

Exploring diseases/traits and blood proteins causally related to expression of ACE2, the putative receptor of SARS-CoV-2: A Mendelian Randomization analysis highlights tentative relevance of diabetes-related traits

Shitao RAO, Alexandria LAU and Hon-Cheong SO

Supplementary Text

Mendelian randomization methods

Brief introduction to MR

MR makes use of genetic variants as “instruments” to represent the exposure of interest, and aims to infer causal relationship between the exposure and outcome (1). Susceptibility genetic variants for the exposure are usually derived from genome-wide association studies (GWAS), which is still observational in nature. However, in an MR analysis, in essence we are considering the ‘genetically predicted’ exposure (e.g. genetically-predicted lipid level or probability of a disease) as the risk factor, hence it is different from more conventional case-control/cohort studies.

MR is usually much less prone to reverse causation, when compared to conventional case-control or cohort studies with (non-genetic) risk factors, as genetic variants are fixed at conception and precedes the outcomes. Secondly, conventionally measured exposures are often associated with a wide range of behavioral, social and physiological factors that confound associations with outcomes. Genetic variants, in general, are less likely to be strongly associated with a wide variety of confounders such as age or environmental factors. Empirical evidence suggests that generally there is a lack of substantial confounding of genetic variants with factors that usually confound exposures in conventional epidemiological studies (2-4). In addition, in most MR analyses (including the present study), multiple genetic variants are usually employed as instruments for the exposure. As argued by Smith and Ebrahim (see <https://www.ncbi.nlm.nih.gov/books/NBK62433/>), this may serve as another mechanism to reduce the effects of confounding, as it is unlikely that a large number of instruments are all subject to confounding. Heterogeneity/outlier detection tests may also detect violation of instrumental variable or MR model assumptions. Such tests were also carried out in this study; we did not find evidence of heterogeneity among our significant findings.

However, we note that the above is not foolproof. MR may not achieve the same level of confounding control as the ‘gold-standard’ randomized controlled trials (RCT). In addition, MR reflects long-term exposure of the risk factor (as genetic variants are fixed at birth), which may not reflect effects in a shorter term. MR may provide one way to guide the study of causal relationship between risk factors and outcomes, but there are also limitations. For more detailed discussions, we refer the readers to relevant reviews on the topic (2; 5-8).

Approaches to MR analysis

We conducted MR primarily with the ‘inverse-variance weighted’ (MR-IVW) (9) and Egger regression (MR-Egger) (10) approaches, which are among the most widely used MR methods. For exposure with only one instrument, the Wald ratio method was used. One of the concerns of MR is horizontal pleiotropy, in which the genetic instruments have effects on the outcome other than through effects on the exposure. Of note, MR-Egger gives valid estimates of causal effects in the presence of imbalanced or directional horizontal pleiotropy. In addition, significance of the MR-Egger intercept can be used to judge whether significant directional/imbalanced pleiotropy is present.

For selected traits with stronger evidence of association, we also performed further analysis by GSMR and MR-RAPS. GSMR (<http://cnsgenomics.com/software/gsmr/>) can take into account of (imbalanced) horizontal pleiotropy but is based on a different principle from MR-Egger. It excludes ‘outlier’ or heterogeneous genetic instruments that may contribute to pleiotropy, by the ‘HEIDI-outlier’ method (11). GSMR also employed a slightly different formula from MR-IVW by modelling variance of both $\hat{\beta}_{XG}$ and $\hat{\beta}_{YG}$, and accounts for correlated SNPs (11) by modelling linkage disequilibrium (LD) between SNPs. We tried several r^2 clumping thresholds ($r^2 = 0.001, 0.05, 0.1, 0.15, 0.2$) for GSMR, with SNP correlation matrices extracted from 1000 Genomes European samples. The authors showed by extensive simulations (11) that GSMR produced well-calibrated test statistics under the null, and that the causal effect estimates are unbiased under H_1 .

MR-RAPS is another MR analysis methodology which can take into account multiple weak instruments by a robust procedure known as ‘Robust Adjusted Profile Score’ (RAPS). Details of MR-RAPS were described in Zhao et al. (12). The authors performed simulations and showed that under certain assumptions, the method is able to produce relatively unbiased causal estimates with proper 95% CI coverage rate, while also having lower variance in the estimates. More specifically, MR-RAPS is valid under ‘systematic and idiosyncratic pleiotropy’, in which most pleiotropic effects are normally distributed with mean zero but some SNPs are allowed to exhibit much larger pleiotropic effects (12). Based on the strength of MR-RAPS to include weak instruments, we employed a more relaxed p-value threshold for SNP selection (0.01) for MR-RAPS. The threshold was chosen as the authors’ simulations showed that unbiased causal estimates can be achieved by MR-RAPS, up to a p-threshold of 0.01 [Table 5 of ref(12)].

A note on ethnicity of GWAS samples and MR analysis

Most GWAS included in this work were based on predominantly European samples. However, subjects of other ethnicities were included in some samples. Ancestry can be a potential confounder that may lead to violation of MR assumptions, however adjustment for ancestry can be made which can avoid spurious associations(5). As shown in Table S2a, proper adjustment by principal component analysis or mixed models have been performed for GWAS prior to MR analysis.

We believe that spurious causal associations are unlikely; however, it is still possible for genetic

associations to differ across ethnicities, which may affect the causal estimates of MR. For example, this may occur if some SNP-exposure or SNP-outcome associations are weaker/stronger in one ethnic group than another. However, in type 2 diabetes, most studies to date did not detect significant effect heterogeneity across ethnicities for the majority of loci (13; 14). In addition, if genetic associations of some SNPs differ substantially across ethnicities, we may expect the causal effect conferred by some genetic instruments to deviate from others more than expected by chance. This appears to be not the case as we did not observe significant heterogeneity in causal estimates across SNPs for most traits (Table 1). However, it is difficult to assess the possibility of differential genetic effects across ethnicities for all traits, hence we consider it as a potential limitation.

Highlights of pathways (that are enriched for proteins causally linked to *ACE2* expression)

Here we highlight a few pathways which were enriched for proteins causally linked to *ACE2* expression in MR. Note that this is an exploratory analysis and the definitive role of these pathways remain to be elucidated in further experimental studies.

One of the top-ranked pathways is cytokine- cytokine receptor interaction. There was evidence from previous studies that inflammatory processes and *ACE2* activity are closely related (reviewed in (15)), but their relationship is complex. Some experimental studies showed that the *ACE2*/Angiotensin-(1-7)/Mas receptor axis is associated with reduction in cytokine release and inhibition of tissue fibrosis in various diseases. Nevertheless, clinical evidence is lacking and the mechanism remains to be understood (16). Another highlighted pathway involves HIF (hypoxia-inducible factor) signaling. Interestingly, in a study on pulmonary artery smooth muscle cells, Zhang et al. showed that *ACE2* mRNA and protein levels increased during the early stages of hypoxia. The levels reduced afterwards after HIF-1 α accumulation (17).

Another pathway involves VEGFA-VEGFR2 Signaling. A recent animal study showed that VEGF could influence *ACE2*. Using CD34⁺ cells derived from mice, Joshi et al. demonstrated that VEGF could mimic the effects of hypoxia on CD34⁺ cells, leading to increased activity of *ACE2* (18).

References

1. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* 2003;32:1-22
2. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89-98
3. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human genetics* 2008;123:15-33
4. Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;4:e352
5. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and

checklist for clinicians. *BMJ* 2018;362:k601

6. Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart* 2017;103:1400-1407

7. Evans DM, Smith GD. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annual Review of Genomics and Human Genetics* 2015;16:327-350

8. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *International journal of epidemiology* 2003;32:1-22

9. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology* 2013;37:658-665

10. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology* 2015;44:512-525

11. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR, Yang J. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature communications* 2018;9:224

12. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *arXiv preprint arXiv:180109652* 2018;

13. Ng MC, Shriner D, Chen BH, Li J, Chen WM, Guo X, Liu J, Bielinski SJ, Yanek LR, Nalls MA, Comeau ME, Rasmussen-Torvik LJ, Jensen RA, Evans DS, Sun YV, An P, Patel SR, Lu Y, Long J, Armstrong LL, Wagenknecht L, Yang L, Snively BM, Palmer ND, Mudgal P, Langefeld CD, Keene KL, Freedman BI, Mychaleckyj JC, Nayak U, Raffel LJ, Goodarzi MO, Chen YD, Taylor HA, Jr., Correa A, Sims M, Couper D, Pankow JS, Boerwinkle E, Adeyemo A, Doumatey A, Chen G, Mathias RA, Vaidya D, Singleton AB, Zonderman AB, Igo RP, Jr., Sedor JR, Consortium F, Kabagambe EK, Siscovick DS, McKnight B, Rice K, Liu Y, Hsueh WC, Zhao W, Bielak LF, Kraja A, Province MA, Bottinger EP, Gottesman O, Cai Q, Zheng W, Blot WJ, Lowe WL, Pacheco JA, Crawford DC, e MC, Consortium D, Grundberg E, Mu TC, Rich SS, Hayes MG, Shu XO, Loos RJ, Borecki IB, Peyser PA, Cummings SR, Psaty BM, Fornage M, Iyengar SK, Evans MK, Becker DM, Kao WH, Wilson JG, Rotter JI, Sale MM, Liu S, Rotimi CN, Bowden DW, Consortium ME-aotDiAA. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS genetics* 2014;10:e1004517

14. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjie C, Huffman JE, Assimes TL, Judy RL, Huang J. Discovery of 318 novel loci for type-2 diabetes and related micro-and macrovascular outcomes among 1.4 million participants in a multi-ethnic meta-analysis. *medRxiv* 2019:19012690

15. Ravinder Reddy G, Stephen C, Madhav B. ACE and ACE2 in Inflammation: A Tale of Two Enzymes. *Inflammation & Allergy - Drug Targets (Discontinued)* 2014;13:224-234

16. Thiago Ruiz Rodrigues P, Natalia Pessoa R, Aline Silva M, Antônio Lúcio T, Ana Cristina S-e-S. The Anti-Inflammatory Potential of ACE2/Angiotensin-(1-7)/Mas Receptor Axis: Evidence from Basic and Clinical Research. *Current Drug Targets* 2017;18:1301-1313

17. Zhang R, Wu Y, Zhao M, Liu C, Zhou L, Shen S, Liao S, Yang K, Li Q, Wan H. Role of HIF-1 α in the

regulation ACE and ACE2 expression in hypoxic human pulmonary artery smooth muscle cells. American journal of physiology Lung cellular and molecular physiology 2009;297:L631-640

18. Joshi S, Wollenzien H, Leclerc E, Jarajapu YP. Hypoxic regulation of angiotensin-converting enzyme 2 and Mas receptor in human CD34⁺ cells. Journal of Cellular Physiology 2019;234:20420-20431