

COVID-19 mortality is associated with pre-existing impaired innate immunity in health conditions

Matthew E Lee¹, Yung Chang², Navid Ahmadinejad¹, Crista E Johnson-Agbakwu³, Celeste Bailey³, Li Liu^{Corresp. 4}

¹ College of Health Solutions, Arizona State University, Phoenix, AZ, United States

² School of Life Sciences, Arizona State University, Tempe, AZ, United States

³ Valleywise Health Medical Center, Phoenix, AZ, United States

⁴ College of Health Solutions, Arizona State University, Pheonix, AZ, United States

Corresponding Author: Li Liu
Email address: liliu@asu.edu

COVID-19 can be life-threatening to individuals with chronic diseases. To prevent severe outcomes, it is critical that we comprehend pre-existing molecular abnormalities found in common health conditions that predispose patients to poor prognoses. In this study, we focused on fourteen pre-existing health conditions for which age-and-sex-adjusted hazard ratios of COVID-19 mortality have been documented. For each health condition, we analyzed transcriptomics data archived before the pandemic to identify differentially expressed genes between affected individuals and unaffected individuals. We then tested if fold changes of these genes in pre-existing conditions were correlated with hazard ratios of COVID-19 mortality to discover molecular risk factors. We found that upregulated expression of 70 genes and downregulated expression of 181 genes in pre-existing health conditions were correlated with increased risk of COVID-19 related death. These genes were significantly enriched with endoplasmic reticulum and mitochondria function, proinflammatory reaction, interferon production, and programmed cell death that participate in viral replication and innate immune responses to viral infections. These results collectively suggest that impaired innate immunity in pre-existing health conditions is associated with increased hazard of COVID-19 mortality. The discovered molecular risk factors are potential prognostic biomarkers and targets for therapeutic intervention.

Introduction

COVID-19 was declared a global pandemic by the World Health Organization (WHO) as of March 11, 2020 [1]. The outbreak has seen over 264 million cases and over 5.2 million deaths as of December 2021, and these numbers are still rising [2]. The clinical spectrum of illness ranges from asymptomatic or mild infection to severe pneumonia and death. Well documented risk factors for COVID-19 severity and fatality include age, sex, race, social determinants, and pre-existing health conditions [3-9]. The fatality rate stratified by age groups steadily increases from 0.2 per 100,000 patients in children (<14 years old) to 1797.8 in elders (≥85 years) [4]. Age-adjusted fatality rate in men is 1.4 times higher than in females [5]. More importantly, people with pre-existing health conditions are susceptible to extreme outcomes. While diagnosis rates of COVID-19 are nearly equal for patients with and without comorbidities, those with a comorbidity account for 83.29% of COVID-19 deaths [6]. Not all comorbidities have the same impact on COVID-19 prognosis. By linking primary care records of >17 million adults to 10,926 COVID-19 related deaths, the OpenSAFELY project estimated age-and-sex-adjusted hazard ratios (HR) of COVID-19 related deaths for 23 groups of pre-existing health conditions [7]. It reported that patients with organ transplant were at the highest risk (HR=6.00) and those with high blood pressure were at the lowest risk (HR=1.09). However, the biological mechanisms underlying such distinct impacts are largely unknown, which impedes the development of effective interventions to improve clinical outcomes.

Independently, mechanistic studies of COVID-19 pathogenesis have found that dysregulated biological processes in hosts play an important role in disease severity. It is widely recognized that COVID-19 patients associated with cytokine storms have high mortality [10]. Serum levels of inflammatory factors, such as interleukins and C-reactive protein, have been proposed as prognostic markers [11-13]. Early metabolic responses to infections also show signature differences between patients with favorable outcomes and patients with unfavorable outcomes [14, 15]. Given that comorbidities and molecular dysregulations influence COVID-19 severity, we must question what molecular abnormalities associated with pre-existing conditions predispose COVID-19 patients to poor prognosis and to what extent.

To answer this question, an intuitive approach would be to examine molecular profiles of COVID-19 patients before SARS-CoV-2 infection and correlate with prognoses after infection. However, this strategy is not feasible because before-infection samples of COVID-19 patients are rarely collected. To circumvent this obstacle, we propose to link molecular dysregulations in pre-existing conditions to COVID-19 prognosis at the health condition level instead of the individual level, making use of summary statistics from epidemiology studies and eliminating the need for individual specific data.

Our strategy is based on the fact that the COVID-19 mortality rate varies with pre-existing conditions. We hypothesize that such variations are associated with molecular dysregulations frequently observed in multiple health conditions. To test this hypothesis, we need quantitative data of COVID-19 mortality rates and molecular profiles obtained

from people with various health conditions, which fortunately are readily available. Specifically, the OpenSAFELY study has published the HRs of COVID-19 mortality for 23 groups of pre-existing conditions [7]. For each of these conditions, molecular profiles of affected and unaffected individuals can be found in public repositories, such as the Gene Expression Omnibus (GEO) database [16]. It is widely acknowledged that patient gene expression profiles reflect the underlying pathological processes [17]. While different diseases target different tissues and organs, transcriptomes of peripheral blood cells are informative about systematic changes of a person's overall health [18, 19]. Therefore, we chose to examine peripheral blood transcriptomes in this study. By correlating transcriptional dysregulations with HRs of COVID-19 mortality, we discovered molecular risk factors that predispose COVID-19 patients to severe outcomes. We further analyzed functional relationships of these risk genes, which converged onto impaired innate immunity as a systemic mechanism that weakens host defense against SARS-CoV-2 infection.

While we focused on gene expression risk factors in this study, our analytic approach is applicable to other omics-level profiles, such as epigenetic and metabolomic data. Integration of these discoveries will allow for better prediction of severe outcomes of COVID-19 and inform the development of preventative measures to reduce fatality. Furthermore, the long-term sequelae of COVID-19 survivors are currently unknown. A greater apprehension of the disease mechanisms in the context of comorbidities will serve for future evaluation of the health impact of COVID-19 on patients with chronic diseases.

Materials & Methods

Data sets

The OpenSAFELY project reported age-sex-adjusted HRs of COVID-19 related deaths for 23 groups of pre-existing health conditions. We excluded the two cancer groups (solid tumors and hematological malignancies) due to the extremely high heterogeneity of cancers [20]. For each remaining health condition, we searched the GEO database [16] to identify transcriptomics studies involving affected individuals (cases) and unaffected individuals (controls, Fig. 1A). We limited our queries to peripheral blood samples as a *modus operandi* of removing confounders related to different tissue types and encapsulating disease characteristics at a systemic level. We further limited our query to microarray-based transcriptomic profiles to reduce technical variance. If multiple data sets were available for a health condition, we chose the one with the largest sample size. We downloaded the normalized gene expression values.

Identify dysregulated gene expression in pre-existing health conditions

Given a transcriptomic data set, we used the Student's t-test to compare expression levels of each gene in cases versus controls (Fig. 1B). Specifically, we used the `t.test()` function in R and set the following parameters (`alternative="two.sided"`, `paired=F`). Aiming to be inclusive at this step, we considered genes with nominal P value <0.05 to be differentially expressed. If multiple probes on the microarray represented the same gene, we kept the one with the lowest P value and removed the other ones to avoid redundancy. For a differentially expressed gene (DEG), we computed the fold change

(FC) as the ratio of the mean expression level in cases over controls. For a non-DEG, we set the FC to one. If a gene was differentially expressed in at least four health conditions, it was designated a recurrent DEG (rDEG).

Identify molecular risk factors of COVID-19 mortality

For each rDEG, we tested if its FCs in pre-existing conditions were correlated with HRs of COVID-19 mortality using the Pearson correlation test (Fig. 1C). Specifically, we used the `cor.test()` function in R and set the following parameters (`method="pearson", use="pairwise.complete.obs"`). We corrected for multiple comparisons by converting nominal P values to false discovery rates (FDRs) using the Benjamini-Hochberg method [21], as implemented in the `p.adjust()` function in R. $FDR < 0.05$ indicated significant molecular risk factors. $FDR > 0.05$ but a nominal P value < 0.01 indicated marginal risk factors. The positive or negative sign of a Pearson correlation coefficient (PCC) indicated that upregulated gene expression or downregulated gene expression in pre-existing conditions increased the risk of COVID-19 mortality, respectively.

Functional categorization and analysis

We classified molecular risk factors into overlapping gene sets based on annotations of biological processes in the Gene Ontology database and pathways in the KEGG database. For each gene set, we tested if it was overrepresented using the Fisher's exact test. We corrected for multiple comparisons by converting nominal P values to FDRs. We built association networks of enriched gene sets ($FDR < 0.05$) and examined their relationships using the `cnetplot()` function from the R/ `enrichplot` package [22] (Fig. 1D).

Experimental validations

To validate if the identified molecular risk factors at the gene level reflected the molecular changes at the protein level observed in SARS-CoV-2 infections, we used the published proteomic data set from an in vitro infection experiment [23]. In this experiment, human Caco-2 cells were infected with SARS-CoV-2 viruses. Infected cells and mock controls, each with three replicates, were cultured for 24 hours. Quantifications of 6,381 proteins were obtained and compared between the infected cells and the mock control cells with two-group t tests. We obtained the normalized protein quantification data and the pre-computed t-test p values from the Supplementary Table 2 in the published study [23]. Given a risk factor gene, we searched the proteomics data set for the matching protein based on the gene symbol. To validate a risk factor gene, we required that its transcriptional dysregulation pattern was consistent with the dysregulated protein expression pattern (Fig. 1E). Specifically, if upregulation of the risk factor gene was associated with an increased COVID-19 mortality rate, the matching protein shall also be upregulated in SARS-CoV-2 infected cells as compared to mock control cells. If downregulation of the risk factor gene was associated with an increased COVID-19 mortality rate, we required the matching protein also to be downregulated in SARS-CoV-2 infected cells as compared to mock control cells.

R source codes used in this study are available at https://github.com/liliulab/COVID19_Mortality_Association/.

Results

DEGs in pre-existing health conditions

Our search of the GEO database found qualified transcriptomic data for fourteen health conditions. For each health condition, we identified DEGs with Student t-test $P < 0.05$. This lenient cutoff allowed us to be as inclusive as possible at this step. On average, each health condition was associated with 5,777 DEGs (range 1215 to 17,688). Most of the DEGs were downregulated in cases as compared to controls (mean FCs range from 0.003 to 0.160). Among a total of 25,552 genes analyzed, we found 11,930 rDEGs, that is, those that were differentially expressed in at least four health conditions. Table 1 presents the summary statistics of DEGs.

Pre-existing expression dysregulations increase COVID-19 death risks

For each rDEG, we tested if its FCs in different health conditions were correlated with HRs for COVID-19 mortality. For health conditions where this gene was not differentially expressed, we set the FCs to 1 and included them in the correlation test, as well. Among a total of 11,930 DEGs, we found no significant molecular risk factor that passed the stringent $FDR < 0.05$ threshold. However, 231 genes passed the Pearson correlation test $P < 0.01$ threshold and were considered as marginal molecular risk factors. Among them, upregulated expression of 70 genes and downregulated expression of 181 genes increased risk of COVID-19 related death (Supplementary Table 1).

The *RPS28* gene had the most significant correlation P value (0.0003). Its FCs in pre-existing conditions were positively correlated with HRs of COVID-19 mortality (PCC= 0.83, Fig. 2A). *RPS28* encodes a component of the 40S subunit of the ribosome where a cell synthesizes proteins. It was differentially expressed in six health conditions, including rheumatoid arthritis, chronic obstructive pulmonary disease, alcoholic hepatitis, multiple sclerosis, HIV, and chronic kidney disease. As its FC increased from 0.98 to 1.14, the HR of COVID-19 mortality increased from 1.30 to 3.48. Furthermore, the list of molecular risk factors contained seven additional genes that encode ribosomal components (*RPLP1*, *RPLP2*, *RPL13*, *RPL23A*, *RPL30*, *RPL38*, and *RPS11*). Except for *RPL30*, upregulation of these genes consistently increased the HR of COVID-19 mortality (P range= 0.001 to 0.009, PCC range= 0.71 to 0.78, Fig. 2B). This is in accordance with the positive viral infection-regulating roles of ribosomal proteins [24]. Notably, three ribosomal proteins in our list are required for early virus accumulation [25], though mechanistic studies in SARS-CoV-2 are still lacking.

The *POLR3GL* gene showed the most significant negative correlation (P= 0.0008, PCC= -0.81, Fig. 2C). *POLR3GL* encodes a subunit of RNA polymerase III that catalyzes the transcription of DNA into RNA. It induces production of interferon (IFN- α/β) to inhibit virus replication [26, 27]. Consistent with this function, pre-existing downregulation of *POLR3GL* in eight health conditions (Alzheimer's disease, chronic kidney disease, alcoholic hepatitis, chronic obstructive pulmonary disease, HIV, multiple sclerosis, rheumatoid arthritis, and type-2 diabetes) increased the risk of COVID-19 related death. Furthermore, the list of risk factors contained two additional genes, *LSM14A* and *IFRD1*,

that regulate interferon signaling. For these genes, downregulation increased HRs of COVID-19 mortality ($P = 0.003$ and 0.005 , $PCC = -0.74$ and -0.71 , respectively, Fig. 2D). Interestingly, we did not find interferons as molecular risk factors, presumably due to their transient expression profiles.

Functional groups enriched with risk factors

We classified the list of marginal molecular risk factors into functional gene sets based on Gene Ontology and KEGG annotations and then performed enrichment analysis. At $FDR < 0.05$, these molecular risk factors were significantly enriched in ten biological processes and three pathways (Table 2). Most of these gene sets were related to viral transcription, mRNA processing and metabolism, protein synthesis, and endoplasmic reticulum (ER) function.

We then built an association network and examined the functional relationships of the enriched gene sets and the molecular risk factors. We observed three clusters (Fig. 3A). Each cluster is composed of highly interconnected gene sets. Crosstalk between clusters is indicated with individual genes bridging these clusters.

The first cluster consisted of gene sets involved in viral transcription, translational initiation, mRNA catabolic process, and proteins targeting ER. These gene sets shared eight common risk factor genes encoding ribosomal components that function in ER. Except *RPL30*, upregulation of all these genes correlated with increased HRs of COVID-19 mortality. Conversely, for most of the other genes (7 out of 9) in this cluster,

downregulation correlated increased HRs of COVID-19 mortality, including two genes (*SEC62* and *SGTB*) that target ER and participate in degradation of misfolded proteins. Therefore, pre-existing abnormal functions in ER, specifically upregulated protein synthesis and downregulated degradation of unfolded protein, were associated with high risk of COVID-19 death.

The second cluster consisted of a single gene set involved in RNA splicing. For almost all genes (11 out of 14) in this cluster, downregulation was associated with increased HRs of COVID-19 mortality. In COVID-19 patients, host RNA splicing was significantly disrupted by SARS-CoV-2 [28]. Our observation suggests that pre-existing downregulation of RNA splicing genes can potentially aggravate such disruptions.

The third cluster consisted of a single gene set participating in the interleukin-12-mediated signaling pathway. Noticeably, one of the risk factor genes in this cluster, *PPIA*, has been shown to act as a potential mediator between human SARS coronavirus nucleoprotein and BSG/CD147 during the process of invasion of host cells by the virus [29]. Consistent with this previous study, we found that pre-existing upregulation of this gene increased COVID-19 mortality risk in nine common health conditions (Fig. 3B). Interestingly, interleukin-12, a pro-inflammatory cytokine (*IL12A* and *IL12B* genes) was not a molecular risk factor. It was dysregulated in two health conditions but did not meet the criterion of being a rDEG, which required dysregulation in at least four health conditions. Its receptor *IL12RB1* was a near miss, showing a positive correlation between FC and HR in five health conditions ($P=0.079$, $PCC=0.48$).

229

230 Crosstalk between the second and third clusters is via *HNRNPA2B1*, which binds
231 heterogeneous nuclear RNA (hnRNA) and subsequently induces IFN- α/β production to
232 inhibit virus replication.

233

234 Two other hnRNA binding proteins, *HNRNPH3* and *HNRNPDL*, were also molecular risk
235 factors. For all three genes, pre-existing downregulation increased COVID-19 mortality
236 risk, presumably by blocking IFN- α/β production, which compromises innate immunity
237 (Fig. 3C).

238

239 **Dysregulated cell death and mitochondrial functions**

240 The innate immune system is intrinsically connected with programmed cell death [30] and
241 mitochondrial functions [31]. In accordance with this, our list of molecular risk factors
242 contained 8 genes involved in apoptosis, 7 genes involved in autophagy, and 20 genes
243 involved in mitochondrial function (Supplementary Table 2). Although these gene sets did
244 not pass the stringent threshold of $FDR < 0.05$ in the enrichment analysis, they were
245 overrepresented in several biological processes with borderline nominal P values,
246 including “mitochondrial RNA metabolism” ($P=0.001$), “regulation of apoptotic signaling
247 pathway” ($P=0.05$), “mitochondrial organization” ($P=0.04$), “autophagy of mitochondrion”
248 ($P=0.08$), and “autophagosome assembly” ($P=0.08$).

249

250 To examine how these dysregulated processes correlated to COVID-19 mortality, we built
251 another association network (Fig. 4A). Autophagy and apoptosis are two mechanisms of

programmed cell death that inhibit each other [32]. Our association network contains six genes that co-regulate these processes. In normal conditions, these genes keep autophagy and apoptosis in balance. Specifically, *BNIP3*, *FBXW7*, *RAB7A*, *STX17*, and *UBQLN2* induce autophagy and suppress apoptosis, which is counteracted by *FLCN* [33, 34]. Pre-existing dysregulations of these genes converged to a common pattern, i.e., suppressed autophagy and escalated apoptosis jointly increased the risk of COVID-19 related death (Fig. 4B). Furthermore, downregulation of *BNIP3* and *FBXW7* suggest compromised mitochondrial organization and reduced mitochondrial autophagy. As disrupted mitochondrial functions, suppressed autophagy, and escalated apoptosis are commonly found in COVID-19 patients [35-37], our results imply if such dysregulations are present prior to SARS-CoV-2 infection, patients are prone to poor prognosis.

Validation in SARS-CoV2 infected samples

We validated the identified risk factor genes using a published proteomics data set from an in vitro study [23]. In this experiment, human Caco-2 cells were infected with SARS-CoV-2 viruses. Infected cells and mock controls, each with three replicates, were cultured for 24 hours. Quantifications of 6,381 proteins were obtained. We cross-referenced these proteins with the molecular risk factors based on gene symbols and found 105 gene-protein pairs. For each pair, we first examined if the protein had significantly different expression levels in infected cells compared to mock control cells using the published pre-computed t-test P value <0.05 as the threshold. We then examined if the protein dysregulation pattern was consistent with the dysregulation patterns of the risk factor genes, which would exacerbate the pre-existing aberrations and lead to poor prognosis.

275

276 We found that 11 proteins were significantly upregulated in the infected cells compared
277 to mock control cells. Among them, three (*RPS28*, *RPLP2*, and *PDE12*) had matching
278 risk factor genes showing consistent patterns, i.e., transcriptional upregulations increased
279 the COVID-19 mortality risk (Fig. 5A). At 24 hours after infection, the expression level of
280 these proteins was significantly higher than that in mock controls by 9.5% to 12.2% (P
281 values range 0.00027 to 0.019).

282

283 Similarly, we found that 11 proteins were significantly downregulated in the infected cells
284 compared to mock control cells. Among them, six (*ADK*, *DHFR*, *METAP2*, *PIN4*, *SC5D*,
285 and *TMEM263*) had matching risk factor genes showing consistent patterns, i.e.,
286 transcriptional downregulations increased the COVID-19 mortality risk (Fig. 5B). At 24
287 hours after infection, the expression level of the corresponding protein was significantly
288 lower than that in mock controls by 5.3% to 28.6% (P values range 0.0004 to 0.04).

289

290 In summary, we validated 9 molecular risk factor genes by showing the concordance
291 between transcriptional dysregulations caused by pre-existing medical conditions and the
292 protein dysregulations caused by SARS-CoV-2 infection. The joint effects plausibly
293 modulated the clinical outcomes of COVID-19 patients.

294

295 Discussion

296

297 The drastically different disease progression and prognosis among COVID-19 patients

with pre-existing health conditions challenge clinical management of this life-threatening disease. In this study, we integrated publicly available transcriptomics data of common health conditions and COVID-19 epidemiology data to study the molecular mechanisms underlying this complex problem. Our analyses revealed that pre-existing transcriptional dysregulations frequently observed in multiple health conditions increased risk of severe COVID-19 outcomes, plausibly via impairing host innate immunity, as discussed below.

Innate immunity is an integral part of the body's defense system which responds to invading pathogens, as well as damage caused by chronic health conditions [38]. Abnormalities of many biological processes may alter innate immune responses, which in turn reduces host defense against SARS-CoV-2 infection [39].

ER and mitochondria are crucial organelles that lie at the crossroad between host physiological functions and viral infection. On the one hand, various chronic disorders interfere with ER homeostasis and mitochondrial dynamics, giving rise to chronic inflammation that subsequently activates the body's innate immune system [40, 41]. On the other hand, invading viruses hijack ER for entry into host cells and assembly of viral genomes [42]. Mitochondrial dynamics are subverted to benefit the intracellular survival of viruses [42]. Emerging evidence suggests that coronavirus infection, including SARS and COVID-19, triggers ER stress and viral replication within mitochondrial structures [43-45]. If these organelles are already compromised prior to infection, further disruptions of their crucial functions by COVID-19 will likely lead to severe outcomes. Our study supports this hypothesis as we found genes targeting ER and mitochondria were enriched

in the list of molecular risk factors. These genes include *SEC62* and *SGTB* that degrade unfolded proteins in ER; *UBE2J2*, *COPS5*, *MBTPS2*, and *PPIA* that respond to unfolded proteins; and 20 genes that participate in various aspects of mitochondrial functions.

Innate immune responses to virus-infected cells include autophagy to isolate and clear infected viruses or apoptosis and necroptosis to eliminate infected cells. Viruses have evolved mechanisms to inhibit these surveillance processes so that infected cells would not be cleared efficiently, and the viruses could spread [46]. Suppressed autophagy and escalated apoptosis have been reported in severe COVID-19 cases [36, 37]. In this study, we discovered transcriptional dysregulations that predisposed patients to such distortion. Interestingly, two of these genes, *FBXW7* and *BNIP3*, determine the cell fate via autophagy of mitochondria and apoptosis, implying the important role of mitochondrial functions in COVID-19 severity.

Both chronic inflammation in pre-existing health conditions and acute inflammation in pathogen infections regulate interferon production. Reduced antiviral interferon response has been associated with excessive proinflammatory responses in COVID-19 and emerges as a clinical determinant of COVID-19 severity [47, 48]. Our results also imply that interferon production and signaling are suppressed in several health conditions via downregulation of *POLR3GL*, *LSM14A*, *IFRD1*, and three hnRNA binding proteins (*HNRNPA2B1*, *HNRNPH3*, and *HNRNPDL*). Dysregulation of these genes disrupts type I interferon signaling pathways. Furthermore, we found upregulation of *PPIA* and *MTAP* genes that activate IL12-mediated signal pathway in which proinflammatory cytokines

IL12 mediate the innate immune response [49]. However, we did not find any cytokines, either proinflammatory or anti-inflammatory, to be molecular risk factors. Although several cytokines were dysregulated in the health conditions we examined, there was no uniform correlation with HR of COVID-19 mortality.

Limitations of our study include the lack of individual risk factors passing a stringent statistical threshold and no consideration of multivariate effects. Although our analysis identified marginal molecular risk factors passing the nominal P value cutoff, none had a significant FDR after correction for multiple comparisons, which disqualified them as prognostic markers. However, analysis using the aggregation of these risk factor genes discovered significantly enriched biological processes, with the best $FDR < 10^{-4}$ (Table 2). Therefore, we are confident that chronic ER stress and immune dysregulation in pre-existing health conditions increased risk of COVID-19 mortality. Our analyses were based on univariate models, in which we examined the expression levels of each gene separately. Because multiple genes are dysregulated concurrently and a combination of them contributes to COVID-19 prognosis, a more realistic model should consider their combined effect. However, because the transcriptomics data were derived from individual patients and HRs of COVID-19 mortality were from summary statistics of an epidemiology study, we chose to use univariate models that are more straightforward to interpret. Lastly, our analyses focused on associating pre-existing transcriptional dysregulations to the ultimate outcome of COVID-19 assessed on mortality. However, dynamic changes of molecular and cellular processes are expected during the clinical course of COVID-19, which are potentially influenced by pre-existing conditions, as well.

367

368 Our novel analytical approach integrates epidemiology data and omics data to discover
369 molecular risk factors. While we focus on transcriptional regulation in this study, an
370 immediate next step is to apply this approach to other molecular changes, including
371 genetic variation, epigenetic modification, and metabolic perturbation to investigate their
372 roles in COVID-19 pathogenesis. As before-infection samples of COVID-19 patients are
373 difficult to acquire, integration of existing multi-omics data and epidemiology data hold
374 promise to accelerate the discovery of diagnostic and therapeutic markers to improve the
375 management of COVID-19 disease.

376

377 **Conclusions**

378

379 Our study illuminates that gene expression dysregulations in pre-existing health
380 conditions that impair innate immunity are molecular risk factors of COVID-19 related
381 death. The individual risk factor genes or gene sets are potential mediators in disease
382 pathogenesis. These findings allow for better prediction of severe outcome, inform the
383 development of preventative measures to reduce fatality, and inform the evaluation of
384 long-term health impact of COVID-19 in different populations.

385

386 **Acknowledgements**

387

388 We thank Dr. Terry Christenson for his professional editing of the manuscript.

References

1. World Health Organization. Coronavirus disease (COVID-19) 2021 [cited 2021 Apr 22]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
2. World Health Organization. WHO Coronavirus (COVID-19) Dashboard 2021 [cited 2021 Apr 22]. Available from: <https://covid19.who.int/>
3. Takagi H. Risk and protective factors of SARS-CoV-2 infection. *J Med Virol.* 2021;93(2):649-51.
4. Center for Disease Control and Prevention. Provisional Mortality Data - United States, 2020. Centers for Disease Control and Prevention 2021 [cited 2021 Apr 22]. Available from: <https://www.cdc.gov/mmwr/volumes/>
5. Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, Liu S, Yang JK. Gender differences in patients With COVID-19: focus on severity and mortality. *Front Public Health.* 2020;8:152.
6. Makary M. Risk factors for COVID-19 mortality among privately insured patients: A claims data analysis. FAIR Health, Inc. 2020.
7. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L, Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020;584(7821):430-6.
8. Center for Disease Control and Prevention. Certain Medical Conditions and Risk for Severe COVID-19 Illness. Centers for Disease Control and Prevention 2021 [cited 2021 Apr 22]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>
9. Johnson-Agbakwu CE, Ali NS, Oxford CM, Wingo S, Manin E, Coonrod DV. Racism, COVID-19, and health inequity in the USA: a call to action. *J Racial Ethn Health Disparities.* 2020.
10. Hojyo S, Uchida M, Tanaka K, Hasebe R, Tanaka Y, Murakami M, Hirano T. How COVID-19 induces cytokine storm with high mortality. *Inflamm Regen.* 2020;40:37.
11. Benard A, Jacobsen A, Brunner M, Krautz C, Klosch B, Swierzy I, Naschberger E, Podolska MJ, Kouhestani D, David P, Birkholz T, Castellanos I, Trufa D, Sirbu H, Vetter M, Kremer AE, Hildner K, Hecker A, Edinger F, Tenbusch M, Muhl-Zurbes P, Steinkasserer A, Richter E, Streeck H, Berger MM, Brenner T, Weigand MA, Swirski FK, Schett G, Grutzmann R, Weber GF. Interleukin-3 is a predictive marker for severity and outcome during SARS-CoV-2 infections. *Nat Commun.* 2021;12(1):1112.
12. Szarpak L, Ruetzler K, Safiejko K, Hampel M, Pruc M, Kanczuga-Koda L, Filipiak KJ, Jaguszewski MJ. Lactate dehydrogenase level as a COVID-19 severity marker. *Am J Emerg Med.* 2020.
13. Liu T, Zhang J, Yang Y, Ma H, Li Z, Zhang J, Cheng J, Zhang X, Zhao Y, Xia Z, Zhang L, Wu G, Yi J. The role of interleukin-6 in monitoring severe case of coronavirus disease 2019. *EMBO Mol Med.* 2020;12(7):e12421.

14. Xiao N, Nie M, Pang H, Wang B, Hu J, Meng X, Li K, Ran X, Long Q, Deng H, Chen N, Li S, Tang N, Huang A, Hu Z. Integrated cytokine and metabolite analysis reveals immunometabolic reprogramming in COVID-19 patients with therapeutic implications. *Nat Commun.* 2021;12(1):1618.
15. Ayres JS. A metabolic handbook for the COVID-19 pandemic. *Nat Metab.* 2020;2(7):572-85.
16. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5.
17. Kurreck J, Stein CA. *Molecular medicine: an introduction*: John Wiley & Sons; 2015.
18. Mesko B, Poliska S, Szegedi A, Szekanecz Z, Palatka K, Papp M, Nagy L. Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets. *BMC Med Genomics.* 2010;3:15.
19. Aziz H, Zaas A, Ginsburg GS. Peripheral blood gene expression profiling for cardiovascular disease assessment. *Genomic Med.* 2007;1(3-4):105-12.
20. Chandrashekar P, Ahmadinejad N, Wang J, Sekulic A, Egan JB, Asmann YW, Kumar S, Maley C, Liu L. Somatic selection distinguishes oncogenes and tumor suppressor genes. *Bioinformatics.* 2020;36(6):1712-7.
21. 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(1):289-300.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-7.
23. Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, Cinatl J, Munch C. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature.* 2020;583(7816):469-72.
24. Li S. Regulation of ribosomal proteins on viral infection. *cells.* 2019;8(5).
25. Campos RK, Wong B, Xie X, Lu YF, Shi PY, Pompon J, Garcia-Blanco MA, Bradrick SS. RPLP1 and RPLP2 are essential flavivirus host factors that promote early viral protein accumulation. *J Virol.* 2017;91(4).
26. Chiu YH, Macmillan JB, Chen ZJ. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell.* 2009;138(3):576-91.
27. Samuel CE. Antiviral actions of interferons. *Clin Microbiol Rev.* 2001;14(4):778-809.
28. Banerjee AK, Blanco MR, Bruce EA, Honson DD, Chen LM, Chow A, Bhat P, Ollikainen N, Quinodoz SA, Loney C, Thai J, Miller ZD, Lin AE, Schmidt MM, Stewart DG, Goldfarb D, De Lorenzo G, Rihn SJ, Voorhees RM, Botten JW, Majumdar D, Guttman M. SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell.* 2020;183(5):1325-39 e21.
29. Chen Z, Mi L, Xu J, Yu J, Wang X, Jiang J, Xing J, Shang P, Qian A, Li Y, Shaw PX, Wang J, Duan S, Ding J, Fan C, Zhang Y, Yang Y, Yu X, Feng Q, Li B, Yao X, Zhang Z, Li L, Xue X, Zhu P. Function of HAb18G/CD147 in invasion of host cells by severe acute respiratory syndrome coronavirus. *J Infect Dis.* 2005;191(5):755-60.
30. Birge RB, Ucker DS. Innate apoptotic immunity: the calming touch of death. *Cell Death Differ.* 2008;15(7):1096-102.

31. West AP, Shadel GS, Ghosh S. Mitochondria in innate immune responses. *Nat Rev Immunol.* 2011;11(6):389-402.
32. Thorburn A. Apoptosis and autophagy: regulatory connections between two supposedly different processes. *Apoptosis.* 2008;13(1):1-9.
33. Khandia R, Dadar M, Munjal A, Dhama K, Karthik K, Tiwari R, Yattoo MI, Iqbal HMN, Singh KP, Joshi SK, Chaicumpa W. A comprehensive review of autophagy and its various roles in infectious, non-infectious, and lifestyle diseases: current knowledge and prospects for disease prevention, novel drug design, and therapy. *cells.* 2019;8(7).
34. Possik E, Jalali Z, Nouet Y, Yan M, Gingras MC, Schmeisser K, Panaite L, Dupuy F, Kharitidi D, Chotard L, Jones RG, Hall DH, Pause A. Folliculin regulates ampk-dependent autophagy and metabolic stress survival. *PLoS Genet.* 2014;10(4):e1004273.
35. Ajaz S, McPhail MJ, Singh KK, Mujib S, Trovato FM, Napoli S, Agarwal K. Mitochondrial metabolic manipulation by SARS-CoV-2 in peripheral blood mononuclear cells of patients with COVID-19. *Am J Physiol Cell Physiol.* 2021;320(1):C57-C65.
36. Gassen NC, Papies J, Bajaj T, Emanuel J, Dethloff F, Chua RL, Trimpert J, Heinemann N, Niemeyer C, Weege F, Honzke K, Aschman T, Heinz DE, Weckmann K, Ebert T, Zellner A, Lennarz M, Wyler E, Schroeder S, Richter A, Niemeyer D, Hoffmann K, Meyer TF, Heppner FL, Corman VM, Landthaler M, Hocke AC, Morkel M, Osterrieder N, Conrad C, Eils R, Radbruch H, Gialvalisco P, Drosten C, Muller MA. SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. *Nat Commun.* 2021;12(1):3818.
37. Paolini A, Borella R, De Biasi S, Neroni A, Mattioli M, Lo Tartaro D, Simonini C, Franceschini L, Cicco G, Piparo AM, Cossarizza A, Gibellini L. Cell death in coronavirus infections: uncovering its role during COVID-19. *Cells.* 2021;10(7).
38. Horiguchi H, Loftus TJ, Hawkins RB, Raymond SL, Stortz JA, Hollen MK, Weiss BP, Miller ES, Bihorac A, Larson SD, Mohr AM, Brakenridge SC, Tsujimoto H, Ueno H, Moore FA, Moldawer LL, Efron PA, Sepsis, Critical Illness Research Center I. Innate immunity in the Persistent Inflammation, Immunosuppression, and Catabolism Syndrome and its implications for therapy. *Front Immunol.* 2018;9:595.
39. Schultze JL, Aschenbrenner AC. COVID-19 and the human innate immune system. *Cell.* 2021;184(7):1671-92.
40. Ozcan L, Tabas I. Role of endoplasmic reticulum stress in metabolic disease and other disorders. *Annu Rev Med.* 2012;63:317-28.
41. Bhatti JS, Bhatti GK, Reddy PH. Mitochondrial dysfunction and oxidative stress in metabolic disorders - A step towards mitochondria based therapeutic strategies. *Biochim Biophys Acta Mol Basis Dis.* 2017;1863(5):1066-77.
42. Inoue T, Tsai B. How viruses use the endoplasmic reticulum for entry, replication, and assembly. *Cold Spring Harb Perspect Biol.* 2013;5(1):a013250.
43. Chan CP, Siu KL, Chin KT, Yuen KY, Zheng B, Jin DY. Modulation of the unfolded protein response by the severe acute respiratory syndrome coronavirus spike protein. *J Virol.* 2006;80(18):9279-87.

44. Banerjee A, Czinn SJ, Reiter RJ, Blanchard TG. Crosstalk between endoplasmic reticulum stress and anti-viral activities: a novel therapeutic target for COVID-19. *Life Sci.* 2020;255:117842.
45. Zhang X, Yang Z, Long X, Sun Q, Wang F, Wang P-h, Li X, Kuang E. Sequential ER stress and inflammatory responses are induced by SARS-CoV-2 ORF3 through ERphagy. *bioRxiv.* 2020. doi:10.1101/2020.11.17.387902
46. Kvensakul M. Viral infection and apoptosis. *Viruses.* 2017;9(12).
47. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA, tenOever BR. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell.* 2020;181(5):1036-45 e9.
48. Zheng HY, Zhang M, Yang CX, Zhang N, Wang XC, Yang XP, Dong XQ, Zheng YT. Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cell Mol Immunol.* 2020;17(5):541-3.
49. Trinchieri G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nat Rev Immunol.* 2003;3(2):133-46.

Figure 1

Analysis workflow.

(A) Compile microarray-based transcriptomic data sets for common health conditions. We included case-control studies using peripheral blood samples. **(B)** Find differentially expressed genes (DEGs) for each health condition. Recurrent DEGs were those differentially expressed in at least four health conditions. **(C)** Perform correlation tests to identify molecular risk factors, which are pre-existing expression dysregulations that increase the risk of COVID-19 related death. **(D)** Examine functional relationships of molecular risk factors via enrichment and network analyses. **(E)** Validate molecular risk genes using published data from an *in vitro* SARS-CoV-2 infection study [23].

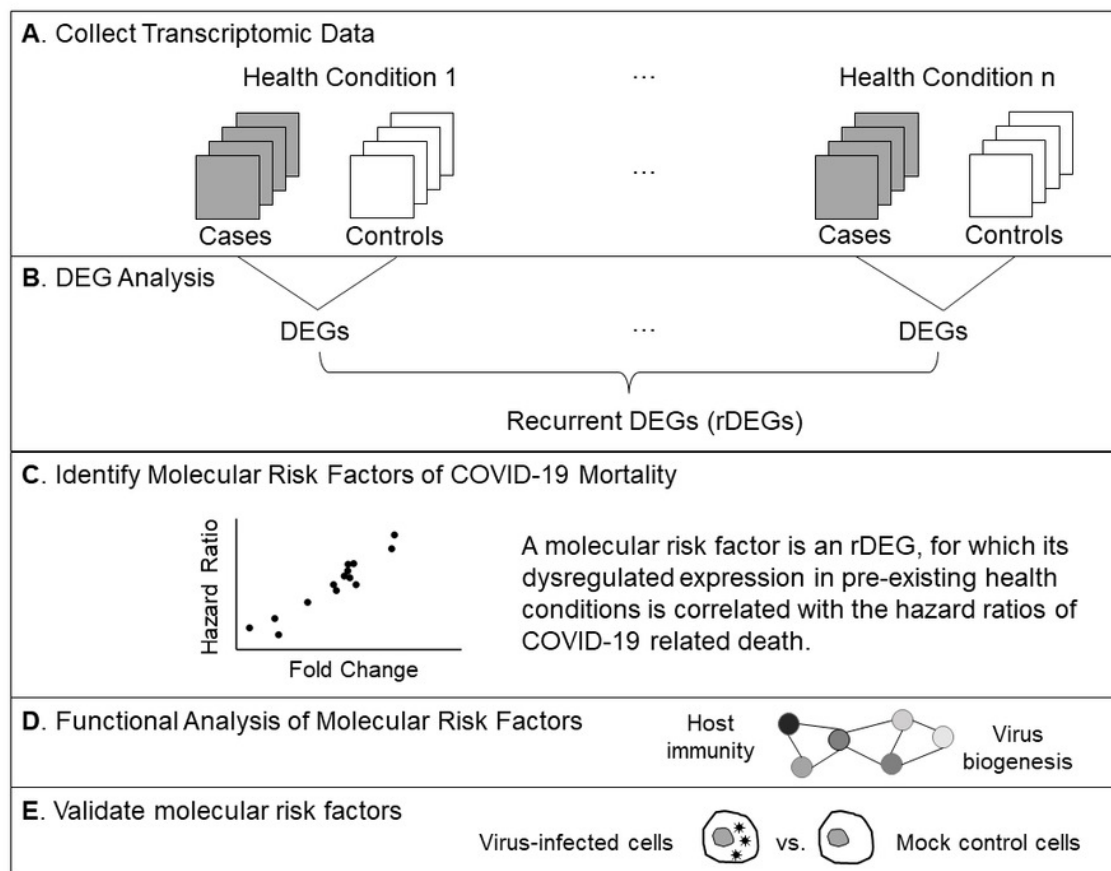


Figure 2

Scatter plots of selected genes showing correlations between fold change (FC) and hazard ratio (HR).

(A) *RPS28* had the lowest Pearson correlation test *P* values and positive PCC values. The fourteen health conditions are represented by different symbols (AD: Alzheimer's disease, AH: alcoholic hepatitis, Asthma: asthma, CAD: coronary artery disease, CKD: chronic kidney disease, COPD: chronic obstructive pulmonary disease, Ethnicity: Latino vs. Caucasian, HIV: human immunodeficiency virus, HT: hypertension, Lupus: lupus, MS: multiple sclerosis, Obese: obesity, RA: rheumatoid arthritis, and T2D: type-2 diabetes). Black symbols indicate health conditions in which a given gene was differentially expressed. Gray symbols indicate health conditions in which a given gene was not differentially expressed. Broken lines represent linear fits between FC and HR. **(B)** Seven genes coding ribosomal proteins consistently show positive correlations between expression FCs and HRs of COVID-19 mortality. **(C)** *POLR3GL* had the lowest Pearson correlation test *P* values and negative PCC values. It is involved in interferon production, providing anti-viral innate immunity. **(D)** Two additional genes involved in interferon signaling show negative correlations between FC and HR.

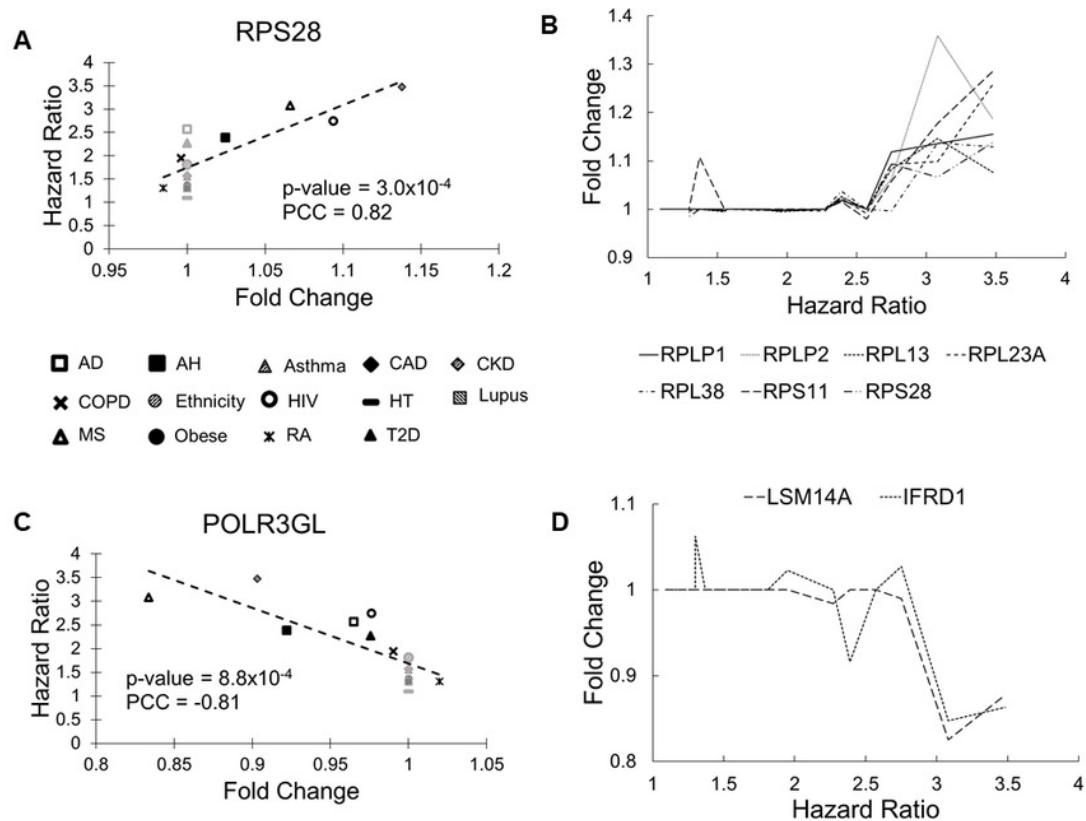


Figure 3

Functional groups enriched with molecular risk factor genes.

(A) The association network of gene sets with $FDR < 0.05$ from enrichment analysis. This network contained two types of nodes, with gene sets represented by open circles and individual risk factor genes by colored dots. An edge links a risk factor gene to its associated gene set. The size of an open circle is proportional to the number of risk factor genes connected to it. The color of a dot indicates if its upregulation is positively (red) or negatively (blue) correlated with HR of COVID-19 mortality. **(B)** Upregulation of *PPIA* increased risk of COVID-19 mortality. *PPIA* facilitates virus replication. **(C)** Downregulation of *HNRNPA2B1* increased risk of COVID-19 mortality. *HNRNPA2B1* inhibits virus replication.

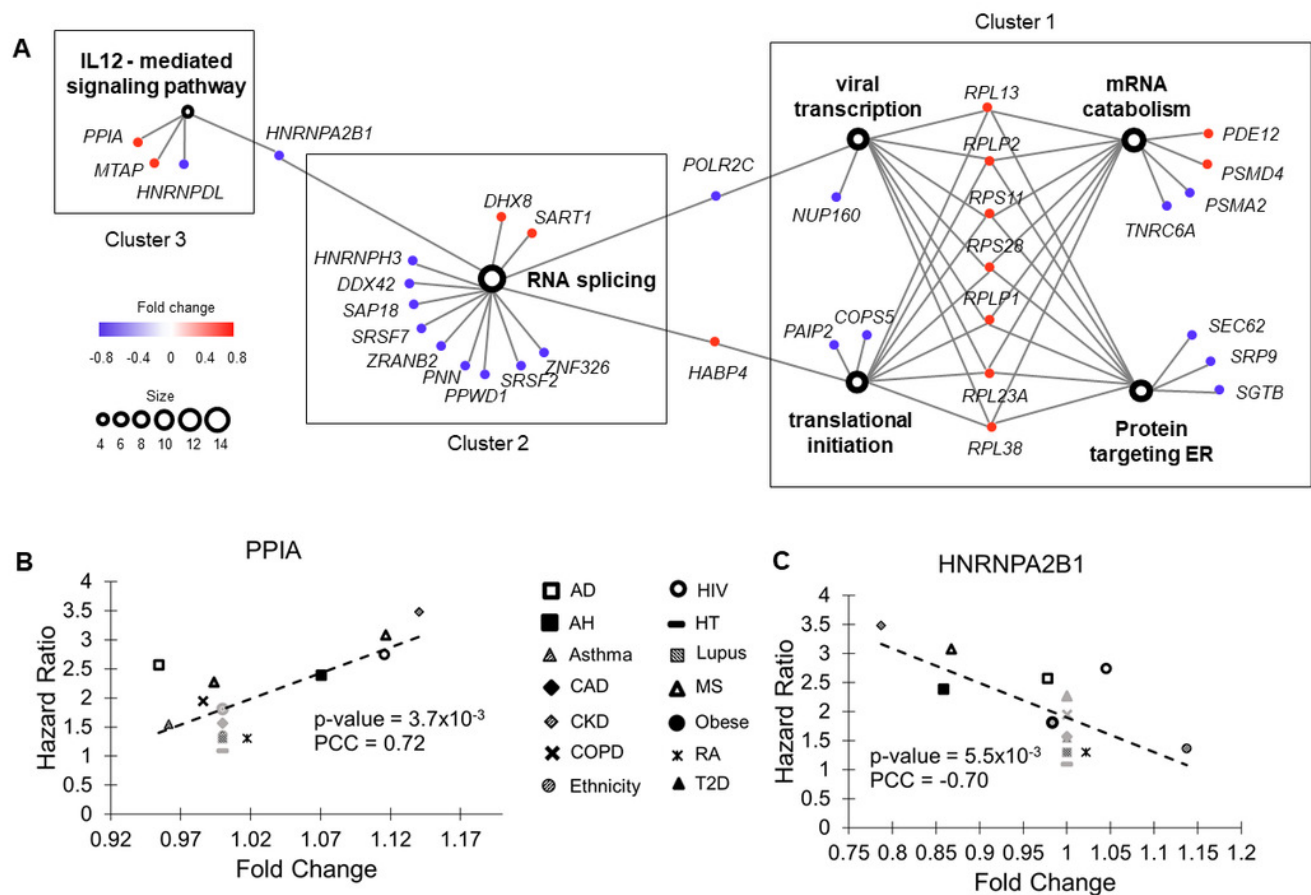


Figure 4

Genes involved in autophagy, apoptosis, and mitochondrial functions.

(A) The association network of gene sets with nominal P values <0.1 from enrichment analysis. This network contained two types of nodes, with gene sets represented by open circles and individual risk factor genes by colored dots. An edge links a risk factor gene to its associated gene set. The color of a dot indicates if its upregulation is positively (red) or negatively (blue) correlated with HR of COVID-19 mortality. **(B)** Five genes (blue lines) inducing autophagy and suppressing apoptosis show negative correlations between expression FCs and HRs of COVID-19 mortality. One gene (red line) suppressing autophagy and inducing apoptosis shows positive correlations between expression FCs and HRs of COVID-19 mortality.

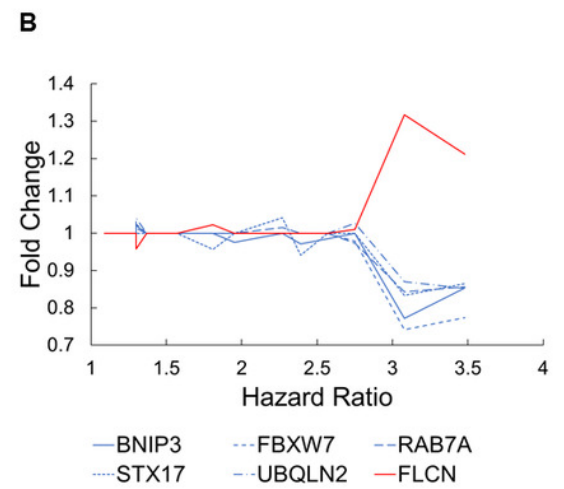
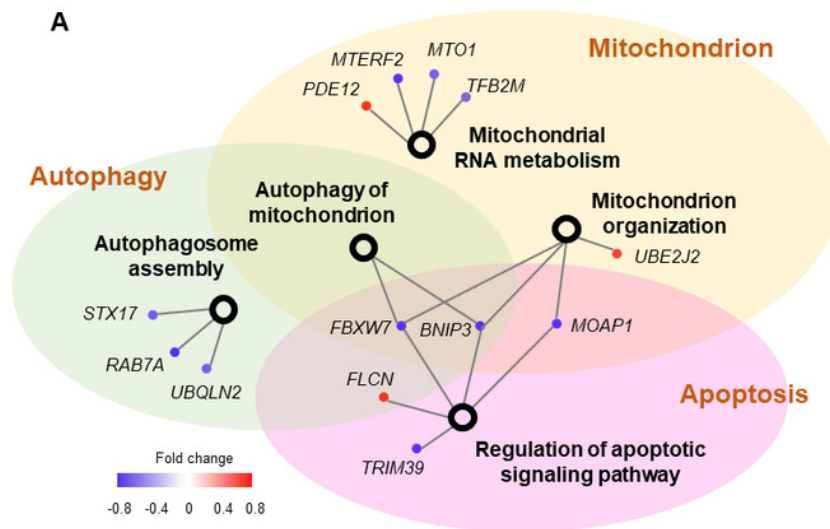


Figure 5

Protein expression of risk factor genes. Boxplot of protein abundance (log scale) in mock controls and SARS-CoV-2 infected cells.

(A) The plot includes genes for which upregulation increased risk of COVID-19 related death.

(B) The plot includes genes for which downregulation increased risk of COVID-19 related death.

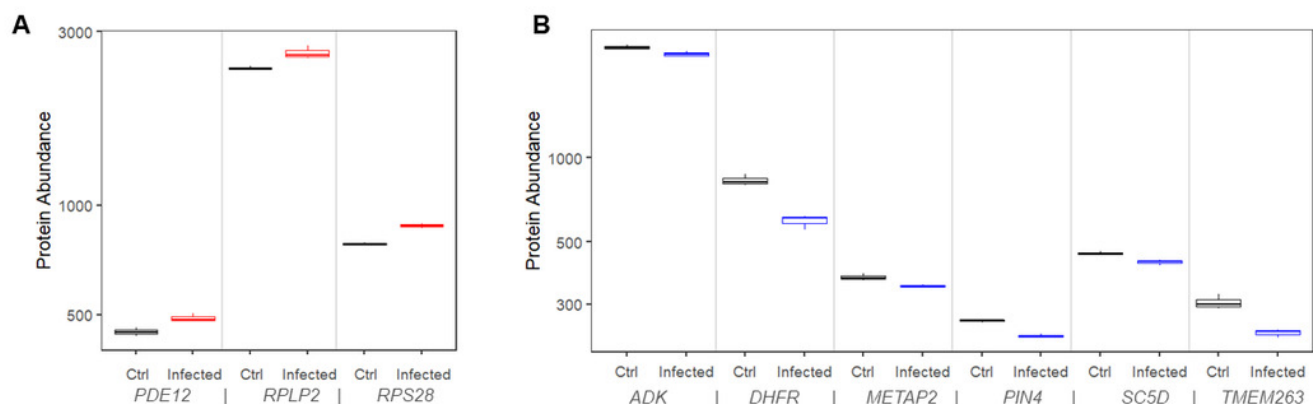


Table 1 (on next page)

Data sets and DEGs of fourteen pre-existing condition.

1 **Table 1:** Data sets and DEGs of fourteen pre-existing conditions

Health Condition	Gene Expression			COVID-19 Mortality *	
	GEO Accession	DEG Count	Mean FC	Disease Group	HR
Alzheimer's disease	GSE63060	5,413	0.99	Stroke or dementia	2.57
Asthma	GSE110551	1,215	1.01	Asthma	1.55
Coronary artery disease	GSE12288	1,289	1.10	Chronic heart disease	1.57
Chronic obstructive pulmonary disease	GSE42057	2,054	0.99	Respiratory disease excluding asthma	1.95
Hispanic ethnicity	GSE30101	4,249	1.13	Ethnicity	1.37
HIV	GSE104640	1,1503	1.00	Immunosuppressive condition	2.75
Hypertension	GSE135111	413	1.53	High blood pressure or diagnosed hypertension	1.09
Lupus	GSE37356	2,289	1.00	Rheumatoid arthritis, lupus or psoriasis	1.30
Obesity	GSE110551	3,905	1.00	BMI >35	1.81
Rheumatoid Arthritis	GSE93272	9,902	1.01	Rheumatoid arthritis, lupus or psoriasis	1.30
Type-2 Diabetes	GSE65561	4,155	0.99	Diabetes	2.27
Chronic Kidney Disease	GSE37171	17,688	1.01	Reduced kidney function (eGFR<30)	3.48
Multiple sclerosis	GSE21942	7,735	1.12	Other neurological disease	3.08
Alcoholic Hepatitis	GSE28619	9,489	1.05	Liver disease	2.39

2 * as reported in the OpenSAFELY Project [7]

3

Table 2(on next page)

Significantly enriched gene sets

Table 2: Significantly enriched gene sets.

ID	Description	Enrich	FDR	Count	Genes
KEGG Pathways					
hsa03010	Ribosome	4.77	0.00026	8	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/RPL30/RPL23A</i>
hsa00563	GPI-anchor biosynthesis	10.87	0.0025	3	<i>PIGP/PIGA/PIGG</i>
hsa05171	Coronavirus disease COVID-19	3.25	0.0031	8	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/RPL30/RPL23A</i>
GeneOntology Biological Process					
GO:0045047	protein targeting to ER	8.55	0.00036	10	<i>RPL38/RPS11/SEC62/RPS28/RPLP1/RPLP2/SRP9/RPL13/SGTB/RPL23A</i>
GO:0006613	co-translational protein targeting to membrane	8.33	0.00087	9	<i>RPL38/RPS11/SEC62/RPS28/RPLP1/RPLP2/SRP9/RPL13/RPL23A</i>
GO:0022626	cytosolic ribosome	7.34	0.0040	8	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/RPL30/RPL23A</i>
GO:0006413	translational initiation	5.23	0.0055	10	<i>HABP4/RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/PAIP2/COPS5/RPL23A</i>
GO:0019083	viral transcription	5.13	0.013	9	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/NUP160/POLR2C/RPL23A</i>
GO:00116607	nuclear speck	3.45	0.014	13	<i>PNN/HABP4/CBLL1/HBP1/ACADM/SRSF2/SRSF7/RFXAP/DDX42/SART1/POLDIP3/NRIP1/SAP18</i>
GO:0000184	nuclear-transcribed mRNA catabolic process	5.89	0.031	7	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/RPL23A</i>
GO:0008380	RNA splicing	3.01	0.032	14	<i>PNN/HABP4/DHX8/ZRANB2/HNRNPH3/HNRNPA2B1/ZNF326/SRSF2/SRSF7/DDX42/PPWD1/SART1/POLR2C/SAP18</i>
GO:0044391	ribosomal subunit	4.32	0.040	8	<i>RPL38/RPS11/RPS28/RPLP1/RPLP2/RPL13/RPL30/RPL23A</i>
GO:0031307	mitochondrial outer membrane component	15.40	0.049	3	<i>SYNJ2BP/TOMM20/BNIP3</i>