medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Exploring diseases/traits and blood proteins causally related to expression of ACE2, the putative receptor of 2019-nCov: A Mendelian Randomization analysis

Shitao RAO[1*], Alexandria LAU[1*] and Hon-Cheong SO[1,2,3^]

[1]School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong

[2]KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research of Common Diseases, Kunming Institute of Zoology and The Chinese University of Hong Kong, China

[3]CUHK Shenzhen Research Institute, Shenzhen, China

**these authors contributed equally to this work*

^Corresponding author

Correspondence to: Hon-Cheong So, Lo Kwee-Seong Integrated Biomedical Sciences Building, The Chinese University of Hong Kong, Shatin, Hong Kong. Tel: +852 3943 9255; E-mail: hcso@cuhk.edu.hk

Abstract

The novel coronavirus 2019-nCoV has caused major outbreaks in many parts of the world. A better understanding of the pathophysiology of COVID-19 is urgently needed. Clinically, it is important to identify who may be susceptible to infection and identify treatments for the disease.

There is good evidence that ACE2 is a receptor for 2019-nCoV, and studies also suggested that high expression of ACE2 may increase susceptibility to infection. Here we conducted a phenome-wide Mendelian randomization (MR) study to prioritize diseases/traits and blood proteins that may be causally linked to *ACE2* expression in the lung. Expression data was based on GTEx. We also explored drug candidates whose targets overlapped with the top-ranked proteins in MR analysis, as these drugs could potentially alter *ACE2* expression and may be clinically relevant. Notably, MR is much less vulnerable to confounding and reverse causality compared to observational studies.

The most consistent finding was a tentative causal association between diabetes-related traits and increased *ACE2* expression. Based on one of the largest GWAS on type II diabetes (T2DM) to date (N=898,130), we found that T2DM is causally linked to raised *ACE2* expression (beta=0.1835, 95% CI 0.0853-0.2817; p=2.49E-4; GSMR method). Significant associations (at nominal level; $p<0.05$) was also observed across multiple datasets, with different analytic methods, and for both type I and II diabetes. Other diseases/traits having nominal significant associations with increased *ACE2* included inflammatory bowel disease, (ER+) breast and lung cancers, asthma, smoking and elevated ALT, among others. We also uncovered a number of plasma/serum proteins potentially linked to altered *ACE2* expression, and the top enriched pathways included cytokine-cytokine-receptor interaction, VEGF signaling, JAK-STAT signaling etc. We also explored drugs that target some of the top-ranked proteins in the MR analysis.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In conclusion, the current MR analysis reveals diseases/traits and blood proteins that may causally affect *ACE2* expression, which in turn may influence susceptibility to the infection. The proteome-wide MR analysis may shed light on the molecular mechanisms underlying *ACE2* expression, and may help guide drug repositioning in the future. Nevertheless, we stress that further studies are required to verify our findings due to various limitations and the exploratory nature of some analyses.

Introduction

A novel coronavirus named 2019-nCoV was detected from patients in Wuhan city, China [1,2] at the end of 2019. The virus has since caused an outbreak of coronavirus disease 2019 (COVID-19) not only presenting in China but also spreading worldwide [3-5]. A total of 87,137 of confirmed cases have been reported as at 01-03-2020, according to the latest WHO situation report (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2). Although most of patients were still concentrated in China, the total number of infected subjects in other countries has exceeded 7000, with confirmed cases in 58 countries. COVID-19 has already caused 2,977 deaths worldwide as at 01-03-2020. Considering the severity of COVID-19 outbreak all over the world, it is urgent to seek solutions to control the spread of the disease to susceptible groups, and to identify effective treatments. A better understanding of the pathophysiology of the disease is also urgently needed.

There have been ongoing efforts to characterize the clinical features of the illness. Through integrating three recent reports of confirmed cases, we estimated that over one quarter of patients had a previous history of comorbidity conditions, including hypertension (12.9%), diabetes (5.4%), cardiovascular disease (4.1%), respiratory system diseases (2.4%) and malignancy (0.5%) (Supplementary Table 1) [4-6]. Furthermore, various human organs such as lung, heart, kidney and bladder were reported to be vulnerable to the virus *via* analyzing the single-cell RNA sequencing datasets [7]. Nevertheless, it is still unclear whether the above or other comorbid diseases would lead to increased susceptibility to COVID-19, and if so, what the underlying mechanisms may be. In addition, since most clinical studies on the disease were observational in nature, it may be difficult to discern causality as many known or unknown confounders may be present (e.g. age, sex, other diseases, medications received, smoking/drinking history etc.). These confounders may lead to spurious associations between the exposure (e.g. comorbid conditions) and the outcome (e.g. susceptibility to infection, severity of illness etc.).

According to four lines of evidence, Wan *et al.* speculated that the host receptor of 2019-nCoV was angiotensin-converting enzyme 2 (ACE2) [8]. ACE2 has been established a receptor for SARS-CoV [9,10], and the same protein might regulate 2019-nCoV's capacity for cross-species and human-to-human transmissions [8]. Besides, Zhou *et al.* also confirmed this finding in HELA cell lines [11]. Subsequently, Wrapp *et al.* observed that the ACE2 protein could bind to the virus spike ecto-domain with around 15 nM affinity, which is up to 20-fold

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

higher than the binding affinity to the previous SARS-CoV spike glycoprotein [12]. This provides strong evidence that ACE2 is a target of the novel coronavirus.

A number of studies have looked into the relationship between *ACE2* expression level and coronavirus infection. For example, Li *et al.* overexpressed ACE2 protein in 293T and Vero E6 cell lines. Over-expression of ACE2 lead to more efficient viral replication, but it was blocked by anti-ACE2 antibody in a dose-dependent manner [9]. A further study also confirmed that susceptibility to SAR-CoV correlates with ACE2 expression on cell lines [13]. In a subsequent work, Jia et al. showed that undifferentiated airway epithelial cells that express little ACE2 were poorly infected by SARS-CoV, while well-differential cells expressing higher ACE2 were readily infected [14]. Taking the evidence together, expression level of the ACE2 protein is associated with susceptibility to SARS-CoV infection.

Since ACE2 has been supported by multiple studies as a receptor for 2019-nCov, it is reasonable to speculate that higher ACE2 expression in relevant tissues (e.g. the lung) may lead to increased susceptibility to infection. As discussed above, studies have indeed supported this hypothesis for SARS-CoV. While further studies are required, revealing diseases/traits that are causally associated with altered ACE2 expression may shed light on why certain subjects are more susceptible to the infection and the underlying mechanisms (whether the increased susceptibility is mediated via ACE2).

In a related work, recently Cai [15] performed an analysis with TCGA datasets and showed that smoking is associated with elevated expression of *ACE2* in the lung, which may be associated with greater susceptibility to infection or more severe illness. However, there are several limitations as detailed by the author. For example, the samples studied are derived from patients with lung cancer, which may not be fully reflective of the expression in normal lung tissues. Another potential limitation is that it is difficult to control for all confounders, as smoking may be related to other unhealthy life habits [16] and multiple comorbid diseases [17,18].

Here we conducted a phenome-wide Mendelian randomization (MR) study to explore diseases or traits that may be *causally* linked to increased *ACE2* expression in the lung. We will employ MR for causal inference. MR makes use of genetic variants as “instruments” to represent the exposure of interest, and infers causal relationship between the exposure and the outcome[19]. MR is much less susceptible to confounding bias and reverse causality when compared to observational studies.

The concept of MR is akin to randomized controlled trials (RCT). For example, consider a study on the causal effect of lipid on the risk of a certain disease. Subjects who have inherited lipid-lowering alleles at a locus (or a set of such alleles at multiple loci) will have lower lipid levels on average, which is analogous to receiving lipid-lowering drugs in an RCT[20]. The random allocation of alleles at conception is similar to random assignment of treatment in an RCT. As a result, the chance of spurious associations due to known or unknown confounders is reduced due to the ‘random assignment’ of treatment or risk factor. Another important point to

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

note is that MR can be conducted with summary statistics from genome-wide association studies (GWAS), which are now widely available and often of very large sample sizes.

In this study, we wish to answer the following question: What conditions or traits may lead to increased *ACE2* expression, which may in turn result in greater susceptibility to 2019-nCov infection? Since COVID-19 is a new disease and prior knowledge is lacking, we employed a *phenome-wide* approach in which a large variety of traits are studied for causal associations with *ACE2* expression. This analysis may help to prioritize resources for better prevention of the infection in those susceptible subjects.

In addition to diseases, we also studied serum/plasma proteins as exposure, as they may point to potential molecular mechanisms underlying ACE2 expression, and may serve as potential predictive or prognostic biomarkers. It has also been suggested that such proteome-wide studies may help to reveal drug repositioning candidates[21], through the search for drugs that target the top-ranked proteins. For example, if a protein is found to casually increase the risk of a disease by MR, by the definition of causality, blocking the protein will lead to reduced disease risks. In our study, by finding plasma/serum proteins *causally* linked to *ACE2* expression, we may find drugs that will alter *ACE2* expression, which in turn may be useful for treatment.

Methods

GWAS data

Exposure data

To perform the phenome-study, we made use of the latest IEU GWAS database (https://gwas.mrcieu.ac.uk/), which contains up to 111,908,636,549 genetic associations from 31,773 GWAS summary datasets (as at 26th Feb 2020). The database was retrieved via the R package "TwoSampleMR" (ver 0.5.1). MR analysis was conducted with the same package. Due to the extremely huge number of traits in the database, we performed some pre-selection to the list of traits/diseases before full analysis. Briefly, we selected the following categories of traits: (1) Traits listed as priority 1 (high priority) and labelled as "Disease" or "Risk factor" (81 and 71 items respectively); (2) traits labelled as "protein" as described above (3371 items); (3) (selected) traits from the UK Biobank, as it is one of the largest source of GWAS data worldwide (with sample size ~ 500,000). We consider that a proportion of traits have presumably low prior probability of association with respiratory infections, and others are less directly clinically relevant. To reduce computational burden and for ease of interpretation, a proportion of UK Biobank (UKBB) traits were filtered. More specifically, we excluded GWAS data of diseases or traits related to the following: eye or hearing problems, orthopedic and trauma-related conditions (except autoimmune diseases), skin problems (except systemic or autoimmune diseases), perinatal and obstetric problems, operation history, medication history (as confounding by indication is common and may affect the validity of results[22]), diet/exercise habit (as accuracy of information

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

cannot be fully guaranteed and recall bias may be present), other socioeconomic features (such as type of jobs). A total of 425 UKBB traits were retained for final analysis.

GWAS of UKBB were based on analysis results from the Neale Lab (https://sites.google.com/broadinstitute.org/ukbbgwasresults/) and from MRC-IEU. GWAS analysis was performed using linear models with adjustment for population stratification; details of the analytic approach is given in the following links: https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas, http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas and https://doi.org/10.5523/bris.pnoat8cxo0u52p6ynfaekeigi. For binary outcomes, we converted the regression coefficients obtained from the linear model to those under a logistic model, based on methodology presented in [23]. The SE under a logistic model was derived by the delta method (see supplementary text of [23], equation 37).

Outcome data

Regarding the outcome, we are interested in the expression of *ACE2*. While ideally one should study the protein expression in the lung, such data is scarce and corresponding genotype data (required for MR) is not available. Here we focus on the gene expression of *ACE2* in the lung (N = 515). ACE2 protein levels appear to be relatively well-correlated with mRNA levels across tissues, based on the Human Proteome Atlas (https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue). We retrieved GWAS summary data from the GTEx database, one of the largest databases to date with both genotype and expression data for a large variety of tissues. For details of GTEx please refer to [24].

Mendelian randomization (MR) analysis

Here we performed two-sample MR in which the instrument-exposure and instrument-outcome associations were estimated in different samples. We conducted MR primarily with the 'inverse-variance weighted' (MR-IVW)[25] and Egger regression (MR-Egger)[26] approaches, which are among the most widely used MR methods. 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 imbalanced pleiotropy is present. MR was performed on (approximately) independent SNPs with r^2 threshold of 0.001, following default settings in TwoSampleMR. We only included SNPs passing genome-wide significance ($p<5e-8$) as instruments. For exposure with only one instrument, the Wald ratio method was used. For analysis with less than 3 genetic instruments, we employed MR-IVW since MR-Egger cannot be reliably performed. As will be detailed in the results section, for selected trait(s) 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 [27].

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 [27]. MR-RAPS is another MR analysis methodology which can take into account multiple weak instruments by a robust procedure. Details of MR-RAPS were described in Zhao et al. [28].

We also performed analysis with plasma/serum proteins as exposure. Besides MR analysis on individual proteins, we also performed pathway analysis by ClueGO [29]. Hypergeometric tests were conducted on the top-ranked proteins (with $p<0.05$). In addition, we searched for drugs with targets overlapping with the top-ranked proteins. Drug targets were defined based on the DrugBank database. Our aim is uncover drug candidates leading to alteration of *ACE2* expression, which may be therapeutically relevant.

Results

MR analysis for diseases and clinically relevant traits

MR results are presented in Tables 1 and 2 (full results shown in Tables S2 and S3). Traits were shown if any of the three methods (MR-IVW, MR-Egger, Wald ratio) showed nominally significant ($p<0.05$) results. For traits that do not show evidence for directional pleiotropy (p-value of Egger Intercept>0.05), we shall primarily report the results from MR-IVW, as generally the SE of causal estimates is larger with MR-Egger [30] (hence power is weaker). Results from MR-Egger will be presented if there is significant directional pleiotropy.

Diabetes-related traits

Remarkably, a number of top-ranked results were related to diabetes. We observed totally five diabetes-related traits that showed nominally significant MR results, and they were all positively associated with *ACE2* expression. Three are related to diagnosis of diabetes (including both type I and II) in the UKBB. Another one (id: ieu-a-23) was based on a trans-ethnic meta-analysis in 2014 [31], which had no overlap with the UKBB sample. The finding of a nominally significant result in this dataset can therefore be considered as an independent replication of the UKBB result.

We also observed that starting insulin within one year of diagnosis, which was only assessed within diabetic subjects, was casually associated with increased *ACE2* expression. Early use of insulin may indicate type I diabetes as the underlying diagnosis or more severe or late-stage disease for type II diabetic subjects [32].

In view of the consistent causal associations with diabetes or related traits, we further searched for GWAS summary statistics that have not been included in the IEU GWAS database. We found another publicly available dataset from the DIAGRAM Consortium, based on a recent meta-analysis of type II diabetes by Mahajan et al. [33]. For a more in-depth analysis, we also employed GSMR and MR-RAPS in addition to IVW and Egger. The full results are presented in Table 2. Reassuringly, with the exception of MR-Egger (which is relatively less

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

powerful [30]), all other methods showed (nominally) significant results. GSMR reported the lowest p-value of 2.49E-04 (beta = 0.1835, 95% CI: 0.0853 to 0.2817). While this study [33] has partial overlap with the trans-ethnic analysis in 2014 [31], the consistent associations provide further support to a causal link between diabetes and expression of *ACE2*.

Regarding the effect size of casual associations, since the exposures were binary, the regression coefficients (beta) from MR may be roughly interpreted as average change in the outcome (increase in normalized expression level) per 2.72-fold increase in the prevalence of the exposure [34]. For type II diabetes, or self-reported diabetes from UKBB which presumably comprised mainly type II diabetes, the causal estimates ranged from ~0.1621 to 0.1835. These estimates were reasonably close despite different datasets being used. The causal estimate from type I diabetes was slightly lower and estimated to be ~0.1006.

Other disease/traits

As shown in Table 1, a number of other disease/traits also showed (nominally) significant results. Several neoplasms, such as breast and lung cancer, may be associated with increased *ACE2* expression. We also observed that several autoimmune disorders, especially inflammatory bowel diseases may be casually associated with *ACE2* expression. Interestingly, asthma and tobacco use also showed nominal significant associations with higher *ACE2* expression. As for other traits, high alanine aminotransferase (ALT), commonly associated with liver diseases, may be related to elevated *ACE2* expression. Other commonly measured blood measures that may lead to altered *ACE2* expression also included red cell distribution width (often associated with iron-deficiency, folate or B12 deficiency anemia), basophil percentage (inverse relationship), calcium level, urate level, HDL-cholesterol and LDL cholesterol (inverse relationship).

MR results with plasma/serum proteins as exposure

Full results are shown in Table S2 and the enriched pathways are shown in Table 3 and Table S4. Since a large number of proteins are involved, we only highlight a few top pathways here. Some of top pathways include cytokine and cytokine receptor interaction, VEGFA-VEGF2 signaling pathway, JAS-STAT signaling pathway etc. Table 4 and S5 shows the list of drugs whose targets overlap with the top-ranked proteins. Note that the tables do *not* explicitly discern the direction of effects of the drugs. A few drugs target more than one protein. If they are ranked by the number of proteins targeted, the top drugs are fostamatinib, copper, zinc and zonisamide, which target >=3 proteins.

Discussions

Diseases/traits causally linked to ACE2 expression

In this study we have employed Mendelian Randomization (MR) to uncover diseases/traits that may be causally linked to *ACE2* expression levels in the lung, which in turn may influence susceptibility to the infection. We believe such analysis is of value as observational studies are more prone to confounding bias. Also, in

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

practice, it is very difficult to organize a comprehensive clinical study of many different risk factors/diseases, evaluate if they are risk factors for COVID-19. Such studies may also be limited by the lack of relevant clinical data for some patients.

From our analysis, the most consistent finding was the casual link between diabetes (and related traits) with *ACE2* expression, which was supported by multiple datasets and different analytic approaches. Other results were more tentative, but may be worthy of further studies. For example, several neoplasms (e.g. breast and lung cancers) and autoimmune diseases, elevated ALT, asthma and smoking all showed nominally significant and positive associations with *ACE2* expression. If the findings are replicated and confirmed in further studies, there may be clinical implications.

For example, identification of those at greater risk may help to guide the prioritization of resources to reduce infection risks in susceptible groups. Also, it is likely that vaccines may be developed in the near future; in the lack of resources, susceptible groups may be prioritized to receive vaccination to maximize cost-effectiveness. In a similar vein, if resources are limited, the more susceptible subjects may receive higher priority for diagnostic testing for the infection. As far as treatment is concerned, if certain conditions such as diabetes indeed increases susceptibility via ACE2, then drugs targeting at this gene/protein may be particularly useful for this patient subgroup. For example, human recombinant ACE2 has been proposed as a treatment and is under clinical trial for COVID-19[35]. It will be interesting to see if the drug may be more beneficial for DM patients. More generally speaking, if DM is causally linked to elevated *ACE2* and potentially increased susceptibility to infection, then anti-diabetic drugs or improved glycemic control may reverse the process. Interestingly, a few studies have shown that metformin may reduce mortality from lower respiratory disease in diabetic patients{Ho, 2019 #59;Mendy, 2019 #60}. It will be intriguing to know if metformin (or other anti-diabetic medications) may be clinically beneficial in preventing or reducing the severity of disease in diabetic or non-diabetic subjects.

Here we further discuss on a few disease/traits also supported by previous studies. First, as discussed in the introduction, a recent study [15] also suggested smoking was associated with higher *ACE2* expression. Our analysis adds further support to the hypothesis, but we further showed that the relationship may be casual. As shown in Table S1, a number of COVID-19 cases (~5.4%) were also comorbid with diabetes mellitus (DM). Similarly, DM was also common in patients infected with MERS-CoV [36,37]. Kulcsar *et al.* built a mouse model susceptible to MERS-CoV infection and induced type 2 DM using a high-fat diet. They found that, if affected by the virus, these diabetic mice suffered from a prolonged phase of disease and delayed recovery, which might be due to a dysregulated immune response [38]. With regards to comorbidity with cancers, Liang *et al.* recently carried out a nationwide analysis of 1,590 patients with laboratory-confirmed COVID-19 and suggested that cancer patients might have a higher infection risk than those without [39]. The study also reported higher risk of severe complications in such patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

There are several limitations in our analysis of associated diseases/traits with *ACE2* expression. A major limitation is that the sample size for GTEx is relatively modest, which limits the power of MR analysis. However, to our knowledge, GTEx is probably the largest database with both genotype and expression data for lung tissues.

Another point we wish to emphasize is that we consider this work as largely an **exploratory** rather than confirmatory study. Our main purpose is to *prioritize* diseases, traits or proteins with potential causal links with *ACE2* expression, and hence possibly increased susceptibility to 2019-nCov infection. Owing to relatively modest sample size of the outcome dataset of GTEx (N = 515), we expect the power to be modest. Also in view of the exploratory and hypothesis-generating nature of this study, we have not implemented stringent multiple testing procedures such as Bonferroni correction. On the other hand, we examined the consistency of the observed associations across different datasets, and considered those supported by more than one set of data as relatively more trustworthy or robust, similar to the approach adopted by Pendergrass et al. [40]. However, we emphasize that our findings will require further replications and support by further clinical and experimental studies.

On the other hand, we also wish to highlight some results could be false negatives. The main reason for false negatives is the limited sample size of GTEx, and that for some exposure traits, the number of instruments available may be small. For example, in our analysis, we do not find evidence of hypertension or blood pressure, history of coronary heart disease and stroke to be casually linked to *ACE2* expression, although patients with severe infections have been reported to be enriched for these comorbidities.

On ACE2 expression and exploring drug candidates

As discussed above, increased expression of *ACE2* appears to be correlated with susceptibility to SARS-Cov and 201-nCoV infection. Nevertheless, the consequences of altered *ACE2* expression may be rather complex. Kuba et al. reported that the Spike protein of the SARS-CoV down-modulated *ACE2* expression [10], which may lead to heightened risks of acute lung injury. Another study [41] suggested ACE2 may protect against acute pulmonary failure by blocking the renin-angiotensin signaling pathway. However, whether the same may apply to 2019-nCov is still unknown. If this is the case, then one may hypothesize that for *unaffected* individuals or those without (or with minimal) lung involvement yet (as could be the case for some subjects at the early stage of disease), a *lower* level of *ACE2* expression on lung cells may be beneficial in reducing susceptibility to more sustained infection by reducing viral entry. However, for patients with severe lung involvement or at risk of acute lung injury, *higher* level of *ACE2* expression may prevent risk of acute lung failure. Therefore, it may be clinically relevant to identify both types of drugs, i.e. those leading to elevated *ACE2* expression as well as those leading to reduced expression. Further studies are warranted to clarify the role of ACE2 and whether drugs targeting ACE2 may be therapeutically useful.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The drugs we highlighted in this study may help researchers to *prioritize* repositioning candidates for further studies, given the huge cost in developing a brand-new drug and that detailed investigations on every existing mediation will be impractical. We briefly discuss a few drugs highlighted by our analysis. Fostamatinib targets the largest number (seven) of proteins potentially linked to *ACE2* expression. According to DrugBank, it serves as an inhibitor for all these proteins, and all were linked to elevated *ACE2* expression in the present MR analysis except one. This drug has been approved for treating Immune Thrombocytopenic Purpura (ITP), and is a spleen tyrosine kinase inhibitor[42]. There has been trials on rheumatoid arthritis (RA)[43] and IgA nephropathy as well[42]. Interestingly, recent studies by the company BenevolentAI[44,45] employed a proprietary knowledge graph approach and found several repositioning candidates [46,47]. Baricitinib, a JAK 1/2 inhibitor approved for RA, was suggested as a top candidate. The drug was proposed on its action on AAK1 which is a regulator of endocytosis, although how AAK1 was prioritized as target was not described in the study. Fostamatinib, which we prioritized in this study, also inhibits JAK1, JAK2 and AAK1 [48] based on curations from DrugBank and was shown to be effective for RA [43]. Of note, two other JAK-STAT signaling inhibitors were recommended by Stebbing et al.[45], while in our analysis JAK-STAT signaling is among the top 10 pathways enriched for top proteins linked to *ACE2* expression. Another candidate highlighted by Richardson et al.[44], sunitinib, was also top-listed by our MR-based analysis. We have employed a rather different algorithm based on causal inference, when compared to the approach by BenevolentAI. The concordance between different studies provides additional support to the usefulness our MR-based approach, and the drugs with converging evidence by different approaches may be more likely to be true candidates. Zinc was also a top-listed candidate in our study, and it has been reported to reduce the risk of lower respiratory tract infections in some studies, e.g. [49], although further studies are required as the evidence is not firm.

Despite some interesting findings, due to limited knowledge of how the drugs act on the targets and their directions of effect, as well as the pathophysiology of COVID-19, we consider our results as exploratory findings which require further investigations. We note that a number of drugs may act on more than one target, but the exact pharmacological action on each target is often unclear; the overall direction and magnitude of effect of each drug may not be easily determined and must be verified in further studies. We emphasize that the drugs highlighted in this work are meant to *prioritize* suitable candidates to speed up discovery for treatments, and are *not* supposed to be applied to clinical practice or trials yet. However, due to the potential huge cost and extreme urgency of developing new therapies, we believe that any drug repositioning/discovery attempt that may improve the success rate even by a small margin may still be much valuable.

Moreover, we should stress that this study does *not* address what factors may aggravate or ameliorate *CoV-induced changes in ACE2 levels* (i.e. the expression changes *as a result* of CoV infection). This involves complex interaction between the virus, the ACE2 receptor and other downstream pathways, and could not be predicted by the present analysis per se. Our findings from MR mainly reflect diseases/traits/proteins causally linked to *ACE2* expression in *uninfected* subjects, as the outcome dataset (GTEx) is composed of such subjects.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Finally, on a methodological note, we have employed MR in a different manner in most present studies. Usually MR is used to identify causal risk factors with a disease as the outcome, for which GWAS data for the disease is available. Here we presented a new analytic approach; we made use of existing knowledge of a key receptor of an infectious agent to uncover risk factors as well as repositioning candidates. This analytic framework may also be applied to other diseases, especially when a target can be identified but genomic data for the disease is limited.

Conclusions

Notwithstanding the limitations, we have identified several diseases and traits which may be causally related to *ACE2* expression the lung, which in turn may mediate susceptibility to 2019-nCoV infection. In addition, our proteome-wide MR analysis revealed proteins that could lead to changes in *ACE2* expression. Subsequent drug repositioning analysis highlighted several candidates that may warrant further investigations. We stress that most of the findings require replications and validation in further studies, especially the part on drug repositioning. Nevertheless, we believe this work is of value in view of the urgency to address the outbreak of 2019-nCoV.

Acknowledgements

We would like to thank Prof. Stephen Tsui for computing support. This study was partially supported by the Lo Kwee Seong Biomedical Research Fund, an NSFC grant and a Chinese University of Hong Kong Direct Grant. We also thank Mr Carlos Chau for assistance in part of the analysis.

Author contributions

Conceived and designed the study: HCS, SR. Supervised the study: HCS. Data analysis: HCS (lead), SR, AL. Data interpretation: SR, AL, HCS. Drafted the manuscript: HCS (lead), with input from AL and SR.

Conflicts of interest

The authors declare no conflict of interest.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References:

1. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565-574 (2020).
2. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **382**, 727-733 (2020).
3. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* (2020).
4. Guan, W.-j. *et al.* Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv*, 2020.02.06.20020974 (2020).
5. Novel-Coronavirus-Pneumonia-Emergency-Response-Epidemiology-Team. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**, 145-151 (2020).
6. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497-506 (2020).
7. Zou, X. *et al.* The single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to Wuhan 2019-nCoV infection. *Front. Med.*, [Epub ahead of print] (2020).
8. Wan, Y., Shang, J., Graham, R., Baric, R.S. & Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol* (2020).
9. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-4 (2003).
10. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* **11**, 875-9 (2005).
11. Zhou, P. *et al.* Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv*, 2020.01.22.914952 (2020).
12. Wrapp, D. *et al.* Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *bioRxiv*, 2020.02.11.944462 (2020).
13. Hofmann, H. *et al.* Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochemical and Biophysical Research Communications* **319**, 1216-1221 (2004).
14. Jia, H.P. *et al.* ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol* **79**, 14614-21 (2005).
15. Cai, G. Bulk and Single-Cell Transcriptomics Identify Tobacco-Use Disparity in Lung Gene Expression of ACE2, the Receptor of 2019-nCov. *Preprints* **2020020051**(2020).
16. Ng, R., Sutradhar, R., Yao, Z., Wodchis, W.P. & Rosella, L.C. Smoking, drinking, diet and physical activity—modifiable lifestyle risk factors and their associations with age to first chronic disease. *International journal of epidemiology* (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

17. Bielicki, P., Trojnar, A., Sobieraj, P. & Wąsik, M. Smoking status in relation to obstructive sleep apnea severity (OSA) and cardiovascular comorbidity in patients with newly diagnosed OSA. *Advances in respiratory medicine* **87**, 103-109 (2019).
18. Azibou, A. *et al.* Smoking, comorbidity, anxiety and depression in patients with MS and their relation with motor and/or cognitive fatigue. *MULTIPLE SCLEROSIS JOURNAL* **25**, 428-429 (2019).
19. Smith, G.D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1-22 (2003).
20. Bennett, D.A. & Holmes, M.V. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart* **103**, 1400-1407 (2017).
21. Schmidt, A.F. *et al.* Genetic drug target validation using Mendelian randomization. *bioRxiv*, 781039 (2019).
22. Bellera, C. *et al.* A two-stage model in a Bayesian framework to estimate a survival endpoint in the presence of confounding by indication. *Stat Methods Med Res* **27**, 1271-1281 (2018).
23. Lloyd-Jones, L.R., Robinson, M.R., Yang, J. & Visscher, P.M. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics* **208**, 1397-1408 (2018).
24. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics* **50**, 956-967 (2018).
25. Burgess, S., Butterworth, A. & Thompson, S.G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-65 (2013).
26. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).
27. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* **9**, 224 (2018).
28. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D.S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *arXiv preprint arXiv:1801.09652* (2018).
29. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)* **25**, 1091-1093 (2009).
30. Burgess, S. & Thompson, S.G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* **32**, 377-389 (2017).
31. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
32. Thomas, N.J. *et al.* Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *The lancet. Diabetes & endocrinology* **6**, 122-129 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

33. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nature Genetics* **50**, 559-571 (2018).
34. Burgess, S. & Labrecque, J.A. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *European Journal of Epidemiology* **33**, 947-952 (2018).
35. APEIRON's respiratory drug product to start pilot clinical trial to treat coronavirus disease COVID-19 in China. Available at https://pipelinereview.com/index.php/2020022673884/Proteins-and-Peptides/APEIRONs-respiratory-drug-product-to-start-pilot-clinical-trial-to-treat-coronavirus-disease-COVID-19-in-China.html.
36. Banik, G.R., Alqahtani, A.S., Booy, R. & Rashid, H. Risk factors for severity and mortality in patients with MERS-CoV: Analysis of publicly available data from Saudi Arabia. *Virol Sin* **31**, 81-4 (2016).
37. Alqahtani, F.Y. *et al.* Prevalence of comorbidities in cases of Middle East respiratory syndrome coronavirus: a retrospective study. *Epidemiol Infect*, 1-5 (2018).
38. Kulcsar, K.A., Coleman, C.M., Beck, S.E. & Frieman, M.B. Comorbid diabetes results in immune dysregulation and enhanced disease severity following MERS-CoV infection. *JCI insight* **4**, e131774 (2019).
39. Liang, W. *et al.* Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *The Lancet Oncology* (2020).
40. Pendergrass, S.A. *et al.* The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genetic epidemiology* **35**, 410-422 (2011).
41. Imai, Y. *et al.* Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112-6 (2005).
42. Markham, A. Fostamatinib: First Global Approval. *Drugs* **78**, 959-963 (2018).
43. Kang, Y. *et al.* Efficacy and Safety of Multiple Dosages of Fostamatinib in Adult Patients With Rheumatoid Arthritis: A Systematic Review and Meta-Analysis. *Front Pharmacol* **10**, 897 (2019).
44. Richardson, P. *et al.* Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* **395**, E30-E31 (2020).
45. Stebbing, J. *et al.* COVID-19: combining antiviral and anti-inflammatory treatments. *The Lancet Infectious Diseases* (2020).
46. Richardson, P. *et al.* Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *The Lancet* **395**, e30-e31 (2020).
47. Stebbing, J. *et al.* COVID-19: combining antiviral and anti-inflammatory treatments. *The Lancet Infectious Diseases* **0**(2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.04.20031237; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

48. Rolf, M.G. *et al.* In vitro pharmacological profiling of R406 identifies molecular targets underlying the clinical effects of fostamatinib. *Pharmacol Res Perspect* **3**, e00175 (2015).

49. Roth, D.E., Richard, S.A. & Black, R.E. Zinc supplementation for the prevention of acute lower respiratory infection in children in developing countries: meta-analysis and meta-regression of randomized trials. *International Journal of Epidemiology* **39**, 795-808 (2010).

Table 1 Overall MR analysis results achieving nominal significance (p<0.05), with diseases/traits as exposure and ACE2 lung expression as outcome

id	trait	nsnps	b_ IVW	se_ IVW	pval_ IVW	b_ Egger	se_ Egger	pval_ Egger	b_ Wald	se_ Wald	pval_ Wald	Egger_ intercept	pval_ intercept
Diseases as exposure													
Diabetes-related													
ukb-b-10753	Diabetes diagnosed by doctor	58	0.1754	0.0723	0.0152	0.0024	0.1584	0.9879	-	-	-	0.0196	0.2251
ukb-b-12948	Non-cancer illness code, self-reported: diabetes	49	0.1621	0.0766	0.0343	0.2030	0.1760	0.2548	-	-	-	-0.0047	0.7974
ukb-b-10694	Diagnoses - secondary ICD10: E10.9 Type I Diabetes mellitus without complications	3	0.1006	0.0495	0.0423	0.1842	0.1468	0.4284	-	-	-	-0.0523	0.6536
ieu-a-23	Type 2 diabetes	25	0.2103	0.1034	0.0421	0.1600	0.4038	0.6957	-	-	-	0.0053	0.8986
ukb-b-8388	Started insulin within one year diagnosis of diabetes	7	0.0763	0.0354	0.0310	0.0770	0.0751	0.3523	-	-	-	-0.0005	0.9914
Neoplasms													
ukb-b-8549	Diagnoses - main ICD10: D24 Benign neoplasm of breast	1	-	-	-	-	-	-	0.3768	0.1650	0.0223	-	-
ukb-d-D12	Diagnoses - main ICD10: D12 Benign neoplasm of colon, rectum, anus and anal canal	12	-0.2861	0.1177	0.0150	-0.0575	0.4634	0.9038	-	-	-	-0.0312	0.6210
ukb-d-C3_RESPIRATORY_INTRATHORACIC	Malignant neoplasm of respiratory system and intrathoracic organs	3	0.5137	0.1976	0.0093	0.7114	0.7363	0.5109	-	-	-	-0.0562	0.8213
ukb-d-C34	Diagnoses - main ICD10: C34 Malignant neoplasm of bronchus and lung	1	-	-	-	-	-	-	0.3627	0.1661	0.0290	-	-
ieu-a-1134	ER+ Breast cancer (GWAS)	7	0.1759	0.0755	0.0198	0.5389	0.2953	0.1276	-	-	-	-0.0944	0.2594
ieu-a-1013	Glioma	3	0.2000	0.0958	0.0367	0.4379	0.7370	0.6587	-	-	-	-0.0705	0.7997
Autoimmue disorders													
ukb-b-18194	Non-cancer illness code, self-reported: ankylosing spondylitis	3	0.0865	0.0358	0.0157	0.0171	1.2239	0.9911	-	-	-	0.0586	0.9639

ieu-a-32	Ulcerative colitis	29	0.0256	0.0596	0.6671	**0.5855**	0.1775	0.0027	-	-	-	-0.1049	**0.0027**
ieu-a-292	Inflammatory bowel disease	107	0.0034	0.0429	0.9376	**0.2458**	0.1034	0.0192	-	-	-	-0.0309	**0.0113**
ieu-a-30	Crohn's disease	41	0.0106	0.0390	0.7864	**0.1961**	0.0896	0.0346	-	-	-	-0.0443	**0.0268**
ieu-a-31	Inflammatory bowel disease	49	-0.0138	0.0499	0.7827	**0.2866**	0.1327	0.0360	-	-	-	-0.0510	**0.0191**
Other diseases													
ukb-b-17219	Diagnoses - secondary ICD10: J45.9 Asthma, unspecified	19	0.2638	0.1253	0.0353	0.8765	0.4270	0.0558	-	-	-	-0.0635	0.1517
ukb-b-5115	Diagnoses - secondary ICD10: Z72.0 Tobacco use	3	0.9175	0.3818	0.0163	0.6241	2.7179	0.8563	-	-	-	0.0236	0.9303
ukb-b-6514	Non-cancer illness code, self-reported: hiatus hernia	2	0.8635	0.3822	0.0239	-	-	-	-	-	-	-	-
ukb-b-8848	Diagnoses - main ICD10: K44.9 Diaphragmatic hernia without obstruction or gangrene	1	-	-	-	-	-	-	1.0439	0.4627	0.0241	-	-
ukb-b-1386	Diagnoses - main ICD10: K21.9 Gastro-oesophageal reflux disease without oesophagitis	1	-	-	-	-	-	-	-0.8653	0.4214	0.0400	-	-
ukb-d-I9_DISVEINLYMPH	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified	14	-0.2693	0.1194	0.0241	-0.1381	0.3021	0.6558	-	-	-	-0.0165	0.6448
ukb-d-M48	Diagnoses - main ICD10: M48 Other spondylopathies	1	-	-	-	-	-	-	0.7776	0.3238	0.0163	-	-
ukb-d-N13	Diagnoses - main ICD10: N13 Obstructive and reflux uropathy	1	-	-	-	-	-	-	0.6746	0.3417	0.0483	-	-
Other risk factors or clinically relevant traits as exposure													
ukb-b-15272	Ever manic/hyper for 2 days	2	1.1349	0.3862	0.0033	-	-	-	-	-	-	-	-
ukb-d-30620_raw	Alanine aminotransferase	91	0.0474	0.0177	0.0074	0.0806	0.0389	0.0413	-	-	-	-0.0124	0.3411
ukb-d-30070_irnt	Red blood cell (erythrocyte) distribution width	225	0.2428	0.1047	0.0204	0.3487	0.1885	0.0657	-	-	-	-0.0044	0.5001
ukb-d-30220_irnt	Basophill percentage	77	-0.5036	0.2281	0.0273	-0.5837	0.4200	0.1687	-	-	-	0.0026	0.8211
ukb-d-30780_raw	LDL direct	126	-0.2718	0.1676	0.1050	-0.7577	0.2667	0.0053	-	-	-	0.0195	0.0207
ukb-d-30680_raw	Calcium	152	0.2020	1.8845	0.9146	-10.38	4.0074	0.0105	-	-	-	0.0308	0.0034
ukb-d-30830_raw	SHBG	163	0.0044	0.0057	0.4428	0.0216	0.0105	0.0424	-	-	-	-0.0164	0.0552
ieu-a-793	Urate	4	0.0256	0.0116	0.0271	0.0263	0.0124	0.1671	-	-	-	-0.0079	0.8839

ieu-a-1034	Height	4	0.6362	0.3202	0.0469	2.0601	7.3837	0.8064	-	-	-	-0.1177	0.8648
ieu-a-299	HDL cholesterol	84	0.0842	0.1291	0.5145	0.5630	0.2402	0.0215	-	-	-	-0.0258	0.0204
ieu-a-1089	Body mass index	1	-	-	-	-	-	-	3.6262	1.2369	0.0034	-	-

Nsnps, number of SNPS. b, beta (causal estimate); se, standard error; pval, p-value. IVW, inverse-variance weighted approach.

Table 2 Further MR analysis results for type II diabetes based on 2018 Mahajan et al.

	beta	se	pval	Egger Intercept	Intercept pval	n_pleio	nsnps
MR-IVW	0.1772	0.0595	2.91E-03	-	-	-	196
MR-Egger	-0.0390	0.1264	0.7578	0.0159	0.0545	-	196
GSMR	0.1835	0.0501	2.49E-04	-	-	-	391
MR-RAPS	0.0571	0.0268	3.28E-02	-	-	-	4748

The exposure GWAS dataset on type II diabetes was based on Mahajan et al. (2018).Nature genetics, 50(11), 1505-1513. This study was not covered in IEU GWAS database.

Table 3 Top 10 enriched pathways for significant proteins in MR analysis

GOID	GOTerm	Ontology Source	Term PVal	Term Pval (Bonf corrected)	Associated Genes
KEGG:04060	Cytokine-cytokine receptor interaction	KEGG_27.02.2019	1.82E-06	7.29E-05	*[CCL25, CTF1, CX3CL1, CXCL12, IL15RA, IL22, IL34, IL37, LTA, LTBR, OSM, TNFSF4, TNFSF8]*
WP:3888	VEGFA-VEGFR2 Signaling Pathway	WikiPathways_27.02.2019	7.79E-06	3.11E-04	*[ACACB, BCL2L1, CFL1, EEA1, F3, IGFBP7, JAG1, KDR, PIK3CA, PTPN1, TXNIP]*
WP:254	Apoptosis	WikiPathways_27.02.2019	1.07E-04	4.29E-03	*[BCL2L1, BIRC5, DIABLO, IGF1, LTA, MCL1]*
WP:3614	Photodynamic therapy-induced HIF-1 survival signaling	WikiPathways_27.02.2019	2.99E-04	1.19E-02	*[BCL2L1, BIRC5, HK1, MCL1]*
R-HSA:399954	Sema3A PAK dependent Axon repulsion	REACTOME_Pathways_27.02.2019	3.40E-04	1.36E-02	*[CFL1, PAK3, PLXNA1]*

R-HSA:2173782	Binding and Uptake of Ligands by Scavenger Receptors	REACTOME_Pathways_27.02.2019	4.89E-04	1.96E-02	*[FTH1, HP, STAB1, STAB2]*
KEGG:04630	JAK-STAT signaling pathway	KEGG_27.02.2019	5.61E-04	2.24E-02	*[BCL2L1, CTF1, IL15RA, IL22, MCL1, OSM, PIK3CA]*
KEGG:04672	Intestinal immune network for IgA production	KEGG_27.02.2019	8.84E-04	3.53E-02	*[CCL25, CXCL12, IL15RA, LTBR]*
WP:3657	Hematopoietic Stem Cell Gene Regulation by GABP alpha/beta Complex	WikiPathways_27.02.2019	9.00E-04	3.60E-02	*[BCL2L1, FLT3, MCL1]*
WP:3872	Regulation of Apoptosis by Parathyroid Hormone-related Protein	WikiPathways_27.02.2019	0.00090	0.03602	*[BCL2L1, MCL1, PIK3CG]*

Pval, p-vlaue; Bonf, Bonferroni correction.

Table 4 Drugs with targets overlapping with significant proteins from MR analysis (only FDA-approved drugs shown)

Drug	No. of proteins targeted	Targets (that overlap with proteins significant in MR analysis)						
Fostamatinib	7	ZAP70	FLT3	HIPK3	KDR	MST1R	PAK3	PIK3CG
Copper	6	CFL1	S100A2	PARK7	AHSG	APOD	CBX5	
Zinc	4	S100A2	AHSG	C8A	APLP2			
Zonisamide	3	CA4	CA9	CA10				
Benzthiazide	2	CA4	CA9					
Hyaluronic acid	2	LAYN	STAB2					
Hydroflumethiazide	2	CA4	CA9					
Isosorbide	2	BCL2L1	MCL1					
Midostaurin	2	KDR	FLT3					
Nintedanib	2	KDR	FLT3					
Ponatinib	2	FLT3	KDR					

Sodium carbonate	2	CA4	CA9
Sorafenib	2	KDR	FLT3
Sunitinib	2	KDR	FLT3