Completed STROBE-MR Checklist

Manuscript: Integrative proteogenomic, transcriptomic and structural prioritization of therapeutic targets for anorexia nervosa

Page numbers refer to the supplied Manuscript.docx rendered as a 31-page manuscript PDF. The original uploaded checklist contained content from a different study; all entries below were remapped to the current manuscript. Status labels identify items that are reported, partly reported, not explicitly reported or not applicable.

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
1	TITLE and ABSTRACT	Indicate Mendelian randomization (MR) as the study design in the title and/or abstract if that is a main purpose of the study	1-2	Reported - Abstract Methods: cis-pQTL instruments were evaluated against the PGC2017 and PGC2019 AN GWAS datasets using drug-target Mendelian randomization and Bayesian colocalization.
2	INTRODUCTION	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	2-3	Reported - Introduction: AN treatment gap and biology; protein and transcript QTLs are the molecular exposures; MR is used to reduce confounding and reverse causation, with colocalization and SMR/HEIDI addressing LD and pleiotropy.
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	3, 5	Reported - Introduction objective: establish a transparent multi-layer prioritization framework and identify candidates for experimental validation. Methods prespecify IVW P<0.05, concordant beta direction and PP.H4>=0.70 as prioritization criteria; the three MR assumptions are not explicitly restated here.
4	METHODS - 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:	3-8	Reported - Methods sections 1.1-1.7 describe the summary-statistics integration design and all analysis layers. Data resources are summarized in Table 1 and Additional file 1: Table S34.
4a	Study design and data sources	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	3-4	Reported - Methods 1.1: PGC2017 and PGC2019 AN GWASs, plasma proteogenomic meta-analysis, eQTLGen Phase I release and GTEx v10 brain tissues are identified; release dates and cohort metadata are provided in Table 1 and Table S34.
4b	Study design and data sources	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	3-4, 25	Reported - Methods 1.1 reports AN case/control counts, 38 proteogenomic cohorts and up to 78,664 participants, eQTLGen 31,684 participants from 37 cohorts and GTEx tissue sample ranges. No a priori power calculation is reported.
4c	Study design and data sources	Describe measurement, quality control and selection of genetic variants	4-5	Reported - Methods 1.2-1.3: P<5e-8, F>=10, LD clumping (r2=0.001; 10,000 kb), GRCh38/Ensembl harmonization, allele alignment, palindromic-variant filtering and minimum overlap requirements.
4d	Study design and data sources	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	3-4	Reported - Methods 1.1: exposures are genetically predicted plasma proteins or transcripts; the outcome is AN case-control status from PGC2017 and PGC2019 GWAS summary statistics; source metadata are in Table 1 and Table S34.
4e	Study design and data sources	Provide details of ethics committee approval and participant informed consent, if relevant	20	Not applicable - Not applicable: the study uses publicly available summary-level genetic data. The manuscript states "Ethics approval and consent to participate: Not applicable."
5	Assumptions	Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well as assumptions for any additional or sensitivity analysis	4-8	Partly reported - Instrument strength, LD control, colocalization, pleiotropy annotation and sensitivity analyses are described in Methods 1.2-1.7, but the three core IV assumptions are not explicitly stated as a numbered set.
6	Statistical methods: main analysis	Describe statistical methods and statistics used	4-15	Reported - Methods 1.2-1.7 describe MR estimators, Bayesian colocalization, SMR/HEIDI, sensitivity analyses, eQTL follow-up, enrichment, PPI, druggability and docking; Results 2.1-2.9 report their outputs.
6a	Statistical methods: main analysis	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)	4-6, 25-27	Partly reported - Beta, SE, odds ratios, P values, F statistics and variance explained are used; the primary pQTL MR and GTEx expression effects are interpreted per genetically predicted exposure increment. A single consolidated units/scaling statement is not provided.

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
6b	Statistical methods: main analysis	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	4-5	Reported - Variants were selected at P<5e-8, required F>=10, clumped at r2=0.001 over 10,000 kb, harmonized with action=2 and analyzed with ratio/variance-weighted estimators; coloc used regional variants without MR clumping.
6c	Statistical methods: main analysis	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	4-8	Reported - IVW was prespecified as the primary estimator; MR-Egger, weighted median, simple/weighted mode and Wald ratio were also used. No covariates were fitted to summary-level genetic associations.
6d	Statistical methods: main analysis	Explain how missing data were addressed	5-7, 9	Partly reported - The manuscript reports successful, skipped and failed assay counts and states when instruments were unavailable; however, a general missing-summary-statistics rule is not explicitly stated.
6e	Statistical methods: main analysis	If applicable, indicate how multiple testing was addressed	5, 7, 9-13	Partly reported - Assay-level FDR<0.10 and Benjamini-Hochberg FDR<0.05 are reported for selected analyses; the manuscript does not provide one unified multiplicity plan across all layers.
7	Assessment of assumptions	Describe any methods or prior knowledge used to assess the assumptions or justify their validity	6, 10-13	Partly reported - Steiger directionality, F statistics, Cochran Q, MR-Egger intercept, MR-PRESSO, HEIDI and exact-rsID pleiotropy screening assess instrument validity and potential violations. The core IV assumptions themselves are not separately articulated.
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)	6-8, 10-15	Reported - Core-target sensitivity analyses, pQTL/eQTL SMR/HEIDI, blood eQTL, GTEx brain MR/coloc, pleiotropy, enrichment, STRING, druggability and exploratory docking are described and reported.
9	Software and pre-registration	Software and pre-registration	4-8	Partly reported - The computational tools used across the analysis layers are named; version and setting details are incomplete in places.
9a	Software and pre-registration	Name statistical software and package(s), including version and settings used	4-8	Partly reported - PLINK, TwoSampleMR, coloc.abf, STRING, CB-Dock2, Open Targets and ChEMBL are named. Some package versions/settings are not fully specified.
9b	Software and pre-registration	State whether the study protocol and details were pre-registered (as well as when and where)	-	Not reported - No preregistration statement is reported.
10	RESULTS - Descriptive data	Descriptive data	3-9	Reported - Participant and assay resources are described in Methods 1.1-1.2 and screening counts are reported in Methods 1.2 and Results 2.1.
10a	Descriptive data	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram	5, 9	Reported - The 1,161-assay screen reports successful, skipped and failed counts for each AN dataset; 141 assay-level candidates were selected for regional follow-up and non-core TRAIL assays were excluded from the eight-target set.
10b	Descriptive data	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g., means, SDs, proportions)	3-6, 26-27	Reported - Sample sizes are in Table 1; beta, SE, OR and F statistics are reported in Methods and Tables 2-4. Individual-level means/SDs are not applicable to summary-statistics analyses.
10c	Descriptive data	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	3-4	Not reported - The contributing GWAS and proteogenomic meta-analysis sources are described, but heterogeneity across their contributing cohorts is not newly assessed in this manuscript.
10d	Descriptive data	For two-sample MR: (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	3-4, 19, 26	Partly reported - The manuscript states that PGC2017 and PGC2019 partially overlap and are treated as complementary rather than independent replication datasets. The number of exposure-outcome participant overlaps is not quantified.
11	Main results	Main results	9-15, 26-28	Reported - Results 2.1-2.9 and Tables 2-5 provide the integrated target evidence, SMR/HEIDI, GTEx, pleiotropy, network, druggability and docking findings.
11a	Main results	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale	4-5, 9-12, 28-29	Reported - Instrument-level exposure/outcome records are provided in Additional file 1: Tables S11-S12 and S27-S30; assay-level and target-level MR results are reported in Results 2.1-2.5.

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
11b	Main results	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	9-12, 26-27	Reported - Results and Tables 2 and 4 report beta, OR, 95% CI and P values, including the replicated NCAM1 protein-level signal and cerebellar CRIM1 GTEx estimates.
11c	Main results	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-	Not applicable - Not applicable/not reported because the analysis uses summary-level genetic effects on AN case-control risk and does not estimate absolute individual risk.
11d	Main results	Consider plots to visualize results (e.g., forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	9-15, 30-31	Reported - Figures 2-7 visualize integrated MR/coloc evidence, representative loci, CRIM1 brain localization, pleiotropy, network/enrichment and docking results.
12	Assessment of assumptions	Assessment of assumptions	10-13	Reported - Assumption-related statistics are reported for the main pQTL, blood eQTL and GTEx analyses and in the supplementary tables.
12a	Assessment of assumptions	Report the assessment of the validity of the assumptions	10-13	Reported - Steiger directionality, F statistics, Q, Egger, MR-PRESSO, HEIDI and exact-instrument pleiotropy results are reported; single-SNP analyses are identified as unable to support heterogeneity testing.
12b	Assessment of assumptions	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I2, Q statistic or E-value)	10-13, 28-29	Reported - Cochran Q, MR-Egger intercept, MR-PRESSO, HEIDI, F statistics, Steiger results and GWAS Catalog pleiotropy summaries are provided in Tables S5-S10, S22 and S24-S26.
13	Sensitivity analyses and additional analyses	Sensitivity analyses and additional analyses	6, 10-15	Reported - Sensitivity and follow-up analyses are described in Methods 1.4-1.7 and Results 2.2-2.9.
13a	Sensitivity analyses and additional analyses	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	6, 10-13	Reported - Steiger, Q, Egger, MR-PRESSO, HEIDI, alternate transcript/tissue analyses and pleiotropy annotation were used to assess robustness and potential violations.
13b	Sensitivity analyses and additional analyses	Report results from other sensitivity analyses or additional analyses	10-15	Reported - Blood and brain eQTL follow-up, colocalization, enrichment, PPI, druggability and docking provide additional evidence layers and are reported in Results 2.3-2.9.
13c	Sensitivity analyses and additional analyses	Report any assessment of direction of causal relationship (e.g., bidirectional MR)	6, 10, 12	Reported - Steiger filtering favored the exposure-to-AN direction in the core pQTL and GTEx analyses. Bidirectional MR was not performed.
13d	Sensitivity analyses and additional analyses	When relevant, report and compare with estimates from non-MR analyses	10-15	Not applicable - No independent non-MR causal estimates are reported; colocalization, SMR/HEIDI, enrichment, PPI, annotation and docking are complementary analyses rather than alternative causal estimators.
13e	Sensitivity analyses and additional analyses	Consider additional plots to visualize results (e.g., leave-one-out analyses)	9-15, 30-31	Reported - Figures 2-7 provide the principal visual summaries; the manuscript also reports leave-one-out analysis where estimable.
14	DISCUSSION - Key results	Summarize key results with reference to study objectives	15-20	Reported - Discussion and Conclusions summarize NCAM1 as the leading protein-level candidate, CRIM1 as the leading exploratory transcript-level candidate and CALCA as a pathway-level repurposing hypothesis.
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	19	Reported - Limitations include partial GWAS overlap, reconstructed eQTL SMR beta/SE, single-SNP/low-SNP analyses, incomplete FDR support for enrichment and computational-only docking; the manuscript identifies replication and perturbation needs.
16	Interpretation	Interpretation	15-20	Reported - The Discussion interprets the findings in biological, translational and clinical context while repeatedly distinguishing hypotheses from validated therapeutic effects.
16a	Interpretation	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	15-20	Reported - The manuscript states that results prioritize experimental hypotheses rather than clinically validated targets and discusses uncertainty from colocalization, single-SNP instruments and pleiotropy.

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
16b	Interpretation	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	16-18	Reported - Mechanistic discussion covers NCAM1 adhesion/plasticity, CRIM1 extracellular-matrix and developmental biology, CALCA/CGRP signaling and SORT1 lipid biology; causal language is qualified by instrument and colocalization limitations.
16c	Interpretation	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	17-20	Reported - The manuscript frames CALCA/CGRP as pathway-level repurposing potential and states that no target is clinically validated or treatment-ready; experimental validation is required.
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	4, 19	Not explicitly reported - The European LD reference panel is stated, but ancestry, exposure-period and exposure-level generalizability are not explicitly discussed.
18	OTHER INFORMATION - 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	20	Not reported - A Funding heading is present, but specific funding sources and funder roles are not reported.
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	4, 20, 26-31	Partly reported - Data availability states that derived results and metadata are in the article and Additional file 1, while PGC2017, PGC2019, eQTLGen and GTEx summary data are publicly available. A public code repository is not identified.
20	Conflicts of Interest	All authors should declare all potential conflicts of interest	20	Reported - Declaration of competing interest: the authors report no known competing financial interests or personal relationships.

Checklist source: Skrivankova VW et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization (STROBE-MR) Statement. JAMA. 2021;326(16):1614-1621. Explanation and elaboration: BMJ. 2021;375:n2233.
This completed checklist is a reporting aid and does not alter the manuscript.