Title: Multi-omics study revealing tissue-dependent putative mechanisms of SARS-CoV-2 drug targets on viral infections and complex diseases

Authors: Jie Zheng[1†]*, Yuemiao Zhang[2†]*, Yi Liu[1], Denis Baird[1], Xingzi Liu[2], Lin Wang[3], Hong Zhang[2], George Davey Smith[1,4], Tom R Gaunt[1,4]

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

[4] 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 causal evidence with drug trials to prioritize drug targets for COVID-19 treatment.

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 result in less drug resistance but could also exhibit unintended effects on other complex diseases. Here we pooled 353 candidate drug targets of COVID-19 from clinical trial registries and the literature and estimated their putative causal effects in 11 SARS-CoV-2 related tissues on 622 complex human diseases. By constructing a disease atlas of drug targets for COVID-19, we prioritise 726 target-disease associations with evidence of causality using robust Mendelian randomization (MR) and colocalization evidence (http://epigraphdb.org/covid-19/ctda/). Triangulating these MR findings with historic drug trial information and the druggable genome, we ranked and prioritised three genes *DHODH*, *ITGB5* and *JAK2* targeted by three marketed drugs (Leflunomide, Cilengitide and Baricitinib) which may have repurposing potential with careful risk assessment. This study evidences the value of our integrative approach in prioritizing and repurposing drug targets, which will be particularly applicable when genetic association studies of COVID-19 are available in the near future.

Introduction

Drug development is expensive and time-consuming. To address this, studies have begun to promote the development of disciplined pipelines for early therapeutic Research and Development (R&D) (1) and support the role of genetics in predicting drug trial success (2). In recent work we have demonstrated the value of proteome-wide and tissue-specific transcriptome-wide studies in prioritizing target-indication pairs that are more likely to be successful in clinical trials (3)(4).

The recent outbreak of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of novel coronavirus disease (COVID-19), has become a global pandemic. The current treatment strategy is focused on maintaining life and organ functions of patients. Drug development is a component of our response to both COVID-19 and potential future coronavirus epidemics. At this time, 153 clinical trials have been set up (information from ClinicalTrial.gov), including Hydroxychloroquine, Chloroquine and Baricitinib. However, in this rapidly developing situation, only preliminary results of some early stage trials (5) and observational studies (6) have been reported, and the efficacy of these drugs against COVID-19 is as yet unproven (7). In addition to COVID-19 trials, alternative evidence for COVID-19 drug target development has been provided in two ways: i) by determining the host proteins which interact with SARS-CoV-2; and ii) by exploring the potential translation of SARS-CoV targets to SAR-CoV-2 therapeutics. The first of these has been addressed by a recent study which identified 332 human proteins interacting with 26 SARS-CoV-2 proteins using affinity purification-mass spectrometry in HEK293 cell line. This gives the potential means for an important potential anti-coronavirus approach—blocking the interaction between human proteins and SARS-CoV-2 (8). The second relevant strand of evidence comes from SARS-CoV (2003), a closely related coronavirus, which may provide drug targets for general coronavirus interventions. One such study has identified 59 mouse genes associated with SARS-CoV using mouse models (9), of which 44 could be mapped to human genes. The above potential drug targets, including *DHODH* as the target gene for Leflunomide, may offer anti-virus therapeutic mechanisms through inhibiting human proteins interacting with SARS-CoV-2 rather than the virus itself, which will potentially result in less drug resistance and generalisability to other coronaviruses (8). However, they could also exhibit unintended beneficial or adverse effects on other complex diseases. Systematic assessment of the putative causal effects of these drug targets on the human phenome using genetics will provide key information on the on-target beneficial and adverse effects of these COVID-19 targets.

To inform the COVID-19 drug development process, we applied our drug prioritization pipeline to 353 drug targets which are believed to potentially interact with SARS-CoV-2. We attempted to test their causal effects on other viral infection phenotypes and explore their potential on-target beneficial and adverse effects on complex diseases. To achieve these aims, we built a disease atlas for human proteins and genes interacting with SARS-CoV-2 using plasma proteome-wide and tissue-specific transcriptome-wide Mendelian randomization (MR) studies, which includes 372,482 unique estimates of target effects on disease. This represents 353 targets versus 49 viral infection phenotypes, 501 complex diseases and 72 disease related phenotypes evaluated using target gene expression data from 11 SARS-CoV-2 related human tissues. We further triangulate the estimates of target effect on disease with drug trial information and the druggable genome to prioritise the most promising drugs targets with good repurposing potential and few predicted adverse effects. To enable rapid queries, results of all analyses are available in an open access online platform (http://epigraphdb.org/covid-19/ctda/).

Results

Characteristics of genetic instruments for COVID-19 drug targets

We retrieved drug targets potentially relevant to COVID-19 from three resources (**Figure 1**): 1) target genes of 11 drugs under trials for COVID-19 treatment from *ClinicalTrials.gov* (**Supplementary Table 1**); 2) 332 human proteins interacting with SARS-CoV-2 proteins in human cell lines (8); 3) 44 genes associated with SARS-CoV in a mouse model (9). After removing duplicates, 380 unique targets were selected (**Supplementary Table 2**).

The genetic predictors (instruments) for the 380 target genes in 11 SARS-CoV-2 related tissues were extracted from 4 recent expression and protein quantitative trait locus (eQTL and pQTL respectively) studies (3)(10)(11)(12). From these studies 1493 instruments for 353 drug targets were available for our proteome/transcriptome-wide association analysis (**Supplementary Table 3**), while the remaining 27 targets have no robustly associated genetic variants.

Disease atlas of COVID-19 drug targets

The target-disease atlas of the 353 drug targets was built using two-sample MR (13)(14) and colocalization analysis (15), evaluating evidence for their causal effects on 49 viral infection phenotypes, 501 complex diseases and 72 disease related phenotypes (**Supplementary Table 4**). 45,590 target-disease associations were tested in plasma proteome and transcriptome in whole blood ($P< 1.1x10^{-6}$ 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. Overall, 372,482 target-disease associations were estimated in the 11 COVID-19 relevant tissues (see the list of tissues in Methods). Overall, we observed 833 target-disease associations with robust MR evidence in the 11 tested tissues. 726 of the 833 (87.2%) associations also showed strong colocalization evidence (colocalization probability > 70%) (**Supplementary Table 5, 6, 7** and **8**), making these the most reliable findings of this study. Of these, 366 associations were obtained using a single cis instrument in a Wald ratio model (16), 327 were obtained using a single trans instrument and 33 were estimated using multiple instruments in an inverse variance weighted (IVW) model (17). The remaining 107 (12.8%) associations had evidence from MR but did not have strong evidence of colocalization (probability<70%; **Supplementary Table 9**), emphasizing the importance of this approach to address confounding by linkage disequilibrium (LD) in phenome-wide association studies.

Findings from our target-disease atlas can be used to conduct hypothesis-driven investigations of tissue-dependent effects of target expression on certain diseases. **Figure 2** illustrates the effects of target expression in different tissues on four diseases from our atlas: Crohn's disease (A); hypertension (B); hay fever, allergic rhinitis or eczema (C); and diabetes (D). Tissue-specific associations were observed between 11 to 17 of the target genes and these four diseases (**Supplementary Table 10**). Since these target genes that encode COVID-19-interacting proteins also appear to have causal effects on these diseases, it would be important to carefully assess their potential beneficial and/or adverse effects on these complex diseases in any future drug target prioritization for COVID-19.

Target – viral infection phenotype associations

In the analysis of the association between the 353 drug targets and 49 viral infection phenotypes, we identified 2 associations with robust MR ($P<1.1x10^{-6}$) and colocalization evidence (probability >70%) using gene expression data in whole blood, including *NEU1* associated with

chronic hepatitis and *DPY19L1* associated with viral enteritis. Three additional drug targets were suggestively associated with 3 viral infection phenotypes ($P<1x10^{-3}$) (**Supplementary Table 5**). This included *JAK2* (the target gene for a marketed drug, Baricitinib) associated with chronic hepatitis.

Target – complex disease associations in blood

In the primary analysis in plasma proteome and whole blood transcriptome, we identified 45 potential causal effects of protein levels on disease traits (**Supplementary Table 6**) and 430 potential causal effects of transcript levels on disease traits (**Supplementary Table 7**), with both MR and colocalization evidence. These 475 causal estimates included 95 targets on 105 diseases and disease related phenotypes. **Figure 3** illustrates the MR results of gene expression level of the *JAK2* gene on 501 complex diseases and 72 phenotypes using whole blood data from the eQTLGen consortium. In this example, the gene expression level of *JAK2* was predicted to influence 9 diseases/phenotypes (**Supplementary Table 7**), such as atopy, Type 1 diabetes, inflammatory bowel disease and obesity, which suggested that this target gene is potentially pleiotropic.

Target-complex disease associations in SARS-CoV-2 related tissues

As a follow-up analysis, we further identified 249 potential tissue-dependent causal effects of target on disease with robust MR and colocalization evidence. This covered 52 unique genes based on instruments for expression in 7 tissues against 47 unique diseases (**Supplementary Table 8**). Among the 29 tissue-dependent effects of target on disease with robust MR and colocalization evidence in more than one tissue, 27 of them showed the same direction of effect across whole blood and multiple other tissues. Only the genetically instrumented gene expression level of *KPNA1* and *SRP19* in whole blood and in testis showed opposite direction of effects against diastolic blood pressure and heel bone mineral density respectively. **Figure 4** illustrates an example of the tissue-dependent effect of *DHODH* gene on low-density lipoprotein cholesterol (LDL-C). Inhibition of *DHODH* gene is the functional mechanism of a marketed drug, Leflunomide. This drug has been suggested as a potential treatment for COVID-19 (18)(19). In our MR analysis, we observed that the expression level of *DHODH* was specifically associated with LDL-C, which means it did not suffer a major issue of pleiotropy across the phenome. Moreover, the expression level of *DHODH* in a wide range of tissues (e.g. lung, testis and colon, in which the SARS-CoV-2 receptor angiotensin converting enzyme II (ACE2) is also highly expressed) were all found to be positively associated with LDL-C.

Drug targets prioritised

We further integrated the 726 target-disease MR associations with clinical trial information from Open Targets (20), ChEMBL (21), DrugBank (22), Drug-gene-interaction (DGI) (23), the druggable genome (24) and ClinicalTrial.gov to prioritise drug targets against COVID-19. The 726 associations were further scored based on their putative causal evidence across multiple tissues and their druggability. More specifically, 4 categories of scores were setup for the prioritisation, including omics score, infection score, trial score and druggability score. Each of the scores was scaled between 0 and 1 (**Supplementary Table 11**). Of the 726 target-disease associations, only 499 of them were unique target-disease pairs (due to the same pair appearing in multiple tissues). Two of the 499 associations received high scores in three of the four

categories (making these the highest priority targets), while 77 scored highly in two categories, and 97 scored highly in one category. 323 of the original 499 associations had low scores in all categories and are therefore lowest priority (**Supplementary Table 12**). The top 5 target genes were *DHODH*, *ITGB5*, *JAK2*, *TFRC* and *POR*, scoring 1 in at least two categories and more than zero in at least one other category. *ITGB5* is the target for Cilengitide, which was primarily developed for treating cancers such as glioblastoma. Our study prioritised this target in three categories (omics score=1, druggable score=1, trial score=0.75, infection score=0), which implies that it merits further investigation as a potential target for COVID-19 treatment. In addition, genetically predicted *ITGB5* transcript levels across multiple tissues showed a consistent negative association with three blood pressure phenotypes, which suggests anti-hypertensive potential. None of the 5 top targets were strongly associated with potentially acute comorbidities for COVID-19 treatment such as cardiac dysrhythmias or allergic diseases (excepting the potential beneficial effect of *JAK2* on allergic diseases), which would have raised concerns about their suitability as COVID-19 targets.

As a contrasting example, *TLR9*, the target for a marketed drug, Hydroxychloroquine, showed druggable evidence (druggable score=1, trial score=1) but with a low score for MR evidence across omics/tissues (omics score=0, infection score=0). Despite the reported antiviral activity of Chloroquine against COVID-19 in vitro (25), a recent study emulating a target trial suggested no evidence of any antiviral activity or clinical benefit combining Hydroxychloroquine and Azithromycin for the treatment of severe COVID-19 patients (26). A recent systematic review and meta-analysis of observational evidence also suggested no clinical benefits of Hydroxychloroquine for treatment of COVID-19 patients (27). In addition to the efficacy concerns previously raised for Hydroxychloroquine, our phenome-wide MR results further suggested that reduced *TLR9* expression increases risk of cardioembolic stroke, asthma and some immune related diseases (**Supplementary Table 13**), consistent with some concerns previously expressed in the literature (28).

Discussion

Genetic methods such as MR and colocalization provide a rapid and cost-effective approach to prioritise drug targets in the early stages of drug development (3)(29), providing an indication of efficacy and potential on-target beneficial or adverse effects (30)(31). In this study, we integrated a list of 380 coronavirus-related drug targets from COVID-19 trials and in vitro experiments. We pooled data on genetic associations with these drug targets from 4 recent molecular phenotype genome-side association studies (GWASs) (3)(10)(11)(12), identifying 353 drug targets for which there were reliable genetic predictors which could be used as instruments in MR. By applying tissue-specific MR of gene expression and plasma protein levels, we were able to estimate the causal effects of these targets on 49 viral infection phenotypes, 501 complex diseases and 72 disease related phenotypes, in total estimating 372,482 effects. Of these, potentially causal effects in 726 target-phenotype pairs were estimated with robust MR and colocalization evidence. These results provide evidence of on-target safety issues and/or beneficial effects of the COVID-19 drug targets. To enable this evidence to be widely accessible for drug target research, we constructed an open access browser allowing rapid queries of the drug targets, genetic predictors of the target genes and the disease atlas (http://epigraphdb.org/covid-19/ctda/).

This study supports drug target prioritization for treatment of COVID-19 and other diseases in three ways: 1) investigating potential on-target safety issues for approved drugs under trials for COVID-19 treatment; 2) prioritizing drug targets that merit further evaluation for COVID-19; 3) identifying tissue-specific effects of the drug targets on the human phenome.

First, our study provides evidence of the potential on-target safety issues of marketed drugs in trials for COVID-19 treatment. For example, Baricitinib, a cyclin G-associated kinase inhibitor, has been suggested as a drug for COVID-19 treatment (32) and is now in trials for efficacy evaluation (CTID: NCT04320277). It was proposed to reduce systemic inflammation by inhibiting JAK2 protein. In our study, we observed a putative causal effect of *JAK2* expression on chronic hepatitis, which aligned with literature evidence of hepatitis B reactivation following ruxolitinib treatment (inhibitor of JAK2) (33). Our phenome-wide association analysis further suggested that reducing the gene expression of *JAK2* was not strongly associated with increased risk of acute conditions (cardiac dysrhythmias or allergic diseases). Although JAK2 reduction may increase weight and risk of inflammatory bowel disease, potential Baricitinib use for COVID-19 treatment will be short-term, and these side effects are probably inconsequential.

Second, our prioritization approach highlighted three drugs with potential for COVID-19 treatment, including Leflunomide (target: *DHODH*), Cilengitide (target: *ITGB5*), and E2.3 (under Phase I trial, target: *TFRC*). Leflunomide is an immune suppressive drug used for autoimmune diseases, such as rheumatoid arthritis (34). It also showed antiviral effect against a wide range of DNA and RNA viruses, such as herpes simplex virus (35), human cytomegalovirus (36)(37), polyoma BK virus (38)(39), Junín virus (40), influenza A virus, Zika virus and Ebola virus. Recent studies also suggested its anti-SARS-CoV-2 effect (18)(19). In addition, our MR finding suggested the lipid lowering potential of Leflunomide use, which merits further investigation. For Cilengitide (*ITGB5*), our MR analyses suggested that reduced gene expression level of *ITGB5* showed a consistent effect on lowering systolic and diastolic blood pressure level and lowering hypertension risk. This implies its potential anti-hypertensive effect. Hypertension has been suggested as a potential risk factor for complications of COVID-

19 (41). Cilengitide could be a promising candidate drug for COVID-19 treatment with anti-hypertensive potential as well.

Third, the tissue-specific results suggested that expression levels of some targets across multiple tissues were associated with the same phenotypes, while some other targets showed apparent tissue-specific effects on diseases/phenotypes. For example, expression levels of *DHODH* and *ITGB5* in multiple tissues showed similar causal effects on LDL-C and blood pressure respectively. When designing trials based on these drug targets, careful pre-clinical investigation is needed to understand the tissue-specific mechanisms of the targets on tested indications.

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/) and in the UK Biobank (42), no such GWAS of COVID-19 phenotypes has yet been performed. Our evaluation of potential targets therefore excludes estimated effects of these targets on COVID-19 disease risk or progression. In the near future such data is likely to become available in a number of large-scale population-based biobanks across the world, including the UK Biobank, China Kadoorie Biobank (43), HUNT study (44), FinnGen (https://www.finngen.fi/fi), DeCODE (https://www.decode.com/) and the Million Veteran Program (45). These biobanks are designed to study long-term health conditions, but with data linkage between these biobanks and electronic health records and/or COVID-19 data (as has recently happened with UK Biobank), these biobanks could be tremendous resources to rapidly generate genetic association data in epidemic and pandemic situations. The recent initial GWAS of COVID-19 found no GWAS hits, which highlight the importance of seriously considering potential bias of the data. When the unbiased data are available, the drug target information and the genetic predictors of the 353 drug targets we curated in this study as well as the drug target prioritization pipeline could provide even more valuable insights into the potential drug targets for infectious disease treatment. 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. 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, we can’t predict whether successful intervention would impact on risk of infection, progression/severity of disease or other disease characteristics relevant to public health (e.g. viral shedding).

In conclusion, this study identified 726 putatively causal effects between the 353 COVID-19 related targets and the human phenome using MR and colocalization approaches. These associations support causality but do not prove it, as horizontal pleiotropy remains an alternative possibility. Our study provides both a statistical genetics pipeline and an openly available platform to prioritise potential COVID-19 targets on the basis of MR and existing trial evidence. The platform also evaluated repurposing opportunities for approved drugs on other indications. Our approach and platform will provide additional value once genetic association studies of COVID-19 are available in the near future.

Materials and Methods

Selection of drug targets of COVID-19

In this study, drug targets against COVID-19 were selected from three resources: i) a list of 11 drugs under trials for COVID-19 treatment were extracted from *ClinicalTrials.gov* (https://clinicaltrials.gov/ct2/results?cond=COVID-19;). 10 of these 11 drugs were mapped to their target genes using Drug-Gene-Interaction (DGI) database (http://dgidb.org/) (23) and CHEMBL database (21) (**Supplementary Table 1**); ii) 332 human proteins interacting with SARS-CoV-2 proteins from Gordon et al. (8); iii) 44 genes associated with SARS-CoV from Gralinski et al. (9). After de-duplicating these, 380 unique drug targets were selected for our study (**Supplementary Table 2**).

Selection of genetic instruments for the target genes

Next, gene and protein expression levels of these targets were looked up from four resources: protein expression levels in plasma from four studies (29)(46)(47)(48) implemented in Zheng et al. (3), gene expression levels in whole blood from eQTLGen consortium (10), tissue specific gene expression levels in 7 tissues from the GTEx consortium (11) and gene expression levels in two kidney tissues from Gilles et al. (12). 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 2**).

The genetic variants, genes and proteins were further mapped to genome build GRCh37.p13 coordinates and 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 with 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, 1493 instruments were kept for the genetic analyses of this study (**Supplementary Table 3**).

Identifying cis and trans instruments

We split instruments into two groups: 1) 1269 cis-acting instruments within a 500Kb window from each side of the leading pQTL of the protein/gene; 2) 224 instruments outside the 500Kb window of the leading pQTL 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 3**).

Association of genetic variants with human phenotypes

We obtained effect estimates for the association of the proteins and genes with viral infection phenotypes and complex human phenotypes using GWAS summary statistics which were obtained from the GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics) (49), IEU GWAS database (http://gwas.mrcieu.ac.uk/) (14) and SAIGE UK Biobank data release

(https://www.leelabsg.org/resources) (50). We used the following inclusion criteria to select complex phenotypes to be analysed:

1. The GWAS with the greatest expected statistical power (e.g. largest sample size or number of cases) when multiple GWAS records of the same disease or quantitative phenotype were available.
2. GWAS with effect sizes, standard errors and effect alleles for all tested variants (i.e. full GWAS summary statistics available).

After selection, 49 viral infection phenotypes and 501 complex diseases were defined as primary outcomes. 72 quantitative phenotypes were defined as secondary outcomes (**Supplementary Table 4**).

Causal inference and sensitivity analyses

Mendelian randomization analysis

In the initial MR analysis, 55 instruments of 39 plasma proteins (N=3,301) and 833 instruments of 331 gene transcripts in whole blood (N=31,684) were used as the exposures to proxy the effect of the 353 drug targets for COVID-19 (**Supplementary Table 3**), which maximized the statistical power of the study and supported us to obtain an overall view of the target-phenotype associations across multiple disease areas. 49 viral infection phenotypes, 501 complex diseases and 72 quantitative phenotypes were used as the outcomes. Due to missingness of available exposure or outcome data, the causal relations of 353 targets on 622 human phenotypes were tested in this study. We selected a P-value threshold of 0.05, corrected for the number of independent tests, as our threshold for prioritising MR results for follow up analyses (number of tests= 45,590; $P< 1.1x10^{-6}$). The Wald ratio (16) method was used to obtain MR effect estimates for targets with only one instrument, where the Wald ratio MR effect estimates were sensitive to the particular choice of instrument. For targets with two or more instruments, the IVW method (17) was used to estimate the causal effects.

Colocalization analysis

Results that survived the Bonferroni corrected threshold ($P< 1.1x10^{-6}$) in the MR analysis were evaluated using colocalization analysis, which estimates the posterior probability of each genomic locus containing a single variant affecting both the target and the phenotype (15). The prior probabilities that a variant is equally associated with each phenotype ($p1=1\times10^{-4}$; $p2=1\times10^{-4}$) and both phenotypes jointly ($p12=1\times10^{-5}$) were used for this analysis. 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 same colocalization approach was applied to cis- and trans- instruments. 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".

Tissue Specificity analysis

The functional receptor of SARS-CoV-2, ACE2 (51), 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 (52) and kidney tubules (53). The presence of SARS-CoV-2 in the alimentary tract for longer than in the respiratory system (54) 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 (11)(12). 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 3**). The same MR and colocalization analysis pipeline were applied for the tissue specific analysis.

Analysis software

The MR analyses (including Wald ratio, single SNP MR, IVW) were conducted using the MR-Base TwoSampleMR R package (github.com/MRCIEU/TwoSampleMR) (14). 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).

Triangulation of evidence with drug trial information

For the 154 drug targets with suggestive MR and colocalization evidence (P of MR < $1x10^{-5}$ and colocalization probability >70%), we mapped the drug targets with related drug names using four platforms: OpenTargets (20), ChEMBL (21), DrugBank (22) and DGI platforms (23). After this analysis, we mapped 40 targets to their targeted drugs (**Supplementary Table 14**). The trial information of these 40 drugs were further looked up from ClinicalTrial.gov or WHO ATC (https://www.whocc.no/atc_ddd_index/) websites, which includes the indication, the trial status (recruiting or completed), the action type of the drug (inhibitor or antagonist), the ClinicalTrial.gov ID (CTID) or ATC ID and maximum trial phase.

We further mapped the targets to the previously reported "druggable genome" (24). 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 COVID-19 drug targets, 16 targets were mapped to tier 1, 4 targets to tier 2 and 17 targets to tier 3 (**Supplementary Table 12**).

Evidence-based drug targets prioritisation

For the 726 target-disease associations with suggestive MR and colocalization evidence, we prioritised the COVID-19 related drug targets by four separate categories of scores: 1) Omics

score (0 or 1), which estimates whether a specific target-disease association showed robust MR evidence in both proteome and transcriptome levels or in more than one tissue; 2) Infection score (0 or 1), which refers to whether the target showed strong association with at least one of the viral infection phenotypes we tested; 3) Druggable score (0 to 1), which refers to the druggability of the target (data from Finnan et al); 4) Trial score (0 to 1), which refers to the trial stage of the target (e.g. Pre-clinical to Phase 4). More details of the score system is listed in **Supplementary Table 11**.

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

We have made all 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 372,482 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 name, outcome disease name, and the tissues 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. Drugs in trials for COVID-19 treatment and their target genes.

TableS2. The available genetic information for SARS-CoV-2 target genes from pQTL and tissue specific eQTL resources.

TableS3. The genetic predictors and association information of the SARS-CoV-2 target genes.

TableS4. The list of human phenotypes been used as outcomes for the omics study.

TableS5. The target-viral infection phenotype associations with suggestive Mendelian randomization and colocalization evidence.

Note: The target- infection disease associations with Mendelian randomization p value < 10^{-3} were included in this table.

TableS6. The plasma protein-phenotype associations with strong MR and colocalization evidence.

Note: The target-disease associations passed Bonferroni corrected threshold (1.1×10^{-6}) and with robust colocalization evidence (probability > 70%) were included in this table.

TableS7. The blood gene expressions-phenotype associations with strong MR and colocalization evidence.

Note: The target-disease associations passed Bonferroni corrected threshold (1.1×10^{-6}) and with robust colocalization evidence (probability > 70%) were included in this table.

TableS8. The tissue-specific gene expressions-phenotype associations with strong MR and colocalization evidence.

Note: The target-disease associations passed Bonferroni corrected threshold (1.1×10^{-6}) and with robust colocalization evidence (probability > 70%) were included in this table.

TableS9. The target-disease associations with strong MR evidence but without strong colocalization evidence.

Note: The target-disease associations passed Bonferroni corrected threshold ($1.1x10^{-6}$) but without robust colocalization evidence (probability < 70%) were included in this table.

TableS10. Number of unique targets each tested human phenotype associated with.

Note: The target-phenotype associations counted in this table were associations passed Bonferroni corrected threshold ($1.1x10^{-6}$) and with robust colocalization evidence (probability > 70%).

TableS11. The score system been used to prioritize drug targets for COVID-19.

Note: Four categories of scores were setup, including omics score, infection score, trial score and druggability score. Each of the scores was scaled between 0 and 1.

TableS12. The prioritization scores for the top target-phenotype associations.

Note: For each target-phenotype association, its four scores were listed in this table together with its drug trial information.

TableS13. The top target-phenotype association for one drug target gene TLR9.

Note: The exposure here is the gene expression level of TLR9 in blood and testis, the outcomes were human phenotypes, the association beta, standard error and p value were reported in this table.

TableS14. 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. The druggability information were extracted from Finan et al (24).

References and Notes:

1. Plenge RM. Disciplined approach to drug discovery and early development. Sci Transl Med. 2016 Jul 27;8(349):349ps15.

2. 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.

3. Zheng J, Haberland V, Baird D, Walker V, Haycock P, Gutteridge A, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. 2019. BioRxiv. https://www.biorxiv.org/content/10.1101/627398v1

4. 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.

5. 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

6. 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. MedRxiv. https://www.medrxiv.org/content/10.1101/2020.04.10.20060699v1.abstract

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

8. 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

9. 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.

10. 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. 2018. BioRxiv, https://www.biorxiv.org/content/10.1101/447367v1

11. 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. 2019. BioRxiv. https://www.biorxiv.org/content/10.1101/787903v1

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

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

18. Robinson PC, Yazdany J. The COVID-19 Global Rheumatology Alliance: collecting data in a pandemic. Nat Rev Rheumatol. 2020 Apr 2.

19. Xiong R, Zhang L, Li S, Sun Y, Ding M, Wang Y, et al. Novel and potent inhibitors targeting DHODH, a rate-limiting enzyme in de novo pyrimidine biosynthesis, are broad-spectrum antiviral against RNA viruses including newly emerged coronavirus SARS-CoV-2. 2020. BioRxiv. https://www.biorxiv.org/content/10.1101/2020.03.11.983056v1

20. 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.

21. 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.

22. 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.

23. 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.

24. 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).

25. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020 Mar;30(3):269–71.

26. Molina JM, Delaugerre C, Goff JL, Mela-Lima B, Ponscarme D, Goldwirt L, et al. No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. Med Mal Infect. 2020 Mar 30.

27. Shamshirian A, Hessami A, Heydari K, Alizadeh-Navaei R, Ebrahimzadeh MA, Ghasemian R, et al. Hydroxychloroquine Versus COVID-19: A Rapid Systematic Review and Meta-Analysis. medRxiv. 2020 Apr 20;2020.04.14.20065276.

28. Chen C-Y, Wang F-L, Lin C-C. Chronic hydroxychloroquine use associated with QT prolongation and refractory ventricular arrhythmia. Clin Toxicol. 2006;44(2):173–5.

29. 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.

30. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017 Oct;14(10):577–90.

31. Sekula P, Del Greco M F, Pattaro C, Köttgen A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. J Am Soc Nephrol. 2016 Nov;27(11):3253–65.

32. Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet. 2020 Feb 15;395(10223):e30–1.

33. Caocci G, Murgia F, Podda L, Solinas A, Atzeni S, La Nasa G. Reactivation of hepatitis B virus infection following ruxolitinib treatment in a patient with myelofibrosis. Leukemia. 2014 Jan;28(1):225–7.

34. Sanders S, Harisdangkul V. Leflunomide for the treatment of rheumatoid arthritis and autoimmunity. Am J Med Sci. 2002 Apr;323(4):190–3.

35. Josephson MA, Javaid B, Kadambi PV, Meehan SM, Williams JW. Leflunomide in Solid Organ Transplantation and Polyoma Virus Infection. In: Ahsan N, editor. Polyomaviruses and Human Diseases. New York, NY: Springer New York; 2006. p. 255–65.

36. Ubel PA, Baron J, Asch DA. Social responsibility, personal responsibility, and prognosis in public judgments about transplant allocation. Bioethics. 1999 Jan;13(1):57–68.

37. Evers DL, Wang X, Huong S-M, Andreoni KA, Huang E-S. Inhibition of human cytomegalovirus signaling and replication by the immunosuppressant FK778. Antiviral Res. 2005 Jan;65(1):1–12.

38. Bernhoff E, Tylden GD, Kjerpeseth LJ, Gutteberg TJ, Hirsch HH, Rinaldo CH. Leflunomide inhibition of BK virus replication in renal tubular epithelial cells. J Virol. 2010 Feb;84(4):2150–6.

39. Jeffers-Francis LK, Burger-Calderon R, Webster-Cyriaque J. Effect of Leflunomide, Cidofovir and Ciprofloxacin on replication of BKPyV in a salivary gland in vitro culture system. Antiviral Res. 2015 Jun;118:46–55.

40. Sepúlveda CS, García CC, Damonte EB. Antiviral activity of A771726, the active metabolite of leflunomide, against Junín virus. J Med Virol. 2018 May;90(5):819–27.

41. Gandhi RT, Lynch JB, Del Rio C. Mild or Moderate Covid-19. N Engl J Med. 2020 Apr 24.

42. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015 Mar;12(3):e1001779.

43. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011 Dec;40(6):1652–66.

44. Holmen J, Midthjell K, Krüger Ø, Langhammer A, Holmen TL, Bratberg GH, et al. The Nord-Trøndelag Health Study 1995-97 (HUNT 2). Norsk epidemiologi. 2003;13(1).

45. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol. 2016 Feb;70:214–23.

46. 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.

47. 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.

48. 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.

49. 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.

50. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018 Sep;50(9):1335–41.

51. 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.

52. 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.

53. 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. MedRxiv. https://www.medrxiv.org/content/10.1101/2020.03.04.20031120v3.abstract

54. 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.

Acknowledgments and 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).

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.

Competing interests: No competing interests

Figures:

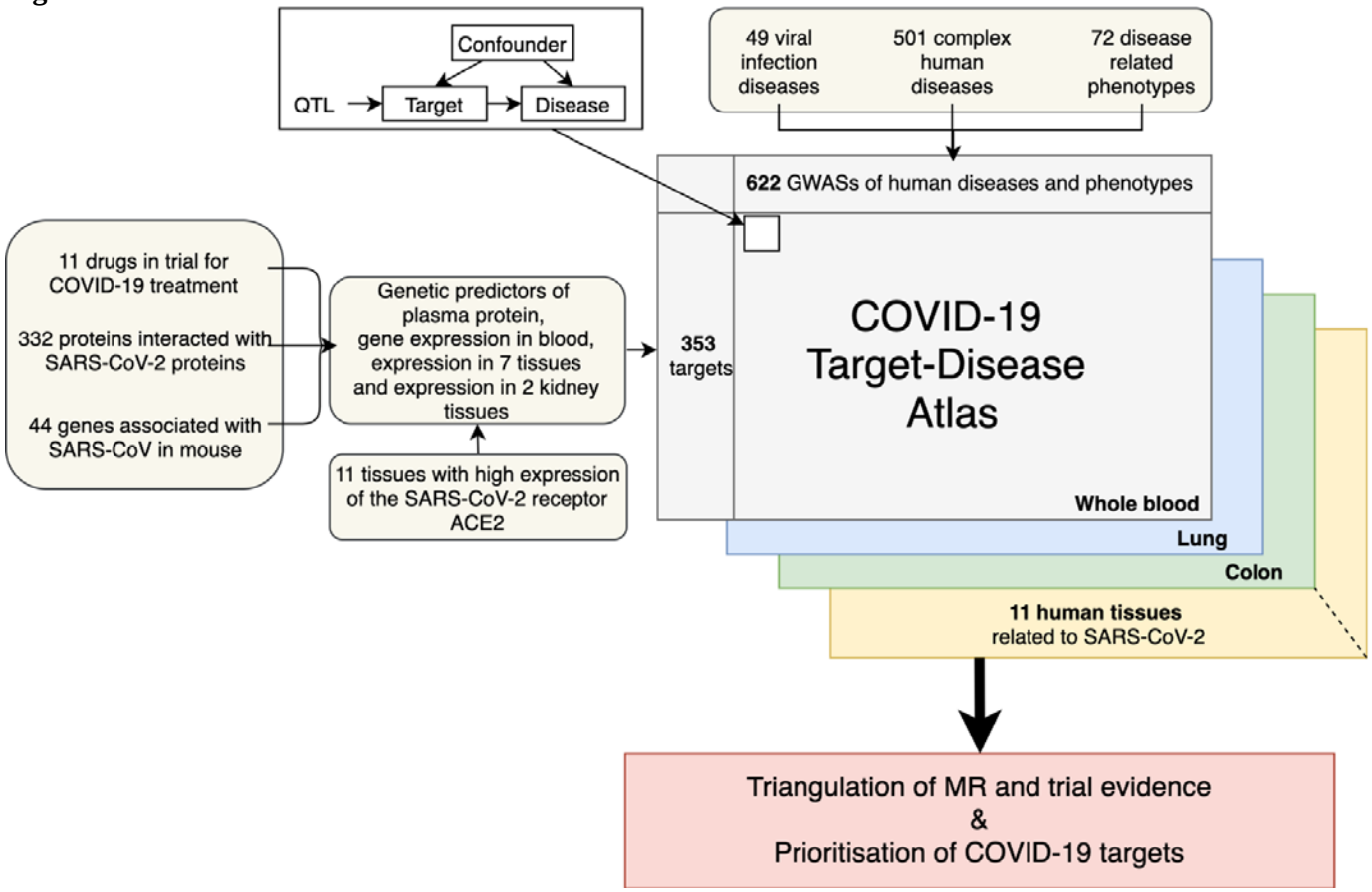
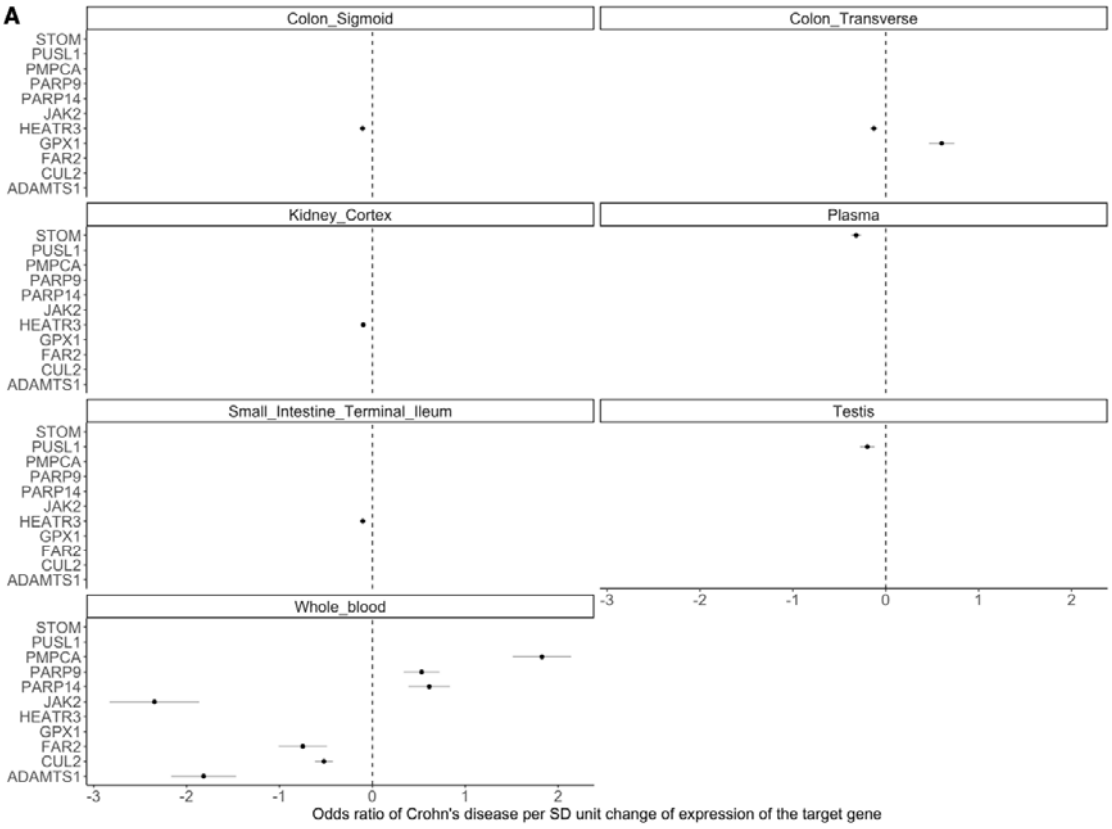
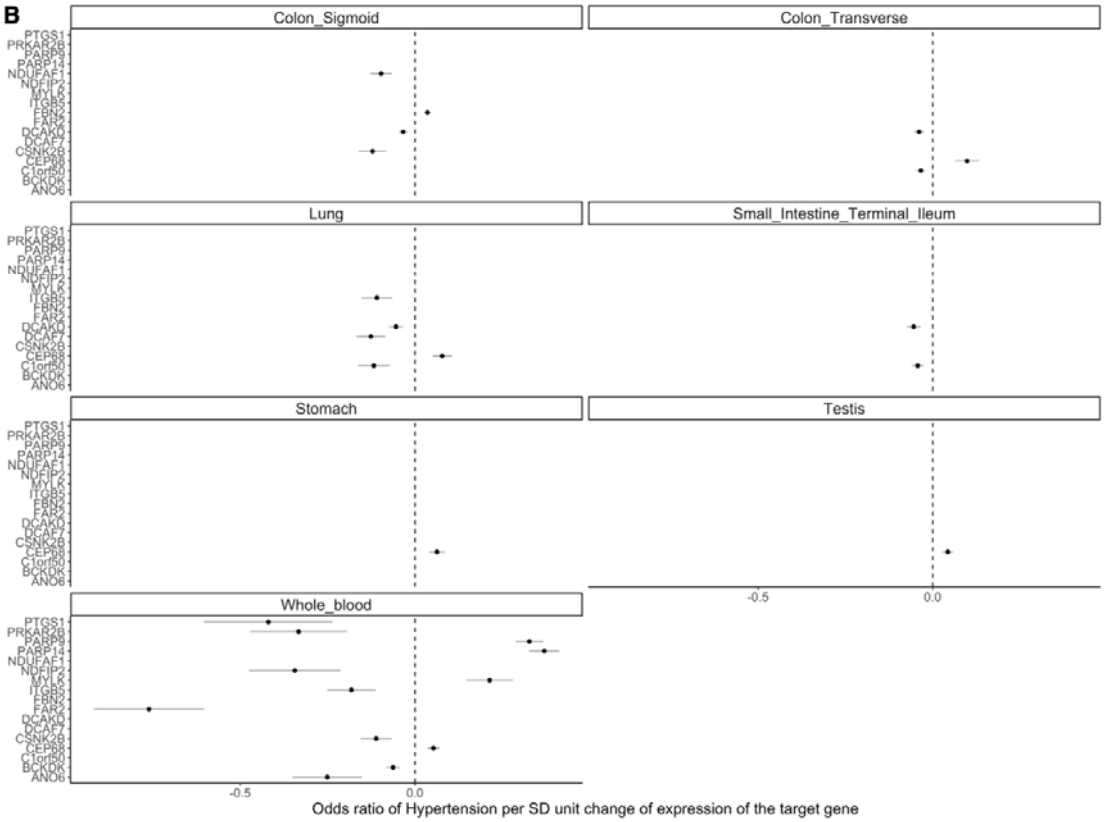


Figure 1. Study design of this multi-omics MR study of COVID-19 targets on human phenome and drug target prioritization. The diagram on the top left corner is the Mendelian randomization model been used in this study. The QTL refers to the genetic variant(s) robustly associated with the protein/gene expression levels of the COVID-19 target genes.


A
Colon_Sigmoid
Colon_Transverse
Kidney_Cortex
Plasma
Small_Intestine_Terminal_Ileum
Testis
Whole_blood
STOM
PUSL1
PMPCA
PARP9
PARP14
JAK2
HEATR3
GPX1
FAR2
CUL2
ADAMTS1
-3
-2
-1
0
1
2
Odds ratio of Crohn's disease per SD unit change of expression of the target gene


B
Colon_Sigmoid
Colon_Transverse
Lung
Small_Intestine_Terminal_Ileum
Stomach
Testis
Whole_blood
PTGS1
PRKAR2B
PARP9
PARP14
NDUFAF1
NDFIP2
MYLK
ITGB5
FBN2
FAR2
DCAKD
DCAF7
CSNK2B
CEP68
C1orf50
BCKDK
ANO6
-0.5
0.0
Odds ratio of Hypertension per SD unit change of expression of the target gene

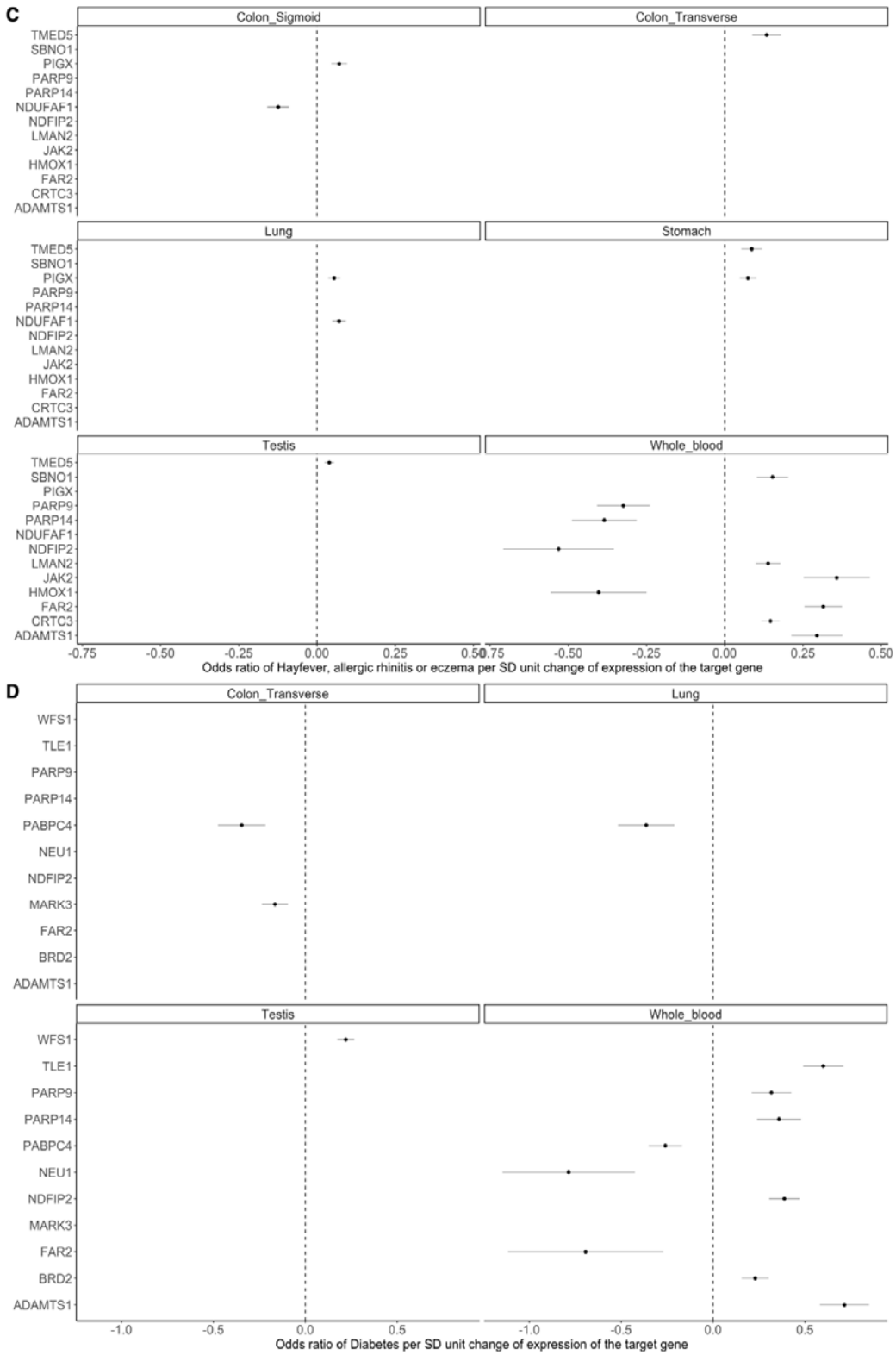


Figure 2. Forest plot of the Mendelian randomization associations of targets on 4 disease entities across multiple tissues.

The four forest plots refer to the MR associations of Crohn's disease (A), hypertension (B), hay fever/allergic rhinitis or eczema (C) and diabetes (D). The X-axis is the odds ratio of the MR estimates.

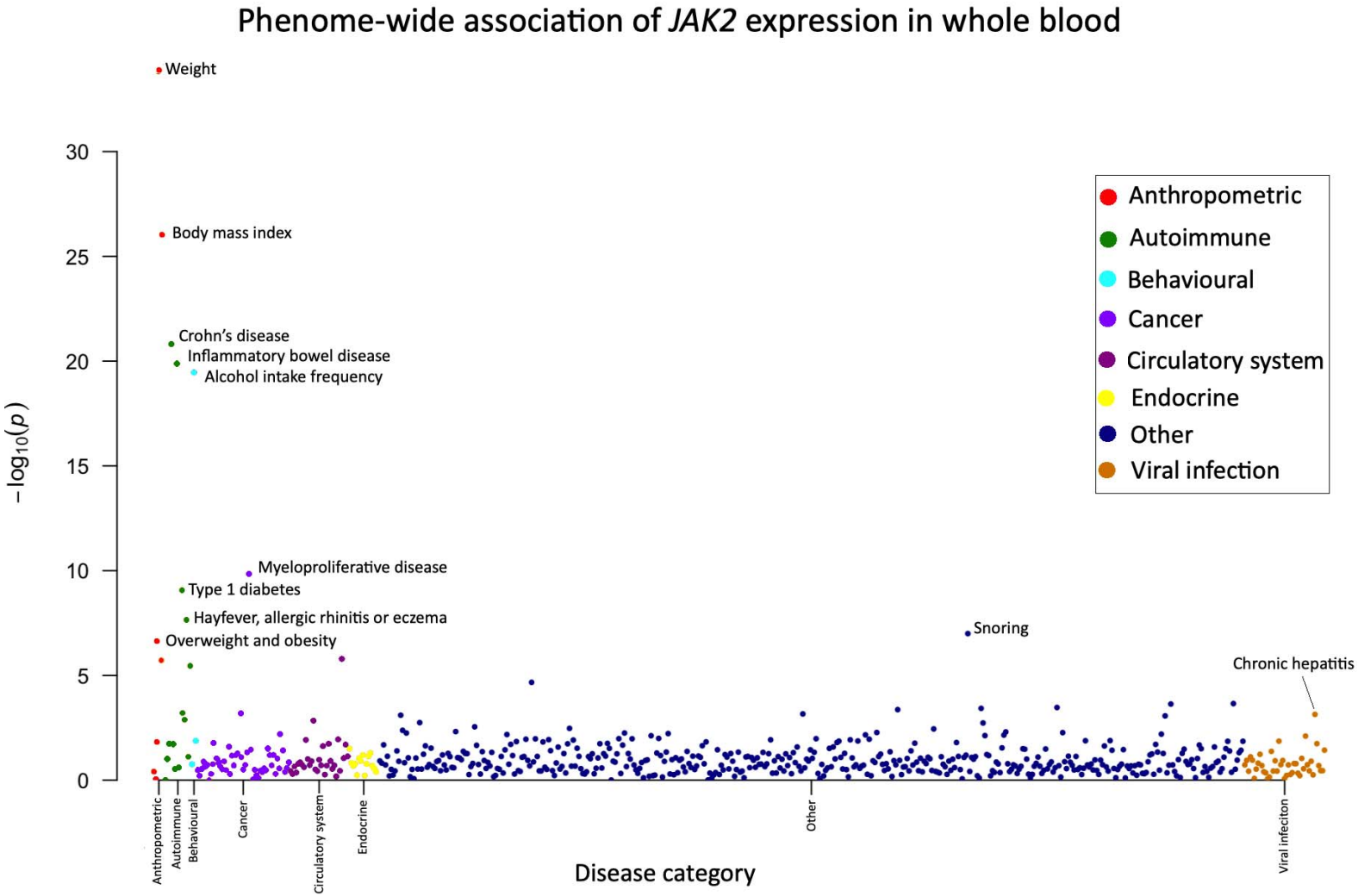


Figure 3. Manhattan plot demonstrating phenome-wide associations of *JAK2* expression in whole blood on human phenotypes. Different colors refer to different disease areas.

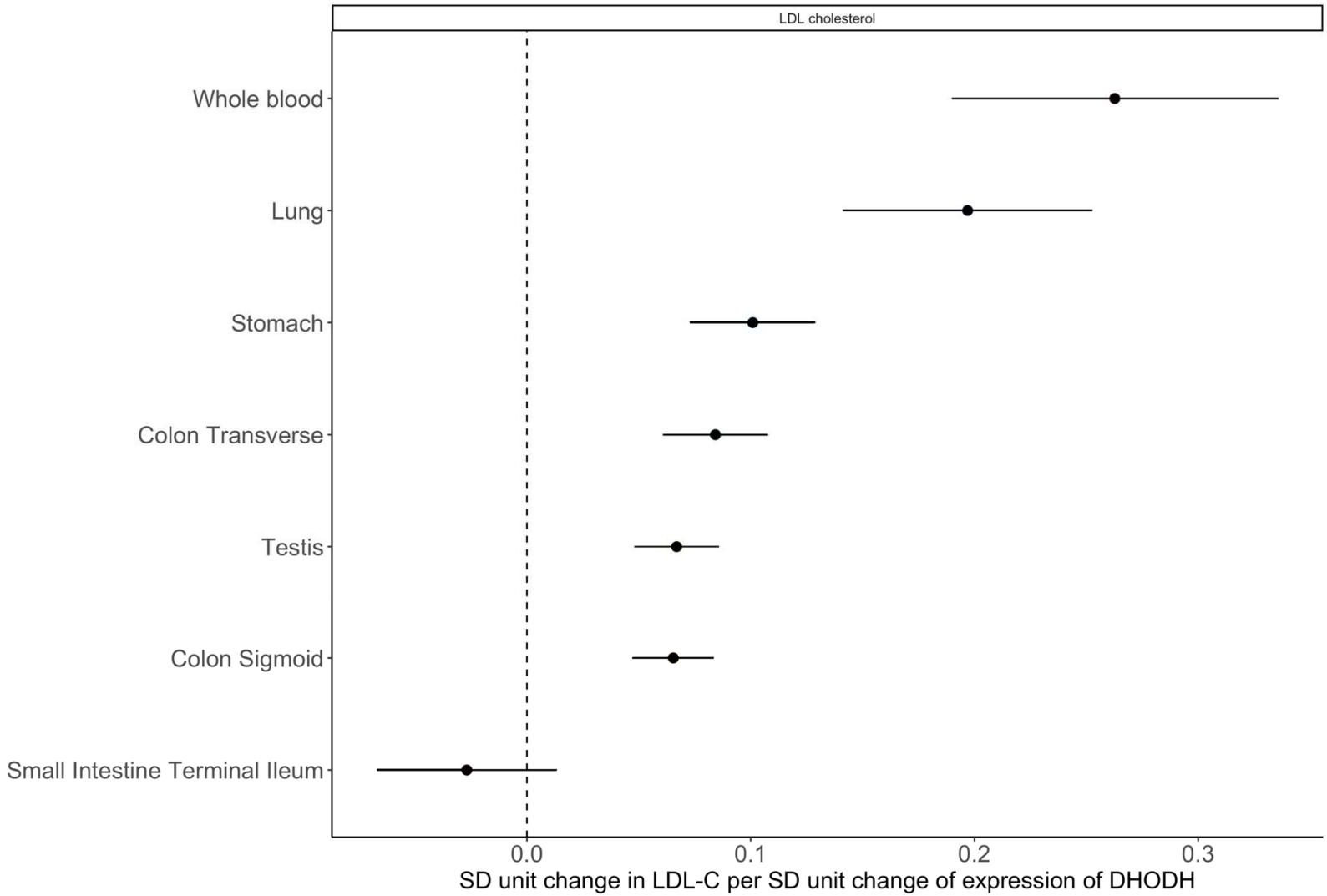


Figure 4. Forest plot demonstrating the tissue-dependent effect of *DHODH* expression on LDL cholesterol.