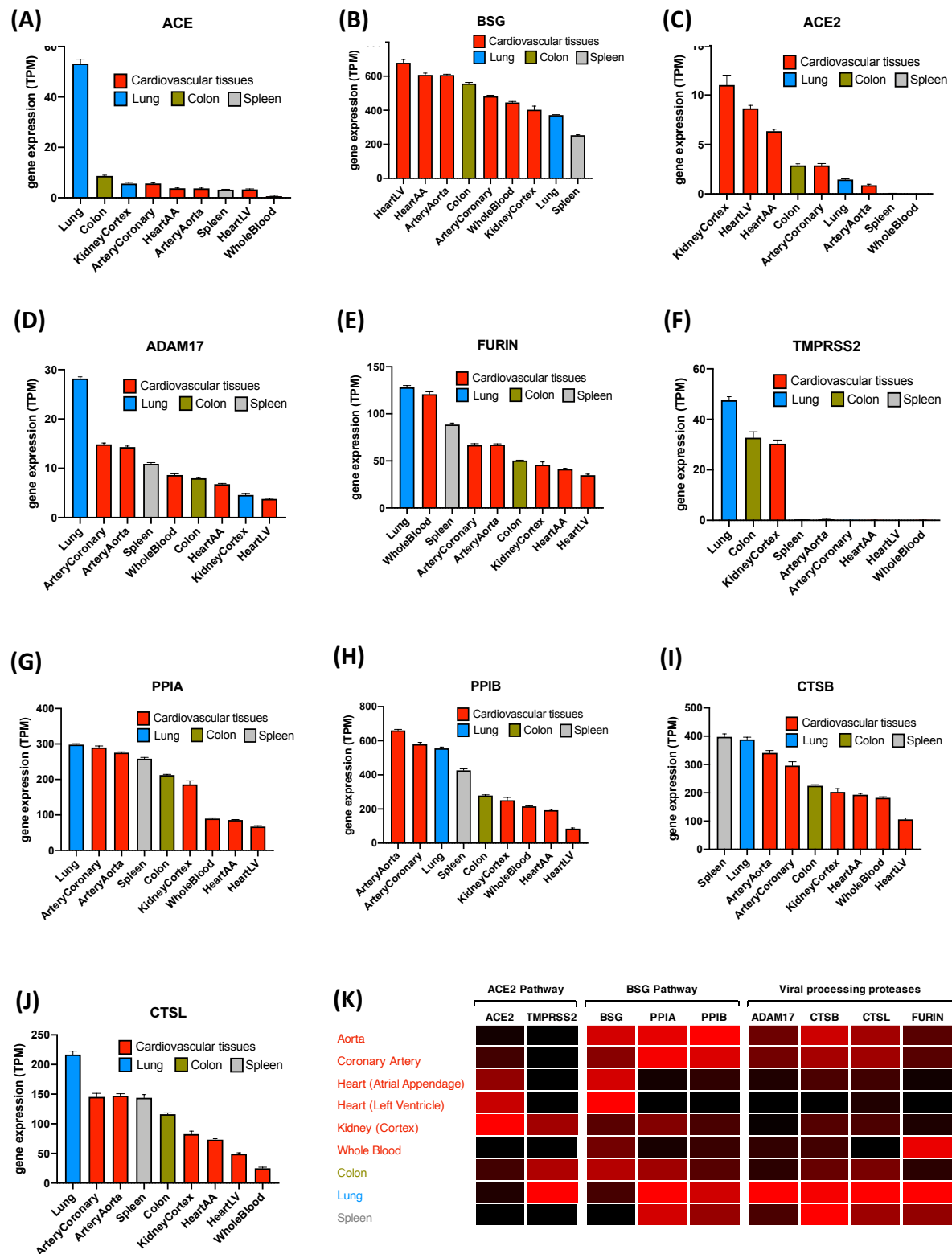


Figure 3:

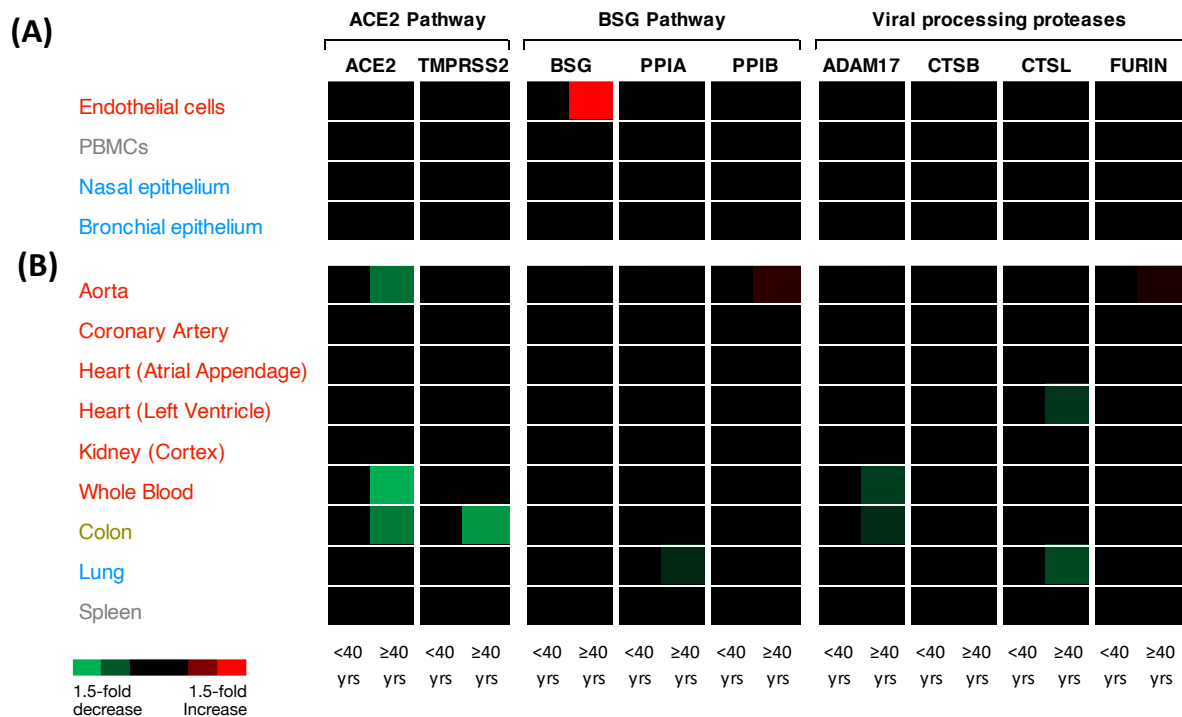


Figure 4:

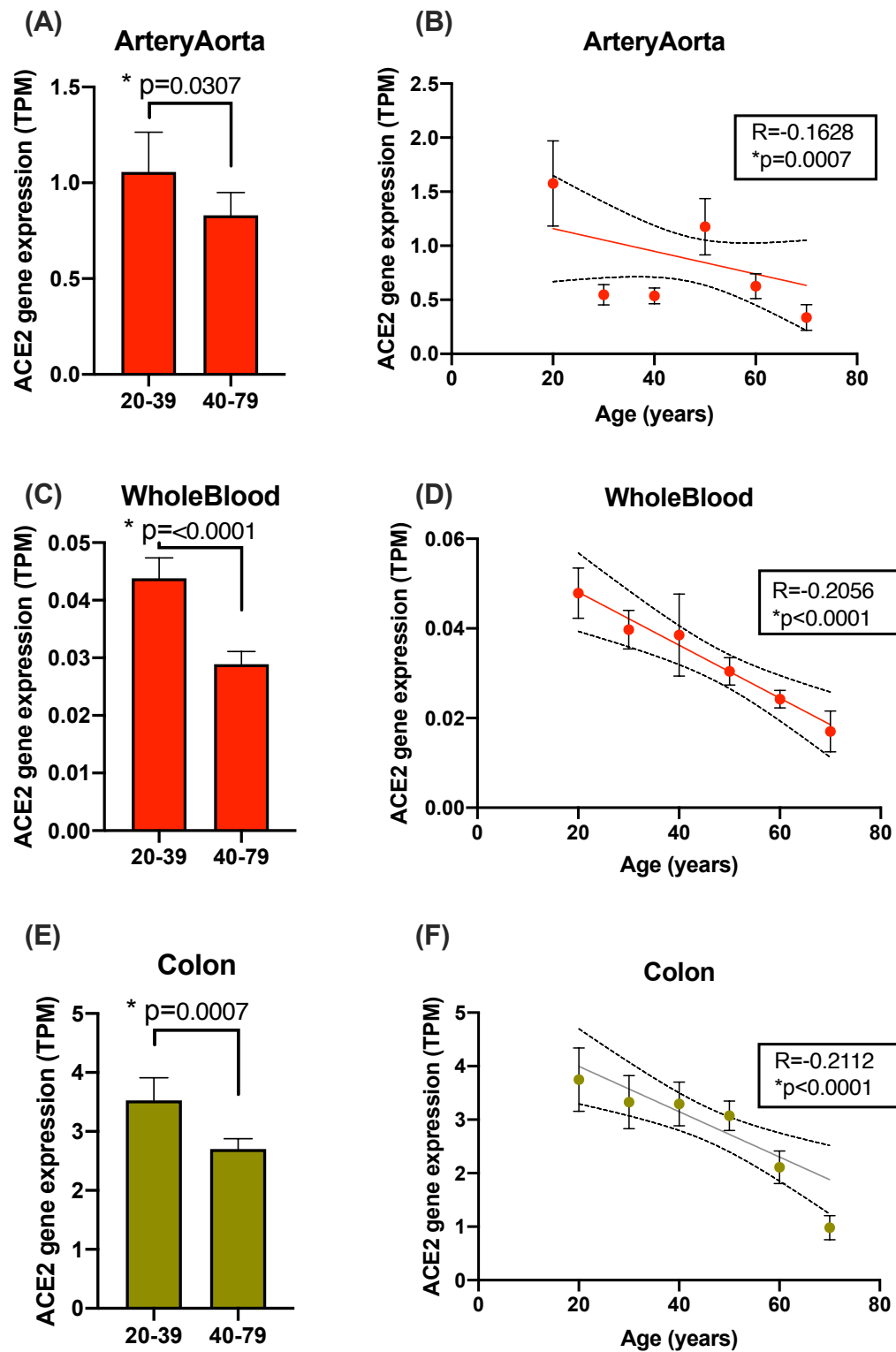
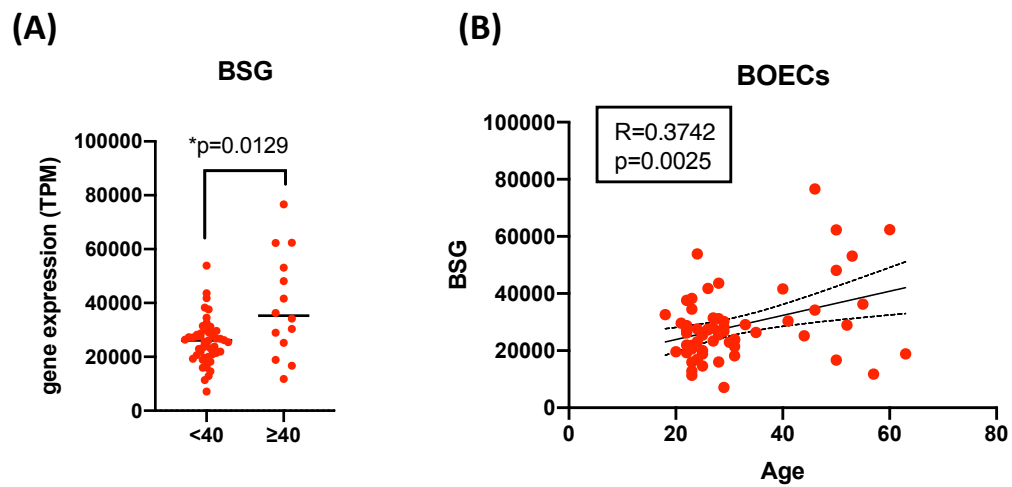


Figure 5:



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. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).
- 2 Ulrich, H. & Pillat, M. M. CD147 as a Target for COVID-19 Treatment: Suggested Effects of Azithromycin and Stem Cell Engagement. *Stem Cell Rev Rep*, doi:10.1007/s12015-020-09976-7 (2020).
- 3 Wang, K. *et al.* SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. *BioRxiv preprint.*, doi:10.1101/2020.03.14.988345 (2020).
- 4 Coutard, B. *et al.* The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* **176**, 104742, doi:10.1016/j.antiviral.2020.104742 (2020).
- 5 Chen, Z. *et al.* Function of HAb18G/CD147 in invasion of host cells by severe acute respiratory syndrome coronavirus. *J Infect Dis* **191**, 755-760, doi:10.1086/427811 (2005).
- 6 Pushkarsky, T. *et al.* CD147 facilitates HIV-1 infection by interacting with virus-associated cyclophilin A. *Proc Natl Acad Sci U S A* **98**, 6360-6365, doi:10.1073/pnas.111583198 (2001).
- 7 Watanabe, A. *et al.* CD147/EMMPRIN acts as a functional entry receptor for measles virus on epithelial cells. *J Virol* **84**, 4183-4193, doi:10.1128/JVI.02168-09 (2010).
- 8 Prakash, P. *et al.* Human Cyclophilin B forms part of a multi-protein complex during erythrocyte invasion by Plasmodium falciparum. *Nat Commun* **8**, 1548, doi:10.1038/s41467-017-01638-6 (2017).
- 9 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, doi:10.1038/s41591-020-0868-6 (2020).
- 10 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* **18**, 1-6, doi:10.1016/j.omtm.2020.05.013 (2020).
- 11 Radzikowska, U. *et al.* Distribution of ACE2, CD147, CD26 and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. *Allergy*, doi:10.1111/all.14429 (2020).
- 12 Teuwen, L. A., Geldhof, V., Pasut, A. & Carmeliet, P. COVID-19: the vasculature unleashed. *Nat Rev Immunol*, doi:10.1038/s41577-020-0343-0 (2020).
- 13 Leisman, D. E., Deutschman, C. S. & Legrand, M. Facing COVID-19 in the ICU: vascular dysfunction, thrombosis, and dysregulated inflammation. *Intensive Care Med* **46**, 1105-1108, doi:10.1007/s00134-020-06059-6 (2020).
- 14 Mestres, G. *et al.* Risk of peripheral arterial thrombosis in COVID-19. *J Vasc Surg*, doi:10.1016/j.jvs.2020.04.477 (2020).
- 15 Lodigiani, C. *et al.* Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. *Thromb Res* **191**, 9-14, doi:10.1016/j.thromres.2020.04.024 (2020).
- 16 Middeldorp, S. *et al.* Incidence of venous thromboembolism in hospitalized patients with COVID-19. *J Thromb Haemost*, doi:10.1111/jth.14888 (2020).
- 17 Wichmann, D. *et al.* Autopsy Findings and Venous Thromboembolism in Patients With COVID-19. *Ann Intern Med*, doi:10.7326/M20-2003 (2020).

- 18 Somani, S. *et al.* Characterization of Patients Who Return to Hospital Following Discharge from Hospitalization For COVID-19. *medRxiv*, doi:10.1101/2020.05.17.20104604 (2020).
- 19 Su, Y. B. *et al.* Cardiovascular Manifestation and Treatment in COVID-19. *J Chin Med Assoc*, doi:10.1097/JCMA.0000000000000352 (2020).
- 20 Wang, Y. *et al.* Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. *Radiology*, 200843, doi:10.1148/radiol.2020200843 (2020).
- 21 Mo, X. *et al.* Abnormal pulmonary function in COVID-19 patients at time of hospital discharge. *Eur Respir J*, doi:10.1183/13993003.01217-2020 (2020).
- 22 Huertas, A. *et al.* Endothelial cell dysfunction: a major player in SARS-CoV-2 infection (COVID-19)? *Eur Respir J*, doi:10.1183/13993003.01634-2020 (2020).
- 23 Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913 e907, doi:10.1016/j.cell.2020.04.004 (2020).
- 24 Gu, J. *et al.* Multiple organ infection and the pathogenesis of SARS. *J Exp Med* **202**, 415-424, doi:10.1084/jem.20050828 (2005).
- 25 Ng, L. F. *et al.* Detection of severe acute respiratory syndrome coronavirus in blood of infected patients. *J Clin Microbiol* **42**, 347-350, doi:10.1128/jcm.42.1.347-350.2004 (2004).
- 26 Wang, W. *et al.* Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*, doi:10.1001/jama.2020.3786 (2020).
- 27 Young, B. E. *et al.* Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA*, doi:10.1001/jama.2020.3204 (2020).
- 28 Chang, L., Yan, Y. & Wang, L. Coronavirus Disease 2019: Coronaviruses and Blood Safety. *Transfus Med Rev* **34**, 75-80, doi:10.1016/j.tmr.2020.02.003 (2020).
- 29 Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* **369**, m1985, doi:10.1136/bmj.m1985 (2020).
- 30 .
doi:https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892085/disparities_review.pdf.
- 31 Paschalaki, K. E. & Randi, A. M. Recent Advances in Endothelial Colony Forming Cells Toward Their Use in Clinical Translation. *Front Med (Lausanne)* **5**, 295, doi:10.3389/fmed.2018.00295 (2018).
- 32 George, P. M. *et al.* Evidence for the involvement of type I interferon in pulmonary arterial hypertension. *Circ Res* **114**, 677-688, doi:10.1161/CIRCRESAHA.114.302221 (2014).
- 33 Ahmetaj-Shala, B. *et al.* A bioassay system of autologous human endothelial, smooth muscle cells, and leukocytes for use in drug discovery, phenotyping, and tissue engineering. *FASEB J* **34**, 1745-1754, doi:10.1096/fj.201901379RR (2020).
- 34 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).
- 35 Wang J., Huang M, Torre, M. & et al. ComBat-Seq: batch effect adjustment for RNA-Seq count data. *bioRxiv*, doi:<https://doi.org/10.1101/2020.01.13.904730>.

- 36 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).
- 37 Ng, K. K. & Vane, J. R. Fate of angiotensin I in the circulation. *Nature* **218**, 144-150, doi:10.1038/218144a0 (1968).
- 38 Ryan, J. W., Ryan, U. S., Schultz, D. R., Whitaker, C. & Chung, A. Subcellular localization of pulmonary antiotensin-converting enzyme (kininase II). *Biochem J* **146**, 497-499, doi:10.1042/bj1460497 (1975).
- 39 Chen, J. *et al.* Individual variation of the SARS-CoV-2 receptor ACE2 gene expression and regulation. *Aging Cell*, doi:10.1111/accel.13168 (2020).
- 40 Chen, L., Li, X., Chen, M., Feng, Y. & Xiong, C. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. *Cardiovasc Res* **116**, 1097-1100, doi:10.1093/cvr/cvaa078 (2020).
- 41 Song, J. *et al.* Systematic analysis of ACE2 and TMPRSS2 expression in salivary glands reveals underlying transmission mechanism caused by SARS-CoV-2. *J Med Virol*, doi:10.1002/jmv.26045 (2020).
- 42 Dai, Y. J. *et al.* A profiling analysis on the receptor ACE2 expression reveals the potential risk of different type of cancers vulnerable to SARS-CoV-2 infection. *Ann Transl Med* **8**, 481, doi:10.21037/atm.2020.03.61 (2020).
- 43 Li, M. Y., Li, L., Zhang, Y. & Wang, X. S. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty* **9**, 45, doi:10.1186/s40249-020-00662-x (2020).
- 44 Paniri, A., Hosseini, M. M. & Akhavan-Niaki, H. First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. *J Biomol Struct Dyn*, 1-18, doi:10.1080/07391102.2020.1767690 (2020).
- 45 Ganier, D.-H. X., Harun N *et al.*, CD147 (BSG) but not ACE2 expression is detectable in vascular endothelial cells within single cell RNA sequencing datasets derived from multiple tissues in healthy individuals. *bioRxiv*, <https://doi.org/10.1101/2020.1105.1129.123513> (2020).
- 46 Acosta Saltos, F. & Acosta Saltos, A. Entry of SARS-CoV2 through the Basal Surface of Alveolar Endothelial Cells – A Proposed Mechanism Mediated by CD147 in COVID-19. *Preprints.org*, doi:doi: 10.20944/preprints202005.0359.v1.
- 47 Yoon, H. E. *et al.* Age-Associated Changes in the Vascular Renin-Angiotensin System in Mice. *Oxid Med Cell Longev* **2016**, 6731093, doi:10.1155/2016/6731093 (2016).
- 48 Xie, X., Chen, J., Wang, X., Zhang, F. & Liu, Y. Age- and gender-related difference of ACE2 expression in rat lung. *Life Sci* **78**, 2166-2171, doi:10.1016/j.lfs.2005.09.038 (2006).
- 49 Schulman, I. H. *et al.* Altered renal expression of angiotensin II receptors, renin receptor, and ACE-2 precede the development of renal fibrosis in aging rats. *Am J Nephrol* **32**, 249-261, doi:10.1159/000318607 (2010).
- 50 Santesmasses D, C. P., Zenin AR *et al.*,
COVID-19 is an emergent disease of aging.
<https://www.medrxiv.org/content/10.1101/2020.04.15.20060095v1.full.pdf> (2020).
- 51 AlGhatrif, M., Cingolani, O. & Lakatta, E. G. The Dilemma of Coronavirus Disease 2019, Aging, and Cardiovascular Disease: Insights From Cardiovascular Aging Science. *JAMA Cardiol*, doi:10.1001/jamacardio.2020.1329 (2020).

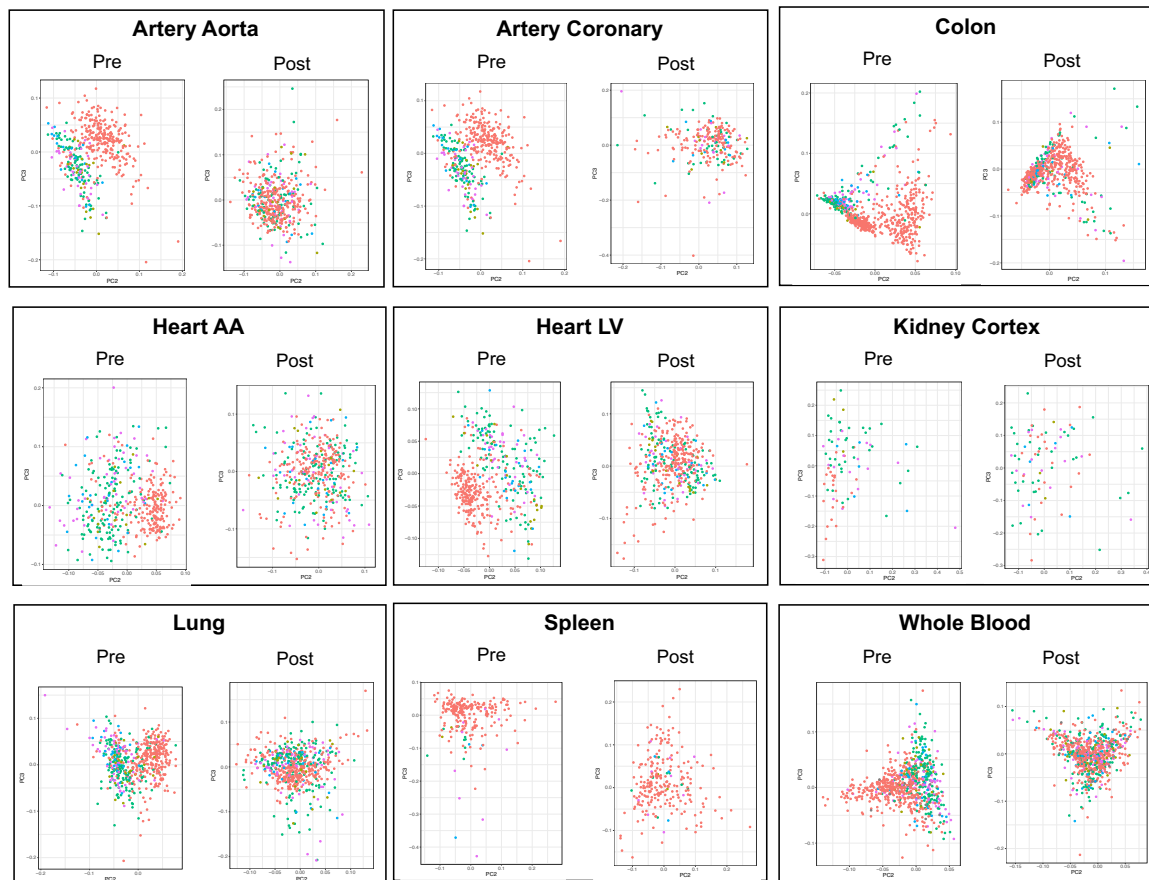
- 52 Li, J. *et al.* Expression of cyclophilin A and CD147 during skin aging. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **36**, 203-211, doi:10.3969/j.issn.1672-7347.2011.03.003 (2011).
- 53 Pennings, G. J. & Kritharides, L. CD147 in cardiovascular disease and thrombosis. *Semin Thromb Hemost* **40**, 747-755, doi:10.1055/s-0034-1390001 (2014).
- 54 Satoh, K. *et al.* Basigin mediates pulmonary hypertension by promoting inflammation and vascular smooth muscle cell proliferation. *Circ Res* **115**, 738-750, doi:10.1161/CIRCRESAHA.115.304563 (2014).
- 55 Kosugi, T., Maeda, K., Sato, W., Maruyama, S. & Kadomatsu, K. CD147 (EMMPRIN/Basigin) in kidney diseases: from an inflammation and immune system viewpoint. *Nephrol Dial Transplant* **30**, 1097-1103, doi:10.1093/ndt/gfu302 (2015).
- 56 Chiu, P. F. *et al.* Cyclophilin A and CD147 associate with progression of diabetic nephropathy. *Free Radic Res* **52**, 1456-1463, doi:10.1080/10715762.2018.1523545 (2018).
- 57 Guillot, S. *et al.* Increased extracellular matrix metalloproteinase inducer (EMMPRIN) expression in pulmonary fibrosis. *Exp Lung Res* **32**, 81-97, doi:10.1080/01902140600710512 (2006).
- 58 George, P. M. & Mitchell, J. A. Defining a pathological role for the vasculature in the development of fibrosis and pulmonary hypertension in interstitial lung disease. *Am J Physiol Lung Cell Mol Physiol* **317**, L431-L433, doi:10.1152/ajplung.00330.2019 (2019).
- 59 George, P. M., Wells, A. U. & Jenkins, R. G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir Med*, doi:10.1016/S2213-2600(20)30225-3 (2020).

Supplementary Table 1:

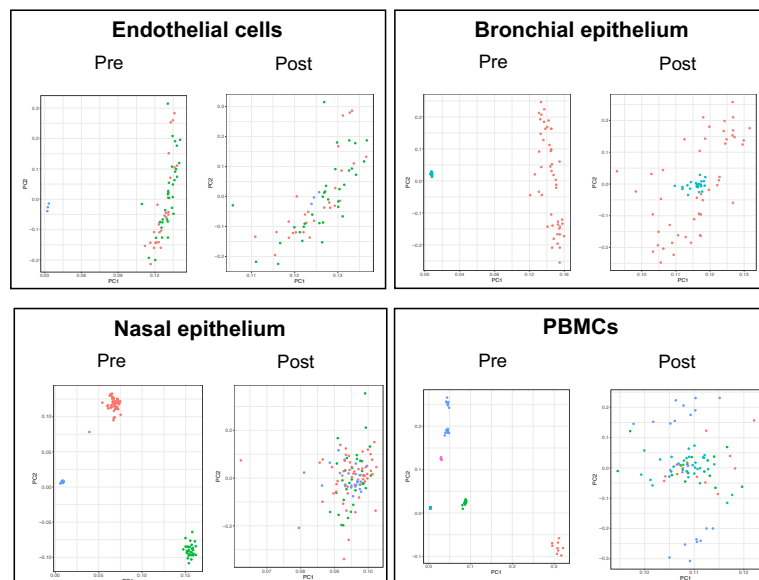
Cell type	Study ID	Gender		Donors in age category		Mean age (yrs)	Datasource weblink
BOECs (late outgrowth endothelial cells)	GEOD9877	M	12	<40	18	38.1	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-9877/
		F	15	≥40	9		
	GEOD22688	M	18	<40	32	25.0	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-22688/
		F	15	≥40	1		
	GEOD38961	M	1	<40	2	35.7	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-38961/
		F	2	≥40	1		
PBMCs (whole blood derived)	GEOD49641	M	10	<40	1	60.4	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-49641/
		F	8	≥40	17		
	GEOD17114	M	7	<40	9	36.7	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-17114/
		F	7	≥40	5		
	GEOD6645	M	4	<40	0	58.3	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66465
		F	0	≥40	4		
	MEXP1635	M	0	<40	5	43.1	https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1635/
		F	9	≥40	4		
	GEOD38958	M	24	<40	0	69.3	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-38958/
		F	11	≥40	35		
Nasal (scrapings/brushings)	GEOD11348	M	11	<40	31	21.0	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-11348/
		F	20	≥40	0		
	EMTAB4015	M	32	<40	0	55.5	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4015/
		F	22	≥40	54		
	GEOD16008	M	11	<40	18	33.5	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-16008/
		F	15	≥40	8		
Bronchial (brushings)	GEOD16008	M	11	<40	18	33.8	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-16008/
		F	15	≥40	8		
	GEOD19667	M	N/A	<40	26	39.1	https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-19667/
		F	N/A	≥40	22		

Supplementary Figure 1:

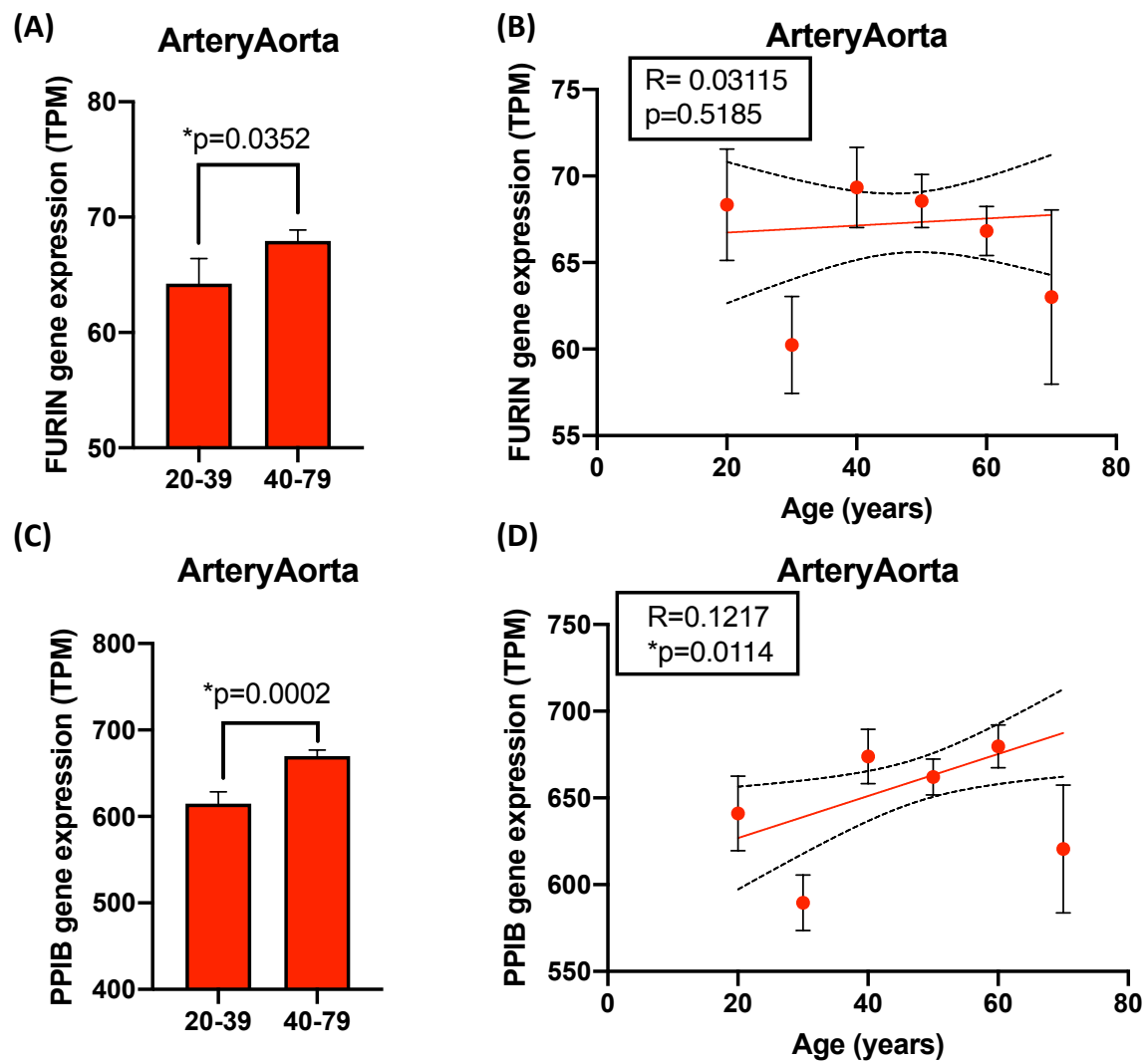
(A) Tissues



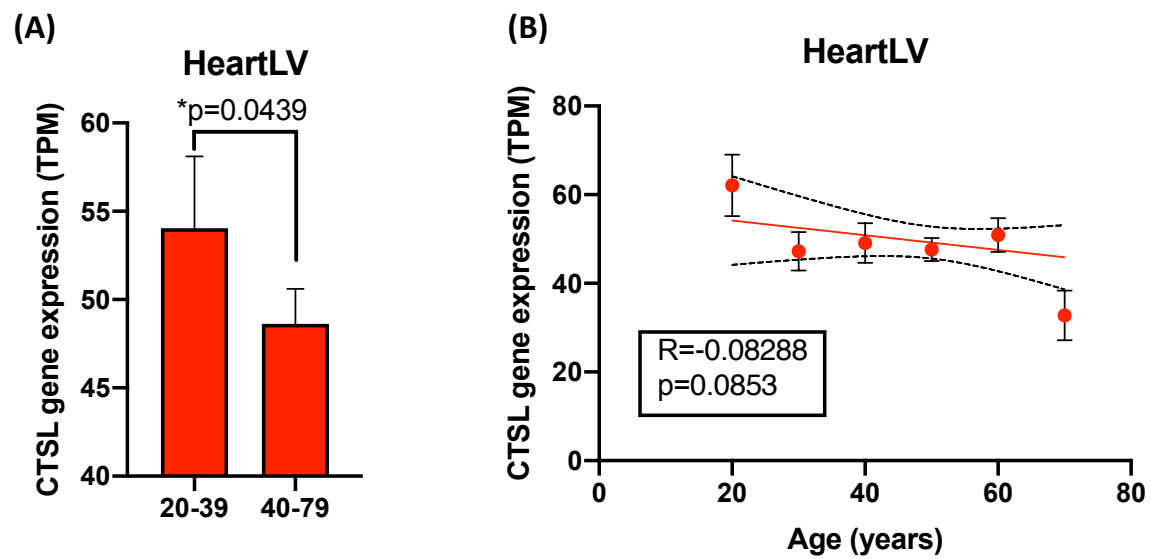
(B) Cells



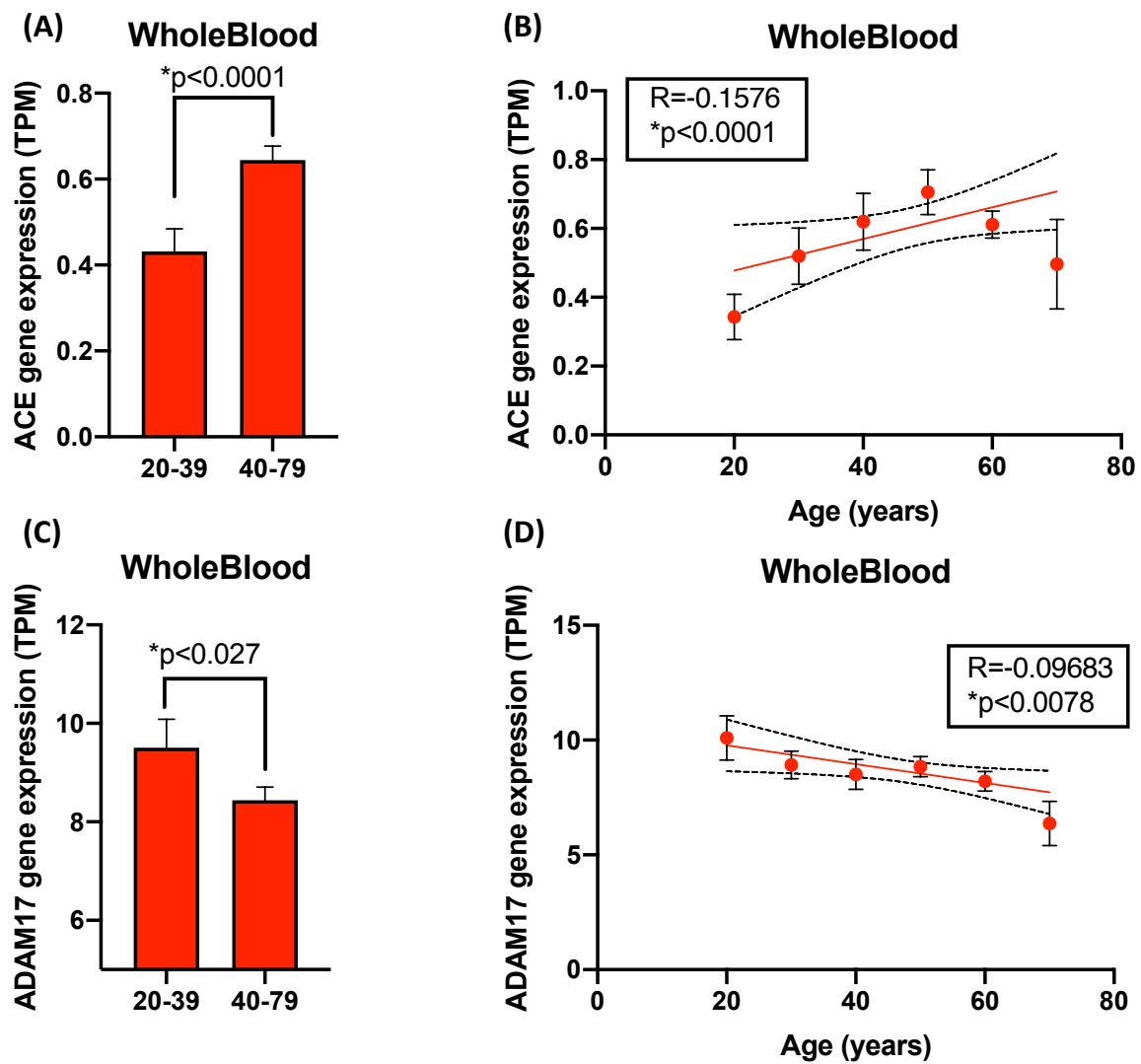
Supplementary Figure 2:



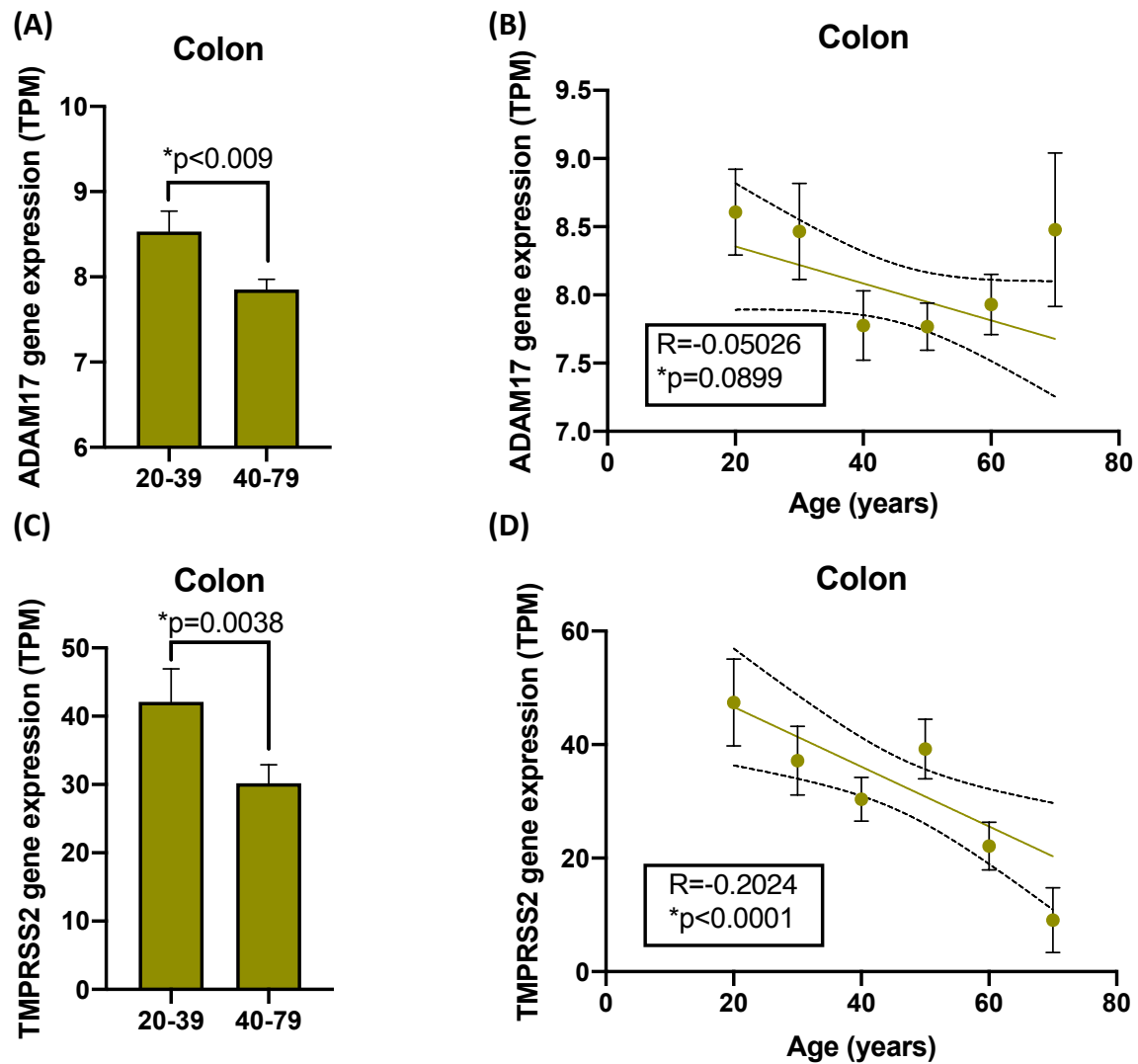
Supplementary Figure 3:



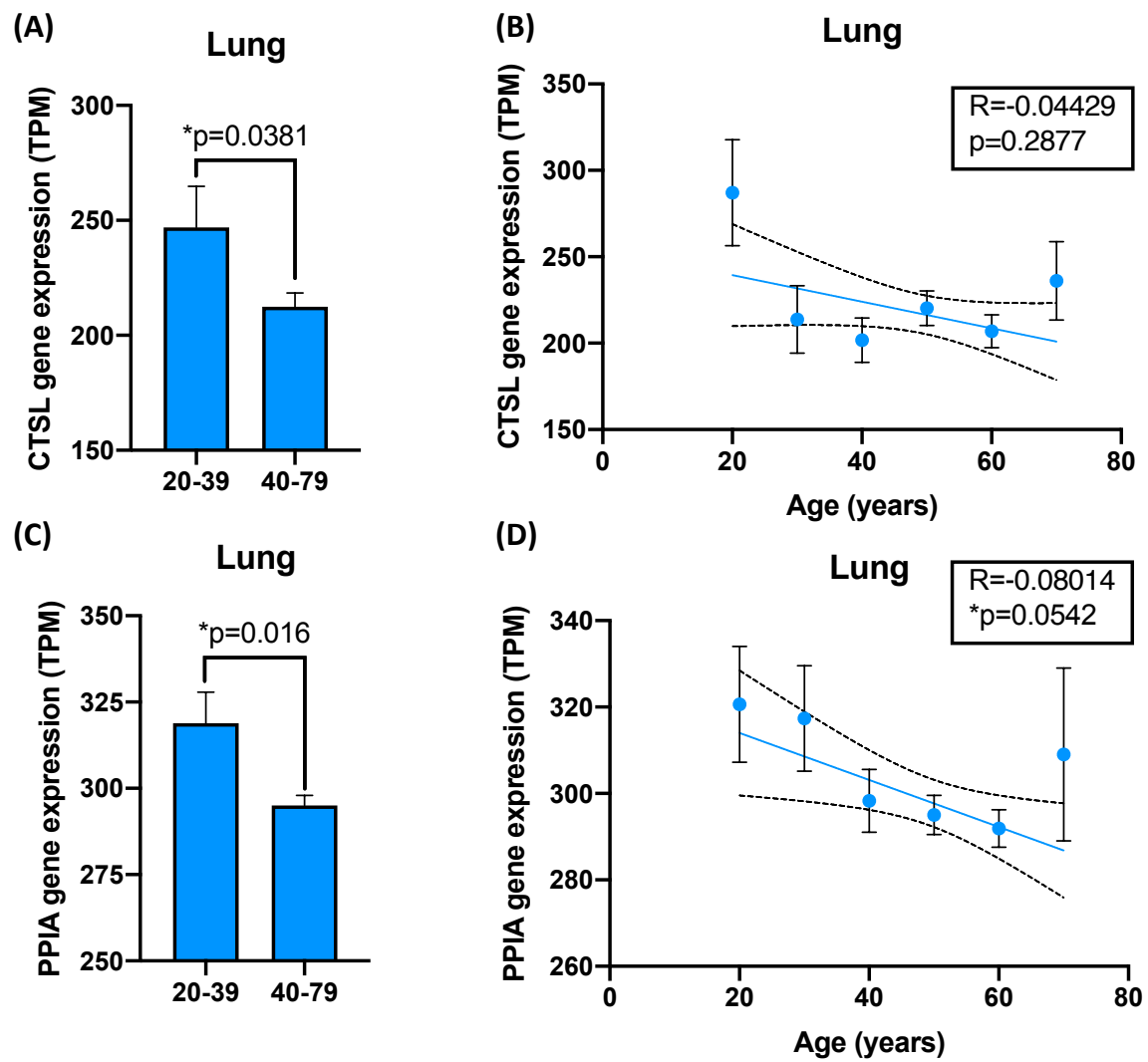
Supplementary Figure 4:



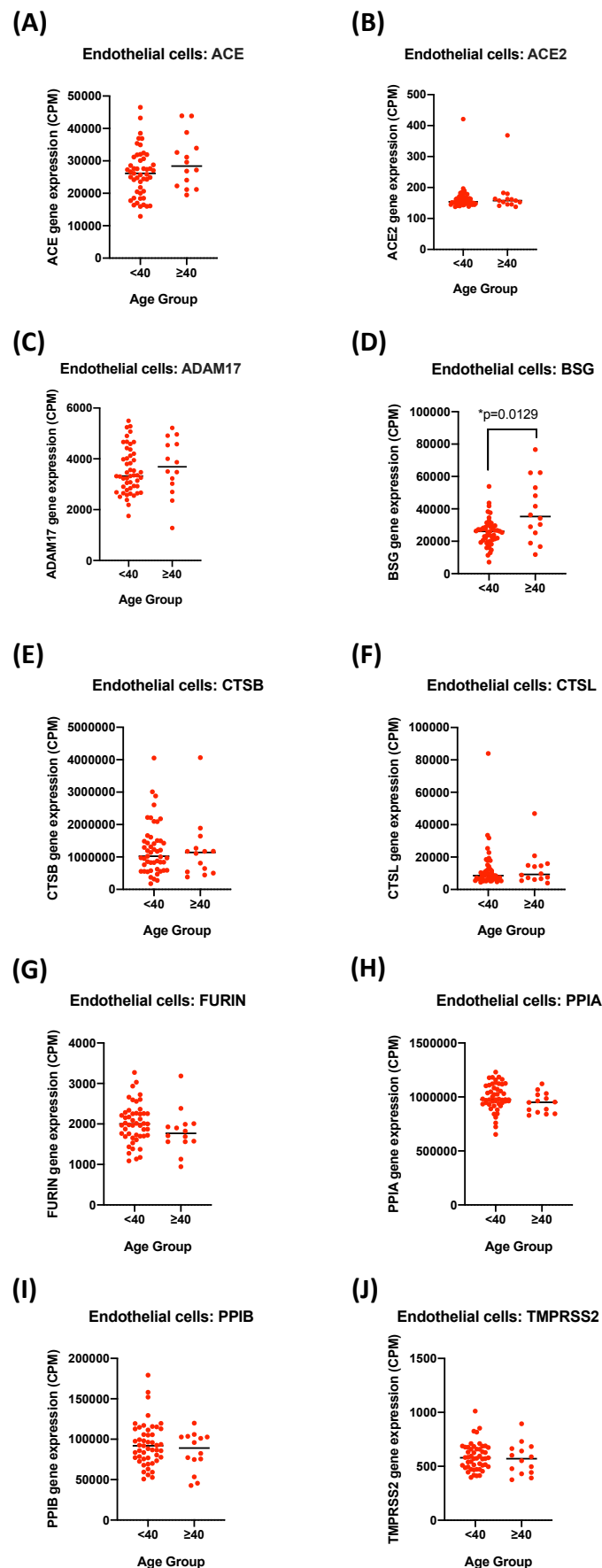
Supplementary Figure 5:



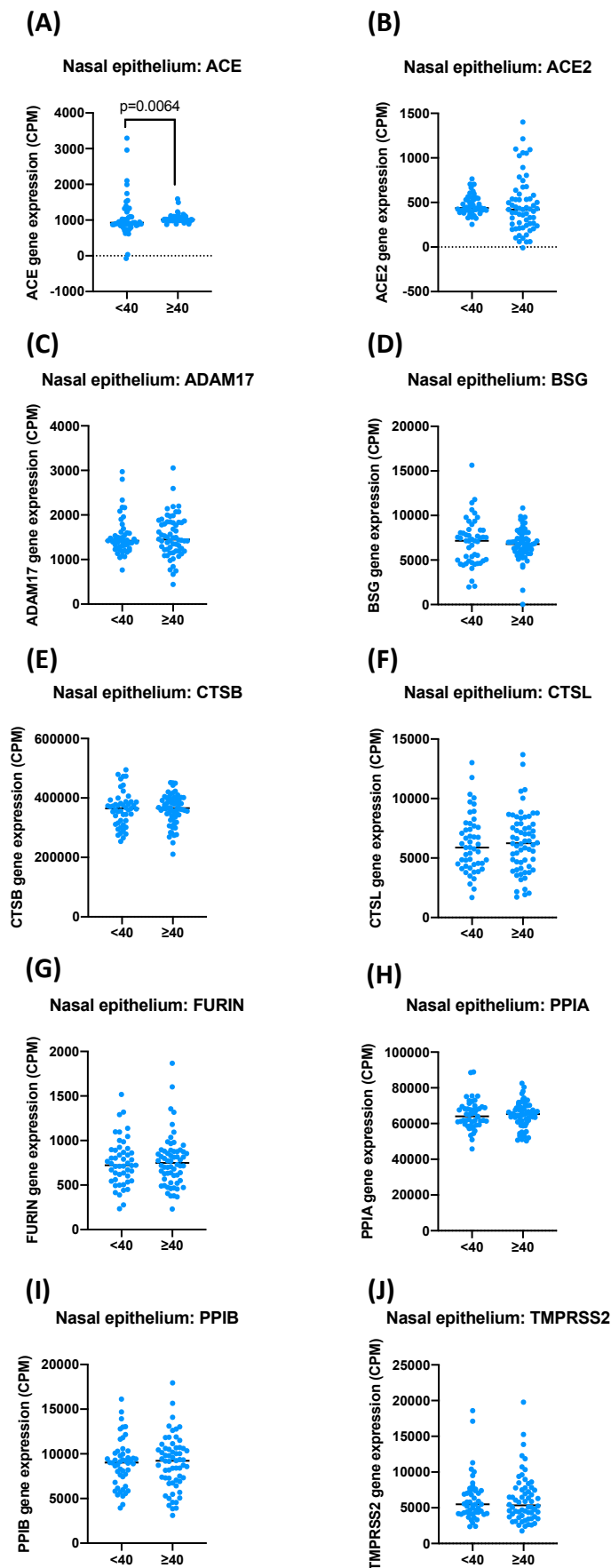
Supplementary Figure 6:



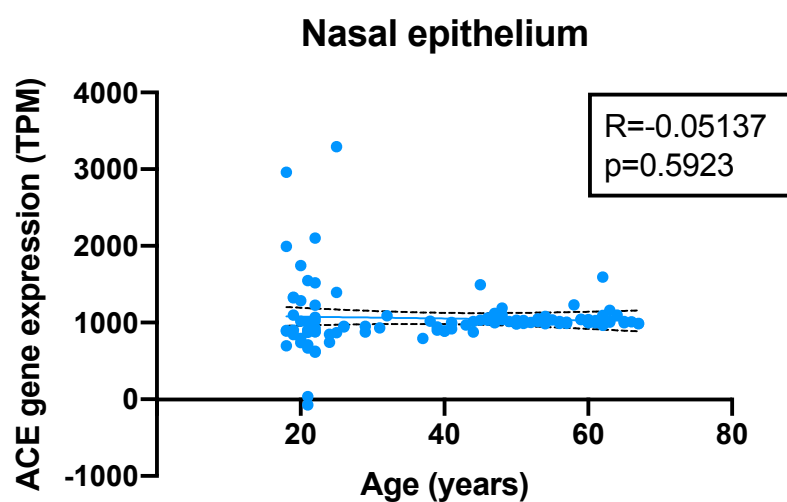
Supplementary Figure 7:



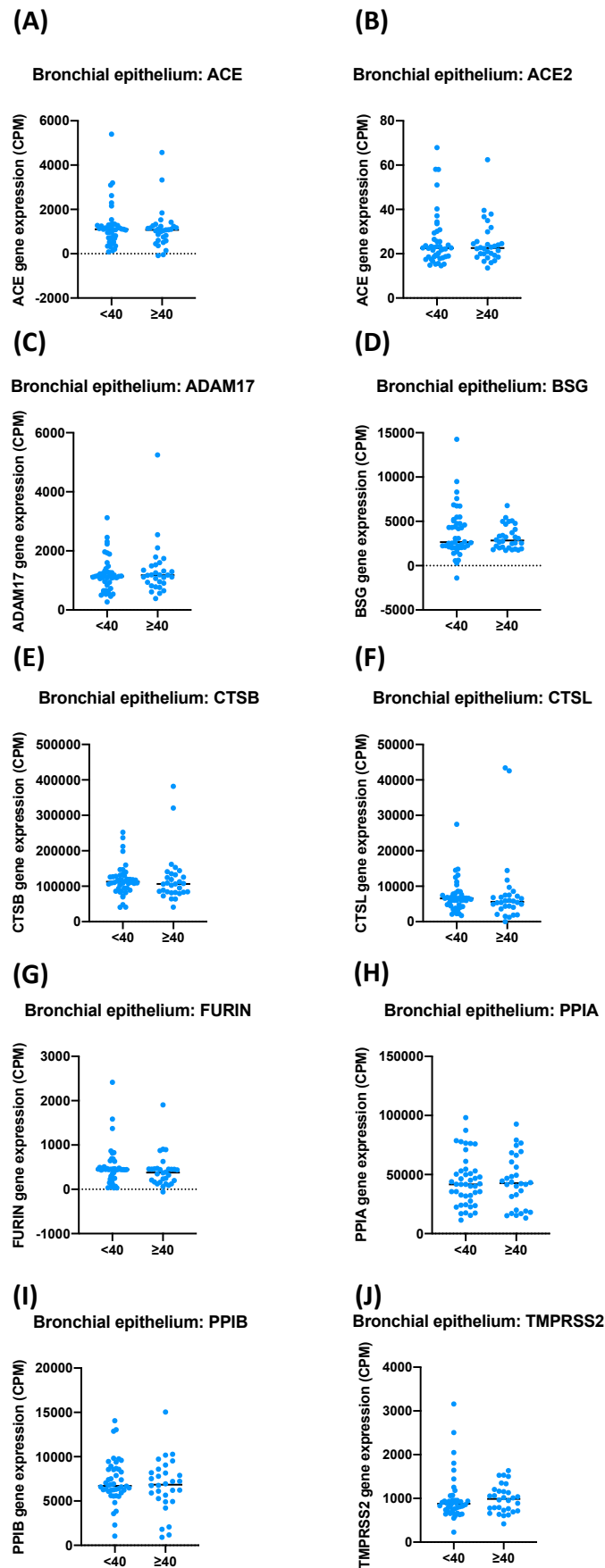
Supplementary Figure 8:



Supplementary Figure 9:



Supplementary Figure 10:



Supplementary Figure 12:

