ACE2 Expression is Increased in the Lungs of Patients with Comorbidities Associated with Severe COVID-19

Summary: A comprehensive transcriptome analysis of over 700 lung samples revealed that the expression of ACE2, which encodes for the entry receptor of SARS-CoV-2, is increased in the lungs of patients with some of the comorbidities associated with severe COVID-19.

Bruna G.G. Pinto[1#], Antonio E.R. Oliveira[1#], Youvika Singh[1], Leandro Jimenez[1], Andre N A. Gonçalves[1], Rodrigo L.T. Ogava[1], Rachel Creighton[2], Jean Pierre Schatzmann Peron[3,4], Helder I. Nakaya[1,4,*]

[1]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

[2]Department of Bioengineering, University of Washington, Seattle, WA, USA

[3]Neuroimmune Interactions Laboratory – Department of Immunology – Institute of Biomedical Sciences - University of Sao Paulo, Sao Paulo, Brazil. CEP 05508-000

[4]Scientific Platform Pasteur USP, São Paulo, Brazil

[*]Corresponding author

[#]These authors contribute equally to this work.

Accepted Manuscript

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com.

Abstract

Patients who died from COVID-19 often had comorbidities, such as hypertension, diabetes, and chronic obstructive lung disease. Although angiotensin-converting enzyme 2 (ACE2) is crucial for SARS-CoV2 to bind and enter host cells, no study has systematically assessed the ACE2 expression in the lungs of patients with these diseases. Here, we analyzed over 700 lung transcriptome samples of patients with comorbidities associated with severe COVID-19 and found that ACE2 was highly expressed in these patients, compared to control individuals. This finding suggests that patients with such comorbidities may have higher chances of developing severe COVID-19. Correlation and network analyses revealed many potential regulators of ACE2 in the human lung, including genes related to histone modifications, such as HAT1, HDAC2, and KDM5B. Our systems biology approach offers a possible explanation for increase of COVID-19 severity in patients with certain comorbidities.

Keywords: Angiotensin converting enzyme 2, COVID-19, SARS-CoV-2, KDM5B

Accepted Manuscript

Background

Recent studies of the epidemiological characteristics of COVID-19 have revealed that severe infection is more likely in people with an existing chronic medical condition. Two independent studies of infected populations in Wuhan, China found that approximately half the subjects infected with COVID-19 had an existing comorbidity (*1, 2*). In a study of 1099 patients across mainland China, 38.7% of patients with comorbidities progressed to severe infection (*3*). 2020), and in a study of 52 inpatients in Wuhan, 67% of patients with comorbidities died (*2*). The most common comorbidities reported in these studies were hypertension, diabetes, cerebrovascular disease, chronic obstructive lung disease, and coronary heart disease (*1-3*). Other comorbidities such as carcinoma, chronic kidney disease, chronic liver disease, digestive system disease, and nervous system disease have also been reported in patients with COVID-19 (*1, 2, 4*). A better understanding of the link between these conditions and COVID-19 infection is required to inform better treatment and prevention interventions.

The molecular mechanism responsible for the increased disease severity in patients with these comorbidities is not fully understood, but previous studies suggest a role for angiotensin-converting enzyme 2 (ACE2) (*5*). ACE2 is a membrane protein required for SARS-CoV2 to bind and enter cells (*6-8*). After binding, viral entry is facilitated by the activation of the viral spike glycoprotein and cleavage of the C-terminal segment of ACE2 by proteases like TMPRSS2 and FURIN that are readily expressed in lung tissue (*9-11*). ACE2 is only moderately expressed in healthy lung tissue compared to the heart, kidneys, and testes (*12*), but

Accepted Manuscript

staining of lung tissue sections from adults with pulmonary hypertension has revealed increased ACE2 protein in the endothelium of pulmonary arteries, compared to healthy controls (*13*). A comprehensive analysis of single-cell RNA-seq datasets revealed that ACE2 was co-expressed with TMPRSS2 within ileal absorptive enterocytes, nasal goblet secretory cells, and lung type II pneumocytes (*14*). ACE2 upregulation has also been observed in animal models of liver fibrosis (*15*). However, the reason for this upregulation remains unclear, and a link to other COVID-19 comorbidities has not been determined.

Here, we showed that the expression of the gene encoding the ACE2 receptor in lung tissue is upregulated by diseases representing comorbidities along with COVID-19. We also used systems biology approaches including co-expression analysis, meta-analysis, and network analysis to determine a potential cause of the ACE2 upregulation. From this analysis, we found that ACE2 expression could be regulated by enzymes that modify histones, including KDM5B. This identification of a common molecular mechanism of increased COVID-19 severity in patients with diverse comorbidities could direct the development of interventions to reduce the infection risk and disease severity in this population.

Accepted Manuscript

Methods

Literature curation

Relevant scientific literature related to key COVID-19 morbidities was retrieved from PubMed on March 16, 2020 using the query terms “pulmonary hypertension,” “chronic obstructive pulmonary disease,” “hypertension,” “smoking,” “pulmonary fibrosis,” and “asthma”. For terms returning more than 100,000 papers (“hypertension,” “smoking,” and “asthma”), only the most recent 100,000 papers were analyzed. Abstracts were annotated to identify all genes, diseases, and species appearing in the title or abstract using the PubTator Central API (*16*). This open source tool uses TaggerOne for disease annotations, GNormPlus for gene annotations, and SR4GN for species annotations (*16*). The data were filtered to retain only papers containing a human species annotation. Annotation of the abstract text identified 6 relevant disease MeSH terms: “Autoimmune Diseases,” “Cardiovascular Diseases,” “Familial Primary Pulmonary Hypertension,” “Hypertension,” “Hypertension, Pulmonary,” and “Renal Insufficiency, Chronic”, which were used in Figure 1. Next, every possible combination of gene and disease annotation within the title and abstract of each paper was generated. Only gene-disease associations supported by at least four documents, and those with a proximity less than or equal to the median sentence length of the paper section were retained.

Gene IDs were converted to gene symbols using the biomaRt R package (*17, 18*), and disease IDs were converted to disease MeSH terms using the Entrez Programming Utilities to query the Entrez database provided by the National Center for Biotechnology Information. The data was then further filtered to retain disease

Accepted Manuscript

MeSH terms relevant to reported clinical COVID-19 comorbidities (*3*). Redundant terms were collapsed using fuzzy string matching. The final gene-disease data set was used to generate a network utilizing Gephi software where the nodes were genes and diseases, and the edge weight was determined by the number of analyzed papers containing the gene-disease combination (*19*).

Meta-Analysis

We manually curated Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/) on March 16, 2020 to find lung transcriptome datasets related to “Pulmonary Arterial Hypertension” (PAH), “Chronic Obstructive Pulmonary Disease” (COPD), and “Smoking”. Author-normalized expression values and metadata from these datasets were downloaded using the GEOquery package (*20*). We performed differential expression analyses between patients with a disease and the control individuals (see Table S1) using the limma package (*21*). The gene symbol for each probe was obtained from the annotation file (*22*). Probes that matched the same gene symbol were collapsed by taking the one with the lowest p-value. Meta-analysis was performed with the MetaVolcanoR package (*23*) by combining the p-values using Fisher's method. To adjusting for multiple comparisons, we calculated the False Discovery Rate (FDR) to identify the differentially expressed genes (FDR < 0.05). For enrichment analyses, we utilized the EnrichR tool (*24*) with the “GO Biological Process 2018” and “BioPlanet 2019” databases. We then selected pathways with a P-value adjusted for multiple comparisons lower than 0.05. The network was created in Cytoscape (*25*).

Accepted Manuscript

Author-normalized FPKM expression values of *ACE2* gene in COPD patients and in subjects with normal spirometry were downloaded from GEO (accession ID: GSE57148). A single t-test was performed between COPD patients and controls (P = 0.000359).

Pearson correlation between the expression of ACE2 and all other genes in each of the seven lung transcriptome studies was performed. The p-values were then combined using Fisher's method, and an FDR correction was applied to adjust for multiple comparisons.

For the epigenetics analysis, we run the EnrichR tool (*24*) with the "ENCODE and ChEA Consensus TFs from ChIP-X" and "Epigenomics Roadmap" databases on the genes negatively or positively correlated with ACE2. Pathways with a P-value adjusted for multiple comparisons lower than 10^{-10} were selected. We utilized the Encode Roadmap browser (http://www.roadmapepigenomics.org/) from the Roadmap Epigenomics Project database (*26*) to identify the H3K27ac, H3K4m1, and H3K4m3 marks of histone acetylation and methylation with the corresponding p-values in the Lung of ENCODE donor STL002 (Roadmap alias E096). The figure showing the ACE2 locus and the marks was edited using CorelDraw software.

Accepted Manuscript

Results

To identify the genes highly associated with key comorbidities of severe COVID-19 (*1, 3*), we mined all relevant scientific literature of these human diseases. Specifically, over 8,000 abstracts were gathered from PubMed by querying titles and abstracts for the terms “pulmonary hypertension,” “chronic obstructive pulmonary disease,” “hypertension,” “smoking,” “pulmonary fibrosis,” or “asthma” (Figure 1a). Several relevant terms, such as “autoimmune diseases” and “cardiovascular diseases” were excluded from the PubMed query because of the breadth of literature published in these fields. Annotation of the abstract text identified 6 relevant disease MeSH terms: “Autoimmune Diseases,” “Cardiovascular Diseases,” “Familial Primary Pulmonary Hypertension,” “Hypertension,” “Hypertension, Pulmonary,” and “Renal Insufficiency, Chronic.” (*3*) (Figure 1a). Our text-mining analysis revealed 804 genes highly associated with one or more COVID-19 morbidities (Figure 1b). Among those genes, 26 were associated with four or more diseases (Figure 1c). Although ACE2 was known to be related to “Cardiovascular Diseases”, “Familial Primary Pulmonary Hypertension”, “Hypertension, Pulmonary”, and “Hypertension”, none of the articles containing this gene-disease association studied how ACE2 expression was altered in the lungs of patients with these diseases.

Based on the list of key comorbidities of severe COVID-19 (*1, 3*), we searched for lung transcriptome datasets available at GEO repository. We identified seven lung transcriptome studies of patients with either Chronic Obstructive Pulmonary Disease (COPD) or Pulmonary Arterial Hypertension (PAH), as well as smoking volunteers, compared to individuals who were non-smoking volunteers, were downloaded and used in our meta-analysis (Table S1). For each study, we performed differential expression analysis between patients and control individuals (Table S1). By

Accepted Manuscript

combining the p-values obtained in all the seven comparisons, we were able to identify 1,740 and 938 genes that were, respectively, up- and down-regulated in the disease (Figure 2a). Enrichment analysis using these differentially expressed genes revealed several pathways associated with inflammatory processes, metabolism, and ER stress. Among the pathways enriched with down-regulated genes, there were “vasculogenesis” and “regulation of Notch signaling pathway” (figure 2b). The “viral life cycle” pathway, which describes the processes utilized by viruses to ensure survival and to attach and enter the host cells was enriched with up-regulated genes (Figure 2b). ACE2 was included in this pathway, as well as 25 other genes (Figure 2c). One of these genes is *RAB1A*. Rab GTPases are involved in the replication of many viruses infecting humans (*27*), but have not been associated with SARS-CoV-2 life cycle yet. Both *TMPRSS2* gene, which is required for SARS-CoV-2 cell entry (*5*), and *FURIN* gene, which cleaves SARS-CoV-2 spike glycoprotein (*28*) were not differentially expressed in most of the lung transcriptome. However, both genes were highly expressed in lung (data not shown), suggesting that the levels of ACE2 may be the limiting factor for viral infection.

We then decided to investigate whether the gene encoding the ACE2 receptor was specifically up-regulated in the lungs of patients having one of these morbidities (Figure 3a). In a lung RNA-seq dataset (Table S1), we compared ACE2 expression between patients with COPD and subjects with normal spirometry (*30*). Again, the expression of ACE2 was significantly up-regulated in the disease compared to controls (Figure 3b). In fact, ACE2 was significantly up-regulated in 6 out of 7 lung transcriptome studies (Figure 3c), suggesting that patients who have COPD or PAH, and even people who smoke, may have higher chances of developing severe COVID-19.

Accepted Manuscript

Co-expression analyses can provide useful insights about the functional role of genes and their regulatory mechanisms (*31*). We performed Pearson correlation between the expression of ACE2 and all other genes in each of the seven lung transcriptome studies (Table S1), combined the p-values using Fisher's method, and applied an FDR correction (Figure 4a). This approach identified 544 and 173 genes with positive and negative correlation with ACE2, respectively (Figure 4a). Several of these genes were related to histone modifications, such as HAT1, HDAC2, KDM5B, among others (Figure 4a). Among the positively correlated genes, we found ADAM10 that regulates ACE2 cleavage in human airway epithelia (*32*) and TLR3 that plays a key role in the innate response to SARS-CoV or MERS-CoV infection (*33*).

Pathway enrichment analysis revealed that several of the genes positively associated with ACE2 were regulated by KDM5B, and by specific histone acetylation (H3K27ac) and histone methylation (H3K4me1 and H3K4me3) (Figure 4b). In fact, KMD5B demethylates lysine 4 of histone H3 (i.e. H3K4) and is involved in transcriptional regulation and DNA repair (*34*). We then checked in the Roadmap Epigenomics Project database (*26*) to see whether ACE2 locus contained ChIP-seq information for these histone markers. In the human lung, peaks for H3K4me1 and H3K4me3, as well as H3K27ac, were identified in ACE2 locus (Figure 4c), suggesting that ACE2 may be epigenetically regulated in the lung.

Accepted Manuscript

Discussion

We showed here that patients with co-morbidities which have very distinct mechanisms have increased expression of ACE2 in the lungs. Although our findings did not include COVID-19 infection data, we suggest that the higher expression of ACE2 in the lungs is associated with higher chances of developing the severe form of COVID19, by facilitating the SARS-CoV-2 entry into lung cells during the infection. In fact, COVID19 patients classified as severe cases displayed higher viral loads in nasopharyngeal swab samples during the early stages of disease onset compared to mild patients (*35*).

The current diabetes pandemic (*36*) could be worsening the SARS-CoV-2 pandemic by increasing the comorbidities associated with severe COVID-19. As we did not find lung transcriptome samples from patients with type 2 diabetes, we could not directly test whether ACE2 expression was increased in patients with diabetes, compared to healthy controls. However, our text-mining approach revealed that IL-6 and INS genes were associated with all the diseases we searched. The *INS* gene encodes the insulin hormone, and insulin is associated with the NAD-dependent histone deacetylase Sirtuin 1 (SIRT1) (*37*). We found that SIRT1 was up-regulated in the lung of patients with severe COVID-19 comorbidities in 4 out 7 studies (data not

Accepted Manuscript

shown). Clarke et al (*38*) have demonstrated that, under conditions of cell energy stress, SIRT1 can epigenetically regulate ACE2. Others too have shown that non-steroidal anti-inflammatory drugs may inhibit the SIRT1 deacetylase activity (*39*), which in turn could impact ACE2 expression.

The “viral life cycle” pathway that was enriched with up-regulated genes in patients with severe COVID-19 comorbidities contains several genes other than ACE2 that can be potentially important for SARS-CoV-2 cell cycle and invasion/attachment. These include *RAB1A* gene, whose product promotes the replication of Vaccinia virus (*40*). Also, RAB1A is important for Herpes Simplex Virus 1 Secondary Envelopment (*41*), and is required for assembly of classical swine fever virus particle (*42*). It is possible that SARS-CoV-2 utilizes RAB1A as well.

The fact that ACE2 gene is located in the X chromosome, and the initial findings showing that older males with comorbidities are more likely to be have severe COVID-19 compared to females (*1*), indicate that ACE2 expression in the lung may be sex-biased. Although no significant sexual differences was found in the activity of ACE2 in mouse lung (*43*), in rats, the levels of ACE2 were dramatically reduced with aging in both genders, but with significantly higher ACE2 expression in old female rats than male (*44*).

Although the mechanisms by which *ACE2* is up-regulated in patients with severe COVID-19 comorbidities were not addressed, our analysis may shed some light on the subject. Among the genes whose expression was positively correlated with ACE2, we detected genes associated with epigenetic regulation of gene transcription. For instance, HAT and HDAC modulate chromatin and DNA condensation by changing histone acetylation status, thus permitting gene

Accepted Manuscript

transcription. This could be happening in lung tissue, facilitating ACE2 expression, as observed during lung cancer and COPD.

KDM5B is associated with infection of hepatitis B virus (*45*). In breast cancer cells, blockage of KDM5 triggers a robust interferon response that results in resistance to infection by DNA and RNA viruses (*46*). This finding suggests that KDM5 demethylases are potential targets for preventing SARS-CoV-2 infection.

COVID-19 may kill between 5.6% and 15.2% of people infected with SARS-CoV-2 (*47*). Drug treatments that lower this mortality rate may save many thousands of lives. Our systems biology approach offers putative gene targets for treating and preventing severe COVID-19 cases.

Accepted Manuscript

References

1. N. Chen *et al.*, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507-513 (2020).
2. F. Zhou *et al.*, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*, (2020).
3. W. J. Guan *et al.*, Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*, (2020).
4. C. Huang *et al.*, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506 (2020).
5. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 10 (2020).
6. K. Kuba *et al.*, A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* **11**, 875-879 (2005).
7. R. Yan *et al.*, Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science*, (2020).
8. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
9. M. H. Vaarala, K. S. Porvari, S. Kellokumpu, A. P. Kyllonen, P. T. Vihko, Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. *J Pathol* **193**, 134-140 (2001).
10. A. Shulla *et al.*, A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* **85**, 873-882 (2011).
11. G. Simmons *et al.*, Different host cell proteases activate the SARS-coronavirus spike-protein for cell-cell and virus-cell fusion. *Virology* **413**, 265-274 (2011).
12. S. R. Tipnis *et al.*, A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J Biol Chem* **275**, 33238-33243 (2000).
13. C. Orte, J. M. Polak, S. G. Haworth, M. H. Yacoub, N. W. Morrell, Expression of pulmonary vascular angiotensin-converting enzyme in primary and secondary plexiform pulmonary hypertension. *J Pathol* **192**, 379-384 (2000).
14. C. G. K. Ziegler *et al.*, SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell*, (2020).

Accepted Manuscript

15. M. L. Huang *et al.*, Upregulation of angiotensin-converting enzyme (ACE) 2 in hepatic fibrosis by ACE inhibitors. *Clin Exp Pharmacol Physiol* **37**, e1-6 (2010).
16. C. H. Wei, A. Allot, R. Leaman, Z. Lu, PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res* **47**, W587-W593 (2019).
17. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
18. S. Durinck *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005).
19. M. Bastian, S. Heymann, M. Jacomy, paper presented at the International AAAI Conference on Weblogs and Social Media, 2009.
20. S. Davis, P. S. Meltzer, GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846-1847 (2007).
21. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
22. D. S. de Lima *et al.*, Long noncoding RNAs are involved in multiple immunological pathways in response to vaccination. *Proc Natl Acad Sci U S A* **116**, 17121-17126 (2019).
23. C. Prada, D. Lima, H. I. Nakaya. (Bioconductor, 2019).
24. E. Y. Chen *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
25. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
26. X. Zhou *et al.*, The Human Epigenome Browser at Washington University. *Nat Methods* **8**, 989-990 (2011).
27. P. Spearman, Viral interactions with host cell Rab GTPases. *Small GTPases* **9**, 192-201 (2018).
28. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, (2020).
29. J. S. Seo *et al.*, The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* **22**, 2109-2119 (2012).
30. W. J. Kim *et al.*, Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. *Int J Genomics* **2015**, 206937 (2015).
31. L. E. Cardozo *et al.*, webCEMiTool: Co-expression Modular Analysis Made Easy. *Front Genet* **10**, 146 (2019).
32. H. P. Jia *et al.*, Ectodomain shedding of angiotensin converting enzyme 2 in human airway epithelia. *Am J Physiol Lung Cell Mol Physiol* **297**, L84-96 (2009).
33. A. L. Totura *et al.*, Toll-Like Receptor 3 Signaling via TRIF Contributes to a Protective Innate Immune Response to Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio* **6**, e00638-00615 (2015).
34. S. Pippa *et al.*, Small Molecule Inhibitors of KDM5 Histone Demethylases Increase the Radiosensitivity of Breast Cancer Cells Overexpressing JARID1B. *Molecules* **24**, (2019).
35. Y. Liu *et al.*, Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect Dis*, (2020).
36. A. Toniolo *et al.*, The diabetes pandemic and associated infections: suggestions for clinical microbiology. *Rev Med Microbiol* **30**, 1-17 (2019).
37. F. Liang, S. Kume, D. Koya, SIRT1 and insulin resistance. *Nat Rev Endocrinol* **5**, 367-373 (2009).
38. N. E. Clarke, N. D. Belyaev, D. W. Lambert, A. J. Turner, Epigenetic regulation of angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress. *Clin Sci (Lond)* **126**, 507-516 (2014).
39. G. Dell'Omo *et al.*, Inhibition of SIRT1 deacetylase and p53 activation uncouples the anti-inflammatory and chemopreventive actions of NSAIDs. *Br J Cancer* **120**, 537-546 (2019).

Accepted Manuscript

40. T. Pechenick Jowers *et al.*, RAB1A promotes Vaccinia virus replication by facilitating the production of intracellular enveloped virions. *Virology* **475**, 66-73 (2015).
41. H. L. Zenner, S. Yoshimura, F. A. Barr, C. M. Crump, Analysis of Rab GTPase-activating proteins indicates that Rab1a/b and Rab43 are important for herpes simplex virus 1 secondary envelopment. *J Virol* **85**, 8012-8021 (2011).
42. J. Lin *et al.*, Rab1A is required for assembly of classical swine fever virus particle. *Virology* **514**, 18-29 (2018).
43. J. Liu *et al.*, Sex differences in renal angiotensin converting enzyme 2 (ACE2) activity are 17beta-oestradiol-dependent and sex chromosome-independent. *Biol Sex Differ* **1**, 6 (2010).
44. X. Xie, J. Chen, X. Wang, F. Zhang, Y. Liu, Age- and gender-related difference of ACE2 expression in rat lung. *Life Sci* **78**, 2166-2171 (2006).
45. X. Wang *et al.*, Hepatitis B virus X protein induces hepatic stem cell-like features in hepatocellular carcinoma by activating KDM5B. *World J Gastroenterol* **23**, 3252-3261 (2017).
46. L. Wu *et al.*, KDM5 histone demethylases repress immune response via suppression of STING. *PLoS Biol* **16**, e2006134 (2018).
47. D. Baud *et al.*, Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis*, (2020).

Accepted Manuscript

Figure legends

Figure 1. Literature curation of genes associated with key COVID-19 morbidities. a. Text-mining of 8,727 abstracts identified six relevant disease MeSH terms associated with a total of 804 genes. The number of genes associated with each disease MeSH term in at least four abstracts is shown in the pie chart. **b.** The knowledge-based network of COVID-19 morbidities. The network shows the diseases (red nodes) and genes (purple nodes) from panel a. The edges represent an association between a disease and a gene. The size of the nodes is proportional to its degree. **c.** Genes associated with four or more COVID-19 morbidities.

Figure 2. Meta-analysis of lung transcriptomes of patients with COVID-19 morbidities. a. Meta-analysis of seven differential expression analyses. Meta-volcano tool was used to combine the p-values of seven studies (Table S1) and to identify the differentially expressed genes (FDR < 0.01). **b.** Pathway enrichment analysis. Pathways from the "GO Biological Process 2018" database with Adjusted P-value < 0.05 were selected to create the network. The width of edges is proportional to the number of genes shared by two pathways (nodes). The size and color of nodes are proportional to the $-\log_{10}$ Adjusted P-value. **c.** Genes from the "viral life cycle" pathway that were up-regulated in human diseases. The colors in the heat map represent the $\log_2$ fold-change between patients and control individuals.

Accepted Manuscript

Figure 3. ACE2 is up-regulated in patients with lung diseases. a. Analysis of ACE2 expression in lung transcriptome datasets of patients with human pulmonary diseases. **b.** ACE2 expression in patients with COPD. The boxplot on the right shows the difference between COPD patients (red dots) and control individuals (blue dots). Student t-test P-value is indicated. **c.** ACE2 is up-regulated in patients with COVID-19 morbidities. Each bar represents the log2 expression fold-change between patients and control individuals. The error bars indicate the 95% confidence interval. Bars in red represent a p-value < 0.05 and in grey a non-significant p-value. The original studies are indicated and can be found in Table S1.

Figure 4. Insights of ACE2 regulation in the lung. a. Genes whose expression is correlated with ACE2 in the lung. Selected genes that were negatively (blue) or positively (red) correlated with ACE2 are highlighted. **b.** Pathway enrichment analysis using the ACE2-positively correlated genes. Pathways from the "ChIP-X Enrichment Analysis" and "Epigenomics Roadmap" databases with Adjusted P-value < 10^{-10} were selected. The size of the red circles is proportional to the - $\log_{10}$ Adjusted P-value of the enrichment. **c.** ACE2 locus contains marks of histone acetylation and methylation. The plot was modified from the WashU EpiGenome Browser using E096 lung. The peaks corresponding to each histone modification and the p-values of the marks are indicated.

Accepted Manuscript

Declaration of Interest

The author reports no conflicts of interest in this work.

Ethical Approval

Not applicable as we utilized publicly available data.

Acknowledgments

We would like to thank Tiago Lubiana for his valuable inputs.

Funding

HIN is supported by CNPq (313662/2017-7) and the São Paulo Research Foundation (FAPESP; grants 2017/50137-3, 2012/19278-6, 2018/14933-2, 2018/21934-5 and 2013/08216-2). RC was supported by an NSF Graduate Research Fellowship (DGE-1256082), NSF Graduate Research Opportunities Worldwide Fellowship, and by the STD & AIDS Research Training Fellowship Program (NIH 2T32AI007140-41).

Author contributions

BGGP, AERO, YS, LJ, ANAG, RLTO, RC, HIN performed the analyses. JPSP and HIN interpreted the results. All authors helped with the writing of the manuscript.

Accepted Manuscript

Corresponding author information:

Helder I Nakaya, PhD

Av. Prof. Lúcio Martins Rodrigues, 370, block C, 4th floor

São Paulo – SP - Brazil

CEP: 05508-020

Phone: + 55 11 2648-1130

e-mail: hnakaya@usp.br

Accepted Manuscript

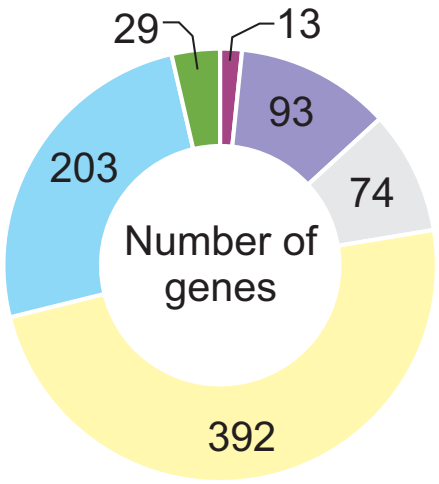
a

PubMed

Query: “pulmonary hypertension”, “chronic obstructive pulmonary disease”, “hypertension”, “smoking”, “pulmonary fibrosis”, or “asthma”

8,727 abstracts

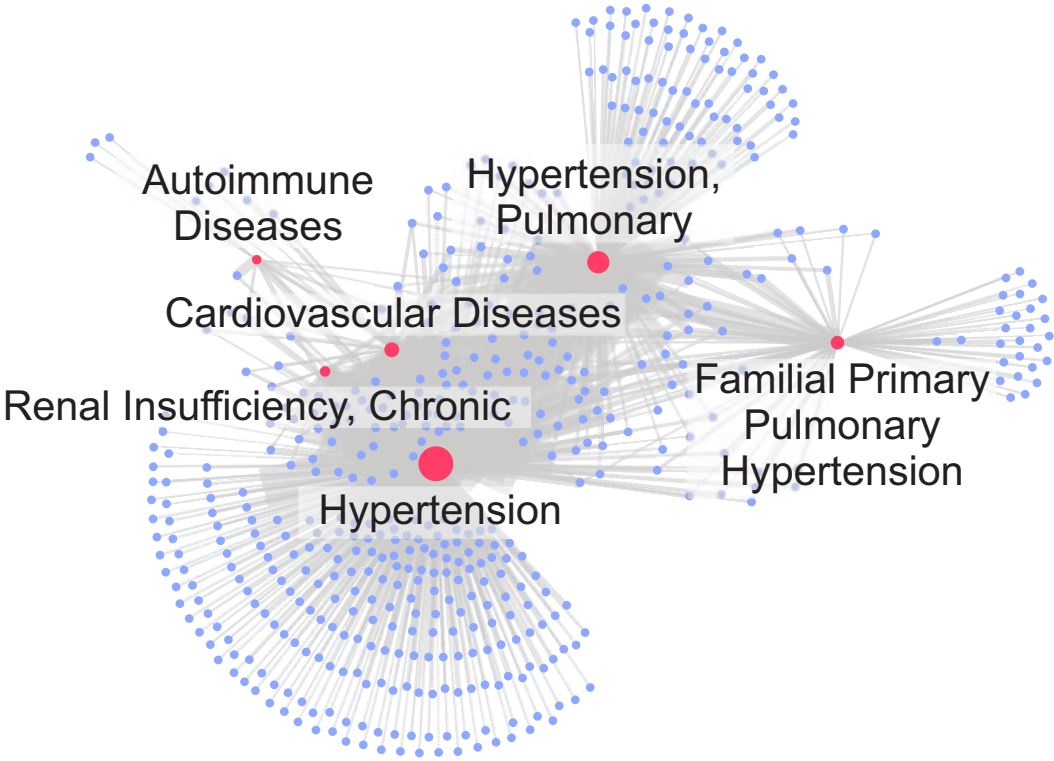
Text-mining analysis



Annotated Diseases:

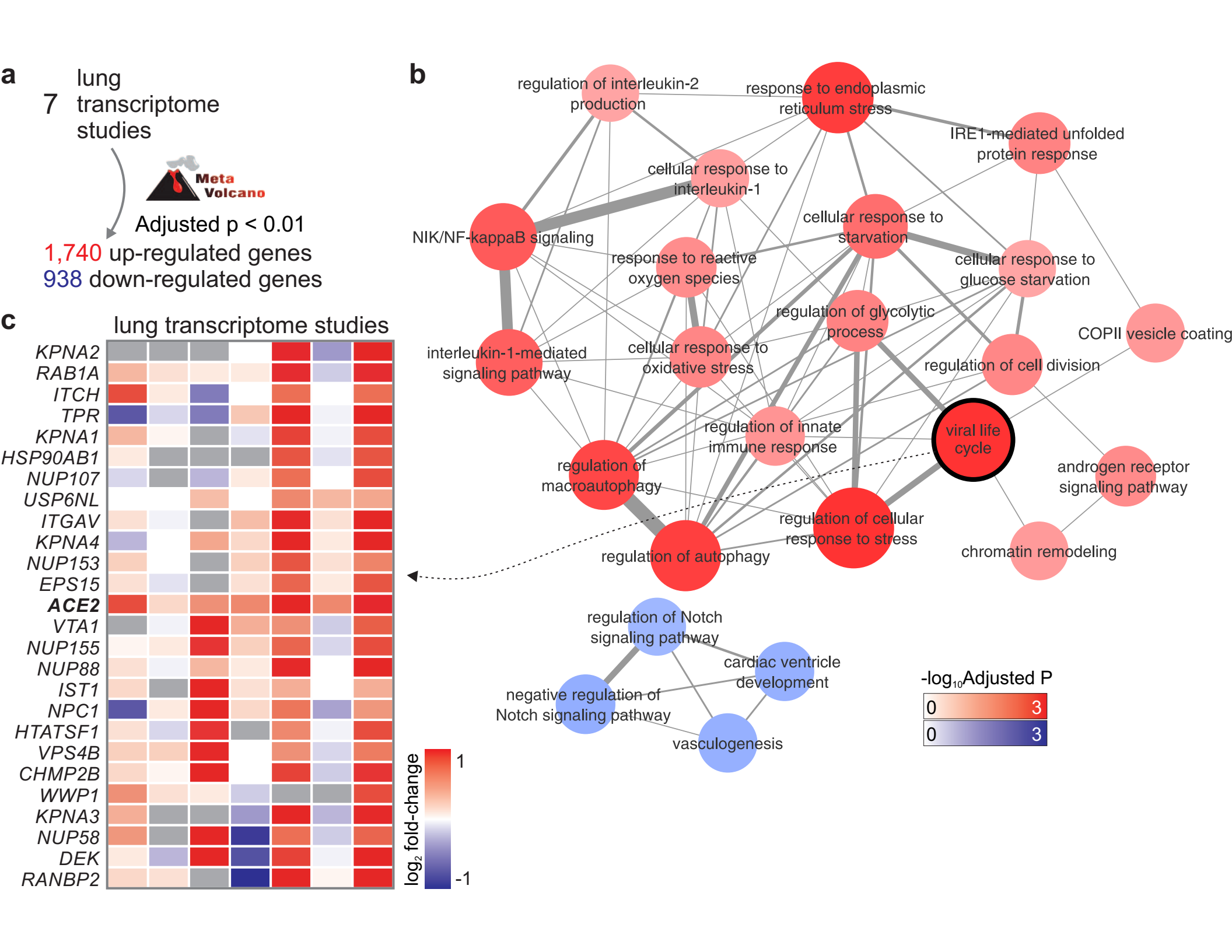
- Autoimmune Diseases
- Cardiovascular Diseases
- Familial Primary Pulmonary Hypertension
- Hypertension
- Hypertension, Pulmonary
- Renal Insufficiency, Chronic

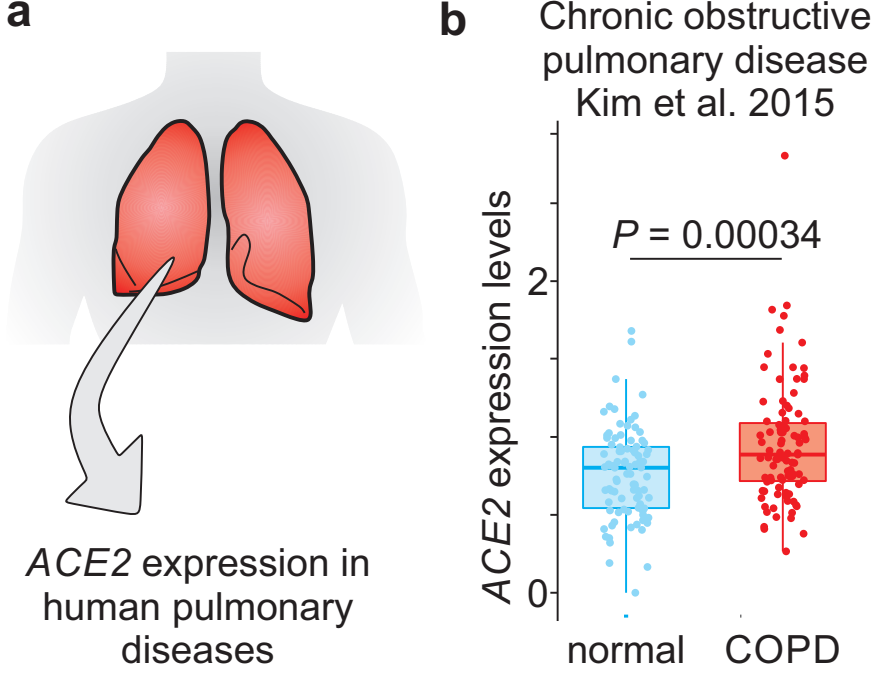
b

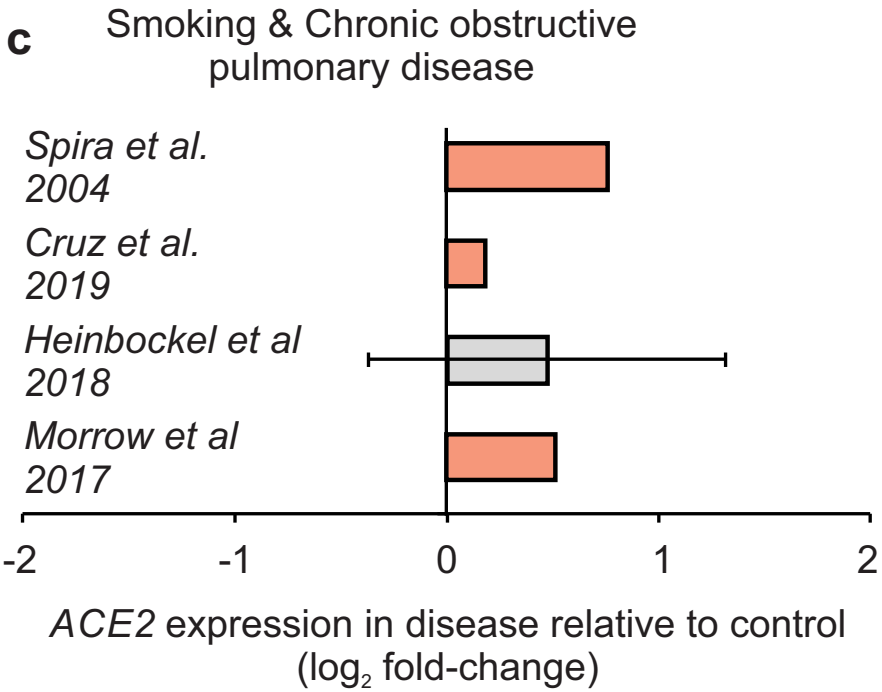


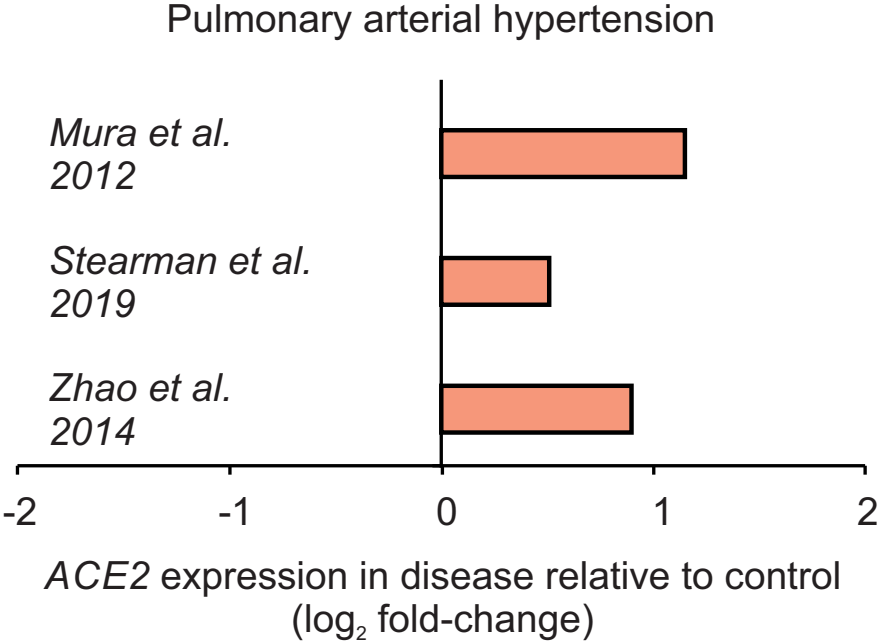
c

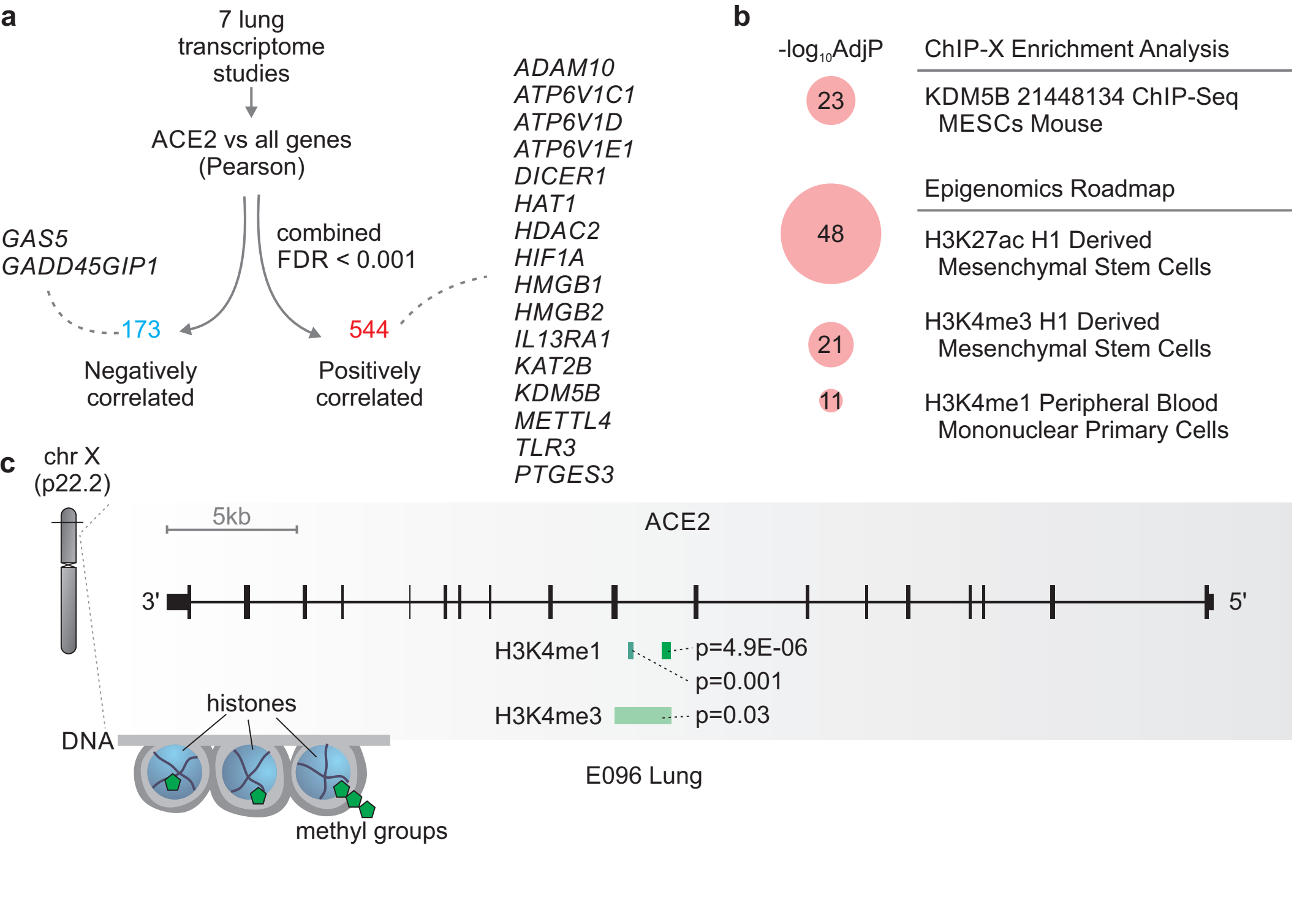
# diseases	
6	IL6; INS
5	REN; VEGFA
4	ACE2; AGT; CD59; COX8A; CRP; CST3; CD4; IFNA1; TGFB1; TNF; ACE; EDN1; EGFR; ENG; GDF15; LGALS3; MIF; MMP9; NPPB; PDE5A; PPARG; PTH


a
7 lung transcriptome studies
Meta Volcano
Adjusted p < 0.01
1,740 up-regulated genes
938 down-regulated genes
b
regulation of interleukin-2 production
response to endoplasmic reticulum stress
IRE1-mediated unfolded protein response
cellular response to interleukin-1
NIK/NF-kappaB signaling
cellular response to starvation
response to reactive oxygen species
cellular response to glucose starvation
regulation of glycolytic process
COPII vesicle coating
interleukin-1-mediated signaling pathway
cellular response to oxidative stress
regulation of cell division
regulation of innate immune response
viral life cycle
regulation of macroautophagy
androgen receptor signaling pathway
regulation of cellular response to stress
regulation of autophagy
chromatin remodeling
regulation of Notch signaling pathway
cardiac ventricle development
negative regulation of Notch signaling pathway
vasculogenesis
-log10Adjusted P
0
3
0
3
c
lung transcriptome studies
KPNA2
RAB1A
ITCH
TPR
KPNA1
HSP90AB1
NUP107
USP6NL
ITGAV
KPNA4
NUP153
EPS15
ACE2
VTA1
NUP155
NUP88
IST1
NPC1
HTATSF1
VPS4B
CHMP2B
WWP1
KPNA3
NUP58
DEK
RANBP2
log2 fold-change
1
-1


a
b
Chronic obstructive pulmonary disease
Kim et al. 2015
ACE2 expression levels
P = 0.00034
2
0
normal
COPD
ACE2 expression in human pulmonary diseases


c
Smoking & Chronic obstructive pulmonary disease
Spira et al. 2004
Cruz et al. 2019
Heinbockel et al 2018
Morrow et al 2017
-2
-1
0
1
2
ACE2 expression in disease relative to control (log2 fold-change)


Pulmonary arterial hypertension
Mura et al. 2012
Stearman et al. 2019
Zhao et al. 2014
-2
-1
0
1
2
ACE2 expression in disease relative to control (log2 fold-change)


a
7 lung transcriptome studies
ACE2 vs all genes (Pearson)
combined FDR < 0.001
GAS5
GADD45GIP1
173
Negatively correlated
544
Positively correlated
ADAM10
ATP6V1C1
ATP6V1D
ATP6V1E1
DICER1
HAT1
HDAC2
HIF1A
HMGB1
HMGB2
IL13RA1
KAT2B
KDM5B
METTL4
TLR3
PTGES3
b
-log10AdjP
23
48
21
11
ChIP-X Enrichment Analysis
KDM5B 21448134 ChIP-Seq MESCs Mouse
Epigenomics Roadmap
H3K27ac H1 Derived Mesenchymal Stem Cells
H3K4me3 H1 Derived Mesenchymal Stem Cells
H3K4me1 Peripheral Blood Mononuclear Primary Cells
c
chr X (p22.2)
5kb
ACE2
3'
5'
H3K4me1
p=4.9E-06
p=0.001
H3K4me3
p=0.03
histones
DNA
methyl groups
E096 Lung