

Structured Metabolic Genetic Axes Reveal Distinct Lipid and Ketone-Body Programs in Alzheimer's and Parkinson's Disease

Jing Shen

The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University

Yuxuan Shen

The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University

Gang Yu

The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University

Lijun Pang

The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University

Chenxu Xiao

xcx5443@gmail.com

The Affiliated Jiangsu Shengze Hospital of Nanjing Medical University

Article

Keywords: metabolomics, lipid metabolism, ketone bodies, Alzheimer disease, Parkinson disease, Genomic SEM, single-cell genomics

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11247856/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Metabolic dysregulation has been implicated in Alzheimer disease (AD) and Parkinson disease (PD), yet circulating metabolites are embedded within highly correlated metabolic networks. This structure complicates single-metabolite analyses and motivates genetic approaches that model shared metabolic architecture directly.

Methods

We applied Genomic SEM to NMR metabolite GWAS summary statistics to construct lipid and non-lipid latent metabolic factor GWAS. Six factors were screened against AD, PD, and Lewy body dementia using cross-trait LDSC, followed by MiXeR, pleiotropy-informed locus discovery, colocalization, transcriptomic prioritization, single-cell projection, virtual knockout, and target annotation.

Results

Among 18 factor-disease comparisons, three survived FDR correction: an HDL-core axis positively correlated with AD ($r_g = 0.142$, $FDR = 0.002$), a ketone-body axis inversely correlated with PD ($r_g = -0.106$, $FDR = 0.015$), and a TG-rich/VLDL-remodeling axis inversely correlated with PD ($r_g = -0.072$, $FDR = 0.031$). Component-level LDSC supported the biological traceability of these axes. Downstream analyses linked HDL-core/AD to perivascular support, ketone-body/PD to distributed neuronal-glia energy adaptation, and TG/VLDL-remodeling/PD to lysosomal and membrane-trafficking programs. Target annotation prioritized experimentally approachable genes including *GRK4*, *PRMT7*, and *TMEM175*.

Conclusions

Structured metabolic genetic axes reveal selective and biologically distinct links between circulating metabolism and neurodegenerative disease, providing a framework for mechanism-focused follow-up beyond single-metabolite association.

Introduction

Neurological disorders are a leading contributor to global disease burden, and age-related neurodegenerative diseases account for an increasing share of this burden as populations age [1–2]. Alzheimer disease (AD) and Parkinson disease (PD), the two most common neurodegenerative disorders, remain clinically and pathologically anchored by protein aggregation, neuronal loss, and progressive functional decline [3–5]. However, genetic, cellular, and molecular studies increasingly indicate that these diseases arise from interacting biological systems extending beyond isolated

proteinopathies. AD is defined by amyloid-beta plaques and tau-containing neurofibrillary tangles, yet its clinical course is shaped by synaptic failure, endosomal-lysosomal dysfunction, innate immune activation, vascular injury, and genetic background [6–9]. PD is anchored by dopaminergic neurodegeneration and alpha-synuclein pathology, but its genetic architecture also implicates lysosomal biology, mitochondrial function, membrane trafficking, lipid handling, and inflammatory regulation [10–11]. In parallel, epidemiological and metabolomic studies have repeatedly linked neurodegenerative disease to circulating metabolic differences, but these associations can be difficult to interpret because measured metabolites are influenced by age, diet, medication, disease stage, comorbidity, and reverse causation [12–14]. This creates a need for genetic approaches that can test whether systemic metabolic variation aligns with disease architecture in neurodegeneration and whether such alignments can be traced to brain cell contexts and tractable mechanisms.

Lipid metabolism is one of the most prominent examples of this broader disease model. In AD genetics, lipid processing and innate immunity are repeatedly implicated alongside amyloid and tau biology, with apolipoprotein E (APOE) remaining the strongest common genetic determinant of late-onset disease risk [8, 15]. Recent single-cell and lipidomic studies have strengthened this connection by showing that APOE4 can promote lipid-droplet accumulation in disease-associated microglia, while cell-resolved lipid atlases increasingly relate neurodegenerative pathology to specific lipid classes and cellular states [15–16]. In PD, lipid biology intersects with alpha-synuclein membrane binding, lysosomal storage pathways, sphingolipid metabolism, mitochondrial stress, and ferroptosis-related mechanisms [11, 17]. These findings support a lipid-neurodegeneration framework, but they also expose an analytical challenge: lipid traits do not represent a single exposure. High-density lipoprotein (HDL) particles, apolipoproteins, triglyceride-rich remnants, cholesteryl esters, free cholesterol, fatty acids, and lipoprotein particle size capture distinct biological programs and may align with neurodegenerative disease in different directions.

Non-lipid metabolism adds a further layer of heterogeneity. Metabolomic studies in AD and PD have reported alterations in amino-acid metabolism, glucose utilization, mitochondrial and tricarboxylic-acid-cycle-related pathways, purine and xanthine metabolism, microbiome-linked metabolites, and ketone-body biology [18–21]. Ketone bodies such as beta-hydroxybutyrate and acetoacetate are particularly relevant because they are not merely alternative fuels; they also participate in oxidative-stress responses, inflammatory signaling, protein quality control, and metabolic adaptation under impaired glucose utilization [22–25]. Yet metabolomic findings in neurodegeneration often vary across platforms, biospecimens, medication states, disease stages, and analytic designs. Thus, both lipid and non-lipid evidence point to the same conceptual challenge: metabolism is relevant to neurodegeneration, but individual metabolites often do not provide a stable or mechanistically complete unit of analysis.

Large-scale metabolite genome-wide association studies (GWAS) provide an opportunity to address this challenge genetically. Nuclear magnetic resonance (NMR)-based biobank studies have mapped common and rare genetic determinants of circulating metabolic traits, confirming that metabolite GWAS recover biologically plausible metabolic genes while revealing extensive pleiotropy and correlation across

metabolite classes [26–28]. This creates both opportunity and complexity. Genetically informed metabolite studies and Mendelian-randomization analyses have begun to nominate blood metabolites related to AD or PD risk [8, 26]. However, testing hundreds of correlated metabolites individually can fragment shared metabolic programs into partially redundant associations, obscure directionally opposing subcomponents, and make downstream gene or target prioritization unstable. Genomic structural equation modelling (Genomic SEM) offers a principled approach for modeling this covariance structure directly from GWAS summary statistics, estimating latent genetic factors that summarize shared genetic variation across related traits, generating factor-level GWAS, and evaluating SNP-level heterogeneity through the SNP-level heterogeneity statistic Q_{SNP} [29].

The unresolved question is therefore not simply whether metabolism and neurodegeneration are associated. Rather, it is whether correlated metabolic traits can be reorganized into genetically interpretable latent axes, whether those axes show disease-specific alignment with AD, PD, or Lewy body dementia (LBD), and whether the resulting relationships can be traced to biologically coherent loci, genes, cell contexts, and follow-up targets. Here, we used Genomic SEM to derive lipid and non-lipid metabolic factor GWAS from circulating metabolite summary statistics and integrated genome-wide overlap, locus-level refinement, transcriptomic prioritization, single-cell contextualization, virtual perturbation, and target annotation. This design reframed metabolite-disease analysis from a one-metabolite-at-a-time association screen into a structured genetic framework for defining disease-aligned metabolic programs with distinct biological and translational implications.

Methods

Study design and data sources

We designed this study to test whether genetically structured metabolic programs show disease-specific alignment with major neurodegenerative disorders and whether these alignments can be followed from latent metabolic factors to loci, genes, brain cell contexts, perturbational signatures, and target-tractability annotations. The analysis proceeded in four linked stages: metabolic factor construction, factor GWAS generation and validation, cross-disease genetic-overlap screening, and downstream locus-to-target interpretation.

Insert Figure 1 near here

The primary metabolic input was the $m\eta_EUR$ summary-statistic release from a large NMR metabolite GWAS of 249 circulating metabolic traits measured with the Nightingale Health platform in the Estonian Biobank and UK Biobank [27,30]. NMR refers to nuclear magnetic resonance spectroscopy, the high-throughput assay used to quantify circulating lipids, lipoprotein subclasses, fatty acids, amino acids, ketone bodies, glycolysis-related metabolites, and related biomarkers. The $m\eta_EUR$ release meta-analyzed European-ancestry Estonian Biobank participants and the European-ancestry UK Biobank subset, with a reported combined sample size of 599,249 for the meta-analysis resource. External

metabolic validation used an independent NMR metabolite GWAS by Karjalainen et al., which analyzed 233 circulating metabolic biomarkers in up to 136,016 participants from 33 cohorts [26]. This validation resource measured a highly overlapping set of NMR metabolic biomarkers but was not identical to the discovery resource; validation therefore relied on curated exact matches, terminology variants, and same-domain proxies without assuming one-to-one identity for every phenotype (Supplementary Methods).

Neurodegenerative disease comparisons used European-ancestry GWAS summary statistics for AD from the European Alzheimer & Dementia Biobank stage-I analysis (39,106 clinically diagnosed AD cases, 46,828 proxy-AD cases, and 401,577 controls) [8], PD from the European-ancestry comparison file derived from a multi-ancestry PD meta-analysis (63,555 cases, 17,700 proxy cases, and 1,746,386 controls) [10], and LBD from a genome-sequencing association study of European-ancestry participants (2,591 cases and 4,027 controls) [31]. Single-cell, eQTL, target-annotation, and public disease-state summary resources are described below and in the Supplementary Methods.

Summary-statistic harmonization and metabolic factor construction

All GWAS summary statistics underwent standard harmonization and quality control before downstream analyses, including allele and genome-build alignment, removal of invalid or duplicated variants, and linkage disequilibrium score regression (LDSC)-compatible filtering where required (Supplementary Methods). Metabolic traits entered multivariable modeling only if they showed sufficient single nucleotide polymorphism (SNP) heritability signal in univariate LDSC ($h^2_Z > 4$; Supplementary Methods).

The metabolic modeling universe was restricted to non-proportion traits with robust univariate LDSC signal, yielding 112 traits. This universe was split into lipid-dominant and non-lipid candidate panels. Within each domain, staged reduction retained indicators that preserved metabolic breadth, avoided near-redundancy, and supported stable latent factor structure. The final retained set consisted of 8 lipid traits and 8 non-lipid traits (Supplementary Tables S1-S3). Multivariable LDSC estimated the genetic covariance and correlation structure among retained indicators before formal factor modeling [32]. Exploratory factor analysis and exploratory structural equation modeling were used during model selection, whereas final retained structures were evaluated with Genomic SEM [29]. The lipid panel yielded a TG-rich/VLDL-remodeling axis, an HDL-core axis, and a cholesteryl-ester/structural lipid axis. The non-lipid panel yielded a ketone-body axis, an amino-acid axis, and an energy-bridge axis spanning acetate, glucose, and lactate (Supplementary Tables S4-S7; Supplementary Methods).

Factor GWAS, validation, and disease-overlap analyses

GWAS summary statistics for the six latent factors were generated with Genomic SEM using the final retained measurement models and harmonized SNP-level inputs [29]. Factor-level effective sample size was estimated from exported summary statistics using a SNP-level variance approximation across common variants, because latent-trait GWAS do not inherit a single univariate source sample size (Supplementary Methods; Supplementary Table S8). The final factor GWAS were re-evaluated with univariate and bivariate LDSC, including SNP counts after munging and LD-score merge, mean chi-square statistics, genomic inflation, LDSC intercepts, and intercept ratios [32] (Supplementary Tables S9-S10).

We evaluated factor validity through internal metabolic validation against retained model indicators and validation-only same-domain support traits, leave-one-out sensitivity analyses, and external metabolite GWAS validation (Supplementary Tables S11-S18). Q_SNP statistics were retained as a SNP-level heterogeneity annotation to distinguish common-factor signal from indicator-specific effects (Supplementary Tables S19-S22).

All six factor GWAS were compared with AD, PD, and LBD in a complete 6-by-3 bivariate LDSC screen [32]. False discovery rate (FDR) correction was used to account for multiple factor-disease comparisons [33]. Because LDSC genetic correlation alone does not describe the amount or directionality of shared polygenic architecture, the causal mixture model MiXeR was applied to the three factor-disease pairs prioritized from the LDSC screen [34]. MiXeR summarized shared polygenicity, effect concordance, and Dice overlap alongside LDSC estimates (Supplementary Tables S23 and S25; Supplementary Methods).

Locus, gene, single-cell, and target analyses

For each focal factor-disease branch, pleiotropy-informed conjunctive false discovery rate (conjFDR) analysis was used to identify jointly associated loci [35]. Regional follow-up was then performed with coloc and pairwise colocalization (PWCoCo) to evaluate whether prioritized regions showed support for shared local association signals or more complex conditional architectures [36-38] (Supplementary Tables S26-S29). Candidate genes were prioritized by integrating positional and regulatory mapping, Functional Mapping and Annotation (FUMA), causal transcriptome-wide association study (cTWAS), bulk-brain summary-data-based Mendelian randomization (SMR), Genotype-Tissue Expression (GTEx) brain-tissue SMR, and brain cell-type SMR [39-43] (Supplementary Tables S30-S41; Supplementary Methods).

To place the retained factor-disease branches into disease-relevant cellular contexts, focal factor GWAS were projected into disease-state single-cell or single-nucleus transcriptomic atlases using scPagwas [44]. The AD and PD atlases, case-status fields, downsampling, pathway framework, and cell-type significance procedures are described in the Supplementary Methods (Supplementary Tables S42-S47). Targeted scTenifoldKnn virtual knockout analyses were then performed in selected gene-cell combinations supported jointly by genetic prioritization and scPagwas results [45] (Supplementary Tables S48-S54). Finally, prioritized genes were annotated with Open Targets, DGIdb, Human Protein

Atlas, chemical-probe, protein-interaction, and AlphaFold-derived resources to assess target class, drug-gene evidence, binding or modality support, brain-expression context, and existing disease-knowledge links [46-50] (Supplementary Tables S55-S59). Processed public AD proteomic and PD transcriptomic summaries were queried as supportive disease-state checks [51-52] (Supplementary Tables S60-S61; Supplementary Methods).

Results

Genomic SEM derived validated lipid and non-lipid metabolic factor GWAS

Among the 249 available NMR-derived traits, 112 non-proportion traits showed adequate SNP heritability for multivariable modeling ($h^2 Z > 4$) and were retained as the shared discovery universe. These traits were separated into 94 lipid-related and 18 non-lipid traits, followed by staged refinement to compact eight-indicator panels for each branch. This refinement prioritized traits with robust heritability, interpretable metabolic coverage, limited near-redundancy, and stable factor-model behavior, with full selection details provided in the Supplementary Information. Pairwise and multivariable LDSC confirmed that the final retained traits were sufficiently correlated to support latent-factor modeling while preserving interpretable within-branch structure (Supplementary Tables S1-S5; Supplementary Information).

Formal Genomic SEM evaluation supported a three-factor solution for both branches. The lipid final-8 model showed comparative fit index (CFI) = 0.957 and standardized root mean square residual (SRMR) = 0.088, and the non-lipid final-8 model showed CFI = 0.973 and SRMR = 0.102, supporting use of these models as compact representations of shared metabolic genetic architecture while recognizing that they are not perfectly discrete biochemical entities (Fig. 2A,B; Supplementary Tables S6-S7). Based on their loading patterns and biochemical content, the six factor GWAS are reported using descriptive axis labels: lipid8_F1 as the triglyceride (TG)-rich/very-low-density lipoprotein (VLDL)-remodeling axis, lipid8_F2 as the HDL-core axis, lipid8_F3 as the cholesteryl-ester/structural lipid axis, nonlipid8_F1 as the ketone-body axis, nonlipid8_F2 as the amino-acid axis, and nonlipid8_F3 as the energy-bridge axis (Supplementary Table S2). The six exported factor GWAS showed clear polygenic signal, with approximately 1.143 million SNPs retained after munging and LD-score filtering, mean chi-square values ranging from 1.525 to 2.849, lambda GC from 1.303 to 1.775, and LDSC intercepts from 1.024 to 1.177. Mean effective sample sizes were 276,215 for the TG-rich/VLDL-remodeling axis, 532,584 for the HDL-core axis, 236,386 for the cholesteryl-ester/structural lipid axis, 101,905 for the ketone-body axis, 123,638 for the amino-acid axis, and 34,657 for the energy-bridge axis (Supplementary Tables S8-S9).

Internal LDSC validation showed that the factor GWAS recaptured their intended metabolic domains. The TG-rich/VLDL-remodeling axis correlated strongly with triglycerides in medium HDL ($r_g = 0.923$), VLDL particle diameter ($r_g = 0.946$), monounsaturated fatty acids ($r_g = 0.955$), and triglycerides in small VLDL ($r_g = 0.966$). The HDL-core axis correlated strongly with apolipoprotein A1 ($r_g = 0.948$) and HDL

cholesteryl esters ($r_g = 0.983$). The ketone-body axis correlated with acetoacetate ($r_g = 0.778$) and 3-hydroxybutyrate ($r_g = 0.985$), and the amino-acid axis correlated with leucine ($r_g = 0.973$), valine ($r_g = 0.981$), and phenylalanine ($r_g = 0.517$). Leave-one-out sensitivity supported the stability of the lipid final-8 model, and external metabolite validation confirmed that factor GWAS aligned with matched or same-domain traits in the independent Karjalainen et al. resource (Fig. 2C; Supplementary Tables S11-S18; Supplementary Information).

Genome-wide factor and heterogeneity scans were summarized for all six factors. Factor-level genome-wide significant lead loci were observed across all six GWAS. Q_SNP results varied across factors, with no genome-wide heterogeneity for the HDL-core and cholesteryl-ester/structural lipid axes but appreciable heterogeneity for several remaining axes. We therefore treated Q_SNP as an annotation layer that distinguishes factor-consistent from potentially trait-specific loci (Supplementary Tables S19-S22).

Insert Figure 2 near here

Three factor-disease branches survived the cross-disease screen

We evaluated all six metabolic factor GWAS against AD, PD, and LBD using cross-trait LDSC. Among 18 factor-disease comparisons, three survived FDR correction. The HDL-core lipid axis showed a positive genetic correlation with AD ($r_g = 0.142$, $SE = 0.037$, $P = 1.30 \times 10^{-4}$, $FDR = 0.002$). In contrast, the ketone-body axis showed an inverse genetic correlation with PD ($r_g = -0.106$, $SE = 0.034$, $P = 0.002$, $FDR = 0.015$), and the TG-rich/VLDL-remodeling axis also showed an inverse genetic correlation with PD ($r_g = -0.072$, $SE = 0.026$, $P = 0.005$, $FDR = 0.031$). The remaining 15 comparisons, including all LBD analyses, did not survive FDR correction (Fig. 2D; Supplementary Table S23).

We next used targeted component-level LDSC to relate the prioritized latent-axis signals back to their measured metabolic indicators, thereby distinguishing shared domain-level signal from component-specific contributions. For the HDL-core-AD branch, the factor association ($r_g = 0.142$, $P = 1.30 \times 10^{-4}$) was directionally concordant with ApoA1 ($r_g = 0.165$, $P = 4.11 \times 10^{-4}$), HDL cholesteryl esters ($r_g = 0.131$, $P = 1.53 \times 10^{-4}$), and additional HDL-related traits, all of which showed positive genetic correlations with AD. For the ketone-body-PD branch, the factor ($r_g = -0.106$, $P = 0.002$) and 3-hydroxybutyrate ($r_g = -0.102$, $P = 0.003$) showed concordant inverse correlations with PD, whereas acetoacetate was weaker ($r_g = -0.059$, $P = 0.067$). For the TG-rich/VLDL-remodeling-PD branch, the factor association ($r_g = -0.072$, $P = 0.005$) was directionally concordant with monounsaturated fatty acids, VLDL particle diameter, and triglycerides in small VLDL. These checks supported the interpretation that Genomic SEM reorganized correlated metabolic traits into biologically traceable domains while preserving information about which measured components contributed most strongly to each disease-aligned axis (Supplementary Tables S24 and S24b).

The cross-trait LDSC intercepts for factor-disease analyses were close to zero, ranging from -0.012 to 0.006 across all 18 comparisons, arguing against substantial inflation from sample overlap or residual cross-trait confounding. MiXeR then distinguished the overlap architecture of the three focal branches. For the HDL-core-AD branch, MiXeR estimated a positive effect correlation with a selective overlap profile ($\pi_{12} = 2.91 \times 10^{-5}$; Dice = 0.243). For the ketone-body-PD branch, the overlap fraction was smaller ($\pi_{12} = 2.50 \times 10^{-5}$; Dice = 0.148), but the estimated effect correlation was strongly negative ($\rho_{\text{beta}} = -0.881$), consistent with compact but directionally coherent inverse sharing. By contrast, the TG-rich/VLDL-remodeling-PD branch showed the broadest shared polygenic architecture ($\pi_{12} = 1.71 \times 10^{-4}$; Dice = 0.800) despite only a modest negative LDSC r_g , indicating extensive overlap in causal variant space with limited net genome-wide directional concordance (Fig. 2E; Supplementary Table S25).

Shared-locus and gene-prioritization analyses identified branch-specific candidate programs

ConjFDR analysis identified shared loci for each retained branch. At $\text{conjFDR} < 0.05$, the HDL-core-AD branch identified 5,862 SNPs across 147 independent loci, the ketone-body-PD branch identified 2,405 SNPs across 50 independent loci, and the TG-rich/VLDL-remodeling-PD branch identified 7,545 SNPs across 143 independent loci (Fig. 3A,B; Supplementary Tables S26-S27). Coloc analysis first identified a stringent subset of regions with strong single-signal sharing evidence, including five regions for HDL-core-AD, four for ketone-body-PD, and four for TG-rich/VLDL-remodeling-PD at $\text{PP.H4} > 0.8$. Because many pleiotropic loci showed more complex local architecture, PWCoCo was then used as a conditional follow-up layer; the integrated coloc/PWCoCo framework retained 65 prioritized regions across the three focal branches (Fig. 3C; Supplementary Tables S28-S30; Supplementary Information).

Integrated gene prioritization reduced broad locus signals to candidate sets supported by convergent evidence. This was not a strict intersection across all methods; rather, genes were tiered by the breadth and consistency of support from positional/regulatory mapping, FUMA, cTWAS, bulk-brain SMR, GTEx brain-tissue SMR, and brain cell-type SMR (Fig. 3D). The HDL-core-AD branch retained 44 core genes, including *YPEL3*, *INO80E*, *CAB39L*, *MED24*, *CDIPT*, *TBX6*, *LRRC37A2*, *ARL17B*, *DGKQ*, *HLA-DQB1*, *TMEM175*, *SLC26A1*, and *SLC12A9*. The ketone-body-PD branch retained 14 core genes, with the highest-confidence subset including *LRPAP1*, *DUS2*, *PRMT7*, *CENPV*, *GRK4*, *ZSWIM7*, *ELP5*, and *DDX55*. The TG-rich/VLDL-remodeling-PD branch retained 50 core genes, including *MAEA*, *PCGF3*, *SLC26A1*, *TMEM175*, *DGKQ*, *LRPAP1*, *LRRC37A2*, *CTSB*, *KAT8*, *GRK4*, *SHROOM3*, *ALKBH5*, *STX4*, and *WNT3* (Supplementary Tables S31-S41).

Insert Figure 3 near here

Single-cell projection, virtual knockout, and target annotation supported interpretable biological branches

scPagwas projection placed each focal branch into disease-relevant cellular contexts. For HDL-core-AD, pericytes were the only FDR-significant cell type (FDR = 0.014), with pathway patterns implicating cholesterol metabolism, Wnt signaling, cGMP-PKG signaling, actin-cytoskeleton regulation, and galactose metabolism. For ketone-body-PD, 13 of 21 tested cell types reached FDR significance, led by GABAergic neurons (FDR = 0.011), with additional signals in oligodendrocyte precursor cells, astrocytes, VIP GABAergic interneurons, endothelial cells, pericytes, and oligodendrocytes. For TG-rich/VLDL-remodeling-PD, pericytes were again the only FDR-significant cell type (FDR = 0.006), with ranked pathways including glycan biosynthesis, galactose metabolism, propanoate metabolism, autophagy, adherens junction, amino-acid metabolism, and calcium signaling (Fig. 4B-H; Supplementary Tables S42-S47).

Targeted scTenifoldKnk virtual knockout analyses further prioritized cell-context hypotheses. In HDL-core-AD, *KANSL1* knockout in pericytes produced 73 perturbed genes and overlapped with cGMP-PKG, HIF-1, relaxin, transcriptional misregulation, and central carbon metabolism pathways. In ketone-body-PD, *LRPAP1*, *PRMT7*, *GRK4*, and *DUS2* produced substantial perturbation profiles across pericytes, oligodendrocyte precursor cells, and VIP GABAergic interneurons, with recurrent overlap in synaptic, inflammatory, spliceosomal, and sodium-handling pathways. In TG-rich/VLDL-remodeling-PD, *TMEM175* produced a compact pericyte perturbation profile linking ribosomal and platelet-activation pathways, whereas *LRRC37A2* showed broader perturbation in oligodendrocyte precursor cells. These analyses were interpreted as prioritization of experimentally testable mechanisms (Fig. 4I,J; Supplementary Tables S48-S54).

Insert Figure 4 near here

Target annotation highlighted a focused set of experimentally approachable genes at the end of the discovery chain. *GRK4* showed the strongest overall target profile, with approved drug-gene interaction evidence, kinase-family tractability, protein-interaction support, and available structure information. *PRMT7* showed small-molecule and chemical-probe support, including SGC3027 and DS-437, and *TMEM175* showed convergent membrane-target and small-molecule support. *DUS2*, *DGKQ*, *LRPAP1*, and *LRRC37A2* showed intermediate tractability, whereas *ARL17B*, *KANSL1*, and *KNOP1* were supported mainly by structure or network context with limited direct binding evidence (Fig. 5A,C,D; Supplementary Tables S55-S59). As a supportive disease-state check, all KNK-prioritized PD genes were present in a public PD midbrain-neuron transcriptomic meta-signature, with *LRPAP1*, *GRK4*, *KNOP1*, *TMEM175*, *KANSL1*, and *PRMT7* among the higher-ranked prioritized genes (Fig. 5B; Supplementary Tables S60-S61; Supplementary Information).

Insert Figure 5 near here

Discussion

This study addressed a specific gap between metabolomics and neurodegenerative disease genetics: whether correlated circulating metabolic traits can be reorganized into genetically interpretable axes and

whether these axes show disease-specific links to neurodegeneration that can be followed from genome-wide overlap to loci, genes, cellular contexts, perturbational programs, and tractable targets. The central contribution is a structured metabolic-genetic framework that distinguishes shared metabolic programs from dispersed single-metabolite associations. Three observations define the main contribution. First, metabolic genetic architecture did not behave as a single broad metabolic liability. Instead, lipid and non-lipid traits resolved into six validated latent factor GWAS with distinct metabolic profiles. Second, the neurodegenerative disease screen was selective: among 18 factor-disease comparisons, only the HDL-core axis with AD, the ketone-body axis with PD, and the TG-rich/VLDL-remodeling axis with PD survived FDR correction. Third, downstream evidence suggested that these three branches were biologically nonredundant, mapping to different combinations of perivascular lipid support, neuronal-glial energy adaptation, lysosomal or membrane-trafficking stress, and experimentally tractable signaling or regulatory nodes.

The factor-construction stage is important because it changes the unit of analysis. Most metabolite-disease studies test individual measured metabolites or genetically predicted metabolites, which is powerful for discovery but can be difficult to interpret when hundreds of traits are highly correlated or represent partially overlapping biochemical programs. Large metabolite GWAS have shown that circulating metabolites are shaped by extensive shared genetic architecture and pleiotropy across metabolite classes [26–27]. Our results show that this covariance can be used as analytic signal instead of treated only as a nuisance. The six factor GWAS retained strong polygenic signal, recaptured their defining indicators in internal LDSC validation, and aligned with external metabolite GWAS resources. The fact that only three of 18 factor-disease comparisons were significant argues against a generic conclusion that metabolism broadly correlates with neurodegeneration. Instead, the study points to branch-specific metabolic architecture. Targeted component-level LDSC further supported this interpretation by showing that correlated, partially redundant metabolic traits could be organized through shared genetic covariance into biologically interpretable metabolic axes, while still remaining traceable to their measured indicator traits [29, 53].

The AD-associated HDL-core branch was defined primarily by ApoA1, HDL cholesteryl esters, and closely related HDL cholesterol traits, and showed a positive genetic correlation with AD. Targeted single-trait LDSC checks supported this direction across ApoA1, HDL cholesteryl esters, total HDL cholesterol, and HDL cholesterol subfractions, indicating that the factor-level result was not driven solely by an abstract latent score. This finding is compatible with genetic evidence linking lipid-related traits to AD risk, including Mendelian-randomization analyses of blood lipid fractions [54]. At the same time, observational studies suggest that the HDL-dementia relationship is complex, nonlinear, and context dependent, and should not be reduced to the conventional cardiovascular interpretation of HDL-C [15, 55].

The downstream evidence provides a plausible biological context for this HDL-core AD signal. MiXeR suggested a selective overlap profile with limited evidence for broad shared polygenicity, scPagwas localized the branch to pericytes, and *KANSL1*-centered virtual knockout implicated vascular signaling,

hypoxia adaptation, transcriptional regulation, and central carbon metabolism pathways. This pattern supports a perivascular-support model in which HDL-core genetic architecture intersects with AD through vascular and glial regulatory states. Such an interpretation is compatible with the view that AD is shaped by the neurovascular unit, blood-brain barrier integrity, pericyte function, and APOE-linked vascular dysfunction [9, 56–57]. Candidate genes including *KANSL1*, *LRRC37A2*, *ARL17B*, *DGKQ*, *TMEM175*, and *SLC26A1* point toward transcriptional regulation, membrane traffic, stress adaptation, and local vascular remodeling. Together, these findings suggest that part of the lipid-AD relationship may be organized through HDL-related vascular-support biology, not peripheral lipid abundance alone.

The PD-associated ketone-body branch was different in both direction and cellular breadth. It linked PD liability inversely to a ketone-body genetic axis, with MiXeR suggesting compact but directionally coherent overlap. Ketone bodies are increasingly discussed as both metabolic fuels and signaling molecules that can influence mitochondrial function, oxidative stress, inflammatory signaling, and brain energy resilience [22–23]. Brain energetic failure and mitochondrial dysfunction are core themes in neurodegenerative disease biology, and therapeutic strategies aimed at rescuing brain energy metabolism have been proposed across age-related neurodegenerative disorders [11, 19]. The inverse PD association observed here supports a genetic link between ketone-body metabolic architecture and disease-relevant energy-adaptive biology, while leaving the therapeutic implications of altering circulating ketone levels unresolved.

The downstream layers support this more nuanced model. Unlike the AD HDL-core branch, the ketone-body PD branch showed significant scPagwas signals across multiple cell types, led by GABAergic neuronal contexts but extending to oligodendrocyte precursor cells, astrocytes, endothelial cells, pericytes, and oligodendrocytes. This distributed pattern is consistent with the idea that metabolic adaptation in PD is not restricted to one cell class. Selective vulnerability in neurodegeneration depends on both cell-autonomous and non-cell-autonomous mechanisms, including neuronal energetic demand, glial support, inflammatory signaling, and local circuit context [58]. The core genes *LRPAP1*, *PRMT7*, *GRK4*, and *DUS2* further shifted the interpretation away from ketone abundance itself and toward signaling, transcriptional regulation, RNA handling, and proteostatic stress. *PRMT7* is biologically plausible in this framework because protein arginine methyltransferases participate in stress response and RNA biology, and dysregulated RNA processing is increasingly recognized as a cross-disease mechanism in neurodegeneration [59]. *GRK4* provides a signaling-oriented entry point whose relevance in this study comes from convergence across genetic prioritization, cell-context projection, perturbation profiles, and tractability annotation.

The PD-associated TG-rich/VLDL-remodeling branch provides a third pattern: broad polygenic sharing with PD, but only modest net LDSC correlation. This combination suggests widespread shared variant space with partially offsetting directions. Biologically, the gene, single-cell, and virtual-knockout layers converged on lysosomal, membrane-trafficking, degradative, adhesion, and pericyte stress-response biology. This is relevant to PD, where lysosomal dysfunction, autophagy, alpha-synuclein handling, and membrane trafficking are central disease mechanisms [11, 17]. The appearance of *TMEM175* in this

branch is notable because *TMEM175* is a lysosomal ion channel genetically linked to PD and increasingly considered a tractable lysosomal target [10, 60]. In our data, *TMEM175* appeared within a broader remnant-lipid branch that also included *DGKQ*, *LRR37A2*, *LRPAP1*, *CTSB*, *ALKBH5*, and *GRK4*, supporting a model in which triglyceride-rich/remnant lipid architecture intersects with PD through endolysosomal stress handling and neurovascular or neuroimmune coupling.

The two lipid branches should therefore not be collapsed into a single lipid-metabolism result. The HDL-core AD branch and the TG-rich/VLDL-remodeling PD branch both involve lipid-related factors and pericyte contexts, but their inferred biology differs. The AD branch is more compatible with HDL-core, apolipoprotein, and perivascular-support biology. The PD branch is more compatible with triglyceride-rich remnant handling, lysosomal stress, membrane traffic, autophagy, and adhesion-related responses. This distinction is an advantage of the latent-factor design: a single lipid score or broad serum lipid category would likely have obscured these differences.

The multi-layered gene and cell analyses strengthened the biological interpretation of these branches by testing whether locus-level signals, transcriptome-based prioritization, single-cell projection, and virtual perturbation pointed to compatible mechanisms. This convergence is especially useful in neurodegenerative disease, where disease mechanisms are spatially and cellularly heterogeneous and where the same locus can influence different processes across brain cell types or disease stages [58]. The convergence was not uniform across every method, but the strongest branches repeatedly pointed to coherent biological programs: vascular support for the HDL-core AD branch, distributed energy and stress adaptation for the ketone-body PD branch, and lysosomal or trafficking stress for the TG-rich/VLDL-remodeling PD branch. This layered evidence does not replace experimental validation, but it provides a structured basis for selecting genes and cell contexts for functional follow-up.

The target-annotation layer adds a practical translational endpoint to this framework. Beyond broad pathway nomination, the analysis highlighted a small set of experimentally approachable nodes embedded within branch-specific biology. *GRK4* emerged as a signaling-oriented PD target, *PRMT7* as a regulatory and RNA-biology target with existing probe support, and *TMEM175* as a lysosomal or membrane-trafficking target. These genes are attractive because they are annotated as tractable and supported by genetic, cellular, and perturbational evidence within specific metabolic-disease branches.

Several limitations should be acknowledged. First, all principal analyses were based on GWAS summary statistics and therefore inherit the ancestry composition, phenotype definitions, sample overlap patterns, and power limitations of the source datasets. Although LDSC intercepts argued against substantial cross-trait inflation in the disease screen, summary-statistic analyses cannot fully resolve individual-level confounding or mediation. Second, latent metabolic factors are statistical summaries of shared genetic architecture, not isolated biochemical pathways. Their interpretation depends on loadings, internal validation, external validation, and downstream convergence, and their signs are arbitrary up to model orientation. Third, the single-cell and virtual-knockout analyses should be interpreted as context-prioritization and hypothesis-generation layers, not direct causal proof. Colocalization can be sensitive to

regional architecture and multiple causal variants; cTWAS and SMR depend on the quality and relevance of expression reference panels; scPagwas projects GWAS signal into cell contexts but does not establish cell-autonomous causality; and scTenifoldKnn is an in silico perturbation method without experimental knockout. Fourth, the external disease-state expression and proteomic checks were intentionally treated as supportive because they relied on processed public resources with incomplete gene coverage and did not constitute a harmonized de novo validation analysis. Finally, the current analyses do not establish temporal ordering, disease-stage specificity, or causal mediation from metabolic programs to neurodegeneration.

In summary, this study suggests that the genetic relationship between metabolism and neurodegeneration is structured, selective, and branch-specific. By moving from individual metabolites to latent metabolic factor GWAS, we identified three disease-aligned metabolic axes and followed them through polygenic overlap, shared loci, candidate genes, single-cell contexts, perturbational signatures, and target annotations. The resulting framework supports three working models: an HDL-core/perivascular-support axis for AD, a ketone-body/neuronal-glial energy-adaptation axis for PD, and a TG-rich remnant/lysosomal-trafficking axis for PD. These models remain hypotheses requiring experimental validation, but they provide a more coherent foundation for mechanistic follow-up than single-metabolite association alone.

Abbreviations

AD: Alzheimer disease

APOE: apolipoprotein E

CFI: comparative fit index

conjFDR: conjunctive false discovery rate

cTWAS: causal transcriptome-wide association study

DGIdb: Drug-Gene Interaction Database

FDR: false discovery rate

FUMA: Functional Mapping and Annotation

GTEx: Genotype-Tissue Expression

GWAS: genome-wide association study

HDL: high-density lipoprotein

HPA: Human Protein Atlas

LBD: Lewy body dementia

LDSC: linkage disequilibrium score regression

MiXeR: causal mixture model

NDD: neurodegenerative disease

NMR: nuclear magnetic resonance

OPC: oligodendrocyte precursor cell

PD: Parkinson disease

PWCoCo: pairwise colocalization

Q_SNP: SNP-level heterogeneity statistic

rg: genetic correlation

scPagwas: single-cell pathway-based genome-wide association study

scTenifoldKnk: single-cell TenifoldKnk

SEM: structural equation modelling

SMR: summary-data-based Mendelian randomization

SNP: single nucleotide polymorphism

SRMR: standardized root mean square residual

TG: triglyceride

TRS: trait-relevance score

UMAP: uniform manifold approximation and projection

VLDL: very-low-density lipoprotein

Declarations

Ethics approval and consent to participate

This study used publicly available or controlled-access summary-level genetic, transcriptomic, proteomic, and annotation resources. No new individual-level human participant data were collected. Ethical approval and informed consent were obtained by the original contributing studies.

Consent for publication

Not applicable.

Availability of data and materials

All source GWAS summary statistics, single-cell/single-nucleus transcriptomic datasets, eQTL resources, and public disease-state resources used in this study are available from the original studies, repositories, or databases cited in the manuscript, Supplementary Methods, and Supplementary Tables. The source GWAS and single-cell datasets were not redistributed because they remain subject to the data-access terms, licenses, and usage policies of the original data providers. The GWAS summary statistics for the six latent metabolic genetic factors generated in this study, including three lipid factors and three non-lipid factors, have been deposited at Zenodo under DOI: <https://doi.org/10.5281/zenodo.20174020> . Derived analysis summaries, source identifiers, accession numbers, and intermediate result tables are provided in the Supplementary Tables. Large intermediate files generated during downstream analyses are available from the corresponding authors upon reasonable request for reproducibility purposes.

Code availability

All analysis scripts used for GWAS processing, LD score regression, GenomicSEM factor construction, pleiotropy analyses, colocalization, transcriptome-wide association analysis, and single-cell integration are available at: <https://github.com/duanxian0301/metabolism>.

Declaration of competing interests

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. All authors have completed the journal's disclosure requirements and report no conflicts of interest related to this study.

Funding

The authors declare that no specific funding was received for this work.

Authors' contributions

Jing Shen: Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualization.
Yuxuan Shen: Data curation, Formal analysis, Validation, Visualization. Gang Yu: Software, Data curation, Validation, Investigation. Lijun Pang: Supervision, Project administration, Writing - review and editing.
Chenxu Xiao: Conceptualization, Supervision, Methodology, Writing - review and editing, Project administration. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the investigators and participants of the metabolomics, neurodegenerative disease GWAS, single-cell, eQTL, proteomic, transcriptomic, and target-annotation resources used in this study.

Declaration of generative AI and AI-assisted technologies in the writing process

During preparation of this work, the authors used AI-assisted tools to support language editing and formatting. After using these tools, the authors reviewed and edited the content and take full responsibility for the scientific accuracy of the manuscript.

References

1. GBD 2021 Nervous System Disorders Collaborators. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol.* 2024;23:344–381. doi:10.1016/S1474-4422(24)00038-3.
2. Livingston, Gill, Huntley, Jonathan, Sommerlad, Andrew, et al.. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020;396:413–446. doi:10.1016/S0140-6736(20)30367-6.
3. McKhann, Guy M., Knopman, David S., Chertkow, Howard, Hyman, Bradley T., et al.. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups. *Alzheimer's & Dementia* 2011;7:263–269. doi:10.1016/j.jalz.2011.03.005.
4. Bloem, Bastiaan R., Okun, Michael S., Klein, Christine. Parkinson's disease. *Lancet* 2021;397:2284–2303. doi:10.1016/S0140-6736(21)00218-X.
5. Postuma, Ronald B., Berg, Daniela, Stern, Matthew, Poewe, Werner, et al.. MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders* 2015;30:1591–1601. doi:10.1002/mds.26424.
6. Scheltens, Philip, De Strooper, Bart, Kivipelto, Miia, Holstege, Henne, Ch'etelat, Gael, Teunissen, Charlotte E., et al.. Alzheimer's disease. *The Lancet* 2021;397:1577–1590. doi:10.1016/S0140-6736(20)32205-4.

7. De Strooper, Bart, Karran, Eric. Alzheimer's disease: from cellular pathology to therapeutic strategies. *Cell* 2016;164:603–615. doi:10.1016/j.cell.2016.01.028.
8. Bellenguez, Celine, et al.. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics* 2022;54:412–436. doi:10.1038/s41588-022-01024-z.
9. Sweeney, Melanie D., Sagare, Abhay P., Zlokovic, Berislav V.. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nature Reviews Neurology* 2018;14:133–150. doi:10.1038/nrneurol.2017.188.
10. Kim, Joo-Yeon, et al.. Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. *Nature Genetics* 2024. doi:10.1038/s41588-024-01669-0.
11. Coukos, Robert, Krainc, Dimitri. Key genes and convergent pathogenic mechanisms in Parkinson disease. *Nature Reviews Neuroscience* 2024;25:393–413. doi:10.1038/s41583-024-00812-2.
12. Toledo, Jon B., Arnold, Matthias, Kastenmuller, Gabi, et al.. Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimer's & Dementia* 2017;13:965–984. doi:10.1016/j.jalz.2017.01.020.
13. Mapstone, Mark, Cheema, Aneela K., Fiandaca, Massimo S., et al.. Plasma phospholipids identify antecedent memory impairment in older adults. *Nature Medicine* 2014;20:415–418. doi:10.1038/nm.3466.
14. Hatano, Taku, Saiki, Shinji, Okuzumi, Ayami, Mohny, Robert P., Hattori, Nobutaka. Identification of novel biomarkers for Parkinson's disease by metabolomic technologies. *Journal of Neurology, Neurosurgery & Psychiatry* 2016;87:295–301. doi:10.1136/jnnp-2014-309676.
15. Blumenfeld, Jessica, Yip, Oscar, Kim, Min Joo, Huang, Yadong. Cell type-specific roles of APOE4 in Alzheimer disease. *Nature Reviews Neuroscience* 2024;25:91–110. doi:10.1038/s41583-023-00776-9.
16. Haney MS, et al. APOE4/4 is linked to damaging lipid droplets in Alzheimer's disease microglia. *Nature*. 2024;628:154–161. doi:10.1038/s41586-024-07185-7.
17. Nixon, Ralph A., Rubinshtein, David C.. Mechanisms of autophagy-lysosome dysfunction in neurodegenerative diseases. *Nature Reviews Molecular Cell Biology* 2024;25:926–946. doi:10.1038/s41580-024-00757-5.
18. Kaddurah-Daouk, Rima, Rozen, Sergey, Matson, William, Han, Xia, Hulette, Christine M., Burke, James R., et al.. Metabolomic changes in autopsy-confirmed Alzheimer's disease. *Alzheimer's & Dementia* 2013;9:309–320. doi:10.1016/j.jalz.2012.08.015.
19. Colca JR, Finck BN. Metabolic mechanisms connecting Alzheimer's and Parkinson's diseases: potential avenues for novel therapeutic approaches. *Front Mol Biosci*. 2022;9:929328. doi:10.3389/fmolb.2022.929328.
20. Paul KC, Zhang K, Walker DI, Sinsheimer JS, Yu Y, Kusters C, et al. Untargeted serum metabolomics reveals novel metabolite associations and disruptions in amino acid and lipid metabolism in Parkinson's disease. *Mol Neurodegener*. 2023;18:100. doi:10.1186/s13024-023-00694-5.

21. Zhang JD, Xue C, Kolachalama VB, Donald WA. Interpretable machine learning on metabolomics data reveals biomarkers for Parkinson's disease. *ACS Cent Sci.* 2023;9:1035–1045. doi:10.1021/acscentsci.2c01468.
22. Newman, John C., Verdin, Eric. Ketone bodies as signaling metabolites. *Trends in Endocrinology and Metabolism* 2017;28:42–52. doi:10.1016/j.tem.2016.09.002.
23. Norwitz, Nicholas G., Hu, Mou, Clarke, Kieran. Ketone bodies as signaling metabolites in health and disease. *Current Opinion in Endocrinology, Diabetes and Obesity* 2020. doi:10.1097/MED.0000000000000567.
24. Cunnane, Stephen C., Courchesne-Loyer, Anne, St-Pierre, Veronique, Vandenberghe, Cedric, et al.. Can ketones compensate for deteriorating brain glucose uptake during aging? Implications for the risk and treatment of Alzheimer's disease. *Annals of the New York Academy of Sciences* 2016;1367:12–20. doi:10.1111/nyas.12999.
25. Youm, Yun-Hee, Nguyen, Kim Y., Grant, Ryan W., Goldberg, Emily L., et al.. The ketone metabolite beta-hydroxybutyrate blocks NLRP3 inflammasome-mediated inflammatory disease. *Nature Medicine* 2015;21:263–269. doi:10.1038/nm.3804.
26. Karjalainen, Juha, et al.. Genome-wide characterization of circulating metabolic biomarkers. *Nature* 2024;628:130–138. doi:10.1038/s41586-024-07148-y.
27. Tambets R, Kronberg J, Abner E, Võsa U, Rahu I, Taba N, et al. Genome-wide association study for circulating metabolites in 619,372 individuals. *medRxiv.* 2024. doi:10.1101/2024.10.15.24315557.
28. Solovieff, Nadia, Cotsapas, Chris, Lee, Phil H., Purcell, Shaun M., Smoller, Jordan W.. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* 2013;14:483–495. doi:10.1038/nrg3461.
29. Grotzinger, Andrew D., Rhemtulla, Mijke, de Vlaming, Ronald, Ritchie, Stuart J., Mallard, Travis T., Hill, W. David, et al.. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* 2019;3:513–525. doi:10.1038/s41562-019-0566-x.
30. Bycroft, Clare, Freeman, Colin, Petkova, Desislava, et al.. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–209. doi:10.1038/s41586-018-0579-z.
31. Chia, Ruth, et al.. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nature Genetics* 2021;53:294–303. doi:10.1038/s41588-021-00785-3.
32. Bulik-Sullivan, Brendan K., Loh, Po-Ru, Finucane, Hilary K., Ripke, Stephan, Yang, Jian, Patterson, Nick, et al.. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* 2015;47:291–295. doi:10.1038/ng.3211.
33. Benjamini, Yoav, Hochberg, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* 1995;57:289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

34. Frei, Oleksandr, Holland, Dominic, Smeland, Olav B., et al.. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications* 2019;10:2417. doi:10.1038/s41467-019-10310-0.
35. Andreassen, Ole A., Thompson, Wesley K., Schork, Andrew J., et al.. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genetics* 2013;9:e1003455. doi:10.1371/journal.pgen.1003455.
36. Giambartolomei, Claudia, Vukcevic, Damjan, Schadt, Eric E., Franke, Lude, Hingorani, Aroon D., Wallace, Chris, et al.. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genetics* 2014;10:e1004383. doi:10.1371/journal.pgen.1004383.
37. Wallace, Chris. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genetics* 2021;17:e1009440. doi:10.1371/journal.pgen.1009440.
38. Hormozdiari, Farhad, van de Bunt, Martijn, Segre, Ayellet V., et al.. Colocalization of GWAS and eQTL signals detects target genes. *American Journal of Human Genetics* 2016;99:1245–1260. doi:10.1016/j.ajhg.2016.10.003.
39. Watanabe, Kyoko, Taskesen, Erdogan, van Bochoven, Arjen, Posthuma, Danielle. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017;8:1826. doi:10.1038/s41467-017-01261-5.
40. Zhao, Siming, Crouse, Wesley, Qian, Sheng, Luo, Kaixuan, Stephens, Matthew, He, Xin. Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. *Nature Genetics* 2024;56:336–347. doi:10.1038/s41588-023-01648-9.
41. Zhu, Zhihong, Zhang, Futao, Hu, Han, Bakshi, Andrew, Robinson, Matthew R., Powell, Joseph E., et al.. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* 2016;48:481–487. doi:10.1038/ng.3538.
42. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;369:1318–1330. doi:10.1126/science.aaz1776.
43. Bryois, Julien, et al.. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nature Neuroscience* 2022;25:1104–1112. doi:10.1038/s41593-022-01128-z.
44. Ma Y, Deng C, Zhou Y, Zhang Y, Qiu F, Jiang D, et al. Polygenic regression uncovers trait-relevant cellular contexts through pathway activation transformation of single-cell RNA sequencing data. *Cell Genomics* 2023;3:100383. doi:10.1016/j.xgen.2023.100383.
45. Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, et al. scTenifoldKnk: an efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns*. 2022;3:100434. doi:10.1016/j.patter.2022.100434.
46. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Research* 2021;49:D1302-D1310. doi:10.1093/nar/gkaa1027.

47. Freshour, Sharon L., Kiwala, Soma, Cotto, Kelsy C., et al.. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Research* 2021;49:D1144–D1151. doi:10.1093/nar/gkaa1084.
48. Uhlen, Mathias, Fagerberg, Linn, Hallstrom, Bjorn M., et al.. Tissue-based map of the human proteome. *Science* 2015;347:1260419. doi:10.1126/science.1260419.
49. Jumper, John, Evans, Richard, Pritzel, Alexander, et al.. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–589. doi:10.1038/s41586-021-03819-2.
50. Tunyasuvunakool, Kathryn, Adler, Jonas, Wu, Zachary, et al.. Highly accurate protein structure prediction for the human proteome. *Nature* 2021;596:590–596. doi:10.1038/s41586-021-03828-1.
51. Pichet Binette, Alexa, et al.. Proteomic changes in Alzheimer's disease associated with progressive A β plaque and tau tangle pathologies. *Nature Neuroscience* 2024;27:1880–1891. doi:10.1038/s41593-024-01737-w.
52. van den Hurk, Mark, Lau, Shong, Marchetto, Maria C., Mertens, Jerome, Stern, Shani, Corti, Olga, et al.. Druggable transcriptomic pathways revealed in Parkinson's patient-derived midbrain neurons. *npj Parkinson's Disease* 2022;8:134. doi:10.1038/s41531-022-00400-0.
53. Grotzinger, Andrew D., de la Fuente, Javier, Davies, Gail, Nivard, Michel G., Tucker-Drob, Elliot M.. Transcriptome-wide and stratified genomic structural equation modeling identify neurobiological pathways shared across diverse cognitive traits. *Nature Communications* 2022;13:6280. doi:10.1038/s41467-022-33724-9.
54. Proitsi P, Lupton MK, Velayudhan L, Newhouse S, Fogh I, et al. Genetic predisposition to increased blood cholesterol and triglyceride lipid levels and risk of Alzheimer disease: a Mendelian randomization analysis. *PLoS Med.* 2014;11:e1001713. doi:10.1371/journal.pmed.1001713.
55. Westerlin Kjeldsen E, Thomassen JQ, Juul Rasmussen I, Nordestgaard BG, Tybjaerg-Hansen A, Frikke-Schmidt R. Plasma high-density lipoprotein cholesterol and risk of dementia: observational and genetic studies. *Cardiovasc Res.* 2022;118:1330–1343. doi:10.1093/cvr/cvab164.
56. Zlokovic, Berislav V.. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nature Reviews Neuroscience* 2011;12:723–738. doi:10.1038/nrn3114.
57. Montagne, Axel, Nation, Daniel A., Sagare, Abhay P., et al.. APOE4 leads to blood-brain barrier dysfunction predicting cognitive decline. *Nature* 2020;581:71–76. doi:10.1038/s41586-020-2247-3.
58. Kampmann, Martin. Molecular and cellular mechanisms of selective vulnerability in neurodegenerative diseases. *Nature Reviews Neuroscience* 2024;25:351–371. doi:10.1038/s41583-024-00806-0.
59. Wolozin, Benjamin, Ivanov, Pavel. Stress granules and neurodegeneration. *Nature Reviews Neuroscience* 2019;20:649–666. doi:10.1038/s41583-019-0222-5.
60. Wie, Juyoung, et al.. A growth-factor-activated lysosomal K⁺ channel regulates Parkinson's pathology. *Nature* 2019;569:294–299. doi:10.1038/s41586-019-1181-3.

Figures

Figure 1. Overview of study design and analytical workflow

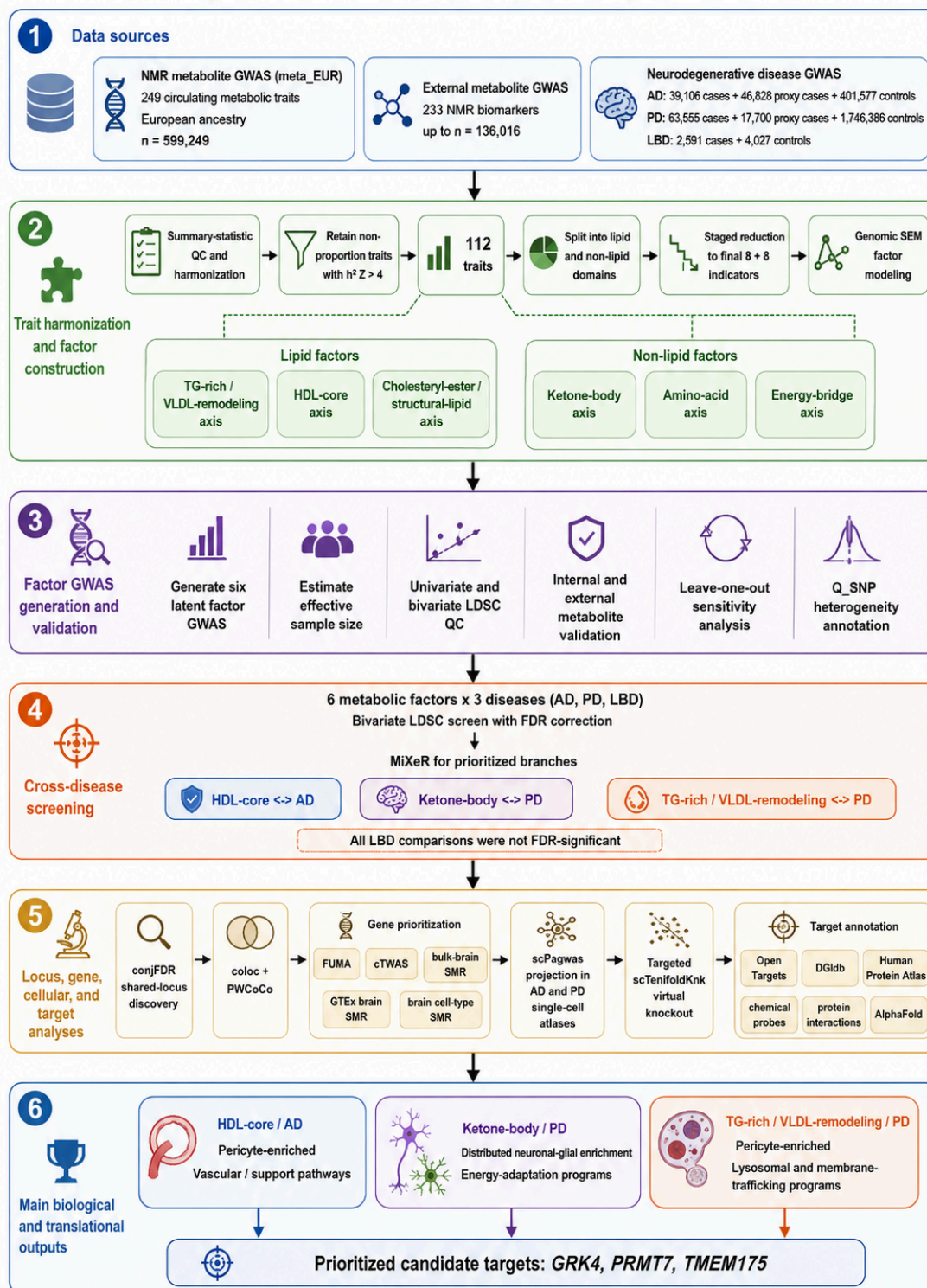


Figure 1

Overview of the analytical framework.

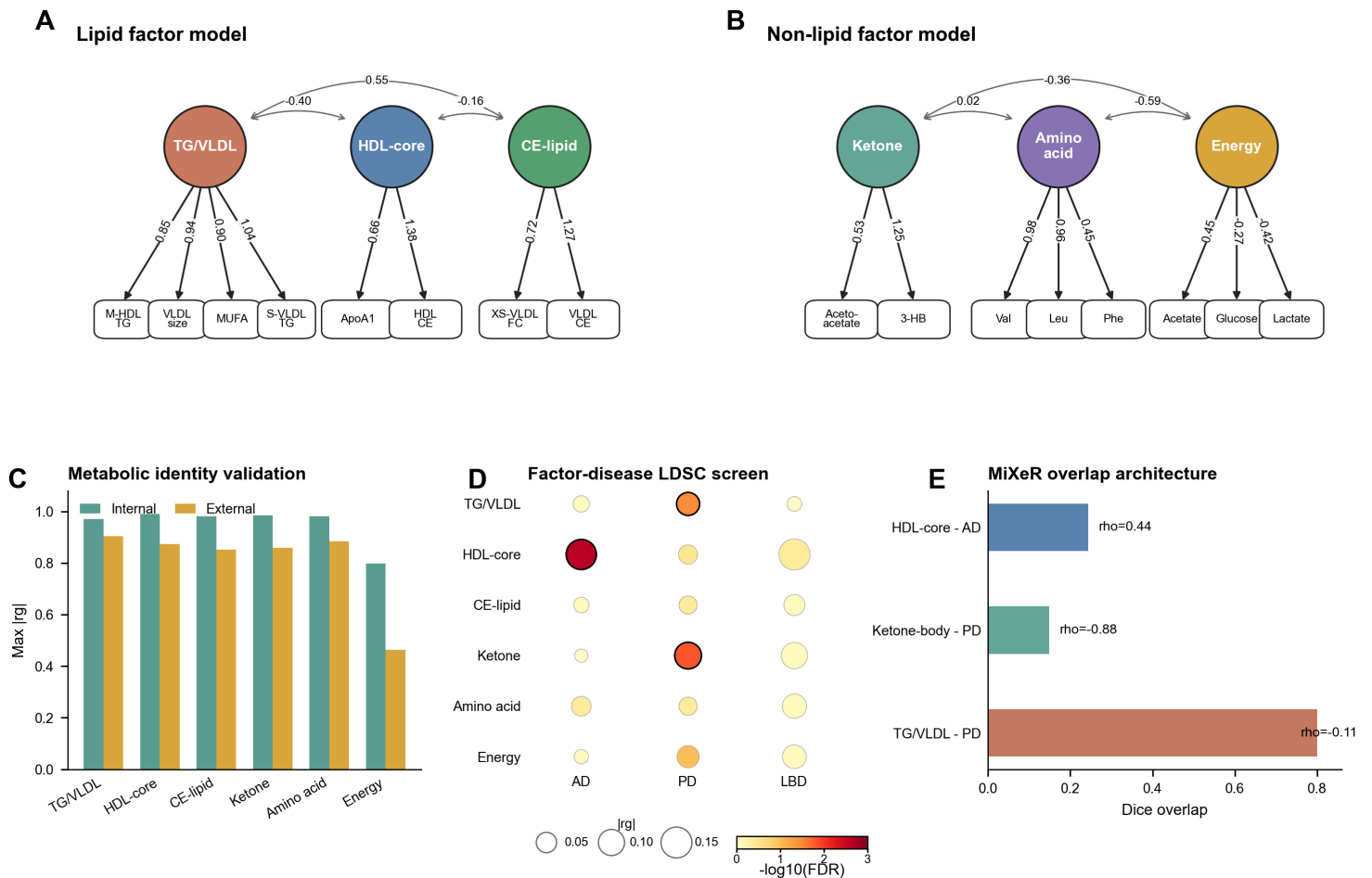


Figure 2

Genetic construction, validation, and disease screening of lipid and non-lipid metabolic factors.

(A, B) GenomicSEM factor structures for the lipid and non-lipid metabolic domains. Metabolic traits are shown as observed indicators and latent metabolic factors as circles. Standardized factor loadings are shown on indicator paths, and latent factor correlations are shown between factors. (C) Genetic validation of the derived factor GWASs against their internal and supporting metabolic traits, summarized by bivariate LDSC genetic correlations. (D) Cross-trait LDSC screening between the six metabolic factor GWASs and neurodegenerative disease GWASs. Circle size represents the absolute genetic correlation ($|rg|$), and color represents the FDR-adjusted significance level. Three factor-disease pairs passed the FDR threshold and were taken forward: HDL-core with AD, ketone-body with PD, and TG/VLDL-core with PD. (E) MiXeR-based genetic-overlap estimates for the prioritized metabolic factor-disease pairs, showing the extent of shared polygenic architecture beyond genome-wide significant loci. AD, Alzheimer's disease; PD, Parkinson's disease; LDSC, linkage disequilibrium score regression; rg , genetic correlation; FDR, false discovery rate; MiXeR, causal mixture model.

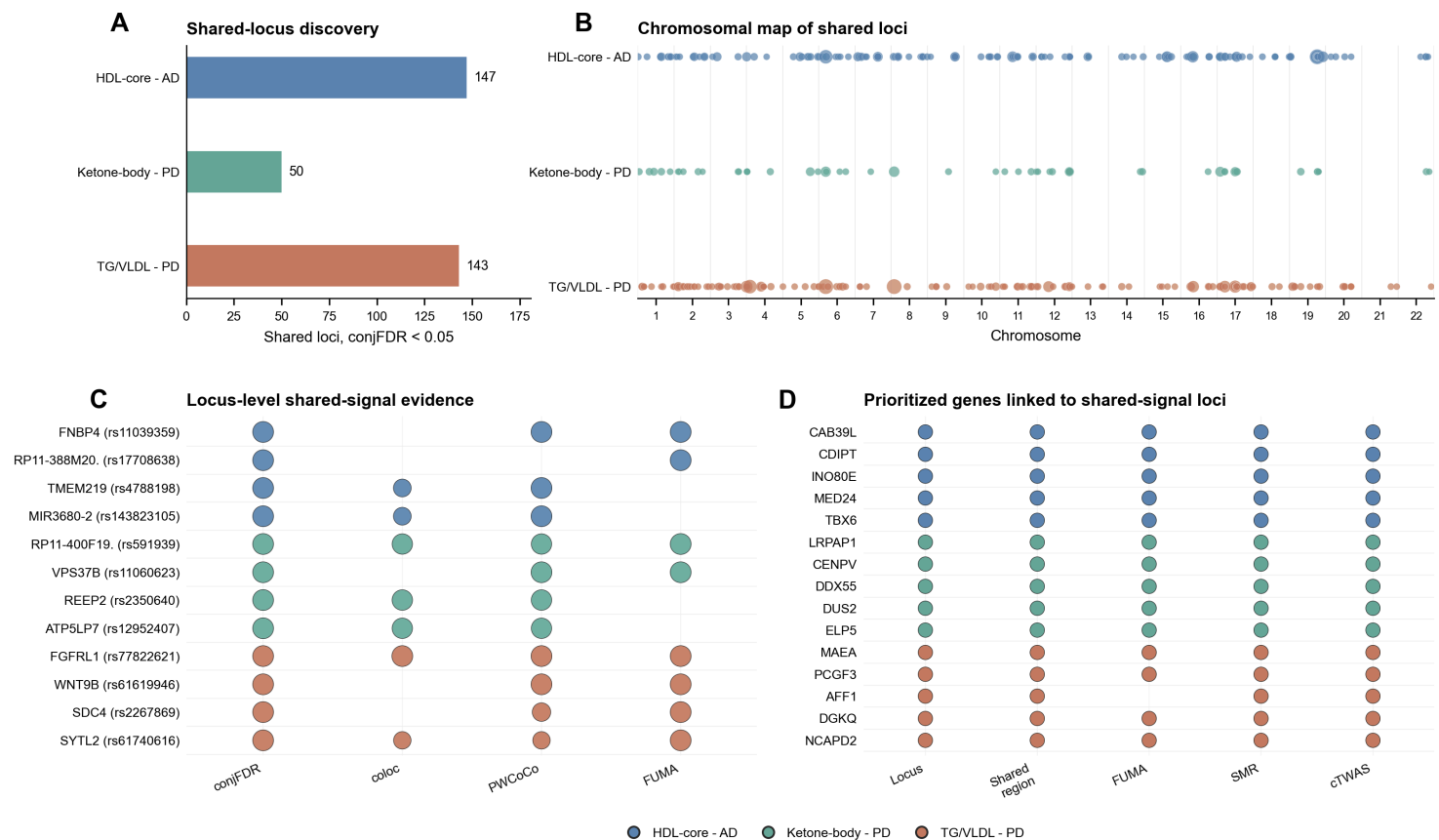


Figure 3

Shared loci and convergent gene prioritization across the three metabolic factor-NDD axes.

(A) Summary of shared-locus discovery across the prioritized factor-disease pairs, including pleiotropic loci and downstream mapped genes. (B) Chromosomal distribution of shared loci identified for HDL-core/AD, ketone-body/PD, and TG/VLDL-core/PD. (C) Locus-level evidence supporting shared genetic signals, integrating conjunctural FDR, colocalization, PWCoCo, and FUMA annotation. Gene names in this panel denote nearest or mapped genes used to label shared regions rather than independent proof of gene causality. (D) Prioritized genes linked to shared-signal regions after integrating locus annotation with regulatory evidence, including FUMA mapping, SMR, and cTWAS. Colocalization and PWCoCo evidence in this panel indicate that the gene lies within or maps to a shared-signal region supported by locus-level evidence. AD, Alzheimer's disease; PD, Parkinson's disease; conjFDR, conjunctural false discovery rate; FUMA, Functional Mapping and Annotation; SMR, summary-data-based Mendelian randomization; cTWAS, causal transcriptome-wide association study; PWCoCo, pairwise colocalization.

summary-data-based Mendelian randomization; cTWAS, causal transcriptome-wide association study; FDR, false discovery rate.

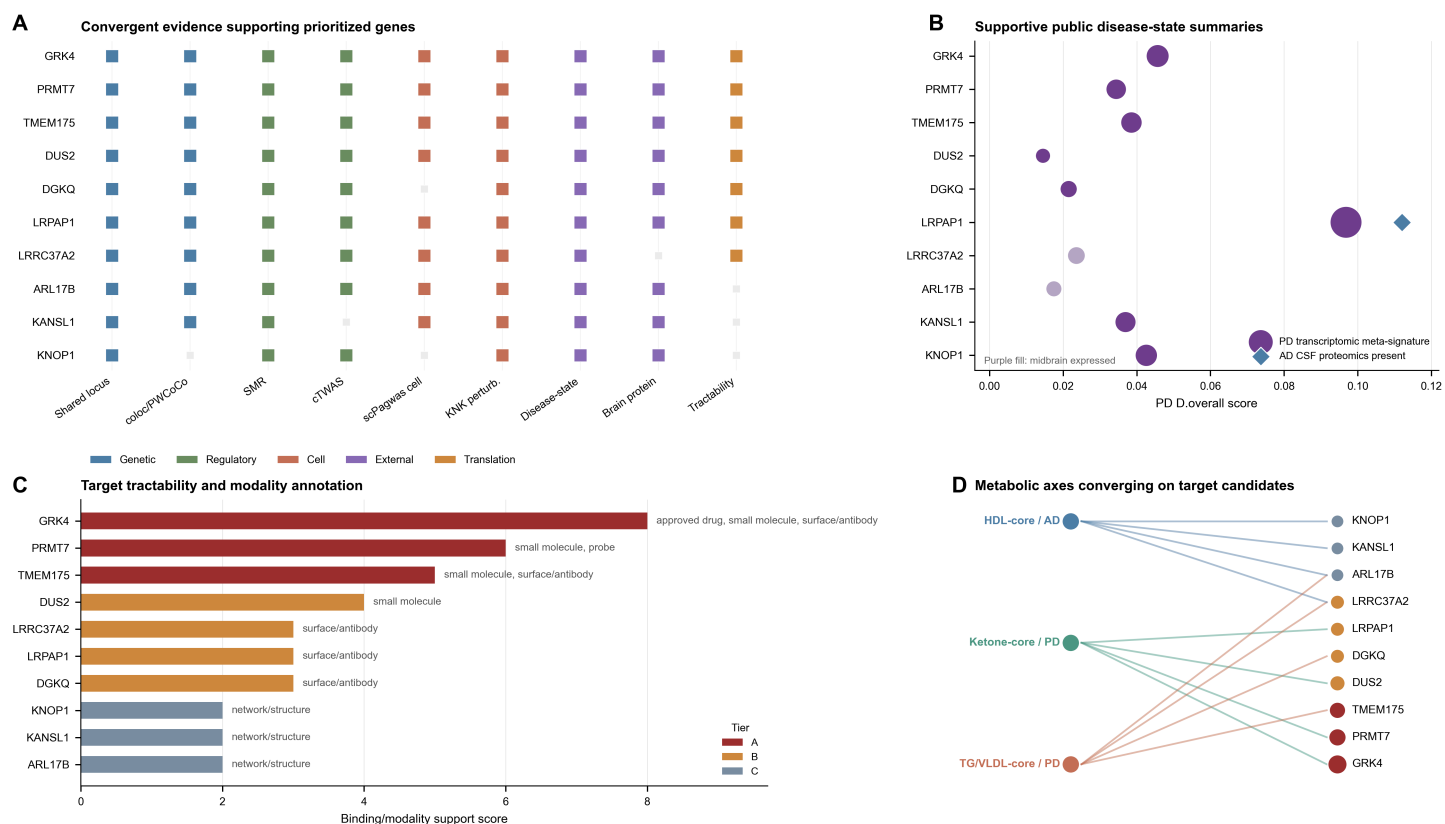


Figure 5

Translational prioritization of metabolic factor-linked candidate genes.

(A) Evidence-convergence matrix for prioritized target genes. Columns summarize genetic evidence, regulatory evidence, cell-level support, external disease-state support, and translational annotation. The SMR column represents the union of GTEx SMR, bulk-brain SMR, and cell-type SMR support. Filled squares indicate the presence of support from the corresponding evidence layer. (B) Supportive public disease-state summaries for prioritized genes. Circles show PD transcriptomic meta-signature support using the D.overall score, with darker purple indicating midbrain expression. Diamonds indicate genes present in the AD CSF proteomics summary. These analyses were used as supportive validation rather than primary causal evidence. (C) Target tractability and modality annotation for prioritized genes, integrating approved drug-gene evidence, small-molecule binding or pocket support, chemical probe availability, surface or antibody modality support, protein interaction evidence, and structural availability. Bar length represents the binding/modality support score, and color indicates the tractability tier. (D) Network-style summary linking the three prioritized metabolic factor-disease axes to candidate target genes. Node color reflects the metabolic axis or tractability tier, and edges indicate that a gene was prioritized within the corresponding axis. AD, Alzheimer's disease; PD, Parkinson's disease; SMR,

summary-data-based Mendelian randomization; GTEx, Genotype-Tissue Expression project; CSF, cerebrospinal fluid.

Supplementary Files

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

- [supplementarymethods.docx](#)
- [supplementaryinformation.docx](#)
- [supplementarytables.xlsx](#)