Title: Multi-omics study revealing putative drug targets of COVID-19 severity and other viral infection diseases

Authors: Jie Zheng[1†]*, Yuemiao Zhang[2†]*, Yi Liu[1], Denis Baird[1], Mohd Anisul Karim[3,4], Maya Ghoussaini[3,4], Jeremy Schwartzentruber[3,4], Ian Dunham[3,4,5], Benjamin Elsworth[1], Katherine Roberts[6], Hannah Compton[6], Felix Miller-Molloy[6], Xingzi Liu[2], Lin Wang[7], Hong Zhang[2], George Davey Smith[1,8], Tom R Gaunt[1,8]

Affiliations:

[1] MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom

[2] Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China, Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Beijing, People's Republic of China.

[3] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

[4] Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

[5] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

[6] Bristol Medical School, University of Bristol, 5 Tyndall Avenue, Bristol, BS8 1UD, United Kingdom.

[7] Department of Microbiology and Infectious Disease Centre, School of Basic Medical Sciences, Peking University Health Science Centre, Beijing, China.

[8] NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol, United Kingdom.

* To whom correspondence should be addressed:

Jie Zheng: jie.zheng@bristol.ac.uk

Yuemiao Zhang: zhangyuemiao@bjmu.edu.cn

† Equal contribution

One Sentence Summary:

Integrating multi-omic evidence to prioritise potential drug targets and identify molecular mechanisms for COVID-19 severity.

Abstract:
Drug target prioritization for new targets and drug repurposing of existing drugs for COVID-19 treatment are urgently needed for the current pandemic. COVID-19 drugs targeting human proteins will potentially relate to different host responses involving multiple molecular pathways. Here we implemented a novel COVID-19 drug target prioritisation protocol by integrating multi-omics Mendelian randomization (MR) of COVID-19 severity with colocalization. By integrating novel MR and existing evidence, we prioritised 353 candidate drug targets with clinical and/or literature evidence of host-coronavirus interaction and one additional target (ENTPD5) with robust MR and colocalization evidence on COVID-19 severity. We further conducted phenome-wide MR of these prioritised targets on 622 complex diseases in 11 SARS-CoV-2 related tissues, which we identified 726 potential causal effects on other diseases, providing information on potential beneficial and adverse effects. This study demonstrated the value of our integrative approach in prioritising drug targets and mechanisms involved in COVID−19 severity. This provides the first steps towards evaluating intervention targets worthy of follow-up for coronavirus and viral infections.

Introduction

The outbreak of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of novel coronavirus disease 2019 (COVID-19), has been a global pandemic. It is expected to infect a large portion of the world population with a wide spectrum of manifestations, ranging from completely asymptomatic carriers (1) to critical respiratory failure, multi-organ dysfunction, and death (2). This heterogeneity may relate to different host responses involving multiple human proteins and pathways. Considering the relatively high mortality rate of patients with severe disease (3), search for optimal treatments, such as novel drugs to improve survival rate of severe patients is a key objective to reduce the impact of both the current and potential future coronavirus epidemics.

Currently, both clinical trials and experimental studies are ongoing for developing new drugs or repurposing existing drugs for COVID-19. For example, 1,444 clinical trials are currently at status of recruiting or completed (ClinicalTrial.gov searched on December 6th using COVID-19), including remdesivir (4). In parallel with these drug trials, one recent study identified a set of human proteins interacting with 26 SARS-CoV-2 proteins using affinity purification-mass spectrometry in HEK293 cell line by determining the host proteins which interact with SARS-CoV-2 (5). Another study has identified 44 genes associated with SARS-CoV, a closely related coronavirus with SARS-CoV-2, using mouse models (6), which may provide drug targets for general coronavirus interventions. However, in this rapidly developing situation, apart from some early stage trials (7) and observational studies (8), only dexamethasone has been reported to reduce the death rate significantly in severe hospitalized COVID-19 patients in a large randomized controlled trial - RECOVERY (9). Till 2 October 2020, nearly 318 of 405 therapeutic drugs in development for COVID-19 were tested in clinical trials, but their efficacy is as yet unproven (10)(11).

Genetic, transcriptomic and proteomic ("omics") data could play a role in supporting efficient drug development for various diseases including COVID-19. Recent studies have begun to support the role of genetics in predicting drug trial success (12). Some multi-omics studies have further demonstrated the value of molecular quantitative trait locus (QTL) studies in repurposing existing targets to additional indications as well as prioritizing novel drug targets (13)(14). One approach to utilising molecular QTL data is Mendelian randomization (MR) (15). MR uses genetic variants as instrumental variables to estimate the effect of an exposure (e.g. measured levels of a protein or a gene transcript) on an outcome (e.g. COVID-19 severity) (16), which could prioritise drug targets cost-effectively. MR can also be applied to identify potential mediators (e.g. coagulation), linking the exposure (e.g. a protein) with the outcome (e.g. COVID-19) in a mediation/two-step MR framework (17). In addition, phenome-wide MR could be applied to identify potential drug repurposing opportunities and adverse effects (18).

Recent GWASs (19)(20)(21) and MR studies (22)(23)(24)(25) utilising data from the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/) and GenOMICC consortium have identified a set of genes (e.g. *ABO*, *OAS1*, *IFNAR2*, *IL10RB*) associated with various COVID-19 phenotypes. However, there are issues that require consideration: i) some genes and proteins (e.g. ABO) are well-known to be pleiotropic, which may violate the "no pleiotropy" (or "exclusion restriction") assumption of MR; ii) when comparing severe COVID-19 cases with population controls as an outcome, it is not possible to separate the causal effects of becoming infected from

any causal effects on disease progression after infection (despite these potentially being separate mechanisms) (26); iii) collider bias (also known as selection bias, sampling bias or ascertainment bias) (27). Each of these could induce spurious associations between the target and COVID-19. Careful instrument and outcome selection are needed for the drug target MR of COVID-19.

To prioritise potential drug targets and molecular mechanisms for COVID-19 severity as well as identify potential beneficial and adverse effects, we combined genetic association information of 1,002 proteins and 15,944 transcripts from eight studies (28)(29)(30)(31)(32)(33)(34)(35) and applied two-sample MR (36) and colocalization analyses (37) to estimate the causal effects of these targets on COVID-19 severity and 622 complex human traits. To enable rapid queries, results of all analyses are available in an open access online platform (https://epigraphdb.org/covid-19/ctda/).

Results

Summary of the study

Figure 1 illustrates the design of this study: i) selection of drug targets potentially relevant to COVID-19 from five resources; ii) selection of the COVID-19 severity data and other diseases data; iii) MR analysis of the selected targets on COVID-19 severity; iv) phenome-wide association study of prioritised targets to identify potential beneficial and adverse effects of these targets on 622 complex human traits; v) follow-up mediation/two-step and bidirectional MR analyses of the prioritised targets with MR and colocalization evidence.

Selection of drug targets for COVID-19

We retrieved drug targets from two resources: i) By using the instrument selection tier system described by Zheng et al (13), we selected 1,002 protein targets with robust genetic predictors and less concerns about pleiotropy (Tier 1 or 2 instruments) from five studies (28)(29)(30)(31)(32) (**Supplementary Table 1**); ii) we selected 16,059 genes with robust genetic instruments to predict their gene expression levels using data from the eQTLGen consortium (33) (**Supplementary Table 2**). To further prioritise potential drug targets for COVID-19, we integrated evidence from three resources: i) target genes of 11 drugs under trials for COVID-19 treatment from *ClinicalTrials.gov* (**Supplementary Table 3**); ii) 332 human proteins interacting with SARS-CoV-2 proteins in human cell lines (5); iii) 44 genes associated with SARS-CoV in a mouse model (6). After removing duplicates, 380 unique targets were selected (**Supplementary Table 4**), of which 353 targets have genetic predictors (instruments) in either blood or in any of the 11 SARS-CoV-2 related tissues (data from Zheng et al, eQTLGen and GTEX, **Supplementary Table 5**). For these targets, both non-pleiotropic cis and trans instruments were considered to maximise the number of targets we included in the MR analyses.

Characteristics of COVID-19 severity

The genetic association data for COVID-19 was obtained from the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/). In total, GWASs of six COVID-19 related phenotypes were provided by the initiative. We excluded three COVID-19 GWASs using population as controls and two additional GWASs using self-reported data as either cases or controls, since these comparisons don't distinguish between asymptomatic infected, exposed but uninfected and unexposed. Using these data as outcomes for MR, we will not be able to separate the causal

effects of exposure to virus, infection and progression of COVID-19 after infection (which may represent separate mechanisms) (**Supplementary Figure 1**). Even after excluding these studies, we note that collider bias (also known as selection bias, sampling bias or ascertainment bias) could influence our understanding of COVID-19 disease risk and severity (27). Collider bias can induce spurious associations between the exposure (e.g. levels of a target) and the outcome (e.g. COVID-19 severity) (38). For example, COVID-19 testing could be a collider (common consequence) of both asthma and COVID-19 severity, since people with respiratory illnesses may be more likely to seek COVID-19 testing, as are people with more severe symptoms of COVID-19. When we sample individuals based on the collider (COVID-19 testing), that is equivalent to conditioning on the collider (26). This could induce a spurious association between asthma-associated genetic variants (and protein levels) and severity of COVID-19 (**Supplementary Figure 2**). Despite this limitation, we nevertheless used the GWAS of hospitalized COVID-19 vs. not hospitalized COVID-19 (928 confirmed hospitalized COVID-19 patients, 2,028 confirmed non hospitalised COVID-19 patients**)** as the dataset to represent COVID-19 severity in our study, since it provides the most meaningful comparison to explore effects on severity. The summary statistics of the genetic associations with COVID-19 severity were used as the outcome dataset for the MR analysis (**Supplementary Table 6**).

Estimated causal effects of targets on COVID-19 severity

We conducted two-sample MR (36) to comprehensively estimate the causal effects of 1,002 plasma proteins and 16,059 transcripts on COVID-19 severity (**Supplementary Figure 1**). For the MR analysis of 1,002 plasma proteins, we were able to successfully perform MR on COVID-19 severity for 837 of them due to missingness of genetic variants in the COVID-19 GWAS (Bonferroni corrected $P = 5.97\times10^{-5}$). The protein level of ENTPD5 (represented by a cis protein QTL) showed a positive MR association with COVID-19 severity after multiple testing correction (odds ratio [OR] of COVID-19 severity per standard deviation change of protein level= 2.07, 95%CI=1.47 to 2.92, $P=3.29\times10^{-5}$; **Supplementary Table 7**). For the MR analysis of 16,059 transcripts, 14,036 of could be analysed with COVID-19 severity using MR (Bonferroni corrected $P = 3.56\times10^{-6}$). None of the transcript MR results showed robust MR evidence after correction for multiple testing (best MR p value= 4.17×10^{-5}, **Supplementary Table 8**).

For the 353 prioritised targets with experimental or trial evidence, all of them showed no strong MR evidence of a causal effect on COVID-19 severity. Only expression of 10 genes (*AATF, NEU1, NUP210, PIGS, PLOD2, PMPCA, RETREG3, SPART, SRP72, TLE3*) and levels of one protein (NPC2) showed nominal association on COVID-19 severity (we used a lenient P value threshold of $P < 0.05$, maximizing the number of possible genes analysed but also allowing readers to filter out associations should they wish to apply a more stringent threshold, **Supplementary Table 9**). All of these 11 targets were reported to interact with SARS-CoV-2 proteins in human cell lines (**Supplementary Table 4**).

Genetically predicted associations between a protein or gene and COVID-19 severity may arise any of the four scenarios: causality, reverse causality, confounding by linkage disequilibrium (LD) between the leading SNPs for proteins and phenotypes or horizontal pleiotropy (**Supplementary Figure 3**). For the target, ENPTD5, with strong MR evidence of a causal effect on COVID-19 severity, we conducted single-trait colocalization (37) to distinguish causality

from confounding by LD. This analysis showed strong evidence (colocalization probability=99.9%) of a shared genetic effect between ENPTD5 protein levels and COVID-19 severity (**Supplementary Table 10A**). In addition, we estimated the causal effects of COVID-19 severity on protein expression level of ENTPD5 using bidirectional MR (39) to explore the influence of potential reverse causality (with two conditional independent COVID-19 severity associated SNPs [$P<1\times10^{-6}$] used as instruments and the protein level of ENTPD5 used as outcome). This analysis showed little MR evidence to support reverse causality (OR= 0.94, 95%CI=0.81 to 1.11, P= 0.47; **Supplementary Table 10B**). In addition, the phenome-wide MR analysis (MR-PheWAS) of ENTPD5 provided no strong MR evidence of protein level of ENTPD5 associated with any of the 622 tested diseases or infection traits at a Bonferroni-corrected threshold of 8.04×10^{-5} (lowest MR P value=1.41×10^{-4}; **Figure 3**), which suggested that horizontal pleiotropy is less likely to be an issue for this target.

Estimated beneficial and adverse effects of the prioritised targets on other complex diseases

To identify potential beneficial and/or adverse effects of ENTPD5 (prioritised by omics MR of COVID-19 severity) and the 353 additional targets (prioritised by literature evidence), we conducted MR-PheWAS (18) and colocalization (37) of the protein and gene expression level of these targets on 622 traits, which including 49 viral infection phenotypes, 501 human diseases and 72 disease related risk factors (e.g. blood lipids) (**Supplementary Table 11**). 45,938 target-disease associations were tested in plasma proteome and/or transcriptome in whole blood ($P<1.09\times10^{-6}$ at a Bonferroni-corrected threshold). Where data was available, we also tested the tissue-specific effects of gene expression of the same targets on the outcome phenotypes using data from GTEX. Overall, 372,830 target-disease associations were estimated in the 11 COVID-19 relevant tissues (see the list of tissues in Methods).

As mentioned above, protein level of ENTPD5 was not strongly associated with any of the 622 traits. Considering a more inclusive threshold (a lenient P value threshold of P<0.05), the MR-PheWAS suggested that reducing protein levels of ENTPD5 may increase the risk of lipid metabolism disorders (OR=0.93, 95%CI=0.90 to 0.97, $P=1.41\times10^{-4}$) and may increase the risk of joint disorders such as arthropathies (OR= 0.94, 95%CI=0.91 to 0.98, P= 0.002; **Figure 3** and **Supplementary Table 12A**). The MR-PheWAS also showed that genetically predicted ENTPD5 expression level was not associated with any disease (e.g. asthma) that might influence likelihood of seeking COVID-19 testing (the collider) (see **Supplementary Figure 2** for more information). With these analyses we did not find evidence that the MR result of expression level of ENTPD5 on COVID-19 severity was influenced by collider bias, although this does not exclude it as a possibility.

For the 353 targets prioritised by literature evidence, we observed 833 target-disease associations with robust MR evidence in the 11 tested tissues. Using the same omics data as the MR analysis, 726 of the 833 (87.2%) associations also showed strong colocalization evidence (colocalization probability > 70%) (**Supplementary Table 12B**), making these as more reliable findings of this study. Of these, 366 associations were obtained using a single cis instrument in a Wald ratio model (40), 327 were obtained using a single trans instrument and 33 were estimated using multiple instruments in an inverse variance weighted (IVW) model (36). The remaining 107 (12.8%) associations had evidence from MR but did not have strong evidence of colocalization

(probability<70%; **Supplementary Table 12C**), emphasizing the importance of this approach to address confounding by linkage disequilibrium (LD) in phenome-wide association studies.

Integration of evidence for the prioritised drug targets

We further integrated the 726 target-disease MR associations with clinical trial information from Open Targets (41), ChEMBL (42), DrugBank (43), Drug-gene-interaction (DGI) (44), the druggable genome (45) and ClinicalTrial.gov to prioritise drug targets against COVID-19. Of the 726 target-disease associations, 499 were unique target-disease pairs (due to the same pair appearing in multiple tissues; **Supplementary Table 13**). Further integrating these trial and literature evidence with the target-COVID-19 severity associations estimated in this study, we found that all targets showed no strong MR evidence of a causal effect on COVID-19 severity. Only 11 of them, including *PLOD2*, showed nominal effects. *PLOD2* gene is the target for three approved drugs: sirolimus, ribavirin and apicidin (**Supplementary Table 13**). The direction of effect between gene expression of *PLOD2* and disease risk was consistent with the therapeutic mechanism of these drugs. The MR-PheWAS of gene expression of *PLOD2* suggested that the target may be associated with a few blood cell and lipid traits but is not strongly associated with any disease outcome. For the same reason explained in the ENPTD5 example, the MR-PheWAS did not show clear evidence of collider bias influencing our MR of gene expression of *PLOD2* on COVID-19 severity.

Follow-up analyses of the prioritised drug target – ENTPD5

Given the potential causal effect of protein expression level of ENTPD5 on COVID-19 severity was supported by both MR and colocalization analyses, we further investigated this target with some additional follow-up analyses.

Shared causal effect among ENTPD5 gene expression and protein levels on COVID-19 severity

To further confirmed the ENTPD5 finding using plasma proteome data, we estimated the causal effect of the gene expression level of *ENTPD5* on COVID-19 severity using data from eQTLGen and observed strong MR ($P=5.66\times10^{-5}$) and colocalization (probability=99.9%) evidence (**Supplementary Table 10C**). We then estimated the shared causal effect among gene and protein expression of ENTPD5 and COVID-19 severity using multi-trait colocalization analysis (46). We observed robust colocalization evidence to support shared genetic aetiology of the three traits (shared probability=95.4%; **Figure 2**; **Supplementary Table 10D**), which further strengthens the evidence level of the association between ENTPD5 and COVID-19 severity.

Causal mediator linking ENTPD5 with COVID-19 severity

We further investigated potential mediators linking protein level of ENTPD5 with COVID-19 using two-step MR (17). We first selected candidate mediators using phenome-wide association study (PheWAS) of the protein (rs57731447) and gene expression (rs118113209) instruments of ENTPD5 (using the IEU OpenGWAS database (47)). Four modifiable traits associated with the protein and/or gene expression instruments of ENTPD5 were selected as candidate mediators (HDL cholesterol, C-reactive protein, coagulation factor VIII and body fat percentage; **Supplementary Table 14**). We then conducted a two-step MR to link ENTPD5 with the mediators and COVID-19 severity (**Supplementary Figure 4**). In the first step of the two-step MR, we formally estimated the effect of gene and protein expression of ENPTD5 on mediators,

confirmed the causal effects of protein level of ENTPD5 on these four candidate mediators (**Supplementary Table 15A**). In step 2 of the two-step MR, we estimated the effect of these four candidate mediators on COVID-19 severity but found little evidence to support their causal relationships (**Supplementary Table 15B**). We thus extended step 2 of the two-step MR by including traits similar to the four candidate mediators, which included lipid traits and lipoproteins, coagulation traits as well as traits related to body weight. We observed good evidence to support a causal role of increased D-dimer on increasing COVID-19 severity (OR=2.36, 95%CI=1.03 to 3.69, P value=0.001) as well as suggestive evidence of increased platelet count on increasing COVID-19 severity (OR=1.46, 95%CI=1.04 to 2.04, P value=0.027; **Supplementary Table 15C**). Moreover, we observed a protein-protein interaction between coagulation factor VIII (F8) and D-dimer (FGA/FGB/FGG) (data from StringDB implemented in EpiGraphDB). Although we did not identify a single mediator in this analysis, the two-step MR result and protein-protein interaction information suggested that coagulation may be involved in the causal effect of ENPTD5 protein expression on COVID-19 severity.

In addition, to explore the influence of potential reverse causality, we estimated the causal effects of the mediators on protein level of ENTPD5 using bidirectional MR (39) (with the mediators used as exposures and the protein expression level of ENTPD5 used as outcome) and found little evidence to support such reverse relationships, except for triglyceride levels showing evidence of a negative effect on ENTPD5 protein expression levels (**Supplementary Table 15D**).

Discussion

Genetic, transcriptomic and proteomic data provide a cost-effective approach to prioritise drug targets in the early stages of drug development (13)(28). In this study, we applied multi-omics MR, two-step MR and MR-PheWAS methods in a comprehensive in silico pipeline to prioritise drug targets for COVID-19 severity. Our multi-omics MR and colocalization study of 16,059 transcripts and 1,002 proteins identified putative causal effect of protein level of ENTPD5 on COVID-19 severity. Follow-up two-step MR analyses suggested that coagulation may be involved in the association between ENPTD5 protein levels and COVID-19 severity. Further integrating trial and experimental evidence, we investigated 353 targets for COVID-19. The MR-PheWAS of these 353 targets showed 726 target-disease associations with strong MR and colocalization evidence. However, MR of the gene or protein expression levels of these targets showed no strong evidence of a causal effect on COVID-19 severity. Moreover, the drug targets prioritised in this study were expected to offer therapeutic mechanisms through changing human proteins rather than SARS-CoV-2 itself. We therefore seriously considered the unintended beneficial or adverse effects of these targets on other complex diseases using MR-PheWAS. Our MR-PheWAS analysis suggested that the prioritised targets are unlikely to have a major adverse effect on complex human diseases. To enable the evidence to be widely accessible for COVID-19 drug target research, we constructed an open access browser allowing rapid queries of the drug targets, genetic predictors of the target genes and the target-COVID-19 associations (http://epigraphdb.org/covid-19/ctda/).

The *ENTPD5* gene, located on chromosome 14, encodes the NTPase Ectonucleoside Triphosphate Diphosphohydrolase 5, an enzyme mediating extracellular catabolism. A specific role of ENTPD5 is to promote host protein N-glycosylation for proper protein folding (information from Open Targets). SARS-CoV-2 spike protein is extensively N-glycosylated (48) and blocking viral N-glycan biosynthesis was shown to inhibit viral entry (49). Taylor et al. suggested *ENTPD5* as a gene associated with severe COVID-19 (50). In this study, our multi-omics MR and colocalization analyses suggested that both gene and protein expression levels of *ENTPD5* showed a positive effects on COVID-19 severity. The existing evidence plus our MR findings suggested that future studies are needed to examine the direct role of ENTPD5 inhibition on SARS-CoV-2 viral entry and, subsequently, in the prevention of severe COVID-19. In addition, the follow-up two-step MR analysis did not identify any single mediator but suggested that coagulation dysfunction could be a potential mechanism of ENTPD5 on COVID-19 severity. In the literature, coagulation dysfunction was reported to be observed in hospitalized / severe COVID-19 patients (51)(52)(53)(54), including mild thrombocytopenia (55), increased D-dimer levels (56) and increased thromboembolic events (57)(58). In a candidate-driven genetic association study of severe COVID-19, genetic loci known to be associated with coagulation have been identified, which revealed that SARS-CoV-2 engages activation of coagulation cascades (59). An association between anticoagulation and improved survival was observed in patients with severe COVID-19 even after adjusting for mechanical ventilation (60)(61). Till the 19th of December 2020, 25 randomized controlled trials studying the effect of anticoagulation agents (i.e. Bivalirudin, Enoxaparin, Rivaroxaban and Aspirin) on COVID-19 are under investigation (https://clinicaltrials.gov/).

For the 353 targets with trial or literature evidence, the MR provided no strong evidence of a causal effect of the expressions of these targets showed on COVID-19 severity. Only 11 targets

showed nominated associations. This included *PLOD2*, which is the target for three marketed drugs: sirolimus, ribavirin and apicidin. *PLOD2*, located on chromosome 3, is found in the endoplasmic reticulum and functions to catalyse the hydroxylation of lysyl residues in collagen-like peptides. Its protein, PLOD2, was revealed to interact with the SARS-CoV-2 protein open reading frame 8 (ORF8) (62), which was identified as containing the most pathogenic variants of all coronavirus proteins (63). Another gene (*SRP72*) located on chromosome 4, is the only subunit of the signal recognition particle (SRP) that can be phosphorylated. The SRP mediates the transportation of proteins to the plasma membrane. In the context of coronavirus, the SRP mediates the transportation of proteins to the endoplasmic reticulum (64). SRP72 protein also forms part of the viral mRNA translation and gene expression pathways. In the context of COVID-19, SRP72 has been identified as a human protein that interacts with the nonstructural protein 8 (nsp8) of SARS-CoV-2 (65). This SRP-nsp8 interaction may contribute to the reconfiguration of the endoplasmic reticulum during SARS-CoV-2 infection (5).

Some limitations of our analysis are worth noting. Whilst initiatives are underway to collect genetic information for COVID-19 patients (e.g. the COVID-19 host genetics initiative, https://covid-19genehostinitiative.net/), the recent initial GWASs of COVID-19 found few genetic association signals, which highlights the importance of sample size and statistical power. In addition, recent MR studies of COVID-19 identified a few gene targets associated with COVID-19 (19)(20)(21)(22)(23)(24), but these studies downplayed consideration of potential biases in the data. For example, if we use a mixture of population samples as controls (which include unexposed individuals and asymptomatic and untested cases), we will not be able to distinguish the causal effects of being exposed to SARS-CoV2, infected by the virus, or progressing to severe COVID-19 after infection (26). A GWAS using an unbiased population sample screened to detect infection of COVID-19 (e.g. from antibody tests) will be helpful to disentangle this issue, but such a GWAS does not yet exist. Our study excluded the five available COVID-19 GWAS datasets (with populations or self-reported data as controls) to minimise the influence of mixed controls. We only used hospitalised COVID-19 vs. non-hospitalised COVID-19 as the outcome for MR, which had the most reliable definition of cases and controls. However, given the GWAS of hospitalized COVID-19 vs. not hospitalized COVID-19 was still selected on the basis of COVID-19 status, collider bias could still be an issue (see **Supplementary Figure 2**). To estimate the influence of potential collider bias, we conducted MR-PheWAS of the 353 prioritised targets and ENTPD5 on 622 human diseases and phenotypes. The prioritised targets such as ENTPD5 and PLOD2 did not show much evidence of a causal effect on major diseases (e.g. asthma), which suggested that gene/protein level change of these targets are not obviously acting through these disease on selection (e.g. COVID-19 testing). However, the targets prioritised by our study should still be carefully reviewed in future trials before clinical application. A second limitation is that the drug targets evaluated in this study were proxied using a limited number of instruments, which means the putative causal effects rely on one or two genetic instruments. Thus, these associations support causality but do not prove it, as horizontal pleiotropy remains an alternative possibility. In addition, even though our results suggest some biological links between the target and diseases, these only provide evidence for the very first step of the drug development process. Finally, whilst these are plausible targets for COVID-19 severity, we cannot predict whether successful intervention would impact on risk of infection or other disease characteristics relevant to public health (e.g. viral shedding).

In conclusion, this study prioritised ENTPD5 as a drug target for COVID-19 severity using an integrated analytical protocol including MR and colocalization approaches. ENTPD5 is related to anti-coagulation, which suggests a potential molecular pathway for COVID-19 treatment. This study evidences the value of genetic approaches in drug target validation and drug repurposing. This provides the very first step towards evaluating intervention targets that are worth following-up for coronavirus and viral infections.

Materials and Methods

Selection of genetic instruments for COVID-19 drug targets

First, genetic association information of gene and protein expression levels were looked up from two resources: i) protein expression levels in plasma from five studies (28)(29)(30)(31)(32) implemented in Zheng et al. (13) ; ii) gene expression levels in whole blood from eQTLGen consortium (33). The genetic variants, genes and proteins were further mapped to genome build GRCh37.p13 coordinates. We used the following criteria to select genetic instruments:

1. We selected SNPs that were associated with any protein or gene expression (using a P-value threshold $\leq 5x10^{-8}$) in at least one of the four GWASs, including both cis and trans instruments.
2. We then conducted LD clumping for the instruments using the 1000 Genome European samples as reference panel, which was implemented in the TwoSampleMR R package (14) to identify independent instruments for each protein/gene. We used $r^2 < 0.001$ as the threshold to exclude correlated instruments in the cis (or trans) gene region.

After instrument selection, 1,674 conditioning distinguished pQTL signals of 1,002 proteins (**Supplementary Table 1**) and 39,630 conditioning distinguished eQTLs signals for 16,059 transcripts (**Supplementary Table 2**) were kept as instruments for the genetic analyses of this study. To identify cis and trans instruments, we split instruments into two groups: 1) 29,610 cis-acting instruments within a 500Kb window from each side of the leading QTL of the protein/transcript; 2) 10,020 instruments outside the 500Kb window of the leading QTL were designated as trans instruments. Whilst trans instruments may be more prone to pleiotropy, their inclusion could increase statistical power as well as the scale of the study. Therefore, for the proteins and genes with cis instruments or trans instruments, we conducted MR analyses using both sets of instruments (**Supplementary Table 1 and 2**).

Second, we selected additional drug targets that were prioritised by evidence from the following three resources: i) a list of 11 drugs under trials for COVID-19 treatment were extracted from *ClinicalTrials.gov* (**Supplementary Table 3**). These drugs were mapped to their target genes using Drug-Gene-Interaction (DGI) database (http://dgidb.org/) (44) and CHEMBL database (42); ii) human proteins interacting with SARS-CoV-2 proteins from Gordon et al. (5); iii) genes associated with SARS-CoV from Gralinski et al. (6). After de-duplicating these, 380 unique drug targets were selected for our study (**Supplementary Table 4**). Next, gene and protein expression levels of these targets were looked up from four resources: protein expression levels in plasma from four studies (28)(29)(30)(31)(32) implemented in Zheng et al. (13), gene expression levels in whole blood from eQTLGen consortium (33), tissue specific gene expression levels in 7 tissues from the GTEx consortium (34) and gene expression levels in two kidney tissues from Gilles et al. (35). After the mapping step, 353 drug targets with genetic variants robustly associated with the transcripts and/or proteins were included as the start point of the instrument selection (**Supplementary Table 5**).

For any of the 353 targets, we mapped the drug targets with related drug names using four platforms: OpenTargets (41), ChEMBL (42), DrugBank (43) and DGI platforms (44). After this analysis, we mapped one target, PLOD2, to the drugs which target it (**Supplementary Table 9**).

We further mapped the targets to the previously reported "druggable genome" (45). This study stratified the potential drug targets from across the genome into three tiers. Tier 1 (1427 genes) included efficacy targets of approved small molecules and biotherapeutic drugs, as well as targets modulated by clinical-phase drug candidates; tier 2 was composed of 682 genes encoding proteins closely related to drug targets, or with associated drug-like compounds; and tier 3 contained 2370 genes encoding secreted or extracellular proteins, distantly related proteins to approved drug targets. For the 11 additional prioritised COVID-19 drug targets, one target was mapped to tier 1, 2 targets to tier 3 (**Supplementary Table 9**).

Selection of genetic association information for COVID-19 severity

The genetic association information of COVID-19 severity were obtained from the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/). Among the six available genetic association data listed in **Supplementary Table 6**, we considered the hospitalized COVID-19 vs. not hospitalized COVID-19 as the most suitable dataset to represent COVID-19 severity. The major advantage of this data compared to other datasets (e.g. COVID-19 vs. population) is that all participants of this genetic analysis were confirmed COVID-19 cases, therefore this analysis will be less bias by data selection (e.g. there will be some unconfirmed COVID-19 cases in the population so it is not a perfect control dataset). We used the GWAS summary statistics of the hospitalized COVID-19 vs. not hospitalized COVID-19 as the outcome for the MR analysis. This GWAS combined European samples from three large scale biobanks: UK Biobank, DECODE and FinnGen (928 confirmed hospitalized COVID-19 patients, 2,028 confirmed non hospitalised COVID-19 patients; **Supplementary Table 6**).

Causal inference and sensitivity analyses

Mendelian randomization analysis

In the initial MR analysis, 1,674 instruments of 1,002 plasma proteins and 39,630 instruments of 16,059 gene transcripts in whole blood were used as the exposures to proxy the effect of these targets on COVID-19, which maximized the statistical power of the study and supported us to obtain an overall view of the target-COVID-19 associations. The hospitalized COVID-19 vs. not hospitalized COVID-19 were used as the outcome to proxy COVID-19 severity. Due to missingness of available exposure or outcome data, the causal relations of expression of 837 proteins and 14,036 transcripts on COVID-19 severity were tested in this study. We selected a P-value threshold of 0.05, corrected for the number of tests, as our threshold for prioritising MR results for follow up analyses (N_protein= 837, P_protein< 5.97×10^{-5}; N_transcript= 14,036, P_transcript <3.56×10^{-6}) The Wald ratio (40) method was used to obtain MR effect estimates for protein and/or transcript targets with only one instrument. For protein and/or transcript targets with two or more instruments, the inverse variance weighted (IVW) method (36) was used to estimate the causal effects.

Colocalization analysis

For the drug targets - COVID-19 associations, results that survived the Bonferroni corrected threshold (P_protein< 5.97×10^{-5} or P_transcript< 3.56×10^{-6}) in the MR analysis were evaluated using colocalization analysis (the "coloc" R package), which estimates the posterior probability of each genomic locus containing a single variant affecting both the target and the phenotype (37). We used the default prior probabilities that a variant is equally associated with each

phenotype ($p1=1\times10^{-4}$; $p2=1\times10^{-4}$) and both phenotypes jointly ($p12=1\times10^{-5}$). A posterior probability of > 70% for the colocalization hypothesis in this analysis would suggest that the two association signals are likely to colocalize within the test region (noted as "Colocalised"). The rest of the target-phenotype associations were noted as "Not colocalised".

Tissue Specificity analysis

The functional receptor of SARS-CoV-2, ACE2 (66), is highly expressed in multiple organs, including gastrointestinal tract, gallbladder, testis, and kidney. This is consistent with the fact that whilst SARS-CoV-2 infection primarily manifests with acute respiratory illness, SARS-CoV-2 can also be detected in faeces (67) and kidney tubules (68). The presence of SARS-CoV-2 in the alimentary tract for longer than in the respiratory system (69) suggests that the intestine may be a hidden reservoir of SARS-CoV-2. We therefore set out to explore differences in potential target effects in different tissues. In order to understand the tissue specific effects of the candidate targets of COVID-19 on human phenotypes, we selected the 9 tissues in which ACE2 is highly expressed, including testis, lung, kidney cortex, kidney glomerular, kidney tubulointerstitial, stomach, colon transverse, small intestine terminal ileum and colon sigmoid.

Tissue specific gene expression data of the 353 targets in each selected tissue were obtained from two studies: GTEx V8 and Gillies et al (34)(35). After selection, 580 instruments of 218 gene transcripts were selected in the 9 tissues, which included 141 instruments for 132 gene transcripts in testis, 125 instruments for 115 transcripts in lung, 20 instruments for 20 transcripts in kidney cortex, 6 instruments for 6 in kidney glomerular, 8 instruments for 8 transcripts in kidney tubulointerstitial, 71 instruments for 67 transcripts in stomach, 84 instruments for 81 transcripts in colon sigmoid, 98 instruments for 96 transcripts in colon transverse and 41 instruments for 39 transcripts in small intestine terminal ileum (**Supplementary Table 6**). The same MR and colocalization analysis pipeline were applied for the tissue specific analysis.

Follow-up analyses of prioritised targets for COVID-19 severity

For protein and/or transcript targets with robust MR (Bonferroni corrected threshold for tested proteins=5.97×10^{-5}; Bonferroni corrected threshold for tested transcripts=3.56×10^{-6}) and colocalization (probability >80%) evidence, we conducted a set of follow-up analyses to further confirm the association between targets and COVID-19 severity as well as to identify potential mediators/mechanisms linking the targets with COVID-19 severity.

Multi-trait colocalization analysis of prioritised targets

We first explore whether the causal variants were shared across transcriptome, proteome and COVID-19 severity for top MR findings. We applied multi-trait colocalization implemented in the moloc R package (46). The default prior probabilities of 1×10^{-4} for any one layer of association, 1×10^{-6} for any two layers of associations and 1×10^{-7} for colocalization of all three layers of associations were used in the moloc analysis. An overall colocalization probability of three traits (Pa,bc+Pab,c+Pac,b+Pabc) > 80% would suggest that the three association signals are likely to colocalize within the test region.

Mediation/two-step Mendelian randomization analysis of prioritised targets

For targets showing robust evidence from either protein or transcript MR, we further identified the potential causal mediators linking the prioritised targets with COVID-19 severity using two-

step MR (17). In first step, we used the protein / expression QTLs as instruments to estimate the effect of the proteins/transcripts on the potential mediator (e.g. an intermediate trait such as coagulation). In second step, we used the SNPs associated with the mediator as instruments to estimate the effect of the mediator on COVID-19 severity (**Supplementary Figure 3**). To make a comprehensive selection of candidate mediators (intermediate traits), we conducted a phenome-wide scan of the protein and transcript instruments (QTLs) in all 34,494 available human traits in the IEU Open GWAS database (47). For four traits showing suggestive association with the QTLs (P<0.001), we conducted mediation/two-step MR for these mediators.

Bidirectional MR to identify potential reverse causality

We further estimated the potential reverse causality of the mediators on ENTPD5 using bi-directional MR (39). The prioritised mediators were treated as the exposures and ENTPD5 protein level were selected as the outcome of this MR analysis.

Analysis software

The MR analyses (including Wald ratio, IVW, two-step MR and bidirectional MR) were conducted using the TwoSampleMR R package (github.com/MRCIEU/TwoSampleMR) (70). The MR results were plotted as Manhattan plots and forest plots using code derived from the ggplot2 package in R (https://cran.r-project.org/web/packages/ggplot2/index.html).

MR PheWAS of prioritised targets for COVID-19 severity

For the 353 prioritised drug targets with trial or experimental evidence and targets with robust MR evidence of association with COVID-19 severity, we further conducted a phenome-wide MR analysis (MR-PheWAS) to identify potential beneficial and/or adverse effects of these targets on other human diseases. The QTLs for the prioritised targets were chosen as the exposures for the MR-PheWAS. For the outcomes, 49 viral infection phenotypes from the GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics) (71), 504 human diseases and 72 disease related traits (e.g. blood lipids) selected from the IEU Open GWAS database (47) were selected as outcomes for this analysis (**Supplementary Table 11**). These 573 outcomes were selected using the following inclusion criteria:

- The GWAS with the greatest expected statistical power (e.g. largest sample size / number of cases) when multiple GWAS records of the same disease / risk factor were available in the Open GWAS database.
- GWAS with betas, standard errors and effect alleles for all tested variants (i.e. full GWAS summary statistics available)

For the MR-PheWAS results that survived the Bonferroni corrected threshold ($P< 1.1x10^{-6}$), we further evaluated them using colocalization analysis, which estimates the posterior probability of each genomic locus containing a single variant affecting both the target and the phenotype (37). The same prior probabilities applied for the previous colocalization were applied here. A posterior probability of > 70% for the colocalization hypothesis in this analysis would suggest that the two association signals are likely to colocalize within the test region (noted as "Colocalised"). In addition, we conducted an approximate colocalization analysis (called LD check), in which the LD between the sentinel variant of each target and the 30 strongest SNPs in the region associated with the phenotype were checked. LD $r^2 > 0.7$ between the sentinel variant

and any of the 30 SNPs associated with the phenotype was used as evidence of colocalization (noted as "LD Check"). The rest of the target-phenotype associations were noted as "Not colocalised".

COVID-19 Target-Disease Atlas (CTDA) browser of the EpiGraphDB platform

We have made all MR results openly available to browse or download at the COVID-19 Target-Disease Atlas (CTDA) browser within the EpiGraphDB platform (http://epigraphdb.org/covid-19/ctda/). This includes 14,873 unique target-COVID-19 severity associations evidence for the omics MR as well as 372,830 unique target-disease associations evidence for 353 targets on 622 diseases/phenotypes in 11 SARS-CoV-2 related tissues. Users are able to query the study results by the targeted gene/protein name and QTL SNPs via the online platform, and the results are presented in searchable tables as well as volcano plots. In addition, users can programmatically access the results using the /covid-19/ctda endpoints in the application programming interface (API) of EpiGraphDB via http://api.epigraphdb.org/.

Supplementary Materials

TableS1. The available pQTL instruments of 1,002 plasma proteins.

TableS2. The available eQTL instruments of 16,059 blood transcripts.

TableS3. Drugs in trials for COVID-19 treatment and their target genes.

TableS4. The available genetic information for SARS-CoV-2 target genes from pQTL and tissue specific eQTL resources.

TableS5. The genetic predictors and association information of the SARS-CoV-2 target genes.

TableS6. The genome-wide association studies of the 6 COVID-19 phenotypes.

TableS7. The plasma protein-COVID-19 severity associations with suggestive Mendelian randomization evidence (P<0.05).

Note: The target-COVID-19 severity associations with Mendelian randomization p value < 10^{-2} were included in this table.

TableS8. The blood transcription-COVID-19 severity associations with suggestive Mendelian randomization evidence (P<0.05).

Note: The target-COVID-19 severity associations with Mendelian randomization p value < 10^{-2} were included in this table.

TableS9. The target-COVID-19 severity associations with suggestive Mendelian randomization evidence (P<0.05).

TableS10. The colocalization results for proteins, expression levels of ENTPD5 and COVID-19 severity.

Note: The target-disease associations passed Bonferroni corrected threshold (6.0×10^{-5}) were tested further by colocalization analysis.

TableS11. The list of 622 human phenotypes been used as outcomes for the phenome-wide mendelian randomization analysis.

TableS12. The phenome-wide Mendelian randomization results for the 12 prioritised drug targets.

TableS13. The triangulation table mapping drug target gene with drug name, the drug trial and druggability information.

Note: The target genes were mapped to the drugs using DGI, OpenTargets, ChEMBL and/or Drugbank databases. The drug trial information for each drug was identified from DGI, OpenTargets, ChEMBL, Drugbank and/or ClinicalTrial.gov.

TableS14. Phenome-wide association scan of the pQTL and eQTL instruments of ENTPD5.

TableS15. Two-step Mendelian randomization analysis linking ENTPD5 with the potential mediators and COVID-19 severity

Note: Four modifiable traits showed association signals with the pQTL and eQTL of ENTPD5 were selected as candidate mediators (HDL cholesterol, C-reactive protein, coagulation factor VIII and body fat percentage).

Funding:
JZ is funded by a Vice-Chancellor Fellowship from the University of Bristol. This research was also funded by the UK Medical Research Council Integrative Epidemiology Unit (MC_UU_00011/1, MC_UU_00011/4). This study was funded/supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol (TRG). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. YMZ is supported by National Natural Science Foundation of China (81800636). HZ is supported by the University of Michigan Health System–Peking University Health Science Center Joint Institute for Translational and Clinical Research (BMU2017JI007). This work was supported by a Wellcome Trust Institutional Translational Partnership Award (209739/Z/17/Z).

Competing interests: No competing interests

Availability of data
The data used in this study is available via the IEU Open GWAS platform (https://gwas.mrcieu.ac.uk/). The results of this study is available via the EpiGraphDB platform http://epigraphdb.org/covid-19/ctda/

Code availability
The code to conduct Mendelian randomization analysis was available via the TwoSampleMR R package (https://mrcieu.github.io/TwoSampleMR/).

Acknowledgments

This research was conducted using the UK Biobank resource under application number 15825. The UK Biobank received ethnical approval from the research ethnic committee (REC reference for UK biobank 11/NW/0382) and participants provided written informed consent.

Author contributions:
JZ and YMZ selected the drug targets; JZ performed the Mendelian randomization analysis; JZ and DB performed the colocalization analysis; JZ conducted the triangulation between MR and drug trials; JZ and YMZ conducted the drug target prioritisation; YL developed the database and web browser; JZ and YMZ wrote the manuscript; DB, YL, LW, XZL, HZ and TRG reviewed the paper and provided key comments; JZ, YMZ and TRG conceived and designed the study and oversaw all analyses.

References:

1. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection□: A Narrative Review. Ann Intern Med. 2020 Sep 1;173(5):362–7.

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020 Feb 15;395(10223):497–506.

3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054–62.

4. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Final Report. N Engl J Med. 2020 Oct 8;

5. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020 Apr 30

6. Gralinski LE, Ferris MT, Aylor DL, Whitmore AC, Green R, Frieman MB, et al. Genome Wide Identification of SARS-CoV Susceptibility Loci Using the Collaborative Cross. PLoS Genet. 2015 Oct;11(10):e1005504.

7. Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. N Engl J Med. 2020 Apr 10

8. Mahevas M, Tran V-T, Roumier M, Chabrol A, Paule R, Guillaud C, et al. No evidence of clinical efficacy of hydroxychloroquine in patients hospitalized for COVID-19 infection with oxygen requirement: results of a study using routinely collected data to emulate a target trial. 2020.

9. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med. 2020 Jul 17

10. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020 Apr 13

11. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2020 Oct 6

12. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015 Aug;47(8):856–60.

13. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020 Oct;52(10):1122–31.

14. Richardson TG, Hemani G, Gaunt TR, Relton CL, Davey Smith G. A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. Nat Commun. 2020 Jan 10;11(1):185.

15. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22.

16. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep. 2017 Nov 22;4(4):330–45.

17. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012 Feb;41(1):161–76.

18. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci Rep. 2015 Nov 16;5:16645.

19. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. 2020 Oct 15;383(16):1522–34.

20. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in Covid-19. Nature. 2020 Dec 11

21. Mohd Karim M, Ghoussaini M, Dunham I. Mining a GWAS of Severe Covid-19. N Engl J Med. 2020 Nov 24;383(26).

22. Zhou S, Butler-Laporte G, Nakanishi T, Morrison D, Afilalo J, Afilalo M, et al. A Neanderthal OAS1 isoform protects against COVID-19 susceptibility and severity: Results from Mendelian randomization and case-control studies. 2020.

23. Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. medRxiv. 2020 Sep 4

24. Gaziano L, Giambartolomei C, Pereira AC, Gaulton A, Posner DC, Swanson SA, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. medRxiv. 2020 Nov 23;2020.11.19.20234120.

25. Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, Justice AAE, et al. Common genetic variants identify therapeutic targets for COVID-19 and individuals at high risk of severe disease. 2020.

26. Paternoster L, Tilling KM, Smith GD. Genetic Epidemiology And Mendelian Randomization For Informing Disease Therapeutics: Conceptual And Methodological Challenges. 2017.

27. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun. 2020 Nov 12;11(1):5749.

28. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018 Jun;558(7708):73–9.

29. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS Genet. 2017 Apr;13(4):e1006706.

30. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017 Feb 27;8:14357.

31. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun. 2018 Aug 15;9(1):3268.

32. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018 Aug 2

33. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. BioRxiv. 2018

34. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. BioRxiv. 2019

35. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. Am J Hum Genet. 2018 Aug 2;103(2):232–44.

36. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013 Nov;37(7):658–65.

37. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014 May;10(5):e1004383.

38. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol [Internet]. 2017 Sep 27

39. Timpson NJ, Nordestgaard BG, Harbord RM, Zacho J, Frayling TM, Tybjærg-Hansen A, et al. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. Int J Obes. 2011 Feb;35(2):300–8.

40. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008 Apr 15;27(8):1133–63.

41. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. Nucleic Acids Res. 2017 Jan 4;45(D1):D985–94.

42. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019 Jan 8;47(D1):D930–40.

43. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D668-72.

44. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res. 2018 Jan 4;46(D1):D1068–73.

45. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017 Mar 29;9(383).

46. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2018 Aug 1;34(15):2538–45.

47. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. BioRxiv. 2020

48. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020 Jul 17;369(6501):330–3.

49. Yang Q, Hughes TA, Kelkar A, Yu X, Cheng K, Park S, et al. Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. Elife. 2020 Oct 26;9.

50. Krystyna Taylor SD Matthew Pearson, James Kozubek, Marcin Pawlowski, Claus Erik Jensen, Zbigniew Skowron, Gert Lykke Møller, Mark Strivens, Steve Gardner. Analysis of Genetic Host Response Risk Factors in Severe COVID-19 Patients. Precision Life. 2020;

51. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020 May;8(5):475–81.

52. Zhang L, Yan X, Fan Q, Liu H, Liu X, Liu Z, et al. D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. J Thromb Haemost. 2020 Jun;18(6):1324–9.

53. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med. 2020 Jul 1;180(7):934–43.

54. Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. J Thromb Haemost. 2020 May;18(5):1094–9.

55. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta. 2020 Jul;506:145–8.

56. Lippi G, Favaloro EJ. D-dimer is Associated with Severity of Coronavirus Disease 2019: A Pooled Analysis. Thromb Haemost. 2020 May;120(5):876–8.

57. Lillicrap D. Disseminated intravascular coagulation in patients with 2019-nCoV pneumonia. J Thromb Haemost. 2020 Apr;18(4):786–7.

58. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. N Engl J Med. 2020 Apr 23;382(17):e38.

59. Ramlall V, Thangaraj PM, Meydan C, Foox J, Butler D, Kim J, et al. Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. Nat Med. 2020 Oct;26(10):1609–15.

60. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost. 2020 Apr;18(4):844–7.

61. Paranjpe I, Fuster V, Lala A, Russak AJ, Glicksberg BS, Levin MA, et al. Association of Treatment Dose Anticoagulation With In-Hospital Survival Among Hospitalized Patients With COVID-19. J Am Coll Cardiol. 2020 Jul 7;76(1):122–4.

62. Sirpilla O, Bauss J, Gupta R, Underwood A, Qutob D, Freeland T, et al. SARS-CoV-2-Encoded Proteome and Human Genetics: From Interaction-Based to Ribosomal Biology Impact on Disease and Risk Processes. J Proteome Res. 2020 Nov 6;19(11):4275–90.

63. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet. 2020 Aug 29;396(10251):603–11.

64. Luirink J, Sinning I. SRP-mediated protein targeting: structure and function revisited. Biochim Biophys Acta. 2004 Nov 11;1694(1–3):17–35.

65. Sicari D, Chatziioannou A, Koutsandreas T, Sitia R, Chevet E. Role of the early secretory pathway in SARS-CoV-2 infection. J Cell Biol. 2020 Sep 7;219(9).

66. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270–3.

67. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med. 2020 Mar 5;382(10):929–36.

68. Diao B, Wang C, Wang R, Feng Z, Tan Y, Wang H, et al. Human Kidney is a Target for Novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection. 2020.

69. Wu Y, Guo C, Tang L, Hong Z, Zhou J, Dong X, et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol Hepatol. 2020 Mar 19

70. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018 May 30;7.

71. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019 Jan 8;47(D1):D1005–12.

Figures:

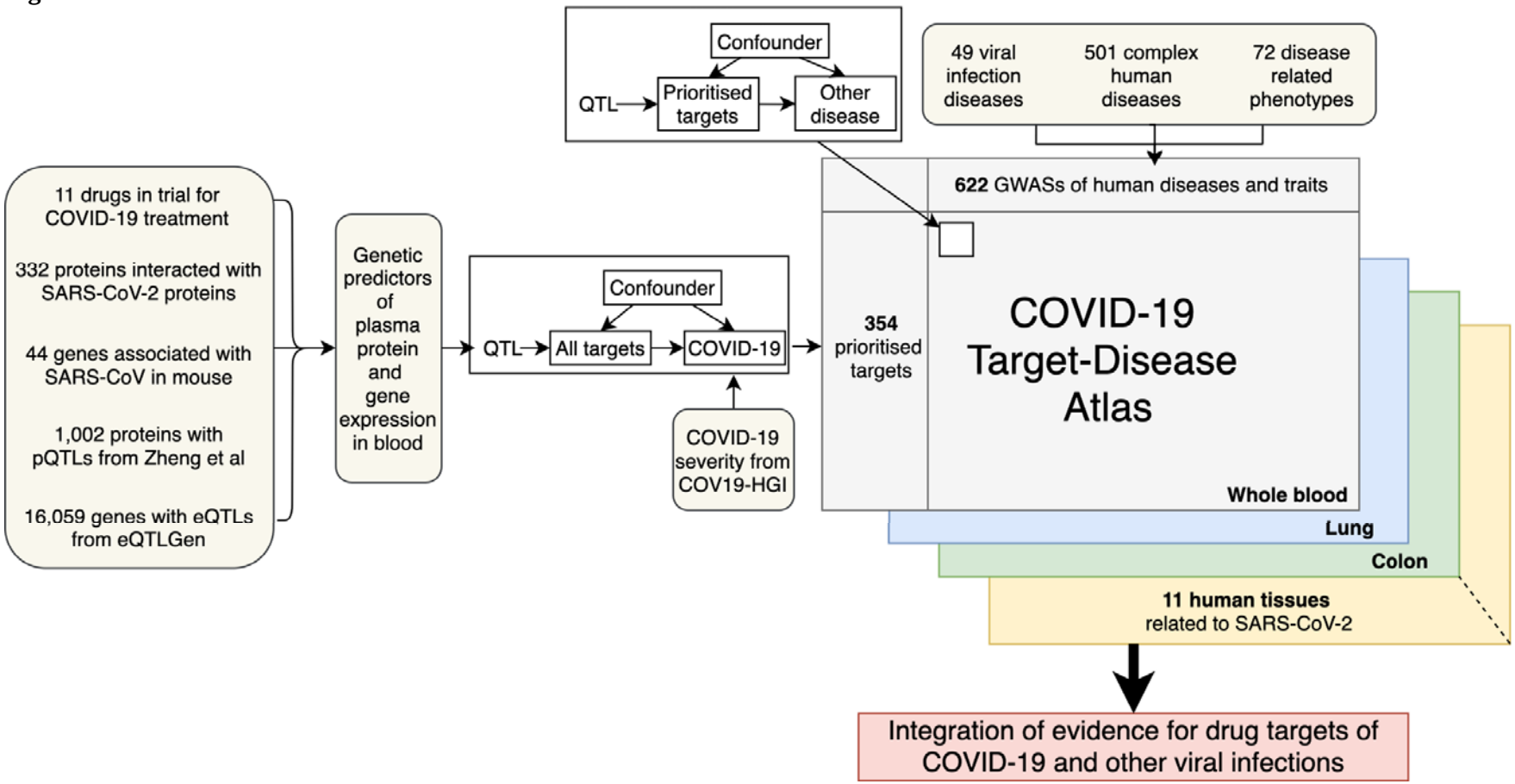


Figure 1. Study design of the multi-omics MR study to prioritise drug targets for COVID-19 severity.

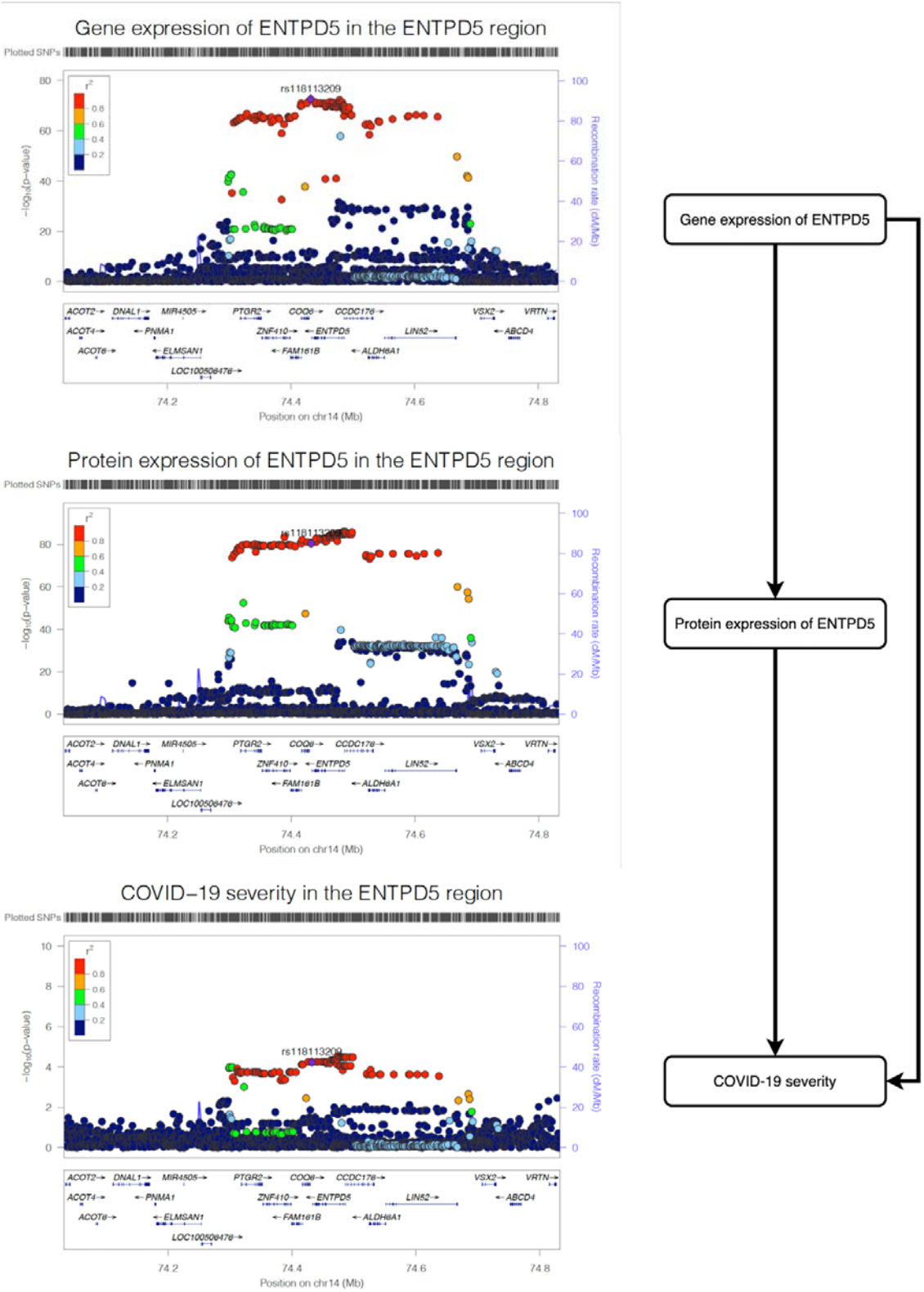


Figure 2. Regional plot of gene expression of *ENTPD5*, protein expression of ENTPD5 and COVID-19 severity in the *ENTPD5* region. The three regional plots refer to the genetic associations of (A), gene expression of *ENTPD5*; (B) protein expression of ENTPD5;

(C) and COVID-19 severity. The X-axis is the chromosome and position of the *ENTPD5* region.

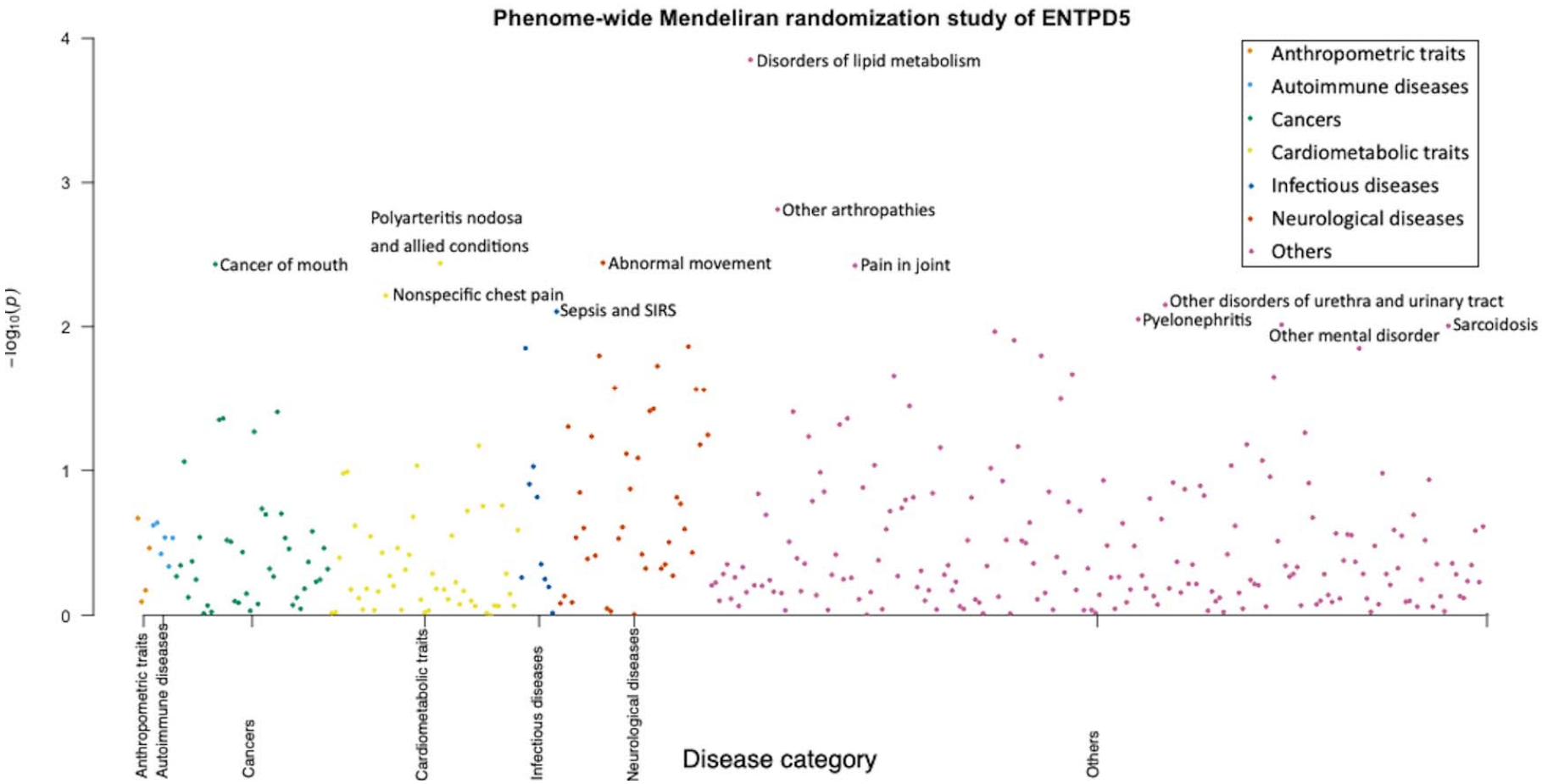
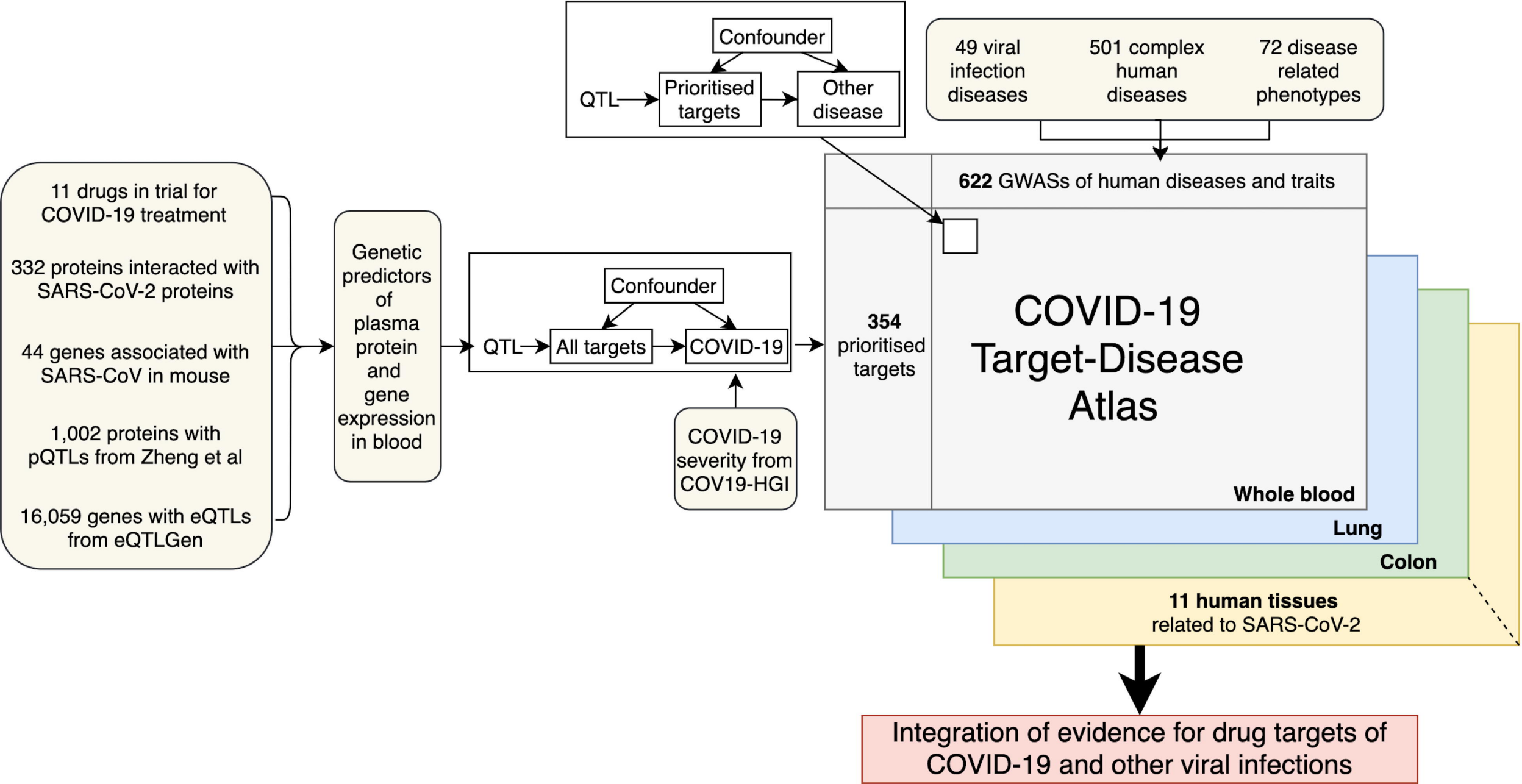
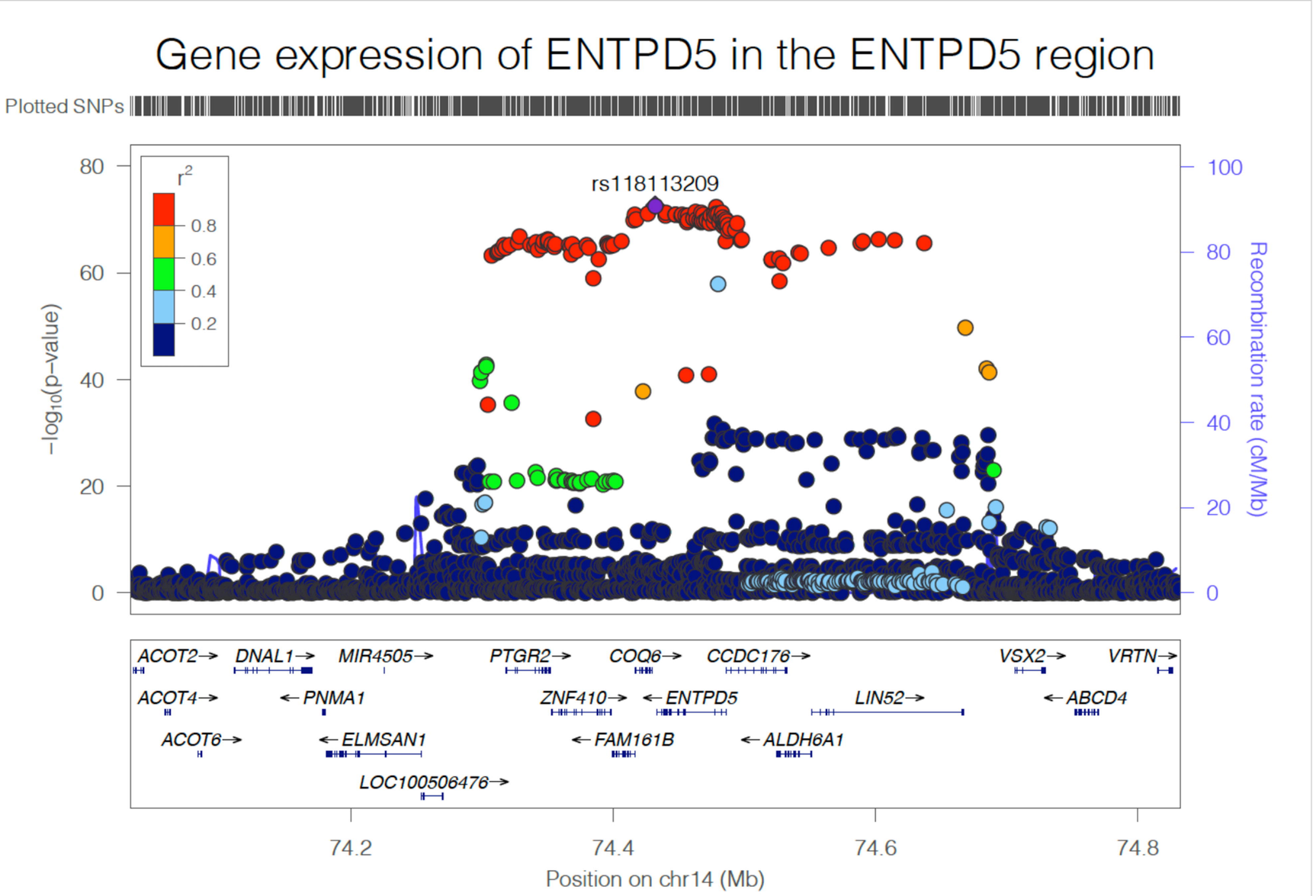
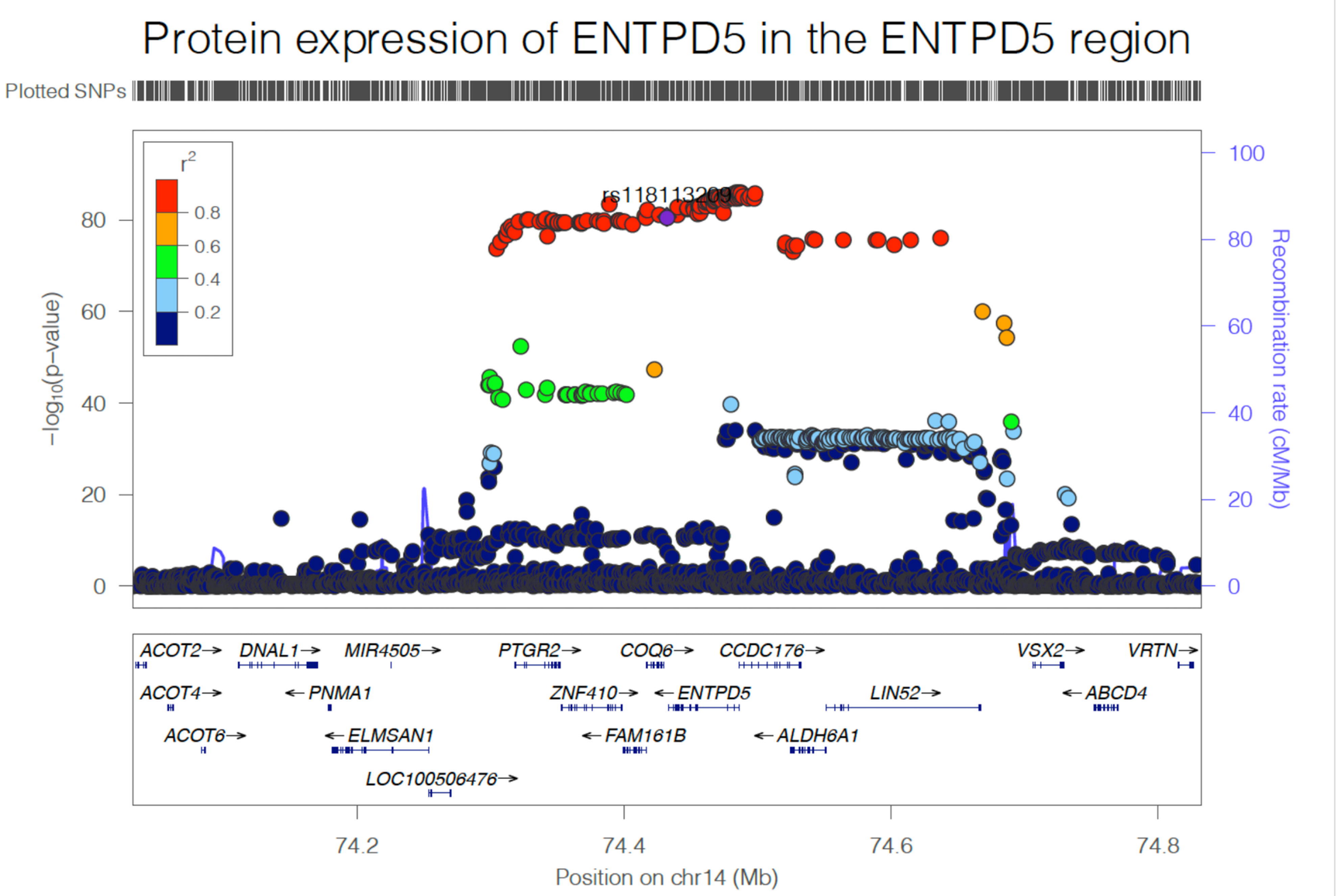
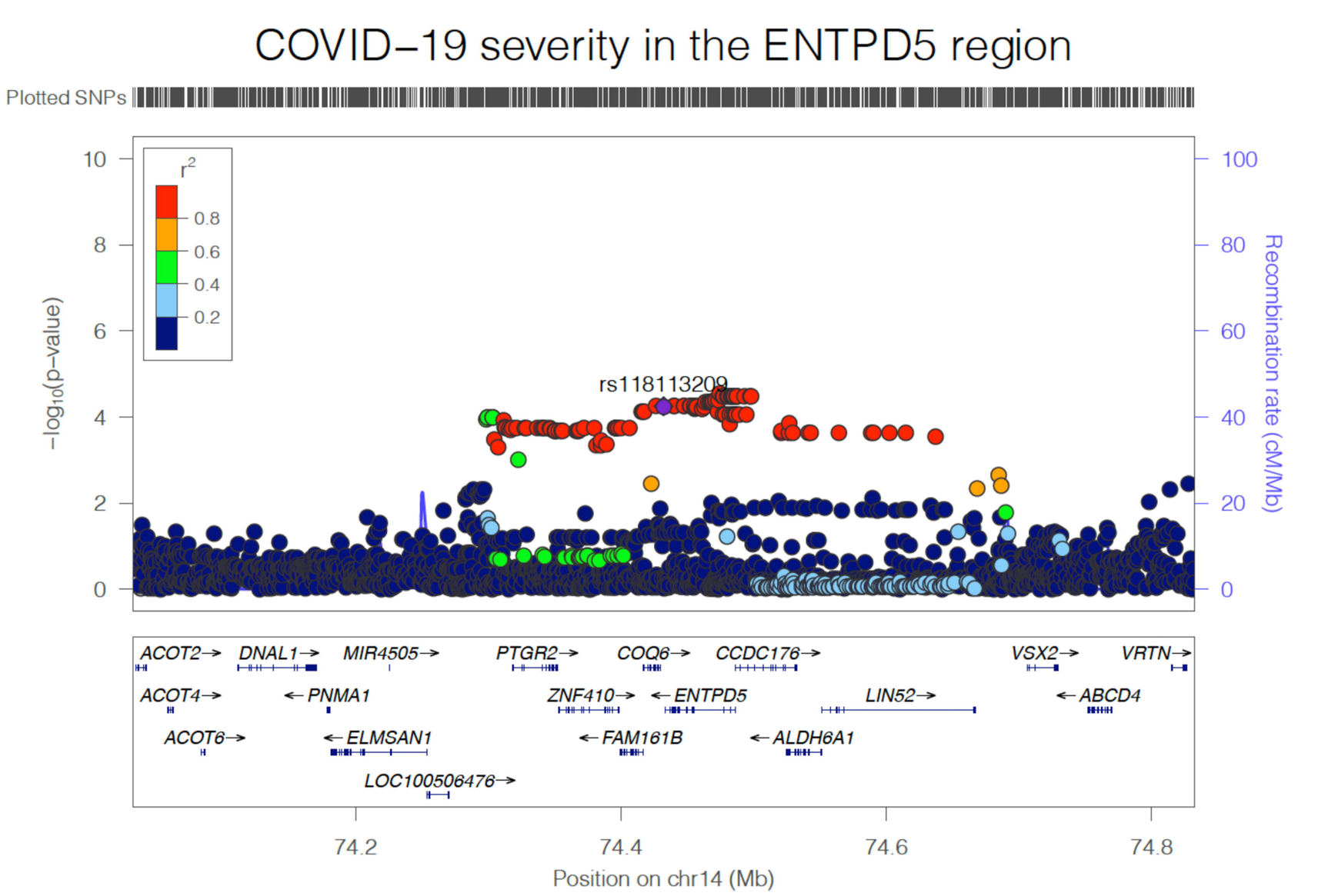


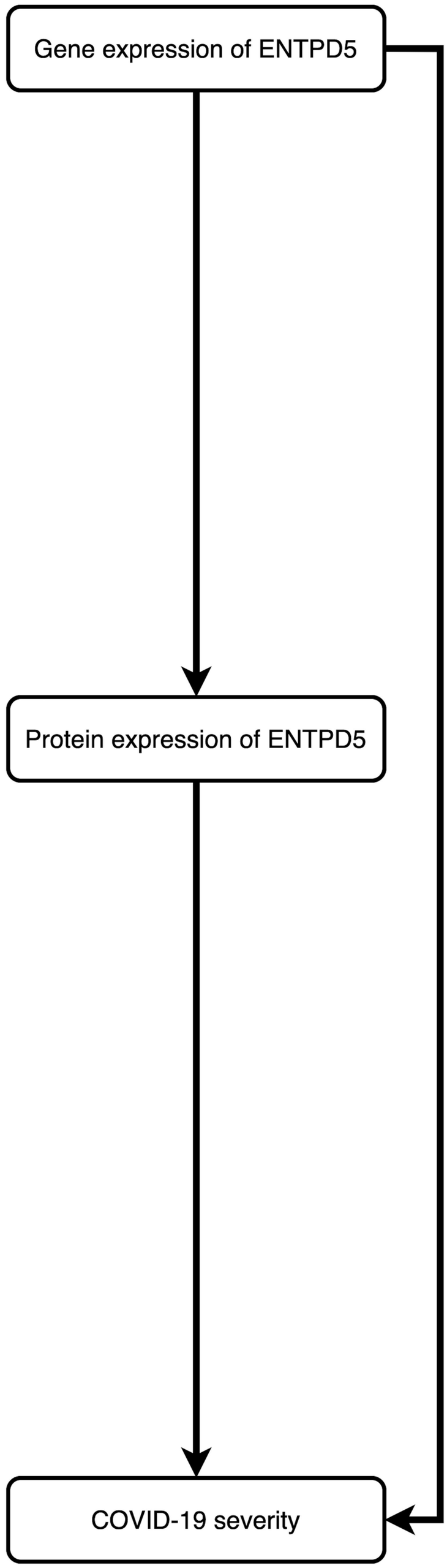
Figure 3. Manhattan plot demonstrating phenome-wide Mendelian randomization associations of protein expression level of ENTPD5 on human phenotypes. Different colors refer to different disease areas. For the 622 tested traits, none of them passed the Bonferroni-corrected threshold is 8.04×10^{-5}. Therefore, ENTPD5 are less likely to be influenced by horizontal pleiotropy.

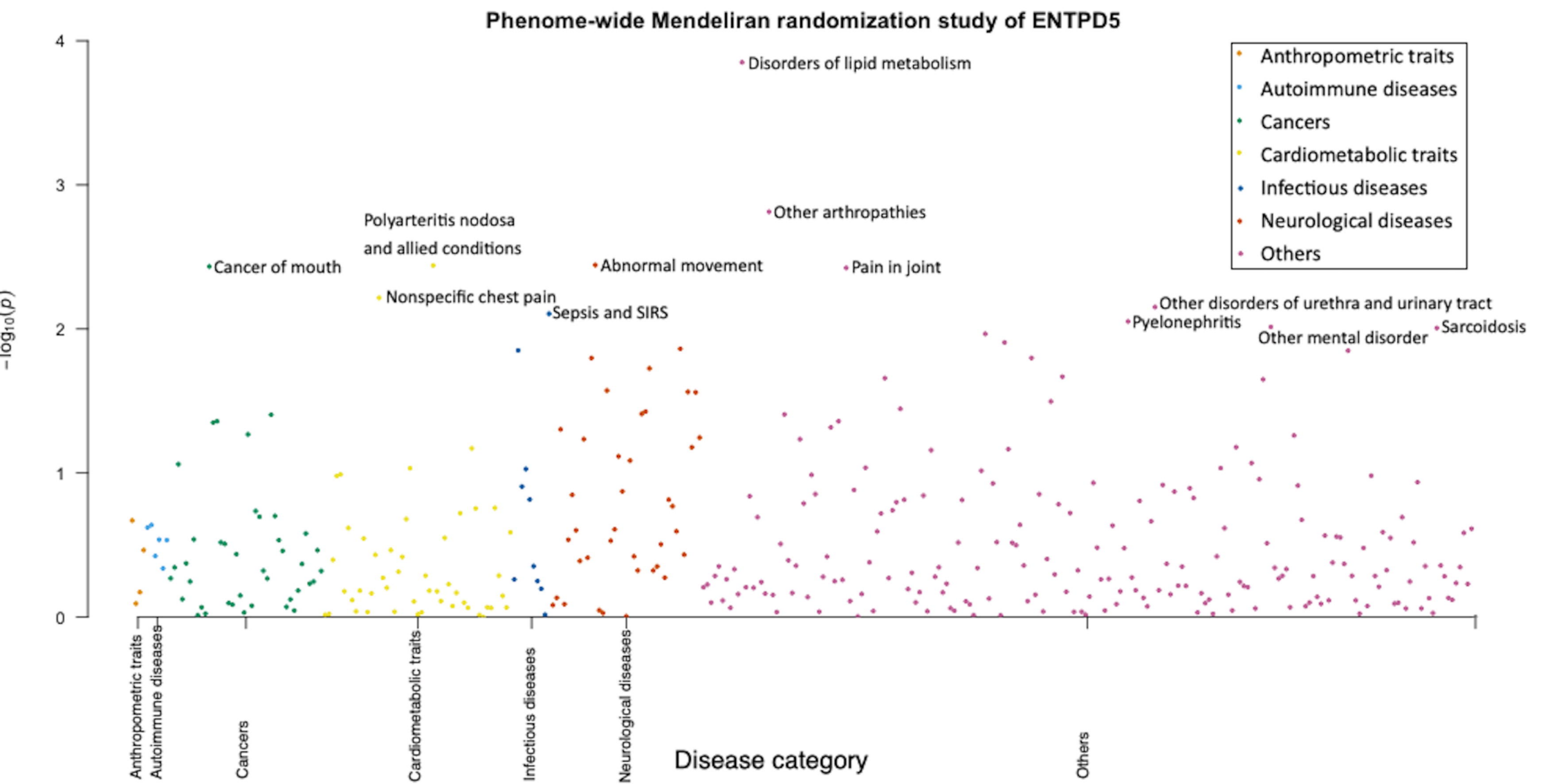

Confounder
QTL
Prioritised targets
Other disease
49 viral infection diseases
501 complex human diseases
72 disease related phenotypes
11 drugs in trial for COVID-19 treatment
332 proteins interacted with SARS-CoV-2 proteins
44 genes associated with SARS-CoV in mouse
1,002 proteins with pQTLs from Zheng et al
16,059 genes with eQTLs from eQTLGen
Genetic predictors of plasma protein and gene expression in blood
Confounder
QTL
All targets
COVID-19
COVID-19 severity from COV19-HGI
622 GWASs of human diseases and traits
354 prioritised targets
COVID-19 Target-Disease Atlas
Whole blood
Lung
Colon
11 human tissues related to SARS-CoV-2
Integration of evidence for drug targets of COVID-19 and other viral infections


Gene expression of ENTPD5 in the ENTPD5 region
Plotted SNPs
rs118113209
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ACOT2→ DNAL1→ MIR4505→ PTGR2→ COQ6→ CCDC176→ VSX2→ VRTN→
ACOT4→ ←PNMA1 ZNF410→ ←ENTPD5 LIN52→ ←ABCD4
ACOT6→ ←ELMSAN1 ←FAM161B ←ALDH6A1
LOC100506476→
74.2 74.4 74.6 74.8
Position on chr14 (Mb)


Protein expression of ENTPD5 in the ENTPD5 region
Plotted SNPs
rs118113209
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ACOT2→ DNAL1→ MIR4505→ PTGR2→ COQ6→ CCDC176→ VSX2→ VRTN→
ACOT4→ ←PNMA1 ZNF410→ ←ENTPD5 LIN52→ ←ABCD4
ACOT6→ ←ELMSAN1 ←FAM161B ←ALDH6A1
LOC100506476→
74.2 74.4 74.6 74.8
Position on chr14 (Mb)


COVID−19 severity in the ENTPD5 region
Plotted SNPs
rs118113209
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ACOT2→ DNAL1→ MIR4505→ PTGR2→ COQ6→ CCDC176→ VSX2→ VRTN→
ACOT4→ ←PNMA1 ZNF410→ ←ENTPD5 LIN52→ ←ABCD4
ACOT6→ ←ELMSAN1 ←FAM161B ←ALDH6A1
LOC100506476→
74.2 74.4 74.6 74.8
Position on chr14 (Mb)


Gene expression of ENTPD5
Protein expression of ENTPD5
COVID-19 severity


Phenome-wide Mendeliran randomization study of ENTPD5
Anthropometric traits
Autoimmune diseases
Cancers
Cardiometabolic traits
Infectious diseases
Neurological diseases
Others
Disorders of lipid metabolism
Other arthropathies
Polyarteritis nodosa
and allied conditions
Cancer of mouth
Abnormal movement
Pain in joint
Nonspecific chest pain
Sepsis and SIRS
Other disorders of urethra and urinary tract
Pyelonephritis
Other mental disorder
Sarcoidosis
$-\log_{10}(p)$
0
1
2
3
4
Anthropometric traits
Autoimmune diseases
Cancers
Cardiometabolic traits
Infectious diseases
Neurological diseases
Others
Disease category