

STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies^{1 2}

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
1	TITLE and ABSTRACT	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study	1	Evaluating the causal effect of circulating proteome on the glycemic traits: Evidence from Mendelian randomization
INTRODUCTION				
2	Background	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question	4	Exploring the mechanisms underlying abnormal glycemic traits is important for further deciphering type 2 diabetes and characterizing novel drug targets. The MR assumption is met by utilizing protein quantitative trait loci (pQTLs) as instrumental variables for their respective proteins
3	Objectives	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects	5	The study aimed to identify circulating proteins causally related to glycemic traits through MR-based analytical framework, providing clues for the molecular pathological basis and potential avenues for diabetes treatment.
METHODS				
4	Study design and data sources	Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:		
	a)	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	6	The exposures in this study consisted of circulating proteins derived from the ten largest proteomic GWAS. The screening criteria are as follows: sample size >500, and measured proteins >50
	b)	Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis	6	The exposures in this study consisted of circulating proteins derived from the ten largest proteomic GWAS. The screening criteria are as follows: sample size >500, and measured proteins >50
	c)	Describe measurement, quality control and selection of genetic variants	6	First, single nucleotide polymorphisms (SNPs) were selected if they were associated with serum proteins according to the suggested p-value thresholds in the corresponding GWAS study (Supplementary Table 1). Second, to mitigate issues related to linkage disequilibrium (LD), SNPs within the human major histocompatibility complex (MHC) region (chr6: from 26 Mb to 34 Mb) were

				excluded from the analysis[22]. Third, LD clumping was conducted to identify independent pQTLs for each protein ($r^2 > 0.01$ and upstream/downstream distance $< 5000\text{kb}$). Last, instrumental SNPs associated with ≥ 5 proteins were excluded as these instruments are considered high pleiotropic nature.
	d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	7	The outcomes were genetically predicted glycemic traits including FG, 2hGlu, FI, and HbA1c. Summary data on the four glycemic traits were derived from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) GWAS (30% non-European ancestry)
	e)	Provide details of ethics committee approval and participant informed consent, if relevant	7	The data used in our study were obtained from public databases, ethical approval and informed consent were obtained from the respective institutional review committee.
5	Assumptions	Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis	6	First, single nucleotide polymorphisms (SNPs) were selected if they were associated with serum proteins according to the suggested p-value thresholds in the corresponding GWAS study (Supplementary Table 1). Second, to mitigate issues related to linkage disequilibrium (LD), SNPs within the human major histocompatibility complex (MHC) region (chr6: from 26 Mb to 34 Mb) were excluded from the analysis[22]. Third, LD clumping was conducted to identify independent pQTLs for each protein ($r^2 > 0.01$ and upstream/downstream distance $< 5000\text{kb}$). Last, instrumental SNPs associated with ≥ 5 proteins were excluded as these instruments are considered high pleiotropic nature.
6	Statistical methods: main analysis	Describe statistical methods and statistics used		
	a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)	8	The Wald ratio or the inverse variance weighted method were used to estimate the causal relationship when there was 1 or ≥ 2 pQTLs, respectively.
	b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	8	To address the potential of pleiotropy, instruments associated with ≥ 2 proteins were excluded in a sensitivity analysis.

	c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples	8	The Wald ratio or the inverse variance weighted method were used to estimate the causal relationship when there was 1 or ≥ 2 pQTLs, respectively.
	d)	Explain how missing data were addressed		
	e)	If applicable, indicate how multiple testing was addressed		
7	Assessment of assumptions	Describe any methods or prior knowledge used to assess the assumptions or justify their validity	8	The weighted median method will be used when heterogeneity between multiple instruments is detected by Cochrane's Q-tests, which method provide up to 50% invalid instruments [25]. Moreover, when horizontal pleiotropy was detected, the MR-Egger method will be conducted to correct the pleiotropy [24]
8	Sensitivity analyses and additional analyses	Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)	9	To evaluate potential confounding by linkage disequilibrium (LD), we examined whether cis-pQTLs and all pQTLs of the MR-prioritized proteins were associated with glycemic traits or if they were in LD with a distinct causal variant for glycemic traits. We used Bayesian colocalization analysis (the 'coloc' R package) to estimate the posterior probability of each genomic locus that contains a single variant affecting both the protein and the glycemic traits
9	Software and pre-registration			
	a)	Name statistical software and package(s), including version and settings used	7	Two-sample MR analyses were conducted using the "TwoSampleMR" R package
	b)	State whether the study protocol and details were pre-registered (as well as when and where)		
RESULTS				
10	Descriptive data			
	a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram	13	The pQTLs from the combination of ten GWAS were used to construct genetic instruments. Supplementary Figure 1 presents the selection procession.
	b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)	13	There were 4901 proteins (2643 unique proteins) with cis-pQTLs and 4466 proteins (3112 unique

			proteins) with trans-pQTLs. 2560 of the total 6807 proteins were affected by both cis-pQTL and trans-pQTLs, while 2341 proteins and 1906 proteins were only act as cis instruments and trans instruments, respectively.
	c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	
	d)	For two-sample MR: <ul style="list-style-type: none"> i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples ii. Provide information on the number of individuals who overlap between the exposure and outcome studies 	8-9 The Bonferroni correction was performed for each outcome, with p values set as 0.05 divided by the number of proteins tested
11	Main results		
	a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale	13 Overall, after Bonferroni correction of the thresholds, 36 protein-traits associations were determined (8 associated with FG, 3 associated with FI and 25 associated with HbA1c) (Table 1, Figure 2, Supplementary Table 4).
	b)	Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference	14 In the cis/trans-pQTLs MR analyses, we observed 12, 36, 14, and 55 proteins associated with 2hGlu, FG, FI, and HbA1c, respectively.
	c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
	d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	
12	Assessment of assumptions		
	a)	Report the assessment of the validity of the assumptions	
	b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2 , Q statistic or E-value)	
13	Sensitivity analyses and additional analyses		

	a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	14	all significant associations identified in the main analysis remain robust, with an additional 4 associations were newly identified (Supplemental table 6 and 7).
	b)	Report results from other sensitivity analyses or additional analyses	14	The results of Multivariate Mendelian Randomization of cis-pQTLs and cis+trans-pQTLs are presented in Supplementary Tables 11 and 12.
	c)	Report any assessment of direction of causal relationship (e.g., bidirectional MR)	15	The results of Steiger filtering analysis presented that all the associations identified by two-sample analysis and sensitivity analysis with cis-pQTLs had the correct causal directions from the proteins to the glycemic traits
	d)	When relevant, report and compare with estimates from non-MR analyses		
	e)	Consider additional plots to visualize results (e.g., leave-one-out analyses)		

DISCUSSION

14	Key results	Summarize key results with reference to study objectives	18	In summary, based on the MR analyses, 33 unique proteins were identified using cis-pQTLs, of which 11 proteins had strong evidence of colocalization, and 93 unique proteins were identified using all the pQTLs. Moreover, in-silico analyses were conducted to explore the mechanisms and potential therapeutic targets of identified proteins and to validate the causal inference.
15	Limitations	Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them	21-22	This study has some limitations. First, this analysis was limited to Europeans, which limits the generalizability of our findings to other populations. Second, although this study included a large number of plasma proteins, we may have inadvertently overlooked important proteins that lacked genetic tools. Third, because this study was based on pooled data, we cannot assess the risk of non-specific binding of proteins or the variability of protein measurements. Fourth, it is possible that cis-pQTL coding variants altering the amino acid sequence of the encoded protein have no actual effect on the function of proteins, instead they only impact the quantitative protein assay. Nevertheless, PAV assessment for the cis-pQTLs in this study provided clues for this issue
16	Interpretation			

	a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	18	This study is the first one to evaluate the causal association between plasma proteins and glycemic traits, which differs from previous MR studies focusing on the association between plasma proteins and the risk of diabetes and diabetic complications
	b)	Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions	18	Of the proteins prioritized by our cis-pQTL MR analysis, some of these have been reported previously to be associated with the pathophysiology of diabetes. For example, the insulin receptor (INSR) serves as a critical mediator in the intricate insulin signaling cascade. Perturbations within these pathways significantly contribute to the pathogenesis of insulin resistance, a complex disorder characterized by compromised insulin responsiveness[36]
	c)	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	18	Perturbations within these pathways significantly contribute to the pathogenesis of insulin resistance, a complex disorder characterized by compromised insulin responsiveness[36]. R-Spondin 3 (RSPO3) impacts body fat distribution and adipose cell biology and predicts type 2 diabetes [37].
17	Generalizability	Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure	21	First, this analysis was limited to Europeans, which limits the generalizability of our findings to other populations.
OTHER INFORMATION				
18	Funding	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based	23	This work is supported by the National Natural Science Foundation of China (NSFC, 82103933) and the Scientific Research Level Upgrading Project of Anhui Medical University (2021xkjT006).
19	Data and data sharing	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where	12-13	In this study, we used the selected summary-level data of European ancestry. The data used in our study were obtained from public databases, ethical approval and informed consent were obtained from the respective institutional review committee. The data that support the findings of this study are available from the corresponding author upon reasonable request.
20	Conflicts of Interest	All authors should declare all potential conflicts of interest	23	Conflict of interests: None.

This checklist is copyrighted by the Equator Network under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license.

1. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) Statement. JAMA. 2021;under review.
2. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration. BMJ. 2021;375:n2233.