PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Drug target Mendelian randomization applied to metabolic dysfunction-associated steatotic liver disease: Opportunities and Challenges
AUTHORS	Shan, Luo; Zheng, Ming-Hua; Wong, Vincent Wai-Sun; Au Yeung, Shiu Lun

VERSION 1 - REVIEW

REVIEWER NAME	Yuan, Shuai
REVIEWER AFFILIATION	Karolinska Institutet, Institute of Environmental Medicine
REVIEWER CONFLICT OF INTEREST	None
DATE REVIEW RETURNED	29-Jul-2024

GENERAL COMMENTS	Luo et al wrote a review paper to describe drug target discovery for metabolic dysfunction-associated steatotic liver disease using Mendelian randomization. In this study, they comprehensively summarized points of designing and appraising a drug-target Mendelian randomization study and presented examples to show the opportunities and challenges in this filed. The paper is well-written. The reviewer has a few points. 1. Page 4, line 20: “should not share unmeasured common causes with the outcome (independence)”. Do you mean only residual confounding? It should be both forms of confounding. 2. Page 4, line 26. Please include “genetically predicted or proxied” before “a measure of pharmacological perturbation of the relevance drug target”. 3. Page 5, line 3. Please consider revising “Although less common now”. Maybe it is not. Basically, all protein MR studies used this strategy, like using SNPs in coding region. 4. Given that these MR approaches are heavily based on accuracy of MASLD GWAS (mostly phenotyping of the disease), please discuss this point (defining the disease based on ICD, biopsy, imaging, liver enzymes, and so on) and possible related traits, like liver enzymes. Also, for different types of MASLD. 5. Figure 3. For factorial MR analysis, it is more efficient to use continuous data to test interactions? Do you think it will be more clinically relevant if we define the cutoff point according to some clinically used threshold, but not arbitrarily by median?

REVIEWER NAME	Zhou, Ang
REVIEWER AFFILIATION	University of Cambridge, MRC Biostatistics Unit
REVIEWER CONFLICT OF INTEREST	I have no conflicts of interest to declare

DATE REVIEW RETURNED	15-Aug-2024

GENERAL COMMENTS	The review by Luo, S. et al summarized key aspects of drug-target Mendelian randomization study design, and highlighted its potential for the discovery, repositioning, and safety assessment of drug targets in metabolic dysfunction-associated steatotic liver disease. I have the following comments: • Sensitivity analyses are mentioned on page 5 (lines 57-59) and 6 (lines 8-9). Please clarify what sensitivity analyses you are referring to? Are you referring to pleiotropy-robust methods, such as MR Egger, mode- and median-based method etc? If so, they may not be reliable in the cis-MR setting (PMID: 32760811) because these methods assume SNPs to be independent in terms of violating the IV assumptions. SNPs in the same gene region are likely to either all be valid IVs or all be invalid. • Page 7 lines 5-9 'To address this issue, researchers often select multiple candidate variants in partial linkage disequilibrium (LD) as instruments to increase statistical power and facilitate sensitivity analyses in MR.' Again, 'to facilitate sensitivity analyses' may not be a valid motivation for the development of variable selection techniques you mentioned; Improving statistical power certainly is. Another motivation worth mentioning is that using all variants (i.e., where no variable selection process is involved) within the same gene region can lead to numerical instability due to multicollinearity between SNPs. • Page 5, line 60 'When multiple candidate variants are presented within the protein-encoding gene,…..'. You may consider changing 'candidate' variants to 'causal' variants. • Page 6, line 52, '….when using protein abundance as the instruments,…..' should this be 'when using protein abundance as the exposures,….'? • 'Weights' is mentioned several times on page 7 without further explanation. Please elaborate on what is used as weights, and how they are applied in the cis-MR analyses. • Page 7 lines 21-23 "Colocalization of cis-eQTL and cis-pQTL exhibiting the same regulatory direction would further support a druggable protein as candidate target.37" It is not clear to me how this fits in the paragraph, which, if I understand correctly, focuses on using gene expression data as exposure. Please clarify. • Page 7 lines 50-55 For the binary outcomes, please clarify whether they are the outcomes of interest or phenotypes that lie in the pathway between exposure and outcome. I assume it is the latter. • Page 8 line 15 For the genome-wide and drug-target MR analyses, are you referring to the MR analyses on the phenotypes downstream of the druggable proteins? If so, you may need to clarify this to avoid confusion. • Page 8 line 41 '….highlighting triglyceride-lowering drug targets…' should it be 'triglyceride-lowering drug'? • Page 8 lines 50-54 Please clarify why there is a discrepancy between mechanism-specific effects of LPL activation and systemic triglyceride level reduction? I ask this because based on what you have described here the evidence seem to be consistent: o Triglyceride-NAFLD association: 'Accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD'. o LPL-NAFLD association: '…not all therapeutic targets aim at lowering plasma triglycerides reduce the risk of NAFLD, except the drug target LPL (lipoprotein lipase).'.

	• Page 10 lines 9-11 Genetic risk score is mentioned here for the first time. You may need a sentence to clarify what it is. • Page10 line 35 A full stop is missing between ‘scenarios’ and ‘More’. • Page 10 line 54 It is not clear to me how the example of IL-6 receptor inhibition supports your earlier statement: ‘To successfully reposition a drug, it is essential to identify genetic instruments influencing the molecular mechanisms of action of the drug.’ Please clarify. • Page 13 line 35 There is a typo. Please change ‘state-of-art’ to ‘state-of-the-art’.

VERSION 1 – AUTHOR RESPONSE

Associate Editor

Comments to the Author:

Thank you for your submission. The reviewers have some minor points - please

consider these carefully in revising the manuscript.

We thank the reviewers’ valuable comments concerning our manuscript (egastro 2024-100114). A cleaned manuscript and a track changed copy are provided. The

major corrections in the manuscript and the point-by-point responses to the

Reviewer’s comments are in italic as follows, where changes made in the manuscript

are underlined.

Reviewer: 1

Comments to the Author

Luo et al wrote a review paper to describe drug target discovery for metabolic

dysfunction-associated steatotic liver disease using Mendelian randomization. In this

study, they comprehensively summarized points of designing and appraising a drug target Mendelian randomization study and presented examples to show the

opportunities and challenges in this filed. The paper is well-written. The reviewer has

a few points.

1. Page 4, line 20: “should not share unmeasured common causes with the outcome

(independence)”. Do you mean only residual confounding? It should be both forms of

confounding.

Thank you for your comment. We have now revised the assumptions of Mendelian

randomization to align with the literature.1

From: “Should not share unmeasured common causes with the outcome

(independence)."

To: "The association of instrument in the outcome is not confounded (independence)."

2. Page 4, line 26. Please include "genetically predicted or proxied" before "a measure of pharmacological perturbation of the relevance drug target".

Thank you, we have now included as you suggested.

"In the context of drug-target MR, the genetically predicted or proxied exposure represents a measure of pharmacological perturbation of the relevance drug target."

3. Page 5, line 3. Please consider revising "Although less common now". Maybe it is not. Basically, all protein MR studies used this strategy, like using SNPs in coding region.

2

Thank you for your comment. We have now revised our statement to address your concerns, as follows:

From: "Although less common now, this approach was also used in earlier MR studies before the emergence of large genome-wide association studies (GWAS), such as obesity (FTO)2

and alcohol consumption (ALDH2 and ADH1B).3,4

"

To: "Prior to the emergence of large-scale genome-wide association studies (GWAS), earlier MR studies also employed this approach in selecting instruments, such as FTO variants for obesity,2

and ALDH2 and ADH1B variants for alcohol consumption.3,4"

4. Given that these MR approaches are heavily based on accuracy of MASLD GWAS (mostly phenotyping of the disease), please discuss this point (defining the disease based on ICD, biopsy, imaging, liver enzymes, and so on) and possible related traits, like liver enzymes. Also, for different types of MASLD.

Thank you for raising this important point. We acknowledge that the accuracy and validity of MR investigation are heavily based on the quality of data derived from

GWAS and the accurate phenotyping of the disease in question. We have now included a discussion on this topic on pages 14-15:

Variation in phenotyping of MASLD

The validity of an MR study is directly related to the quality of GWAS data, in particular the classification of the disease and trait measurement. Liver biopsy is the reference standard for diagnosing and staging hepatic diseases. However, its invasive nature limits its usage, especially in paediatric population. Amongst non invasive methods, magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF) stands out for its superior accuracy in detecting and quantifying liver steatosis.5 Nevertheless, its high cost and limited availability can restrict its widespread application, especially amongst large scale epidemiologic studies. Liver enzymes, such as alanine aminotransferase (ALT) and aspartate aminotransferase (AST), are routinely assessed in primary care to screen for liver fibrosis. However, these biomarkers are non-specific, and do not always correlate liver disease severity.5 Alternatively, the use of International Classification of Diseases (ICD) codes in electronic health records diagnosis provides a feasible approach for population-based diagnosis but may lead to misclassification and is insufficient for assessing disease severity.

MASLD comprises a spectrum of progressive liver conditions, ranging from isolated hepatic steatosis to metabolic dysfunction-associated steatohepatitis. Variants that contribute solely to the progression of steatohepatitis, fibrosis, or cirrhosis without promoting initial occurrence of steatosis may remain unidentified. To enhance the utility of drug-target MR, conducting GWAS on steatotic liver disease and its subcategories,6 diagnosed histologically or by imaging could provide more informative insights into treatment responses and disease progression.7-9

3

5. Figure 3. For factorial MR analysis, it is more efficient to use continuous data to test interactions? Do you think it will be more clinically relevant if we define the cutoff point according to some clinically used threshold, but not arbitrarily by median?

Thank you for raising this important point regarding the definition of cut-off point in factorial MR analysis. We have now expanded this discussion in “Informing adjunctive treatment using factorial Mendelian randomization” on page 10.

From: “Despite large sample size in genetic studies, statistical power to detect a statistical interaction using a dichotomized genetic risk score at median is typically low and may not be justified in all scenarios.10 More recent implementations additionally consider using continuous genetic risk scores to increase the statistical power to detect an interaction.11,12

”

To: “A previous methodological review has indicated that, despite large sample size in previous applications of factorial MR, the statistical power to detect interactions using a dichotomized genetic risk score at median is generally inefficient.10 More recent applications using continuous genetic risk scores have shown improved efficiency compared to using dichotomized scores.

11,12 Efficiency can be optimized

by maximizing the difference in the mean levels of risk factors across sufficiently large groups to detect statistical interactions. This can be achieved either by identifying a natural break in the risk factor distribution or by establishing a threshold that divide the population into equal-sized groups as much as possible.10”

Reviewer: 2

Comments to the Author

The review by Luo, S. et al summarized key aspects of drug-target Mendelian randomization study design, and highlighted its potential for the discovery, repositioning, and safety assessment of drug targets in metabolic dysfunction associated steatotic liver disease.

I have the following comments:

1) Sensitivity analyses are mentioned on page 5 (lines 57-59) and 6 (lines 8-9). Please clarify what sensitivity analyses you are referring to? Are you referring to pleiotropy-robust methods, such as MR Egger, mode- and median-based method etc? If so, they may not be reliable in the cis-MR setting (PMID: 32760811) because

these methods assume SNPs to be independent in terms of violating the IV assumptions. SNPs in the same gene region are likely to either all be valid IVs or all be invalid.

Thank you for your comments and we apologize for the lack of clarity regarding sensitivity analyses. We have now clarified this section to avoid misunderstanding.

From: "… The putative causal effect estimate can be obtained by Wald ratio using single instrument; however, this approach precludes certain sensitivity analyses that require multiple instruments. When multiple candidate variants are presented within

4

the protein-encoding gene, single-variant MR may not capture all the genetic effects in the region, potentially leading to a loss of statistical power. To address this issue, researchers often select multiple candidate variants in partial linkage disequilibrium (LD) as instruments to increase statistical power and facilitate sensitivity analyses in MR. Various techniques have been introduced for this purpose, including stepwise pruning, conditional analysis, principal component analysis, factor analysis, and Bayesian variable section.13

"

To: "… The putative causal effect estimate can be obtained using the Wald ratio with a single instrument. However, this approach precludes the applications of robust methods for sensitivity analyses that require multiple instruments, such as the weighted median and MR-Egger, which are commonly employed in genome-wide MR analyses. In cis-MR, multiple cis-variants may be available if the GWAS is sufficiently large. However, using variants from the same gene region could violate assumptions underlying sensitivity analyses due to shared pleiotropy and non independence. Therefore, genetic colocalization analysis is often used as a complementary analysis to cis-MR to assess potential biases arising from linkage disequilibrium (LD).14 When multiple causal variants are presented within a protein encoding gene, single-variant MR may not adequately capture all genetic effects in the region, potentially leading to a loss of statistical power. Conversely, including all genetic variants from the same gene region may result in numerical instability due to multicollinearity among the variants.13 To enhance statistical power, researchers

often select multiple candidate variants that are in partial LD as instruments. Various techniques have been introduced for this purpose, including stepwise-pruning, conditional analysis, principal component analysis, factor analysis, and Bayesian variable section.13

2) Page 7 lines 5-9 ‘To address this issue, researchers often select multiple candidate variants in partial linkage disequilibrium (LD) as instruments to increase statistical power and facilitate sensitivity analyses in MR.’ Again, ‘to facilitate sensitivity analyses’ may not be a valid motivation for the development of variable selection techniques you mentioned; Improving statistical power certainly is. Another motivation worth mentioning is that using all variants (i.e., where no variable selection process is involved) within the same gene region can lead to numerical instability due to multicollinearity between SNPs.

Thank you for your comments. We agree that the primary motivation for selecting multiple candidate variants in partial LD is to increase statistical power. Additionally, we also acknowledge that employing all variants within the same gene region can induce numerical instability due to multicollinearity among variants. We have now addressed these points in detail in our response to your first comment.

3) Page 5, line 60 ‘When multiple candidate variants are presented within the protein-encoding gene,…..’. You may consider changing ‘candidate’ variants to ‘causal’ variants.

5

We appreciate your suggestion to replace the term ‘candidate’ with ‘causal’ variants for improved clarity. This change has been made in the text.

4) Page 6, line 52, ‘….when using protein abundance as the instruments,…..’ should this be ‘when using protein abundance as the exposures,….’?

Thank you for identifying the typo. It has been corrected, “Ideally, a drug target MR will be most reliable when using protein abundance as the exposure, …”

5) ‘Weights’ is mentioned several times on page 7 without further explanation. Please elaborate on what is used as weights, and how they are applied in the cis-MR analyses.

Thank you for your comments. We have now clarified the weight as follows:

"…, researchers can use other phenotypes upstream or downstream of the druggable protein to search for relevant genetic variants and corresponding weights for MR analyses. The weights refer to the effect sizes of these genetic variants on the relevant phenotypes."

6) Page 7 lines 21-23 "Colocalization of cis-eQTL and cis-pQTL exhibiting the same regulatory direction would further support a druggable protein as candidate target.37" It is not clear to me how this fits in the paragraph, which, if I understand correctly, focuses on using gene expression data as exposure. Please clarify.

Thank you for your comment, we agree that the statement regarding the colocalization of cis-eQTL and cis-pQTL does not fit well within the context of using eQTL data as exposure. Therefore, we have removed this statement to avoid confusion.

7) Page 7 lines 50-55 For the binary outcomes, please clarify whether they are the outcomes of interest or phenotypes that lie in the pathway between exposure and outcome. I assume it is the latter.

Thank you for your comments. We have now clarified:

"Alternatively, genetic variants may be weighted by their association with a binary disease outcome, which is an intermediate phenotype in the pathway between the exposure and the outcome of interest."

8) Page 8 line 15 For the genome-wide and drug-target MR analyses, are you referring to the MR analyses on the phenotypes downstream of the druggable proteins? If so, you may need to clarify this to avoid confusion.

6

Thank you for your comment. To avoid confusion, we have revised the text as follows:

From: "Combining genome-wide and drug-target MR analyses can address complementary research questions, where possible. The former helps establish a causal relationship between a modifiable exposure and the outcome, while the latter investigates whether perturbation of drug class on this exposure could mitigate the

outcome risk."

To: "Combining genome-wide and drug-target MR analyses, where feasible, can address complementary research questions. The genome-wide MR establishes a causal relationship between a modifiable exposure, which is typically a downstream phenotype of the drug (e.g., LDL-C), and the outcome. In contrast, the drug-target MR investigates whether perturbations of this exposure through a specific drug target or mechanism (e.g., LDL-C reduction via HMGCR inhibition) could mitigate the outcome risk."

9) Page 8 line 41 '….highlighting triglyceride-lowering drug targets…' should it be 'triglyceride-lowering drug'?

Thank you for your suggestion. We have revised the sentence to read, "highlighting triglyceride-lowering drugs as potential treatments."

10) Page 8 lines 50-54 Please clarify why there is a discrepancy between mechanism-specific effects of LPL activation and systemic triglyceride level reduction? I ask this because based on what you have described here the evidence seem to be consistent:

o Triglyceride-NAFLD association: 'Accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD'.

o LPL-NAFLD association: '…not all therapeutic targets aim at lowering plasma triglycerides reduce the risk of NAFLD, except the drug target LPL (lipoprotein lipase).'.

Thank you for bring this important point to our attention. We acknowledge that if genetic variants associated with the same exposure have effects on the outcomes in opposite directions, this can indicate mechanism-specific effects.15 We apologize that our original phrasing may have caused confusion, and we have now revised it as follows:

From: "In the context of MASLD, corresponding genetic studies have uncovered susceptible genes such as PNPLA3, TM6SF2, GCKR, and LPL.16 17 For example, LPL is the main enzyme regulating the hydrolysis of triglycerides-rich lipoproteins,18

which provided plausible biological relevance to excessive hepatic triglycerides accumulation as a common pathological feature of MASLD, and highlighting

7

triglyceride-lowering drugs as potential treatments.19 Accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD.20-23 However, recent drug-target MR study suggests not all therapeutic targets aim at lowering plasma triglycerides reduce the risk of NAFLD, except the drug target LPL (lipoprotein lipase).21 The discrepancy between mechanism-specific effects of LPL activation and systemic triglyceride level reduction might imply a unique mechanism of action on MASLD that is independent of its triglyceride-lowering effects. This highlights the opportunities to develop new drugs targeting LPL activation as their central mechanistic effect to reduce MASLD risk. Moreover, understanding the underlying biological mechanisms, such as insulin resistance, can further enhance the development of these LPL-targeting drug.24

”

To: “Of note, it is plausible that targeting the same phenotype via different mechanisms can yield varying effects on outcomes. For example, accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD.20-23 However, recent drug-target MR studies have found that among various therapeutic targets (PPARA, ANGPTL3, ANGPTL4, APOC3 and LPL) aimed at lowering plasma triglycerides, only LPL has been shown to reduce the risk of NAFLD.21,25 LPL, the main enzyme regulating the hydrolysis of triglycerides-rich lipoproteins,18 is biological relevance to the excessive accumulation of hepatic triglycerides, a characteristic pathological feature of MASLD.19 Thus, targeting LPL activation may represent a unique therapeutic strategy to mitigate MASLD risk. Furthermore, insights into mechanistic nuances within broader drug indication categories (e.g., triglycerides-modifying agents) can enhance our understanding of the underlying biological pathways through metabolic profiling.26,27 A deeper understanding of these mechanisms could aid in refining drug development strategies, ultimately leading to more targeted and effective

interventions for the outcome."

11) Page 10 lines 9-11 Genetic risk score is mentioned here for the first time. You may need a sentence to clarify what it is.

Thank you. We have clarified the concept of genetic risk score and took this opportunity to explain the factorial design in more detail as follows:

From: "Utilizing MR approach in a factorial design can help investigate interactions between risk factors and explore interactions between pharmacological interventions on an outcome, similar to a factorial RCT.10 In the simplest 2 □ 2 factorial design, the study population can be divided into four groups using dichotomized genetic risk scores at their median, ensuring balanced numbers of participants across each group."

To: "Utilizing MR approach in a factorial design can help investigate the interactions between two (or potentially more) distinct exposures, as well as explore interactions between pharmacological interventions on an outcome, similar to a factorial RCT.10 In a factorial MR study, participants are categorised into different levels of each

8

exposure based on their genetic risk score, which is the sum of risk alleles corresponding to the exposure of interest, with or without weighting by their genetic associations with the exposure. In the simplest 2 □ 2 factorial MR design, the study population can be divided into four groups using dichotomized genetic risk scores at their median, ensuring balanced numbers of participants across each group."

12) Page10 line 35 A full stop is missing between 'scenarios' and 'More'.

Thank you. We have now added a full stop accordingly.

13) Page 10 line 54 It is not clear to me how the example of IL-6 receptor inhibition supports your earlier statement: 'To successfully reposition a drug, it is essential to identify genetic instruments influencing the molecular mechanisms of action of the drug.' Please clarify.

Thank you for your valuable feedback. The example of IL-6 receptor inhibition in the context of COVID-19, illustrates how leveraging genetic instruments can validate the repositioning of existing drugs based on their underlying molecular mechanisms,

particularly in reducing inflammatory signalling pathway. Furthermore, this example provides a foundation for discussing the challenges faced in drug repositioning, such as the unavailability of plausible genetic instruments for certain drug targets, unclear mechanisms of action, or when a drug exerts it effects through multiple pathways, as observed with SGLT2 inhibitors, aspirin and metformin.28-31

We have clarified the text as follows:

From: "… To successfully reposition a drug, it is essential to identify genetic instruments influencing the molecular mechanisms of action of the drug. For example, the genetic mimicry IL-6 receptor inhibition, which recapitulates the known downstream inflammatory signaling pathways (e.g., raising soluble IL-6 receptor and circulating IL-6 levels, and reducing C-reactive protein and fibrinogen), has been suggested to have a protective effect against COVID-19,32,33 which were subsequently corroborated by RCTs.34

Drug-target MR is also particularly promising for MASLD prevention and management. In order to address the epidemic of MASLD and prevent its complications, including cirrhosis and hepatocellular carcinoma, pharmacological interventions are being evaluated in clinical trials. These treatments include drug targeting energy efficiency and disposal (e.g., GLP1-RA and sodium-glucose cotransporter 2 (SGLT2) inhibitors),28-30 and those targeting lipotoxic liver injury and the resulting inflammation and fibrosis (e.g., metformin and aspirin).31 However, challenges arise where plausible genetic instruments are unavailable for certain drug targets, the drug's mechanisms of action is unclear, or when a drug exerts it effects through multiple pathways."

To: "… To successfully reposition a drug, it is essential to understand its known molecular mechanisms of action and to identify plausible genetic instruments that can mimic these effects. For example, genetic mimicry IL-6 receptor inhibition

9

recapitulates established downstream inflammatory signalling pathways, such as increased soluble IL-6 receptor and circulating IL-6 levels, and along with decreased C-reactive protein and fibrinogen. This mechanism has been suggested to confer

protective effects against COVID-19,32,33 as subsequently corroborated by RCTs.34 Similarly, drug target repositioning holds promising for addressing the epidemic of MASLD and preventing its complications, including cirrhosis and hepatocellular carcinoma. Current pharmacological interventions under evaluation include those targeting energy efficiency and disposal, such as GLP1-RA and sodium-glucose cotransporter 2 (SGLT2) inhibitors,28-30 as well as those mitigating lipotoxic liver injury and associated inflammation and fibrosis, such as metformin and aspirin.31 However, challenges may arise when plausible genetic instruments are unavailable for certain drug targets, when the drug's mechanisms of action are unclear, or when a drug exerts it effects through multiple pathways."

14) Page 13 line 35 There is a typo. Please change 'state-of-art' to 'state-of-the-art'.

Thank you for spotting this typo, we have corrected it accordingly.

10

References

1. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nat Rev Methods Primers. Feb 10 2022;2doi:10.1038/s43586-021-00092-5

2. Kivimaki M, Smith GD, Timpson NJ, et al. Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study. Eur Heart J. Oct 2008;29(20):2552-60. doi:10.1093/eurheartj/ehn252

3. Au Yeung SL, Jiang C, Cheng KK, et al. Is aldehyde dehydrogenase 2 a credible genetic instrument for alcohol use in Mendelian randomization analysis in Southern Chinese men? Int J Epidemiol. Feb 2013;42(1):318-28. doi:10.1093/ije/dys221

4. Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ. Jul 10 2014;349:g4164. doi:10.1136/bmj.g4164

5. European Association for the Study of the Liver. Electronic address eee, Clinical Practice Guideline P, Chair, representative EGB, Panel m. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and

prognosis - 2021 update. J Hepatol. Sep 2021;75(3):659-689. doi:10.1016/j.jhep.2021.05.025

6. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol. Dec 2023;79(6):1542-1556. doi:10.1016/j.jhep.2023.06.003

7. Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, et al. Multiomics study of nonalcoholic fatty liver disease. Nat Genet. Nov 2022;54(11):1652-1663. doi:10.1038/s41588-022-01199-5

8. Chen Y, Du X, Kuppa A, et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. Nat Genet. Oct 2023;55(10):1640-1650. doi:10.1038/s41588-023-01497-6

9. Anstee QM, Darlay R, Cockell S, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(☆). J Hepatol. Sep 2020;73(3):505-515. doi:10.1016/j.jhep.2020.04.003

10. Rees JMB, Foley CN, Burgess S. Factorial Mendelian randomization: using genetic variants to assess interactions. Int J Epidemiol. Aug 1 2020;49(4):1147-1158. doi:10.1093/ije/dyz161

11. Georgakis MK, Malik R, Burgess S, Dichgans M. Additive Effects of Genetic Interleukin-6 Signaling Downregulation and Low-Density Lipoprotein Cholesterol Lowering on Cardiovascular Disease: A 2x2 Factorial Mendelian Randomization Analysis. J Am Heart Assoc. Jan 4 2022;11(1):e023277. doi:10.1161/JAHA.121.023277

12. Cupido AJ, Reeskamp LF, Hingorani AD, et al. Joint Genetic Inhibition of PCSK9 and CETP and the Association With Coronary Artery Disease: A Factorial Mendelian Randomization Study. JAMA Cardiol. Sep 1 2022;7(9):955-964. doi:10.1001/jamacardio.2022.2333

13. Gkatzionis A, Burgess S, Newcombe PJ. Statistical methods for cis Mendelian randomization with two-sample summary-level data. Genet Epidemiol. Feb 2023;47(1):3-25. doi:10.1002/gepi.22506

14. Zuber V, Grinberg NF, Gill D, et al. Combining evidence from Mendelian

randomization and colocalization: Review and comparison of approaches. Am J Hum Genet. May 5 2022;109(5):767-782. doi:10.1016/j.ajhg.2022.04.001

11

15. Burgess S, Cronjé HT. Incorporating biological and clinical insights into variant choice for Mendelian randomisation: examples and principles. eGastroenterology. 2024;2(1):e100042. doi:10.1136/egastro-2023-100042

16. Trepo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. J Hepatol. Jun 2020;72(6):1196-1209. doi:10.1016/j.jhep.2020.02.020

17. Ghodsian N, Abner E, Emdin CA, et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non alcoholic fatty liver disease. Cell Rep Med. Nov 16 2021;2(11):100437. doi:10.1016/j.xcrm.2021.100437

18. Wu SA, Kersten S, Qi L. Lipoprotein Lipase and Its Regulators: An Unfolding Story. Trends Endocrinol Metab. Jan 2021;32(1):48-61. doi:10.1016/j.tem.2020.11.005

19. Semova I, Biddinger SB. Triglycerides in Nonalcoholic Fatty Liver Disease: Guilty Until Proven Innocent. Trends Pharmacol Sci. Mar 2021;42(3):183-190. doi:10.1016/j.tips.2020.12.001

20. Xie JR, Huang HK, Liu ZN, et al. The associations between modifiable risk factors and nonalcoholic fatty liver disease: A comprehensive Mendelian randomization study. Hepatology. Mar 2023;77(3):949-964. doi:10.1002/hep.32728

21. Li Z, Zhang B, Liu Q, et al. Genetic association of lipids and lipid-lowering drug target genes with non-alcoholic fatty liver disease. EBioMedicine. Apr 2023;90:104543. doi:10.1016/j.ebiom.2023.104543

22. Yuan S, Chen J, Li X, et al. Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study. Eur J Epidemiol. Jul 2022;37(7):723-733. doi:10.1007/s10654-022-00868-3

23. Gagnon E, Manikpurage HD, Mitchell PL, et al. Large-scale metabolomic profiling and incident non-alcoholic fatty liver disease. iScience. Jul 21 2023;26(7):107127. doi:10.1016/j.isci.2023.107127

24. Geldenhuys WJ, Lin L, Darvesh AS, Sadana P. Emerging strategies of targeting lipoprotein lipase for metabolic and cardiovascular diseases. Drug Discov Today. Feb 2017;22(2):352-365. doi:10.1016/j.drudis.2016.10.007

25. Landfors F, Henneman P, Chorell E, Nilsson SK, Kersten S. Drug-target Mendelian randomization analysis supports lowering plasma ANGPTL3, ANGPTL4, and APOC3 levels as strategies for reducing cardiovascular disease risk. Eur Heart J Open. May 2024;4(3):oeae035. doi:10.1093/ehjopen/oeae035

26. Wang Q, Oliver-Williams C, Raitakari OT, et al. Metabolic profiling of angiopoietin-like protein 3 and 4 inhibition: a drug-target Mendelian randomization analysis. Eur Heart J. Mar 21 2021;42(12):1160-1169. doi:10.1093/eurheartj/ehaa972

27. Richardson TG, Leyden GM, Wang Q, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. PLoS Biol. Feb 2022;20(2):e3001547. doi:10.1371/journal.pbio.3001547

28. Kuchay MS, Krishan S, Mishra SK, et al. Effect of Empagliflozin on Liver Fat in Patients With Type 2 Diabetes and Nonalcoholic Fatty Liver Disease: A Randomized Controlled Trial (E-LIFT Trial). Diabetes Care. Aug 2018;41(8):1801-1808. doi:10.2337/dc18-0165

29. Mantovani A, Byrne CD, Targher G. Efficacy of peroxisome proliferator activated receptor agonists, glucagon-like peptide-1 receptor agonists, or sodium glucose cotransporter-2 inhibitors for treatment of non-alcoholic fatty liver disease: a systematic review. Lancet Gastroenterol Hepatol. Apr 2022;7(4):367-378. doi:10.1016/S2468-1253(21)00261-2

12

30. Newsome PN, Buchholtz K, Cusi K, et al. A Placebo-Controlled Trial of Subcutaneous Semaglutide in Nonalcoholic Steatohepatitis. N Engl J Med. Mar 25 2021;384(12):1113-1124. doi:10.1056/NEJMoa2028395

31. Neuschwander-Tetri BA. Therapeutic Landscape for NAFLD in 2020. Gastroenterology. May 2020;158(7):1984-1998 e3. doi:10.1053/j.gastro.2020.01.051

32. Bovijn J, Lindgren CM, Holmes MV. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19. Lancet Rheumatol. Nov

2020;2(11):e658-e659. doi:10.1016/S2665-9913(20)30345-3

33. Larsson SC, Burgess S, Gill D. Genetically proxied interleukin-6 receptor inhibition: opposing associations with COVID-19 and pneumonia. Eur Respir J. Jan 2021;57(1)doi:10.1183/13993003.03545-2020

34. Brown MJ, Alazawi W, Kanoni S. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. N Engl J Med. Sep 16 2021;385(12):1147. doi:10.1056/NEJMc2108482

VERSION 2 – REVIEW

REVIEWER NAME	Yuan, Shuai
REVIEWER AFFILIATION	Karolinska Institutet, Institute of Environmental Medicine
REVIEWER CONFLICT OF INTEREST	I have no competing interests.
DATE REVIEW RETURNED	05-Sep-2024

GENERAL COMMENTS	Thank you for this great revision. Congrats!

REVIEWER NAME	Zhou, Ang
REVIEWER AFFILIATION	University of Cambridge, MRC Biostatistics Unit
REVIEWER CONFLICT OF INTEREST	I have no competing interest.
DATE REVIEW RETURNED	06-Sep-2024

GENERAL COMMENTS	Thank you for responding to all comments. All comments have been suitably addressed and I have no further comments.

VERSION 2 – AUTHOR RESPONSE

Associate Editor

Comments to the Author:

Thank you for your submission. The reviewers have some minor points - please consider these carefully in revising the manuscript.

We thank the reviewers' valuable comments concerning our manuscript (egastro-2024-100114). A cleaned manuscript and a track changed copy are provided. The major corrections in the manuscript and the point-by-point responses to the Reviewer's comments are in italic as follows, where changes made in the manuscript are underlined.

Reviewer: 1

Comments to the Author

Luo et al wrote a review paper to describe drug target discovery for metabolic dysfunction-associated steatotic liver disease using Mendelian randomization. In this study, they comprehensively summarized points of designing and appraising a drug-target Mendelian randomization study and presented examples to show the opportunities and challenges in this filed. The paper is well-written. The reviewer has a few points.

1. Page 4, line 20: "should not share unmeasured common causes with the outcome (independence)". Do you mean only residual confounding? It should be both forms of confounding.

Thank you for your comment. We have now revised the assumptions of Mendelian randomization to align with the literature.[1]

From: "Should not share unmeasured common causes with the outcome (independence)."

To: "The association of instrument in the outcome is not confounded (independence)."

2. Page 4, line 26. Please include "genetically predicted or proxied" before "a measure of pharmacological perturbation of the relevance drug target".

Thank you, we have now included as you suggested.

"In the context of drug-target MR, the genetically predicted or proxied exposure represents a measure of pharmacological perturbation of the relevance drug target."

3. Page 5, line 3. Please consider revising "Although less common now". Maybe it is not. Basically, all protein MR studies used this strategy, like using SNPs in coding region.

Thank you for your comment. We have now revised our statement to address your concerns, as follows:

From: "Although less common now, this approach was also used in earlier MR studies before the emergence of large genome-wide association studies (GWAS), such as obesity (FTO)[2] and alcohol consumption (ALDH2 and ADH1B).[3,4]"

To: "Prior to the emergence of large-scale genome-wide association studies (GWAS), earlier MR studies also employed this approach in selecting instruments, such as FTO variants for obesity,[2] and ALDH2 and ADH1B variants for alcohol consumption.[3,4]"

4. Given that these MR approaches are heavily based on accuracy of MASLD GWAS (mostly phenotyping of the disease), please discuss this point (defining the disease based on ICD, biopsy, imaging, liver enzymes, and so on) and possible related traits, like liver enzymes. Also, for different types of MASLD.

Thank you for raising this important point. We acknowledge that the accuracy and validity of MR investigation are heavily based on the quality of data derived from GWAS and the accurate phenotyping of the disease in question. We have now included a discussion on this topic on pages 14-15:

Variation in phenotyping of MASLD

The validity of an MR study is directly related to the quality of GWAS data, in particular the classification of the disease and trait measurement. Liver biopsy is the reference standard for diagnosing and staging hepatic diseases. However, its invasive nature limits its usage, especially in paediatric population. Amongst non-invasive methods, magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF) stands out for its superior accuracy in detecting and quantifying liver steatosis.[5] Nevertheless, its high cost and limited availability can restrict its widespread application, especially amongst large scale epidemiologic studies. Liver enzymes, such as alanine aminotransferase (ALT) and aspartate aminotransferase (AST), are routinely assessed in primary care to screen for liver fibrosis. However, these biomarkers are non-specific, and do not always correlate liver disease severity.[5] Alternatively, the use of International Classification of Diseases (ICD) codes in electronic health records diagnosis provides a feasible approach for population-based diagnosis but may lead to misclassification and is insufficient for assessing disease severity.

MASLD comprises a spectrum of progressive liver conditions, ranging from isolated hepatic steatosis to metabolic dysfunction-associated steatohepatitis. Variants that contribute solely to the progression of steatohepatitis, fibrosis, or cirrhosis without promoting initial occurrence of steatosis may remain unidentified. To enhance the

utility of drug-target MR, conducting GWAS on steatotic liver disease and its subcategories,[6] diagnosed histologically or by imaging could provide more informative insights into treatment responses and disease progression.[7-9]

5. Figure 3. For factorial MR analysis, it is more efficient to use continuous data to test interactions? Do you think it will be more clinically relevant if we define the cutoff point according to some clinically used threshold, but not arbitrarily by median?

Thank you for raising this important point regarding the definition of cut-off point in factorial MR analysis. We have now expanded this discussion in "Informing adjunctive treatment using factorial Mendelian randomization" on page 10.

From: "Despite large sample size in genetic studies, statistical power to detect a statistical interaction using a dichotomized genetic risk score at median is typically low and may not be justified in all scenarios.[10] More recent implementations additionally consider using continuous genetic risk scores to increase the statistical power to detect an interaction.[11,12] "

To: "A previous methodological review has indicated that, despite large sample size in previous applications of factorial MR, the statistical power to detect interactions using a dichotomized genetic risk score at median is generally inefficient.[10] More recent applications using continuous genetic risk scores have shown improved efficiency compared to using dichotomized scores.[11,12] Efficiency can be optimized by maximizing the difference in the mean levels of risk factors across sufficiently large groups to detect statistical interactions. This can be achieved either by identifying a natural break in the risk factor distribution or by establishing a threshold that divide the population into equal-sized groups as much as possible.[10]"

Reviewer: 2

Comments to the Author

The review by Luo, S. et al summarized key aspects of drug-target Mendelian randomization study design, and highlighted its potential for the discovery, repositioning, and safety assessment of drug targets in metabolic dysfunction-associated steatotic liver disease.

I have the following comments:

1) Sensitivity analyses are mentioned on page 5 (lines 57-59) and 6 (lines 8-9). Please clarify what sensitivity analyses you are referring to? Are you referring to pleiotropy-robust methods, such as MR Egger, mode- and median-based method etc? If so, they may not be reliable in the cis-MR setting (PMID: 32760811) because these methods assume SNPs to be independent in terms of violating the IV

assumptions. SNPs in the same gene region are likely to either all be valid IVs or all be invalid.

Thank you for your comments and we apologize for the lack of clarity regarding sensitivity analyses. We have now clarified this section to avoid misunderstanding.

From: "… The putative causal effect estimate can be obtained by Wald ratio using single instrument; however, this approach precludes certain sensitivity analyses that require multiple instruments. When multiple candidate variants are presented within the protein-encoding gene, single-variant MR may not capture all the genetic effects in the region, potentially leading to a loss of statistical power. To address this issue, researchers often select multiple candidate variants in partial linkage disequilibrium (LD) as instruments to increase statistical power and facilitate sensitivity analyses in MR. Various techniques have been introduced for this purpose, including stepwise-pruning, conditional analysis, principal component analysis, factor analysis, and Bayesian variable section.[13]"

To: "… The putative causal effect estimate can be obtained using the Wald ratio with a single instrument. However, this approach precludes the applications of robust methods for sensitivity analyses that require multiple instruments, such as the weighted median and MR-Egger, which are commonly employed in genome-wide MR analyses. In cis-MR, multiple cis-variants may be available if the GWAS is sufficiently large. However, using variants from the same gene region could violate assumptions underlying sensitivity analyses due to shared pleiotropy and non-independence. Therefore, genetic colocalization analysis is often used as a complementary analysis to cis-MR to assess potential biases arising from linkage disequilibrium (LD).[14] When multiple causal variants are presented within a protein-encoding gene, single-variant MR may not adequately capture all genetic effects in the region, potentially leading to a loss of statistical power. Conversely, including all genetic variants from the same gene region may result in numerical instability due to multicollinearity among the variants.[13] To enhance statistical power, researchers often select multiple candidate variants that are in partial LD as instruments. Various techniques have been introduced for this purpose, including stepwise-pruning, conditional analysis, principal component analysis, factor analysis, and Bayesian variable section.[13]

2) Page 7 lines 5-9 'To address this issue, researchers often select multiple candidate variants in partial linkage disequilibrium (LD) as instruments to increase statistical power and facilitate sensitivity analyses in MR.' Again, 'to facilitate sensitivity analyses' may not be a valid motivation for the development of variable selection techniques you mentioned; Improving statistical power certainly is. Another motivation worth mentioning is that using all variants (i.e., where no variable selection process is involved) within the same gene region can lead to numerical instability due to multicollinearity between SNPs.

Thank you for your comments. We agree that the primary motivation for selecting multiple candidate variants in partial LD is to increase statistical power. Additionally, we also acknowledge that employing all variants within the same gene region can induce numerical instability due to multicollinearity among variants. We have now addressed these points in detail in our response to your first comment.

3) Page 5, line 60 'When multiple candidate variants are presented within the protein-encoding gene,…..'. You may consider changing 'candidate' variants to 'causal' variants.

We appreciate your suggestion to replace the term 'candidate' with 'causal' variants for improved clarity. This change has been made in the text.

4) Page 6, line 52, '….when using protein abundance as the instruments,…..' should this be 'when using protein abundance as the exposures,….'?

Thank you for identifying the typo. It has been corrected, "Ideally, a drug target MR will be most reliable when using protein abundance as the exposure, …"

5) 'Weights' is mentioned several times on page 7 without further explanation. Please elaborate on what is used as weights, and how they are applied in the cis-MR analyses.

Thank you for your comments. We have now clarified the weight as follows:

"…, researchers can use other phenotypes upstream or downstream of the druggable protein to search for relevant genetic variants and corresponding weights for MR analyses. The weights refer to the effect sizes of these genetic variants on the relevant phenotypes."

6) Page 7 lines 21-23 "Colocalization of cis-eQTL and cis-pQTL exhibiting the same regulatory direction would further support a druggable protein as candidate target.37" It is not clear to me how this fits in the paragraph, which, if I understand correctly, focuses on using gene expression data as exposure. Please clarify.

Thank you for your comment, we agree that the statement regarding the colocalization of cis-eQTL and cis-pQTL does not fit well within the context of using eQTL data as exposure. Therefore, we have removed this statement to avoid confusion.

7) Page 7 lines 50-55 For the binary outcomes, please clarify whether they are the outcomes of interest or phenotypes that lie in the pathway between exposure and outcome. I assume it is the latter.

Thank you for your comments. We have now clarified:

"Alternatively, genetic variants may be weighted by their association with a binary disease outcome, which is an intermediate phenotype in the pathway between the exposure and the outcome of interest."

8) Page 8 line 15 For the genome-wide and drug-target MR analyses, are you referring to the MR analyses on the phenotypes downstream of the druggable proteins? If so, you may need to clarify this to avoid confusion.

Thank you for your comment. To avoid confusion, we have revised the text as follows:

From: "Combining genome-wide and drug-target MR analyses can address complementary research questions, where possible. The former helps establish a causal relationship between a modifiable exposure and the outcome, while the latter investigates whether perturbation of drug class on this exposure could mitigate the outcome risk."

To: "Combining genome-wide and drug-target MR analyses, where feasible, can address complementary research questions. The genome-wide MR establishes a causal relationship between a modifiable exposure, which is typically a downstream phenotype of the drug (e.g., LDL-C), and the outcome. In contrast, the drug-target MR investigates whether perturbations of this exposure through a specific drug target or mechanism (e.g., LDL-C reduction via HMGCR inhibition) could mitigate the outcome risk."

9) Page 8 line 41 '….highlighting triglyceride-lowering drug targets…' should it be 'triglyceride-lowering drug'?

Thank you for your suggestion. We have revised the sentence to read, "highlighting triglyceride-lowering drugs as potential treatments."

10) Page 8 lines 50-54 Please clarify why there is a discrepancy between mechanism-specific effects of LPL activation and systemic triglyceride level reduction? I ask this because based on what you have described here the evidence seem to be consistent:

o Triglyceride-NAFLD association: ‘Accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD’.

o LPL-NAFLD association: ‘…not all therapeutic targets aim at lowering plasma triglycerides reduce the risk of NAFLD, except the drug target LPL (lipoprotein lipase).’.

Thank you for bring this important point to our attention. We acknowledge that if genetic variants associated with the same exposure have effects on the outcomes in opposite directions, this can indicate mechanism-specific effects.[15] We apologize that our original phrasing may have caused confusion, and we have now revised it as follows:

From: “In the context of MASLD, corresponding genetic studies have uncovered susceptible genes such as PNPLA3, TM6SF2, GCKR, and LPL.[16 17] For example, LPL is the main enzyme regulating the hydrolysis of triglycerides-rich lipoproteins,[18] which provided plausible biological relevance to excessive hepatic triglycerides accumulation as a common pathological feature of MASLD, and highlighting triglyceride-lowering drugs as potential treatments.[19] Accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD.[20-23] However, recent drug-target MR study suggests not all therapeutic targets aim at lowering plasma triglycerides reduce the risk of NAFLD, except the drug target LPL (lipoprotein lipase).[21] The discrepancy between mechanism-specific effects of LPL activation and systemic triglyceride level reduction might imply a unique mechanism of action on MASLD that is independent of its triglyceride-lowering effects. This highlights the opportunities to develop new drugs targeting LPL activation as their central mechanistic effect to reduce MASLD risk. Moreover, understanding the underlying biological mechanisms, such as insulin resistance, can further enhance the development of these LPL-targeting drug.[24]”

To: “Of note, it is plausible that targeting the same phenotype via different mechanisms can yield varying effects on outcomes. For example, accumulating evidence from genome-wide MR studies consistently suggests a positive association between plasma triglycerides and NAFLD.[20-23] However, recent drug-target MR studies have found that among various therapeutic targets (PPARA, ANGPTL3, ANGPTL4, APOC3 and LPL) aimed at lowering plasma triglycerides, only LPL has been shown to reduce the risk of NAFLD.[21,25] LPL, the main enzyme regulating the hydrolysis of triglycerides-rich lipoproteins,[18] is biological relevance to the excessive accumulation of hepatic triglycerides, a characteristic pathological feature of MASLD.[19] Thus, targeting LPL activation may represent a unique therapeutic strategy to mitigate MASLD risk. Furthermore, insights into mechanistic nuances within broader drug indication categories (e.g., triglycerides-modifying agents) can enhance our understanding of the underlying biological pathways through metabolic profiling.[26,27] A deeper understanding of these mechanisms could aid in refining drug

development strategies, ultimately leading to more targeted and effective interventions for the outcome."

11) Page 10 lines 9-11 Genetic risk score is mentioned here for the first time. You may need a sentence to clarify what it is.

Thank you. We have clarified the concept of genetic risk score and took this opportunity to explain the factorial design in more detail as follows:

From: "Utilizing MR approach in a factorial design can help investigate interactions between risk factors and explore interactions between pharmacological interventions on an outcome, similar to a factorial RCT.[10] In the simplest 2 × 2 factorial design, the study population can be divided into four groups using dichotomized genetic risk scores at their median, ensuring balanced numbers of participants across each group."

To: "Utilizing MR approach in a factorial design can help investigate the interactions between two (or potentially more) distinct exposures, as well as explore interactions between pharmacological interventions on an outcome, similar to a factorial RCT.[10] In a factorial MR study, participants are categorised into different levels of each exposure based on their genetic risk score, which is the sum of risk alleles corresponding to the exposure of interest, with or without weighting by their genetic associations with the exposure. In the simplest 2 × 2 factorial MR design, the study population can be divided into four groups using dichotomized genetic risk scores at their median, ensuring balanced numbers of participants across each group."

12) Page10 line 35 A full stop is missing between 'scenarios' and 'More'.

Thank you. We have now added a full stop accordingly.

13) Page 10 line 54 It is not clear to me how the example of IL-6 receptor inhibition supports your earlier statement: 'To successfully reposition a drug, it is essential to identify genetic instruments influencing the molecular mechanisms of action of the drug.' Please clarify.

Thank you for your valuable feedback. The example of IL-6 receptor inhibition in the context of COVID-19, illustrates how leveraging genetic instruments can validate the repositioning of existing drugs based on their underlying molecular mechanisms, particularly in reducing inflammatory signalling pathway. Furthermore, this example provides a foundation for discussing the challenges faced in drug repositioning, such as the unavailability of plausible genetic instruments for certain drug targets, unclear

mechanisms of action, or when a drug exerts it effects through multiple pathways, as observed with SGLT2 inhibitors, aspirin and metformin.[28-31]

We have clarified the text as follows:

From: "… To successfully reposition a drug, it is essential to identify genetic instruments influencing the molecular mechanisms of action of the drug. For example, the genetic mimicry IL-6 receptor inhibition, which recapitulates the known downstream inflammatory signaling pathways (e.g., raising soluble IL-6 receptor and circulating IL-6 levels, and reducing C-reactive protein and fibrinogen), has been suggested to have a protective effect against COVID-19,[32,33] *which were subsequently corroborated by RCTs.*[34]

Drug-target MR is also particularly promising for MASLD prevention and management. In order to address the epidemic of MASLD and prevent its complications, including cirrhosis and hepatocellular carcinoma, pharmacological interventions are being evaluated in clinical trials. These treatments include drug targeting energy efficiency and disposal (e.g., GLP1-RA and sodium-glucose cotransporter 2 (SGLT2) inhibitors),[28-30] *and those targeting lipotoxic liver injury and the resulting inflammation and fibrosis (e.g., metformin and aspirin).*[31] *However, challenges arise where plausible genetic instruments are unavailable for certain drug targets, the drug's mechanisms of action is unclear, or when a drug exerts it effects through multiple pathways."*

To: "… To successfully reposition a drug, it is essential to understand its known molecular mechanisms of action and to identify plausible genetic instruments that can mimic these effects. For example, genetic mimicry IL-6 receptor inhibition recapitulates established downstream inflammatory signalling pathways, such as increased soluble IL-6 receptor and circulating IL-6 levels, and along with decreased C-reactive protein and fibrinogen. This mechanism has been suggested to confer protective effects against COVID-19,[32,33] *as subsequently corroborated by RCTs.*[34]

Similarly, drug target repositioning holds promising for addressing the epidemic of MASLD and preventing its complications, including cirrhosis and hepatocellular carcinoma. Current pharmacological interventions under evaluation include those targeting energy efficiency and disposal, such as GLP1-RA and sodium-glucose cotransporter 2 (SGLT2) inhibitors,[28-30] *as well as those mitigating lipotoxic liver injury and associated inflammation and fibrosis, such as metformin and aspirin.*[31] *However, challenges may arise when plausible genetic instruments are unavailable for certain drug targets, when the drug's mechanisms of action are unclear, or when a drug exerts it effects through multiple pathways."*

14) Page 13 line 35 There is a typo. Please change 'state-of-art' to 'state-of-the-art'.

Thank you for spotting this typo, we have corrected it accordingly.

References

1. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Methods Primers*. Feb 10 2022;2doi:10.1038/s43586-021-00092-5
2. Kivimaki M, Smith GD, Timpson NJ, et al. Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study. *Eur Heart J*. Oct 2008;29(20):2552-60. doi:10.1093/eurheartj/ehn252
3. Au Yeung SL, Jiang C, Cheng KK, et al. Is aldehyde dehydrogenase 2 a credible genetic instrument for alcohol use in Mendelian randomization analysis in Southern Chinese men? *Int J Epidemiol*. Feb 2013;42(1):318-28. doi:10.1093/ije/dys221
4. Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ*. Jul 10 2014;349:g4164. doi:10.1136/bmj.g4164
5. European Association for the Study of the Liver. Electronic address eee, Clinical Practice Guideline P, Chair, representative EGB, Panel m. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol*. Sep 2021;75(3):659-689. doi:10.1016/j.jhep.2021.05.025
6. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *J Hepatol*. Dec 2023;79(6):1542-1556. doi:10.1016/j.jhep.2023.06.003
7. Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, et al. Multiomics study of nonalcoholic fatty liver disease. *Nat Genet*. Nov 2022;54(11):1652-1663. doi:10.1038/s41588-022-01199-5
8. Chen Y, Du X, Kuppa A, et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat Genet*. Oct 2023;55(10):1640-1650. doi:10.1038/s41588-023-01497-6
9. Anstee QM, Darlay R, Cockell S, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(☆). *J Hepatol*. Sep 2020;73(3):505-515. doi:10.1016/j.jhep.2020.04.003
10. Rees JMB, Foley CN, Burgess S. Factorial Mendelian randomization: using genetic variants to assess interactions. *Int J Epidemiol*. Aug 1 2020;49(4):1147-1158. doi:10.1093/ije/dyz161
11. Georgakis MK, Malik R, Burgess S, Dichgans M. Additive Effects of Genetic Interleukin-6 Signaling Downregulation and Low-Density Lipoprotein Cholesterol Lowering on Cardiovascular Disease: A 2x2 Factorial Mendelian Randomization Analysis. *J Am Heart Assoc*. Jan 4 2022;11(1):e023277. doi:10.1161/JAHA.121.023277
12. Cupido AJ, Reeskamp LF, Hingorani AD, et al. Joint Genetic Inhibition of PCSK9 and CETP and the Association With Coronary Artery Disease: A Factorial Mendelian Randomization Study. *JAMA Cardiol*. Sep 1 2022;7(9):955-964. doi:10.1001/jamacardio.2022.2333
13. Gkatzionis A, Burgess S, Newcombe PJ. Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genet Epidemiol*. Feb 2023;47(1):3-25. doi:10.1002/gepi.22506
14. Zuber V, Grinberg NF, Gill D, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am J Hum Genet*. May 5 2022;109(5):767-782. doi:10.1016/j.ajhg.2022.04.001

15. Burgess S, Cronjé HT. Incorporating biological and clinical insights into variant choice for Mendelian randomisation: examples and principles. *eGastroenterology*. 2024;2(1):e100042. doi:10.1136/egastro-2023-100042
16. Trepo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. *J Hepatol*. Jun 2020;72(6):1196-1209. doi:10.1016/j.jhep.2020.02.020
17. Ghodsian N, Abner E, Emdin CA, et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep Med*. Nov 16 2021;2(11):100437. doi:10.1016/j.xcrm.2021.100437
18. Wu SA, Kersten S, Qi L. Lipoprotein Lipase and Its Regulators: An Unfolding Story. *Trends Endocrinol Metab*. Jan 2021;32(1):48-61. doi:10.1016/j.tem.2020.11.005
19. Semova I, Biddinger SB. Triglycerides in Nonalcoholic Fatty Liver Disease: Guilty Until Proven Innocent. *Trends Pharmacol Sci*. Mar 2021;42(3):183-190. doi:10.1016/j.tips.2020.12.001
20. Xie JR, Huang HK, Liu ZN, et al. The associations between modifiable risk factors and nonalcoholic fatty liver disease: A comprehensive Mendelian randomization study. *Hepatology*. Mar 2023;77(3):949-964. doi:10.1002/hep.32728
21. Li Z, Zhang B, Liu Q, et al. Genetic association of lipids and lipid-lowering drug target genes with non-alcoholic fatty liver disease. *EBioMedicine*. Apr 2023;90:104543. doi:10.1016/j.ebiom.2023.104543
22. Yuan S, Chen J, Li X, et al. Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study. *Eur J Epidemiol*. Jul 2022;37(7):723-733. doi:10.1007/s10654-022-00868-3
23. Gagnon E, Manikpurage HD, Mitchell PL, et al. Large-scale metabolomic profiling and incident non-alcoholic fatty liver disease. *iScience*. Jul 21 2023;26(7):107127. doi:10.1016/j.isci.2023.107127
24. Geldenhuys WJ, Lin L, Darvesh AS, Sadana P. Emerging strategies of targeting lipoprotein lipase for metabolic and cardiovascular diseases. *Drug Discov Today*. Feb 2017;22(2):352-365. doi:10.1016/j.drudis.2016.10.007
25. Landfors F, Henneman P, Chorell E, Nilsson SK, Kersten S. Drug-target Mendelian randomization analysis supports lowering plasma ANGPTL3, ANGPTL4, and APOC3 levels as strategies for reducing cardiovascular disease risk. *Eur Heart J Open*. May 2024;4(3):oeae035. doi:10.1093/ehjopen/oeae035
26. Wang Q, Oliver-Williams C, Raitakari OT, et al. Metabolic profiling of angiopoietin-like protein 3 and 4 inhibition: a drug-target Mendelian randomization analysis. *Eur Heart J*. Mar 21 2021;42(12):1160-1169. doi:10.1093/eurheartj/ehaa972
27. Richardson TG, Leyden GM, Wang Q, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol*. Feb 2022;20(2):e3001547. doi:10.1371/journal.pbio.3001547
28. Kuchay MS, Krishan S, Mishra SK, et al. Effect of Empagliflozin on Liver Fat in Patients With Type 2 Diabetes and Nonalcoholic Fatty Liver Disease: A Randomized Controlled Trial (E-LIFT Trial). *Diabetes Care*. Aug 2018;41(8):1801-1808. doi:10.2337/dc18-0165
29. Mantovani A, Byrne CD, Targher G. Efficacy of peroxisome proliferator-activated receptor agonists, glucagon-like peptide-1 receptor agonists, or sodium-glucose cotransporter-2 inhibitors for treatment of non-alcoholic fatty liver disease: a systematic review. *Lancet Gastroenterol Hepatol*. Apr 2022;7(4):367-378. doi:10.1016/S2468-1253(21)00261-2

30. Newsome PN, Buchholtz K, Cusi K, et al. A Placebo-Controlled Trial of Subcutaneous Semaglutide in Nonalcoholic Steatohepatitis. *N Engl J Med*. Mar 25 2021;384(12):1113-1124. doi:10.1056/NEJMoa2028395
31. Neuschwander-Tetri BA. Therapeutic Landscape for NAFLD in 2020. *Gastroenterology*. May 2020;158(7):1984-1998 e3. doi:10.1053/j.gastro.2020.01.051
32. Bovijn J, Lindgren CM, Holmes MV. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19. *Lancet Rheumatol*. Nov 2020;2(11):e658-e659. doi:10.1016/S2665-9913(20)30345-3
33. Larsson SC, Burgess S, Gill D. Genetically proxied interleukin-6 receptor inhibition: opposing associations with COVID-19 and pneumonia. *Eur Respir J*. Jan 2021;57(1)doi:10.1183/13993003.03545-2020
34. Brown MJ, Alazawi W, Kanoni S. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. *N Engl J Med*. Sep 16 2021;385(12):1147. doi:10.1056/NEJMc2108482