Original Article

INFECTIOUS Microbes & Diseases

OPEN

Transcriptomics Curation of SARS-CoV-2 Related Host Genes in Mice With COVID-19 Comorbidity: A Pilot Study

Kunkai Su[1#], Xin Huang[2#], Kaijin Xu[1], Weibo Du[1], Danhua Zhu[1], Meifang Yang[1], Wenji Yuan[1], Lanjuan Li[1,✉]

Abstract

The pandemic of coronavirus disease 2019 (COVID-19), a respiratory disease caused by a novel severe acute respiratory syndrome coronavirus-2, is causing substantial morbidity and mortality. Along with the respiratory symptoms, underlying diseases in senior patients, such as diabetes, hypertension, and coronary heart disease, are the most common comorbidities, which cause more severe outcomes and even death. During cellular attachment and entry of severe acute respiratory syndrome coronavirus-2, the key protein involved is the angiotensin I converting enzyme 2 (ACE2), which is located on the membrane of host cells. Here, we aim to curate an expression profile of *Ace2* and other COVID-19 related genes across the available diabetes murine strains. Based on strictly manual curation and bioinformatics analysis of the publicly deposited expression datasets, *Ace2* and other potentially involved genes such as *Furin*, *Tmprss2*, *Ang*, and *Ang2* were examined. We found that *Ace2* expression is rather ubiquitous in three selected diabetes prone strains (db/db, ob/ob and diet-induced obese). With the most abundant datasets present, the liver shows a medium *Ace2* expression level compared with the lungs, pancreatic islets, brain and even T cells. Age is a more critical factor for *Ace2* expression in db/db compared with the other two strains. Besides *Ace2*, the other four host genes showed varied levels of correlation to each other. To accelerate research on the interaction between COVID-19 and underlying diseases, the Murine4Covid transcriptomics database (www.geneureka.org/Murine4Covid) will facilitate the design of research on COVID-19 and comorbidities.

Keywords: ACE2; COVID-19; diabetes; murine model; SARS-CoV-2

Introduction

Coronavirus disease 2019 (COVID-19) has resulted in >2,645,000 infections and >184,000 known deaths globally (up to April 23, 2020, https://coronavirus.jhu.edu/map.html).[1] The etiological agent of this pandemic is a new member of the severe acute respiratory syndrome (SARS) viruses. The novel SARS-coronavirus 2 (SARS-CoV-2) shares ~80% sequence identity at the amino acid level with previous SARS-CoV and Middle East respiratory syndrome coronavirus.[2,3]

The coronaviruses envelope is armed with a Spike protein, which recognizes and binds to the *ang*iotensin-converting enzyme 2 (ACE2) protein on the surface of mammalian cells.[4] Several other host cell surface proteins were computationally modeled, experimentally confirmed or deduced to play important roles in viral attachment, fusion and/or entry. Of these potential targets, FURIN, TMPRSS2, ANG, and ANG2 were most commonly reported recently.[5–8] However, the detailed mechanism of the interaction between coronavirus and the host is still not clear.

Similar to other viral respiratory infections, SARS-CoV-2 or COVID-19 mainly causes damage to the respiratory tract and develops severe pneumonia.[9] Elderly patients and those with underlying diseases are more at risk to develop progressive respiratory failure, which may lead to death.[10] According to recent analyses, besides respiratory diseases, hypertension, cardiovascular diseases, and diabetes were the most prevalent underlying diseases among hospitalized patients and patients dying of COVID-19.[11,12] Besides the study of viral infections, more resources have been placed to shed light on the interaction between underlying diseases and severity of COVID-19. Therefore, the demand for suitable murine models is accumulating. Unlike the infection mouse model, the models for studying underlying diseases and COVID-19 do not require incorporation of humanized ACE2 into the mouse. Therefore, this gives researchers an advantage to rely on currently built strains to evaluate corresponding characteristics. However, there are a lot of strains with different genetic backgrounds to target certain disease. For instance, there are nearly 200 genetically manipulated or diet mediated murine strains for studying diabetes (https://www.jax.org/mouse-search?searchTerm=diabetes). A better way

Editor: Stijn van der Veen

[1] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China,
[2] Biotherapeutics Research Center, the Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China.

✉ Corresponding author: Lanjuan Li, 6A-1501, The First Affiliated Hospital, School of Medicine, Zhejiang University, 310003, Hangzhou, China.
E-mail: ljli@zju.edu.cn

[#] KS and XH contributed equally to this study.

Author contributions: KS and XH conducted the bioinformatics analysis and wrote the manuscript. KX and WD reviewed the clinical comorbidity part. DZ, MY, and WY organized and curated the data. LL conceived and supervised the research.

Funding: This work was supported by the National Key Research and Development Program of China No. 2017YFC1200206 and 2019YFC0840609 (to KS), special fund of State Key Laboratory for Diagnosis and Treatment of Infectious Diseases No. SKLID2019KF05 (to XH) and the National Science and Technology Major Project of China No. 2017ZX10105001 (to KX) and 2017ZX10204401001002 (to MY).

Conflicts of Interest: The authors declare no conflicts of interest.

Infectious Microbes & Diseases (2020) 2:2

Received: 16 April 2020 / Received in final form: 23 April 2020 / Accepted: 26 April 2020

http://dx.doi.org/10.1097/IM9.0000000000000025

to select the most promising strains is of importance for the design of such studies.

Evaluation of expression levels of the known COVID-19 related host genes in the mice will benefit selection, given the limited experience currently available.[13–16] Accumulated data in public databases provides the possibility to do this. More and more researchers make their data online available for review and validation. Although their original design was not specifically targeted for this purpose, the data is still a grand treasure that contains beneficial information. Here we developed a pipeline to extract baseline information of *Ace2* and other COVID-19 related host genes to generate a comprehensive expression profile in murine tissues. The first deployment of this pipeline has been applied to three diabetes murine strains; B6.BKS(D)-*Lepr^{db}*/J (known as db/db), B6.Cg-*^{Lepob}*/J (ob/ob), and C57BL/6J diet-induced obese (DIO).

Results

Three most popular murine models for diabetes studies

All murine strains registered at the Jackson Laboratory were considered in the screening pipeline (Figure 1). A total of 50 strains were given by the JAX search engine using the keyword "Diabetes" and the constraint "Most popular." However, after our manual curation, we found that 37 strains with the word "diabetes" in the introduction are not specially maintained for diabetes studies. After removing these nonspecific strains, the 13 remaining strains were used to retrieve the Gene Expression Omnibus (GEO) database in The National Center for Biotechnology Information. Only stains with adequate datasets deposited were kept to ensure the possibility to cover most of the 11 selected tissues in next steps. For further analyses, we selected the top three strains: B6.BKS(D)-*Lepr^{db}*/J (db/db), B6.Cg-*^{Lepob}*/J (ob/ob), and C57BL/6J DIO (DIO). The expression profiles were manually curated and downloaded.[17–32]

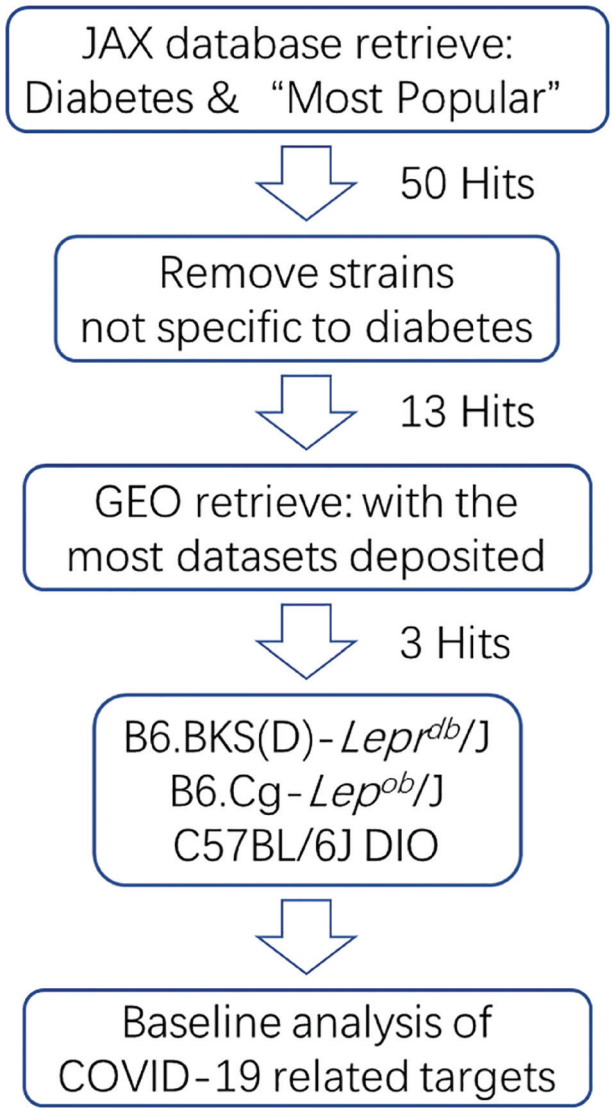

Figure 1. Pipeline to retrieve the most important diabetes murine strains.

Ace2 expression in diabetes murine tissues by array-based profiling

Distribution of *Ace2* in murine tissues is ubiquitous, ranging from the immediately targeted lung to the barrier isolated brain (Figure 2). To make the results more comprehensive, not only COVID-19 or diabetes related tissues were included. The fact that T cells exhibited the highest expression levels was a novel finding. However, this finding was supported by only one dataset. Db/db has been the most popular strain in previous studies, yet it lacks a qualified dataset for the lungs and other respiratory tissues. Obviously, most db/db related studies have focused on metabolic related tissues, such as the pancreas, liver, and adipose tissue. For the liver, which organ contains the most adequate deposited samples, its *Ace2* levels ranked in the medium range of all profiled tissues, which is consistent with other two mouse strains. For the ob/ob strain, *Ace2* levels were less variable compared with db/db and DIO, and the highest levels were found in the lungs. DIO exhibited the least number of datasets among the three strains. However, it provided a validation set for *Ace2* levels for the lung data obtained from the ob/ob mice, the brain data from the db/db mice and other tissues.

Ace2 expression changes according to age, tissue, and strains

As for diabetes, metabolic disorder is the most concerning process in humans. Therefore, we paid specific attention to metabolic related organs, such as the liver, pancreas, and muscle in this study. The dataset GEO Series (GSE) 43691, which contained the highest number of samples deposited in GEO, harbored 96 samples from the liver and muscle tissue under different kinds of treatment. Besides the baseline, the dataset GSE43691 offered us an opportunity to explore more details of *Ace2* expression. As shown in Figure 3, *Ace2* expression patterns varied extensively. The db/db strain exhibited a more consistent pattern in all designated groups, while DIO had the worst in-group performance. Expression levels in the liver and muscle tissue also showed different patterns depending on age, which might be taken into consideration in designing steps for future studies.

Correlation between Ace2 and other COVID-19 related host genes

Ace2 is not the only host gene reported to be related to infection in COVID-19. *Furin*, *Tmprss2*, *Ang*, and *Ang2* were also included in this study. Correlation of their expression levels

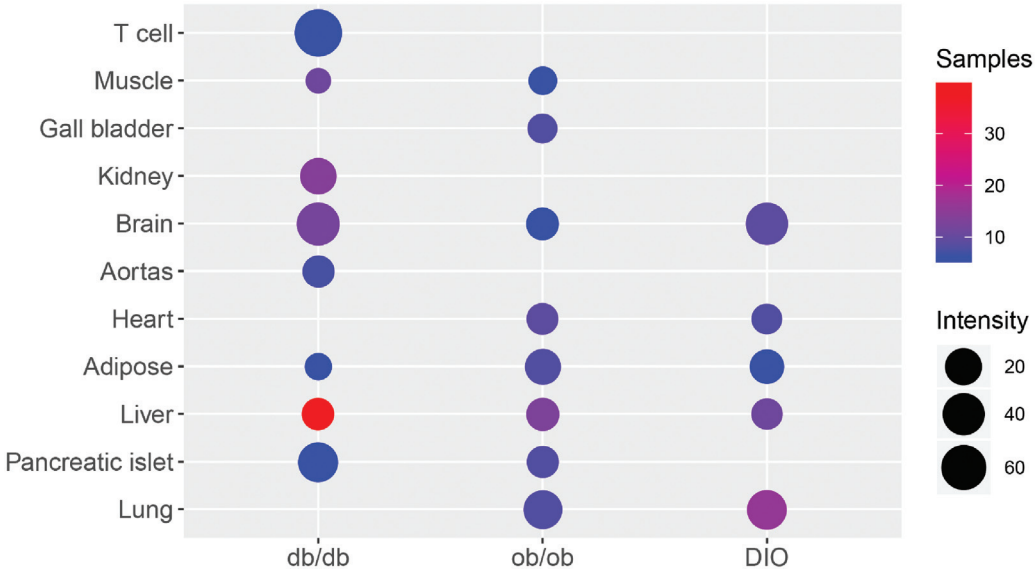

Figure 2. Summary of *Ace2* expression in mouse tissues based on publicly available transcriptomics datasets. Different sizes of circles represent normalized expression levels of *Ace2*, and scaled colors of circles represent number of samples curated in certain tissues of strains. A consistent expression in the lungs, pancreatic islets, liver, adipose tissue, heart, aortas, brain, kidney, gall bladder, muscle, and T cells was observed across all datasets. Blank in situ represents no qualified dataset available. Intensity is normalized and shown as /(10^3 intensity of geometric mean of *Gapdh* and *Actb*). Ace2: angiotensin I converting enzyme 2.

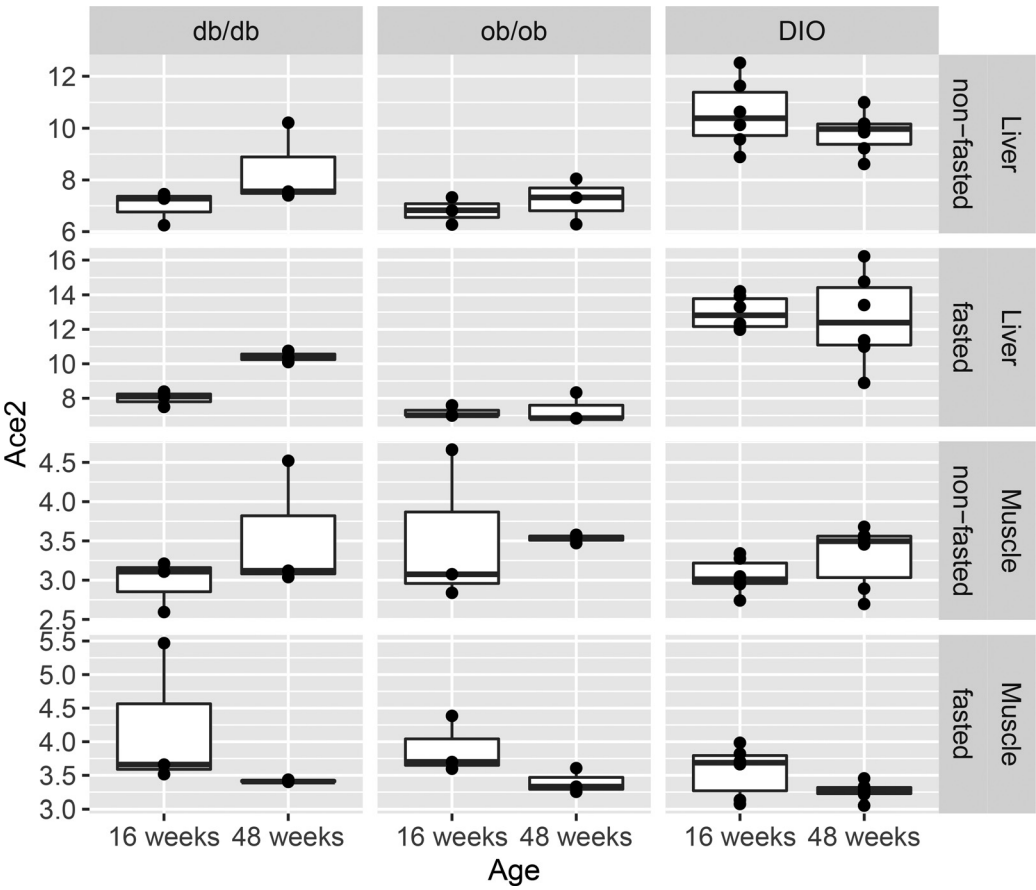

Figure 3. *Ace2* expression changes in age, strains, and organs. *Ace2* expression changes according to conditions of different strains, age, tissues, and diets. All data are from a single dataset. Ace2: angiotensin I converting enzyme 2.

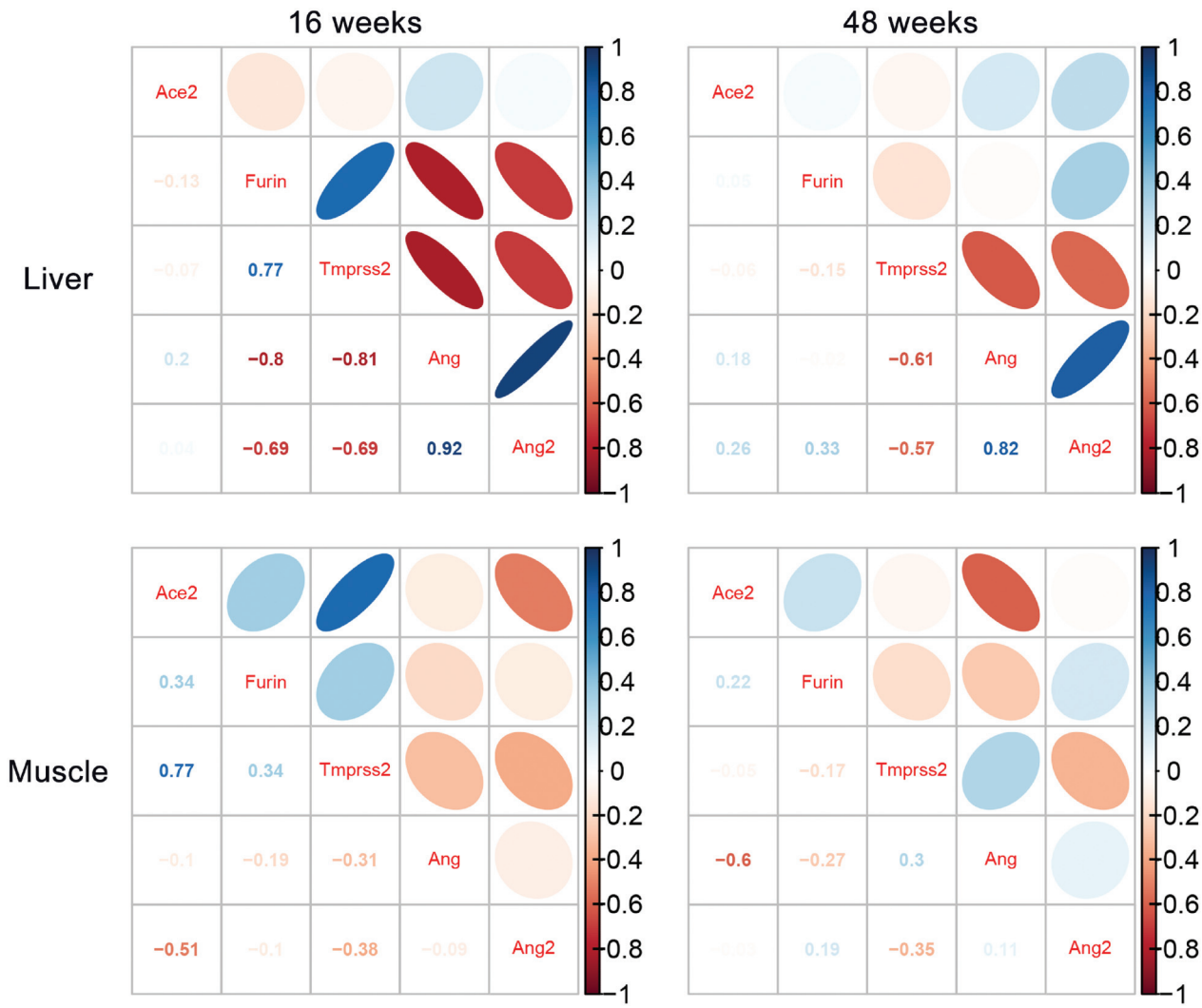

Figure 4. Correlation of five reported COVID-19 related targets. Correlation is displayed according to the diagonal order of *Ace2*, *Furin*, *Tmprss22*, *Ang*, and *Ang2*. The top half shows graphical demonstration and the bottom half shows the corresponding correlation coefficient. Scaled color is shown in the legend bar. Ace2: angiotensin I converting enzyme 2; COVID-19: coronavirus disease 2019.

is of particular relevance (Figure 4). In livers from younger (16 weeks) mice, *Ace2* expression was not found to be significantly correlated with any of these four genes. However, *Furin*, *Tmprss2*, *Ang*, and *Ang2* expression levels were highly correlated with each other and the correlation of the latter three genes was still observed in older (48 weeks) mice. The correlation between *Ang* and *Ang2* expression was the strongest (0.92 in the liver of 16-week-old mice and 0.82 the liver of 48-week-old mice). In contrast, all 5 selected genes did not show a strong correlation in muscle tissue.

Discussion

In this study, a comprehensive expression profile of COVID-19 related host genes (*Ace2*, *Furin*, *Tmprss2*, *Ang*, and *Ang2*) was established and analyzed in diabetes murine models as a pilot study. The baseline distribution of these genes across tissues and strains will provide important information to facilitate studies focusing on the interaction between COVID-19 and comorbidities, such as diabetes, hypertension, and coronary heart disease.

Previous studies verified the ubiquitous distribution of *Ace2* in different types of human tissues, which led to concerns on the potential tissue range of SARS-CoV-2.[7,15,33] Our analysis shows more variability in *Ace2* expression in the included mouse models and tissues, which will require further attention when translating mouse research to the human situation. The relationship between *Ace2* and age is still controversial. The phenomenon that adults are more vulnerable than children in COVID-19 might suggest that the abundance of *Ace2* expression increases with age.[34] However, results in the present study and other studies do not support this hypothesis,[33,35,36] which indicates that there might be additional key factors, other than *Ace2*, dominating the susceptibility or severity of COVID-19.

The present study employed the "quantile" method by functions in the *limma* package for in-dataset normalization, and the geometric mean of *Actb* and *Gapdh* for correction between datasets. Although *Actb* and *Gapdh* are the most popular reference genes used to compare expression levels of genes of interest in many studies, there are some evidences showing that they might not rank top as reference genes in certain tissues.[37,38] Therefore, to increase reliability, the geometric mean

was used for normalization of expression levels.[39] More potential reference gene candidates should be included in future studies to eliminate underlying biases.

This present pilot study presented COVID-19 related host gene profiles in murine models for diabetes. Hypertension, coronary heart disease, chronic obstructive pulmonary disease, smoking, and kidney diseases are also comorbidities of concern for COVID-19 in hospitalized patients. Therefore, our next step is to expand the analysis to these comorbidities and keep the data updated. We believe this kind of portal of information will provide the clinicians and researchers essential information to support their future studies.

Materials and methods

Murine strain selection

As the biggest provider of genetically defined mouse models for clinical research worldwide, the Jackson lab maintains an online search engine to facilitate the selection (https://www.jax.org/cn/search-intl). The keywords "Diabetes" and constraint "most popular" were used to retrieve the top strains related to diabetes research. Manual curation was performed to remove those with the word "diabetes" in the document but that were not specifically designed for that subject. Thirteen strains were retained to evaluate their abundance in publicly available expression data. The official names or alias of these 13 strains were used as keywords to search in the GEO database, and three strains (B6.BKS(D)-*Lepr^{db}*/J, B6.Cg-*^{Lepob}*/J, and C57BL/6J DIO) were selected based on the number of deposits in the GEO database.

Data cleaning and curation

The expression profiles were downloaded from the GEO database at the National Center for Biotechnology Information. To make the dataset more comprehensive, not only diabetes related organs or tissues were included in this study. According to the reported clinical comorbidity and latent concerns, a total of ten potentially targeted or affected organs were checked in this study. Combined search of official names or aliases and the target organs were used as keywords to find the qualified *GSE* gene expression profiles. Currently, only expression datasets profiled by microarray were included to make them more comparable. After manual curation, well documented profiles with at least three samples in the control groups (untreated or treated with empty vehicle) were kept in the scope.

Data normalization and analysis

The *GSE* gene expression profiles were downloaded by *GEOquery* package and normalized using the "quantile" method by functions in the *limma* package. After annotation by the *AnnoProbe* package, *Ace2* and other COVID-19 related genes were screened for further analysis. Only controlled groups were kept to examine the baseline of the murine strains. To compare different profiles, all the intensive values were normalized by their own geometric means of *Gapdh* and *Actb*. Bubble and correlation plots were generated by *ggplot2* and *Corrplot* package, respectively.

Data availability and updates

The detailed study results are available at www.geneureka.org/Murine4Covid. In addition, more specific target genes and murine strains will be added into the portfolio on request.

Acknowledgments

The authors thank all the contributors from GEO, Github, Bioconductor, and other communities for generously sharing their data and codes.

References

[1] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; doi:10.1016/S1473-3099(20)30120-1.
[2] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; doi:10.1016/S0140-6736(20)30183-5.
[3] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727–733.
[4] Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020; doi:10.1038/s41586-020-2179-y.
[5] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181(2):281–292.e6.
[6] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020; doi:10.1016/j.cell.2020.02.052.
[7] Lukassen S, Lorenz Chua R, Trefzer T, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* 2020; doi:10.15252/embj.20105114.
[8] Danser AHJ, Epstein M, Batlle D. Renin-angiotensin system blockers and the COVID-19 pandemic: at present there is no evidence to abandon renin-angiotensin system blockers. *Hypertension* 2020; doi:10.1161/HYPERTENSIONAHA.120.15082.
[9] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med* 2020;382(13):1199–1207.
[10] Wang L, He W, Yu X, et al. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect* 2020; doi: 10.1016/j.jinf.2020.03.019.
[11] Emami A, Javanmardi F, Pirbonyeh N, Akbari A. Prevalence of underlying diseases in hospitalized patients with COVID-19: a systematic review and meta-analysis. *Arch Acad Emerg Med* 2020;8(1):e35.
[12] Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int J Infect Dis* 2020; doi:10.1016/j.ijid.2020.03.017.
[13] Li R, Qiao S, Zhang G. Analysis of angiotensin-converting enzyme 2 (ACE2) from different species sheds some light on cross-species receptor usage of a novel coronavirus 2019-nCoV. *J Infect* 2020;80(4):469–496.
[14] 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* 2020; doi:10.1093/cvr/cvaa078.
[15] Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med* 2020; doi:10.1007/s11684-020-0754-0.
[16] Xu H, Zhong L, Deng J, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci* 2020; 12(1):8.
[17] Kozuka C, Shimizu-Okabe C, Takayama C, et al. Marked augmentation of PLGA nanoparticle-induced metabolically beneficial impact of gamma-oryzanol on fuel dyshomeostasis in genetically obese-diabetic ob/ob mice. *Drug Deliv* 2017;24(1):558–568.
[18] Schietinger A, Philip M, Krisnawan VE, et al. Tumor-specific T cell dysfunction is a dynamic antigen-driven differentiation program initiated early during tumorigenesis. *Immunity* 2016;45(2):389–401.
[19] Leone V, Gibbons SM, Martinez K, et al. Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. *Cell Host Microbe* 2015;17(5):681–689.
[20] Tilton SC, Karin NJ, Webb-Robertson BJ, et al. Impaired transcriptional response of the murine heart to cigarette smoke in the setting of high fat diet and obesity. *Chem Res Toxicol* 2013;26(7):1034–1042.
[21] Davis RC, van Nas A, Castellani LW, et al. Systems genetics of susceptibility to obesity-induced diabetes in mice. *Physiol Genomics* 2012;44(1):1–13.
[22] Yadav H, Quijano C, Kamaraju AK, et al. Protection from obesity and diabetes by blockade of TGF-beta/Smad3 signaling. *Cell Metab* 2011;14(1):67–79.

[23] Rame JE, Barouch LA, Sack MN, et al. Caloric restriction in leptin deficiency does not correct myocardial steatosis: failure to normalize PPAR{alpha}/PGC1{alpha} and thermogenic glycerolipid/fatty acid cycling. *Physiol Genomics* 2011;43(12):726–738.
[24] Yang JS, Kim JT, Jeon J, et al. Changes in hepatic gene expression upon oral administration of taurine-conjugated ursodeoxycholic acid in ob/ob mice. *PLoS One* 2010;5(11):e13858.
[25] Membrez M, Chou CJ, Raymond F, et al. Six weeks' sebacic acid supplementation improves fasting plasma glucose, HbA1c and glucose tolerance in db/db mice. *Diabetes Obes Metab* 2010;12(12):1120–1126.
[26] Hedbacker K, Birsoy K, Wysocki RW, et al. Antidiabetic effects of IGFBP2, a leptin-regulated gene. *Cell Metab* 2010;11(1):11–22.
[27] Wilson KD, Li Z, Wagner R, et al. Transcriptome alteration in the diabetic heart by rosiglitazone: implications for cardiovascular mortality. *PLoS One* 2008;3(7):e2609.
[28] Mzhavia N, Yu S, Ikeda S, Chu TT, Goldberg I, Dansky HM. Neuronatin: a new inflammation gene expressed on the aortic endothelium of diabetic mice. *Diabetes* 2008;57(10):2774–2783.
[29] Huang K, Rabold R, Abston E, et al. Effects of leptin deficiency on postnatal lung development in mice. *J Appl Physiol (1985)* 2008;105 (1):249–259.
[30] Rink C, Roy S, Khanna S, Rink T, Bagchi D, Sen CK. Transcriptome of the subcutaneous adipose tissue in response to oral supplementation of type 2 Leprdb obese diabetic mice with niacin-bound chromium. *Physiol Genomics* 2006;27(3):370–379.
[31] Schiekofer S, Galasso G, Sato K, Kraus BJ, Walsh K. Impaired revascularization in a mouse model of type 2 diabetes is associated with dysregulation of a complex angiogenic-regulatory network. *Arterioscler Thromb Vasc Biol* 2005;25(8):1603–1609.
[32] Lan H, Rabaglia ME, Stoehr JP, et al. Gene expression profiles of nondiabetic and diabetic obese mice suggest a role of hepatic lipogenic capacity in diabetes susceptibility. *Diabetes* 2003;52(3):688–700.
[33] Hikmet F, Méar L, Uhlén M, Lindskog C. The protein expression profile of ACE2 in human tissues. 2020.
[34] Li X, Xu S, Yu M, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol* 2020; doi: 10.1016/j.jaci.2020.04.006.
[35] Chow RD, Chen S. The aging transcriptome and cellular landscape of the human lung in relation to SARS-CoV-2. 2020.
[36] Booeshagh AS, Pachter L. Decrease in ACE2mRNA expression in aged mouse lung. 2020.
[37] Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3(7):RESEARCH0034.
[38] Dundas J, Ling M. Reference genes for measuring mRNA expression. *Theory Biosci* 2012;131(4):215–223.
[39] Chapman JR, Waldenstrom J. With reference to reference genes: a systematic review of endogenous controls in gene expression studies. *PLoS One* 2015;10(11):e0141853.