

UCSF

UC San Francisco Previously Published Works

Title

Systematic identification of ACE2 expression modulators reveals cardiomyopathy as a risk factor for mortality in COVID-19 patients

Permalink

<https://escholarship.org/uc/item/1xx7752m>

Journal

Genome Biology, 23(1)

ISSN

1474-760X

Authors

Kaur, Navchetan
Oskotsky, Boris
Butte, Atul J
[et al.](#)

Publication Date

2022-12-01

DOI

10.1186/s13059-021-02589-4


Peer reviewed

RESEARCH

Open Access



Systematic identification of ACE2 expression modulators reveals cardiomyopathy as a risk factor for mortality in COVID-19 patients

Navchetan Kaur^{1,2}, Boris Oskotsky^{1,2}, Atul J. Butte^{1,2} and Zicheng Hu^{1,3*} 

* Correspondence: zicheng.hu@ucsf.edu

¹Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA

³Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA

Full list of author information is available at the end of the article

Abstract

Background: Angiotensin-converting enzyme 2 (ACE2) is the cell-entry receptor for SARS-CoV-2. It plays critical roles in both the transmission and the pathogenesis of COVID-19. Comprehensive profiling of ACE2 expression patterns could reveal risk factors of severe COVID-19 illness. While the expression of ACE2 in healthy human tissues has been well characterized, it is not known which diseases and drugs might be associated with ACE2 expression.

Results: We develop GENEVA (GENE Expression Variance Analysis), a semi-automated framework for exploring massive amounts of RNA-seq datasets. We apply GENEVA to 286,650 publicly available RNA-seq samples to identify any previously studied experimental conditions that could be directly or indirectly associated with ACE2 expression. We identify multiple drugs, genetic perturbations, and diseases that are associated with the expression of ACE2, including cardiomyopathy, HNF1A overexpression, and drug treatments with RAD140 and itraconazole. Our joint analysis of seven datasets confirms ACE2 upregulation in all cardiomyopathy categories. Using electronic health records data from 3936 COVID-19 patients, we demonstrate that patients with pre-existing cardiomyopathy have an increased mortality risk than age-matched patients with other cardiovascular conditions. GENEVA is applicable to any genes of interest and is freely accessible at <http://genevatool.org>.

Conclusions: This study identifies multiple diseases and drugs that are associated with the expression of ACE2. The effect of these conditions should be carefully studied in COVID-19 patients. In particular, our analysis identifies cardiomyopathy patients as a high-risk group, with increased ACE2 expression in the heart and increased mortality after SARS-COV-2 infection.

Background

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2). The World Health Organization



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(WHO) declared the COVID-19 outbreak a pandemic on March 11, 2020. As of August 10, 2021, there have been 200 million recorded COVID-19 cases and over 4 million deaths [1].

Angiotensin-converting enzyme 2 (ACE2) is the cell-entry receptor for SARS-CoV-2 [2]. The binding between ACE2 and spike (S) protein of SARS-COV-2 initiates the viral entry into target cells. ACE2 plays key roles in both the transmission and pathogenesis of SARS-CoV-2, as demonstrated by the following lines of evidence: (1) SARS-CoV-2 fails to infect the lung-derived cell line A549 in the absence of ACE2 expression. The infection is restored after overexpressing ACE2 in the cell line [3]. (2) SARS-CoV-2 fails to infect wild-type mice but can infect and cause pneumonia in transgenic mice expressing human ACE2 [4, 5]. (3) COVID19-related tissue damages are detected in organs with ACE2 expression, including lungs, intestines, colons, and hearts [6–8]. (4) ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19, suggesting that the level of ACE2 expression is associated with disease severity [9]. Taking these lines of evidence together, it is crucial to comprehensively characterize the ACE2 expression in human tissues.

To comprehensively profile the expression patterns of ACE2, we not only need to characterize its expression in healthy tissues but also identify diseases, drugs, and genetic perturbations that are associated with ACE2 expression changes. The expression of ACE2 in healthy human tissues has been well characterized by resources such as the Human Cell Atlas and GTEx, with the highest expression detected in the intestine, testis, lung, cornea, heart, kidney, and adipose tissues [10, 11]. However, it is still not clear which diseases and drugs are associated with the ACE2 expression. Since ACE2 expression is tightly linked with the pathogenicity of SARS-COV-2, characterizing the expression pattern of ACE2 in different conditions will help us reveal and explain the risk factors of severe illness from COVID-19.

RNA-sequencing data profiles the full transcriptome of samples. Currently, more than 200,000 human RNA-seq samples are publicly available, providing an unprecedented opportunity for us to examine ACE2 expression in different human cell types under a variety of conditions and treatments. Data harmonization efforts such as ARCHS4 have uniformly preprocessed the RNA-seq data, making them readily available for analysis [12]. However, fully automated analysis of these datasets faces two main obstacles. First, the metadata are non-standardized and are often unstructured, making it difficult to extract experimental conditions from the studies. Second, experimental designs are highly variable. While some studies adopt the simple control-versus-treatment design, other studies are more complicated, involving multiple time points, combination treatments, or stratified cohorts. The heterogeneous design makes it difficult to analyze the datasets using a single statistical model.

Multiple tools have been made to analyze transcriptomics data, including CREEDS [13], scanGEO [14], GEM-TREND [15], StarGEO [16], SIGNATURE [17], SPIED [18], Cell Montage [19], ProfileChaser [20], ExpressionBlast [21], and SEEK [22]. However, the existing tools have several limitations, preventing them from fully exploring the publicly available RNA-sequencing resources. First, some of the tools annotate the metadata manually and are unable to cover the large number of datasets currently available. Second, the tools focus on differential expression analysis between two groups

(e.g., control versus treatment), preventing them from analyzing studies with more complex study designs.

In this study, we developed GENEVA (GENe Expression Variance Analysis), a semi-automated framework for exploring public RNA-seq datasets. For a given gene, GENEVA identifies the most relevant datasets by analyzing the variance of the gene expression. GENEVA visualizes the relevant datasets for detailed manual analysis. GENEVA is scalable and is agnostic to study designs. Using GENEVA, we identified multiple drugs, genetic perturbations, and diseases that modulate the expression of ACE2, including cardiomyopathy, HNF1A overexpression, and drug treatments with RAD140 and itraconazole. Our in-depth meta-analysis of seven datasets reveals increased ACE2 expression in all cardiomyopathy categories. By analyzing the clinical data of 3936 COVID19 patients at UCSF hospital, we demonstrate that patients with pre-existing cardiomyopathy have an increased mortality risk than other patients, including propensity score-matched patients with other cardiovascular conditions.

Results

Analysis of 286,650 RNA-seq samples reveals complex transcriptional networks of ACE2.

Our study leverages human RNA-sequencing data from the ARCHS4 project, containing 286,650 uniformly preprocessed data from 9124 Gene Expression Omnibus (GEO) series [12]. The large number of RNA-sequencing samples provides an unprecedented resource for studying the expression of ACE2 in different human cell types under a variety of conditions and treatments.

We first characterized the transcriptional networks of ACE2 using all 286,650 samples. We calculated the Pearson correlation between ACE2 and all other human genes. Because of the large sample size, most of the correlations are statistically significant, even after multiple testing adjustments. Therefore, we focused on the correlation coefficients themselves as a measure of effect size, rather than the p values or significance. While most of the genes have correlation coefficients near 0, a small set of genes are highly correlated with ACE2, with the highest correlation to be 0.72 between FABP2 and ACE2 (Additional file 1, Fig. S1A and B). The top correlated genes include FABP2, MEP1B, and transcription factors such as HNF4G (Top 30 genes shown in Additional file 1, Fig. S1B-C, and all correlations listed in Additional file 2 Table S1 and Additional file 3 Table S2). The top correlated pathways include multiple pathways related to the digestive process (Top 30 pathways shown in Additional file 1, Fig. S1D, and all pathways listed in Additional file 4, Table S3), consistent with the high expression of ACE2 in small intestines.

Dataset from the GEO database covers a variety of tissue types. To see how the tissue differences affect the correlation between ACE2 and other genes, we examined tissue-specific transcriptome data from the GTEx consortium. We found similar co-expression relationship between ACE2 and other genes using data from GTEx or GEO (correlation = 0.42, p value < 0.0001, Additional file 5, Table S4), suggesting that the co-expression profile between ACE2 and other genes are heavily influenced by the tissue differences. To avoid our analysis being dominated by tissue differences, we evaluated the correlations between ACE2 and other human genes within each RNA-seq dataset. We then calculated the mean and variances of the correlations across all

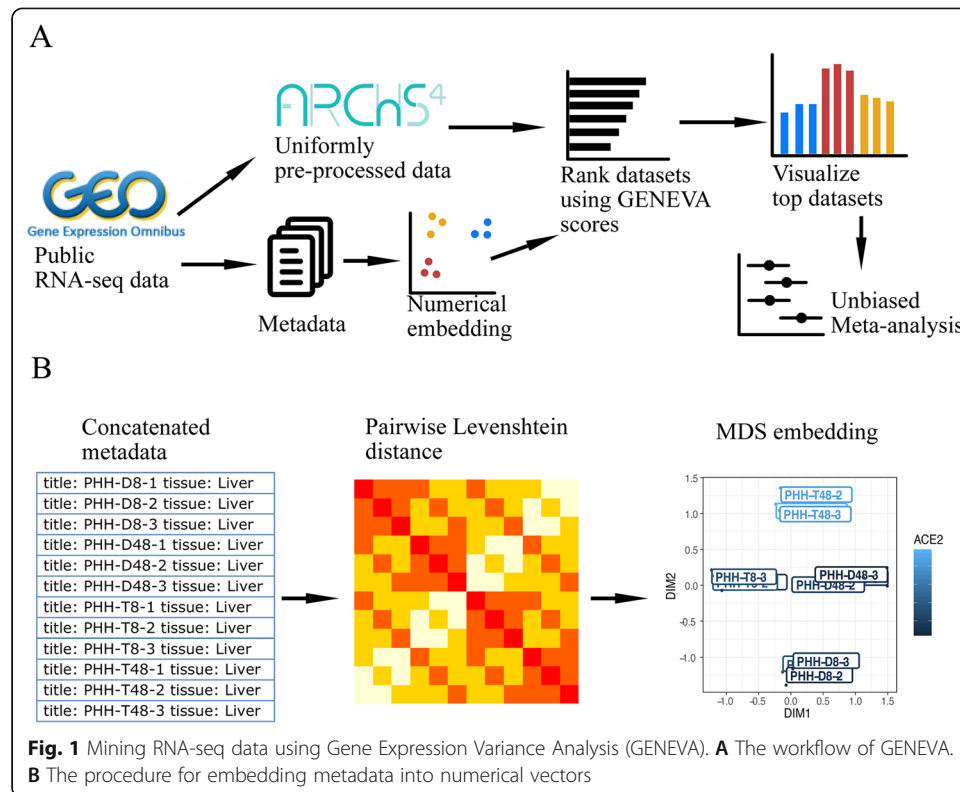
datasets. Our analysis reveals a positive relationship between the mean and variance of the correlation coefficients (Additional file 1, Fig. S1E). While some genes have a high average correlation with ACE2, their correlation with ACE2 is highly variable in individual datasets. The results suggest against a common transcriptional network around the ACE2 gene. Rather, ACE2 is co-regulated with different sets of genes under different conditions.

We adopted a mixed-effect model to estimate the overall association between ACE2 and other genes across studies, allowing us to prioritize genes with relatively conserved correlations with ACE2 across all studies. The top genes include MYO7B, CALML4, and transcription factors such as HNF1A (Top 30 genes shown in Additional file 1, Fig. S1 F-G, and all genes listed in Additional file 6, Table S5 and Additional file 7, Table S6). Pathway analysis identified many metabolic processes to be correlated with ACE2 expression. The findings are consistent with previous observations that ACE2 is involved in glucose metabolism and energy stress responses (Top 30 pathways shown in Additional file 1 Fig. S1 H, and all pathways listed in Additional file 8, Table S7) [23, 24].

Our analysis identified three transcription factors in the hepatocyte nuclear factor family, including HNF4G, HNF1A, and HNF4A (Additional file 1, Fig. S1 C and G). HNF4G has the highest overall correlation with ACE2 while HNF1A has the highest standardized correlation coefficient within studies. We then tested the causal relationship between the transcription factors and ACE2 expression. We identified two RNA-seq datasets that compared human cells with or without genetic perturbation of HNF4G and HNF1A. While HNF4G is positively correlated with ACE2 expression (Additional file 1, Fig. S1 C), overexpression of HNF4G does not lead to significantly increased ACE2 expression (Additional file 1, Fig. S1I). Rather, there is a trend of reduction in ACE2 expression. HNF1A overexpression leads to increased ACE2 expression and HNF1A knockdown reduced ACE2 expression (Additional file 1, Fig. S1J) in LNCaP cells, a prostate cancer cell line. The result is consistent with the positive correlation between HNF1A and ACE2 (Additional file 1, Fig. S1G). A previous study showed that HNF1A induces ACE2 in pancreatic islets [25]. Our result in a prostate cancer cell line further confirmed the role of HNF1A in regulating ACE2. However, it should be noted that HNF1A and ACE2 are not correlated in all RNA-seq datasets (Additional file 1, Fig. S1K), demonstrating the complexity of ACE2 regulation in different tissues.

Gene expression variance analysis reveals diseases and therapeutics that modulate the expression of ACE2

Next, we hope to identify conditions that are associated with the expression of ACE2. We developed a computational framework named GENEVA (Gene Expression Variance Analysis) to identify the most relevant datasets for visualization and detailed manual analysis (Fig. 1A and “Methods”). GENEVA prioritizes the datasets that have a large variance of ACE2 expression. The rationale is that datasets with large ACE2 variance are likely to contain conditions that modulate the ACE2 expression. At the same time, GENEVA controls for the overall heterogeneity of the samples to prioritize datasets in which ACE2 is specifically modulated by experimental conditions rather than due to



tissue type differences. In addition, GENEVA embeds the metadata into numerical space and prioritizes datasets with high correlations between ACE2 expression and the metadata (Fig. 1B). This allows GENEVA to identify datasets in which ACE2 is regulated by experimental conditions rather than randomness or unexplained factors. While our study focuses on ACE2 and its role in COVID-19 disease, GENEVA is applicable to all genes. We created a web application that allows researchers to apply GENEVA to their gene of interest [<http://genevatool.org>].

We tested the significance of the GENEVA scores using a permutation procedure. We randomly shuffle the samples across studies to generate a null distribution. We compared each GENEVA score to the null distribution to calculate the p value. We adjusted for multiple testing using the false discovery rate (FDR) method [26]. We identified 27 significant datasets with FDRs less than 0.05 (Table 1). Interestingly, GENEVA identified HNF1A as an ACE2 modulator, which was also identified in our correlation analysis (Additional file 1, Fig. S1). GENEVA additionally identified multiple drugs and diseases that modulate or are associated with the ACE2 expression, revealing potential risk factors for severe illness from COVID-19.

Here, we highlight three ACE2-modulating conditions, manually picked based on their effect on ACE2 expression and their potential impact on public health. Data from GSE89714 show upregulated expression of ACE2 in hypertrophic cardiomyopathy (Fig. 2A, B). Hypertrophic cardiomyopathy is the most common inherited heart disease, affecting an estimated 15,188,000 individuals (0.2%) worldwide [27]. Our finding is consistent with an increased death rate in COVID-19 patients with heart conditions [28–30] and suggests that higher ACE2 expression can contribute to the increased risk. Data from GSE104177 showed that RAD140, a selective androgen receptor modulator,

Table 1 RNA-seq datasets with GENEVA scores of ACE2 expression that are statistically significant

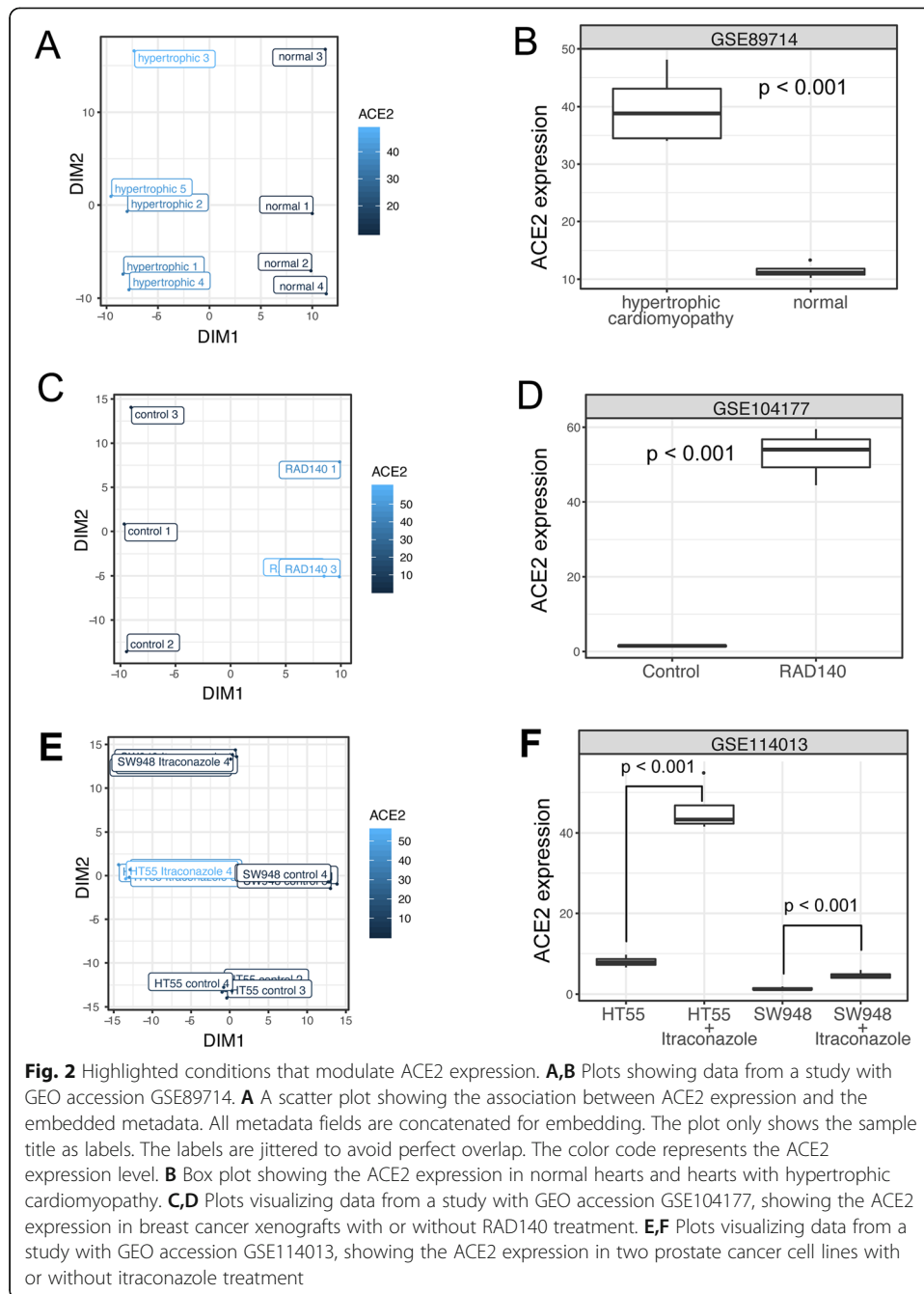
Rank	GSE	Tissue	ACE2 modulating conditions	GENEVA score	FDR
1	GSE46490	ovarian follicle	cell differentiation	87.9	<0.01
2	GSE104177	breast cancer cell line	RAD140 treatment	56.9	<0.01
3	GSE102910	hepatoma cell line	HCV replicon	56.1	<0.01
4	GSE131322	cardiomyocytes	cardiomyopathy	54.5	<0.01
5	GSE89714	heart	cardiomyopathy	29.5	<0.01
6	GSE128967	gastric cancer	FOLFOX treatment	23.1	<0.01
7	GSE69687	Colorectal cancer	PORCN inhibition	21.6	<0.01
8	GSE82207	colonic organoids	cell differentiation	20.1	<0.01
9	GSE134520	stomach	Gastric Metaplasia	18.1	0.013
10	GSE77250	Colonic organoids	cell differentiation	16.1	0.013
11	GSE57401	reninoma tumor	Unexplained	15.8	0.013
12	GSE130521	cardiomyocytes	dimethylxalylglycine	15.2	0.013
13	GSE54066	ES cell	cell differentiation	14.5	0.013
14	GSE116860	kidney	cell differentiation	12.8	0.017
15	GSE133161	human oocyte	Unexplained	12.8	0.017
16	GSE95477	oocyte	cell differentiation	12.4	0.017
17	GSE114013	colorectal tumor	Itraconazole treatment	12.3	0.017
18	GSE104178	colorectal tumor	cancer	12.3	0.017
19	GSE106338	gastric tumor	cancer	11.8	0.017
20	GSE71613	heart	cardiomyopathy	11.8	0.017
21	GSE80463	intestine	Unexplained	11.1	0.017
22	GSE101531	intestine	Unexplained	11.1	0.017
23	GSE114003	hepatocytes	tazarotene treatment	10.2	0.023
24	GSE118548	colon carcinoma cell line	trametinib treatment	10.1	0.023
25	GSE85556	prostate cancer cell line	HNF1A expression	9.6	0.029
26	GSE99321	heart	cardiomyopathy	9.4	0.029
27	GSE142077	kidney	kidney injury	8.9	0.048

induces ACE2 expression in human breast cancer xenografts (Fig. 2C, D). Data from GSE114013 show that itraconazole, an antifungal drug, upregulates ACE2 expression in two colorectal cancer cell lines, HT55 and SW948 (Fig. 2 E, F). These findings suggest that these drugs should be studied with respect to ACE2 expression in lung and cardiac cells and tissues and that patients on these drugs should be studied closely during the pandemic. If these subsequent studies do continue to suggest this effect on increasing ACE2 expression, heightened caution could be warranted when using these drugs during the COVID-19 pandemic.

Joint analysis shows ACE2 upregulation in all types of cardiomyopathy

GENEVA prioritizes datasets with large variances in ACE2 expression. However, the procedure may introduce bias, as studies with small ACE2 variations are ignored. Consider an example in which multiple studies have profiled the effect of a drug. Some studies show that the drug upregulates ACE2 while other studies show that the drug has no effect on ACE2. The effect of the drug will be overestimated if a researcher only includes the studies with positive results. Therefore, after the GENEVA analysis, a joint analysis of all related datasets is required to confirm the findings.

We performed a comprehensive search for datasets related to the three highlighted conditions, including cardiomyopathy, itraconazole treatment, and RAD140 treatment.



We did not find additional datasets related to itraconazole and RAD140 treatment. For cardiomyopathy, we identified a total of 7 datasets. We performed a meta-analysis using a mixed-effect model (with cardiomyopathy as the fixed effect and the dataset as the random effect), taking data from all 7 datasets into account. The result confirmed that ACE2 expression is significantly elevated in heart tissue samples from cardiomyopathy patients (p value < 0.001).

We next examined the ACE2 expression in different types of cardiomyopathy. The most common types of cardiomyopathies include dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM),

arrhythmogenic right ventricular cardiomyopathy (ARVC), and left ventricular non-compaction (LVNC) [31–35]. These cardiomyopathy types have different causes and show distinctive morphology and physiology characteristics. Although previous studies have demonstrated ACE2 upregulation in DCM and HCM [36, 37], how ACE2 is regulated in other types is unknown. Within the 7 datasets, we were able to identify all the common cardiomyopathy types. Our analysis revealed significantly increased ACE2 expression in most of the cardiomyopathy types (Fig. 3), including DCM, HCM, RCM, and LVNC. Although the result of ARVC is not statistically significant, the data show a clear trend of ACE2 upregulation (Fig. 3D). We performed Egger regression and did not observe significant publication bias in the cardiomyopathy datasets [38] (Additional file 1, Figure S2).

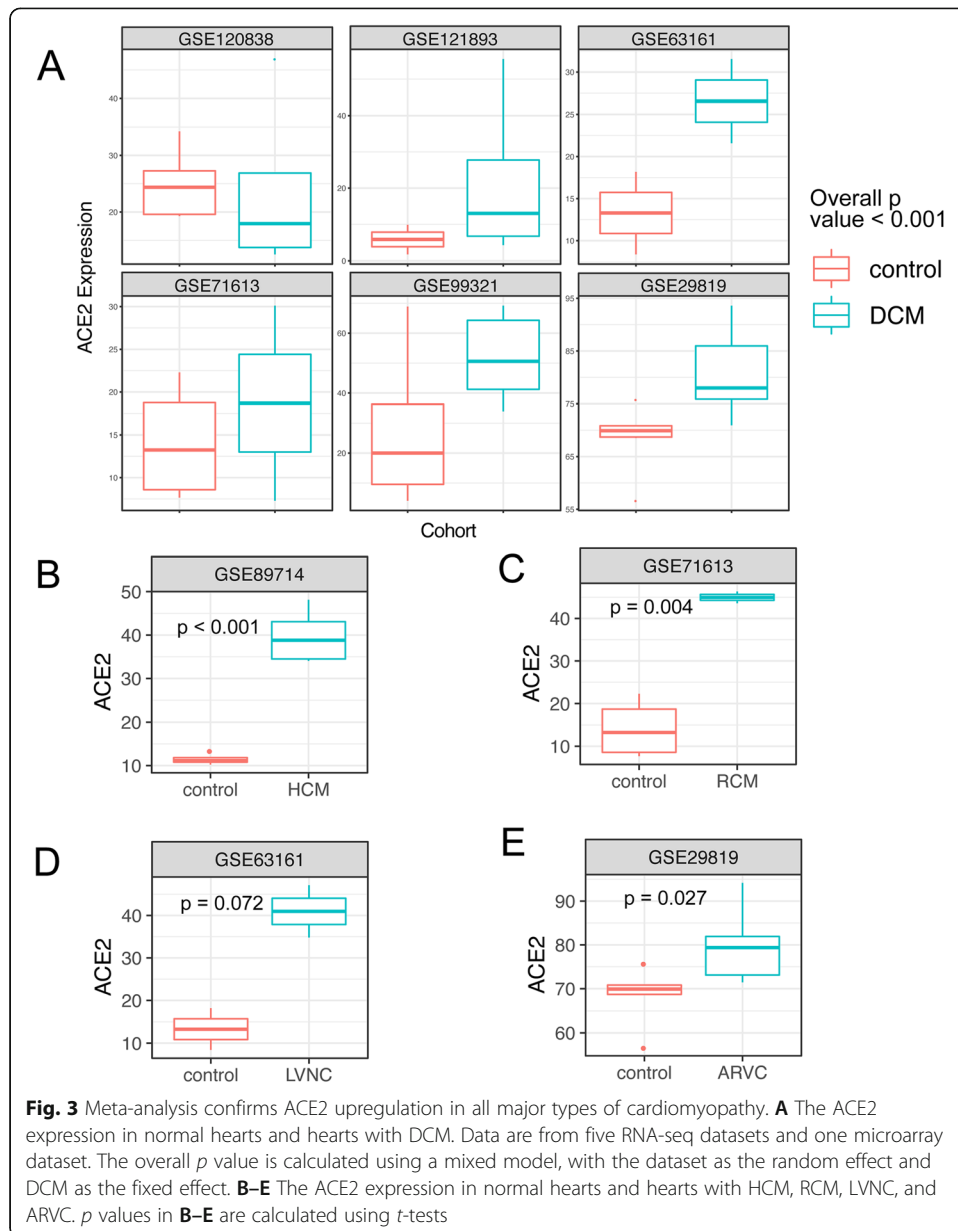
COVID-19 patients with pre-existing cardiomyopathy show an increased mortality rate

While COVID19 patients with cardiovascular conditions show a higher mortality rate, it is not clear how cardiomyopathy, in particular, affects the survival of the patients. Because the ACE2 expression is significantly elevated in the heart of cardiomyopathy patients, we hypothesize that pre-existing cardiomyopathy leads to increased mortality in patients with COVID19.

We identified 3936 COVID19 patients from the electronic health records (EHR) of the University of California San Francisco (UCSF) hospital. We divided the patients into three groups, including patients with pre-existing cardiomyopathy ($N = 43$), patients with other pre-existing cardiovascular diseases ($N = 624$), and patients without cardiovascular diseases ($N = 3269$) (Table 2). The most common non-cardiomyopathy cardiovascular diseases include hypertension ($N = 424$), atherosclerotic heart diseases ($N = 120$), and cardiac arrhythmia ($N = 105$).

We first compared the cardiomyopathy patients to patients without cardiovascular diseases. Patients with cardiomyopathy have a larger proportion of males and older ages. They also have a higher percentage of patients with pre-existing conditions such as cancer, diabetes, and hyperlipidemia. A higher percentage of cardiomyopathy patients have severe COVID-19 disease presentations, including ventilator use, respiratory failure, chest pain, and death (Table 2). We then performed survival analysis to test the effect of cardiomyopathy while controlling for differences in age, gender, and pre-existing conditions using a multivariable Cox proportional-hazards model. We confirmed that cardiomyopathy is significantly associated with the risk of death ($p = 0.004$) (Fig. 4A).

We next compared the cardiomyopathy patients to patients with other cardiovascular diseases. Cardiomyopathy patients have a higher proportion of males compared to patients with other cardiovascular diseases. Age, race, and pre-existing conditions are comparable between the two groups. Again, we observe that a higher percentage of cardiomyopathy patients have severe COVID-19 presentations, including ventilator use, chest pain, and death (Table 2). Multivariable Cox proportional-hazards regression confirms that cardiomyopathy is significantly associated with the risk of death ($p = 0.038$, 438% increase in observed death rate) (Fig. 4A). We further confirmed the increased mortality by comparing cardiomyopathy patients with a propensity score-matched cohort of patients with other cardiovascular diseases (Fig. 4B and Additional file 1, Table S8).



We then examined the survival of cardiomyopathy patients who are COVID-19 negative. We compared the survival of COVID-19-negative cardiomyopathy patients ($N = 2250$) with a propensity score-matched cohort of patients with other cardiovascular diseases ($N = 18,000$). The two cohorts are comparable in demographics and non-cardiovascular diseases (Additional file 1, Table S9). The 5-year mortality rate is only slightly higher in the cardiomyopathy patients ($p = 0.034$, 22% increase in observed death rate) (Fig. 4C). When we consider the patient’s survival at 160 days, a time frame comparable to the COVID-19-positive dataset, there is no significant difference between the survival of the two groups (Fig. 4C).

Taken together, the results show that cardiomyopathy itself does not pose large additional risk of mortality among patients with cardiovascular diseases. Rather, the interaction between SARS-CoV-2 infection and pre-existing cardiomyopathy leads to

Table 2 Demographic and clinical information of COVID-19 patients

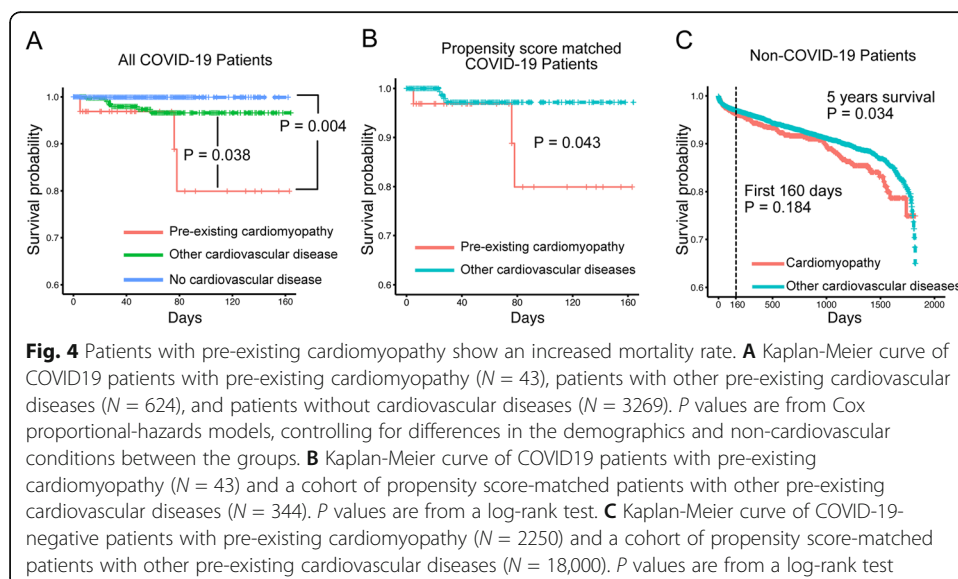
	Cardiomyopathy (N=43)	Other cardiovascular diseases (N=624)	No cardiovascular diseases (N=3269)	Overall (N=3936)
Demographics				
Age				
Mean (SD)	58.6 (17.9)	55.9 (21.0)	35.4 (22.0)***	38.9 (23.1)
Median [Min, Max]	60.0 [18.0, 89.0]	60.0 [0, 100]	33.0 [0, 98.0]	37.0 [0, 100]
Male	33 (76.7%)	296 (47.4%)*	1481 (45.3%)***	1810 (46.0%)
Race				
Asian	12 (27.9%)	103 (16.5%)	489 (15.0%)	604 (15.3%)
African American	5 (11.6%)	61 (9.8%)	203 (6.2%)	269 (6.8%)
MultiRace	1 (2.3%)	13 (2.1%)	59 (1.8%)	73 (1.9%)
Pacific Islander	3 (7.0%)	13 (2.1%)	56 (1.7%)	72 (1.8%)
Unknown	9 (20.9%)	142 (22.8%)	1607 (49.2%)	1758 (44.7%)
White	13 (30.2%)	289 (46.3%)	849 (26.0%)	1151 (29.2%)
Native American	0 (0%)	3 (0.5%)	6 (0.2%)	9 (0.2%)
Pre-existing conditions				
Hyperlipidemia	27 (62.8%)	297 (47.6%)	72 (2.2%)***	396 (10.1%)
Diabetes	27 (62.8%)	281 (45.0%)	136 (4.2%)***	444 (11.3%)
Cancer	10 (23.3%)	145 (23.2%)	67 (2.0%)***	222 (5.6%)
Severe COVID19 presentations				
Ventilator use	24 (55.8%)	148 (23.7%)***	88 (2.7%)***	260 (6.6%)
Respiratory failure	12 (27.9%)	96 (15.4%)	34 (1.0%)***	142 (3.6%)
Chest pain	27 (62.8%)	236 (37.8%)*	140 (4.3%)***	403 (10.2%)
Mortality	3 (7.0%)	8 (1.3%)	1 (0.0%)***	12 (0.3%)

Stars indicate the significant differences when compared the group to the cardiomyopathy group. * P<0.05, ** P<0.01,*** P<0.001

increased mortality in patients. Our transcriptomics analysis suggests that the upregulated ACE2 expression may contribute to the disease severity of COVID-19 in patients with pre-existing cardiomyopathy. However, further mechanistic studies are needed to establish the causal relationship between ACE2 upregulation and mortality in COVID-19 patients.

Discussion

The disease severity of COVID-19 patients varies from asymptomatic to life-threatening. While we do not fully understand the reason behind such variation, it is



clear that the disease severity is determined by multiple factors, including age, gender, the status of the immune system, and the pre-existing conditions [7, 39–41]. ACE2 expression is a key determinant of the disease severity, as shown by multiple studies in humans and in animal models [3, 5, 9]. Therefore, it is critical to identify conditions associated with ACE2 expression, as the information will help us reveal and explain factors associated with increased risk of severe illness from COVID-19.

We leveraged the massive amount of publicly available RNA-seq data to identify the ACE2-modulating conditions. While many tools exist for analyzing bulk GEO data, they are not optimized for this purpose. First, some tools require researchers to search for datasets using keywords, such as the name of a drug or a disease. These tools do not address our needs, as we are looking for any conditions that modulate ACE2 expression. Second, some tools require manual annotation of experimental groups in the studies, which are not scalable and often only cover a small subset of currently available datasets. Finally, the existing tools focus on differential expression analysis of two groups and are unable to address more complex experimental designs.

We address the problems using a variance analysis approach. Instead of comparing two experimental groups, we quantify the variance of the gene expression across all samples in a study. The rationale is that datasets with large ACE2 variances are likely to contain conditions that are associated with ACE2 expression. We improved our approach by using two modifications. First, we numerically embedded the metadata and calculated the regression coefficient R^2 between ACE2 and the embedding. This allows us to prioritize datasets in which the ACE2 variation is associated with metadata. Second, we controlled for the overall heterogeneity of samples in the study, this allows us to prioritize datasets in which ACE2 are specifically modulated rather than as a result of cell-type differences.

Our study identifies multiple diseases, conditions, and genetic perturbations that are associated with ACE2 expression. When interpreting these findings, readers should take into account several limitations of our study. First, many of the conditions are discovered based on data from one study with small sample sizes. These results should be viewed as data-driven hypotheses rather than definitive proofs. Additional data is required to confirm the findings. Second, our analysis does not establish a direct link between COVID-19 risk and the identified ACE2 modulators. Clinical studies are required to test if these ACE2-modulating conditions alter the risk of COVID-19 infection and pathogenesis. Third, many of the RNA-seq datasets are derived from observational studies. The association between conditions and ACE2 expression does not indicate causal relationships. Finally, the datasets profile the ACE2 expression in bulk tissues. Therefore, it is not clear if the variation of ACE2 expression is due to a change in cellular composition or a change in transcriptional regulation. Single-cell analysis is required to identify the cause of ACE2 expression change.

Heart diseases are associated with severe COVID-19 illness through two mechanisms. Pre-existing heart conditions are comorbidities of COVID19 [28, 41]. On the other hand, SARS-COV-2 infection can induce acute myocardial injury [30]. Cardiomyopathy is one of the most common heart diseases. Compared to other age-related heart conditions, cardiomyopathy can affect individuals at any age. However, clinical studies were lacking to specifically characterize the disease severity of COVID-19 in patients with cardiomyopathy. Our transcriptional analysis highlights the significantly increased

ACE2 expression in the hearts of cardiomyopathy patients, suggesting that cardiomyopathy patients are at an increased risk of heart damage caused by COVID-19.

Consistent with the RNA-seq finding, we found that COVID19 patients with pre-existing cardiomyopathy show increased mortality risk than other populations, even compared with age-matched populations with other cardiovascular conditions. While mechanistic research is needed to establish the causal relationship between ACE2 up-regulation and the increased mortality, our result identifies the cardiomyopathy patients as a high-risk group that needs extra protection and care.

This project demonstrated public RNA-sequencing data as a valuable resource for biomedical knowledge discovery. The existing RNA-seq data can be quickly repurposed to address pressing problems, such as identifying the risk factor of COVID19. While this study focused on analyzing the expression of ACE2, researchers can identify modulators for any genes or gene signatures of interest using the GENEVA web portal at genevatool.org.

Conclusions

We applied GENEVA (GENe Expression Variance Analysis) to 28,6650 publicly available RNA-seq samples to identify any previously studied experimental conditions that could directly or indirectly modulate ACE2 expression. We identified multiple drugs, genetic perturbations, and diseases that modulate the expression of ACE2, including cardiomyopathy, HNF1A overexpression, and drug treatments with RAD140 and itraconazole. Our unbiased meta-analysis of seven datasets confirms ACE2 upregulation in all cardiomyopathy categories. Using electronic health records data from 3936 COVID19 patients, we demonstrate that patients with pre-existing cardiomyopathy have an increased mortality risk than matched patients with other cardiovascular conditions.

Methods

Data preparation

We downloaded the uniformly processed RNA-seq data from ARCHS4 website (<https://amp.pharm.mssm.edu/archs4/download.html>) on August 03, 2020. The downloaded data include gene-level count data of 286,650 samples from 9124 datasets and sample-level metadata. We transformed the gene count data into percentile rank data, which reduces the influences of library size, batch effects, and extreme values [42, 43]. We downloaded study-level metadata using the `entrez_search` and `entrez_summary` function from the `rentrez` library [44].

Co-expression analysis

We first calculated the Pearson correlation between ACE2 and other genes using data from all 286,650 samples. For related dataset that share samples with each other, we only included one dataset with the largest sample size and exclude the other related dataset to make sure the studies are independent. Transcription factors are identified by selecting genes with the Gene Ontology term “DNA-binding transcription factor activity (GO:0003700).” We also performed mixed-effect regression to assess the association between ACE2 and other genes while controlling for study differences [Gene ~

ACE2 (fixed effect) + dataset ID (random effect for intercept and slope)]. The lme4 R package was used to fit the mixed-effect models. To identify pathways associated with ACE2, we used the associations with ACE2 (measured by the t statistics of the fixed effect in the mixed-effect model) as a signature. We used the signature to query the Gene Ontology Biological Process database [45]. The fgsea function from the fgsea library was used to calculate the enrichment score [46].

Metadata embedding

We first concatenated the metadata of each sample into a single string, including the title, tissue type, and other characteristics (e.g. demographics, time points, treatment, genetic information, and disease status). We then calculated the pairwise Levenshtein distance between the strings that belong to the same study (GEO series). We applied multidimensional scaling to the pairwise Levenshtein distance and embedded the strings into 2-dimensional space for visualization and downstream analysis.

GENEVA analysis

For a given gene in a given dataset, we first calculated the variance of the gene (VAR_g). We measure the overall heterogeneity of the samples by calculating the average variance of all genes (VAR_m). We run a regression using the expression of the gene as the dependent variable and the embedded metadata as independent variables (expression ~ first embed dimension + second embed dimension). The regression coefficient (R^2) represents the association between the expression of the gene and the embedded metadata. The product between VAR_g and R^2 represents the variance of the gene explained by the embedded metadata. The GENEVA score is defined as VAR_g × R^2 / VAR_m.

To test the significance of the GENEVA scores, we shuffled the samples within each dataset. We then calculated the GENEVA scores of all shuffled datasets to create a null distribution. Given a GENEVA score G , its p value is defined as the probability that the null distribution is greater than G : p value = Prob(null > G). We adjust the p values for multiple testing using the false discovery rate method.

Joint analysis of cardiomyopathy datasets

We searched the gene expression omnibus using the keyword “cardiomyopathy.” We then filter the results to only include studies that (1) profiled the transcriptome of heart tissues from humans and (2) compared cardiomyopathy samples with healthy samples. We identified 7 studies. We used a mixed-effect model to test the effect of cardiomyopathy on ACE2 expression: ACE2 expression ~ study (random effect) + cardiomyopathy status (fixed effect).

To examine the ACE2 expression in different types of cardiomyopathy, we separated the cardiomyopathy samples based on their subtype. We matched the cardiomyopathy samples with healthy controls within the same study. We used unpaired T -tests to test the effect of each cardiomyopathy type on ACE2 expression. Since data from multiple studies are available for dilated cardiomyopathy (DCM), we used a mixed-effect model to test the effect of DCM on ACE2 expression: ACE2 expression ~ study (random effect) + cardiomyopathy status (fixed effect).

To test for publication bias, Egger regression was performed by fitting a linear model using effect size/standard error as the dependent variable and 1/standard error as the dependent variable. The p value of the intercept term was used to assess the significance of the publication bias.

Analysis of electronic health records

The UCSF COVID-19 Data Mart records the clinical information of COVID19 patients and selected control patients using the Observational Medical Outcomes Partnership (OMOP) data format. We identified COVID19 patients from the clinical data using the ICD10 code U07. 1. We identified cardiomyopathy patients using the ICD10 codes I42 and I43, entered before their first COVID19 diagnosis. We identified patients with other cardiovascular diseases using the following ICD10 codes I00 - I99. We compared the cardiomyopathy patient with patients with other cardiovascular diseases and patients without cardiovascular diseases. We tested if the demographic and clinical variables are significantly different between the groups using single variable logistic regressions (cardiomyopathy ~ clinical variable). We used the `coxph` function in the survival R package to perform the survival analysis, controlling for the variables that are significantly different in the logistic regressions. Patient survival time is defined as the time between their first COVID19 diagnosis and their death date. Live patients are censored on the last day of their encounter. In the COVID-19-negative cohort, cardiomyopathy patients are defined as patients whose first cardiovascular-related diagnosis is cardiomyopathy (ICD10 code I42 or I43). Patients with other cardiovascular diseases are defined as patients who have cardiovascular diseases (ICD10 code I00 - I99), but do not have cardiomyopathy. Patient survival time is defined as the time between their first cardiovascular disease and their death. Live patients are censored on the last day of their encounter. To perform propensity score matching, we first calculated the propensity score using logistic regression (cardiomyopathy ~ age + race + gender + non-cardiovascular pre-existing conditions). We subsampled the non-cardiomyopathy cohort so that the distribution of its propensity score matches with the cardiomyopathy cohort.

GENEVA web portal

GENEVA web tool is an open-source application available under GNU General Public License at <http://genevatool.org>. It is implemented in python web framework Django. The source code for the tool is available in the public Git repository at <https://github.com/NavchetanKaur/geneva-webtool>. The tool offers an intuitive interface and user guide on the home page. Users can select either of the two options from “Gene Query” and “Gene Signature Query” and query their gene of interest or set of upregulated and downregulated genes of interest. The results are displayed in tabular form with calculated GENEVA scores. GSE descriptions are further represented in plots and tables.

All experimental methods comply with the Helsinki Declaration.

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02589-4>.

Additional file 1: Supplemental figures and tables. This file including figure S1, figure S2, table S8 and table S9.

Additional file 2: Table S1. The correlation Pearson between ACE2 and other genes across all datasets.

Additional file 3: Table S2. The correlation Pearson between ACE2 and transcription factors across all datasets.

Additional file 4: Table S3. The enriched pathways related to genes that are correlated with ACE2 across all datasets.

Additional file 5: Table S4. Comparing the co-expression relationships of ACE2 in GEO and GTEx datasets.

Additional file 6: Table S5. The association between ACE2 and other genes, measured by the t statistics from mix-effect models that controls for study differences.

Additional file 7: Table S6. The association between ACE2 and transcription factors, measured by the t statistics from mix-effect models that controls for study differences.

Additional file 8: Table S7. The enriched pathways related to genes that are associated with ACE2, measured by the t statistics from mix-effect models that controls for study differences.

Additional file 9: Review history.

Acknowledgements

We thank all researchers who have contributed RNA-seq datasets to the Gene Expression Omnibus. We thank Dr. Alexander Lachmann, Dr. Avi Ma'ayan, and other co-authors for creating ARCHS4, which enabled our study. We thank Dr. Douglas Arneson and Sanchita Bhattacharya for helpful discussion. We thank the UCSF COVID-19 Data Mart for providing the electronic health record data of COVID-19 patients.

Review history

The review history is available as Additional File 9.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

Z.H., B.O., and N.K. designed and implemented GENEVA web tool. Z. H. and N.K. performed the analysis. A.B. gave valuable input and suggestions for analysis. Z.H and A.B. led the study. All authors wrote the manuscript. All author(s) read and approved the final manuscript.

Funding

This work was supported by the National Institute of Allergy and Infectious Diseases (bioinformatics support contract HHSN316201200036W and research grant UH2AI153016) and the common fund of the National Institute of Health (U24DK116214). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

GENEVA web tool is an open-source application available under GNU General Public License at <http://genevatool.org>. The source code for the GENEVA analysis is in the public Git repository at <https://github.com/NavchetanKaur/geneva-webtool>. A stable version of the source code is deposited at Zenodo [47]. The uniformly processed RNA-seq data are available to be explored using the GENEVA webtool and can also be downloaded through the ARCHS4 website (<https://amp.pharm.mssm.edu/archs4/download.html>), including the datasets that are highlighted in the article (GSE89714 [48], GSE104177 [49], GSE114013 [50], GSE120838 [51], GSE121893 [52], GSE63161 [53], GSE71613 [54], GSE99321 [55], and GSE29819 [56]). The clinical data of COVID-19 from UCSF electronic health records is not shared at subject level due to institutional restrictions. The summary statistics of the clinical data is available in the main text or the supplementary materials.

Declarations

Ethics approval and consent to participate

The analysis of the electronic health record data is conducted with approval from the UCSF institutional review board (IRB #: 20-31107).

Competing interests

Atul Butte is a co-founder and consultant to Personalis and NuMedii; consultant to Samsung, Mango Tree Corporation, and in the recent past, 10x Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Facebook, Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, CVS, Nuna Health, Assay Depot, Vet24seven, Regeneron, Sanofi, Royalty Pharma, AstraZeneca, Moderna, Biogen, Paraxel, and Sutro, and several other non-health related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease-specific foundations and associations, and health systems. Atul Butte receives royalty payments through Stanford University, for several patents and other

disclosures licensed to NuMedii and Personalis. Atul Butte's research has been funded by NIH, Northrup Grumman (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervallien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

Author details

¹Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA.

²Department of Pediatrics, University of California, San Francisco, CA, USA. ³Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA.

Received: 10 November 2020 Accepted: 23 December 2021

Published online: 10 January 2022

References

- WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2021 Aug 11]. Available from: <https://covid19.who.int>
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. 2020;181:271–280.e8.
- Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. 2020;181:1036–1045.e9.
- Hassan AO, Case JB, Winkler ES, Thackray LB, Kafai NM, Bailey AL, et al. A SARS-CoV-2 Infection Model in Mice Demonstrates Protection by Neutralizing Antibodies. *Cell*. 2020;182:744–753.e4.
- Bao L, Deng W, Huang B, Gao H, Liu J, Ren L, et al. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature*. Nature Publishing Group. 2020;583:830–3.
- Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. *Am J Gastroenterol* [Internet]. 2020; [cited 2020 Aug 18];115. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7172492/>.
- Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19 in New York City. *New England Journal of Medicine*. Massachusetts Medical Society. 2020;382:2372–4.
- Li J-W, Han T-W, Woodward M, Anderson CS, Zhou H, Chen Y-D, et al. The impact of 2019 novel coronavirus on heart injury: A Systematic review and Meta-analysis. *Progress in Cardiovascular Diseases* [Internet]. 2020; [cited 2020 Aug 18]; Available from: <http://www.sciencedirect.com/science/article/pii/S0033062020300803>.
- Pinto BGG, Oliveira AER, Singh Y, Jimenez L, Gonçalves ANA, Ogawa RLT, et al. ACE2 Expression Is Increased in the Lungs of Patients With Comorbidities Associated With Severe COVID-19. *J Infect Dis*. 2020;222:556–63.
- Consortium TGte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348:648–60.
- Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*. Nature Publishing Group. 2020;26:681–7.
- Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications*. Nature Publishing Group. 2018;9:1366.
- Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nature Communications*. Nature Publishing Group; 2016;7:12846.
- Koeppen K, Stanton BA, Hampton TH. ScanGEO: parallel mining of high-throughput gene expression data. *Bioinformatics*. Oxford Academic. 2017;33:3500–1.
- Feng C, Araki M, Kunimoto R, Tamon A, Makiguchi H, Nijima S, et al. GEM-TREND: a web tool for gene expression data mining toward relevant network discovery. *BMC Genomics*. 2009;10:411.
- Hadley D, Pan J, El-Sayed O, Aljabban J, Aljabban I, Azad TD, et al. Precision annotation of digital samples in NCBI's gene expression omnibus. *Scientific Data*. Nature Publishing Group; 2017;4:170125.
- Chang JT, Gatza ML, Lucas JE, Barry WT, Vaughn P, Nevins JR. SIGNATURE: A workbench for gene expression signature analysis. *BMC Bioinformatics*. BioMed Central; 2011;12:443.
- Williams G. A searchable cross-platform gene expression database reveals connections between drug treatments and disease. *BMC Genomics*. 2012;13:12.
- Fujibuchi W, Kiseleva L, Taniguchi T, Harada H, Horton P. CellMontage: similar expression profile search server. *Bioinformatics*. Oxford Academic. 2007;23:3103–4.
- Engreitz JM, Chen R, Morgan AA, Dudley JT, Mallelwar R, Butte AJ. ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression. *Bioinformatics*. 2011;27:3317–8.
- Zinman GE, Naiman S, Kanfi Y, Cohen H, Bar-Joseph Z. ExpressionBlast: mining large, unstructured expression databases. *Nat Methods*. 2013;10:925–6.
- Zhu Q, Wong AK, Krishnan A, Aure MR, Tadych A, Zhang R, et al. Targeted exploration and analysis of large cross-platform human transcriptomic compendia. *Nature Methods*. Nature Publishing Group. 2015;12:211–4.
- Bernardi S, Tikellis C, Candido R, Tsorotes D, Pickering RJ, Bossi F, et al. ACE2 deficiency shifts energy metabolism towards glucose utilization. *Metab Clin Exp*. 2015;64:406–15.
- Clarke NE, Belyaev ND, Lambert DW, Turner AJ. Epigenetic regulation of angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress. *Clin Sci*. 2014;126:507–16.
- Kb P, Kh C, Vk N, H X, E L. The transcription factor HNF1 α induces expression of angiotensin-converting enzyme 2 (ACE2) in pancreatic islets from evolutionarily conserved promoter motifs [Internet]. *Biochimica et biophysica acta*; 2013 [cited 2020 Jul 31]. Available from: <https://pubmed.ncbi.nlm.nih.gov/24100303/>
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57:289–300.

27. Husser D, Ueberham L, Jacob J, Heuer D, Riedel-Heller S, Walker J, et al. Prevalence of clinically apparent hypertrophic cardiomyopathy in Germany—An analysis of over 5 million patients. *PLoS One* [Internet]. 2018; [cited 2020 Aug 24];13. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5937277/>.
28. Atkins JL, Masoli JAH, Delgado J, Pilling LC, Kuo C-L, Kuchel GA, et al. Preexisting Comorbidities Predicting COVID-19 and Mortality in the UK Biobank Community Cohort. *J Gerontol A Biol Sci Med Sci* [Internet]. [cited 2020 Aug 20]; Available from: <https://academic.oup.com/biomedgerontology/advance-article/doi/10.1093/gerona/glaa183/5873904>
29. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol. American Medical Association*. 2020;5:811–8.
30. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, et al. Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. *JAMA Cardiol. American Medical Association*. 2020;5:802–10.
31. Jefferies JL, Towbin JA. Dilated cardiomyopathy. *Lancet*. 2010;375:752–62.
32. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242–55.
33. Kushwaha SS, Fallon JT, Fuster V. Restrictive Cardiomyopathy. *New England Journal of Medicine. Massachusetts Medical Society*. 1997;336:267–76.
34. Basso C, Corrado D, Marcus FI, Nava A, Thiene G. Arrhythmogenic right ventricular cardiomyopathy. *Lancet*. 2009;373:1289–300.
35. Ross SB, Jones K, Blanch B, Puranik R, McGeechan K, Barratt A, et al. A systematic review and meta-analysis of the prevalence of left ventricular non-compaction in adults. *Eur Heart J. Oxford Academic*. 2020;41:1428–36.
36. Bos JM, Hebl VB, Oberg AL, Sun Z, Herman DS, Teekakirikul P, et al. Marked Up-Regulation of ACE2 in Hearts of Patients With Obstructive Hypertrophic Cardiomyopathy: Implications for SARS-CoV-2–Mediated COVID-19. *Mayo Clin Proc*. 2020; 95:1354–68.
37. Goulter AB, Goddard MJ, Allen JC, Clark KL. ACE2 gene expression is up-regulated in the human failing heart. *BMC Medicine*. 2004;2:19.
38. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ. British Medical Journal Publishing Group*. 1997;315:629–34.
39. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*. 2020;181:1489–1501.e15.
40. Jin J-M, Bai P, He W, Wu F, Liu X-F, Han D-M, et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Front Public Health* [Internet]. *Frontiers*. 2020; [cited 2020 Aug 19];8. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00152/full?utm_source=yxnews&utm_medium=desktop&utm_referrer=https%3A%2F%2Ffyandex.by%2Fnews.
41. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine. Massachusetts Medical Society*. 2020;382:1708–20.
42. Hu Z, Glicksberg BS, Butte AJ. Robust prediction of clinical outcomes using cytometry data. *Bioinformatics*. 2019;35:1197–203.
43. Zwiener I, Frisch B, Binder H. Transforming RNA-Seq Data to Improve the Performance of Prognostic Gene Signatures. *PLOS ONE. Public Library of Science*. 2014;9:e85150.
44. Winter DJ. *retrez2: An R package for the NCBI eUtils API*. *The R Journal*. 2017;9:520–6.
45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–9.
46. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv. Cold Spring Harbor Laboratory*. 2019:060012.
47. Kaur N, Oskotsky B, Butte A, Hu Z. GENEVA web tool. *zenodo*. <https://doi.org/10.5281/zenodo.5735451>. 2021.
48. Li Y, Guo H. Differential gene expressions in the heart of hypertrophic cardiomyopathy patients. *GSE89714. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89714>. 2016.
49. Wang D, He S, Yu Z. Global modulation of signaling pathways by SARM RAD140 in AR/ER+ breast cancer xenografts. *GSE104177. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104177>. 2017.
50. Buczacki SJ. Identifying the molecular mode of action of itraconazole in colorectal cancer. *GSE114013. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114013>. 2018.
51. Marian AJ. Genomic Reorganization of Lamin-Associated Domains in Cardiac Myocytes is Associated with Differential Gene Expression and DNA Methylation in Human Dilated Cardiomyopathy. *GSE120838. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120838>. 2019.
52. Wang L, Yu P, Li Z, Ren Z. Dissecting cell composition and cell-cell interaction network of human disease heart tissue by single-cell sequencing. *GSE121893. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121893>. 2019.
53. Kodo K, Hu S, InanlooRahatloo K, Bernstein D, Wu JC. hiPSCs unravel aberrant TGFβ signaling as an etiology of left ventricular non-compaction. *GSE63161. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63161>. 2016.
54. Schiano C, Costa V, Casamassmi A, Aprile M, Rienzo M, Esposito R, et al. RNA-Sequencing shows novel transcriptomic signatures in failing and non-failing human heart. *GSE71613. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71613>. 2015.
55. Tatman P. RNA-sequencing of pediatric idiopathic dilated cardiomyopathy patients and healthy controls. *GSE99321. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99321>. 2017.
56. Gärtner A, Schwientek P, Ellinghaus P, Summer H, Golz S, Kassner A, et al. Myocardial transcriptome analysis of human arrhythmogenic right ventricular cardiomyopathy (ARVC). *GSE29819. Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29819>. 2011.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.