medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Genetic and Phenotypic Evidence for the Causal Relationship Between Aging and COVID-19

Kejun Ying[1,2,3†], Ranran Zhai[1†], Timothy V. Pyrkov[4], Marco Mariotti[2], Peter O. Fedichev[4,5], Xia Shen*[1,6,7] & Vadim N. Gladyshev*[2]

[1] Biostatistics Group, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China
[2] Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[3] T. H. Chan School of Public Health, Harvard University, Boston, USA
[4] Gero LLC PTE, Singapore
[5] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia
[6] Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom
[7] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

* Correspondence to: xia.shen@ed.ac.uk and vgladyshev@rics.bwh.harvard.edu
[†] These authors contributed equally to this work

Abstract

Epidemiological studies have revealed that the elderly and those with co-morbidities are most susceptible to COVID-19. To understand the genetic link between aging and the risk of COVID-19, we conducted a multi-instrument Mendelian randomization analysis and found that the genetic variation that leads to a longer lifespan is significantly associated with a lower risk of COVID-19 infection. The odds ratio is 0.32 (95% CI: 0.18 to 0.57; $P = 1.3 \times 10^{-4}$) per additional 10 years of life, and 0.62 (95% CI: 0.51 to 0.77; $P = 7.2 \times 10^{-6}$) per unit higher log odds of surviving to the 90th percentile age. On the other hand, there was no association between COVID-19 susceptibility and healthspan (the lifespan free of the top seven age-related morbidities). To examine the relationship at the phenotypic level, we applied various biological aging clock models and detected an association between the biological age acceleration and future incidence and severity of COVID-19 infection for all subjects as well as for the individuals free of chronic disease. Biological age acceleration was also significantly associated with the risk of death in COVID-19 patients. Our findings suggest a causal relationship between aging and COVID-19, defined by genetic variance, the rate of aging, and the burden of chronic diseases.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The coronavirus disease 2019 (COVID-19), caused by severe acute respiratory coronavirus 2 (SARS-CoV-2), first emerged in late 2019 and has led to an unprecedented global health crisis[1]. Notably, the aging population is particularly at risk of COVID-19[2], e.g. in Italy, 88% of the individuals tested positive for COVID-19 were 40 years or older[3]. Moreover, a recent report based on epidemiological data from multiple countries showed that 69% of infections in people over 70-year-old are symptomatic, and this number drops to only 21% for 10- to 19-year-olds[4]. Unsurprisingly, elderly people are also more likely to die from COVID-19, and the case fatality rate for COVID-19 grows exponentially with age[3]. As observational evidence implies a strong link between COVID-19 and age, COVID-19 can be considered a disease of aging[3], and multiple clinical trials using potential lifespan-extending drugs (e.g., metformin, rapamycin, and senolytics) to protect the elderly from COVID-19 have been proposed[5-7]. Although observational data on metformin seems promising[8,9], it is unclear if other lifespan-extending drugs should be prioritized in clinical trials, since the evidence of any causal link between lifespan and COVID-19 susceptibility is still missing.

Mendelian Randomization (MR) is a genetic instrumental variable approach that assesses the causal effect of exposure of interest on an outcome, by ascertaining on genetic variants, e.g., single nucleotide polymorphisms (SNPs), strongly associated with the exposure phenotype. Since the alleles of the genetic variants are naturally randomly allocated at conception, when the genetic effects on the outcome are only mediated through the exposure, the causal effect inferred by MR is, in analogy to randomized clinical trials (RCTs), free of any environmental confounders and reverse causation. Although RCT is considered as a gold standard for establishing causal relationships, MR can provide valuable insights into causality when it is infeasible to perform an RCT or before an RCT is performed[10].

In this study, we performed a multi-SNP MR analysis to elucidate whether and how aging is associated with COVID-19. We considered four lifespan-related traits (parental lifespan, healthspan, longevity, and the combination of these three traits), four measures of epigenetic age acceleration, and four leading genetic risk factors associated with earlier death in humans (Alzheimer Disease (AD), cardiovascular disease (CVD), type 2 diabetes (T2D, and smoking) as exposures and evaluated their causal effects on COVID-19 infection and related phenotypes. To support the argument, we further estimated the biological age acceleration (BAA) in COVID-19 patients from UK Biobank (UKBB) cohort and observed a significant association between the phenotypic indicators of aging progress (aging clocks) and the

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

risk and case fatality rate of COVID-19. Altogether, our findings reveal a causal role of lifespan-related traits in the risk of COVID-19, supporting the idea of using established lifespan-extending drugs as a preventive measure to COVID-19 in the elderly.

Methods

GWAS data for lifespan-related traits and diseases

We studied four lifespan-related traits with publicly available GWAS summary statistics: including parental lifespan[11], healthspan[12], longevity[13], and a multivariate meta-analysis on these three traits, as a measurement of overall healthy aging[14]. We investigated four additional traits genetically correlated with lifespan, using the published case-control studies: Alzheimer's disease[15], coronary artery disease[16], type 2 diabetes[17], and smoking[18] (Table S1).

We also included GWAS for age acceleration measured by four epigenetic clocks, including Hannum age, Horvath age, PhenoAge, and GrimAge[19]. The epigenetic age was measured on 34,449 healthy individuals of European ancestry.

GWAS data for 22 common diseases were from a community-based study, Genetic Epidemiology Research on Adult Health and Aging (GERA)[20], and was analyzed in original GSMR study[21]. There were 60,586 individuals of European ancestry in the GERA data. There is an additional trait "disease count", which represents the number of diseases affecting each individual and the summary statistics of these diseases were adjusted with age, gender, and the first 20 PCs.

We used 1000 Genomes Phase 3 reference (released in 2014 October) to map variants in the GWAS results to rsIDs by chromosome, position, and alleles. Only the autosomal SNPs available in the 1000 Genomes reference panel were used, and the 1000 Genomes European ancestry reference was used to estimate the linkage disequilibrium (LD) among these SNPs. Duplicated rsIDs in the data were removed prior to the analysis.

COVID-19-related traits

To extensively evaluate the genetic effects on COVID-19 risk, we used GWAS summary statistics data from 11 COVID-19-related traits (Table S1). The GWAS results for SARS-COV-2 infection are from the National Institute of Health, Genome-Wide Repository of Associations Between SNPs and Phenotypes (NIH-GRASP), which includes 1,221 positive cases and 458,029 negative/unknown UK Biobank controls. The GWAS summary statistics for severe COVID-19 with respiratory failure is from a genome-wide association study performed in

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1,610 cases and 2,205 controls in Italy and Spain[22]. The rest of the 9 traits are from the COVID-19 Host Genetics Initiative (HGI), with the sample size varies from 1,332 to 1,079,768[23]. Among them, 3 are were from HGI release 2 (May 2020), including COVID-19 Hospitalization (versus non-hospitalized COVID-19), susceptibility (affected versus unaffected population), and COVID-19 predicted by flu-like symptoms; Other 6 traits are were in HGI release 3 (June 2020), including very severe respiratory confirmed COVID-19 (versus the general population), COVID-19 infection (versus negative control or population), hospitalized COVID-19 (versus not hospitalized COVID-19 or population), and predicted COVID from self-reported symptoms (versus predicted or self-reported non-COVID-19).

Mendelian randomization analysis

Mendelian randomization is a method that uses genetic variants as instrumental variables to determine whether an observational association between a risk factor and an outcome is consistent with a potential causal effect[24]. The multi-SNP MR analysis was implemented using GSMR (Generalized Summary-data-based Mendelian randomization) in GCTA[21]. As instruments for each exposure (four lifespan-related traits, four risk factors, and four epigenetic age acceleration traits), we selected near-independent SNPs ($r^2 < 0.1$) with genome-wide significant ($P < 5\times10^{-8}$) association with the exposure. For 22 diseases from GERA community-based study, we selected SNPs with suggestive genome-wide significance ($P < 1\times10^{-6}$) as instruments and performed a separate analysis due to the limited case number in the community-based study. GSMR includes a HEIDI-outlier filter to remove potential pleiotropic SNPs that have effects on the exposures and the outcomes via different pathways. We set its p-value threshold to 0 or 0.01 and tested the remaining SNPs for association with the COVID-19-related traits. The required minimum number of instrumental SNPs for each exposure in the analysis is lowered to 1.

Genetic correlation analysis

We estimated genetic correlations also estimated the genetic correlations between lifespan-related traits, risk factors, epigenetic age acceleration, and COVID-19 using LD score regression (LDSC) and high-definition likelihood (HDL) methods[25,26]. SNPs that are imperfectly imputed (INFO < 0.9) or with low frequency (MAF < 0.05) were removed to reduce statistical noise. LDSC was performed using LDSC software v1.0.1 (https://github.com/bulik/ldsc); and the HDL was performed using R package "HDL" v1.3.8 (https://github.com/zhenin/HDL).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Biological age estimation for UKBB cohorts

All-cause mortality increases exponentially with age, and hence log-linear risk predictors from proportional hazards models can provide natural composite organism state representations characterizing the progress of aging based on biological and physiological measurements. We used two such biological age measures: Phenotypic Age based on blood biochemistry[27], and Dynamic Organism State Indicator (DOSI) based on widely available Complete Blood Counts (CBC) only[28]. The latter is a proxy for the frailty index and is derived from the blood markers only, whereas the Phenotypic Age additionally employs the explicit age. We also used physical activity (the number of steps per day averaged over the week), which is also associated with all-cause mortality and hence can also be viewed as a measure of biological aging[29].

We investigated the association of the incidence of COVID-19 with biological aging acceleration (BAA, which is the difference between the biological age of a person and the average biological age in the cohort of individuals of the same age and sex) using logistic regression. Chronological age and biological sex were used as additional covariates in the analysis.

Following UKBB recommendations, we used the "result" label from the table "COVID-19 test results table" as the proxy of disease severity. This implies that mostly those individuals that showed characteristic COVID-19 symptoms were selected for testing. We investigated BAA associations with the incidence of COVID-19 and its associated fatality using all available cases (All) and separately cohorts of individuals who have (Frail) or do not have (Not Frail) major chronic diseases (from the list including all kinds of cancer, angina pectoris, coronary heart disease, heart attack, heart failure, hypertension, stroke, diabetes, arthritis, bronchitis, and emphysema) at the time of infection.

Results

We applied GSMR to test for potential causal associations between four lifespan-related traits and COVID-19, including lifespan, longevity (i.e. surviving to the 90th percentile), healthspan (time to a first major age-related disease), and the combined lifespan/longevity/healthspan trait (Table S1). We employed summary-level GWAS data[11-14], and selected near-independent SNPs (r^2 threshold = 0.1, the remaining LD is accounted by GSMR method) at a genome-wide significance level ($P_{GWAS} < 5 \times 10^{-8}$) as genetic instruments for each trait. This resulted in 25, 17, 6, and 37 instrumental SNPs for lifespan,

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

healthspan, longevity, and the three combined traits, respectively. For the outcomes, we used 11 different sets of GWAS summary statistic data for COVID-19-related traits from case-control studies (Table S1).

Strikingly, our GSMR analysis showed that long lifespan and longevity were protective against COVID-19 infection based on UKBB reporting (Fig. 1, Table 1). The estimated odds ratio of lifespan was 0.32 (95% CI: 0.18 to 0.57; $P = 1.3 \times 10^{-4}$), indicating that the risk of COVID-19 infection is decreased by 68% with approximately every additional 10 years of life[11]. For longevity, i.e., surviving to the 90th percentile age, the odds ratio was 0.62 (95% CI: 0.51 to 0.77; $P = 7.2 \times 10^{-6}$). As UKBB reporting on COVID-19 infection is largely biased toward severe cases of hospitalized subjects, this measure might also be related to COVID-19 severity and mortality[30]. The predicted longer lifespan and longevity also appeared to be protective against COVID-19 infection in the HGI dataset at a nominal significance level ($P < 0.05$). However, none of the lifespan-related traits showed a significant protective effect on COVID-19 with a severe respiratory disorder or respiratory failure, possibly due to the small sample size. We also estimated the causal effect of genetically-proxied epigenetic age acceleration on the risk of COVID-19. The four epigenetic clock measurements included the Hannum age, Horvath age, PhenoAge, and GrimAge[19]. Notably, GrimAge is trained on mortality and outperforms the other three epigenetic clocks in predicting lifespan[31,32]. In our MR analysis, GrimAge acceleration was found to be a risk factor for COVID-19 infection at a nominal significance level (Table 1, Fig. S1). The odds ratio was 1.25 (95% CI: 1.01 to 1.55; $P = 0.04$), suggesting that subjects with higher GrimAge per year have a 25% higher risk of COVID-19 infection.

Healthspan is defined as the age free of major age-related morbidities. In the healthspan GWAS study, the top seven age-related morbidities were included, i.e., congestive heart failure, myocardial infarction, chronic obstructive pulmonary disease, stroke, dementia, diabetes, and cancer[12]. In our analysis, healthspan did not show a significant effect on COVID-19-related traits. This is unlikely to be due to the power of healthspan GWAS since there were 17 near-independent genome-wide significant SNPs ($P < 5 \times 10^{-8}$), which is the second-most among the five lifespan-related traits. Therefore, we hypothesized that the strong protective effect of longevity against COVID-19 may not be explained by the delayed appearance of age-related morbidities, but rather by decelerated biological age that extends lifespan.

To address this hypothesis, we assessed in parallel the three different risk-based biological age predictions computed for the subjects in the UKBB cohort using blood biochemistry

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Phenotypic Age), Complete Blood Counts (DOSI), and physical activity measurements[27-29]. We found that COVID-19 incidence in all UKBB datasets was significantly associated with BAA of Phenotypic Age, DOSI, and decreased physical activity (Fig. 2, Table 2). The estimated odds ratio of COVID-19 infection is 1.28 (95% CI: 1.25 to 1.31; $P = 8.4 \times 10^{-82}$) and 1.31 (95% CI: 1.26 to 1.38; $P = 9.5 \times 10^{-32}$) for every ten years higher biological age measured by Phenotypic Age and DOSI, respectively. Phenotypic Age and DOSI were also significantly associated with COVID-19 incidence and case fatality independently from the BAA association with chronic diseases, i.e., separately in cohorts of UKBB individuals having (Frail) or not (Not frail) chronic age-related health conditions (Fig. 2, Table 2).

We also observed elevated BAA levels of all measures of biological age (Fig. S2A-G) in cohorts of individuals died from COVID-19 compared to those tested (and most probably suffering from the disease), and, separately, in cohorts of those tested versus the rest of UKBB (and presumed free of the disease). The number of UKBB subjects with data fields required for the Phenotypic Age and DOSI was comparable, and we found that Phenotypic Age comparisons produced a better statistical power. The number of UKBB subjects with physical activity metrics was small, but nevertheless the association of BAA in the form of physical activity deficit and the incidence of COVID-19 was significant.

The loci for AD, CVD, T2D, cancer, and smoking (or lung cancer) explained the most genetic variance of lifespan[11]. To investigate whether these risk factors contribute to the plausible causal association between lifespan and COVID-19, we conducted an additional MR analysis using as exposures late-onset AD, coronary artery disease (CAD, including myocardial infarction, percutaneous transluminal coronary angioplasty, coronary artery bypass grafting, chronic ischemic heart disease, and angina), T2D, and smoking (the number of cigarettes smoked per day) (Fig. 3, Table 1). The GWAS summary statistics were obtained from published GWAS[15-18]. Only late-onset AD was found to significantly increase the risk of COVID-19 after Bonferroni correction for 132 tests ($P = 1.6\times10^{-4}$), with an odds ratio of 1.13 (95% CI: 1.06 to 1.20). We also estimated the genetic correlations between lifespan-related traits and COVID-19 using LD score regression and high-definition likelihood (HDL) methods (Fig. 4)[25,26]. However, the case sample sizes of COVID-19 GWAS studies were too small to support a sufficient power for estimating genetic correlations among these traits. Future genetic analyses utilizing larger sample sizes should provide opportunities to improve these estimates.

To further evaluate the risk factors for COVID-19 infection and severity, we conducted a separate MR analysis using GWAS data of 22 common diseases from GERA[20] (Fig. S3,

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S3). Due to the limited case number in this community-based study, we used SNPs with suggestive genome-wide significance ($P < 1\times10^{-6}$) as instruments. None of the diseases reached the significance threshold after Bonferroni correction ($P = 0.05/253 = 2\times10^{-4}$). Among the nominally significant associations, dyslipidemia, hernia abdominopelvic cavity, peptic ulcer, and peripheral vascular disease were the risk factors for COVID-19 infection; and hypertensive disease, irritable bowel syndrome, and varicose veins were the risk factors for COVID-19 severity (Fig. S3). Interestingly, the disease count, a trait that represents the number of comorbidities affecting each individual, was the risk factor for severe cases of COVID-19 (hospitalization). The odds ratio was 3.57 (95% CI: 1.04 to 12.28, $P = 0.04$), suggesting a subject with roughly every two (one standard deviation) more comorbidities has a 3.6-fold higher risk of having a severe case of COVID-19 (Table S2).

Discussion

In this study, we explored a potential causal relationship between aging and the risk of COVID-19 by conducting a multi-instrument MR analysis using four different lifespan-related traits as exposures and eleven COVID-19-related traits as outcomes. We found that genetically proxied longer lifespan and longevity were significantly associated with the decreased risk of COVID-19 ($P = 1.3 \times 10^{-4}$ and 7.2×10^{-6}, respectively), and further analyses revealed a key role of an elevated biological age and severity of chronic age-related diseases in this association. One of the contributing factors is likely the immune response. The competence of the immune system declines as people age, which is known as "immunosenescence"[33]. The hallmarks of immunosenescence include an impaired response to new antigens, unsustained memory responses, increased autoimmune responses, and prolonged inflammation. As a result, elderly subjects are more susceptible to infectious diseases, including COVID-19, and have a poor response to vaccines[33,34]. On the other hand, it has been reported that the circulating immune cells in centenarians possess unique characteristics that sustain immune responses to infections[35]. Moreover, the offspring of centenarians were shown to have a lower level of inflammation[36], suggesting that the benefits on the immune system in centenarians are heritable. Therefore, a better immunological profile in people with pro-longevity genetics may help support the causal effect of longevity on COVID-19 we observed.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Aging manifests itself as progressive remodeling of the organism state, and hence a great number of biological measurements are associated with age. Several sets of physiological and biological indices have been proposed for quantification of aging progression in the form of a single number – the biological age[37,38], or frailty index[39,40]. One popular approach is to regress relevant variables to predict chronological age and thus produce the "biological age" prediction. Popular Hannum's and Horvath's methylation age-clock models are the widely used examples of such an approach[41,42].

An interesting alternative is to produce the log-linear all-cause mortality estimate with a proportional hazard model and treat the resulting value as a measure of biological age. Phenotypic Age from blood biochemistry markers[27], DOSI from CBC[28], averaged physical activity levels[29], and more sophisticated machine learning algorithms used to predict the risk of death from physical activity time series of wearable devices[43], or even self-reported health questionnaires, are all examples of this approach[44]. We were not able to obtain the epigenetic clock predictions of UKBB subjects due to the lack of corresponding measurements. It has been reported that risk-based biological age predictors often outperform chronological age-based estimations in tasks involving associations with risks of death, lifestyles, and diseases[27,28,43]. This may explain a stronger GrimAge association but non-significant associations of Horvath's and Hannum's biological age models with the incidence of COVID-19 in our MR study.

All reliable biological age predictors are associated with the chronic disease burden, unhealthy lifestyles such as smoking (both overall and in disease-free population), and the future incidence of chronic diseases in healthy subjects[27-29,31,32,37,38,40]. In this work, we also established the association of BAA with the risk of non-chronic diseases, such as COVID-19 and the corresponding case fatality in the UKBB cohort independently of disease burden. The association was significant for BAA measures obtained from blood biochemistry (Phenotypic Age)[27], CBC (DOSI)[28] and mean physical activity (number of steps per day recorded by wearable devices over a week-long period of time[29]; the number of UKBB subjects with physical activity measurements was too low for separate BAA characterization in frail and non-frail cohorts).

Decreased physical activity was associated with an increased risk of COVID-19 in the UKBB cohort. This observation may be interesting on its own since the widespread lockdown measures brought about a dramatic (up to 27.3%, which is 1,432 steps per day, within 30 days) decline of average physical activity[45]. Our association study suggests a more than 10% risk increase corresponding to 1.5 thousand steps per day loss. There are feedback loop

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

effects of decreased mobility on BAA measures, and as such the associated risk adjustments must be taken into account in advanced epidemiological models of lockdown effects.

One advantage of our study design is that all of the BAA predictors are measured prior to the pandemic. Therefore, the association between BAA and the risk of COVID-19 (and probably other dangerous infectious diseases) is free of reverse causation (like in MR) and likely to be causal if there are no other confounders. Thus, our research supports the idea of the pro-active application of anti-aging (that is BAA-reducing) drugs in a prophylaxis mode to protect the biomarker-defined vulnerable individuals. And, reversely, a significant reduction of BA by an experimental drug in a clinical trial (probably as early as phase I) could warrant further clinical studies in elderly subjects.

The association of BAA with case fatality was weaker (only Phenotypic Age BAA exhibited a significant effect). This can be explained by the considerably smaller number of UKBB subjects involved in the statistical analysis (346 of dead individuals compared to 11,619 tested (and presumed sick) and 459,872 overall subjects in UKBB). The case fatality rate increases exponentially with age, and therefore it would be reasonable to expect the association of BAA with the risk of death in COVID-19 patients[3]. We expect future studies to corroborate our findings. Whether or not this association is causative could not be established in our study. Age-dependent severity of COVID-19 has been demonstrated in many epidemiological studies[3,46]. However, we did not observe a significant association between longevity and COVID-19 severity-related traits. This could be largely due to a small number of cases included in GWAS analyses, varying from 536 to 1,610. Moreover, there is an underlying selection bias toward symptomatic cases in the GWAS of COVID-19 positive cases, especially in the UKBB study, which is overrepresented by severe and hospitalized cases[30]. Therefore, our results on UKBB may also be interpreted as a protective effect of longevity on the severe form of COVID-19 infection.

It has been reported that people with comorbidities are more likely to suffer from COVID-19 and have poor prognosis[47]. The most prevalent comorbidities were hypertension, diabetes, and cardiovascular disease[3,47]. Based on our Mendelian randomization analysis, genetically predicted CAD, one of the most common types of cardiovascular disease, increases COVID-19 susceptibility and the chance of being hospitalized after infection at a nominal significance level. However, type 2 diabetes (T2D) did not show a potential causal effect on COVID-19-related traits. There are two possible explanations: (1) the observed link between T2D and COVID-19 in epidemiological studies is confounded by other factors; (2) the power of current COVID-19 GWAS results is limited to reveal its potential causal link to T2D. The

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

genetically predicted Alzheimer's disease (AD) was also shown to be significantly associated with a higher risk of infection. Besides, we observed protective effects of AD on COVID-19 severity and COVID-19 with respiratory failure at a nominal significance level. Although it seems counterintuitive that AD is protective against the severity of COVID-19, currently there is no published data on the risk of AD patients of having respiratory failure with COVID-19. Larger population-based studies are needed to address these questions.

There are multiple clinical trials proposed to employ potential lifespan-extending drugs to protect the elderly from COVID-19, based on promising observational data on metformin[5-9]. However, epidemiological studies are prone to confounding, reverse causation, and various biases, and therefore are an unreliable indicator of the causal associations. MR is a method that utilizes genetic instruments that are robustly associated with exposures, and thus generate more reliable evidence in predicting novel interventions[48]. In our MR study, we established a causal link between aging and COVID-19 infection, thus supporting the idea that lifespan-extending and/or biological age-reversing drugs, as a category, should be considered as a preventive measure in the elderly and prioritized in clinical trials.

Acknowledgements

XS was in receipt of a Swedish Research Council (Vetenskapsrådet) Starting Grant (No. 2017-02543). This research was also supported by NIA grants (to VNG). PF and TP were supported by Gero PTE LLC (Singapore). We thank HGI and NIH-GRASP for the timely release of COVID-19 GWAS summary statistics. We also thank Miss. Hanna Liu for assistance in part of the analysis.

Author contributions

XS initiated the study; VNG, XS, and POF supervised the study; KY, RZ, and TVP performed data analyses; KY, VNG, POF, and XS wrote the manuscript; All authors contributed to manuscript writing.

Competing interest statement

The authors declare no competing financial interests.

References

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).
2 Koff, W. C. & Williams, M. A. Covid-19 and Immunity in Aging Populations - A New Research Agenda. *N Engl J Med*, doi:10.1056/NEJMp2006761 (2020).
3 Santesmasses, D. *et al.* COVID-19 is an emergent disease of aging. *medRxiv*, doi:10.1101/2020.04.15.20060095 (2020).
4 Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*, 1-7, doi:10.1038/s41591-020-0962-9 (2020).
5 Sargiacomo, C., Sotgia, F. & Lisanti, M. P. COVID-19 and chronological aging: senolytics and other anti-aging drugs for the treatment or prevention of corona virus infection? *Aging (Albany NY)* **12**, 6511-6517, doi:10.18632/aging.103001 (2020).
6 Zhavoronkov, A. Geroprotective and senoremediative strategies to reduce the comorbidity, infection rates, severity, and lethality in gerophilic and gerolavic infections. *Aging (Albany NY)* **12**, 6492-6510, doi:10.18632/aging.102988 (2020).
7 Omarjee, L. *et al.* Targeting T-cell senescence and cytokine storm with rapamycin to prevent severe progression in COVID-19. *Clin Immunol* **216**, 108464, doi:10.1016/j.clim.2020.108464 (2020).
8 Bramante, C. *et al.* Observational Study of Metformin and Risk of Mortality in Patients Hospitalized with Covid-19. *medRxiv*, doi:10.1101/2020.06.19.20135095 (2020).
9 Luo, P. *et al.* Metformin Treatment Was Associated with Decreased Mortality in COVID-19 Patients with Diabetes in a Retrospective Analysis. *Am J Trop Med Hyg* **103**, 69-72, doi:10.4269/ajtmh.20-0375 (2020).
10 Roberts, R. Mendelian Randomization Studies Promise to Shorten the Journey to FDA Approval. *JACC Basic Transl Sci* **3**, 690-703, doi:10.1016/j.jacbts.2018.08.001 (2018).
11 Timmers, P. R. *et al.* Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife* **8**, doi:10.7554/eLife.39856 (2019).
12 Zenin, A. *et al.* Identification of 12 genetic loci associated with human healthspan. *Commun Biol* **2**, 41, doi:10.1038/s42003-019-0290-0 (2019).
13 Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* **10**, 3669, doi:10.1038/s41467-019-11558-2 (2019).
14 Timmers, P., Wilson, J. F., Joshi, P. K. & Deelen, J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat Commun* **11**, 3570, doi:10.1038/s41467-020-17312-3 (2020).
15 Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430, doi:10.1038/s41588-019-0358-2 (2019).
16 Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet* **49**, 1385-1391, doi:10.1038/ng.3913 (2017).
17 Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941, doi:10.1038/s41467-018-04951-w (2018).
18 Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244, doi:10.1038/s41588-018-0307-5 (2019).
19 McCartney, D. L. *et al.* Genome-wide association studies identify 137 loci for DNA methylation biomarkers of ageing. doi:10.1101/2020.06.29.133702 (2020).
20 Banda, Y. *et al.* Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1285-1295, doi:10.1534/genetics.115.178616 (2015).
21 Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* **9**, 224, doi:10.1038/s41467-017-02317-2 (2018).
22 Ellinghaus, D. *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med*, doi:10.1056/NEJMoa2020283 (2020).
23 Initiative, C.-H. G. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* **28**, 715-718, doi:10.1038/s41431-020-0636-6 (2020).
24 Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian Randomization. *JAMA* **318**, 1925-1926, doi:10.1001/jama.2017.17219 (2017).
25 Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat Genet*, 1-6, doi:10.1038/s41588-020-0653-y (2020).
26 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

27 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573-591, doi:10.18632/aging.101414 (2018).

28 Pyrkov, T. V. *et al.* Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts ultimate limit of human lifespan. *bioRxiv*, doi:10.1101/618876 (2020).

29 Pyrkov, T. V. *et al.* Quantitative characterization of biological age and frailty based on locomotor activity records. *Aging (Albany NY)* **10**, 2973-2990, doi:10.18632/aging.101603 (2018).

30 Griffith, G. *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. *medRxiv*, 2020.2005.2004.20090506, doi:10.1101/2020.05.04.20090506 (2020).

31 Li, X. *et al.* Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *Elife* **9**, doi:10.7554/eLife.51507 (2020).

32 F. Hillary, R. *et al. Epigenetic clocks predict prevalence and incidence of leading causes of death and disease burden* (Cold Spring Harbor Laboratory, 2020).

33 Goronzy, J. J. & Weyand, C. M. Understanding immunosenescence to improve responses to vaccines. *Nat Immunol* **14**, 428-436, doi:10.1038/ni.2588 (2013).

34 Alpert, A. *et al.* A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat Med* **25**, 487-495, doi:10.1038/s41591-019-0381-y (2019).

35 Hashimoto, K. *et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc Natl Acad Sci U S A* **116**, 24242-24251, doi:10.1073/pnas.1907883116 (2019).

36 Arai, Y. *et al.* in *EBioMedicine* Vol. 2 1549-1558 (2015).

37 Jylhava, J., Pedersen, N. L. & Hagg, S. Biological Age Predictors. *EBioMedicine* **21**, 29-36, doi:10.1016/j.ebiom.2017.03.046 (2017).

38 Lara, J. *et al.* A proposed panel of biomarkers of healthy ageing. *BMC Med* **13**, 222, doi:10.1186/s12916-015-0470-9 (2015).

39 Cardoso, A. L. *et al.* Towards frailty biomarkers: Candidates from genes and pathways regulated in aging and age-related diseases. *Ageing Res Rev* **47**, 214-277, doi:10.1016/j.arr.2018.07.004 (2018).

40 Kojima, G., Iliffe, S. & Walters, K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing* **47**, 193-200, doi:10.1093/ageing/afx162 (2018).

41 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

42 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).

43 Pyrkov, T. V. *et al.* Extracting biological age from biomedical data via deep learning: too much of a good thing? *Sci Rep* 8, 5210, doi:10.1038/s41598-018-23534-9 (2018).

44 Pyrkov, T. V., Fedichev, P. O. & Moskalev, A. Biological Age is a Universal Marker of Aging, Stress, and Frailty. *Healthy Ageing and Longevity* **10**, 23-36, doi:10.1007/978-3-030-24970-0_3 info:doi/10.1007/978-3-030-24970-0_3 (2019).

45 Tison, G. H. *et al.* Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Ann Intern Med*, doi:10.7326/M20-2665 (2020).

46 Bi, Q. *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis* **20**, 911-919, doi:10.1016/S1473-3099(20)30287-5 (2020).

47 Yang, J. *et al.* Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis* **94**, 91-95, doi:10.1016/j.ijid.2020.03.017 (2020).

48 Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* **23**, R89-98, doi:10.1093/hmg/ddu328 (2014).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tables and Figures

Table 1. Mendelian randomization estimates for the association between exposures and risk of COVID-19

Exposure	Outcome	OR	95% CI	Nominal P	Bonferroni Corrected P
Lifespan	UKBB_Covid_infection	0.32	0.18 - 0.57	1.3e-04	0.01600
	HGI_covid_susceptibility	0.53	0.32 - 0.9	1.7e-02	1.00000
	HGI_covid_vs_neg	0.61	0.39 - 0.97	3.5e-02	1.00000
Longevity	UKBB_Covid_infection	0.62	0.51 - 0.77	7.2e-06	0.00095
	HGI_covid_susceptibility	0.75	0.63 - 0.9	1.6e-03	0.20000
Combined	UKBB_Covid_infection	0.16	0.05 - 0.45	5.9e-04	0.07600
	HGI_covid_susceptibility	0.33	0.13 - 0.85	2.2e-02	1.00000
GrimAge_accel	UKBB_Covid_infection	1.25	1.01 - 1.55	4.1e-02	1.00000
CAD	UKBB_Covid_infection	1.19	1.03 - 1.39	2.0e-02	1.00000
	HGI_covid_susceptibility	1.26	1.1 - 1.44	7.2e-04	0.09200
	HGI_hosp_vs_not_hosp_covid	1.40	1.05 - 1.87	2.4e-02	1.00000
AD	UKBB_Covid_infection	1.13	1.06 - 1.2	1.6e-04	0.02100
	Covid_resp_failure	0.83	0.74 - 0.93	8.8e-04	0.11000
	HGI_covid_susceptibility	1.08	1.02 - 1.14	6.2e-03	0.78000
	HGI_covid_flu_like_symp	1.08	1.01 - 1.16	2.4e-02	1.00000
	HGI_hosp_vs_not_hosp_covid	0.87	0.77 - 0.98	1.8e-02	1.00000
	HGI_covid_vs_neg	1.05	1.01 - 1.1	2.6e-02	1.00000
	HGI_covid_predicted	1.06	1 - 1.12	3.9e-02	1.00000

Only the associations that reached nominal significance ($P < 0.05$) are shown.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Association between biological age acceleration and the risk of COVID-19

Biological age measurement	***Outcome***	***OR***	***95% CI***	***P***
Phenotypic Age	COVID19 incidence (All)	1.28	1.25 - 1.31	8.4e-82
	Case fatality (All)	1.19	1.04 - 1.35	1.1e-02
	COVID19 incidence (Not frail)	1.12	1.04 - 1.2	1.9e-03
	Case fatality (Not frail)	1.72	1.17 - 2.51	5.4e-03
	COVID19 incidence (Frail)	1.26	1.23 - 1.3	3.7e-62
DOSI	COVID19 incidence (All)	1.31	1.26 - 1.38	9.5e-32
	COVID19 incidence (Not frail)	1.09	1.01 - 1.19	3.6e-02
	Case fatality (Not frail)	2.44	1.45 - 4.06	7.7e-04
	COVID19 incidence (Frail)	1.35	1.28 - 1.42	3.6e-28
Physical Activity	COVID19 incidence (All)	0.95	0.93 - 0.96	9.1e-19

Only the associations that reached nominal significance ($P < 0.05$) are shown.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

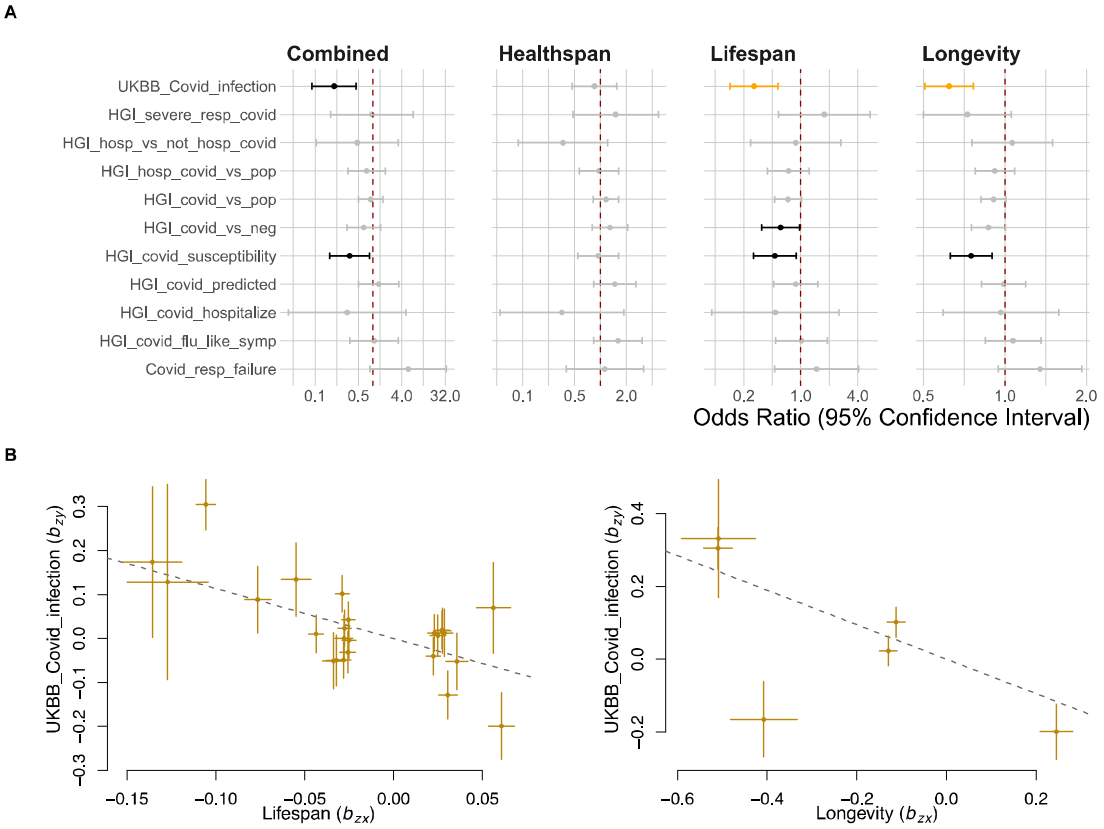

Figure 1. Mendelian randomization analysis investigating the association of genetically proxied lifespan-related traits with the risk of COVID-19. A. Forest plot showing Mendelian randomization estimates for the causal effect of lifespan-related traits on the risk of COVID-19. Error bars show the 95% confidential interval. Significant effects after correcting for 132 tests ($P < 0.05/132$) are in orange. Nominally significant effects ($P < 0.05$) are in black. **B.** Plots of effect sizes of all genetic instruments from GWAS for lifespan and longevity vs. those for UKBB COVID-19. Error bars represent standard errors.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

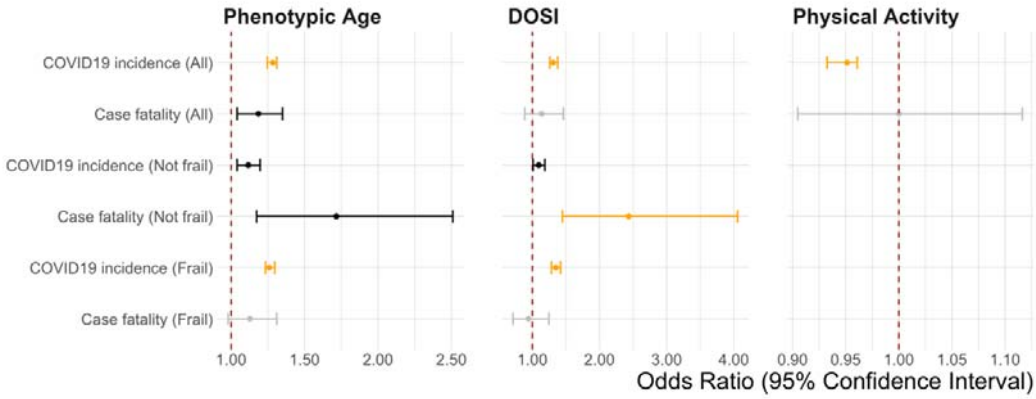

Figure 2. Analysis of association of biological age acceleration with the risk of COVID-19. Error bars show the 95% confidential interval. Significant effects ($P < 0.001$) are in orange. Nominally significant effects ($P < 0.05$) are in black. Odds ratio for Phenotypic Age and Dynamic Organism State Index (DOSI) is given per 10-yr biological age acceleration. Odds ratio for physical activity is given per increase of 1000 steps/day.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

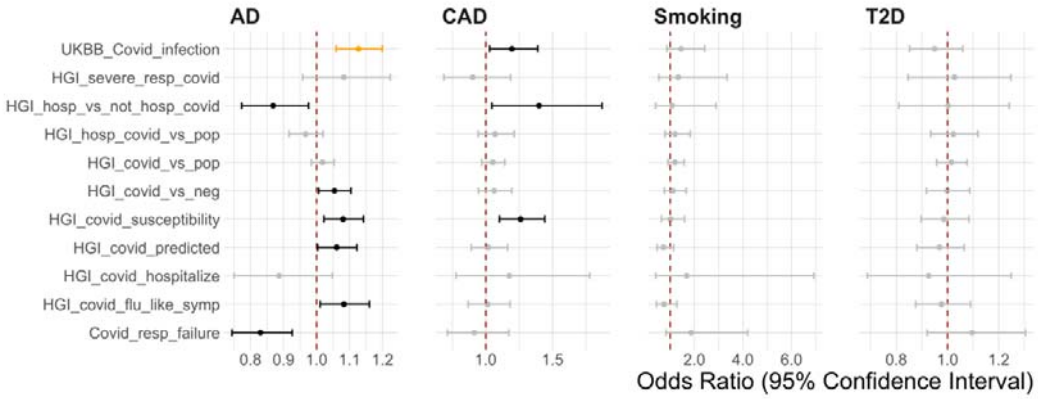

Figure 3. Mendelian randomization analysis investigating the association of genetically proxied age-related diseases with the risk of COVID-19. Error bars show the 95% confidential interval. Significant effects after correcting for 132 tests ($P < 0.05/132$) are in orange. Nominally significant effects ($P < 0.05$) are in black.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20169854; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

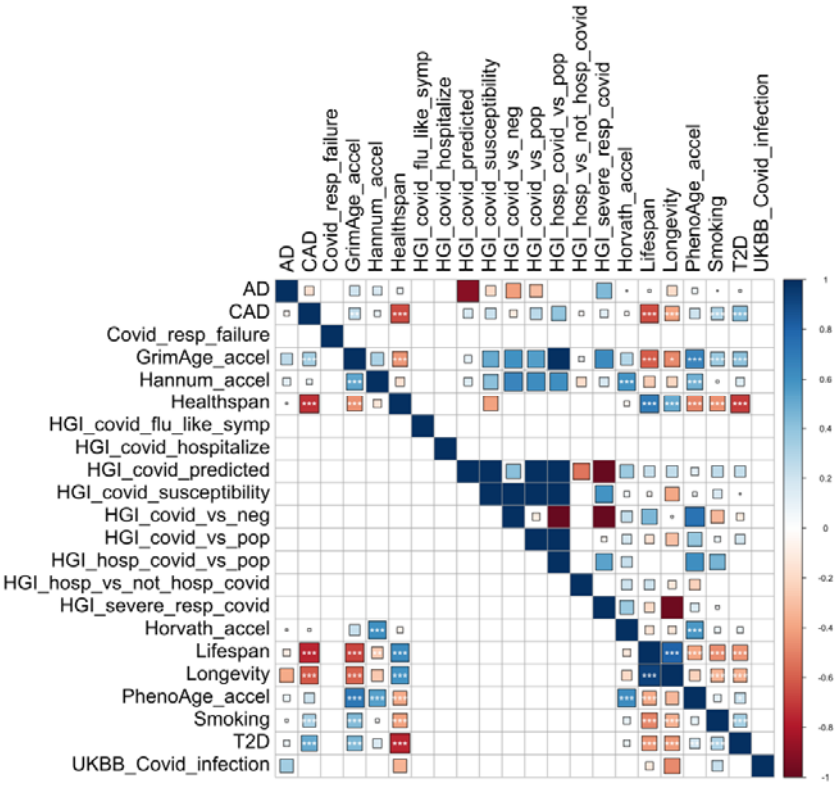

Figure 4. Genetic correlation estimates from HDL and LDSC among phenotypes. Lower triangle: HDL estimates; upper triangle: LDSC estimate. the areas of the squares represent the absolute value of corresponding genetic correlations. The genetic correlation that couldn't be estimated are in blank. P values are corrected using Bonferroni correction for 231 tests, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.