

1 Mining transcriptomics and clinical data reveals ACE2 expression 2 modulators and identifies cardiomyopathy as a risk factor for 3 mortality in COVID-19 patients

4

5 Navchetan Kaur¹, Boris Oskotsky¹, Atul J. Butte¹, and Zicheng Hu^{1,*}

6

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

9 * Correspondence may be addressed to zicheng.hu@ucsf.edu

10

11 Abstract

12 Angiotensin-converting enzyme 2 (ACE2) is the cell-entry receptor for SARS-CoV-2. It plays
13 critical roles in both the transmission and the pathogenesis of the coronavirus disease 2019
14 (COVID-19). Comprehensive profiling of ACE2 expression patterns will help researchers to
15 reveal risk factors of severe COVID-19 illness. While the expression of ACE2 in healthy human
16 tissues has been well characterized, it is not known which diseases and drugs might modulate
17 the ACE2 expression. In this study, we developed GENEVA (GENe Expression Variance
18 Analysis), a semi-automated framework for exploring massive amounts of RNA-seq datasets.
19 We applied GENEVA to 28,6650 publicly available RNA-seq samples to identify any previously
20 studied experimental conditions that could directly or indirectly modulate ACE2 expression. We
21 identified multiple drugs, genetic perturbations, and diseases that modulate the expression of
22 ACE2, including cardiomyopathy, HNF1A overexpression, and drug treatments with RAD140
23 and Itraconazole. Our unbiased meta-analysis of seven datasets confirms ACE2 up-regulation
24 in all cardiomyopathy categories. Using electronic health records data from 3936 COVID19
25 patients, we demonstrate that patients with pre-existing cardiomyopathy have an increased
26 mortality risk than age-matched patients with other cardiovascular conditions. GENEVA is
27 applicable to any genes of interest and is freely accessible at <http://genevatool.org>.

28

29

30 Introduction

31 Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute
32 respiratory syndrome coronavirus 2 (SARS-COV-2). The World Health Organization (WHO)
33 declared the COVID-19 outbreak a pandemic on March 11, 2020. As of October 5, 2020, there
34 have been 34.8 million recorded COVID-19 cases and over 1 million deaths¹.

35

36 Angiotensin-converting enzyme 2 (ACE2) is the cell-entry receptor for SARS-CoV-2². The
37 binding between ACE2 and spike (S) protein of SARS-COV-2 initiates the viral entry into target
38 cells. ACE2 plays key roles in both the transmission and pathogenesis of SARS-CoV-2, as
39 demonstrated by the following lines of evidence: 1) SARS-CoV-2 fails to infect the lung-derived
40 cell line A549 in the absence of ACE2 expression. The infection is restored after overexpressing
41 ACE2 in the cell line³. 2) SARS-CoV-2 fails to infect wild type mice but can infect and cause
42 pneumonia in transgenic mice expressing human ACE2^{4,5}. 3) COVID19 related tissue damages

1 are detected in organs with ACE2 expression, including lungs, intestines, colons, and hearts⁶⁻⁸.
2 4) ACE2 expression is increased in the lungs of patients with comorbidities associated with
3 severe COVID-19, suggesting that the level of ACE2 expression is associated with disease
4 severity⁹. Taking these lines of evidence together, it is crucial to comprehensively characterize
5 the ACE2 expression in human tissues.

6
7 To comprehensively profile the expression patterns of ACE2, we not only need to characterize
8 its expression in healthy tissues but also identify diseases, drugs and genetic perturbations that
9 modulate ACE2 expression. The expression of ACE2 in healthy human tissues has been well
10 characterized by resources such as the Human Cell Atlas and GTEx, with the highest
11 expression detected in intestine, testis, lung, cornea, heart, kidney, and adipose tissues^{10,11}.
12 However, it is still not clear which diseases and drugs modulate the ACE2 expression. Since
13 ACE2 expression is tightly associated with the pathogenicity of SARS-COV-2, identification of
14 ACE2 modulating conditions will help us reveal and explain risk factors of severe illness from
15 COVID-19.

16
17 RNA-sequencing data profiles the full transcriptome of samples. Currently, more than 200,000
18 human RNA-seq samples are publicly available, providing an unprecedented opportunity for us
19 to examine ACE2 expression in different human cell types under a variety of conditions and
20 treatments. Data harmonization efforts such as ARCHS4 have uniformly preprocessed the
21 RNA-seq data, making them readily available for analysis¹². However, fully automated analysis
22 of these datasets faces two main obstacles. First, the metadata are non-standardized and are
23 often unstructured, making it difficult to extract experimental conditions from the studies.
24 Second, experimental designs are highly variable. While some studies adopt the simple control-
25 versus-treatment design, other studies are more complicated, involving multiple time points,
26 combination treatments, or stratified cohorts. The heterogeneous design makes it difficult to
27 analyze the datasets using a single statistical model.

28
29 Multiple tools have been made to analyze transcriptomics data, including CREEDS¹³,
30 scanGEO¹⁴, GEM-TREND¹⁵, StarGEO¹⁶, SIGNATURE¹⁷, SPIED¹⁸, Cell Montage¹⁹,
31 ProfileChaser²⁰, ExpressionBlast²¹ and SEEK²². However, the existing tools have several
32 limitations, preventing them from fully exploring the publicly available RNA-sequencing
33 resources. First, some of the tools annotate the metadata manually and are unable to cover the
34 large number of datasets currently available. Second, the tools focus on differential expression
35 analysis between two groups (e.g., control versus treatment), preventing them from analyzing
36 studies with more complex study designs.

37
38 In this study, we developed GENEVA (GENe Expression Variance Analysis), a semi-automated
39 framework for exploring public RNA-seq datasets. For a given gene, GENEVA identifies the
40 most relevant datasets by analyzing the variance of the gene expression. GENEVA visualizes
41 the relevant datasets for detailed manual analysis. GENEVA is scalable and is agnostic to study
42 designs. Using GENEVA, we identified multiple drugs, genetic perturbations, and diseases that
43 modulate the expression of ACE2, including cardiomyopathy, HNF1A over-expression, and drug
44 treatments with RAD140 and Itraconazole. Our in-depth meta-analysis of seven datasets

1 reveals increased ACE2 expression in all cardiomyopathy categories. By analyzing the clinical
2 data of 3936 COVID19 patients at UCSF hospital, we demonstrate that patients with pre-
3 existing cardiomyopathy have an increased mortality risk than other patients, including
4 propensity score-matched patients with other cardiovascular conditions.

5
6

7 **Results**

8

9 **Analysis of 286650 RNA-seq samples reveals complex transcriptional networks of ACE2.**

10 Our study leverages human RNA-sequencing data from the ARCHS4 project, containing
11 286,650 uniformly pre-processed data from 9,124 Gene Expression Omnibus (GEO) series¹².
12 The large number of RNA-sequencing samples provide an unprecedented resource for studying
13 the expression of ACE2 in different human cell types under a variety of conditions and
14 treatments.

15

16 We first characterized the transcriptional networks of ACE2 using all 286,650 samples. We
17 calculated the Pearson correlation between ACE2 and all other human genes. Because of the
18 large sample size, most of the correlations are statistically significant, even after multiple testing
19 adjustments. Therefore, we focused on the correlation coefficients themselves as a measure of
20 effect size, rather than the p values or significance. While most of the genes have correlation
21 coefficients near 0, a small set of genes are highly correlated with ACE2 (Supplemental Fig.
22 1A). Among 35,238 human genes, 85 (0.24%) genes have correlations higher than 0.5. The top
23 correlated genes include FABP2, MEP1B, and transcription factors such as HNF4G
24 (Supplemental Fig. 1B and C). We examined their expression in different human tissues using
25 GTEx data. We found that these genes are highly expressed in colons and small intestines, an
26 expression pattern similar to ACE2. Similarly, we see high correlations between ACE2 and
27 multiple pathways related to the digestive process (Supplemental Fig. 1D). These results
28 indicate that the correlations are dominated by the overall transcriptomic difference between the
29 digestive system and other human organs.

30

31 To avoid our analysis being dominated by known cellular or organ differences, we evaluated the
32 correlations between ACE2 and other human genes within each RNA-seq dataset. We then
33 calculated the mean and variances of the correlations across all datasets, weighted by the
34 sample sizes. Our analysis reveals a positive relationship between the mean and variance of
35 the correlation coefficients (Supplemental Fig. 1E). While some genes have a high average
36 correlation with ACE2, their correlation with ACE2 is highly variable in individual datasets. The
37 results suggest against a common transcriptional network around the ACE2 gene. Rather,
38 ACE2 is co-regulated with different sets of genes under different conditions.

39

40 We adjusted for the variances by calculating the standardized correlation coefficient, defined as
41 the mean correlation divided by the standard deviation of the correlation. The standardized
42 correlation coefficient allows us to prioritize genes with relatively conserved correlations with
43 ACE2 across all studies. The top genes include MYO7B, PDZK1 and transcription factors such
44 as HNF1A (Supplemental Fig. 1 F and G). Pathway analysis identified many metabolic

1 processes to be correlated with ACE2 expression. The findings are consistent with previous
2 observations that ACE2 is involved in glucose metabolism and energy stress responses^{23,24}.

3
4 Our analysis identified three transcription factors in the hepatocyte nuclear factor family,
5 including HNF4G, HNF1A, and HNF4A (Supplemental Fig. 1 C and G). HNF4G has the highest
6 overall correlation with ACE2 while HNF1A has the highest standardized correlation coefficient
7 within studies. We then tested the causal relationship between the transcription factors and
8 ACE2 expression. We identified two RNA-seq datasets that compared human cells with or
9 without genetic perturbation of HNF4G and HNF1A. While HNF1G is positively correlated with
10 ACE2 expression (Supplemental Fig. 1 C), overexpression of HNF1G does not lead to
11 significantly increased ACE2 expression (Supplemental Fig. 1I). Rather, there is a trend of
12 reduction in ACE2 expression. HNF1A overexpression leads to increased ACE2 expression and
13 HNF1A knockdown reduced ACE2 expression (Supplemental Fig. 1J) in LNCaP cells, a
14 prostate cancer cell line. The result is consistent with the positive correlation between HNF1A
15 and ACE2 (Supplemental Fig. 1G). A previous study showed that HNF1A induces ACE2 in
16 pancreatic islets²⁵. Our result in a prostate cancer cell line further confirmed the role of HNF1A
17 in regulating ACE2. However, it should be noted that HNF1A and ACE2 are not correlated in all
18 RNA-seq datasets (Supplemental Fig. 1K), demonstrating the complexity of ACE2 regulation in
19 different tissues.

20 21 **Gene expression variance analysis reveals diseases and therapeutics that modulate the** 22 **expression of ACE2.**

23
24 Next, we hope to identify conditions that modulate the expression of ACE2. We developed a
25 computational framework named GENEVA (Gene Expression Variance Analysis) to identify the
26 most relevant datasets for visualization and detailed manual analysis (Fig. 1A and Methods).
27 GENEVA prioritizes the datasets that have a large variance of ACE2 expression. The rationale
28 is that datasets with large ACE2 variance are likely to contain conditions that modulate the
29 ACE2 expression. At the same time, GENEVA controls for the overall heterogeneity of the
30 samples to prioritize datasets in which ACE2 is specifically modulated by experimental
31 conditions rather than due to tissue type differences. In addition, GENEVA embeds the meta-
32 data into numerical space and prioritizes datasets with high correlations between ACE2
33 expression and the metadata (Fig. 1B). This allows GENEVA to identify datasets in which ACE2
34 is regulated by experimental conditions rather than randomness or unexplained factors. While
35 our study focuses on ACE2 and its role in COVID-19 disease, GENEVA is applicable to all
36 genes. We created a web application that allows researchers to apply GENEVA to their gene of
37 interest [<http://genevatool.org>].

38
39 We tested the significance of the GENEVA scores using a permutation procedure. We randomly
40 shuffle the samples across studies to generate a null distribution. We compared each GENEVA
41 score to the null distribution to calculate the p-value. We adjusted for multiple-testing using the
42 false discovery rate (FDR) method²⁶. We identified 27 significant datasets with FDRs less than
43 0.05 (Table 1). Interestingly, GENEVA identified HNF1A as an ACE2 modulator, which was
44 also identified in our correlation analysis (Supplemental Fig. 1J). GENEVA additionally identified

1 multiple drugs and diseases that modulate the ACE2 expression, revealing potential risk factors
2 for severe illness from COVID-19.

3
4 Here, we highlight three ACE2 modulating conditions, manually picked based on their effect on
5 ACE2 expression and their potential impact on public health. Data from GSE89714 show up-
6 regulated expression of ACE2 in hypertrophic cardiomyopathy (Fig. 2 A, B). Hypertrophic
7 cardiomyopathy is the most common inherited heart disease, affecting an estimated 15,188,000
8 individuals (0.2%) worldwide²⁷. Our finding is consistent with an increased death rate in COVID-
9 19 patients with heart conditions²⁸⁻³⁰ and suggests that higher ACE2 expression can contribute
10 to the increased risk. Data from GSE104177 showed that RAD140, a selective androgen
11 receptor modulator, induces ACE2 expression in human breast cancer xenografts (Fig. 2 C, D).
12 Data from GSE114013 show that Itraconazole, an antifungal drug, up-regulates ACE2
13 expression in two colorectal cancer cell lines, HT55 and SW948 (Fig. 2 E, F). These findings
14 suggest that these drugs should be studied with respect to ACE2 expression in lung and cardiac
15 cells and tissues, and that patients on these drugs should be studied closely during the
16 pandemic. If these subsequent studies do continue to suggest this effect on increasing ACE2
17 expression, heightened caution could be warranted when using these drugs during the COVID-
18 19 pandemic.

19
20 **Unbiased meta-analysis shows ACE2 up-regulation in all types of cardiomyopathy.**

21
22 GENEVA prioritizes datasets with large variances in ACE2 expression. However, the procedure
23 may introduce bias, as studies with small ACE2 variations are ignored. Consider an example in
24 which multiple studies have profiled the effect of a drug. Some studies show that the drug up-
25 regulates ACE2 while other studies show that the drug has no effect on ACE2. The effect of the
26 drug will be overestimated if a researcher only includes the studies with positive results.
27 Therefore, after the GENEVA analysis, unbiased meta-analyses are required to confirm the
28 findings.

29
30 We performed a comprehensive search for datasets related to the three highlighted conditions,
31 including cardiomyopathy, itraconazole treatment, and RAD140 treatment. We did not find
32 additional datasets related to itraconazole and RAD140 treatment. For cardiomyopathy, we
33 identified a total of 7 datasets. We performed a meta-analysis using a mixed-effect model,
34 taking data from all 7 datasets into account. The result confirmed that ACE2 expression is
35 significantly elevated in heart tissue samples from cardiomyopathy patients (p-value < 0.001).

36
37 We next examine the ACE2 expression in different types of cardiomyopathy. The most common
38 types of cardiomyopathies includes dilated cardiomyopathy (DCM), hypertrophic
39 cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular
40 cardiomyopathy (ARVC) and left ventricular noncompaction (LVNC)³¹⁻³⁵. These cardiomyopathy
41 types have different causes and show distinctive morphology and physiology characteristics.
42 Although previous studies have demonstrated ACE2 up-regulation in DCM and HCM^{36,37}, how
43 ACE2 is regulated in other types is unknown. Within the 7 datasets, we were able to identify all
44 the common cardiomyopathy types. Our analysis revealed significantly increased ACE2

1 expression in most of the cardiomyopathy types (Fig. 3), including DCM, HCM, RCM, and
2 LVNC. Although the result of ARVC is not statistically significant, the data show a clear trend of
3 ACE2 upregulation (Fig. 3D).

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

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

10
11 We identified 3936 COVID19 patients from the electronic health records (EHR) of the University
12 of California San Francisco (UCSF) hospital. We divided the patients into three groups,
13 including patients with pre-existing cardiomyopathy (N = 43), patients with other pre-existing
14 cardiovascular diseases (N = 624), and patients without cardiovascular diseases (N = 3269)
15 (Table 2). The most common non-cardiomyopathy cardiovascular diseases include
16 hypertension (N = 424), atherosclerotic heart diseases (N = 120) and Cardiac arrhythmia (N =
17 105).

18
19 We first compared the cardiomyopathy patients to patients without cardiovascular diseases.
20 Patients with cardiomyopathy have a larger proportion of males and older ages. They also have
21 a higher percentage of patients with preexisting conditions such as cancer, diabetes, and
22 hyperlipidemia. A higher percentage of cardiomyopathy patients have severe COVID-19
23 disease presentations, including ventilator use, respiratory failure, chest pain, and death (Table
24 2). We then performed survival analysis to test the effect of cardiomyopathy while controlling for
25 differences in age, gender, and pre-existing conditions using a multivariable Cox proportional-
26 hazards model. We confirmed that cardiomyopathy is significantly associated with the risk of
27 death ($p = 0.004$) (Figure 4A).

28
29 We next compared the cardiomyopathy patients to patients with other cardiovascular diseases.
30 Cardiomyopathy patients have a higher proportion of males compared to patients with other
31 cardiovascular diseases. Age, race, and pre-existing conditions are comparable between the
32 two groups. Again, we observe that a higher percentage of cardiomyopathy patients have
33 severe COVID-19 presentations, including ventilator use, chest pain, and death (Table 2).
34 Multivariable Cox proportional-hazards regression confirms that cardiomyopathy is significantly
35 associated with the risk of death ($p = 0.038$, 438 % increase in observed death rate) (Figure
36 4A). We further confirmed the increased mortality by comparing cardiomyopathy patients with a
37 propensity score-matched cohort of patients with other cardiovascular diseases (Figure 4B and
38 Table S1).

1
2 We then examined the survival of cardiomyopathy patients who are COVID-19 negative. We
3 compared the survival of COVID-19 negative cardiomyopathy patients (N = 2250) with a
4 propensity score-matched cohort of patients with other cardiovascular diseases (N = 18000).
5 The two cohorts are comparable in demographics and non-cardiovascular diseases (Table S2).
6 The 5-year mortality rate is only slightly higher in the cardiomyopathy patients ($p = 0.034$, 22%
7 increase in observed death rate) (Fig. 4C). When we consider the patient's survival at 160 days,
8 a time frame comparable to the COVID-19 positive dataset, there is no significant difference
9 between the survival of the two groups (Fig. 4C).

10
11 Taken together, the results show that cardiomyopathy itself does not pose large additional risk
12 of mortality among patients with cardiovascular diseases. Rather, the interaction between
13 SARS-CoV-2 infection and pre-existing cardiomyopathy leads to increased mortality in patients.
14 Our transcriptomics analysis suggests that the up-regulated ACE2 expression may contribute to
15 disease severity of COVID-19 in patients with pre-existing cardiomyopathy. However, further
16 mechanistic studies are needed to establish the causal relationship between ACE2 up-
17 regulation and the mortality in COVID-19 patients.

18 19 **Discussion**

20
21 The disease severity of COVID-19 patients varies from asymptomatic to life-threatening. While
22 we do not fully understand the reason behind such variation, it is clear that the disease severity
23 is determined by multiple factors, including age, gender, the status of the immune system, and
24 the pre-existing conditions^{7,38-40}. ACE2 expression is a key determinant of the disease severity,
25 as shown by multiple studies in humans and in animal models^{3,5,9}. Therefore, it is critical to
26 identify conditions that modulate ACE2 expression, as the information will help us reveal and
27 explain factors associated with increased risk of severe illness from COVID-19.

28
29 We leveraged the massive amount of publicly available RNA-seq data to identify the ACE2
30 modulating conditions. While many tools exist for analyzing bulk GEO data, they are not
31 optimized for this purpose. First, some tools require researchers to search for datasets using
32 keywords, such as the name of a drug or a disease. These tools do not address our needs, as
33 we are looking for any conditions that modulate ACE2 expression. Second, some tools require
34 manual annotation of experimental groups in the studies, which are not scalable and often only
35 cover a small subset of currently available datasets. Finally, the existing tools focus on
36 differential expression analysis of two groups and are unable to address more complex
37 experimental designs.

38
39 We address the problems using a variance analysis approach. Instead of comparing two
40 experimental groups, we quantify the variance of the gene expression across all samples in a

1 study. The rationale is that datasets with large ACE2 variances are likely to contain conditions
2 that modulate the ACE2 expression. We improved our approach by using two modifications.
3 First, we numerically embedded the metadata and calculated the regression coefficient R^2
4 between ACE2 and the embedding. This allows us to prioritize datasets in which the ACE2
5 variation is associated with metadata. Second, we controlled for the overall heterogeneity of
6 samples in the study, this allows us to prioritize datasets in which ACE2 are specifically
7 modulated rather than as a result of cell-type differences.

8
9 Our study identifies multiple diseases, conditions, and genetic perturbations that modulate
10 ACE2 expression. When interpreting these findings, readers should take into account several
11 limitations of our study. First, many of the ACE2 modulating conditions are discovered based on
12 data from one study with small sample sizes. These results should be viewed as data-driven
13 hypotheses rather than definitive proofs. Additional data is required to confirm the findings.
14 Second, our analysis does not establish a direct link between COVID-19 risk and the identified
15 ACE2 modulators. Clinical studies are required to test if these ACE2 modulating conditions alter
16 the risk of COVID-19 infection and pathogenesis. Finally, the datasets profile the ACE2
17 expression in bulk tissues. Therefore, it is not clear if the variation of ACE2 expression is due to
18 a change in cellular composition or a change in transcriptional regulation. Single-cell analysis is
19 required to identify the cause of ACE2 expression change.

20
21 Heart diseases are associated with severe COVID-19 illness through two mechanisms. Pre-
22 existing heart conditions are comorbidities of COVID19^{28,40}. On the other hand, SARS-COV-2
23 infection can induce acute myocardial injury³⁰. Cardiomyopathy is one of the most common
24 heart diseases. Compared to other age-related heart conditions, cardiomyopathy can affect
25 individuals at any age. However, clinical studies were lacking to specifically characterize the
26 disease severity of COVID-19 in patients with cardiomyopathy. Our transcriptional analysis
27 highlights the significantly increased ACE2 expression in the hearts of cardiomyopathy patients,
28 suggesting that cardiomyopathy patients are at an increased risk of heart damage caused by
29 COVID-19.

30
31 Consistent with the RNA-seq finding, we found that COVID19 patients with pre-existing
32 cardiomyopathy show increased mortality risk than other populations, even compared with age-
33 matched populations with other cardiovascular conditions. While mechanistic research is
34 needed to establish the causal relationship between ACE2 up-regulation and the increased
35 mortality, our result identifies the cardiomyopathy patients as a high-risk group that needs extra
36 protection and care.

37
38 This project demonstrated public RNA-sequencing data as a valuable resource for biomedical
39 knowledge discovery. The existing RNA-seq data can be quickly repurposed to address
40 pressing problems, such as identifying the risk factor of COVID19. While this study focused on
41 analyzing the expression of ACE2, researchers can identify modulators for any genes or gene
42 signatures of interest using the GENEVA web portal at genevatool.org.

43
44

1 **Methods**

2 **Data preparation**

3 We downloaded the uniformly processed RNA-seq data from ARCHS4 website
4 (<https://amp.pharm.mssm.edu/archs4/download.html>) on August 03, 2020. The downloaded
5 data include gene-level count data of 286650 samples from 9124 datasets and sample-level
6 metadata. We transformed the gene count data into percentile rank data, which reduces the
7 influences of library size, batch effects, and extreme values^{41,42}. We downloaded study-level
8 metadata using the `entrez_search` and `entrez_summary` function from the `rentrez` library⁴³.

9

10 **Correlation analysis**

11 We first calculated the Pearson correlation between ACE2 and other genes using data from all
12 286,650 samples. To calculate intra-study correlation, we selected a subset of studies in which
13 the variance of ACE2 is greater than 10. We then calculated the Pearson correlation within each
14 study. Transcription factors are identified by selecting genes with the Gene Ontology term
15 “DNA-binding transcription factor activity (GO:0003700)”.

16

17 To identify pathways associated with ACE2, we used the pairwise correlations with ACE2 as a
18 signature. We used the signature to query the Gene Ontology Biological Process database⁴⁴.
19 The `fgsea` function from the `fgsea` library was used to calculate the enrichment score⁴⁵.

20

21 **Metadata embedding**

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

28

29 **GENEVA analysis**

30 For a given gene in a given dataset, we first calculated the variance of the gene (VARg). We
31 measure the overall heterogeneity of the samples by calculating the average variance of all
32 genes (VARm). We run a regression using the expression of the gene as the dependent
33 variable and the embedded metadata as independent variables (expression ~ first embed
34 dimension + second embed dimension). The regression coefficient (R^2) represents the
35 association between the expression of the gene and the embedded metadata. The product
36 between VARg and R^2 represents the variance of the gene explained by the embedded
37 metadata. The GENEVA score is defined as $\text{VARg} \times R^2 / \text{VARm}$.

38

39 To test the significance of the GENEVA scores, we shuffled the samples across all datasets.
40 We then calculated the GENEVA scores of all shuffled datasets to create a null distribution.
41 Given a GENEVA score G, its p-value is defined as the probability that the null distribution is
42 greater than G: p-value = $\text{Prob}(\text{null} > G)$. We adjust the p values for multiple testing using the
43 false discovery rate method.

44

1 **Meta-analysis**

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

7
8 To examine the ACE2 expression in different types of cardiomyopathy, we separated the
9 cardiomyopathy samples based on their subtype. We matched the cardiomyopathy samples
10 with healthy controls within the same study. We used unpaired T-tests to test the effect of each
11 cardiomyopathy type on ACE2 expression. Since data from multiple studies are available for
12 dilated cardiomyopathy (DCM), we used a mixed effect model to test the effect of DCM on
13 ACE2 expression: $ACE2 \text{ expression} \sim \text{study (random effect)} + \text{cardiomyopathy status (fixed}$
14 effect) .

15 16 **Analysis of electronic health records**

17 The UCSF COVID-19 Data Mart records the clinical information of COVID19 patients and
18 selected control patients using the Observational Medical Outcomes Partnership (OMOP) data
19 format. We identified COVID19 patients from the clinical data using the ICD10 code U07. 1. We
20 identified cardiomyopathy patients using the ICD10 codes I42 and I43, entered before their first
21 COVID19 diagnosis. We identified patients with other cardiovascular diseases using the
22 following ICD10 codes I00 - I99. We compared the cardiomyopathy patient with patients with
23 other cardiovascular diseases and patients without cardiovascular diseases. We tested if the
24 demographic and clinical variables are significantly different between the groups using single
25 variable logistic regressions ($\text{cardiomyopathy} \sim \text{clinical variable}$). We used the coxph function in
26 the survival R package to perform the survival analysis, controlling for the variables that are
27 significantly different in the logistic regressions. Patient survival time is defined as the time
28 between their first COVID19 diagnosis and their death date. Live patients are censored on the
29 last day of their encounter. In the COVID-19 negative cohort, cardiomyopathy patients are
30 defined as patients whose first cardiovascular-related diagnosis is cardiomyopathy (ICD10 code
31 I42 or I43). Patients with other cardiovascular diseases are defined as patients who have
32 cardiovascular diseases (ICD10 code I00 - I99), but do not have cardiomyopathy. Patient
33 survival time is defined as the time between their first cardiovascular disease and their death.
34 Live patients are censored on the last day of their encounter. To perform propensity score
35 matching, we first calculated the propensity score using logistic regression ($\text{cardiomyopathy} \sim$
36 $\text{age} + \text{race} + \text{gender} + \text{non-cardiovascular pre-existing conditions}$). We subsampled the non-
37 cardiomyopathy cohort so that the distribution of its propensity score matches with the
38 cardiomyopathy cohort.

39 40 41 **GENEVA web portal**

42 GENEVA webportal GENEVA web tool is an open-source application available under GNU
43 General Public License at <http://genevatool.org>. It is implemented in python web framework
44 Django. The source code for the tool is available in the public Git repository at

1 <https://github.com/NavchetanKaur/geneva-webtool>. The tool offers an intuitive interface and
2 user guide on the home page. Users can select either of the two options from “Gene Query” and
3 “Gene Signature Query” and query their gene of interest or set of upregulated and
4 downregulated genes of interest. The results are displayed in tabular form with calculated
5 GENEVA scores. GSE descriptions are further represented in plots and tables.

6

7 **References:**

8

9 1. World Health Organization. Coronavirus Disease (COVID-19) Situation Reports.

10 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (2020).

11 2. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is

12 Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

13 3. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of

14 COVID-19. *Cell* **181**, 1036-1045.e9 (2020).

15 4. Hassan, A. O. *et al.* A SARS-CoV-2 Infection Model in Mice Demonstrates Protection by

16 Neutralizing Antibodies. *Cell* **182**, 744-753.e4 (2020).

17 5. Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* **583**,

18 830–833 (2020).

19 6. Pan, L. *et al.* Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in

20 Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. *Am J Gastroenterol* **115**,

21 (2020).

22 7. Goyal, P. *et al.* Clinical Characteristics of Covid-19 in New York City. *New England Journal*

23 *of Medicine* **382**, 2372–2374 (2020).

24 8. Li, J.-W. *et al.* The impact of 2019 novel coronavirus on heart injury: A Systematic review

25 and Meta-analysis. *Progress in Cardiovascular Diseases* (2020)

26 doi:10.1016/j.pcad.2020.04.008.

27 9. Pinto, B. G. G. *et al.* ACE2 Expression Is Increased in the Lungs of Patients With

28 Comorbidities Associated With Severe COVID-19. *J. Infect. Dis.* **222**, 556–563 (2020).

29 10. Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue

- 1 gene regulation in humans. *Science* **348**, 648–660 (2015).
- 2 11. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells
3 together with innate immune genes. *Nature Medicine* **26**, 681–687 (2020).
- 4 12. Lachmann, A. *et al.* Massive mining of publicly available RNA-seq data from human and
5 mouse. *Nature Communications* **9**, 1366 (2018).
- 6 13. Wang, Z. *et al.* Extraction and analysis of signatures from the Gene Expression Omnibus by
7 the crowd. *Nature Communications* **7**, 12846 (2016).
- 8 14. Koeppen, K., Stanton, B. A. & Hampton, T. H. ScanGEO: parallel mining of high-throughput
9 gene expression data. *Bioinformatics* **33**, 3500–3501 (2017).
- 10 15. Feng, C. *et al.* GEM-TREND: a web tool for gene expression data mining toward relevant
11 network discovery. *BMC Genomics* **10**, 411 (2009).
- 12 16. Hadley, D. *et al.* Precision annotation of digital samples in NCBI's gene expression
13 omnibus. *Scientific Data* **4**, 170125 (2017).
- 14 17. Chang, J. T. *et al.* SIGNATURE: A workbench for gene expression signature analysis. *BMC*
15 *Bioinformatics* **12**, 443 (2011).
- 16 18. Williams, G. A searchable cross-platform gene expression database reveals connections
17 between drug treatments and disease. *BMC Genomics* **13**, 12 (2012).
- 18 19. Fujibuchi, W., Kiseleva, L., Taniguchi, T., Harada, H. & Horton, P. CellMontage: similar
19 expression profile search server. *Bioinformatics* **23**, 3103–3104 (2007).
- 20 20. Engreitz, J. M. *et al.* ProfileChaser: searching microarray repositories based on genome-
21 wide patterns of differential expression. *Bioinformatics* **27**, 3317–3318 (2011).
- 22 21. Zinman, G. E., Naiman, S., Kanfi, Y., Cohen, H. & Bar-Joseph, Z. ExpressionBlast: mining
23 large, unstructured expression databases. *Nat. Methods* **10**, 925–926 (2013).
- 24 22. Zhu, Q. *et al.* Targeted exploration and analysis of large cross-platform human
25 transcriptomic compendia. *Nature Methods* **12**, 211–214 (2015).
- 26 23. Bernardi, S. *et al.* ACE2 deficiency shifts energy metabolism towards glucose utilization.

- 1 *Metab. Clin. Exp.* **64**, 406–415 (2015).
- 2 24. Clarke, N. E., Belyaev, N. D., Lambert, D. W. & Turner, A. J. Epigenetic regulation of
3 angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress.
4 *Clin. Sci.* **126**, 507–516 (2014).
- 5 25. Kb, P., Kh, C., Vk, N., H, X. & E, L. The transcription factor HNF1 α induces expression of
6 angiotensin-converting enzyme 2 (ACE2) in pancreatic islets from evolutionarily conserved
7 promoter motifs. *Biochimica et biophysica acta* vol. 1829
8 <https://pubmed.ncbi.nlm.nih.gov/24100303/> (2013).
- 9 26. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
10 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
11 (*Methodological*) **57**, 289–300 (1995).
- 12 27. Husser, D. *et al.* Prevalence of clinically apparent hypertrophic cardiomyopathy in
13 Germany—An analysis of over 5 million patients. *PLoS One* **13**, (2018).
- 14 28. Atkins, J. L. *et al.* Preexisting Comorbidities Predicting COVID-19 and Mortality in the UK
15 Biobank Community Cohort. *J Gerontol A Biol Sci Med Sci* doi:10.1093/gerona/glaa183.
- 16 29. Guo, T. *et al.* Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus
17 Disease 2019 (COVID-19). *JAMA Cardiol* **5**, 811–818 (2020).
- 18 30. Shi, S. *et al.* Association of Cardiac Injury With Mortality in Hospitalized Patients With
19 COVID-19 in Wuhan, China. *JAMA Cardiol* **5**, 802–810 (2020).
- 20 31. Jefferies, J. L. & Towbin, J. A. Dilated cardiomyopathy. *The Lancet* **375**, 752–762 (2010).
- 21 32. Maron, B. J. & Maron, M. S. Hypertrophic cardiomyopathy. *The Lancet* **381**, 242–255
22 (2013).
- 23 33. Kushwaha, S. S., Fallon, J. T. & Fuster, V. Restrictive Cardiomyopathy. *New England*
24 *Journal of Medicine* **336**, 267–276 (1997).
- 25 34. Basso, C., Corrado, D., Marcus, F. I., Nava, A. & Thiene, G. Arrhythmogenic right
26 ventricular cardiomyopathy. *The Lancet* **373**, 1289–1300 (2009).

- 1 35. Ross, S. B. *et al.* A systematic review and meta-analysis of the prevalence of left ventricular
2 non-compaction in adults. *Eur Heart J* **41**, 1428–1436 (2020).
- 3 36. Bos, J. M. *et al.* Marked Up-Regulation of ACE2 in Hearts of Patients With Obstructive
4 Hypertrophic Cardiomyopathy: Implications for SARS-CoV-2–Mediated COVID-19. *Mayo*
5 *Clin Proc* **95**, 1354–1368 (2020).
- 6 37. Goulter, A. B., Goddard, M. J., Allen, J. C. & Clark, K. L. ACE2 gene expression is up-
7 regulated in the human failing heart. *BMC Medicine* **2**, 19 (2004).
- 8 38. Grifoni, A. *et al.* Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with
9 COVID-19 Disease and Unexposed Individuals. *Cell* **181**, 1489-1501.e15 (2020).
- 10 39. Jin, J.-M. *et al.* Gender Differences in Patients With COVID-19: Focus on Severity and
11 Mortality. *Front. Public Health* **8**, (2020).
- 12 40. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *New England*
13 *Journal of Medicine* **382**, 1708–1720 (2020).
- 14 41. Hu, Z., Glicksberg, B. S. & Butte, A. J. Robust prediction of clinical outcomes using
15 cytometry data. *Bioinformatics* **35**, 1197–1203 (2019).
- 16 42. Zwiener, I., Frisch, B. & Binder, H. Transforming RNA-Seq Data to Improve the
17 Performance of Prognostic Gene Signatures. *PLOS ONE* **9**, e85150 (2014).
- 18 43. Winter, D. J. rentrez: An R package for the NCBI eUtils API. *The R Journal* **9**, 520–526
19 (2017).
- 20 44. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29
21 (2000).
- 22 45. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*
23 060012 (2019) doi:10.1101/060012.
- 24

25 **Acknowledgment:**

1 We thank all researchers who have contributed RNA-seq datasets to the Gene Expression
2 Omnibus. We thank Dr. Alexander Lachmann, Dr. Avi Ma'ayan, and other co-authors for
3 creating ARCHS4, which enabled our study. We thank Dr. Douglas Arneson and Sanchita
4 Bhattacharya for helpful discussion. We thank the UCSF COVID-19 Data Mart for providing the
5 electronic health record data of COVID-19 patients. This work was partially supported by the
6 National Institute of Allergy and Infectious Diseases (Bioinformatics Support Contract
7 HHSN272201200028C). The content is solely the responsibility of the authors and does not
8 necessarily represent the official views of the National Institutes of Health. The analysis of the
9 electronic health record data is conducted with approval from the UCSF institutional review
10 board (IRB #: 20-31107).

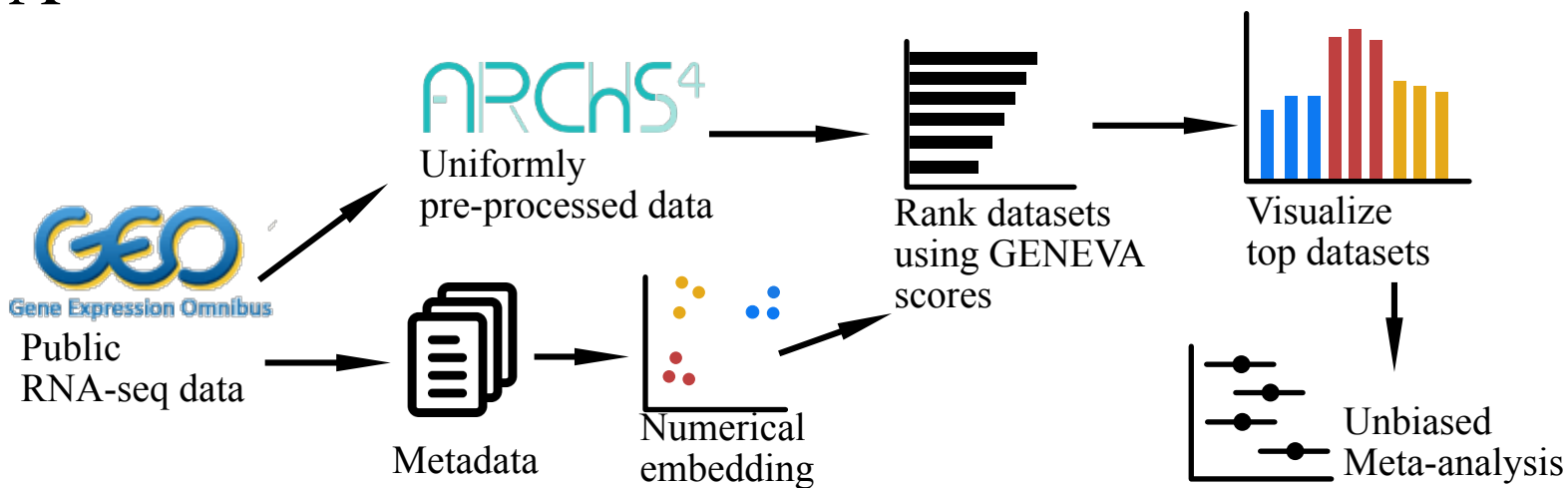
11

12 **Conflicts of Interest:**

13 Atul Butte is a co-founder and consultant to Personalis and NuMedii; consultant to Samsung,
14 Mango Tree Corporation, and in the recent past, 10x Genomics, Helix, Pathway Genomics, and
15 Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health,
16 Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech,
17 Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple,
18 Facebook, Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, CVS, Nuna
19 Health, Assay Depot, Vet24seven, Regeneron, Sanofi, Royalty Pharma, AstraZeneca, Moderna,
20 Biogen, Paraxel, and Sutro, and several other non-health related companies and mutual funds;
21 and has received honoraria and travel reimbursement for invited talks from Johnson and
22 Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum,
23 Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or
24 disease specific foundations and associations, and health systems. Atul Butte receives royalty
25 payments through Stanford University, for several patents and other disclosures licensed to
26 NuMedii and Personalis. Atul Butte's research has been funded by NIH, Northrup Grumman (as
27 the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson
28 Foundation, Leon Lowenstein Foundation, Intervallen Foundation, Priscilla Chan and Mark
29 Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of
30 Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and
31 Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

32

A



B

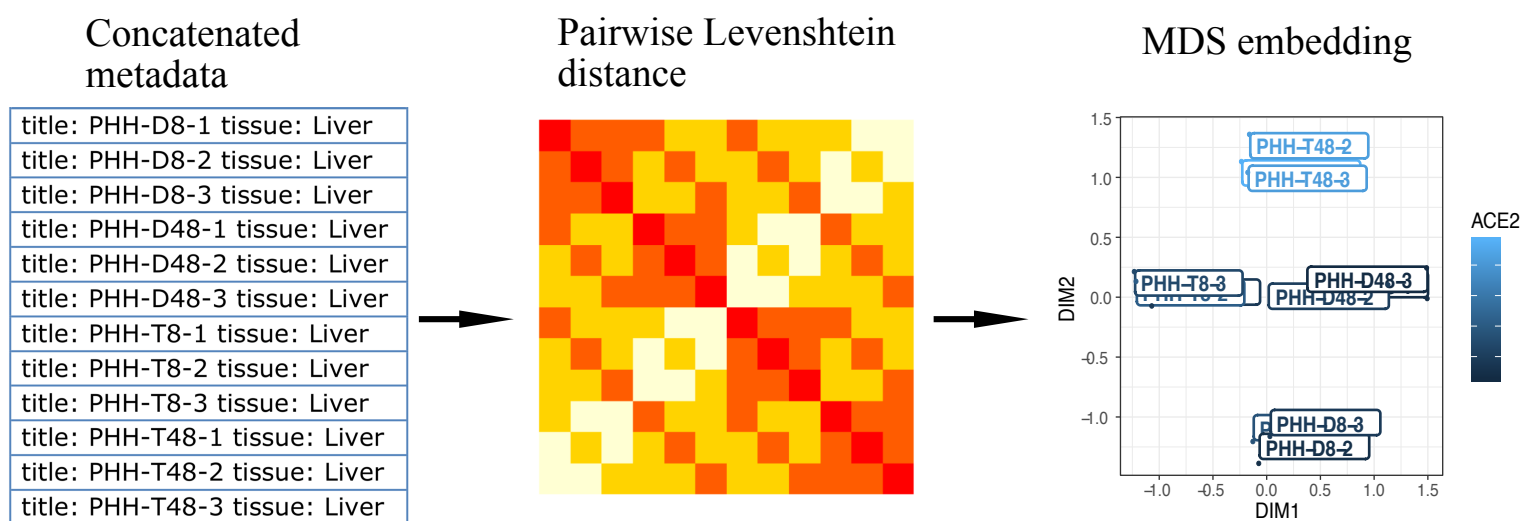


Figure 1: Mining RNA-seq data using Gene Expression Variance Analysis (GENEVA). (A) The workflow of GENEVA. (B) The procedure for embedding metadata into numerical vectors.

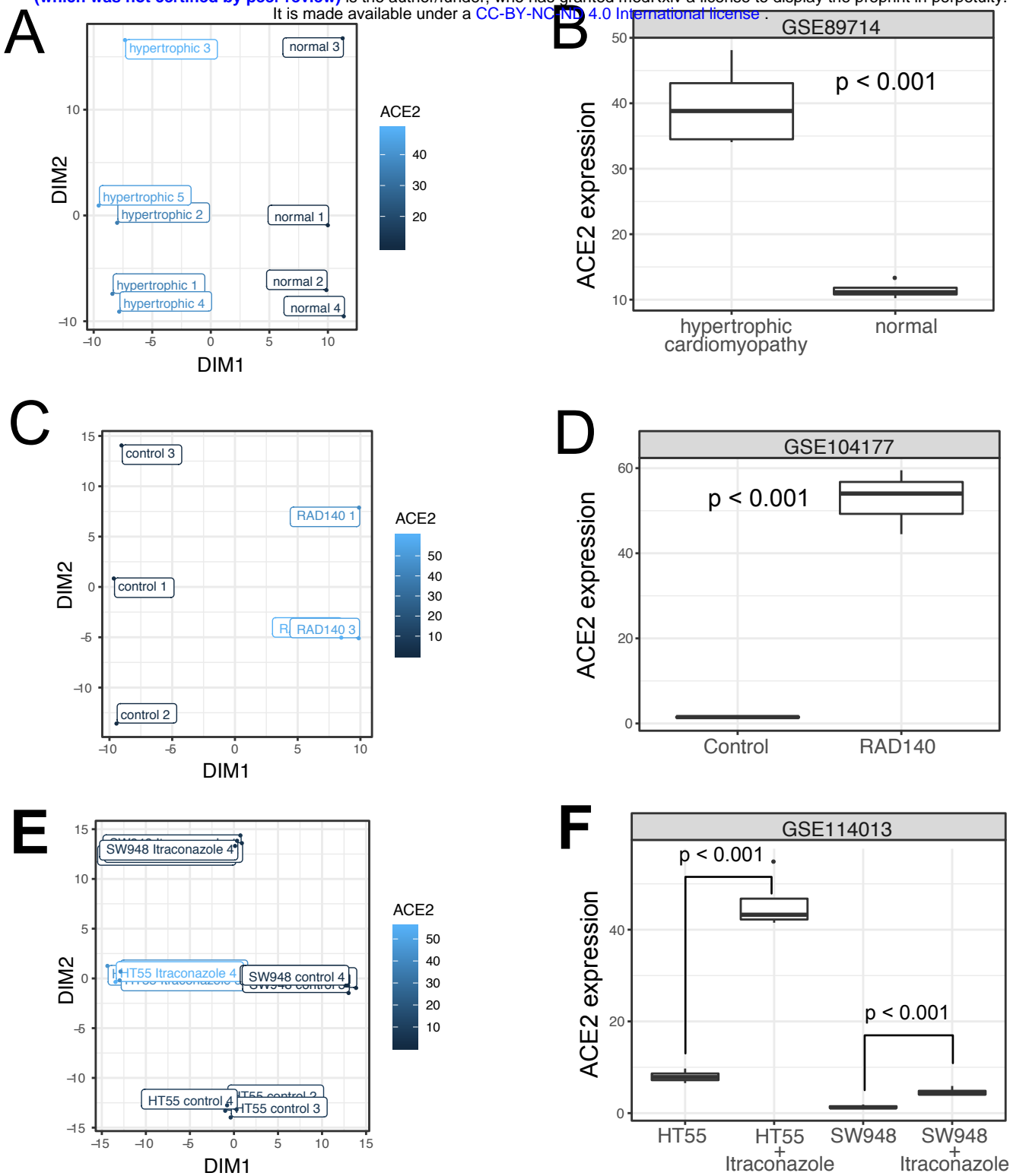


Figure 2: Highlighted conditions that modulate ACE2 expression. (A-B) Plots showing data from a study with GEO accession GSE89714. (A) A scatter plot showing the association between ACE2 expression and the embedded metadata. All metadata fields are concatenated for embedding. The plot only shows the sample title as labels. The labels are jittered to avoid perfect overlap. The color code represents the ACE2 expression level. (B) Box plot showing the ACE2 expression in normal hearts and hearts with hypertrophic cardiomyopathy. (C-D) Plots visualizing data from a study with GEO accession GSE104177, showing the ACE2 expression in breast cancer xenografts with or without RAD140 treatment. (E-F) Plots visualizing data from a study with GEO accession GSE114013, showing the ACE2 expression in two prostate cancer cell lines with or without Itraconazole treatment.

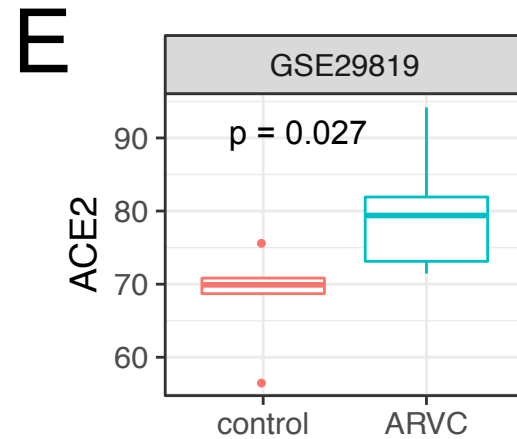
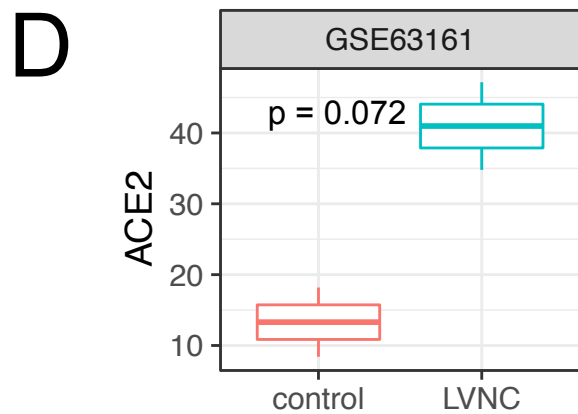
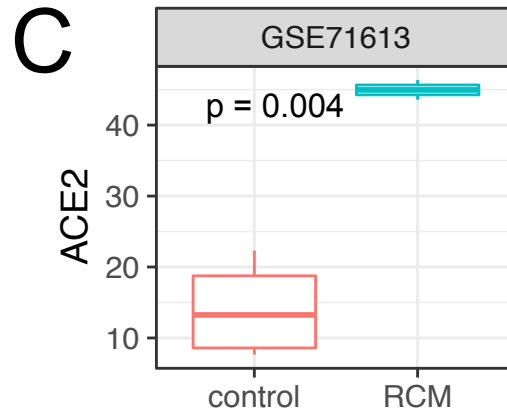
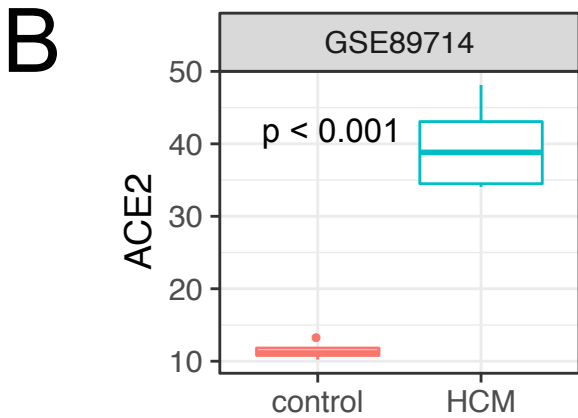
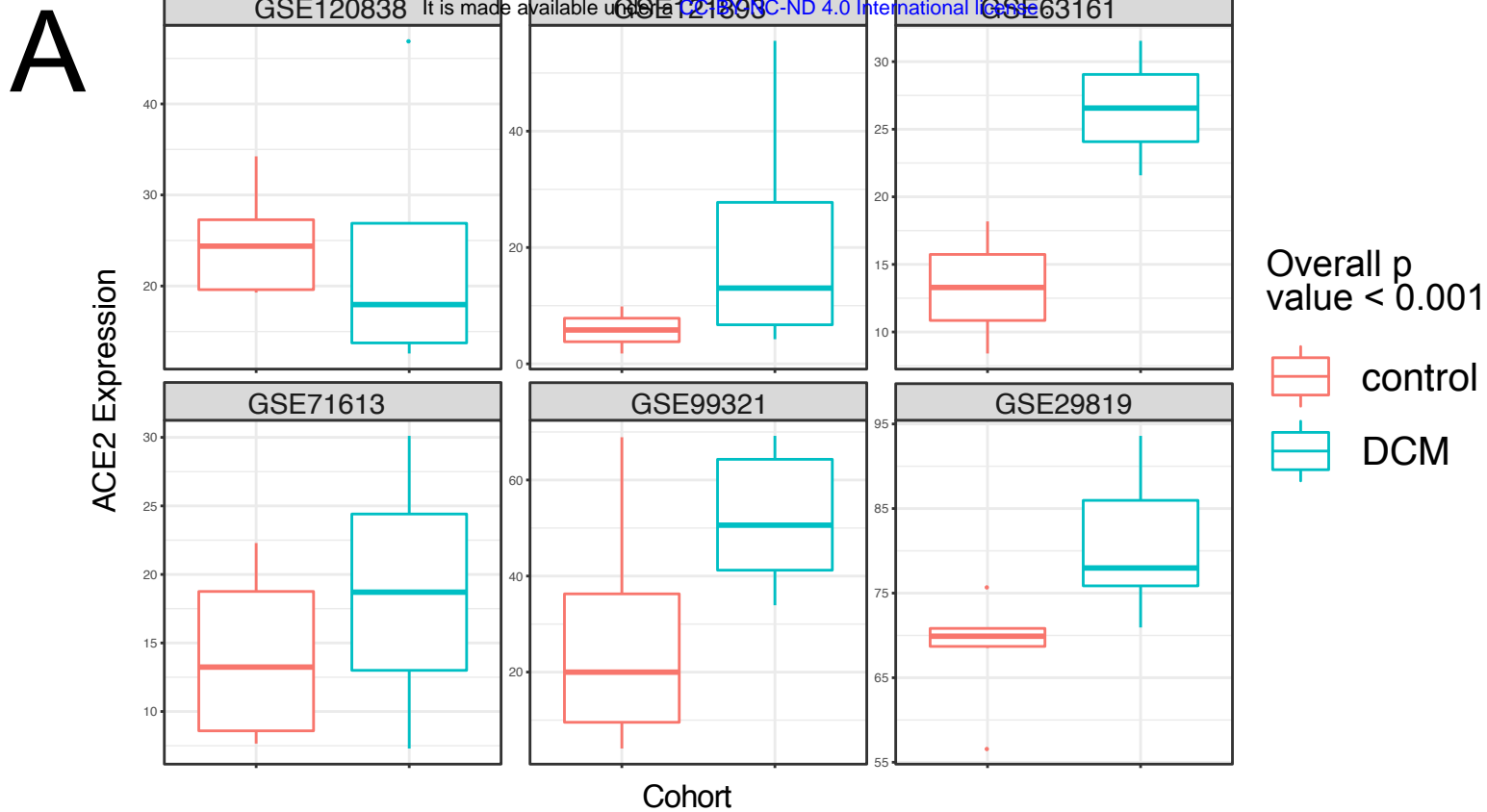


Figure 3: Meta-analysis confirms ACE2 upregulation in all major types of cardiomyopathy. (A) The ACE2 expression in normal hearts and hearts with DCM. Data are from five RNA-seq datasets and one microarray dataset. The overall p-value is calculated using a mixed model, with dataset as the random effect and DCM as the fixed effect. (B-E) The ACE2 expression in normal hearts and hearts with HCM, RCM, LVNC, and ARVC. p values in B-E are calculated using t-tests.

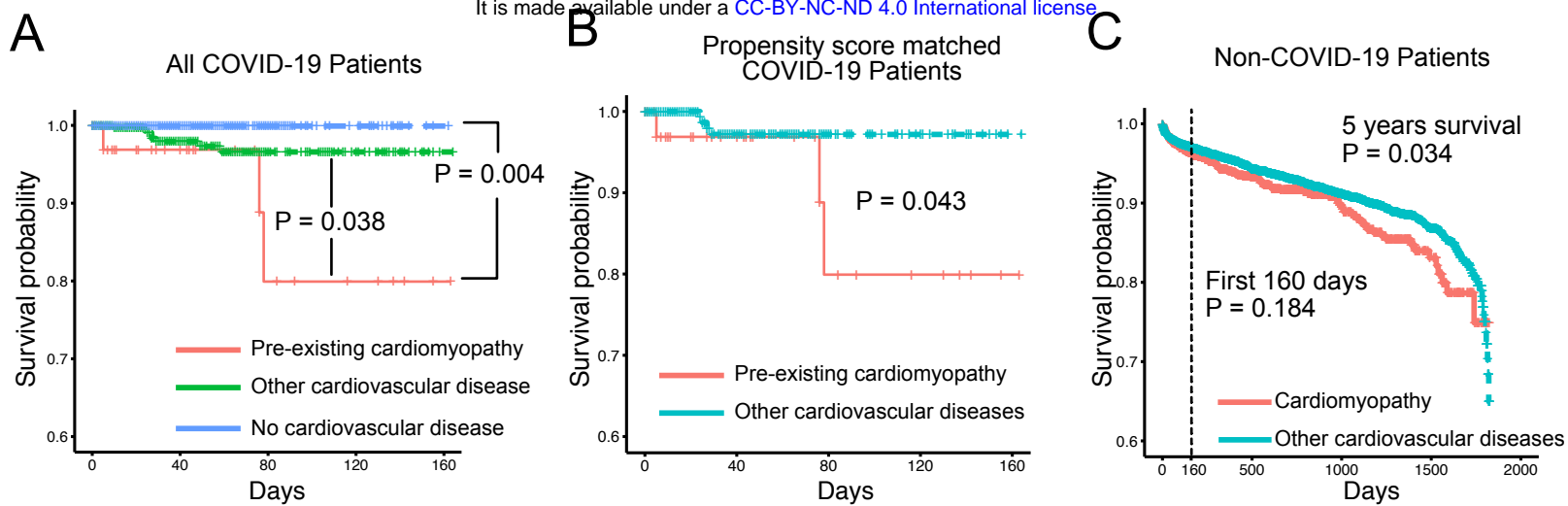


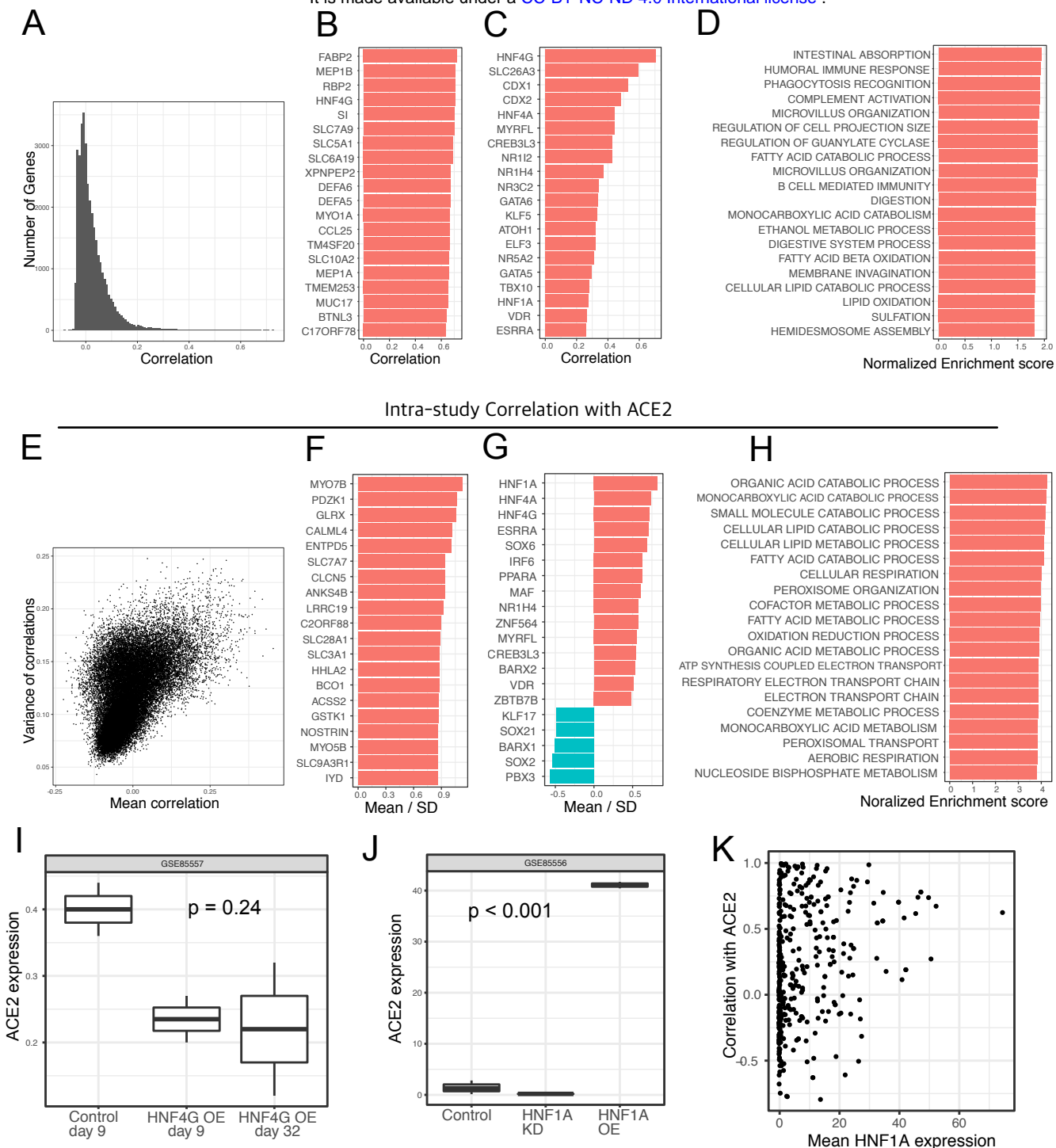
Figure 4: Patients with pre-existing cardiomyopathy show an increased mortality rate. (A) Kaplan Meier curve of COVID19 patients with pre-existing cardiomyopathy (N = 43), patients with other pre-existing cardiovascular diseases (N = 624), and patients without cardiovascular diseases (N = 3269). P values are from Cox proportional-hazards models, controlling for differences in the demographics and non-cardiovascular conditions between the groups. (B) Kaplan Meier curve of COVID19 patients with pre-existing cardiomyopathy (N = 43) and a cohort of propensity score-matched patients with other pre-existing cardiovascular diseases (N = 344). The P value is from a log-rank test. (C) Kaplan Meier curve of COVID-19 negative patients with pre-existing cardiomyopathy (N = 2250) and a cohort of propensity score-matched patients with other pre-existing cardiovascular diseases (N = 18000). P values are from a log-rank test.

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	dimethylxalyglycine	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

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

	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				

Table 2: Demographic and clinical information of COVID-19 patients.



Supplemental Figure 1: RNA-seq analysis reveals complex transcriptional networks of ACE2. We calculated the pairwise Pearson correlations between ACE2 and other genes using all data from 286650 RNA-seq samples (A-D). (A) Histogram of the pairwise correlations. (B) Top 30 genes with the highest correlation with ACE2. (C) Top 30 transcription factors with the highest correlation with ACE2. (D) Top 30 pathways correlated with ACE2 expression. We calculated the pairwise Pearson correlations between ACE2 and other genes using samples in individual datasets (E-H). (E) Scatter plot shows the mean and variance of the correlation between ACE2 and other genes across different datasets. (F) Top 30 genes with the highest standardized correlation with ACE2. The standardized correlation is defined as the average correlation divided by its standard deviation. (G) Top 30 transcription factors with the highest standardized correlation with ACE2. (H) Top 30 pathways with the highest standardized correlation with ACE2. (I) Box plot showing the expression of ACE2 in LNCaP/AR cells with or without HNF4G over expression. (I) Box plot showing the expression of ACE2 in LNCaP/AR cells with or without HNF4A over expression or knock down. (K) scatter plot shows the relationship between HNF4G expression level and its correlation with ACE2.

	pre-existing cardiomyopathy (N=43)	Other cardiovascular diseases (N=344)	Overall (N=387)
age			
Mean (SD)	58.6 (17.9)	57.9 (19.7)	58.0 (19.5)
Median [Min, Max]	60.0 [18.0, 89.0]	60.5 [0, 100]	60.0 [0, 100]
gender			
Female	10 (23.3%)	95 (27.6%)	105 (27.1%)
Male	33 (76.7%)	249 (72.4%)	282 (72.9%)
race			
Asian	12 (27.9%)	77 (22.4%)	89 (23.0%)
Black or African American	5 (11.6%)	33 (9.6%)	38 (9.8%)
MultiRace	1 (2.3%)	9 (2.6%)	10 (2.6%)
Native Hawaiian or Other Pacific Islander	3 (7.0%)	13 (3.8%)	16 (4.1%)
Unknown	9 (20.9%)	74 (21.5%)	83 (21.4%)
White	13 (30.2%)	138 (40.1%)	151 (39.0%)
hyperlipidemia			
Yes	27 (62.8%)	195 (56.7%)	222 (57.4%)
No	16 (37.2%)	149 (43.3%)	165 (42.6%)
diabetes			
Yes	27 (62.8%)	184 (53.5%)	211 (54.5%)
No	16 (37.2%)	160 (46.5%)	176 (45.5%)
cancer			
Yes	10 (23.3%)	81 (23.5%)	91 (23.5%)
No	33 (76.7%)	263 (76.5%)	296 (76.5%)
ventilator use			
Yes	24 (55.8%)	90 (26.2%)	114 (29.5%)
No	19 (44.2%)	254 (73.8%)	273 (70.5%)
respiratory failure			
Yes	12 (27.9%)	62 (18.0%)	74 (19.1%)
No	31 (72.1%)	282 (82.0%)	313 (80.9%)
chest pain			
Yes	27 (62.8%)	125 (36.3%)	152 (39.3%)
No	16 (37.2%)	219 (63.7%)	235 (60.7%)
death			
Yes	3 (7.0%)	4 (1.2%)	7 (1.8%)
No	40 (93.0%)	340 (98.8%)	380 (98.2%)

Table S1: Demographic and clinical information of COVID-19 patients with cardiomyopathy and a propensity score matched cohort of COVID-19 patients with other cardiovascular diseases

	Cardiomyopathy (N=2250)	Other cardiovascular diseases (N=18000)	Overall (N=20250)
age			
Mean (SD)	45.9 (23.3)	46.4 (24.1)	46.3 (24.0)
Median [Min, Max]	50.0 [0, 90.0]	51.0 [0, 91.0]	51.0 [0, 91.0]
gender			
FEMALE	897 (39.9%)	7238 (40.2%)	8135 (40.2%)
MALE	1352 (60.1%)	10758 (59.8%)	12110 (59.8%)
UNKNOWN	1 (0.0%)	4 (0.0%)	5 (0.0%)
race			
American Indian or Alaska Native	9 (0.4%)	73 (0.4%)	82 (0.4%)
Asian	200 (8.9%)	1526 (8.5%)	1726 (8.5%)
Black or African American	190 (8.4%)	1539 (8.6%)	1729 (8.5%)
MultiRace	122 (5.4%)	1018 (5.7%)	1140 (5.6%)
Native Hawaiian or Other Pacific Islander	27 (1.2%)	195 (1.1%)	222 (1.1%)
Other Race	382 (17.0%)	3172 (17.6%)	3554 (17.6%)
Unknown	335 (14.9%)	2449 (13.6%)	2784 (13.7%)
White	985 (43.8%)	8028 (44.6%)	9013 (44.5%)
hyperlipidemia			
Yes	368 (16.4%)	2945 (16.4%)	3313 (16.4%)
No	1882 (83.6%)	15055 (83.6%)	16937 (83.6%)
diabetes			
Yes	370 (16.4%)	2874 (16.0%)	3244 (16.0%)
No	1880 (83.6%)	15126 (84.0%)	17006 (84.0%)
cancer			
Yes	306 (13.6%)	2430 (13.5%)	2736 (13.5%)
No	1944 (86.4%)	15570 (86.5%)	17514 (86.5%)
death			
Yes	123 (5.5%)	807 (4.5%)	930 (4.6%)
No	2127 (94.5%)	17193 (95.5%)	19320 (95.4%)

Table S2: Demographic and clinical information of Non-COVID-19 patients with cardiomyopathy and a propensity score matched cohort of patients with other cardiovascular diseases