

METHODS

Probabilistic assignment of variants into susceptibility vs. severity effects

We consider an assignment of variants reported in Supplementary Table 3 into two groups based on whether they seem to affect susceptibility to corona virus infection (SUC) or severity of the infection (SEV). For this analysis, we meta-analyzed GWAS summary statistics from such versions of B2 (hospitalized covid vs. population) and C2 (infected vs. population) GWAS that only included those studies that had contributed some data to B2. Therefore, all studies included in this analysis had made some effort to distinguish from all the infected individuals those individuals who additionally were hospitalized. The sample sizes of these GWAS were 23,988 cases / 2,834,885 controls for B2 and 114,516 cases / 2,138,237 controls for C2.

We assume that all hospitalized cases (B2 cases) were included among the infected cases (C2 cases) of the corresponding study and that, for each study, the controls of B2 and C2 had a maximum overlap possible given the control counts in the two data sets.

We use a Bayesian framework to compare two models, SUC and SEV, at each SNP. Briefly, we applied priors on effect size, $N(0.0.1^2)$, were integrated out and since we have equal prior of 50% on each model we are essentially just comparing the marginal likelihood of data under two models to get the posterior probability. Intuitively, model SUC represents a variant that associates with susceptibility of infection and hence its effect size is similar in B2 and C2, i.e., $\beta_{C2} \approx \beta_{B2}$. Model SEV represents a variant that affects severity of infection but has little effect on susceptibility to infection. We expect that, for such variants, the effect size in C2 is proportional to its effect in B2, i.e., $\beta_{C2} \approx w_{SEV}\beta_{B2}$, where the constant of proportionality, $w_{SEV} < 1$, depends on the proportion of all C2 cases that are also B2 cases. If we consider C2 analysis as a fixed-effect meta-analysis between B2 and an imaginary “non-hospitalized covid vs.

population” analysis that had no sample overlap with B2 analysis, then $w_{SEV} = \frac{n_{B2}^{(eff)}}{n_{C2}^{(eff)}}$, where

$n_i^{(eff)} = R_i S_i / N_i$ is the effective sample size of study i with R_i the number of controls, S_i the number of cases and $N_i = R_i + S_i$. In our data, $\frac{n_{B2}^{(eff)}}{n_{C2}^{(eff)}} \approx 0.208$, where the effective sample sizes

are computed by summing the effective sample sizes over individual studies of B2 and C2 analyses. Even when controls of B2 analysis and the imaginary “non-hospitalized covid vs. population” analysis overlapped completely, the value of w_{SEV} changes little in these data. After accounting for the overlap in controls, we estimated a value of $w_{SEV} = 0.20$ that we used in the analyses described below.

To derive correlation r_{B2C2} between the estimation errors of the effect size estimators between B2 and C2 analyses, we used formula

$$r_{B2C2} = \frac{\sum_{k=1}^K \sqrt{n_{B2,k}^{(eff)} n_{C2,k}^{(eff)}} r_{B2C2,k}}{\sqrt{\sum_{k=1}^K n_{B2,k}^{(eff)} \sum_{k=1}^K n_{C2,k}^{(eff)}}}$$

where subscript k refers to individual studies. The correlation $r_{B2C2,k}$ between B2 and C2 for study k is computed as in Bhattacharjee et al. (2012)¹:

$$r_{B2C2,k} = \sqrt{n_{B2,k}^{(eff)} n_{C2,k}^{(eff)}} \left(\frac{S_{B2C2,k}}{S_{B2,k} S_{C2,k}} + \frac{R_{B2C2,k}}{R_{B2,k} R_{C2,k}} \right)$$

where $S_{B2C2,k}$ is the number of shared B2 and C2 cases in study k and similarly $R_{B2C2,k}$ is the number of shared controls. By applying this to the data, we estimated $r_{B2C2} = 0.45$.

With these estimates of $w_{SEV} = 0.2$ and $r_{B2C2} = 0.45$, and with the observed data at any one SNP containing effect estimates $(\hat{\beta}_{B2}, \hat{\beta}_{C2})$ and their standard errors (s_{B2}, s_{C2}) , we derive the two models, SUC and SEV, as follows.

Prior distribution for effects is zero-centered bivariate normal distribution

$$\begin{pmatrix} \beta_{B2} \\ \beta_{C2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \theta_i \right), \text{ with}$$

$$\theta_{SUC} = \tau^2 (1 \ 1 - \eta_{SUC} \ 1 - \eta_{SUC} \ 1) \text{ and}$$

$$\theta_{SEV} = \tau^2 (1 \ (1 - \eta_{SEV}) w_{SEV} \ (1 - \eta_{SEV}) w_{SEV} \ w_{SEV}^2).$$

We have used value $\tau = 0.1$ to define the expected effect sizes of the B2 analysis (implying roughly that 95% of the true effect sizes of the risk variants have odds-ratio below 1.2). By tuning the parameters η_{SUC} and η_{SEV} , we can define how much deviation real effects can have from the lines $\beta_{C2} = \beta_{B2}$ and $\beta_{C2} = w_{SEV} \beta_{B2}$ corresponding to models SUC and SEV, respectively. We have set these values in such a way that, under both models, the mean (Euclidean) distance between the effect size and the corresponding line is 0.0025 and 95% of effects are within 0.006 units from the line. This happens when $\eta_{SUC} = 0.001$ and $\eta_{SEV} = 0.013$.

The likelihood for the observed data under both models is Gaussian

$$\begin{pmatrix} \hat{\beta}_{B2} \\ \hat{\beta}_{C2} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_{B2} \\ \beta_{C2} \end{pmatrix}, \Sigma \right), \text{ where } \Sigma = (s_{B2}^2 \ s_{B2} s_{C2} r_{B2C2} \ s_{B2} s_{C2} r_{B2C2} \ s_{C2}^2).$$

It follows from Trochet et al. (2019)² that the marginal likelihood for each model is proportional to a Gaussian density function evaluated at the observed effect size estimates as

$$Pr Pr (DATA | Model i) \propto f_N \left(\begin{pmatrix} \hat{\beta}_{B2} \\ \hat{\beta}_{C2} \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \theta_i + \Sigma \right).$$

We set equal prior probability on each model (i.e. 50% on SUC and 50% on SEV), and consequently the posterior probabilities of the models will be proportional to their marginal likelihoods. These posterior probabilities are reported in Supplementary Table 3

Gene prioritization and Phenome wide association study (PheWAS)

The variant annotation and PheWAS were performed as per methods described in Niemi et. al 2021.

Genetic Correlations and Mendelian Randomization

The genetic correlation, heritability estimates, and Mendelian Randomization were performed as per methods described in Niemi et. al 2021³.

SNP heritability for all three COVID-19 related phenotypes was low (<1% on the observed scale; Supplementary Table 7). To understand which traits are genetically correlated and/or potentially causally related to the three phenotypes, we first estimated genetic correlations with 38 traits (Supplementary Table 8). In addition to what was previously reported, we found positive genetic correlations with two risk factors (depression and insomnia symptoms), two biomarkers (C-reactive protein (CRP) and 25-hydroxyvitamin D), and two disease liability traits (asthma and heart failure).

We next applied two-sample Mendelian Randomization (MR) to infer potential causal relationships between COVID-19 related phenotypes and their genetically correlated traits. Four traits (BMI, type II diabetes, red blood cell count, and height) showed evidence of causal associations after correcting for multiple testing and were robust to potential violations of the underlying assumptions of MR (Supplementary Table 9).

Multivariable Mendelian Randomization (MVMR) was used to estimate the direct effects of body mass index (BMI) and type II diabetes (T2D) on the risk of SARS-CoV-2 phenotypes, by including both exposures within the same model. We selected all independent ($r^2 = 0.001$; $kb = 10000$) genome-wide significant ($p < 5 \times 10^{-8}$) SNPs associated with BMI and type II diabetes; and using the full list of SNPs associated with both BMI and T2D, performed a second clumping procedure ($r^2 = 0.001$; $kb = 10000$) to obtain independent SNPs. After clumping the full list of SNPs from both BMI and T2D and restricting to SNPs found in the SARS-CoV-2 GWASs, a total of 721 SNPs were available for multivariable MR (516 were associated with BMI only, 201 were associated with T2D only, and 4 SNPs overlap between both GWAS). MVMR was performed using the “MVMR” package [1]. The sensitivity analyses conducted using “MVMR” require estimates of the pairwise covariances between each instrument and each exposure, as such, we used the phenotypic correlation between BMI and T2D and summary data to generate estimates of the covariances. Phenotypic correlations between BMI and T2D were estimated from the LDSC regression intercept ($r^2 = 0.13$) [3]. Next, we calculated conditional F-statistics to

evaluate the presence of weak instruments. The conditional F-statistic for both BMI and T2D were >10 , indicating that the selected instruments were strongly associated with their corresponding exposure (Supplementary Table 10b) [1,2]. We then assessed multivariable instrument pleiotropy using the modified Cochran's Q-statistic that accounts for potential weak instrument bias, with evidence of heterogeneity indicative of a violation in the escalation restriction assumption in MR [1,2]. For each of the SARS-CoV-2 phenotypes MVMR models there was evidence of heterogeneity, suggesting the causal estimates from IVW-MVMR may be biased due to the presence of horizontal pleiotropy (Supplementary Table 10c). Inverse weighted multivariable MR model was used to estimate the direct effect of BMI and T2D upon each of the SARS-CoV-2 phenotypes (Supplementary Table 10). When MVMR assumptions are violated, as indicated by the presence of weak instruments or horizontal pleiotropy, it is possible to obtain more robust causal estimates Q- statistic to minimization (Supplementary Table 10).

Code for implementing the MVMR analysis is available at:

https://github.com/marcoralab/multivariate_MR

References

- [1] Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* 48, 713–727. <https://doi.org/10.1093/ije/dyy262> (2019)
- [2] Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat Med* (2021) doi:10.1002/sim.9133.
- [3] Zheng, J., Richardson, T. G., Millard, L., Hemani, G., Elsworth, B. L., Raistrick, C. A., Vilhjalmsson, B., Neale, B. M., Haycock, P. C., Smith, G. D., & Gaunt, T. R. (2018). PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. *GigaScience*, 7(8), giy090. <https://doi.org/10.1093/gigascience/giy090>