Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

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

Xu Lin
Beijing Anding Hospital
Xiang-Yun Yang
Beijing Anding Hospital
Peng-Chong Wang
Beijing Anding Hospital
Yu-Xiao Bi
Beijing Anding Hospital
Zhan-Jiang Li
lizhj8@ccmu.edu.cn

Beijing Anding Hospital

Research Article

Keywords: Anorexia nervosa, Drug-target discovery, pQTL, eQTL, Colocalization, SMR

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11520215/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

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

Xu Lin[a], Xiang-Yun Yang[a], Peng-Chong Wang[a], Yu-Xiao Bi[a], Zhan-Jiang Li[a*]

[a]*Beijing Key Laboratory of Intelligent Drug Research and Development for Mental Disorders; National Clinical Research Center for Mental Disorders;National Center for Mental Disorders; Beijing Anding Hospital, Capital Medical University, Beijing, China*

*Corresponding author at: No. 5 Ankang Hutong Deshengmen Wai, Xicheng District, 100088 Beijing, China. E-mail address: lizhj8@ccmu.edu.cn (Z.-j. Li).

Abstract: Background: Anorexia nervosa (AN) lacks an approved pharmacotherapy for its core psychopathology. We aimed to prioritize genetically supported and therapeutically tractable molecular targets.**Methods:** Cis-pQTL instruments were evaluated against the 2017 and 2019 AN GWAS datasets using drug-target Mendelian randomization and Bayesian colocalization. Eight core targets were subsequently examined using sensitivity analyses, pQTL- and eQTL-based SMR/HEIDI, eQTLGen blood data, GTEx v10 brain tissues, pleiotropy annotation, enrichment analysis, protein-interaction networks, druggability annotation and exploratory molecular docking.**Results:** NCAM1 showed the most coherent protein-level evidence, with directionally concordant nominal IVW associations in both AN GWAS datasets and global PP.H4=0.726 in PGC2017. CRIM1 showed the most reproducible nominal transcript-level pattern across blood SMR and cerebellar GTEx MR analyses, although neither blood nor brain transcript-level colocalization provided strong evidence for a shared causal variant. CALCA received the highest translational annotation score because of established CGRP-pathway pharmacology; this result indicates pathway-level repurposing potential rather than direct evidence that CALCA itself is therapeutically inhibited. SORT1 showed substantial lipid and metabolic pleiotropy, which reduced the specificity of its interpretation.**Conclusions:** NCAM1 is the leading protein-level candidate, CRIM1 is the

leading exploratory transcript-level candidate, and CALCA represents a pathway-level repurposing hypothesis. These findings prioritize experimental hypotheses rather than clinically validated targets and require independent molecular and functional validation.

Keywords: Anorexia nervosa, Drug-target discovery, pQTL, eQTL, Colocalization, SMR

Introduction: Anorexia nervosa (AN) is characterized by persistent restriction of energy intake, significantly low body weight, intense fear of weight gain, and disturbed body-image perception. It often begins during adolescence or early adulthood and is associated with psychiatric and medical morbidity, recurrent or chronic illness, and elevated mortality (Treasure et al.2020;American Psychiatric Association 2013). Current treatment relies on nutritional rehabilitation, medical stabilization and specialized psychotherapy; response is heterogeneous, relapse is common, and no medication has been specifically approved for the core psychopathology of AN (Treasure et al.2020;Attia and Walsh 2009). The neurobiology of AN includes altered reward, interoceptive and cognitive-control processing, making targets that connect disease biology with pharmacological tractability an important translational priority (Treasure et al.2015;Kaye et al.2013).

Genetic studies have changed the biological framing of AN. The first genome-wide significant AN locus was reported in 2017 in 3,495 cases and 10,982 controls, together with genetic correlations with metabolic traits (Duncan et al.2017). A larger 2019 GWAS of 16,992 cases and 55,525 controls identified eight genome-wide significant loci and genetic correlations spanning psychiatric, anthropometric, lipid, glycaemic and physical-activity traits, supporting AN as a metabo-psychiatric disorder (Watson et al.2019). GWAS loci, however, rarely identify the effector protein or the direction of a therapeutically useful intervention.

Proteogenomic and transcriptomic QTL resources provide genetically anchored molecular instruments. Mendelian randomization (MR) can test whether genetically predicted protein or transcript levels are associated with disease risk while reducing confounding and reverse causation (Davey Smith and Hemani 2014;Sanderson et al.2022). Because MR signals can arise from linkage disequilibrium or horizontal pleiotropy, we complemented MR with

Bayesian colocalization (Giambartolomei et al.2014) and summary-data-based MR (SMR) with the HEIDI heterogeneity test (Zhu et al.2016;Verbanck et al.2018). We further examined blood eQTLGen data, brain-specific GTEx v10 eQTLs, annotated pleiotropy, pathway context, PPI structure and ligand docking.

The objective of this study was not to declare a single clinically actionable drug target, but to establish a transparent, multi-layer prioritization framework and identify candidates suitable for subsequent experimental validation. The analysis workflow is shown in Figure 1.

[Insert Fig. 1 here]

1. Research methods and materials

1.1 Study design and data sources

This was a summary-statistics integration study. The AN outcomes were the 2017 and 2019 Psychiatric Genomics Consortium GWASs (PGC2017: 3,495 cases, 10,982 controls, total N=14,477; PGC2019: 16,992 cases, 55,525 controls, total N=72,517) (Duncan et al.2017;Watson et al.2019). The two GWASs partially overlap because PGC2019 incorporated PGC-ED Freeze 1 samples; accordingly, they were used to assess cross-dataset consistency rather than independent replication. The primary protein exposure resource was the multi-cohort plasma proteogenomic meta-analysis by Koprulu et al., which integrated 38 cohorts and up to 78,664 participants, measured circulating proteins primarily with Olink proximity extension assay (PEA), and reported pQTLs for 1,116 proteins (Koprulu et al.2026). The discovery exposure layer consisted of 1,161 cis-pQTL assay files annotated to GRCh38 and a ±1-Mb window around the protein-coding transcription start site. Blood transcript analyses used the eQTLGen Phase I full cis-eQTL release (2019-12-11 cis-eQTLsFDR-ProbeLevel-CohortInfoRemoved-BonferroniAdded; 31,684 participants from 37 cohorts) (Võsa et al.2021), and tissue-specific analyses used GTEx Analysis Release V10 brain tissues (GTEx Consortium 2020;GTEx Consortium 2017). The data resources, sample sizes and analytical roles are summarized in Table 1 and Additional file 1: Table S34; the computational MR platform is described in reference (Hemani et al.2018).

[Insert Table 1 here]

1.2 cis-pQTL instrument selection and protein MR

For each pQTL assay, variants with $P<5\times10^{-8}$ were retained. When multiple records represented the same SNP, the record with the smallest P value was kept. Instrument strength was summarized as $F=\beta^2/SE^2$, and variants with $F\geq10$ were retained (Burgess and Thompson 2011). Linkage disequilibrium clumping used a European reference panel with a 10,000-kb window, $r^2=0.001$ and no additional P-value threshold, implemented with PLINK (Purcell et al.2007). The GRCh38 gene and variant framework was harmonized with Ensembl annotations (Cunningham et al.2022;McLaren et al.2016). Exposure and outcome data were formatted with TwoSampleMR and harmonized using action=2 (Hemani et al.2018). MR estimates were generated with IVW, MR-Egger, weighted median, simple mode, weighted mode and Wald ratio where only one instrument remained; IVW was prespecified as the primary estimate. Robustness interpretation followed established MR-Egger and weighted-median guidance (Burgess and Thompson 2017;Bowden et al.2015;Bowden et al.2016) and the reporting principles of STROBE-MR (Skrivankova et al.2021).

The initial pQTL MR scan processed 1,161 assays for each outcome. PGC2017 yielded 877 successful assays, 240 skipped assays and 44 failed assays (2,793 result rows); PGC2019 yielded 812 successful assays, 278 skipped assays and 71 failed assays (2,212 result rows). The final layered evidence table was built at the assay/source-file level, with target-level best evidence used to summarize multiple assays for the same target. Complete MR, OR, heterogeneity, pleiotropy, instrument and analysis-log files are provided in Additional file 1: Tables S1–S12.

1.3 Bayesian colocalization and target prioritization

Bayesian colocalization was performed for 141 pQTL assays that had been prioritized after the initial pQTL drug-target MR screening, rather than for all available pQTL assays. These assays showed nominal evidence of association with AN in at least one MR analysis and were selected for regional follow-up using the same two AN GWAS datasets. pQTL and

AN variants were joined by SNP and harmonized at the allele level. High-frequency palindromic variants (MAF > 0.42) were removed, and at least 50 overlapping variants were required for each analysis. No MR clumping or P-value prefiltering was applied after regional data were assembled. Colocalization was conducted using coloc.abf, with the pQTL sample size represented by the median sample size among overlapping SNPs and case-control proportions of 0.2414 for PGC2017 and 0.2343 for PGC2019. Global PP.H4 was used to assess compatibility with a shared causal variant, whereas SNP-level PP.H4 was reported separately and was not treated as a substitute for global PP.H4.

Layered prioritization used IVW $P < 0.05$ as nominal MR evidence, concordant beta directions across the two AN datasets as replicated directionality, and PP.H4 ≥ 0.70 as strong colocalization. These criteria were used to define candidates for layered follow-up and should not be interpreted as proof of causal therapeutic efficacy. The eight core targets were NCAM1, ADAM22, WFIKKN1, CRIM1, SORT1, DPP6, CALCA and SELPLG.

1.4 Core-target sensitivity analyses

Sensitivity analyses were intentionally restricted to the eight core targets rather than all screened proteins. Steiger directionality, instrument-strength summaries, Cochran'Q heterogeneity, MR-Egger intercept, leave-one-out analysis and MR-PRESSO were performed when the number of independent instruments was sufficient. The Steiger analysis supported the pQTL→AN direction in all 16 core-target/outcome runs (Hemani et al.2017). Tests requiring at least three or four variants were not performed for targets with fewer instruments, and such tests are not interpreted as negative evidence. Steiger directionality results are provided in Additional file 1: Table S22. Cochran's Q heterogeneity and MR-Egger intercept results are provided in Tables S5–S6 and S9–S10, respectively.

1.5 Blood eQTLGen MR, colocalization and SMR/HEIDI

The eight core genes were queried in eQTLGen blood data using the same genome-wide significance threshold ($P<5\times10^{-8}$), clumping parameters and harmonization procedure as the pQTL analysis. CALCA and DPP6 had no qualifying genome-wide significant eQTL instruments and were therefore absent from the primary blood eQTL MR results. Six genes

were retained for eQTL MR and eQTL colocalization. Colocalization used all available cis-eQTL variants rather than MR-clumped instruments. Full blood eQTL MR, SMR/HEIDI and colocalization results are provided in Additional file 1: Tables S15, S16 and S35.

SMR analyses were run with a common pipeline for pQTL and eQTL inputs. Parameters included a ±2-Mb cis window, MAF cutoff 0.01, allele-frequency difference thresholds of 0.2, p_eQTL-SMR=5×10^{-8}, p_HEIDI=1.57×10^{-3}, and $3\leq m\leq20$ HEIDI variants. For eQTL inputs, beta and SE were reconstructed from Z score, MAF and sample size because the supplied eQTL tables did not consistently contain direct beta/SE fields. The inherited column names in the pQTL SMR workbook were standardized here as pSMR and pHEIDI; they do not indicate eQTL biology.

1.6 GTEx v10 brain analyses

Only brain tissues were included from GTEx v10: amygdala (N=317–323), cerebellar hemisphere (N=483–498), cerebellum (N=463–475), cortex (N=469–481), hypothalamus (N=451–455), nucleus accumbens/basal ganglia (N=499–506) and putamen/basal ganglia (N=441–452). Primary instruments required $P<5\times10^{-8}$; a relaxed $P<1\times10^{-5}$ analysis was retained as exploratory. GTEx eQTL MR and Bayesian colocalization were run separately for both AN GWASs. Most estimable GTEx MR associations were single-variant Wald ratios, so heterogeneity, pleiotropy and leave-one-out analyses were generally not identifiable. Complete GTEx MR and colocalization outputs are provided in Additional file 1: Tables S17–S18.

1.7 Pleiotropy, enrichment, PPI, druggability and docking

Core pQTL instruments were queried using exact rsID matching in the GWAS Catalog and using PhenoScanner (Buniello et al.2019;Staley et al.2016). The uploaded PhenoScanner output tables were empty, indicating a query or retrieval failure rather than absence of pleiotropy; this analysis is therefore reported as unavailable. The GWAS Catalog screen was treated as a partial annotation because exact rsID searches do not capture unqueried or unreported associations. Gene-level GWAS Catalog summaries, matched associations and PhenoScanner status are supplied in Additional file 1: Tables S24–S26.

Over-representation analysis used 125 MR targets against a 1,031-target universe and offline MSigDB-derived Hallmark, Gene Ontology, KEGG, Reactome and WikiPathways annotations (Subramanian et al.2005;Liberzon et al.2015). Hypergeometric P values were corrected by the Benjamini–Hochberg procedure (Reiner et al.2003). STRING was used to create a functional association network (Szklarczyk et al.2023;Shannon et al.2003). Druggability was evaluated as a transparent annotation layer rather than a clinical treatment recommendation, following established target-tractability concepts (Ochoa et al.2021;Hopkins and Groom 2002;Finan et al.2017). For each core target, we recorded the genetic evidence tier, protein class, subcellular localization, brain and blood/plasma relevance, small-molecule tractability, antibody/biologic tractability, PROTAC tractability, clinical or pathway precedent, known drug/pathway examples, safety concerns and repurposing potential. Drug and target annotations were cross-checked against Open Targets and ChEMBL (Ochoa et al.2021;Mendez et al.2019). The final semi-quantitative score was calculated as: genetic-evidence score + tractability score + clinical/pathway-precedence score + pleiotropy penalty. Genetic evidence was scored 0 (no usable support), 1 (exploratory or single-layer evidence), 2 (replicated or multi-omic support) or 3 (strong integrated MR/colocalization support). Tractability was scored 0 (no plausible modality), 1 (limited or indirect modality) or 2 (plausible modality or accessible extracellular/membrane target). Clinical/pathway precedent was scored 0 (none), 1 (preclinical or pathway-level precedent), 2 (established target-class or related-mechanism precedent) or 3 (established clinical pharmacology). The pleiotropy penalty was 0 (no major competing signal), −1 (mild competing traits), −2 (substantial competing architecture) or −3 (strong competing pleiotropic burden). This is an internal prioritization rubric, not a validated clinical score or a standardized Open Targets metric. Direct target druggability was kept separate from pathway-level repurposing. Existing CGRP medicines provide pharmacological precedent for the CGRP pathway linked to CALCA, but they do not demonstrate direct inhibition or therapeutic modulation of the CALCA-encoded products in AN. Full annotations, component scores and source notes are provided in Additional file 1: Tables S19–S20 and S31–S34. Docking used the CB-Dock2

191 web server with an AutoDock Vina-based blind-docking algorithm (Liu et al.2022;Trott and
192 Olson 2010). Receptor structures were obtained from the RCSB Protein Data Bank and
193 ligands from PubChem (Berman et al.2000;Kim et al.2019). The principal systems were the
194 CGRP receptor axis (PDB 3N7S), SORT1/sortilin axis (PDB 4PO7) and SELPLG/PSGL-1
195 axis (PDB 1G1S). CB-Dock2 automatically identified ligand-binding cavities and generated
196 cavity-specific docking boxes; therefore, no common fixed grid box was used. The reported
197 contact residues are pocket-associated predictions and should not be interpreted as confirmed
198 hydrogen-bond or hydrophobic contacts without a dedicated interaction analysis. Complete
199 docking metadata are in Additional file 1: Table S23.

200 **2. Results**

201 *2.1 Protein-level MR and colocalization prioritized eight core targets*

202 Across the 1,161-assay pQTL screen, 672 PGC2017 IVW tests and 565 PGC2019 IVW
203 tests were available. Thirty-five PGC2017 IVW results and 43 PGC2019 IVW results were
204 nominally significant; Five PGC2019 IVW results remained below an assay-level FDR
205 threshold of 0.10. These corresponded to NCAM1, WFIKKN1 and three TRAIL-related
206 assays. The TRAIL-related assays were retained as assay-level FDR-positive findings but
207 were not included among the eight core targets because they did not meet the prespecified
208 cross-layer prioritization criteria. A total of 141 assay-level candidates met the prespecified
209 MR-based screening criterion in at least one AN GWAS. Each assay was subsequently
210 evaluated against both PGC2017 and PGC2019, and all completed colocalization analyses
211 met the minimum overlap requirement. Global PP.H4≥0.70 was observed for one assay in
212 PGC2017 and two assays in PGC2019. The complete assay-level MR and odds-ratio outputs
213 are available in Additional file 1: Tables S1−S4, while the integrated core-target summary is
214 shown in Table 2 and Additional file 1: Table S13. Representative regional concordance
215 panels for selected loci are shown in Figure 3.

216 NCAM1 showed the most coherent cross-dataset protein-level pattern. Its pQTL MR
217 estimates were directionally concordant and nominally significant in both datasets (PGC2017

β=−0.333, OR=0.717, P=0.00489; PGC2019 β=−0.187, OR=0.830, P=0.000651), and PGC2017 colocalization reached PP.H4=0.726 at rs11214489. ADAM22 had PGC2019 MR–colocalization support (P=0.0106; PP.H4=0.874), while WFIKKN1 had PGC2019 support with a positive direction (β=0.257, P=0.000755; PP.H4=0.704). CRIM1, SORT1, DPP6, CALCA and SELPLG showed directionally concordant nominal IVW associations in both datasets but did not reach PP.H4≥0.70.

[Insert Table 2 here]

[Insert Fig. 2 here]

[Insert Fig. 3 here]

2.2 Core-target sensitivity analyses

Sensitivity analyses were performed for the eight core pQTL targets across both AN GWAS datasets. All 16 Steiger tests supported the hypothesized direction from genetically predicted protein levels to AN, and the corresponding instrument-strength assessments indicated adequate strength, with F statistics exceeding the conventional threshold of 10. Cochran's Q tests and MR-Egger intercept tests were successfully performed for the analyzable target–outcome pairs, and none showed significant heterogeneity or evidence of directional horizontal pleiotropy. MR-PRESSO was estimable in five analyses and did not identify significant global distortion or influential outliers. Together, these results provide supportive evidence that the core-target associations were unlikely to be explained by reverse causality, substantial between-instrument heterogeneity, or detectable single-variant pleiotropic effects. Because several targets were instrumented by only one or two independent variants, some sensitivity analyses were not statistically estimable and were therefore not interpreted as negative findings. Complete results are provided in Additional file 1: S5–S6、S9–S10 and S22

2.3 pQTL-SMR and blood eQTL-SMR analyses

pQTL-based SMR and HEIDI analyses were conducted for the eight core protein targets in both AN GWAS datasets. Nominally significant SMR associations with HEIDI support were observed for CRIM1 and CALCA in both PGC2017 and PGC2019. For CRIM1, the SMR associations were β = −0.627, P_SMR = 0.0220 and P_HEIDI = 0.255 in PGC2017, and β = −0.349, P_SMR = 0.00725 and P_HEIDI = 0.067 in PGC2019. CALCA showed similar nominally significant associations in PGC2017 (P_SMR = 0.0314, P_HEIDI = 0.603) and PGC2019 (P_SMR = 0.0298, P_HEIDI = 0.349). SORT1, DPP6 and WFIKKN1 were supported only in PGC2019, with P_SMR values of 0.00861, 0.00796 and 0.000760, respectively, and all corresponding HEIDI tests passed. Although NCAM1 was nominally significant in PGC2019 (P_SMR = 0.00627), its HEIDI test failed the post-processing criterion (P_HEIDI = 0.00521); therefore, this result was not considered robust SMR evidence. No nominally significant pQTL-SMR association was observed for ADAM22 or SELPLG.

Blood eQTL-SMR analyses were performed for the six genes with genome-wide significant eQTL instruments. CALCA and DPP6 were excluded because no qualifying blood eQTL instruments were available. CRIM1 showed the most consistent transcript-level evidence, with nominally significant associations in both PGC2017 (P_SMR = 0.00894, P_HEIDI = 0.663) and PGC2019 (P_SMR = 0.0498, P_HEIDI = 0.311). ADAM22 was nominally significant only in PGC2019 (P_SMR = 0.0339, P_HEIDI = 0.413). The remaining genes did not show nominally significant blood eQTL-SMR associations. No blood eQTL-SMR result remained significant after global FDR correction at 0.05. The complete results are provided in Additional file 1: Tables S14 and S16.

[Insert Table 3 here]

2.4 Blood eQTL MR and colocalization

Among the six genes with available blood eQTL instruments, NCAM1 showed a nominally significant association with AN in PGC2017 (β = 1.107, P = 0.00586), whereas CRIM1 was nominally significant in PGC2019 (β = 0.372, P = 0.00486). ADAM22 was significant only in PGC2019 based on a single-instrument Wald ratio (P = 0.0272). NCAM1

in PGC2017 and CRIM1 in PGC2019 remained significant under the reported global FDR adjustment (FDR=0.0393 for both IVW associations). No blood eQTL MR association showed consistent evidence across both AN GWAS datasets.

Blood eQTL colocalization was available for 12 gene–GWAS pairs, but none reached global PP.H4 ≥ 0.50. Thus, the blood transcript-level results provide exploratory support without strong evidence of shared causal variants with AN. Detailed results are provided in Additional file 1: Tables S15, S16 and S35.

2.5 GTEx v10 brain analyses

GTEx v10 analyses were restricted to seven brain tissues. At the prespecified genome-wide significance threshold ($P<5\times10^{-8}$), 18 single-SNP MR associations were estimable for ADAM22, CRIM1, DPP6 and WFIKKN1. CRIM1 showed the clearest tissue-specific pattern: genetically predicted expression was inversely associated with AN risk in the cerebellar hemisphere for PGC2017 (β=−0.160, P=0.00958; OR=0.852, 95% CI 0.755–0.962) and PGC2019 (β=−0.0592, P=0.0363; OR=0.942, 95% CI 0.892–0.996), and in the cerebellum for PGC2017 (β=−0.221, P=0.00771; OR=0.802, 95% CI 0.681–0.943). The remaining associations were not nominally significant (Table 4). All 18 analyses used one independent SNP; consequently, heterogeneity and horizontal-pleiotropy statistics were not estimable. Steiger filtering supported the exposure-to-AN direction in all 18 analyses. Full results are provided in Additional file 1: Tables S17–S18.

When the eQTL threshold was relaxed to $P<1\times10^{-5}$, 25 one-SNP MR associations were obtained, comprising the 18 standard-threshold rows and seven additional tissue-outcome rows. No additional association reached nominal significance; the three nominal signals remained the CRIM1 associations in cerebellar hemisphere and cerebellum described above. In the 100 GTEx brain eQTL–AN colocalization analyses, no comparison reached PP.H4≥0.50 or PP.H4≥0.70. The largest PP.H4 was 0.226 for CRIM1 in cerebellum with PGC2017, while PP.H3 was 0.286, and the corresponding analyses were classified as favoring distinct signals. These findings support a cerebellar, CRIM1-specific MR signal but

298 do not establish that the eQTL and AN association are driven by the same causal variant (Fig.
299 4).

300 [Insert Table 4 here]

301 [Insert Fig. 4 here]

302 *2.6 Pleiotropy screening*

303 Exact rsID searches in the GWAS Catalog identified 115 associations across 27 core
304 pQTL instrument SNPs (Additional file 1: Tables S24 and S25). SORT1 accounted for 102
305 associations involving 84 traits, predominantly lipid and metabolic phenotypes, including
306 LDL and total cholesterol. This pattern supports applying a pleiotropy penalty and warrants
307 caution in interpreting SORT1 as an AN-specific target. WFIKKN1, CRIM1, NCAM1 and
308 DPP6 showed fewer associations, whereas ADAM22, CALCA and SELPLG had no
309 exact-instrument matches. These null findings do not exclude pleiotropy. PhenoScanner
310 returned no annotated rows and was therefore treated as unavailable rather than as evidence of
311 absent pleiotropy.

312 [Insert Fig. 5 here]

313 *2.7 Enrichment analysis*

314 The 125-target enrichment analysis and the eight-target core analysis did not yield any
315 term with global BH FDR<0.05. Source-specific signals were weaker and database-dependent:
316 the core set had two KEGG terms and ten WikiPathways terms with source-specific
317 FDR<0.10, while all 125 targets showed several GO molecular-function and one Reactome
318 signal at source-specific FDR<0.10. Nominal terms implicated cell adhesion and NCAM1
319 interactions, postsynaptic specialization, TGF-β/growth-factor binding, enzyme-inhibitor
320 activity, vesicle and Golgi biology, secretin-receptor biology and axon-related processes. The
321 DPP6 annotation to dipeptidyl-peptidase activity should be treated cautiously because DPP6
322 is a non-catalytic potassium-channel auxiliary subunit. These analyses provide biological
323 context rather than confirmatory pathway evidence; full enrichment terms are provided in
324 Additional file 1: Table S21, and the representative WikiPathways results are shown in Fig.
325 6A.

2.8 Functional association network and druggability annotation

The STRING network contained 18 nodes and 20 unique edges after deduplication, with functional modules centered on the CALCA/CALCB/RAMP1–3 axis, the ADAM22–CNTNAP2–DPP6–LGI1 axis, and SORT1/WFIKKN1 connections to NTSR1. Because many edges were text-mining dominated, the network is interpreted as a functional association network rather than a purely physical interaction map. CALCA, CALCB and RAMP1–3 were the highest-degree nodes,while most core seeds had degree one (Fig. 6B).

The translational annotation ranked CALCA highest (final score 6), driven by a clinical/pathway-precedence score of 3 for established CGRP pharmacology, despite only exploratory genetic evidence; this should be interpreted as pathway-level repurposing potential rather than evidence that CALCA itself is directly inhibited by current CGRP medicines. NCAM1, WFIKKN1 and ADAM22 each scored 4 because of stronger genetic support and moderate tractability, although direct CNS delivery and target-specific pharmacology remain limitations. DPP6 also scored 4 and was classified as a neuronal-excitability candidate; its non-catalytic DPP-like annotation should not be confused with DPP4 druggability. CRIM1 and SELPLG scored 3 and remain mechanistic or immune/vascular candidates requiring validation. SORT1 scored 1 after a −3 pleiotropy penalty, reflecting extensive lipid, metabolic and cardiovascular associations that reduce target-specific interpretability. Table 5 summarizes the component-score arithmetic, whereas the complete annotation matrix, modality-level notes, safety considerations and source URLs are provided in Additional file 1: Tables S19–S20 and S31–S34.

[Insert Table 5 here]

2.9 Molecular docking

Twelve docking records represented 12 unique compounds.The CGRP receptor axis (PDB 3N7S) showed the most negative predicted scores, including zavegepant (−14.8 kcal/mol), telcagepant (−13.4 kcal/mol), ubrogepant (−11.7 kcal/mol), atogepant (−11.4 kcal/mol) and

rimegepant (−11.4 kcal/mol). The PSGL-1/P-selectin interface (PDB 1G1S) yielded scores from −7.3 to −7.9 kcal/mol, and the SORT1/sortilin–neurotensin axis (PDB 4PO7) yielded −7.5 to −8.8 kcal/mol. The sialyl Lewis X entry is an endogenous reference glycan rather than a therapeutic drug. NCAM1, WFIKKN1, ADAM22, CRIM1 and DPP6 were not docked because no suitable pocket/ligand pair was available in the uploaded workflow. Scores should not be compared across receptors or interpreted as experimental binding affinities. The complete ligand and AutoDock Vina score summary is provided in Additional file 1: Table S23.Representative predicted poses are shown in Fig. 7

[Insert Fig. 6 here]
[Insert Fig. 7 here]

3. Discussion

This study was designed as a triangulation framework rather than as a single-test target list. Starting with 1,161 plasma cis-pQTL assay files, we combined protein-level MR, regional colocalization, core-target sensitivity analyses, blood and brain transcriptomic follow-up, pleiotropy annotation, network biology and exploratory structural modelling. The evidence did not converge on one target across every layer, but this was informative rather than merely negative: the analyses separated three complementary types of translational signal. NCAM1 showed the most coherent cross-dataset protein-level pattern, CRIM1 showed the most reproducible nominal transcript-level pattern, and the CGRP pathway linked to CALCA provided the clearest pathway-level repurposing hypothesis. These signals represent complementary prioritization evidence rather than independent confirmation of a single causal mechanism. Together, these profiles are compatible with the metabo-psychiatric nature of AN (Treasure et al.2015;Kaye et al.2013), in which neural-circuit, neuropeptide, extracellular-matrix and metabolic processes may contribute to risk without being represented by a single molecular mechanism.

NCAM1 provides the clearest protein-level example. Neural cell adhesion molecule 1 is a cell-surface adhesion protein with plausible roles in neurite extension, synaptic organization and activity-dependent plasticity (Kiss and Muller 2001). Genetically higher circulating

NCAM1 was associated with lower AN risk in both PGC datasets, and the pQTL-colocalization posterior reached PP.H4 = 0.726 in PGC2017. A biologically plausible interpretation is that NCAM1-related adhesion and plasticity pathways could influence neural processes relevant to interoceptive integration, reward valuation or body-image processing (Kaye et al.2013). This remains a testable hypothesis, because the plasma instrument does not establish whether the causal mediator is neuronal NCAM1, a shed extracellular isoform or a peripheral correlate of central biology. The PGC2019 pQTL-SMR association was nominally significant but did not pass the post-processing HEIDI criterion, and blood or brain eQTL colocalization was not established. NCAM1 should therefore be prioritized for protein-focused functional experiments, while the transcript-level mechanism and the relevant tissue compartment remain unresolved.

Evidence for the candidate mapped to CRIM1 converged across protein- and transcript-level analyses. CRIM1 protein levels showed nominal associations in pQTL-SMR in both AN GWASs with HEIDI support, whereas CRIM1 expression showed nominal associations in blood eQTL-SMR in both datasets and directionally consistent cerebellar GTEx MR estimates. However, these results were not accompanied by strong pQTL or eQTL colocalization evidence, and the brain MR estimates were based on single-SNP instruments.The target annotation and enrichment results place CRIM1 in an extracellular-matrix and growth-factor-binding context (Iyer et al.2016;Kolle et al.2000). Such biology could provide a mechanistic bridge between neurotrophic signalling, vascular or tissue remodelling and the metabolic adaptation that accompanies AN. Nevertheless, no pQTL or eQTL colocalization posterior crossed the prespecified threshold, and the brain MR estimates were based on single-SNP instruments. The candidate mapped to CRIM1 is therefore biologically coherent but not yet causally resolved. Perturbation experiments should test whether altering CRIM1 expression or CRIM1 protein levels changes neuronal plasticity, extracellular-matrix responses or metabolic phenotypes relevant to AN.

The CGRP pathway linked to CALCA represents a third translational route.CALCA encodes the calcitonin/CGRP precursor, linking the candidate to neuropeptide signalling,

vascular regulation and sensory-interoceptive pathways. The enrichment analysis showed exploratory signals for neuropeptide signalling and nerve-growth-factor-related processes, although none remained significant after global FDR correction. The candidate mapped to CALCA received the highest internal translational score because the CGRP pathway already has clinically used pharmacology (Yuan et al.2019). This score should not be read as evidence that current gepants directly inhibit CALCA: the docking models used receptor-axis structures, not the CALCA protein itself. The appropriate interpretation is therefore a pathway-level repurposing hypothesis, which could be tested with receptor-specific assays, tissue-expression analyses and experimental models of appetite, nociception and interoceptive processing.

Several additional targets add biologically interpretable, but weaker, hypotheses. ADAM22 is a neuronal membrane and synaptic-organization candidate (Fukata et al.2021) with strong PGC2019 colocalization evidence, although its SMR result was not significant. WFIKKN1 is a secreted multidomain protein with protease-inhibitor and growth-factor-binding annotations and was supported by PGC2019 pQTL-MR, colocalization and SMR, suggesting an extracellular trophic-signalling mechanism that merits testing. DPP6 was compatible with a neuronal excitability hypothesis (Malloy et al.2022;Sun et al.2011), whereas SELPLG/PSGL-1 was more naturally connected to immune and vascular adhesion biology. The corresponding docking analysis used the P-selectin interface rather than a direct AN-specific binding site. These candidates broaden the mechanistic landscape from synaptic regulation to extracellular growth-factor control and immune-vascular signalling, but their current evidence is insufficient to rank them as clinically actionable targets.

Evidence integration was particularly important for interpreting SORT1. SORT1 showed replicated pQTL-MR directionality and pQTL-SMR support, and sortilin-related trafficking provides a biologically plausible link to lipoprotein handling and metabolic signalling (Musunuru et al.2010). However, the instruments were heavily represented in lipid, metabolic and cardiovascular GWAS Catalog associations. This pleiotropy is not necessarily irrelevant to AN, which has a documented metabolic component, but it makes it difficult to attribute the

association specifically to SORT1-mediated AN biology. The downgrade is therefore an interpretive safeguard rather than a claim that SORT1 is biologically unrelated to AN. Future work should combine multivariable MR, conditional or fine-mapped colocalization with lipid and body-mass traits, and cellular perturbation to determine whether an AN-relevant effect remains after broader cardiometabolic pathways are accounted for.

The tissue-specific, network and structural analyses help place these molecular signals in context. The GTEx analysis was intentionally restricted to seven brain tissues; CRIM1 associations in cerebellar hemisphere and cerebellum were the most reproducible tissue-level findings. Cerebellar involvement could be relevant to predictive interoception, sensorimotor learning or affective integration (Gaudio et al.2018), but this remains a hypothesis because all primary GTEx MR estimates used one SNP and no brain eQTL-AN colocalization passed PP.H4 >= 0.50. The STRING network and enrichment results provide useful systems-level context, yet several network edges were text-mining supported and no enriched term passed global FDR < 0.05. Similarly, docking of gepants to the CGRP receptor axis, ligands to the SORT1/sortilin-neurotensin axis and compounds at the PSGL-1/P-selectin interface supports pathway plausibility rather than direct target engagement. The druggability annotation consequently separates genetic priority from pharmacological tractability: NCAM1, WFIKKN1 and ADAM22 have stronger genetic rationale but limited direct drug precedent, whereas CALCA has greater pathway precedent but weaker target-level genetic evidence.

The principal strengths of this study are the use of two complementary AN GWAS datasets, explicit instrument-strength filtering, core-target-only sensitivity analyses, separate consideration of global colocalization and SNP-level posterior probabilities, transcript-level and brain-tissue follow-up, and transparent downgrading when pleiotropy or testability limited interpretation. Several limitations should nevertheless guide the next stage of work. The two AN GWAS datasets are partially overlapping and therefore do not constitute independent replication cohorts. Blood eQTL beta and standard-error values used for SMR were reconstructed from summary fields, which may introduce measurement error. Many GTEx and core-target sensitivity analyses were single-SNP or low-SNP analyses, limiting

formal heterogeneity and pleiotropy assessment. Enrichment did not pass global FDR < 0.05, and docking scores remain computational predictions. The most informative next steps are allele-specific and protein-level perturbation of NCAM1 and CRIM1 in neuronal or brain-relevant models, receptor-focused testing of the CALCA/CGRP axis, multivariable and fine-mapped analyses for SORT1, and replication in independent molecular datasets. Accordingly, these results should be used to prioritize testable mechanistic experiments, not to imply clinical efficacy or immediate treatment use.

4. Conclusions

An integrated pQTL–MR, colocalization, transcriptomic, tissue-specific and translational framework prioritized several testable AN target hypotheses. NCAM1 showed the most coherent protein-level evidence across the two AN GWAS datasets. CRIM1 showed the most reproducible nominal transcript-level pattern across blood SMR and cerebellar GTEx MR analyses, but the absence of strong transcript-level colocalization and the predominance of single-SNP instruments limit causal interpretation. The CGRP pathway linked to CALCA showed the strongest pathway-level pharmacological precedent, although this does not establish CALCA as a directly validated therapeutic target. Overall, these findings should guide functional experiments and independent molecular replication rather than be interpreted as evidence of clinical efficacy or treatment readiness.

Ethics approval and consent to participate

Not applicable.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Funding

This study was supported by the Science and Technology Innovation 2030 – Major Project on Brain Science and Brain-like Research (Grant No. 2021ZD0202004).

CRediT authorship contribution statement

Xu Lin: Methodology, Investigation, Formal analysis, Data curation, Writing – original draft.Xiang-Yun Yang: Conceptualization, Writing – review & editing.Peng-Chong Wang: Funding acquisition, Supervision.Yu-Xiao Bi:Formal analysis.Zhan-Jiang Li: Funding acquisition, Project administration, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All derived results and analysis metadata are provided in this article and Additional file 1. The underlying PGC2017, PGC2019, eQTLGen and GTEx summary data are publicly available from the original consortia and data portals cited in Table 1 and Additional file 1: Table S34.

Acknowledgment

We thank the participants and investigators of the Psychiatric Genomics Consortium, eQTLGen, GTEx, and the contributing plasma proteogenomic studies for providing the data used in this research.

Reference

509 American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th
510 ed. Arlington (VA): American Psychiatric Association; 2013.
511 Attia E, Walsh BT. Behavioral management for anorexia nervosa. N Engl J Med.
512 2009;360(5):500-506. doi:10.1056/NEJMct0805569
513 Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res.
514 2000;28(1):235-242. doi:10.1093/nar/28.1.235
515 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments:
516 effect estimation and bias detection through Egger regression. Int J Epidemiol.
517 2015;44(2):512-525. doi:10.1093/ije/dyv080
518 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
519 Randomization with Some Invalid Instruments Using a Weighted Median
520 Estimator. Genet Epidemiol. 2016;40(4):304-314. doi:10.1002/gepi.21965
521 Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published
522 genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic
523 Acids Res. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
524 Burgess S, Thompson SG; CRP CHD Genetics Collaboration. Avoiding bias from weak
525 instruments in Mendelian randomization studies. Int J Epidemiol. 2011;40(3):755-764.
526 doi:10.1093/ije/dyr036
527 Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the
528 MR-Egger method. Eur J Epidemiol. 2017;32(5):377-389.
529 doi:10.1007/s10654-017-0255-x
530 Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. Nucleic Acids Res.
531 2022;50(D1):D988-D995. doi:10.1093/nar/gkab1049
532 Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in
533 epidemiological studies. Hum Mol Genet. 2014;23(R1):R89-R98.
534 doi:10.1093/hmg/ddu328
535 Duncan L, Yilmaz Z, Gaspar H, et al. Significant Locus and Metabolic Genetic Correlations
536 Revealed in Genome-Wide Association Study of Anorexia Nervosa. Am J Psychiatry.

2017;174(9):850-858. doi:10.1176/appi.ajp.2017.16121402

Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017;9(383):eaag1166. doi:10.1126/scitranslmed.aag1166

Fukata Y, Hirano Y, Miyazaki Y, Yokoi N, Fukata M. Trans-synaptic LGI1-ADAM22-MAGUK in AMPA and NMDA receptor regulation. Neuropharmacology. 2021;194:108628. doi:10.1016/j.neuropharm.2021.108628

Gaudio S, Olivo G, Beomonte Zobel B, Schiöth HB. Altered cerebellar-insular-parietal-cingular subnetwork in adolescents in the earliest stages of anorexia nervosa: a network-based statistic analysis. Transl Psychiatry. 2018;8(1):127. Published 2018 Jul 6. doi:10.1038/s41398-018-0173-z

Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. Published 2014 May 15. doi:10.1371/journal.pgen.1004383

GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-213. doi:10.1038/nature24277

GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776

Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13(11):e1007081. Published 2017 Nov 17. doi:10.1371/journal.pgen.1007081

Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408. Published 2018 May 30. doi:10.7554/eLife.34408

Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. 2002;1(9):727-730.

565 doi:10.1038/nrd892
566 Iyer S, Pennisi DJ, Piper M. CRIM1-, a regulator of developmental organogenesis. Histol
567 Histopathol. 2016;31(10):1049-1057. doi:10.14670/HH-11-766
568 Kaye WH, Wierenga CE, Bailer UF, Simmons AN, Bischoff-Grethe A. Nothing tastes as good
569 as skinny feels: the neurobiology of anorexia nervosa. Trends Neurosci.
570 2013;36(2):110-120. doi:10.1016/j.tins.2013.01.003
571 Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical
572 data. Nucleic Acids Res. 2019;47(D1):D1102-D1109. doi:10.1093/nar/gky1033
573 Kiss JZ, Muller D. Contribution of the neural cell adhesion molecule to neuronal and synaptic
574 plasticity. Rev Neurosci. 2001;12(4):297-310. doi:10.1515/revneuro.2001.12.4.297
575 Kolle G, Georgas K, Holmes GP, Little MH, Yamada T. CRIM1, a novel gene encoding a
576 cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate
577 CNS development and organogenesis. Mech Dev. 2000;90(2):181-193.
578 doi:10.1016/s0925-4773(99)00248-8
579 Koprulu M, Smith-Byrne K, Ferolito BR, et al. Multi-cohort proteogenomic analyses reveal
580 genetic effects across the proteome and diseasome. Cell. 2026;189(11):3339-3357.e11.
581 doi:10.1016/j.cell.2026.03.049
582 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
583 Signatures Database (MSigDB) hallmark gene set collection. Cell Syst.
584 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
585 Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CB-Dock2: improved protein-ligand blind
586 docking by integrating cavity detection, docking and homologous template
587 fitting. Nucleic Acids Res. 2022;50(W1):W159-W164. doi:10.1093/nar/gkac394
588 Malloy C, Ahern M, Lin L, Hoffman DA. Neuronal Roles of the Multifunctional Protein
589 Dipeptidyl Peptidase-like 6 (DPP6). Int J Mol Sci. 2022;23(16):9184. Published 2022
590 Aug 16. doi:10.3390/ijms23169184
591 McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol.
592 2016;17(1):122. Published 2016 Jun 6. doi:10.1186/s13059-016-0974-4

593 Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay
594 data. Nucleic Acids Res. 2019;47(D1):D930-D940. doi:10.1093/nar/gky1075
595 Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via
596 SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714-719.
597 doi:10.1038/nature09266
598 Ochoa D, Hercules A, Carmona M, et al. Open Targets Platform: supporting systematic
599 drug-target identification and prioritisation. Nucleic Acids Res.
600 2021;49(D1):D1302-D1310. doi:10.1093/nar/gkaa1027
601 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and
602 population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-575.
603 doi:10.1086/51979520.
604 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false
605 discovery rate controlling procedures. Bioinformatics. 2003;19(3):368-375.
606 doi:10.1093/bioinformatics/btf877
607 Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nat Rev Methods
608 Primers. 2022;2:6. doi:10.1038/s43586-021-00092-5
609 Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated
610 models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-2504.
611 doi:10.1101/gr.1239303
612 Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of
613 Observational Studies in Epidemiology Using Mendelian Randomization: The
614 STROBE-MR Statement. JAMA. 2021;326(16):1614-1621.
615 doi:10.1001/jama.2021.18236
616 Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human
617 genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-3209.
618 doi:10.1093/bioinformatics/btw373
619 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a
620 knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl

621 Acad Sci U S A. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
622 Sun W, Maffie JK, Lin L, Petralia RS, Rudy B, Hoffman DA. DPP6 establishes the A-type
623 K(+) current gradient critical for the regulation of dendritic excitability in CA1
624 hippocampal neurons. Neuron. 2011;71(6):1102-1115. doi:10.1016/j.neuron.2011.08.008
625 Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein
626 association networks and functional enrichment analyses for any sequenced genome of
627 interest. Nucleic Acids Res. 2023;51(D1):D638-D646. doi:10.1093/nar/gkac1000
628 Treasure J, Zipfel S, Micali N, et al. Anorexia nervosa. Nat Rev Dis Primers. 2015;1:15074.
629 Published 2015 Nov 26. doi:10.1038/nrdp.2015.74
630 Treasure J, Duarte TA, Schmidt U. Eating disorders. Lancet. 2020;395(10227):899-911.
631 doi:10.1016/S0140-6736(20)30059-3
632 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new
633 scoring function, efficient optimization, and multithreading. J Comput Chem.
634 2010;31(2):455-461. doi:10.1002/jcc.21334
635 Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in
636 causal relationships inferred from Mendelian randomization between complex traits and
637 diseases. Nat Genet. 2018;50(5):693-698. doi:10.1038/s41588-018-0099-7
638 Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify
639 thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat
640 Genet. 2021;53(9):1300-1310. doi:10.1038/s41588-021-00913-z
641 Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight
642 risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat Genet.
643 2019;51(8):1207-1214. doi:10.1038/s41588-019-0439-2
644 Yuan H, Spare NM, Silberstein SD. Targeting CGRP for the Prevention of Migraine and
645 Cluster Headache: A Narrative Review. Headache. 2019;59 Suppl 2:20-32.
646 doi:10.1111/head.13583
647 Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies
648 predicts complex trait gene targets. Nat Genet. 2016;48(5):481-487. doi:10.1038/ng.3538

Table 1. Summary of genetic resources and analyses

Dataset	Role	Sample size	Analyses
PGC2017 AN GWAS	Outcome	3,495 cases + 10,982 controls	pQTL MR, coloc, SMR/HEIDI
PGC2019 AN GWAS	Outcome	16,992 cases + 55,525 controls	pQTL MR, coloc, SMR/HEIDI
cis-pQTL assay library	Protein exposure	Up to 78,664 participants in source meta-analysis	MR and coloc screen
eQTLGen Phase I blood	Transcript exposure	31,684 participants; 37 cohorts	MR, coloc, SMR/HEIDI
GTEx v10 brain	Tissue-specific transcript exposure	N=317–506 across seven tissues	MR and coloc

Note. PGC2017 and PGC2019 are partially overlapping and are treated as complementary outcome datasets, not independent replication cohorts. The full data-source and software metadata are provided in Additional file 1: Table S34.

Table 2. Core pQTL MR and colocalization evidence

Target	β 2017	PGC2017 P	PGC2017 PP.H4	β 2019	PGC2019 P	PGC2019 PP.H4
NCAM1	-0.333	0.005	0.726	-0.187	6.51e-04	0.377
ADAM22	-0.120	0.103	0.032	-0.107	0.011	0.874
WFIKKN1	0.069	0.578	0.025	0.257	7.55e-04	0.704
CRIM1	-0.464	0.032	0.133	-0.269	0.008	0.198
SORT1	-0.475	0.002	0.057	-0.170	0.027	0.179
DPP6	-0.233	0.046	0.071	-0.146	0.008	0.112
CALCA	-0.550	0.046	0.103	-0.308	0.019	0.061
SELPLG	-0.282	0.048	0.052	-0.160	0.020	0.037

Note. PP.H4 is the global coloc posterior probability. The eight targets were carried forward for molecular validation; target names are standardized to ADAM22 and SELPLG.

Table 3. Selected SMR/HEIDI results

Target gene	pQTL-SMR PGC2017	HEIDI	pQTL-SMR PGC2019	HEIDI	Blood eQTL-SMR PGC2017	HEIDI	Blood eQTL-SMR PGC2019	HEIDI
NCAM1	0.2256	0.432	0.00627*	0.00521†	0.0692	0.408	0.3838	0.638
ADAM22	0.1578	0.309	0.1381	0.57	0.6768	0.843	0.0339*	0.413
WFIKKN1	0.2683	0.878	0.000760*	0.521	0.6435	0.504	0.9436	0.046
CRIM1	0.0220*	0.255	0.00725*	0.067	0.00894*	0.663	0.0498*	0.311
SORT1	0.0652	0.35	0.00861*	0.328	0.4248	0.631	0.5031	0.629
DPP6	0.0599	0.314	0.00796*	0.891	Not available	—	Not available	—
CALCA	0.0314*	0.603	0.0298*	0.349	Not available	—	Not available	—
SELPLG	0.9307	0.692	0.9504	0.433	0.3629	0.206	0.9738	0.988

Note. Summary of pQTL-based SMR and blood eQTL-based SMR analyses for the core targets. SMR analyses were conducted against the PGC2017 and PGC2019 AN GWAS datasets. The table reports nominal SMR P values and corresponding HEIDI P values. Asterisks indicate nominally significant SMR associations ($P_{SMR} < 0.05$), while the dagger indicates a HEIDI failure according to the post-processing criterion ($P_{HEIDI} \leq 0.01$). CALCA and DPP6 were not evaluated in the blood eQTL-SMR analysis because no qualifying genome-wide significant blood eQTL instruments were available. None of the blood eQTL-SMR associations survived global FDR correction at 0.05.

Table 4. GTEx v10 brain eQTL MR associations at $P<5\times10^{-8}$

Gene	Tissue	Outcome	β (SE)	OR (95% CI)	P value	FDR
ADAM22	Cerebellum	PGC2017	0.302 (0.162)	1.353 (0.985-1.856)	0.06164	0.2219
ADAM22	Cerebellum	PGC2019	0.026 (0.076)	1.026 (0.885-1.190)	0.732	0.9412
CRIM1	Cerebellar Hemisphere	PGC2017	-0.160 (0.062)	0.852 (0.755-0.962)	0.009579	0.08621
CRIM1	Cerebellar Hemisphere	PGC2019	-0.059 (0.028)	0.942 (0.892-0.996)	0.03629	0.2177
CRIM1	Cerebellum	PGC2017	-0.221 (0.083)	0.802 (0.681-0.943)	0.00771	0.08621
CRIM1	Cerebellum	PGC2019	-0.073 (0.039)	0.930 (0.862-1.003)	0.05899	0.2219
DPP6	Cerebellar Hemisphere	PGC2017	-0.099 (0.122)	0.905 (0.713-1.150)	0.415	0.8901

Gene	Tissue	Outcome	β (SE)	OR (95% CI)	P value	FDR
DPP6	Cerebellar Hemisphere	PGC2019	-0.045 (0.059)	0.956 (0.850-1.074)	0.4451	0.8901
DPP6	Cerebellum	PGC2017	-0.055 (0.068)	0.946 (0.828-1.081)	0.415	0.8901
DPP6	Cerebellum	PGC2019	-0.025 (0.033)	0.975 (0.914-1.040)	0.4451	0.8901
WFIKKN1	Cerebellar Hemisphere	PGC2017	0.027 (0.062)	1.027 (0.910-1.159)	0.6631	0.9181
WFIKKN1	Cerebellar Hemisphere	PGC2019	0.000 (0.035)	1.000 (0.935-1.070)	0.9959	0.9959
WFIKKN1	Cerebellum	PGC2017	0.021 (0.048)	1.021 (0.929-1.123)	0.6631	0.9181
WFIKKN1	Cerebellum	PGC2019	0.000 (0.027)	1.000 (0.948-1.055)	0.9959	0.9959
WFIKKN1	Hypothalamus	PGC2017	0.020 (0.045)	1.020 (0.933-1.114)	0.6631	0.9181
WFIKKN1	Hypothalamus	PGC2019	0.000 (0.025)	1.000 (0.952-1.051)	0.9959	0.9959
WFIKKN1	Nucleus accumbens basal ganglia	PGC2017	0.022 (0.039)	1.022 (0.946-1.104)	0.5785	0.9181
WFIKKN1	Nucleus accumbens basal ganglia	PGC2019	0.001 (0.022)	1.001 (0.959-1.046)	0.9476	0.9959

Note.β is the change in AN log-odds per 1-SD increase in genetically predicted tissue expression. OR, odds ratio; CI, confidence interval; FDR, Benjamini–Hochberg adjusted P value across all 18 standard-threshold MR rows. Each row was estimated with one SNP (Wald ratio).

Table 5. Druggability component scores and translational interpretation of core targets

Target	Genetic evidence	Tractability	Clinical/pathway precedent	Pleiotropy penalty	Final score
NCAM1	3	2	0	-1	4
WFIKKN1	3	2	0	-1	4
ADAM22	3	2	0	-1	4
CRIM1	2	2	0	-1	3
DPP6	2	2	0	0	4
SORT1	2	2	0	-3	1
CALCA	1	2	3	0	6
SELPLG	1	2	0	0	3

Note. Scores are shown as the semi-quantitative internal rubric: final score = genetic evidence + tractability + clinical/pathway precedent + pleiotropy penalty. Full target annotations, modality notes, safety considerations and source notes are provided in Additional file 1: Tables S19–S20 and S31–S34; the score does not imply clinical efficacy or treatment recommendation.

Additional file 1: Table S1. Mendelian randomization estimates for genetically predicted plasma protein levels and anorexia nervosa risk in the PGC2017 GWAS across multiple MR methods.

Additional file 1: Table S2. Mendelian randomization estimates for genetically predicted plasma protein levels and anorexia nervosa risk in the PGC2019 GWAS across multiple MR methods.

Additional file 1: Table S3. Odds ratios and 95% confidence intervals derived from pQTL MR analyses of plasma proteins and anorexia nervosa in PGC2017.

Additional file 1: Table S4. Odds ratios and 95% confidence intervals derived from pQTL MR analyses of plasma proteins and anorexia nervosa in PGC2019.

Additional file 1: Table S5. Cochran’s Q heterogeneity statistics for pQTL MR instruments across protein targets in PGC2017.

Additional file 1: Table S6. Cochran’s Q heterogeneity statistics for pQTL MR instruments across protein targets in PGC2019.

Additional file 1: Table S7. Instrument-selection and analysis-completion logs for the PGC2017 pQTL MR pipeline, including filtering, F-statistic screening, clumping, harmonization, and retained SNP counts.

Additional file 1: Table S8. Instrument-selection and analysis-completion logs for the PGC2019 pQTL MR pipeline, including filtering, F-statistic screening, clumping, harmonization, and retained SNP counts.

Additional file 1: Table S9. MR-Egger intercept tests assessing directional pleiotropy in PGC2017 pQTL MR analyses.

Additional file 1: Table S10. MR-Egger intercept tests assessing directional pleiotropy in PGC2019 pQTL MR analyses.

Additional file 1: Table S11. Plasma pQTL instrument SNPs retained for PGC2017 analyses, including genomic coordinates, allele information, effect estimates, allele frequencies, and sample sizes.

Additional file 1: Table S12. Plasma pQTL instrument SNPs retained for PGC2019 analyses, including genomic coordinates, allele information, effect estimates, allele frequencies, and sample sizes.

Additional file 1: Table S13. Tiered prioritization of core pQTL targets integrating cross-dataset MR replication, IVW estimates, FDR values, instrument numbers, and pQTL colocalization posterior probabilities.

Additional file 1: Table S14. pQTL-based SMR and HEIDI results for selected targets in PGC2017 and PGC2019, including lead variants, effect estimates, multiple-testing measures, and evidence classifications.

Additional file 1: Table S15. Blood eQTL MR estimates derived from eQTLGen instruments for candidate genes and anorexia nervosa outcomes, including effect sizes, odds ratios, instrument strength, method status, and FDR values.

Additional file 1: Table S16. Blood eQTL SMR and HEIDI results for candidate genes and anorexia nervosa outcomes, including lead variants, effect estimates, FDR values, and evidence classifications.

Additional file 1: Table S17. GTEx v10 brain-tissue eQTL MR results for candidate genes and anorexia nervosa outcomes using primary and exploratory instrument thresholds.

Additional file 1: Table S18. GTEx v10 brain-tissue eQTL – anorexia nervosa colocalization results across gene – tissue – outcome combinations, including posterior probabilities and PP.H4-based interpretations.

Additional file 1: Table S19. Database-based druggability annotations of prioritized targets covering genetic evidence, pleiotropy, protein localization, tissue relevance, modality-specific tractability, clinical precedence, safety, and repurposing potential.

Additional file 1: Table S20. Semi-quantitative druggability scores for prioritized targets combining genetic evidence, tractability, clinical or pathway precedence, and pleiotropy penalties.

Additional file 1: Table S21. Functional-enrichment results for the eight core targets across Gene Ontology biological process, Gene Ontology molecular function, and WikiPathways terms.

Additional file 1: Table S22. Steiger directionality results for core pQTL – anorexia nervosa associations based on variance explained in the protein exposure and disease outcome.

Additional file 1: Table S23. Blind molecular-docking scores for ligands targeting the CGRP receptor, SORT1, and PSGL-1 axes reported as AutoDock Vina binding energies.

Additional file 1: Table S24. Summary of exact-instrument GWAS Catalog associations for core targets, including

trait counts, relevance categories, pleiotropy scores, and qualitative risk ratings.

Additional file 1: Table S25. Individual GWAS Catalog associations for selected pQTL instrument SNPs with genomic, study, trait, and pleiotropy annotations.

Additional file 1: Table S26. PhenoScanner query status for selected pQTL instruments indicating that no annotated rows were returned and that the query should be rerun.

Additional file 1: Table S27. Single-SNP Wald-ratio estimates for pQTL instruments and anorexia nervosa risk in PGC2017 together with combined MR summary results.

Additional file 1: Table S28. Single-SNP Wald-ratio estimates for pQTL instruments and anorexia nervosa risk in PGC2019 together with combined MR summary results.

Additional file 1: Table S29. Harmonized pQTL – anorexia nervosa summary-statistic records for PGC2017 after allele alignment, palindromic-variant assessment, and MR retention filtering.

Additional file 1: Table S30. Harmonized pQTL – anorexia nervosa summary-statistic records for PGC2019 after allele alignment, palindromic-variant assessment, and MR retention filtering.

Additional file 1: Table S31. Public databases and protein resources used to annotate target tractability, biological function, drug interactions, and safety considerations.

Additional file 1: Table S32. Druggability interpretation notes and reporting definitions

Additional file 1: Table S33. README metadata defining the purpose, target scope, interpretation, and cautions for the druggability annotation tables.

Additional file 1: Table S34. Data sources, dataset releases, sample sizes, tissue scopes, and analytical roles for the resources used in this study.

Additional file 1: Table S35. Blood eQTL – anorexia nervosa colocalization results for candidate genes, including overlapping SNP counts, posterior probabilities, lead variants, and PP.H4-based evidence.

Figure legends

Figure 1. Expanded multi-layer target-prioritization workflow. Plasma cis-pQTLs and the two AN GWAS datasets are processed through instrument selection, pQTL MR and Bayesian colocalization, followed by core-target sensitivity analyses, blood and brain eQTL validation, functional annotation, druggability scoring and exploratory docking. PGC2017 and PGC2019 are used to assess cross-dataset consistency because sample overlap is possible.

Figure 2. Integrated protein-level evidence for the eight core targets. The first two columns show −log10 nominal IVW P values and the last two columns show global PP.H4. Global PP.H4 and SNP-level posterior probabilities should not be conflated.

Figure 3. Representative pQTL colocalization loci. LocusCompare panels show three higher-PP.H4 examples (NCAM1 in PGC2017, ADAM22 in PGC2019 and WFIKKN1 in PGC2019) and one lower-PP.H4 contrast locus (SORT1 in PGC2019). Global PP.H4 is reported in the panel labels and should not be conflated with SNP-level posterior probabilities.

Figure 4. Cerebellar localization of the CRIM1 signal in GTEx v10. (A) Lateral scientific brain illustration with colored overlays indicating the two separately analyzed GTEx tissues in which CRIM1 instruments were available: the cerebellum and the cerebellar hemisphere, shown as the lateral subdivision of the cerebellum. The overlays provide anatomical localization and are not MRI segmentations. (B) One-SNP Wald-ratio MR estimates for genetically predicted CRIM1 expression in these tissues and the PGC2017 and PGC2019 anorexia nervosa GWAS outcomes. Circles denote PGC2017 and squares denote PGC2019. Points show odds ratios and horizontal lines show 95% confidence intervals; β, OR, 95% CI and P values are printed beside each estimate. Bold estimates indicate nominal $P<0.05$. $OR<1$ indicates lower anorexia nervosa odds per 1-SD higher genetically predicted CRIM1 expression.

Figure 5. GWAS Catalog pleiotropy among core pQTL instruments. Total GWAS Catalog associations are shown in light blue, with metabolic/lipid associations highlighted in dark blue. SORT1 showed a marked enrichment of lipid and metabolic associations compared with the other core targets.

Figure 6. Integrated functional and structural context for the eight core targets. (A) Exploratory enrichment plot showing the 13 WikiPathways terms for the eight-target set against a 1,031-target universe. Bars are ranked by nominal P value, and colors distinguish terms with source-specific FDR < 0.10 from top-ranked terms that did not meet this threshold. No term passed global Benjamini–Hochberg FDR < 0.05; therefore, this panel is intended as biological context rather than confirmatory pathway evidence. (B) STRING functional-association network containing 18 proteins and 20 unique edges after

deduplication. Text-mining-supported edges do not establish direct physical binding. (C) Exploratory AutoDock Vina scores for 12 docking records representing 12 unique compounds across the CGRP receptor, SORT1 and PSGL-1 axes. More-negative scores indicate better predicted poses within the same computational setup, but do not establish experimental affinity, efficacy, selectivity or target engagement. Enrichment and docking data are provided in Additional file 1: Tables S21 and S23.

Figure 7. Representative exploratory docking poses. The panels show CB-Dock2/AutoDock Vina pose illustrations for the CGRP receptor axis, SORT1/sortilin axis and PSGL-1/P-selectin axis. Pocket-associated contacts are computational predictions and do not establish experimental binding or target engagement.

Figure 1. Overall study design and multi-omic framework for prioritizing potential therapeutic targets for anorexia nervosa

PRIMARY DISCOVERY FRAMEWORK

pQTL-driven causal inference and genetic validation

1. Input data sources

Plasma pQTL resource

38 cohorts; 78,664 participants; >24,000 pQTLs for 1,116 circulating proteins

AN GWAS 2019

16,992 cases; 55,525 controls

AN GWAS 2017

3,495 cases; 10,982 controls

2. Instrument selection

- *cis*-pQTLs (±1 Mb)
- LD clumping and harmonization
- Palindromic SNP handling
- Instrument strength assessment

3. Primary causal analysis and core genetic validation

Drug-target Mendelian randomization

Sensitivity analyses

- Heterogeneity test
- Horizontal pleiotropy assessment
- Leave-one-out analysis
- Steiger directionality test

Cross-dataset comparison

Replication and consistency assessment across AN 2019 and AN 2017

Core genetic validation

- Bayesian colocalization
- SMR / HEIDI

SUPPORTIVE TRANSCRIPTOMIC VALIDATION

Supplementary validation analyses

Blood eQTL replication

- MR
- Colocalization
- SMR / HEIDI

GTEx v10 brain-tissue validation

- MR across brain tissues
- Colocalization

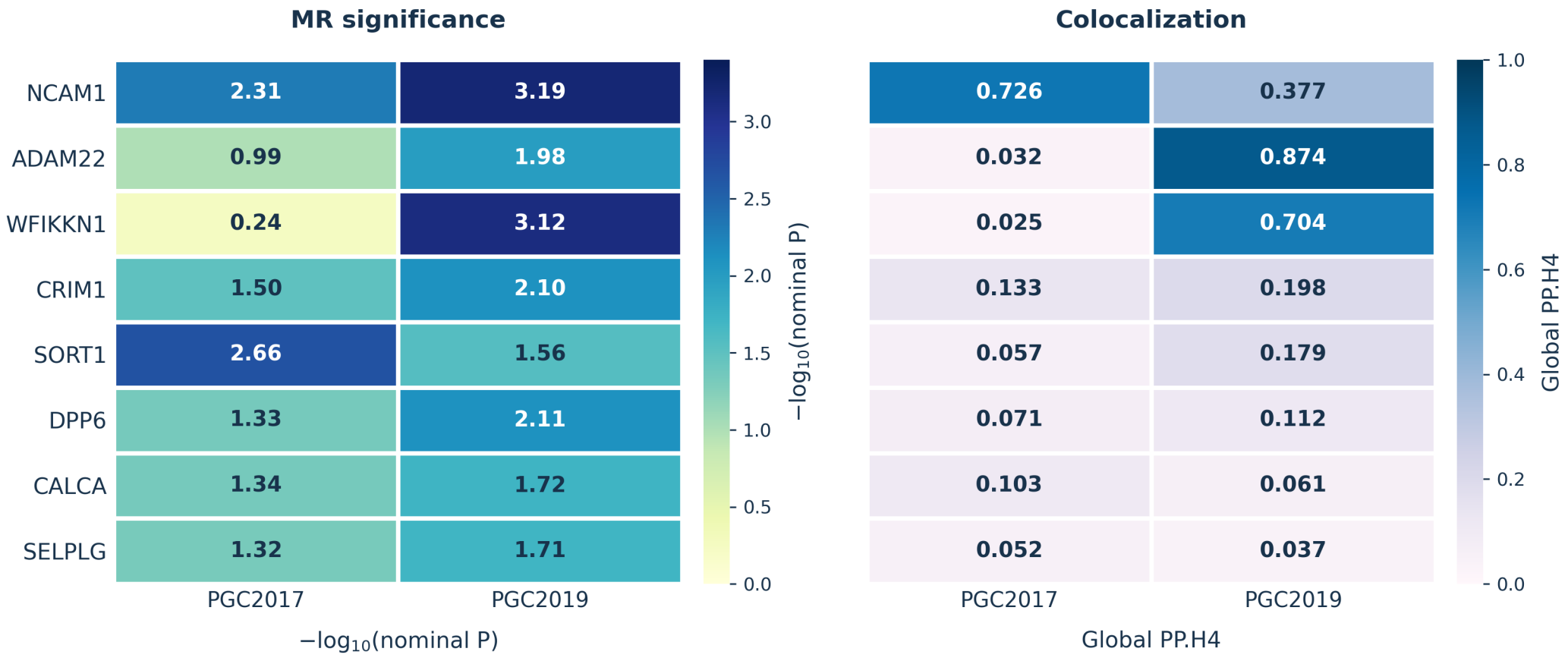
FUNCTIONAL AND TRANSLATIONAL CHARACTERIZATION

- **Functional enrichment analysis**
- **Protein–protein interaction (PPI) network**
- **Druggability assessment**
- **Molecular docking**

Prioritized candidate therapeutic targets for anorexia nervosa

Targets supported by convergent proteomic, transcriptomic, tissue-specific, functional, and structural evidence

Protein-level evidence across the two AN GWAS datasets

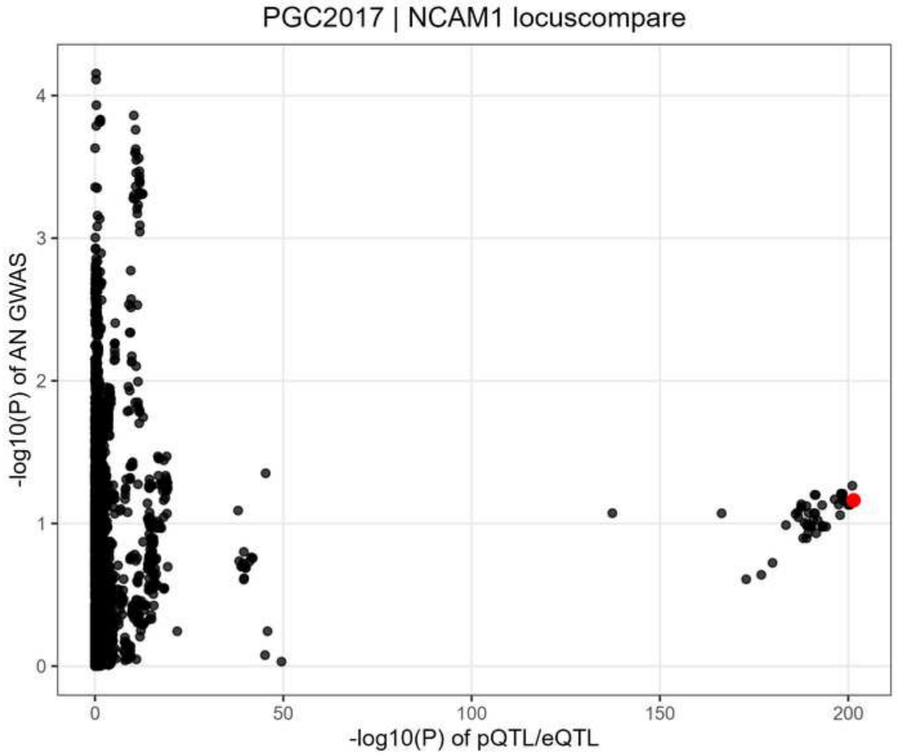


Values are shown in each cell. Global PP.H4 denotes the coloc posterior probability of a shared causal variant.

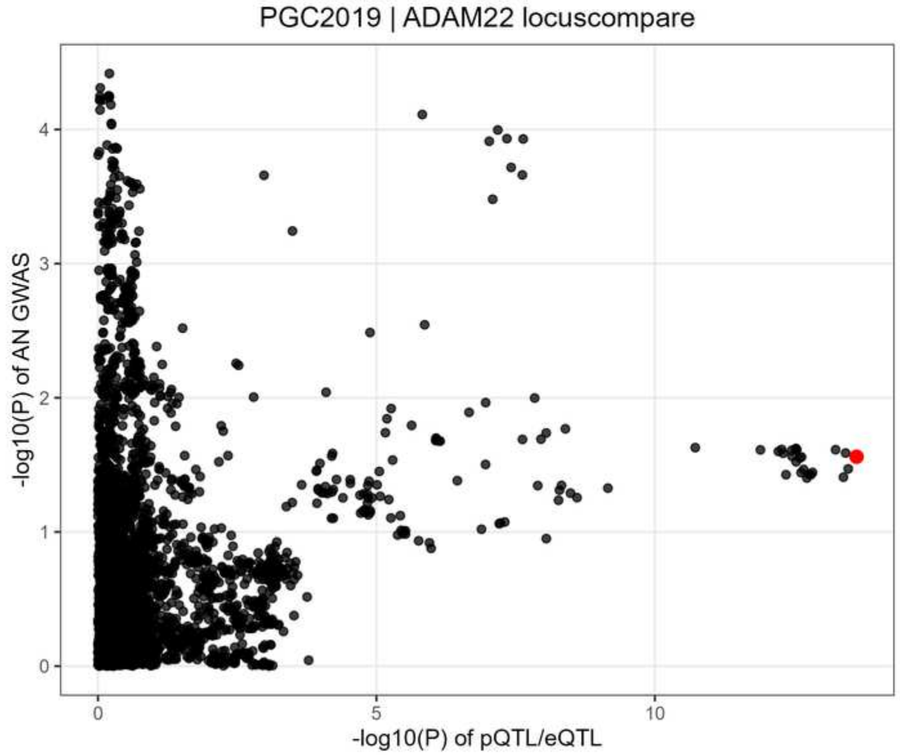
Representative pQTL colocalization loci

LocusCompare displays are shown for three prioritized shared-signal candidates and one lower-PP.H4 contrast locus.

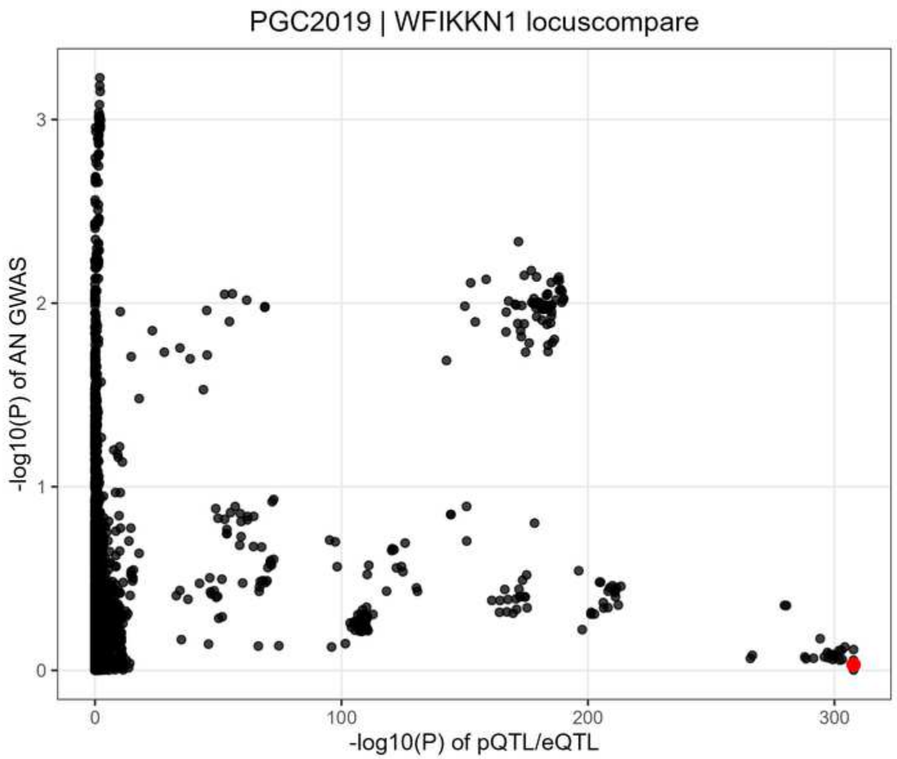
A NCAM1 | PGC2017 | PP.H4 = 0.726



B ADAM22 | PGC2019 | PP.H4 = 0.874



C WFIKKN1 | PGC2019 | PP.H4 = 0.704



D SORT1 | PGC2019 | PP.H4 = 0.179 (contrast)

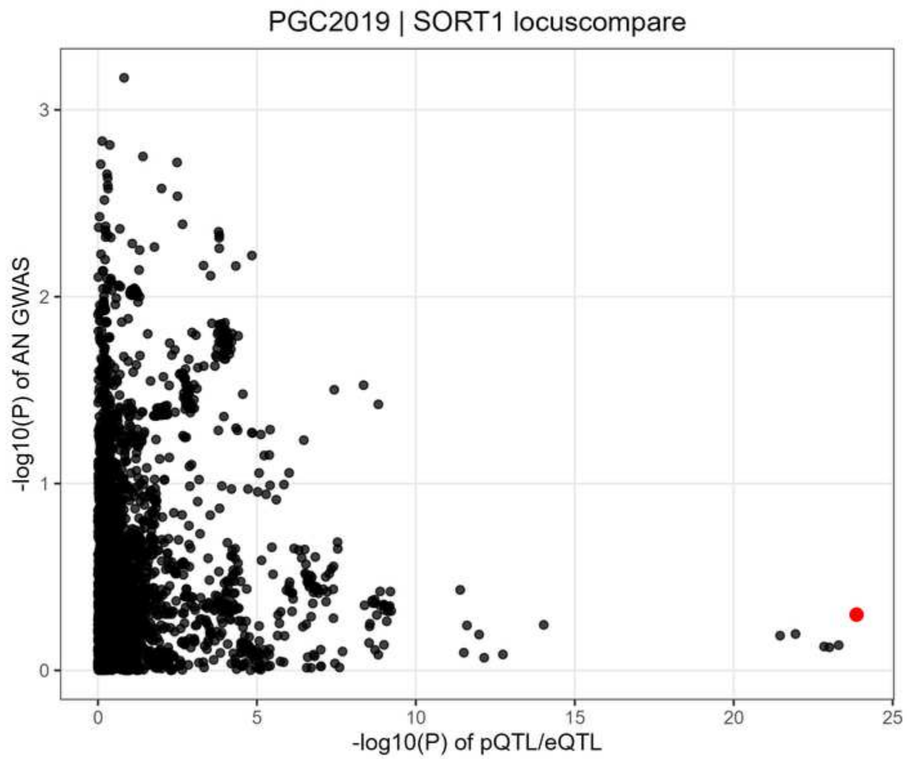
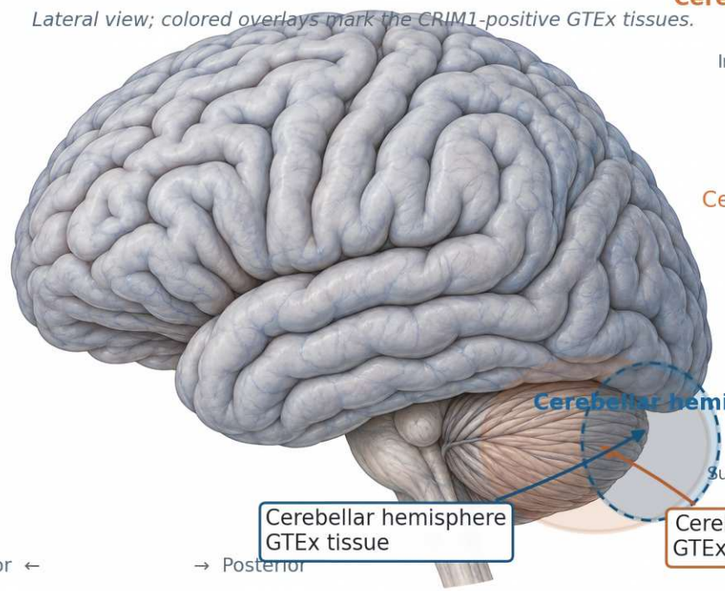


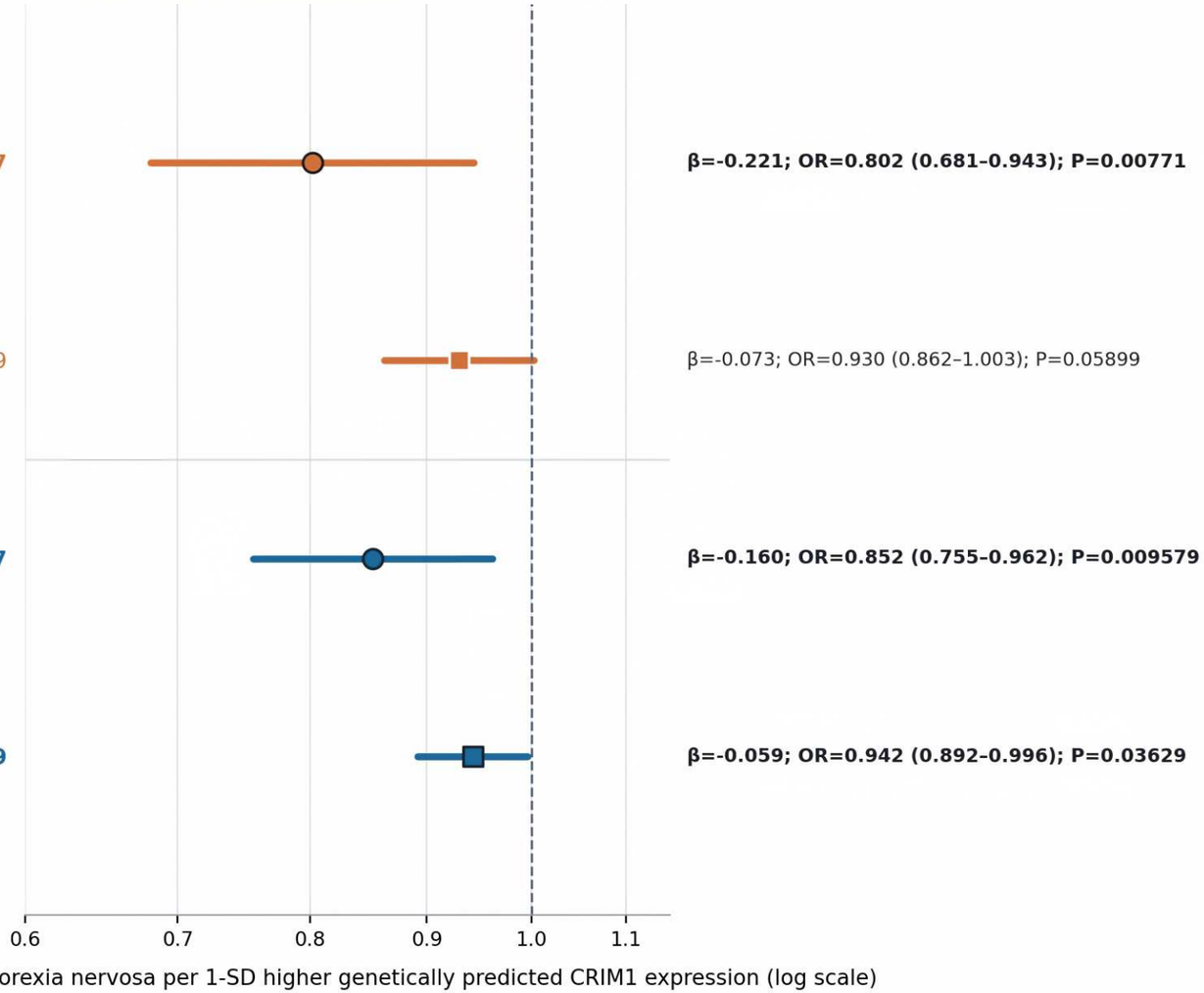
Figure 4. Cerebellar localization of the CRiM1 signal in GTEx v10

A Anatomical localization

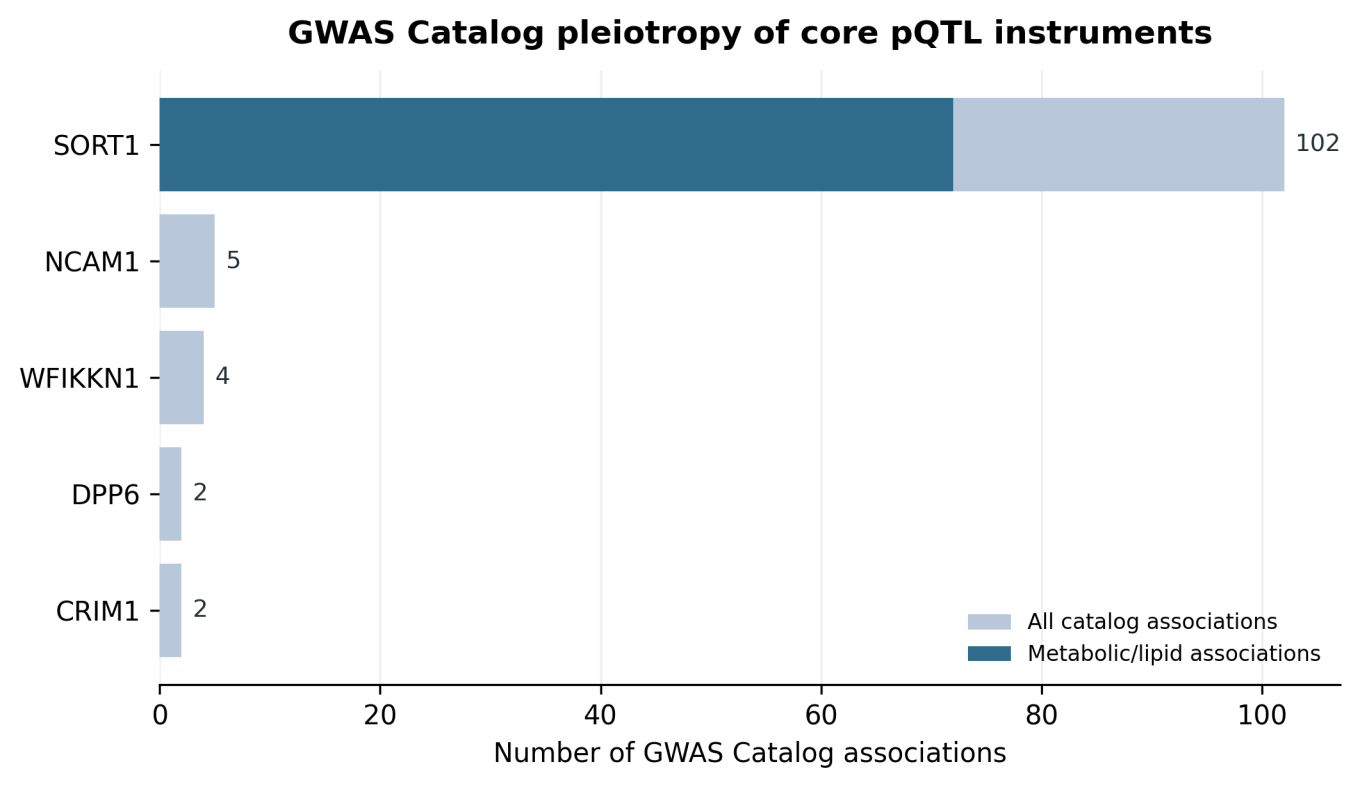
Lateral view; colored overlays mark the CRIM1-positive GTEx tissues.



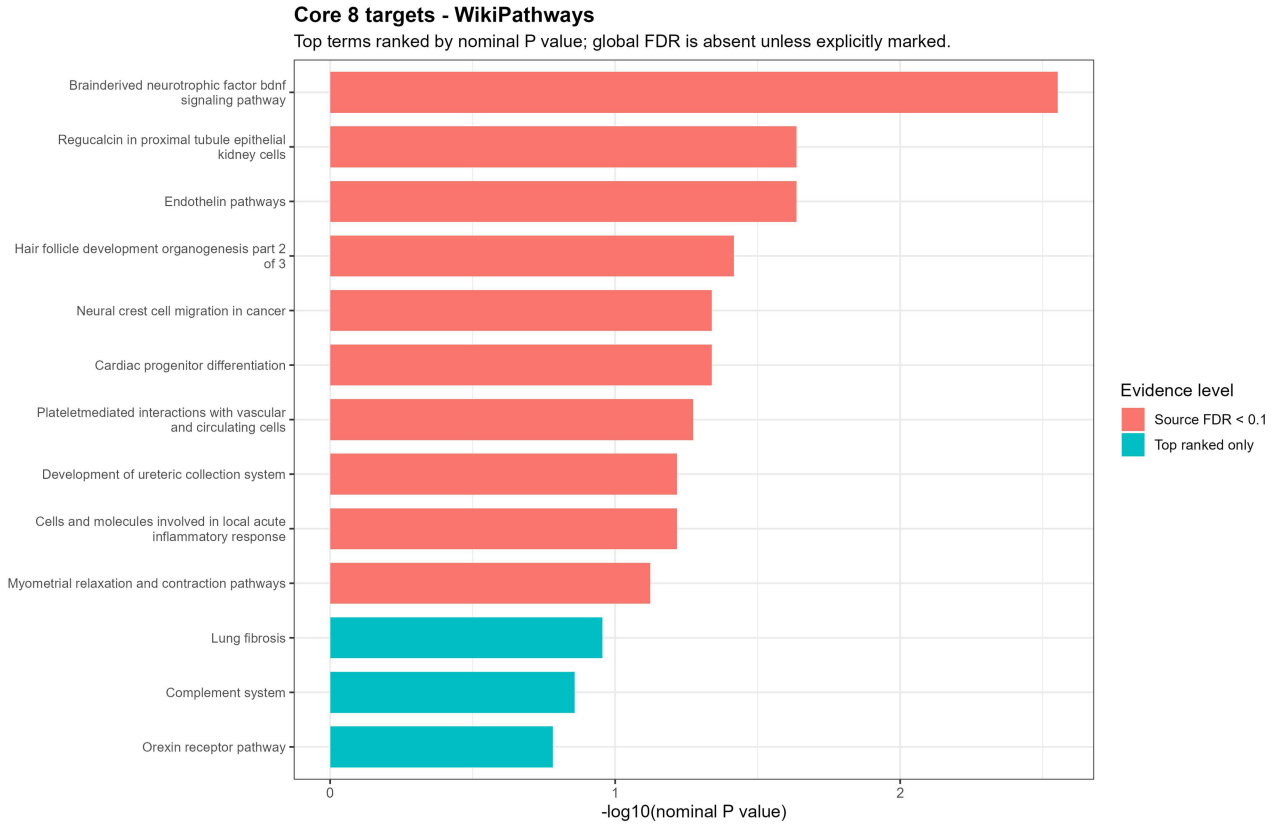
B CRIM1 MR estimates



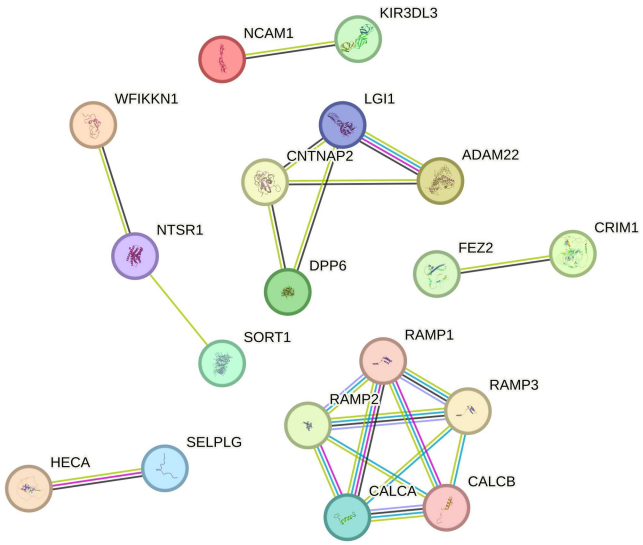
The anatomical render is a lateral scientific illustration; colored overlays provide approximate localization of the two separately analyzed GTEx tissue labels and are not MRI segmentations. Points and lines show odds ratios and 95% confidence intervals. Each estimate used one independent SNP (Wald ratio).


GWAS Catalog pleiotropy of core pQTL instruments
SORT1
NCAM1
WFIKKN1
DPP6
CRIM1
102
5
4
2
2
All catalog associations
Metabolic/lipid associations
0
20
40
60
80
100
Number of GWAS Catalog associations

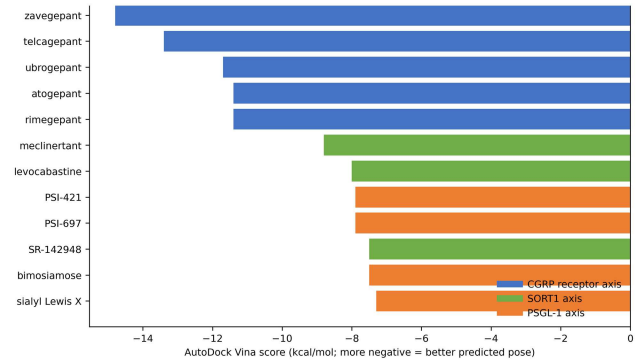
A



B STRING functional association network



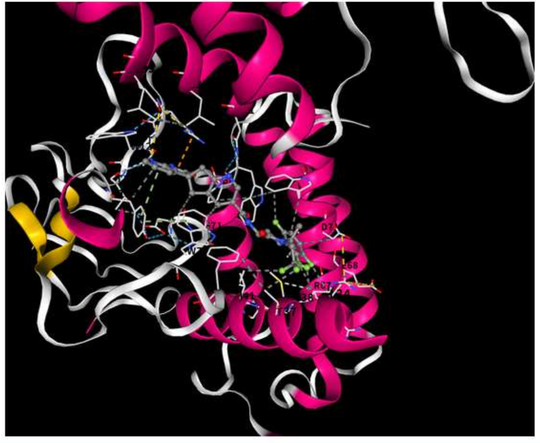
C Exploratory molecular docking scores



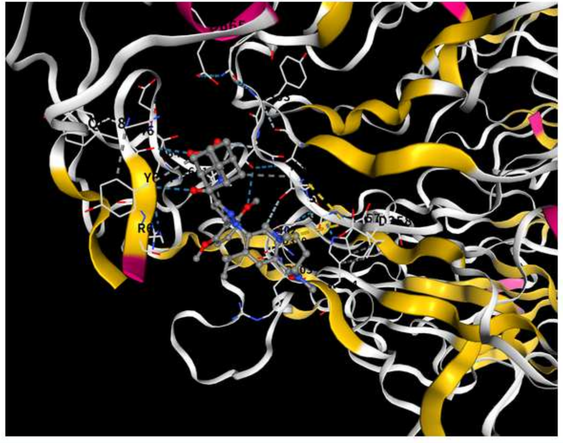
Representative exploratory docking poses

CB-Dock2/AutoDock Vina poses are pathway-axis illustrations; pocket-associated contacts are not experimental binding evidence.

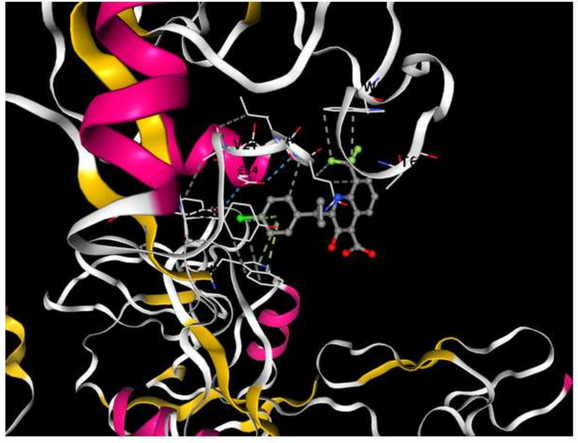
A CGRP receptor axis | atogepant | PDB 3N7S



B SORT1/sortilin axis | SR-142948 | PDB 4PO7



C PSGL-1/P-selectin axis | PSI-421 | PDB 1G1S



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- STROBEMRchecklist.pdf
- Additionalfile1.xlsx