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.

Chronological Age Reshapes the Molecular Landscape of Polygenic Risk Across the Human Lifespan

Nathan Price

nprice@buckinstitute.org

Buck Institute for Research on Aging https://orcid.org/0000-0002-4157-0267

Bartek Nogal

The Buck Institute for Research on Aging

John Earls

The Buck Institute for Research on Aging

Ornit Nahman

The Buck Institute for Research on Aging

Article

Keywords:

Posted Date: April 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9418894/v1

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

Additional Declarations: There is **NO** Competing Interest.

Abstract

An individual's polygenic risk is fixed at conception, yet those pathological consequences manifest dynamically over the lifespan. The molecular mechanisms through which chronological age reshapes static genetic risk into late-onset disease remain largely unknown, obscured by the inability of conventional linear methods to model complex, non-linear interactions in high-dimensional multi-omic data. We combined over 100 PRS with >1,100 metabolomic, proteomic, and clinical features from 1,491 adults using a perturbable variational autoencoder (VAE) that explicitly incorporates chronological age. This framework generated multi-omic maps of genetic risk, contrasting age-residualized and age-included models. Accounting for age qualitatively altered the PRS–omics landscape: age-residualized models emphasized broad metabolite signatures, while age-included models uncovered fewer, more organ-level and clinically interpretable features. The analysis revealed mechanistically coherent patterns spanning hallmarks of aging: impaired antioxidant defense (e.g. reduced heme oxygenase-1 with Alzheimer's PRS), lysosomal dysfuncton(e.g. reduced α-L-iduronidase with Parkinson's PRS), iron-driven inflammation (e.g. elevated ferritin and myeloperoxidase with ALS PRS), and altered immune regulation (e.g. reduced SLAMF1 with cardiovascular PRS). Associations with N-acetylmethionine across cardiometabolic and longevity PRS implicated one-carbon metabolism as a shared axis of disease and resilience, while consistently lower alkaline phosphatase with longevity PRS highlighted a widely used clinical chemistry marker as a potential indicator of healthspan. Reframing polygenic risk as an age-shaped property shifted the emphasis from diffuse metabolite signatures to more clinically interpretable biomarkers and organ-level features. These results demonstrate that polygenic risk is not a static baseline but a dynamic molecular bias modulated by aging. By linking genetics to mechanisms of aging we can identify accessible biomarkers that could inform strategies for monitoring resilience and targeting prevention in age-related disease.

Main

Aging is a complex, dynamic, and inherently nonlinear process shaped by genetics, environmental factors, and stochastic effects [1,2,3,4,5,6]. Studying aging in humans remains challenging: we are long-lived, no single biomarker of biological aging is widely validated, and clinical trial design is complicated by these limitations. Animal models approximate potential interventions and biomarkers to only a limited extent. In parallel, however, genome-wide association studies (GWAS) have grown in size, capturing more complex trait variance, and their derivatives—polygenic risk scores (PRS)—provide increasingly insightful and straightforward tools to probe the genetic underpinnings of aging traits[7,8,9] . Despite the modest variance explained by PRS in many complex traits [10,11], they capture foundational biological pathway information. This information can serve the aging field by providing an efficient and comparatively unbiased framework for generating hypotheses about aging-related disease biomarkers and intervention leads in human populations.

Despite their utility, many current approaches to identifying novel biomarkers or drug targets from GWAS-discovered variants rely on linear models [12,13,14], which are not ideal for capturing the nonlinear

trajectories that characterize age-related processes. In conventional GWAS and related linear models, the additive effect of single variants is assumed, ignoring potential interactions such as epistasis or gene–gene synergy[15]. This assumption arises partly from computational constraints and partly from the complexity of modeling the myriad combinatorial-level effects across millions of SNPs. Such nonlinear associations often escape traditional linear and frequentist methods, as multiple studies have shown the importance of these higher-order interactions for revealing otherwise indiscernible signals in genomic data [16,17,18]. Moreover, pleiotropy is often considered a confounder in linear models, but for multi-omics target and biomarker discovery, pleiotropy can be a feature. When studying age-related phenotypes, PRS frequently span many distinct but interrelated traits [18,19,20,21,22]. This broad genomic footprint can facilitate the identification of multiple pathways involved in complex aging processes. Indeed, recent work by Jabalameli et al. highlights that inescapable pleiotropy and non-linear interactions among longevity traits are central to capturing the additive liability conferred by many aging-related traits [21] . In their report and the work of others, in order to approximate nonlinear interactions among age-related traits, investigators often develop composite scores using approaches such as stacked elastic-net regression or multivariate GWAS as frameworks for suggesting potential drug targets for anti-aging interventions [21,23]. However, since all these methods fundamentally rely on supervised linear regression, we propose that complementary deep learning (DL)-based approaches—capable of efficiently capturing the nonlinear manifestation of genetic risk across phenomic and age spectrums—are now poised to facilitate the discovery of novel targets and biomarkers.

Nevertheless, current data-centric DL approaches in health remain constrained by the availability of suitably large and well-phenotyped cohorts and by the substantial computational resources required. High-dimensional biological data (particularly raw genomic data) often yield a large number of potential predictors (p) relative to the sample size (n), which can lead to overfitting even in well-regularized neural networks. One promising way to address this challenge is to use generative models such as variational autoencoders (VAEs). By embedding thousands of features into a lower-dimensional latent space with a Gaussian prior, VAEs enforce a distributional structure that curbs overfitting—particularly in moderately sized datasets—while still capturing higher-order, nonlinear patterns. These models can robustly characterize synergistic or multi-omic signals that linear methods are prone to overlook, thus uncovering subtle cross-pathway interactions relevant to complex age-related traits [24,25,26] .

Toward this end, we use a perturbable variational autoencoder introduced by Allesoe et al.[27] We apply it to PRS-pleiotropy–driven multi-omics discovery by explicitly simulating genetic risk, where for each person's encoding, we toggle the high-PRS feature from 0 to 1. We then interpret the decoder's output as the PRS-attributable change across omic traits. In this framework, the VAE is first trained on a comprehensive multi-omics dataset (including both high-dimensional omics features and multiple top-risk-decile-binarized PRS) to learn a nonlinear latent space that captures underlying relationships among genetic risk and molecular measures. After training, a given PRS "high risk" state in an individual's input vector is modulated and fed into the encoder. Because the encoder is composed of nonlinear neural layers, a single binary shift in the PRS can induce a notable transformation in that individual's latent

representation, which effectively encodes how the VAE expects a high-risk genotype to interact with the rest of that person's omics features. When this shifted latent embedding is passed through the decoder (the generative stage), the VAE produces a reconstructed multi-omics profile reflective of the high-risk genotype. By comparing this reconstructed profile across a cohort of individuals' original, unperturbed reconstructions, omics features that undergo consistent changes under the model's learned assumptions can be ascertained via an integrated Bayesian decision theory layer for quantifying the posterior evidence that a given omics feature is significantly affected when a PRS is toggled [27,28]. Using this approach, we posit that this VAE model can facilitate the discovery of broad, system-wide effects of genetic risk that may be masked or diluted in conventional linear and/or additive models, thus offering a complementary lens into the mechanistic pathways by which aging-related PRS exert their influence.

Thus, in an effort to examine the non-linear (including age-related) expression of genetic risk, we trained VAEs on a subset of the deeply phenotyped Arivale "scientific wellness" cohort of 1491 individuals and implemented two approaches: one in which we regressed out age from all omics features, and another that allowed the VAE to model potential PRS-age-omics interactions directly by including age as a covariate, along with binarized, age-stratified PRS features (see Methods). These VAE models incorporated between 75-110 PRS and 1,164 omics features, modeling how genetic risk interplays with putative age-related PRS biomarkers. We then systematically toggled binary PRS between high and low risk, thus quantifying the resulting changes across the multi-omics space as described above to infer expected multi-omics signals associated with differential risk.

In this work, we define "high-risk" PRS as those in the top decile. Although this cutoff sacrifices some granularity relative to analyzing the entire PRS distribution, many PRS-based analyses show their most convincing predictions in the extreme quantiles, with substantial uncertainty across the risk spectrum[28,29,30]. This pattern is partly explained by the modest explained variance of many phenotypes and the presence of potential nonlinearities [30,10,16]. Moreover, in the context of a neural network architecture, using a top-decile cutoff creates a more distinct "intervention" or "case," which helps mitigate the computational and overfitting challenges (while boosting the signal-to-noise ratio) that can arise with high-dimensional genomic features (i.e. millions of SNPs per PRS per individual). Although GWAS-derived PRS may still retain some biases (such as those due to residual population stratification) genetic variants themselves are fixed rather than influenced by ongoing confounders, making PRS a relatively objective measure of trait risk. Consequently, leveraging these "high-risk" PRS cutoffs for hypothesis generation is complementary to phenotyped, smaller, clinically ascertained cohorts with narrower ancestry adjustments or shallower phenotyping. Hence, actively toggling the highest polygenic burdens within a perturbable neural network provides an effective and efficient strategy for uncovering biological pathways in aging that purely linear or underpowered phenotype-based models might overlook.

To further evaluate the credibility of the PRS–omics associations revealed by these perturbations, we cross-referenced the genetic variants comprising each PRS for cis protein and expression QTL tags (eQTLs and pQTLs), and where strong genetic instruments were available, performed two-sample

Mendelian randomization in independent cohorts to assess causality. This provided additional mechanistic context, enabling more robust prioritization of biomarker leads, drug targets, and deeper exploration of age-related phenotypes.

Results

VAE Detects Substantially More Pleiotropic Associations Than Linear Regression

We applied a PRS-perturbed variational autoencoder (VAE) to a deeply phenotyped prodromal cohort to identify non-linear, pleiotropic genetic liability signals, benchmarking its performance against linear regression (Fig. 1). Given 75 aging-related PRS, the age-residualized ordinary least squares (OLS) model identified 326 PRS–omics associations after Benjamini-Hochberg (BH) false discovery rate (FDR) at $\alpha = 0.05$, versus 2,219 from the age-residualized VAE (FDR < 0.05 and ≥50% reproducibility; Fig. 1, see Methods and Supplementary Tables 1 and 2). PRS feature-harmonized age-included model analyses yielded 147 vs. 387 associations for OLS and VAE, respectively (Supplementary Tables 3 and 4). We note that the criteria for a reportable VAE association were deliberately stringent, requiring at least 50% reproducibility across 10 independent runs, each comprising 30 model refits (see Methods). This conservative threshold was imposed to mitigate redundancy arising from degeneracy among correlated multi-omic features that may occupy interchangeable positions in latent space (for example, HOMA-IR and fasting insulin). Moreover, the Bayesian decision-theoretic framework used to identify credible associations is itself grounded in aggregation across 30 refits and has been validated against ground-truth benchmarks[27]. When considering the full set of unique associations recovered across all VAE runs, the age-residualized VAE model yields approximately an 18-fold increase in hypothesis-generating associations relative to the corresponding OLS analysis. Overall, 73 unique associations were shared between OLS and VAE models, and these were enriched among the strongest linear associations. Specifically, replicated associations were overrepresented in the top 10% of linear hits ranked by FDR (Fisher's exact test, odds ratio = 2.68, $p = 0.029$). Three associations were supported across all four analyses: venous thromboembolism PRS with body mass index, hypertension PRS with mean arterial blood pressure, and a longevity PGS with surfactant protein D, a circulating pulmonary innate immune protein previously associated with all-cause mortality in elderly populations (Figure 1; Supplementary Tables 1–4)[31].

Chronological Age Reshapes the Genetic Risk Manifestation

To examine how supplying chronological age to the VAE alters PRS–omics associations, we focused on four age-related domains; namely, neurodegeneration (Alzheimer's disease [AD], Parkinson's disease [PD], and amyotrophic lateral sclerosis [ALS]), type 2 diabetes (T2D), cardiovascular disease (CVD), and longevity. We included multiple PRS per domain to maximize signal diversity (Figure 2; Supplementary Table 5). As shown in Figure 2A and B, adding age to the encoder led to three major shifts: (1) a 3.5-fold reduction in total significant associations compared to the age-residualized model, (2) a relative enrichment of clinical chemistry features in the association landscape, and (3) a 42-fold decline in T2D

PRS associations. Figure 2C further illustrates that the top ten most frequently observed associations in the age-included model are exclusively clinical chemistry traits.

Quantitatively, total associations fell from 1,178 (age-residualized) to 335 (age-included). This contraction was driven by metabolomics hits dropping from 688 (58.4 %) to 126 (37.6 %), proteomics from 303 (25.7 %) to 72 (21.5 %), and clinical chemistries from 187 (15.9 %) to 137 (40.9 %) of the total. The residualized model was enriched with fast-turnover metabolite associations, suggesting they capture genetic contributions independent of age. In contrast, when age is provided, the embedding shifts toward organ-level assays—cholesterol subfractions, liver enzymes, blood pressure metrics, and other clinical measures (Fig. 2C). The effect on T2D PRS was most pronounced: these yielded 293 associations without age, but only 6 when age was included (Fig. 2A,B).

Together, these results demonstrate that including chronological age reshapes both the scale and composition of detectable PRS–omics associations. The age-residualized model appears to detect a broad range of age-independent molecular signatures, whereas the age-included model emphasizes fewer, age-conditioned genetic effects in organ-level clinical assays.

Neurodegenerative diseases: shared pathways, age-related signals, and disease-specific signatures

Shared pathways. Across two independent AD PRS and one each for PD and ALS (Supplementary Table 5), we observed relatively convergent shifts in membrane lipid remodeling and oxidative, iron-related biology (Fig. 3). Changes in ether-linked GPE/GPC plasmalogens were associated with high genetic risk for all three domains, which is consistent with reports that plasmalogen dysregulation contributes to synaptic vulnerability in neurodegeneration [32–34]. We also observed consistent changes in the markers of oxidative stress and ferroptosis-related pathways. Inflammatory signaling (e.g., TNF, IL-12B) and a consistent imbalance of ω-6/ω-3 long-chain PUFAs were also observed. Finally, the gut microbiota-linked metabolites, blood-brain barrier integrity-promoting p-cresol glucuronide[35] and metabolically disrupting TMAO[36], were associated with high genetic risk for AD and ALS, respectively (Figs. 4 and 6).

Age-related signals. Age-stratified signals for genetic liability for neurodegeneration indicated when one might expect genetic risk to manifest. In PD, DHA was already lower in <45 year olds and LDL particle number was higher in the same age group, consistent with lipid remodeling in PD[37,38,39]. In AD, lower neuroprotective ω-3 and higher glucose were not evident until midlife (>45 y) in these high PRS carriers[40,41], whereas higher FGF19 and lower homocysteine in <45 y likely represented a compensatory adaptations and/or signatures of resilience[42,43](Fig. 4B). High polygenic risk for ALS associated with lower BMI and BUN/creatinine in >45 y, which aligns with epidemiology of premorbid leanness and renal indices[44,45,46].

Domain-specific signals

Because the number of significant VAE-generated PRS-omic associations was large and beyond the scope of detailed discussion here, we consolidate them in Supplementary Table 6, which includes

literature-based evidence and brief domain-specific commentary for each association to facilitate hypothesis generation. Below, we highlight a subset we deemed had the greatest relevance to established or emerging pathways in human aging pathways.

Alzheimer's disease. We captured over 100 associations that met the FDR threshold and our replicability criteria (see Methods). Of these, we found that heme oxygenase-1 (HMOX1), an eQTL in the APOE-independent late onset Alzheimer's disease (LOAD)-PRS[47] and key cytoprotective enzyme[48], was downregulated across both age-modeling frameworks (Fig. 4). Furthermore, two-sample Mendelian randomization (2SMR) analyses supported a causal effect of genetic AD liability on reduced circulating heme HMOX1 levels. The inverse-variance weighted (IVW) estimate indicated that higher AD genetic risk was associated with significantly lower HMOX1 concentrations ($\beta = -0.069$, SE = 0.023, $p = 0.0025$; Fig. 1S). The MR-Egger regression yielded a directionally consistent and highly significant effect ($\beta = -0.207$, SE = 0.053, $p = 2.7 \times 10^{-4}$; Supplementary Table 7), with a non-zero Egger intercept ($p = 0.0024$; ; Supplementary Table 8) suggesting the presence of directional pleiotropy but without attenuation of the main effect. Furthermore, the Steiger directionality test confirmed that the causal flow was likely from AD genetic liability to decreased HMOX1 rather than the reverse ($p = 2.7 \times 10^{-39}$)(Supplementary Table 9).

We also observed a consistent paradoxical positive association of genetic risk for LOAD with 25-OH vitamin D across age strata[49], though this is likely a pleiotropy-driven, compensatory signal captured by the age-modeled VAE (see Discussion). Along with this pattern, vascular physiology and age-invariant markers (i.e., those withstanding both age-residualized and age-included frameworks) were also consistently associated with AD genetic risk (Fig. 4A). The former included an age-related association with lower diastolic blood pressure, supporting cerebral hypoperfusion as part of dementia-related cognitive decline [50]. Age-invariant associations included higher levels of 3-aminoisobutyrate, CCL24, and carbonic anhydrase VA (a neurodegenerative disease target[51,52]), together with the aforementioned HMOX1 deficit, delineating a potentially core AD-PRS biochemical profile (Fig. 4). In sum, the above highlights a pleiotropy-driven expression of Alzheimer's genetic risk unfolding with age: midlife is characterized by worsening metabolic balance, a decline in vascular tone, and dysregulated immune and lipidemia effects, while earlier adulthood may show subtle compensation, with the convergent evidence from VAE modeling, eQTL colocalization, and 2SMR suggesting that genetically determined LOAD risk exerts a likely upstream influence on HMOX1 levels, consistent with a mechanistically relevant and potentially causal relationship implicating impaired antioxidant defense in AD pathophysiology.

Parkinson's disease. Three age-invariant features replicated across the two age-modeling approaches: α-L-iduronidase (IDUA)(also a eQTL and pQTL), 4-hydroxyphenylpyruvate (4-HPP), and picolinate (Fig. 5). 4-HPP—a key tyrosine-catabolism intermediate—was the strongest metabolite signal and was previously identified as a classifier in cohort-wide blood plasma metabolic profiling of PD patients[53]. Multiple lines of evidence support IDUA involvement in PD and neurodegeneration[54,55], and our data indicate that IDUA reduction may already be present in the prodromal stage among yet-undiagnosed individuals. Moreover,

upon 2S-MR analysis, we observed the genetically-proxied association between IDUA and PD previously reported by Belbasis et al. (see below)[56].

Amyotrophic lateral sclerosis. The ALS-PRS association landscape was characterized by iron-linked oxidative stress, inflammation, and energy–lipid remodeling (Fig. 6). Evidence for this signature is indeed found in the literature, with ferritin, high-sensitivity C-reactive protein (hsCRP), and myeloperoxidase (MPO) elevated in ALS pathology[57,58,59,60]. Furthermore, pharmacologic MPO inhibition has been investigated as a therapeutic[61]. We also observed lower carotene diols, a potential interaction with dietary carotenoid intake, which has been associated with reduced ALS risk[62]. The neurotrophic factors GDNF and artemin were increased with eQTL/pQTL support (Fig. 6), and have mechanistic links in ALS models, including early GDNF surges during denervation[63,64]. Finally, ferritin, MPO, TMAO, and selected cytokines persisted across age-residualized and unresidualized models (Supplementary Data).

Cardiometabolic risk expresses as age-conditioned T2D signals and immune-driven CVD features

Across four T2D PRS (Supplementary Table 5), we observed a broad metabolic signature in the age-residualized VAE, including lipotoxic lipids and depletion of insulin-sensitizing species (Fig. 7). Strikingly, when age was included, >95% of these associations disappeared (Figs. 2A,B and 7A), indicating that the molecular correlates of T2D genetic risk are largely conditioned on chronological age. Yet, within the cardiometabolic-longevity landscape, N-acetylmethionine (NAM) was detected as an age-invariant, cross-trait nexus: NAM was lower with both T2D and CVD PRS and ranks among top incident-CVD predictors in population metabolomics, and may causally protective with respect the T2D[65,66]. In contrast, NAM was higher with longevity PGS (see below). This opposing pattern implies that one-carbon/sulfur-amino-acid metabolism may represent a shared axis that may link cardiometabolic risk to healthspan.

In contrast to T2D, age-modeling of 4 independent CVD PRS (Supplementary Table 5) did not eliminate genetic risk–omic associations and yielded a diverse set of signals spanning immune regulation, canonical clinical markers, and lipid-metabolic pathways (Fig. 7B). The most pronounced was reduced SLAMF1 (its paralog SLAMF7 was also detected) in high-risk carriers, supported by cis-eQTL and pQTL evidence, which is consistent with diminished modulation of macrophage activation in face of cardiotoxicity[67]. Beyond this immune axis, multiple canonical clinical correlates of cardiovascular risk were reliably detected, including a biochemical fingerprint of arterial stiffening reflected in lower diastolic pressure coupled with higher pulse pressure[68] (Fig. 7B). Additional signals highlighted the interplay of lipid and energy metabolism, with dicarboxylic acids and ceramide species enriched in high-CVD risk individuals[69,70]. We also detected potential compensatory signals: lipoprotein lipase was elevated, a trait linked prospectively to lower future CAD incidence[71], and bile-acid handling shifted in a manner that reduced exposure to pro-atherogenic 12α-hydroxylated bile acids[72] (Fig. 7B). Immune regulation also showed signs of adaptive changes, with elevated IL-10 and TRAIL–DR5 signaling (TNFRSF10B), in both

instances suggesting atheroprotection in otherwise healthy individuals with high cardiovascular polygenic risk[72,73,74,75].

Finally, age-invariant CVD signals (Fig. 7C) defined a potentially core prodromal signature: consistent depletion of SLAMF1, reduced chemotactic and adhesion proteins, and loss of amino-acid–derived conjugates, including NAM. Overall, the VAE-derived cardiometabolic footprint indicates that T2D PRS manifest chiefly through age-related metabolic drift, whereas CVD PRS define a prodromal pro-inflammatory phenotype.

Longevity PRS Define an Anti-Inflammatory, Organ-Protective Multi-Omic Signature

Across five longevity-related PGS, the VAE uncovered a multi-omic signature that was suggestive of decelerated aging (Fig. 8). Individuals with higher longevity PGS showed lower chronic inflammation, including reduced hsCRP and broadly lower pro-inflammatory cytokines/chemokines, which is consistent with attenuated inflammaging[76]. Liver enzymes trended lower, with ALP as the highest effect-size and the most stable clinical correlate (Fig. 8). Independent logistic modeling outside the VAE showed lower ALP among those in the top-decile healthspan-PGS ($p = 0.014$), and each SD increase in ALP was associated with lower odds of being in that top-decile (Fig. 2S; $OR = 0.83$, $p = 0.019$), aligning with evidence that ALP tracks frailty/mortality[77]. Glycemic control was also more favorable (lower fasting glucose and HOMA-IR), and this is in line with superior insulin sensitivity seen in long-lived families and the link between rising mid-life HOMA-IR and subsequent mortality[78,79]. Longevity PGS associations with lipid features suggested an anti-inflammatory milieu, including higher long-chain ω-3 (DPA), a marker of lower mortality risk[80,81]. Finally, the aforementioned NAM, which showed a negative association with cardiometabolism-PRS, was elevated in individuals with genetic predisposition to longevity.

Organ-level landscape of longevity PGS associations

Recent plasma-proteome studies show that chronic diseases present as accelerated, organ-specific aging and that protein-based “organ clocks” predict the matching organ’s disease risk[82,83]. In the same vein, our multi-omic profiling with age-modeled VAE and longevity/healthspan PGS mapped onto organ-related signatures whose directions are broadly consistent with slower organ aging (Fig. 9). Of these, the liver axis was most clinically tractable: longevity PGS associated with lower ALT, AST, ALP, and hsCRP—patterns indicative of reduced hepatocellular and systemic inflammatory stress[83,84,85,77]; conjugated bile acids are also tracked with longevity PGS.

Renal indices showed nuanced, age- and PGS-dependent patterns (see Supplementary Tables 1 and 3). Creatinine associations differed by PGS feature; a positive association of BUN in individuals >45 years may reflect subtly higher protein turnover in later life (potential gene×diet interactions) rather than renal impairment[86]. Associations with electrolytes, uric acid, homocitrulline, and renin highlighted the endocrine–renal axis and aligned with reports around organ-aging proteomics frameworks in which

renin–angiotensin signaling contributes to renal risk[82,86,83]. Heart health features included lower systolic and mean arterial pressure, with pulse-pressure differences that depended on age stratum and PGS. Specifically, an age-stratified (>45 years) and age-residualized longevity PGS was associated with higher pulse pressure, whereas the healthspan PGS was associated with lower pulse pressure across the lifespan. However, interpretation of these divergent patterns requires caution, as lifespan- and longevity-related PGS aggregate genetic influences on survival that may reflect adaptive, compensatory, or pleiotropic mechanisms rather than uniformly protective cardiovascular effects. Thus, directionality in such associations may be counterintuitive and context-dependent, underscoring the distinction between genetic determinants of lifespan extension and those that preserve cardiovascular healthspan. Lipoprotein particles (LDL particle number, small LDL, HDL particles), too, showed PGS- and age-model–specific directions, consistent with non-linear lipid dynamics and the decline in lipoprotein functionality with aging and inflammation[83,87,88,89].

The immune milieu pointed to less inflammaging and better insulin sensitivity, with relatively lower pro-inflammatory chemokines/cytokines and more favorable indices of glycemia. Relatedly, adipose signaling was also represented within this favorable metabolic landscape via the adipokine serpin A12. Moreover, brain-relevant signals included positive associations for artemin (a GDNF-family ligand) and gastric intrinsic factor, which supports a neurotrophic and neuroprotective signature with maintained cobalamin absorption[89,90,91]. Finally, associations in matrix remodeling pathways indicated that genetically-determined longevity may involve modulation of extracellular turnover and connective-tissue maintenance.

Taken together, longevity/healthspan genetics are reflected in an organ-level, clinically interpretable profile of lower hepatic and systemic inflammation, endocrine–renal balance, more favorable hemodynamics, context-dependent lipid signaling, preserved glycemic metabolic control, and extracellular-matrix remodeling.

Cross-design triangulation in independent genetic datasets

Importantly, nearly all PRS–omic associations detected by the VAE are supported by existing literature, including causal-inference studies (Supplementary Table 6). That binary PRS-conditioned latent factors recover established disease–omic relationships is notable, especially given that the model received no clinical labels or outcome data. In effect, we detected associations, including relationships previously uncovered only through epidemiologic studies, animal experiments, or interventional trials, solely from polygenic risk and the pleiotropic structure learned within the prodromal latent space.

Given this concordance with prior literature evidence, we elected not to prioritize direct replication via construction of a second, independently trained unsupervised latent space in another observational cohort. While such replication is valuable in many contexts, unsupervised latent representations are known to be sensitive to domain shift, assay platform heterogeneity, and differences in covariate structure, which can complicate one-to-one alignment of latent factors across cohorts[92,93]. Instead, we

used 2SMR, employing external datasets to supply orthogonal, genetically anchored support of the VAE-derived associations. Here, instruments from OpenGWAS were used to model genetic liability for each disease and longevity proxy in bidirectional MR, treating these as exposures and genetically proxied omic features as outcomes; effects were estimated with IVW and MR-Egger assessed horizontal pleiotropy and directional consistency. To mirror the VAE design, multiple GWAS were queried per disease domain.

Across diseases, the forward MR (i.e. disease → omics) replicated several VAE associations, ranging from pleiotropic to likely causal signal. For AD, three independent exposures replicated a lipid signature in which HDL, LDL, omega-6, omega-3, linoleic acid, and sphingomyelins repeatedly showed nominal significance with respect to causality ($p_{IVW} < 0.05$) (Fig 3S-A; Supplementary Tables 7-9). Other features, such as vitamin D, growth hormone 1, and artemin were also significant (Fig 3S-A; Supplementary Tables 7-9). The aforementioned HMOX1 was significant in one AD exposure (Fig 1S; $p_{IVW} = 0.037$), providing directionally-consistent support for the oxidative-stress signal seen in the VAE as a result of genetic liability. Interestingly, MR detected the 3-amino-2-piperidone as a plausible causal signal in the reverse direction, suggesting that higher lifetime exposure may influence the risk of AD (Supplementary Table 7; $p_{IVW} = 0.00497$). Indeed, this analyte's involvement in AD pathology has precedent [94,95].

In PD, forward 2SMR analyses yielded no significant associations, with only DHA approaching the significance under one exposure (Supplementary Table 7). Notably, IDUA, detected by the VAE as a replicated, age-invariant feature and supported by QTL evidence, showed a significant reverse MR association (Supplementary Table 7; $p_{IVW} = 0.0048$), although limited instrument availability precluded MR-Egger sensitivity testing.. For ALS, forward MR replicated a plausibly causal relationship with lower BMI (Fig 3S-C; Supplementary Tables 7-9), whereas reverse MR suggested the same with respect to elevated omega-6, IL-13, IL-22, and artemin. CVD exposure association included n-acetylglycine, cholesterol, hexa-and-octadecadienedioate, pulse pressure, and glycohemoglobin A1c (Fig 3S-D; Supplementary Tables 7-9). For T2D, forward MR replicated a strong pulse pressure signal, 3-hydroxybutrate (BHBA), along with multiple significant pleiotropic associations (Fig 3S-E); reverse T2D MR detected multiple plausibly causal metabolite signals, as well as the protein fms-related tyrosine kinase 3 ligand (Supplementary Tables 7-9). Finally, lifespan and parental-longevity exposures implicated multiple LDL and blood pressure features, and albumin as the most robust features, with hsCRP and ALP also associating across datasets (Fig 3S-E; Supplementary Tables 7-9).

We provide a qualitative assessment of cross-design concordance between prodromal VAE-derived associations and genetically anchored effects estimated via 2SMR in Supplementary Table 10. This analysis is not intended as a rigorous causal replication, as precise cohort matching between VAE training data (a generally healthy, deeply phenotyped cohort) and disease or longevity GWAS is not feasible. Instead, this comparison is meant as a coarse-grained view of how early, subclinical biology captured by polygenic-risk perturbations relates to molecular signatures observed in overt disease or survival-based genetic studies. Across disease domains, replication rates ranged from 21–71%, with

coverage of VAE-identified associations by MR-testable instruments between 63–73% (Supplementary Table 10). Directional concordance was highest in AD (88%), suggesting a potentially closer alignment between prodromal molecular signatures and established disease biology in this domain. However, formally evaluating this hypothesis would require controlled, cohort-matched analyses beyond the scope of the present study. For longevity-related traits, directional concordance was not assessed. Across overlapping features between the longevity VAE and reverse Mendelian randomization analyses, effect directions were heterogeneous. Metabolomic and immune-related features frequently showed concordance between VAE Bayes factors and IVW estimates, whereas cardiometabolic traits (e.g., blood pressure, lipids, glucose) often exhibited discordant directions, consistent with compensatory physiology, survivorship bias, and selection effects inherent to longevity-focused cohorts (Supplementary Data).

Together, these results demonstrate that a perturbable-PRS VAE trained in a generally healthy cohort captures prodromal molecular signals that overlap meaningfully with genetically inferred disease biology, supporting its utility for early detection of pleiotropic and plausibly causal age-related pathways while highlighting potential distinctions between prodromal and overt disease states.

Discussion

This study demonstrates the integration of PRS with age- and context-aware deep learning to uncover early molecular signatures of differential disease risk in apparently healthy individuals, providing candidate biomarkers before symptom onset. Our chief innovation is the use of PRS as the sole genetic anchor to mine pleiotropy through the latent space of the VAE, and coupling this with dense multi-omic data taken across adult lifespan to evaluate the effects of genetics in the context of aging. This approach allows us to identify disease and biomarker relationships, previously uncovered only after years of observational studies, labor-intensive animal experiments, or costly clinical trials, and to propose serum analytes and potential drug targets that warrant further investigation. Our approach builds on the premise that PRS aggregate numerous small-effect variants into a meaningful proxy of inherited disease predisposition[96]. Prior studies have shown that PRS stratification can reveal molecular signals of disease pathogenesis years before overt illness[96,97]. For example, Ritchie et al. demonstrated in ~3,000 healthy adults that PRS for cardiometabolic diseases were associated with 49 plasma proteins, including 12 druggable targets, some mediating subsequent disease events [96]. Here, using a PRS-perturbed VAE framework, we extend this paradigm into the non-linear latent space of a VAE that simultaneously considers the entire multi-omics landscape presented to it, which allowed us to identify multivariate biomarker shifts linked to genetic risk and to highlight pathways that traditional linear regressions analyses or case-control designs miss [98,99]. Importantly, by leveraging multiple domain-specific PRS derived from large, independent, and often ancestry-diverse cohorts, our approach captures a broader spectrum of disease biology, enhances cross-phenotype pleiotropy detection, and reduces overfitting to any single discovery population. Importantly, this strategy is scalable to large biobank datasets and helps to address the challenges of time and costs of longitudinal phenotyping, since

genetic risk can serve as a surrogate exposure to future disease in currently healthy people[100]. Thus, we present an effective approach for early preventive biomarker discovery and therapeutic target prioritization in pre-diagnostic stages of common diseases represented in large-scale GWAS.

A key advantage of the VAE-based latent space modeling is its capacity to capture nonlinear, high-dimensional patterns of biomarker variation that reflect pleiotropy and age-dependent effects. Unlike linear association tests on one biomarker at a time, the VAE learns a compact representation of all molecular features, enabling it to detect subtle coordinated changes induced by PRS perturbations. This is analogous to recent work showing that a VAE trained on healthy transcriptomes could decode disease signatures, uncovering more disease-specific genes than conventional differential expression in most cases[101]. In our study, PRS perturbations in the latent space revealed multi-fold more significant complex biochemical fingerprints for each condition examined compared to OLS linear regressions. These fingerprints often involved pathways known to be pleiotropic and those reflecting the interplay of genetics with aging processes. For example, the latent factors modulated by AD PRS captured not only expected AD-related biomarkers but also signals of metabolic and vascular health, consistent with the idea that genetic risk to AD manifests through diverse, interconnected biological changes in midlife. Because a latent metabolic factor influenced by the PRS can drive systematic changes across many analytes, it can produce a relevant biological signature that would be fragmented or lost in a linear model, where each feature's independent change is estimated in isolation. Such nonlinear integration is especially useful in diseases like Alzheimer's or Parkinson's, where gene variants often have broad systemic effects that a DL model can assimilate into an at-risk profile. Furthermore, because the present VAE model calculates interactions with age, it can encode how the influence of PRS on biomarkers might intensify or attenuate as individuals get older, a feature linear models struggle to capture[102]. This allowed us to map coarse, age-dependent trajectories in the latent space, offering a window into how a disease risk in ostensibly healthy younger vs. older year old individuals might manifest across the multi-omic space. Thus, the VAE's latent space provides a rich manifold of prodromal disease biology, leveraging high-dimensional correlations to detect subtle disease transition patterns (including compensatory biomarker changes) across lifespan that would be invisible in a univariate analysis.

One striking observation was that explicitly modeling age in our framework markedly attenuated many PRS–biomarker associations, particularly for T2D traits. When the VAE was conditioned on age, the signals for the T2D PRS reduced dramatically. This contrast between the age-residualized and age-modeled VAEs indicates that otherwise healthy, high-risk individuals exhibit age-typical metabolic changes earlier in their life course; when chronological age is modeled, the shared variance is attributed to age, revealing that T2D genetic liability is predominantly age-expressed rather than age-independent. In other words, a high PRS for T2D may only translate into observable biomarker perturbations as aging-related processes set in. Indeed,this interpretation aligns with gerotherapeutic rationales for T2D drugs[103,104–106]. In this vein, Escott-Price et al. note that even for neurodegenerative disease PRS, the effect size depends on the age of the GWAS sample used and reflects risk at that particular age[107]. Our results align with this observation, as many PRS-associated biomarkers were age-modulated in the

present analysis. From a geroscience perspective, it suggests many risk biomarkers are not static hallmarks of inherited risk but dynamic features emerging from the interaction of genes with aging biology. Unraveling these gene–age interactions could be key to identifying when in the aging process interventions (such as lifestyle or geroprotective drugs) would most effectively neutralize genetic risk.

Crucially, our study focused on primarily middle-aged individuals who were generally free of overt disease, a design that differentiates our findings from most biomarker studies done in patients or elderly cohorts. By interrogating a healthier population, we were able to capture latent compensatory or early pathological changes rather than end-stage disease effects. For example, traditional case–control biomarker studies in T2D or AD often report markers that are consequences or late-stage correlates of disease, whereas our PRS-based approach in midlife uncovers subtler shifts that generally precede overt symptoms. These shifts can represent the body's compensatory responses to genetic risk. Consistent with this idea, several of our age-stratified observations suggest adaptive metabolic signaling in younger (<45) individuals that becomes pathological later in life. For example, higher FGF19 in the younger group aligns with its known insulin-sensitizing and glucose-lowering effects[108,109], potentially buffering against AD-related insulin resistance[110]. In contrast, the >45 AD risk group showed relatively higher glucose, a shift toward impaired glycemic control. Similarly, lower homocysteine was also particularly pronounced in younger genetic risk carriers and may reflect an adaptation previously observed in APOE4+ AD cohorts[43]. Lipid metabolism features also exhibited an age-related pattern, with reduced omega-3 and linoleic acid (omega-6) in older individuals. Reports indicate that omega-6 fatty acids are elevated in AD cases [111,112], and our 2SMR likewise detected a plausible causal relationship between AD and higher linoleic acid across multiple exposures. Therefore, the observed relative reduction here may reflect a prodromal attempt to normalize the omega-6:omega-3 balance before overt dysregulation emerges, with the above-mentioned adaptive changes hinting at a potential biochemical signature of resilience in individuals who remain cognitively intact despite genetic risk.

Another example of such adaptive or resilient expression of AD genetic liability was the consistently higher vitamin D levels observed across the age spectrum in individuals with high AD PRS. At face value, higher 25(OH)D is a marker of better overall health (additionally, vitamin D deficiency is linked to dementia risk[113]). Yet the fact that our AD-PRS group showed increased vitamin D across age groups despite vitamin D supplementation being residualized out of all the omics features, suggests a compensatory or protective response may be at play. Indeed, immune-related loci within the LOAD-PRS (e.g., CR1, HLA-DRB1, CD33, MS4A6A, INPP5D, SPI1/CELF1), includes HLA-DRB5/DRB1 variation that is in linkage disequilibrium with HLA-DRB1*15:01, which contains a vitamin-D response element, supporting a potential up-regulation of vitamin-D–mediated anti-inflammatory pathways in high-risk, AD-free individuals[47,114,115]. Moreover, other LOAD-PRS loci that affect lipid/HDL biology (ABCA7, CLU) can influence vitamin-D transport, and CLU is inducible by active vitamin D, providing another plausible mechanism for higher circulating 25-OH vitamin D despite genetic risk[116,117,118,119,120,121]. Additionally, while the PRS that associated with elevated 25-OH vitamin D in our cohort did not contain APOE-ε4, the latter has been associated with higher circulating vitamin D levels in humans, possibly reflecting an

evolutionary adaptation that confers some protection against vitamin D deficiency in at-risk genotypes[113,122,123]. Thus, our results suggest this protective buffering dynamic in genetically at-risk individuals may be independent of APOE-ε4 status. In a nonlinear framework like the VAE, we interpret such a shift as part of a pattern suggesting that what appears beneficial may actually be a harbinger of underlying risk dynamics and/or disease transition, as the at-risk individual may be mounting a defense against incipient AD pathology[124]. On the other hand, some PRS-associated changes likely represent early pathological deviations that are still within normal ranges in young adults. For instance, high cardiovascular disease PRS was associated with increased LPL levels in our cohort. Elevated LPL in midlife could be interpreted as a healthy sign, since high LPL activity is known to promote lipid clearance and is associated with lower future CAD risk[71]. However, it may also indicate that high-risk individuals are working harder to turn over lipids, thus revealing an early metabolic strain. Indeed, reduced LPL is a CVD risk factor, but a necessary upregulation of LPL might signal underlying hyperlipidemia pressure [125,126]. In other words, PRS perturbations likely mix protective adjustments and nascent pathological signals. Decoding these would require longitudinal follow-up. However, the VAE's ability to capture these complex patterns highlights its utility in interpreting the pleiotropic effects of genetic risk hypothesis generation.

Notably, several of the biomarker changes we identified align with emerging causal evidence from orthogonal methods like 2SMR, lending credence to their biological significance. One compelling example is HMOX1 in the context of Alzheimer's disease. Our model indicated that high AD PRS individuals have a relative deficit in serum HMOX1 levels. This enzyme, induced by oxidative stress, plays a cytoprotective role in the brain; indeed, functional HMOX1 gene polymorphisms have been linked to higher AD risk[127]. MR analyses in the literature support a protective effect of HMOX1 upregulation in neurodegeneration[127]. Thus, our finding suggests that genetically-at-risk apparently healthy individuals may live in a state of mild HMOX1 under-expression, rendering them more vulnerable to accumulating neural damage. Intriguingly, one modifiable way to boost HMOX1 is physical exercise, which activates Nrf2 and related stress-response pathways [128,129]. Indeed, the well-known cognitive benefits of exercise in APOE-ε4 carriers and others could be partly mediated by restoring such antioxidant defenses[130]. This raises the possibility that exercise or Nrf2-activating compounds might offset AD genetic risk in part by normalizing HMOX1 levels in midlife, a testable hypothesis emerging from our PRS perturbation results [127]. Moreover, widely used drug classes such as statins and selective serotonin reuptake inhibitors (SSRIs) have been shown to upregulate HMOX1 expression, primarily via activation of the Nrf2/HMOX1 antioxidant pathway [131,132]. This finding suggests a potential drug-repurposing opportunity to boost HMOX1's cytoprotective effects in genetically at-risk individuals.

In the Parkinson's disease realm, we found alterations in IDUA, a lysosomal enzyme, among high PRS individuals. Although IDUA is classically known for its role in Hurler syndrome [133], subtle lifelong insufficiency of lysosomal enzymes has been hypothesized to increase PD susceptibility [55,134]. In fact, PD GWAS loci like TMEM175 and RAB29 converge functionally on lysosomal activity, and notably, those

same loci have been associated with plasma IDUA levels [135]. Further, independent multi-omics analyses have linked lower circulating IDUA to higher PD risk[136]. Our results suggest high PD-PRS individuals harbor a latent lysosomal dysfunction signature in the form of lower serum IDUA activity, which may impair lysosomal clearance and drives neuronal glycosaminoglycan accumulation (rescued by CNS enzyme delivery)[54,137,55], and glycosaminoglycans promote α-synuclein aggregation and spread[55,138]. Thus, given our apparently healthy cohort, IDUA augmentation (e.g., via chaperone therapies) warrants evaluation as a pre-symptomatic, prophylactic strategy to delay or prevent PD onset. Together, the HMOX1 and IDUA insights illustrate how in silico PRS perturbation can pinpoint concrete molecular deficits that both explain genetic risk and point to interventions before clinical disease onset.

In addition to the plausibly causal HMOX1 and IDUA associations, each disease area we examined yielded notable PRS-omics associations. In AD, we detected disruptions in lipid and immune markers; in Parkinson's, we observed changes in tyrosine metabolism, anti-inflammatory lipids, and broad metabolism effects that are only beginning to be recognized as features of PD[139,140]. For ALS, ferritin and myeloperoxidase emerged as dominant age-invariant markers. This pattern points to iron-driven oxidative stress and systemic ferro-inflammatory remodeling in high-PRS individuals[141]. The T2D PRS, even though largely quiescent relative to age-drift, still showed a rich baseline signature of lipid dysregulation and immune system related perturbation. Notably, NAM emerged as a consistently replicated biomarker, surviving both age-stratified and age-residualized models and multiple independent PRS as a robust signal of compromised cardio-metabolic state. NAM levels were inversely associated with polygenic risk for T2D and CVD, and positively associated with genetic propensity for longevity, with a consistent direction of effect. Moreover, NAM was also among the top features detected as significant in our independent T2D→omics 2SMR analysis. This robust association suggests that NAM may capture a fundamental node linking genetic liability for metabolic risk and healthy lifespan. Biologically, NAM is an acetylated derivative of the essential sulfur-containing amino acid methionine, placing it in one-carbon and sulfur amino acid metabolism pathways[142]. Perturbations in these pathways are known to influence aging and metabolic homeostasis; for example, restricting dietary methionine improves insulin sensitivity and extends lifespan in model organisms [143,144,145,146]. The persistence of NAM as a protective biomarker across both our VAE-age frameworks involving T2D PRS may further the geroscience hypothesis, where this age-invariant metabolite associated with reduced metabolic risk and increased longevity highlights a potential biochemical link between insulin resistance pathways and the fundamental biology of aging, reinforcing the concept that targeting such conserved pathways could yield benefits for both metabolic health and healthy aging. Mechanistically, proposed gerotherapeutics such as metformin perturb one-carbon metabolism, causing accumulation of folate-cycle intermediates and depletion of glutathione, and effectively acting as an antifolate agent that secondarily activates AMPK and inhibits mTOR[147]. Conversely, mTOR activity itself promotes one-carbon flux by upregulating folate-cycle and serine-synthesis enzymes, while rising levels of the methyl donor S-adenosylmethionine feed back to reinforce mTORC1 signaling[148]. NAM's positioning at this nexus of nutrient sensing, redox balance, and one-carbon–driven anabolic signaling could thus underlie its association with reduced

cardio-metabolic disease risk and increased longevity, raising the possibility that NAM might serve as a potentially novel non-invasive, and dynamic read-out of gerotherapeutic efficacy.

In the CVD PRS omics association landscape, a consistent inverse association of serum SLAMF receptor levels with independent PRS emerged in our analysis. In macrophages, these receptors serve as immune checkpoints that can dampen excessive activation[149]. The pronounced reduction of SLAMF1/7 in high-PRS individuals therefore points to an immunoregulatory deficit aligned with genetic risk. This results aligns with a growing body of literature positioning defective inflammation resolution as a driver of atherosclerotic disease progression[150]. Moreover, in animal models, SLAMF7 deficiency exacerbates inflammatory injury[151], suggesting that loss of this checkpoint may heighten vulnerability to cardiovascular inflammation even before overt disease onset[152]. Thus, given that we observed this association in a generally healthy cohort, it is conceivable that augmenting SLAMF1-SLAMF7 pathways (or tracking their decline) in a subset of high-PRS individuals could help restore immune equilibrium in early disease, highlighting a potentially novel avenue for intervention or risk stratification. Finally, in longevity, our analysis of a longevity PRS revealed a pattern of youthful metabolic biomarker profiles, including lower HOMA-IR, glucose, and blood pressure as well as lower inflammatory cytokines This reinforced that genetic propensity for exceptional longevity correlates with a healthful biomarker milieu[153]. Moreover, we identified a complex neuroprotective signature. Indeed synaptic function has recently been identified as the strongest genetic driver of a functional biological age metric recently developed by Libert et al.[154]. Finally, we identified a commonly assayed liver enzyme, ALP, which correlates with mortality risk in humans[77,155], as an age-invariant, robust, and inexpensive biomarker whose level should be monitored across the lifespan toward healthspan optimization.

This study has several limitations. First, the cross-sectional design of the cohort used herein restricts our ability to make definitive claims about temporal dynamics or aging trajectories. While we observe age-dependent effects in the VAE latent space, these could in part reflect cohort effects, where older and younger individuals may differ due to lifestyle, environment, or cultural influences rather than true biological aging. Thus, longitudinal confirmation in future studies would strengthen our interpretation around age-related effects of genetic risk manifestation. Second, the sample size of ~1,500 individuals where we had deep multi-omics is modest compared to large-scale population biobanks, which may constrain the generalizability of our findings. Third, like many deep learning models, the VAE's interpretability is inherently limited. Although this is partially addressed by validating detected signals with orthogonal approaches such as 2SMR, examining cis-pQTL-supported analytes, and surveying supportive literature, our inferences remain probabilistic and hypothesis-generating rather than confirmatory.

Nonetheless, the VAE framework confers several strengths that mitigate several of the limitations. Its capacity to model high-dimensional, nonlinear interactions enables the extraction of age-related latent patterns that traditional linear or case-control models miss. By incorporating a Gaussian prior, the VAE effectively expands the modeled data space and generalizes beyond the training set, as reflected in the

concordance of our results with existing reports. Moreover, the use of PRS derived from large, diverse GWAS facilitates potentially broader, biologically relevant axes of inherited risk. Operationally, this perturbable VAE functions as a simple “digital twin,” simulating counterfactual increases in genetic liability and reading out predicted multi-omic responses. As such, one would not expect identical latent structure across cohorts, but rather mechanistically consistent patterns. Thus, in choosing 2SMR over re-fitting the VAE in another cross-sectional cohort, we favored a cross-design triangulation that leverages germline instruments fixed at conception and largely independent of cohort-specific behaviors and thus less susceptible to confounding. Indeed, agreement across orthogonal designs strengthens credibility more than agreement across two observational VAEs that may share similar biases. By contrast, a cross-cohort VAE would mainly test portability of an unsupervised latent geometry under domain shift, unaligned covariates, and analyte/genotyping platform differences.

We further note that only a subset of VAE-prioritized features had suitable genetic instrument proxies, so null MR results should be interpreted cautiously (i.e., limited coverage or weak-instrument bias), and thus we do not regard the absence of MR replication as necessarily evidence against a biologically plausible relationship. Furthermore, the VAE was successful in detecting biomarkers in a prodromal cohort of modest size. In contrast, 2SMR leverages case–control disease liability from large GWAS, thus one-to-one biomarker replication is not expected; moreover, our age-included VAE model could not be accommodated in 2SMR. Thus, the associations supported by both VAE and 2SMR likely reflect core pathological processes, whereas VAE-only signals may be highlighting adaptive, age-related, compensatory, or resilience-related biomarker signatures prior to clinical disease manifestation. Because these associations arise from nonlinear structure in the latent space, many of their effect sizes may not map directly onto conventional clinical interpretation, particularly given that many analytes lack well-defined healthy reference ranges and are typically evaluated using linear, context-agnostic frameworks. However, such VAE signals (particularly those derived from the age-included model) may serve as candidate markers for future N-of-1, longitudinal tracking approaches using standardized and sensitive assays, potentially enabling the detection of subtle deviations from an individual’s healthy physiological baselines. Thus, the convergence of VAE-derived signals with genetic evidence across multiple domains provides additional support that the model is surfacing biologically meaningful omic features of prodromal pathology. These triangulated signals prioritize a set of translatable biomarkers and pathways for mechanistic follow-up. Finally, we do not regard our approach as a substitute for randomized controlled trials, but as an efficient and scalable means for generating context-aware hypotheses about age-related disease biology and potential intervention points. The results presented here can thus inform prioritization of candidate biomarkers and potential therapeutic targets for future validation in longitudinal studies or experimental settings.

Methods

Polygenic risk scores (PRS)

PRS sourcing and construction. PRS were calculated for a broad panel of complex traits and diseases using whole-genome sequencing (WGS) data from the Arivale cohort. Most PRS weights were obtained from the PGS Catalog (https://www.pgscatalog.org), using builds based on GRCh37. For traits not represented in the catalog or requiring updated GWAS statistics, we derived scores using the PRS-CS(x) methods (depending on cross-ancestry summary statistics availability), which leverages continuous shrinkage priors and linkage disequilibrium (LD) reference panels to infer posterior effect sizes across ancestries [156,157]. Summary statistics were sourced from the GWAS Catalog and processed according to PRS-CS(x) specifications, and the 1000 Genomes Project Phase 3 reference panel was used to estimate linkage disequilibrium (LD) structure in PRS-CS(x) modeling. In all cases, we chose PRS or summary statistics based on the largest and most diverse available cohorts. All PRS calculations were performed using the pgsc_calc Nextflow pipeline (https://github.com/PGScatalog/pgsc_calc), applied without the built-in ancestry correction. Following score generation, we performed post hoc residualization of PRS values on the top nine ancestry principal components (PCs) to account for population stratification. PRS distributions showing deviation from Hardy–Weinberg equilibrium were excluded prior to score retention. For integration into the VAE framework, raw continuous PRS were binarized such that individuals in the top 10th percentile were assigned a value of 1, and all others were assigned a 0. In VAE models incorporating age effects, PRS were further stratified into age-defined groups using a threshold of 45 years. To maintain computational tractability, we applied a single age threshold, choosing 45 years as the cut-off given consistent evidence that biological aging inflection points occur near this age [102,158]. For each PRS, additional binary features were generated to reflect high-risk individuals aged ≤45 or >45, respectively. These stratified PRS features were included alongside the full-spectrum (age-agnostic) PRS variable. Age-only binary indicators (<45 or >45) were also introduced into the model to support the disentangling of age- versus PRS-driven variance in the latent space. **Age-related domain PRS**. In an attempt to map out heterogeneous genetic architectures and to detect a diverse set of signals for representative age-related traits, we combined multiple PRS per domain where available. For AD, we perturbed two complementary AD PRS: one derived from FinnGen (N ≈ 306 K) that stratified European-ancestry individuals with an OR of 1.64 for AD (independent of first- or second-degree family history)[159] and a family-based genetic risk score from a LOAD cohort (OR 1.29 in Europeans; independent of APOE ε4)[47,160]. For PD, we used a PRS derived from the largest multi-ancestry meta-GWAS to date[161](PGS000903), which aggregated data from ~1.47M individuals across diverse cohorts. For ALS, we used a PRS constructed as part of our previous multi-omic study, derived from large discovery cohorts[97,162]. For T2D, we combined four complementary PRS (PGS000712, PGS000850, PGS000851, PGS004923) to capture distinct aspects of genetic risk, including two scores (PGS000850 and PGS000851) built from SNPs weighted by their effects on insulin sensitivity measures and for first-phase insulin response and fasting pro-insulin[163]. For CVD, we combined four large-scale PRS that capture complementary aspects of genetic liability, including heart failure (PGS001790), CAD (PGS004941), coronary atherosclerosis (PGS002048), and a broad CVD score (PGS002316), totalling nearly 13 million individuals in aggregate. Finally, for longevity, we combined one PGS Catalog PGS (PGS000906)[164] with four additional genome-wide PRS we constructed using PRS-CS from large-scale GWAS meta-analyses.

These included the multivariate genomic scan of healthspan, parental lifespan, and longevity[7], and the genomic SEM–derived multivariate aging factor mvAge[7,23]. The combined sample size of these GWAS cohorts exceeded 2 million individuals.

Multi-omics preprocessing of Arivale cohort

We leveraged data from the Arivale wellness cohort, a consumer "scientific wellness" program that combined personalized lifestyle coaching with dense, longitudinal multi-omic profiling of participants, described in detail elsewhere [165]. These individuals were generally healthy adults (mean age ~48 years, ~59% female) drawn predominantly from the U.S. Pacific Northwest. Multi-omics data included untargeted plasma metabolomics, serum proteomics, and standard clinical chemistries and biometric measurements. Features with >10% missingness were excluded from analysis. For linear modeling applications (e.g., regressions), the remaining missing values were imputed using median-based imputation. No imputation was applied prior to variational autoencoder (VAE) training, as the MOVE framework is designed to natively handle incomplete data matrices [27]. All omics features were adjusted for known technical and biological confounders, including batch effects, vendor platform, season of sample collection, and U.S. state of residence. Additional covariates included the first eight ancestry principal components (PCs), sex, and a curated set of self-reported medication and supplement use variables: cholesterol, blood sugar, blood pressure, thyroid, and psychoactive medications, as well as fish oil, supplemental vitamin D, vitamin B12, multivitamin, and probiotic intake. Age was either included as a covariate (for age-residualized models) or retained as a stratification variable (for age-included models), depending on downstream application. All features were log-transformed (where applicable), z-score standardized across participants, and winsorized to the 0.1st and 99.9th percentiles to reduce the influence of outliers while preserving biological signal. After all preprocessing steps, the final dataset comprised 1,491 individuals with 274 proteomic features, 834 metabolomic features, and 51 clinical chemistry or biometric features retained for modeling.

PRS-omics linear regression modeling

To quantify the linear associations between PRS and individual omics features, we implemented three complementary regression strategies. All omics datasets were first preprocessed as described above. For each of the three omics layers—metabolomics, proteomics, and clinical chemistries—ordinary least squares (OLS) regression models were then used to assess associations with all PRS simultaneously. Analyses were performed separately across three age strata: individuals aged <45, ≥45, and the full cohort. In the first model, omics features were pre-residualized for age (alongside other covariates described above), and regressions were performed on the age-adjusted values without age included as a covariate. In the second model, no residualization was applied; instead, age was excluded as a predictor but used to stratify participants into the aforementioned age groups. In the third model, we explicitly modeled PRS × age interaction effects by including both age and multiplicative interaction terms (PRS × age) as predictors for each omic feature. In all three approaches, results were aggregated into a meta-regression framework, and FDR correction was applied using the Benjamini–Hochberg method ($\alpha = 0.05$).

The BH method remains valid under correlated tests, though it can be conservative in such settings. For the interaction model, both main and interaction term P-values were independently adjusted.

PRS-omics VAE modeling

To identify nonlinear and pleiotropic associations between genetic risk and multi-omic features, we adapted the VAE framework from the MOVE pipeline (https://github.com/RasmussenLab/MOVE)[27]. Input data included above-described standardized continuous features from metabolomics, proteomics, and clinical parameters, alongside binarized PRS (see above), which were treated as categorical variables for perturbation-based analysis. Model training was configured using the MOVE reconstruction.yaml template. The encoder contained a single hidden layer of 2,000 neurons, and the latent space dimensionality ranged from 200 to 300 units. A β-VAE architecture was used with a β value of 1×10^{-4} to modulate the Kullback–Leibler divergence (KLD) penalty. The model was trained for up to 1,000 epochs using the Adam optimizer (learning rate = 1×10^{-4}), a dropout rate of 0.1, early stopping (patience = 10), and warm-up steps for both KLD and batch dilation (each set to 100 epochs). Reconstruction accuracy with optimized hyperparameters exceeded 95% in the training set and 80% in the hold-out set, indicating sufficiently strong fidelity while reducing overfitting. To detect significant associations between PRS perturbations and omic features, we used the Bayes decision-theoretic framework implemented in MOVE [27]. Rather than relying on thresholded reconstruction errors, this method uses posterior predictive distributions to quantify perturbation effects. For each PRS, we performed 10 independent training runs (on top of the 30 model refits built into the MOVE pipeline) and, to balance discovery vs. reliability, reported associations that replicated in at least 50% of models. This replication criterion was used for interpretability; features not meeting this threshold were not necessarily discarded as false positives, as VAE models may substitute functionally redundant features. To elucidate mechanistic insights within a subset of PRS-multiomics associations, we intersected the proteomics associations for respective PRS variant sets with QTLbase v2.0 toward supplying cis eQTL and pQTL support for the associations [166]. To further aid interpretability and confirm model fidelity, known high-R^2 PRS (i.e., Total Bilirubin, LDL cholesterol, and HDL cholesterol) were used as positive controls (see below). Bayes-K values were normalized within each model run to standardize effect size reporting across different models and latent spaces. Latent space projections were visualized by projecting individuals into the 2D space using built in t-sne dimensionality reduction applied to the latent vectors. Feature contributions to the latent space were assessed using a perturbation-based attribution method. For each omic variable, its value was perturbed (set to zero for continuous features, or altered for categorical ones), and the resulting shift in latent space embedding was measured. The magnitude of this shift was used to quantify feature importance [27]. These analyses were used to compare the primary omic drivers across models trained on age-residualized versus age-stratified data.

Benchmarking PRS and VAE models

As described above, we applied OLS regression to benchmark PRS–multi-omics associations across 75 traits, using covariate-residualized and z-scaled omics data to estimate the variance explained by each PRS and to assess linear effect sizes for traits with available phenotypic data (e.g. serum markers,

adiposity measures, etc.) in the Arivale cohort subset . Since the VAE models were unsupervised, the linear regression results informed high-heritability PRS which we deemed as suitable “positive controls” to include in these DL models to help anchor the effect sizes of any pleiotropic associations with these expected canonical associations. The top explained variance features were Direct Bilirubin, HDL, and LDL, whose corresponding PRS explained, respectively, 32.4%, 17.4%, and 17.7% of variance. Both the age-residualized and age-stratified VAE models reliably detected the canonical PRS-biomarker associations with 100% reproducibility, with these non-pleiotropic associations also being among the highest effect size associations (Supplementary Table 1 and 3). Moreover, to ascertain that the signal detected via age-stratified PRS features included in the VAE model was in fact primarily due to the genetic-risk-age interaction and not the age component itself, we also included the isolated age features as part of the perturbable PRS inputs (i.e., Age<45:0→1 and Age>45:0→1). Of a total of 5890 associations, the age features associated with 18 omics features that represented expected trends such as increased levels of the “forever chemical” perfluorooctanesulfonate (PFOS), growth differentiation factor (GDF15), and lower steroid metabolites such as dehydroepiandrosterone sulfate (DHEA-S) with the > 45 age feature (Supplementary Table 3). The identified omics features that associated with these perturbed age features were subsequently removed as confounders from age-stratified PRS-biomarker discovery analyses described below. Moreover, to facilitate mechanistic explainability of the resulting associations, we mapped the PRS-proteomics feature associations against QTLBase2 [166]. Of a total of 892 proteomics associations, 63.3% mapped to a cis eQTL, with the Levene test for homogeneity of variance suggesting that eQTL-mapped features exhibiting suggestively lower variance in the normalized Bayes factors across 10 replicate runs (Figure 1S; $p = 0.071$, $F(1, 890) = 3.26$).

Literature validation and Mendelian randomization analysis

To evaluate whether associations identified by the VAE reflected biologically meaningful relationships rather than noise, we first conducted a structured literature validation step in which each PRS–omic association was systematically cross-referenced against PubMed-indexed experimental, genetic, or epidemiologic studies using a standardized, batched search protocol (ChatGPT Deep Research). This procedure generated evidence summaries for hundreds of associations and enabled domain-level contextualization before any downstream causal inference (see Supplementary Table 6). We then performed bidirectional two-sample Mendelian randomization (2SMR) analyses using the TwoSampleMR R package (v0.5.6) and the MR-Base platform, contingent on the availability of suitable genetic instruments. For each polygenic risk score (PRS) trait (exposure), genome-wide association study (GWAS) summary statistics were obtained from the GWAS Catalog or the IEU OpenGWAS repository, prioritizing large-scale meta-analyses. Single-nucleotide polymorphisms (SNPs) associated with the exposure at genome-wide significance ($P < 5 \times 10^{-8}$) were LD-clumped ($r^2 < 0.001$, 10,000 kb window) to ensure independence, and only exposures with ≥3 independent instruments were retained. Causal estimates were derived primarily using inverse-variance weighted (IVW) regression, with MR-Egger regression used as a sensitivity analysis to assess potential horizontal pleiotropy. For each significant or suggestive IVW result, the directionality and stability of the association were evaluated by comparing the IVW and MR-Egger slopes, focusing on consistency in the sign of effect estimates rather

than MR-Egger P values, which are typically underpowered for moderate pleiotropy. Reverse MR was performed to evaluate whether the observed omic feature could plausibly act upstream of the disease trait. Because the VAE-based associations are pleiotropic in nature and omic outcomes are often highly correlated (e.g., within metabolite or protein families), we did not apply global multiple-testing correction across all omic features. Instead, for each PRS–omic pair, we selected a single representative outcome (based on the strongest instrument set or F-statistic) within each correlated analyte family and interpreted P values in a relative, biologically informed framework. Associations were classified as (i) causal or directional when IVW estimates were significant or suggestive ($P \leq 0.05$ or ≤ 0.1) and directionally consistent with MR-Egger, and (ii) pleiotropic validations when IVW effects were significant or suggestive but showed evidence of heterogeneity, bidirectionality, or MR-Egger disagreement. Collectively, these analyses were intended to support the robustness and biological relevance of the VAE-identified cross-omic relationships, rather than to claim strict causality.

References

1. vB Hjelmborg, J. *et al.* Genetic influence on human lifespan and longevity. *Hum Genet* **119**, 312–321 (2006).
2. Nguyen, X.-M. T. *et al.* Impact of 8 lifestyle factors on mortality and life expectancy among United States veterans: The Million Veteran Program. *Am J Clin Nutr* **119**, 127–135 (2024).
3. Hruby, A. J. *et al.* Beyond genes and environment: mapping biological stochasticity in aging. *Geroscience* **47**, 2835–2850 (2025).
4. Gladyshev, V. N. Aging: progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. *Aging Cell* **15**, 594–602 (2016).
5. Tong, H. *et al.* Quantifying the stochastic component of epigenetic aging. *Nat Aging* **4**, 886–901 (2024).
6. Meyer, D. H. & Schumacher, B. Aging clocks based on accumulating stochastic variation. *Nat Aging* **4**, 871–885 (2024).
7. Timmers, P. R. H. J., Wilson, J. F., Joshi, P. K. & Deelen, J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat Commun* **11**, 3570 (2020).
8. Timmers, P. R. *et al.* Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife* **8**, (2019).
9. Argentieri, M. A. *et al.* Integrating the environmental and genetic architectures of aging and mortality. *Nat Med* **31**, 1016–1025 (2025).
10. Nolte, I. M. *et al.* Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *Eur J Hum Genet* **25**, 877–885 (2017).
11. Zhao, B. & Zou, F. On polygenic risk scores for complex traits prediction. *Biometrics* **78**, 499–511 (2022).

12. Burgess, S. *et al.* Using genetic association data to guide drug discovery and development: Review of methods and applications. *Am J Hum Genet* **110**, 195–214 (2023).
13. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *Am J Hum Genet* **110**, 179–194 (2023).
14. Namba, S. *et al.* A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. *Cell Genom* **2**, 100190 (2022).
15. Fang, G. *et al.* Discovering genetic interactions bridging pathways in genome-wide association studies. *Nature Communications* **10**, 1–18 (2019).
16. Elgart, M. *et al.* Non-linear machine learning models incorporating SNPs and PRS improve polygenic prediction in diverse human populations. *Commun Biol* **5**, 856 (2022).
17. Peng, J. *et al.* DeepRisk: A deep learning approach for genome-wide assessment of common disease risk. *Fundam Res* **4**, 752–760 (2024).
18. Kim, S. B., Kang, J. H., Cheon, M., Kim, D. J. & Lee, B.-C. Stacked neural network for predicting polygenic risk score. *Scientific Reports* **14**, 1–15 (2024).
19. Gunn, S. *et al.* Distribution of 54 polygenic risk scores for common diseases in long lived individuals and their offspring. *Geroscience* **44**, 719–729 (2022).
20. Widen, E., Lello, L., Raben, T. G., Tellier, L. C. A. M. & Hsu, S. D. H. Polygenic Health Index, General Health, and Pleiotropy: Sibling Analysis and Disease Risk Reduction. *Scientific Reports* **12**, 1–13 (2022).
21. Jabalameli, M. R. *et al.* Polygenic prediction of human longevity on the supposition of pervasive pleiotropy. *Scientific Reports* **14**, 1–19 (2024).
22. Van der Auwera, S. *et al.* The genetic predisposition to longevity acts through behavioral phenotypes in females. *Eur Neuropsychopharmacol* **45**, 1–14 (2021).
23. Rosoff, D. B. *et al.* Multivariate genome-wide analysis of aging-related traits identifies novel loci and new drug targets for healthy aging. *Nat Aging* **3**, 1020–1035 (2023).
24. Gomari, D. P. *et al.* Variational autoencoders learn transferrable representations of metabolomics data. *Commun Biol* **5**, 645 (2022).
25. Doncevic, D. & Herrmann, C. Biologically informed variational autoencoders allow predictive modeling of genetic and drug-induced perturbations. *Bioinformatics* **39**, (2023).
26. Holm, N. N. *et al.* amVAE: Age-aware multimorbidity clustering using variational autoencoders. *Comput Biol Med* **186**, 109632 (2025).
27. Allesøe, R. L. *et al.* Discovery of drug-omics associations in type 2 diabetes with generative deep-learning models. *Nat Biotechnol* **41**, 399–408 (2023).
28. Lopez, R., Boyeau, P., Yosef, N., Jordan, M. I. & Regier, J. Decision-Making with Auto-Encoding Variational Bayes. (2020).
29. Ding, Y. *et al.* Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nat Genet* **54**, 30–39 (2022).

30. Lewis, C. M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Med* **9**, 96 (2017).
31. Wulf-Johansson, H. *et al.* Circulating surfactant protein D is associated to mortality in elderly women: a twin study. *Immunobiology* **218**, 712–717 (2013).
32. Lerche, S. *et al.* CSF d18:1 sphingolipid species in Parkinson disease and dementia with Lewy bodies with and without GBA1 variants. *npj Parkinson's Disease* **10**, 1–6 (2024).
33. Dorninger, F., Forss-Petter, S. & Berger, J. From peroxisomal disorders to common neurodegenerative diseases – the role of ether phospholipids in the nervous system. *FEBS Letters* **591**, 2761–2788 (2017).
34. Plasmalogens: A potential therapeutic target for neurodegenerative and cardiometabolic disease. *Progress in Lipid Research* **74**, 186–195 (2019).
35. Stachulski, A. V., Knausenberger, T. B.-A., Shah, S. N., Hoyles, L. & McArthur, S. A host-gut microbial amino acid co-metabolite, -cresol glucuronide, promotes blood-brain barrier integrity. *Tissue Barriers* **11**, 2073175 (2023).
36. Chen, L., Chen, Y., Zhao, M., Zheng, L. & Fan, D. Changes in the concentrations of trimethylamine N-oxide (TMAO) and its precursors in patients with amyotrophic lateral sclerosis. *Scientific Reports* **10**, 1–8 (2020).
37. Alves, B. da S., Schimith, L. E., da Cunha, A. B., Dora, C. L. & Hort, M. A. Omega-3 polyunsaturated fatty acids and Parkinson's disease: A systematic review of animal studies. *J Neurochem* **168**, 1655–1683 (2024).
38. Park, J.-H. *et al.* Association of High-Density Lipoprotein Cholesterol Variability and the Risk of Developing Parkinson Disease. *Neurology* **96**, e1391–e1401 (2021).
39. Huang, X. *et al.* Lower low-density lipoprotein cholesterol levels are associated with Parkinson's disease. *Mov Disord* **22**, 377–381 (2007).
40. Cole, G. M., Ma, Q.-L. & Frautschy, S. A. Omega-3 fatty acids and dementia. *Prostaglandins Leukot Essent Fatty Acids* **81**, 213–221 (2009).
41. Crane, P. K. *et al.* Glucose levels and risk of dementia. *N Engl J Med* **369**, 540–548 (2013).
42. Tang, M. *et al.* Decreased FGF19 and FGF21: possible underlying common pathogenic mechanism of metabolic and cognitive dysregulation in depression. *Front Neurosci* **17**, 1165443 (2023).
43. Ravaglia, G. *et al.* Apolipoprotein E e4 allele affects risk of hyperhomocysteinemia in the elderly. *Am J Clin Nutr* **84**, 1473–1480 (2006).
44. Goutman, S. A. *et al.* Body mass index associates with amyotrophic lateral sclerosis survival and metabolomic profiles. *Muscle Nerve* **67**, 208–216 (2023).
45. Duan Y. & Fan D. [Association of serum uric acid level with amyotrophic lateral sclerosis]. *Zhonghua Yi Xue Za Zhi* **96**, 207–209 (2016).
46. Chiò, A. *et al.* Amyotrophic lateral sclerosis outcome measures and the role of albumin and creatinine: a population-based study. *JAMA Neurol* **71**, 1134–1142 (2014).

47. Tosto, G. *et al.* Polygenic risk scores in familial Alzheimer disease. *Neurology* **88**, 1180–1186 (2017).
48. Origassa, C. S. T. & Câmara, N. O. S. Cytoprotective role of heme oxygenase-1 and heme degradation derived end products in liver injury. *World J Hepatol* **5**, 541–549 (2013).
49. Gallagher, J. C. Vitamin D and aging. *Endocrinol Metab Clin North Am* **42**, 319–332 (2013).
50. Wolters, F. J. *et al.* Cerebral Perfusion and the Risk of Dementia: A Population-Based Study. *Circulation* **136**, 719–728 (2017).
51. Poggetti, V. *et al.* Carbonic Anhydrase Activators for Neurodegeneration: An Overview. *Molecules* **27**, (2022).
52. Lemon, N., Canepa, E., Ilies, M. A. & Fossati, S. Carbonic Anhydrases as Potential Targets Against Neurovascular Unit Dysfunction in Alzheimer's Disease and Stroke. *Front Aging Neurosci* **13**, 772278 (2021).
53. de Lope, E. G. *et al.* Comprehensive blood metabolomics profiling of Parkinson's disease reveals coordinated alterations in xanthine metabolism. *NPJ Parkinsons Dis* **10**, 68 (2024).
54. Favret, J. M., Weinstock, N. I., Feltri, M. L. & Shin, D. Pre-clinical Mouse Models of Neurodegenerative Lysosomal Storage Diseases. *Front Mol Biosci* **7**, 57 (2020).
55. Navarro-Romero, A., Montpeyó, M. & Martinez-Vicente, M. The Emerging Role of the Lysosome in Parkinson's Disease. *Cells* **9**, (2020).
56. Belbasis, L., Morris, S., van Duijn, C., Bennett, D. & Walters, R. Mendelian randomization identifies proteins involved in neurodegenerative diseases. *Brain* **148**, 2412–2428 (2025).
57. Nadjar, Y. *et al.* Elevated serum ferritin is associated with reduced survival in amyotrophic lateral sclerosis. *PLoS One* **7**, e45034 (2012).
58. Lunetta, C. *et al.* Serum C-Reactive Protein as a Prognostic Biomarker in Amyotrophic Lateral Sclerosis. *JAMA Neurol* **74**, 660–667 (2017).
59. Cao, W. & Fan, D. Neutrophils: a subgroup of neglected immune cells in ALS. *Front Immunol* **14**, 1246768 (2023).
60. Xiong, L. *et al.* Inflammation-dependent oxidative stress metabolites as a hallmark of amyotrophic lateral sclerosis. *Free Radic Biol Med* **178**, 125–133 (2022).
61. Writing Committee for the HEALEY ALS Platform Trial *et al.* Verdiperstat in Amyotrophic Lateral Sclerosis: Results From the Randomized HEALEY ALS Platform Trial. *JAMA Neurol* **82**, 333–343 (2025).
62. Fitzgerald, K. C. *et al.* Intakes of vitamin C and carotenoids and risk of amyotrophic lateral sclerosis: pooled results from 5 cohort studies. *Ann Neurol* **73**, 236–245 (2013).
63. Mòdol-Caballero, G. *et al.* Specific Expression of Glial-Derived Neurotrophic Factor in Muscles as Gene Therapy Strategy for Amyotrophic Lateral Sclerosis. *Neurotherapeutics* **18**, 1113–1126 (2021).
64. Faria, M. C. *et al.* Increased plasma levels of BDNF and inflammatory markers in Alzheimer's disease. *J Psychiatr Res* **53**, 166–172 (2014).

65. Lind, L., Fall, T., Ärnlöv, J., Elmståhl, S. & Sundström, J. Large-Scale Metabolomics and the Incidence of Cardiovascular Disease. *J Am Heart Assoc* **12**, e026885 (2023).
66. Yin, X. *et al.* Assessing the potential causal effects of 1099 plasma metabolites on 2099 binary disease endpoints. *Nat Commun* **16**, 3039 (2025).
67. Liu, A. *et al.* SLAMF7 Restrains Pro-Inflammatory Macrophage Activation to Counteract Doxorubicin-Induced Cardiotoxicity. *JACC Basic Transl Sci* 101256 (2025).
68. Niiranen, T. J., Kalesan, B., Mitchell, G. F. & Vasan, R. S. Relative Contributions of Pulse Pressure and Arterial Stiffness to Cardiovascular Disease. *Hypertension* **73**, 712–717 (2019).
69. Steffens, D. C. *et al.* Metabolomic differences in heart failure patients with and without major depression. *J Geriatr Psychiatry Neurol* **23**, 138–146 (2010).
70. Wang, D. D. *et al.* Plasma Ceramides, Mediterranean Diet, and Incident Cardiovascular Disease in the PREDIMED Trial (Prevención con Dieta Mediterránea). *Circulation* **135**, 2028–2040 (2017).
71. Rip, J. *et al.* Serum lipoprotein lipase concentration and risk for future coronary artery disease: the EPIC-Norfolk prospective population study. *Arterioscler Thromb Vasc Biol* **26**, 637–642 (2006).
72. Haeusler, R. A., Astiarraga, B., Camastra, S., Accili, D. & Ferrannini, E. Human insulin resistance is associated with increased plasma levels of 12α-hydroxylated bile acids. *Diabetes* **62**, 4184–4191 (2013).
73. Mallat, Z. *et al.* Protective role of interleukin-10 in atherosclerosis. *Circ Res* **85**, e17–24 (1999).
74. Cartland, S. P. *et al.* TRAIL-Expressing Monocyte/Macrophages Are Critical for Reducing Inflammation and Atherosclerosis. *iScience* **12**, 41–52 (2019).
75. Kelland, E., Patil, M. S., Patel, S., Cartland, S. P. & Kavurma, M. M. The Prognostic, Diagnostic, and Therapeutic Potential of TRAIL Signalling in Cardiovascular Diseases. *Int J Mol Sci* **24**, (2023).
76. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol* **14**, 576–590 (2018).
77. Yan, W., Yan, M., Wang, H. & Xu, Z. Associations of serum alkaline phosphatase level with all-cause and cardiovascular mortality in the general population. *Front Endocrinol (Lausanne)* **14**, 1217369 (2023).
78. Rozing, M. P. *et al.* Human insulin/IGF-1 and familial longevity at middle age. *Aging (Albany NY)* **1**, 714–722 (2009).
79. Lee, J.-H., Jeon, S., Joung, B., Lee, H. S. & Kwon, Y.-J. Associations of Homeostatic Model Assessment for Insulin Resistance Trajectories With Cardiovascular Disease Incidence and Mortality. *Arterioscler Thromb Vasc Biol* **43**, 1719–1728 (2023).
80. Drouin, G., Rioux, V. & Legrand, P. The n-3 docosapentaenoic acid (DPA): A new player in the n-3 long chain polyunsaturated fatty acid family. *Biochimie* **159**, 36–48 (2019).
81. Harris, W. S. *et al.* Blood n-3 fatty acid levels and total and cause-specific mortality from 17 prospective studies. *Nat Commun* **12**, 2329 (2021).

82. Oh, H. S.-H. *et al.* Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
83. Goeminne, L. J. E. *et al.* Plasma protein-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems. *Cell Metab* **37**, 205–222.e6 (2025).
84. Xie, K. *et al.* Loss of Life Expectancy by 10 Years or More From Elevated Aspartate Aminotransferase: Finding Aspartate Aminotransferase a Better Mortality Predictor for All-Cause and Liver-Related than Alanine Aminotransferase. *Am J Gastroenterol* **114**, 1478–1487 (2019).
85. Wassel, C. L., Barrett-Connor, E. & Laughlin, G. A. Association of circulating C-reactive protein and interleukin-6 with longevity into the 80s and 90s: The Rancho Bernardo Study. *J Clin Endocrinol Metab* **95**, 4748–4755 (2010).
86. Levine, M. E. *et al.* Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab* **19**, 407–417 (2014).
87. Zhang, P. *et al.* Trends in LDL-C and Non-HDL-C Levels with Age. *Aging Dis* **11**, 1046–1057 (2020).
88. Munno, M. *et al.* Radical Oxygen Species, Oxidized Low-Density Lipoproteins, and Lectin-like Oxidized Low-Density Lipoprotein Receptor 1: A Vicious Circle in Atherosclerotic Process. *Antioxidants (Basel)* **13**, (2024).
89. Ko, D. T. *et al.* High-Density Lipoprotein Cholesterol and Cause-Specific Mortality in Individuals Without Previous Cardiovascular Conditions: The CANHEART Study. *J Am Coll Cardiol* **68**, 2073–2083 (2016).
90. Ilieva, M. *et al.* Artemin and an Artemin-Derived Peptide, Artefin, Induce Neuronal Survival, and Differentiation Through Ret and NCAM. *Front Mol Neurosci* **12**, 47 (2019).
91. Zhu, M., Lv, T., Peng, M., Sun, H. & Li, Y. PTP-3 regulated by VB12 is important for ageing health in C. elegans. *NPJ Aging* **11**, 10 (2025).
92. Orouji, S., Liu, M. C., Korem, T. & Peters, M. A. K. Domain adaptation in small-scale and heterogeneous biological datasets. *Sci Adv* **10**, eadp6040 (2024).
93. Ranjan, B. *et al.* scConsensus: combining supervised and unsupervised clustering for cell type identification in single-cell RNA sequencing data. *BMC Bioinformatics* **22**, 186 (2021).
94. Yulug, B. *et al.* Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: a randomised, double-blinded, placebo-controlled phase-II trial. *Transl Neurodegener* **12**, 4 (2023).
95. Li, L., Chen, M. & Jiang, F.-C. Design, synthesis, and evaluation of 2-piperidone derivatives for the inhibition of β-amyloid aggregation and inflammation mediated neurotoxicity. *Bioorg Med Chem* **24**, 1853–1865 (2016).
96. Ritchie, S. C. *et al.* Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. *Nat Metab* **3**, 1476–1483 (2021).
97. Wainberg, M. *et al.* Multiomic blood correlates of genetic risk identify presymptomatic disease alterations. *Proc Natl Acad Sci U S A* **117**, 21813–21820 (2020).

98. Li, X. *et al.* Variational Autoencoder-based Model Improves Polygenic Prediction in Blood Cell Traits. *bioRxiv* (2025) doi:10.1101/2025.01.13.632820.
99. Ballard, J. L., Wang, Z., Li, W., Shen, L. & Long, Q. Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min* **17**, 38 (2024).
100. Loesch, D. P. *et al.* Identification of plasma proteomic markers underlying polygenic risk of type 2 diabetes and related comorbidities. *Nat Commun* **16**, 2124 (2025).
101. de Weerd, H. A. *et al.* Latent space arithmetic on data embeddings from healthy multi-tissue human RNA-seq decodes disease modules. *Patterns (N Y)* **5**, 101093 (2024).
102. Shen, X. *et al.* Nonlinear dynamics of multi-omics profiles during human aging. *Nat Aging* **4**, 1619–1634 (2024).
103. Kulkarni, A. S., Gubbi, S. & Barzilai, N. Benefits of Metformin in Attenuating the Hallmarks of Aging. *Cell Metab* **32**, 15–30 (2020).
104. Peng, W., Zhou, R., Sun, Z.-F., Long, J.-W. & Gong, Y.-Q. Novel Insights into the Roles and Mechanisms of GLP-1 Receptor Agonists against Aging-Related Diseases. *Aging Dis* **13**, 468–490 (2022).
105. Yesilyurt-Dirican, Z. E. *et al.* SGLT2 inhibitors as a novel senotherapeutic approach. *NPJ Aging* **11**, 35 (2025).
106. Cao, F., Wu, K., Zhu, Y.-Z. & Bao, Z.-W. Roles and Mechanisms of Dipeptidyl Peptidase 4 Inhibitors in Vascular Aging. *Front Endocrinol (Lausanne)* **12**, 731273 (2021).
107. Escott-Price, V. & Schmidt, K. M. Pitfalls of predicting age-related traits by polygenic risk scores. *Ann Hum Genet* **87**, 203–209 (2023).
108. Tang, M.-J. *et al.* Serum fibroblast growth factor 19 and endogenous islet beta cell function in type 2 diabetic patients. *Diabetol Metab Syndr* **11**, 79 (2019).
109. Carvalho, M. B. de *et al.* The role of FGF19 in metabolic regulation: insights from preclinical models to clinical trials. *Am J Physiol Endocrinol Metab* **327**, E279–E289 (2024).
110. Wei, Z., Koya, J. & Reznik, S. E. Insulin Resistance Exacerbates Alzheimer Disease via Multiple Mechanisms. *Front Neurosci* **15**, 687157 (2021).
111. Loef, M. & Walach, H. The omega-6/omega-3 ratio and dementia or cognitive decline: a systematic review on human studies and biological evidence. *J Nutr Gerontol Geriatr* **32**, 1–23 (2013).
112. Rostamzadeh, A. *et al.* Omega-6/Omega-3 Ratio in Dementia Prevention - Study Design and preliminary data of a feasibility study on lifestyle intervention. *Alzheimer's & Dementia* **20**, e090716 (2025).
113. Zhang, R. *et al.* Factors Influencing Alzheimer's Disease Risk: Whether and How They are Related to the APOE Genotype. *Neurosci Bull* **38**, 809–819 (2022).
114. Cocco, E. *et al.* Vitamin D responsive elements within the HLA-DRB1 promoter region in Sardinian multiple sclerosis associated alleles. *PLoS One* **7**, e41678 (2012).
115. Fenercioglu, A. K. The Anti-Inflammatory Roles of Vitamin D for Improving Human Health. *Curr Issues Mol Biol* **46**, 13514–13525 (2024).

116. Kawatani, K. *et al.* ABCA7 deficiency causes neuronal dysregulation by altering mitochondrial lipid metabolism. *Mol Psychiatry* **29**, 809–819 (2024).
117. Lyssenko, N. N. & Praticò, D. ABCA7 and the altered lipidostasis hypothesis of Alzheimer's disease. *Alzheimers Dement* **17**, 164–174 (2021).
118. Jones, L., Harold, D. & Williams, J. Genetic evidence for the involvement of lipid metabolism in Alzheimer's disease. *Biochim Biophys Acta* **1801**, 754–761 (2010).
119. Surdu, A. M. *et al.* Vitamin D and Its Role in the Lipid Metabolism and the Development of Atherosclerosis. *Biomedicines* **9**, (2021).
120. Marino, M. *et al.* Vitamin D Counteracts Lipid Accumulation, Augments Free Fatty Acid-Induced ABCA1 and CPT-1A Expression While Reducing CD36 and C/EBPβ Protein Levels in Monocyte-Derived Macrophages. *Biomedicines* **10**, (2022).
121. Zhu, Y. *et al.* MEG3 Activated by Vitamin D Inhibits Colorectal Cancer Cells Proliferation and Migration via Regulating Clusterin. *EBioMedicine* **30**, 148–157 (2018).
122. Soares, J. Z. *et al.* Vitamin D Levels, APOE Allele, and MRI Volumetry Assessed by NeuroQuant in Norwegian Adults with Cognitive Symptoms. *J Alzheimers Dis* **79**, 311–321 (2021).
123. Huebbe, P. *et al.* APOE ε4 is associated with higher vitamin D levels in targeted replacement mice and humans. *FASEB J* **25**, 3262–3270 (2011).
124. Mohanad, M., Mohamed, S. K., Aboulhoda, B. E. & Ahmed, M. A. E. Neuroprotective effects of vitamin D in an Alzheimer's disease rat model: Improvement of mitochondrial dysfunction via calcium/calmodulin-dependent protein kinase kinase 2 activation of Sirtuin1 phosphorylation. *Biofactors* **50**, 371–391 (2024).
125. Mead, J. R. & Ramji, D. P. The pivotal role of lipoprotein lipase in atherosclerosis. *Cardiovasc Res* **55**, 261–269 (2002).
126. Kumari, A., Kristensen, K. K., Ploug, M. & Winther, A.-M. L. The Importance of Lipoprotein Lipase Regulation in Atherosclerosis. *Biomedicines* **9**, (2021).
127. Jazwa, A. & Cuadrado, A. Targeting heme oxygenase-1 for neuroprotection and neuroinflammation in neurodegenerative diseases. *Curr Drug Targets* **11**, 1517–1531 (2010).
128. Monir, D. M., Mahmoud, M. E., Ahmed, O. G., Rehan, I. F. & Abdelrahman, A. Forced exercise activates the NrF2 pathway in the striatum and ameliorates motor and behavioral manifestations of Parkinson's disease in rotenone-treated rats. *Behav Brain Funct* **16**, 9 (2020).
129. Alves de Souza, R. W. *et al.* Skeletal muscle heme oxygenase-1 activity regulates aerobic capacity. *Cell Rep* **35**, 109018 (2021).
130. Tokgöz, S. & Claassen, J. A. H. R. Exercise as Potential Therapeutic Target to Modulate Alzheimer's Disease Pathology in APOE ε4 Carriers: A Systematic Review. *Cardiol Ther* **10**, 67–88 (2021).
131. Mansouri, A. *et al.* Antioxidant Effects of Statins by Modulating Nrf2 and Nrf2/HO-1 Signaling in Different Diseases. *J Clin Med* **11**, (2022).

132. Lin, H.-Y. *et al.* Desipramine protects neuronal cell death and induces heme oxygenase-1 expression in Mes23.5 dopaminergic neurons. *PLoS One* **7**, e50138 (2012).
133. Sakuru, R. & Bollu, P. C. Hurler Syndrome. in *StatPearls [Internet]* (StatPearls Publishing, 2023).
134. Dai, L. *et al.* Lysosomal dysfunction in α-synuclein pathology: molecular mechanisms and therapeutic strategies. *Cell Mol Life Sci* **81**, 382 (2024).
135. Gokuladhas, S., Fadason, T., Farrow, S., Cooper, A. & O'Sullivan, J. M. Discovering genetic mechanisms underlying the co-occurrence of Parkinson's disease and non-motor traits. *NPJ Parkinsons Dis* **10**, 27 (2024).
136. Gu, X.-J. *et al.* Expanding causal genes for Parkinson's disease via multi-omics analysis. *npj Parkinson's Disease* **9**, 1–10 (2023).
137. Kida, S. *et al.* Enzyme replacement with transferrin receptor-targeted α-L-iduronidase rescues brain pathology in mucopolysaccharidosis I mice. *Mol Ther Methods Clin Dev* **29**, 439–449 (2023).
138. Chen, M. *et al.* Heparan sulfate proteoglycans mediate prion-like α-synuclein toxicity in Parkinson's in vivo models. *Life Sci Alliance* **5**, (2022).
139. Sian-Hulsmann, J., Riederer, P. & Michel, T. M. Metabolic Dysfunction in Parkinson's Disease: Unraveling the Glucose-Lipid Connection. *Biomedicines* **12**, (2024).
140. Henrich, M. T., Oertel, W. H., Surmeier, D. J. & Geibl, F. F. Mitochondrial dysfunction in Parkinson's disease - a key disease hallmark with therapeutic potential. *Mol Neurodegener* **18**, 83 (2023).
141. Wang, L., Li, C., Chen, X., Li, S. & Shang, H. Abnormal Serum Iron-Status Indicator Changes in Amyotrophic Lateral Sclerosis (ALS) Patients: A Meta-Analysis. *Front Neurol* **11**, 380 (2020).
142. Belete, G. T. *et al.* Metabolomics studies in common multifactorial eye disorders: a review of biomarker discovery for age-related macular degeneration, glaucoma, diabetic retinopathy and myopia. *Front Mol Biosci* **11**, 1403844 (2024).
143. Bou Ghanem, A. *et al.* Exploring the complexities of 1C metabolism: implications in aging and neurodegenerative diseases. *Front Aging Neurosci* **15**, 1322419 (2023).
144. Brown-Borg, H. M., Rakoczy, S., Wonderlich, J. A., Armstrong, V. & Rojanathammanee, L. Altered dietary methionine differentially impacts glutathione and methionine metabolism in long-living growth hormone-deficient Ames dwarf and wild-type mice. *Longev Healthspan* **3**, 10 (2014).
145. Zhu, J., Saikia, G., Zhang, X., Shen, X. & Kahe, K. One-Carbon Metabolism Nutrients, Genetic Variation, and Diabetes Mellitus. *Diabetes Metab J* **48**, 170–183 (2024).
146. Ducker, G. S. & Rabinowitz, J. D. One-Carbon Metabolism in Health and Disease. *Cell Metab* **25**, 27–42 (2017).
147. Corominas-Faja, B. *et al.* Metabolomic fingerprint reveals that metformin impairs one-carbon metabolism in a manner similar to the antifolate class of chemotherapy drugs. *Aging (Albany NY)* **4**, 480–498 (2012).
148. Rosenzweig, A., Blenis, J. & Gomes, A. P. Beyond the Warburg Effect: How Do Cancer Cells Regulate One-Carbon Metabolism? *Front Cell Dev Biol* **6**, 90 (2018).

149. Yurchenko, M. *et al.* SLAMF1 is required for TLR4-mediated TRAM-TRIF-dependent signaling in human macrophages. *J Cell Biol* **217**, 1411–1429 (2018).
150. Cai, B. *et al.* MerTK receptor cleavage promotes plaque necrosis and defective resolution in atherosclerosis. *J Clin Invest* **127**, 564–568 (2017).
151. Wu, Y. *et al.* SLAMF7 regulates the inflammatory response in macrophages during polymicrobial sepsis. *J Clin Invest* **133**, (2023).
152. Olsen, M. B. *et al.* SLAMF1-derived peptide exhibits cardio protection after permanent left anterior descending artery ligation in mice. *Front Immunol* **15**, 1383505 (2024).
153. Revelas, M. *et al.* High polygenic risk score for exceptional longevity is associated with a healthy metabolic profile. *Geroscience* **45**, 399–413 (2023).
154. Libert, S., Chekholko, A. & Kenyon, C. A mathematical model that predicts human biological age from physiological traits identifies environmental and genetic factors that influence aging. *Elife* **13**, (2025).
155. Panichi, V. *et al.* Serum alkaline phosphatase is a strong predictor of mortality in ESKD patients: analysis of the RISCAVID cohort. *J Nephrol* **37**, 1843–1851 (2024).
156. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).
157. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573–580 (2022).
158. Ding, Y. *et al.* Comprehensive human proteome profiles across a 50-year lifespan reveal aging trajectories and signatures. *Cell* (2025) doi:10.1016/j.cell.2025.06.047.
159. Mars, N. *et al.* Systematic comparison of family history and polygenic risk across 24 common diseases. *Am J Hum Genet* **109**, 2152–2162 (2022).
160. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–1458 (2013).
161. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**, 1091–1102 (2019).
162. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* **97**, 1267–1288 (2018).
163. Mansour Aly, D. *et al.* Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nature Genetics* **53**, 1534–1542 (2021).
164. Tesi, N. *et al.* Polygenic Risk Score of Longevity Predicts Longer Survival Across an Age Continuum. *J Gerontol A Biol Sci Med Sci* **76**, 750–759 (2021).
165. Earls, J. C. *et al.* Multi-Omic Biological Age Estimation and Its Correlation With Wellness and Disease Phenotypes: A Longitudinal Study of 3,558 Individuals. *J Gerontol A Biol Sci Med Sci* **74**, S52–S60 (2019).

166. Huang, D. *et al.* QTLbase2: an enhanced catalog of human quantitative trait loci on extensive molecular phenotypes. *Nucleic Acids Res* **51**, D1122–D1128 (2022).

Figures

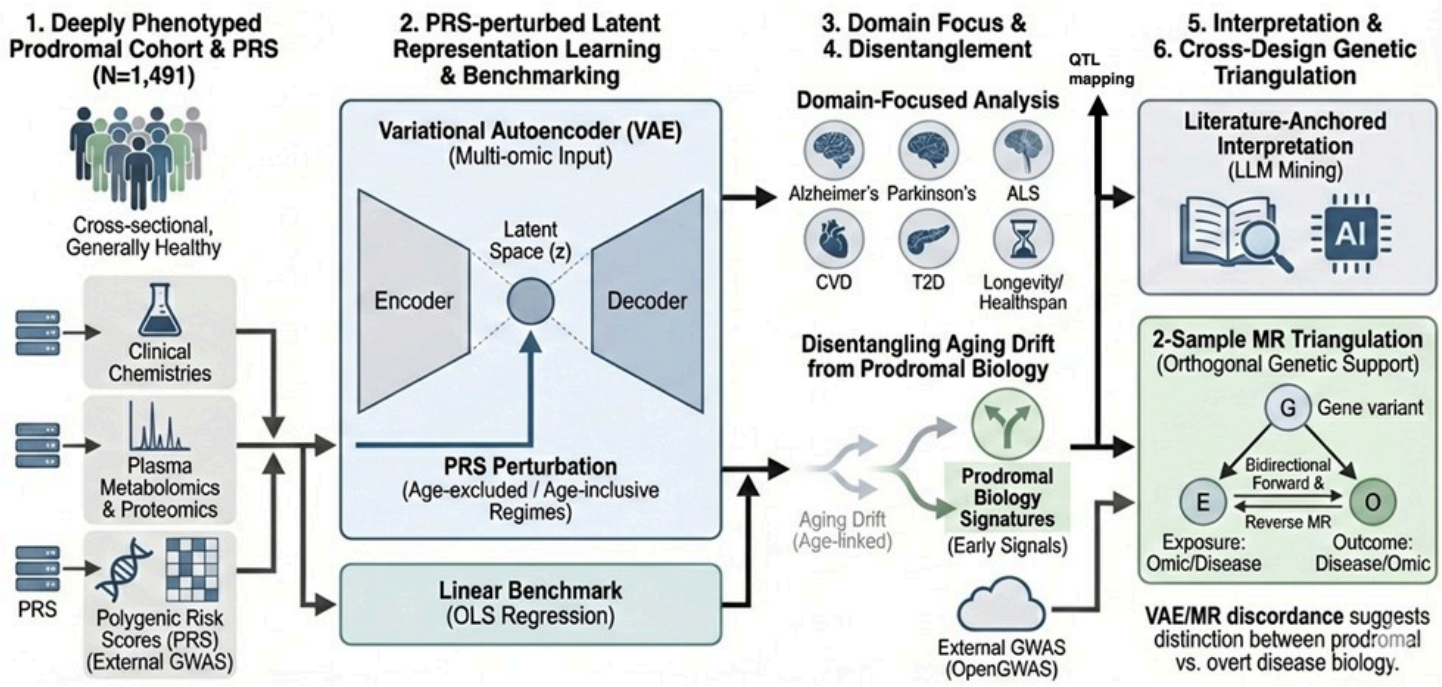


Figure 1

Overview of PRS-perturbed latent modeling and cross-design genetic triangulation.

A deeply phenotyped, cross-sectional prodromal cohort (N = 1,491) with clinical chemistries, plasma metabolomics, and proteomics was integrated with externally derived polygenic risk scores (PRS) spanning disease and aging liabilities. Multi-omic data were modeled using a variational autoencoder (VAE) under age-excluded and age-inclusive perturbation regimes, with ordinary least squares (OLS) regression used as a linear benchmark. Downstream analyses focused on six age-related domains to disentangle age-linked drift from genetically driven prodromal biology and to identify early biomarker signatures. VAE-derived associations were interpreted through systematic literature mining and triangulated using bidirectional two-sample Mendelian randomization (2SMR) with external GWAS, providing orthogonal genetic support and enabling comparison between prodromal and overt disease biology.

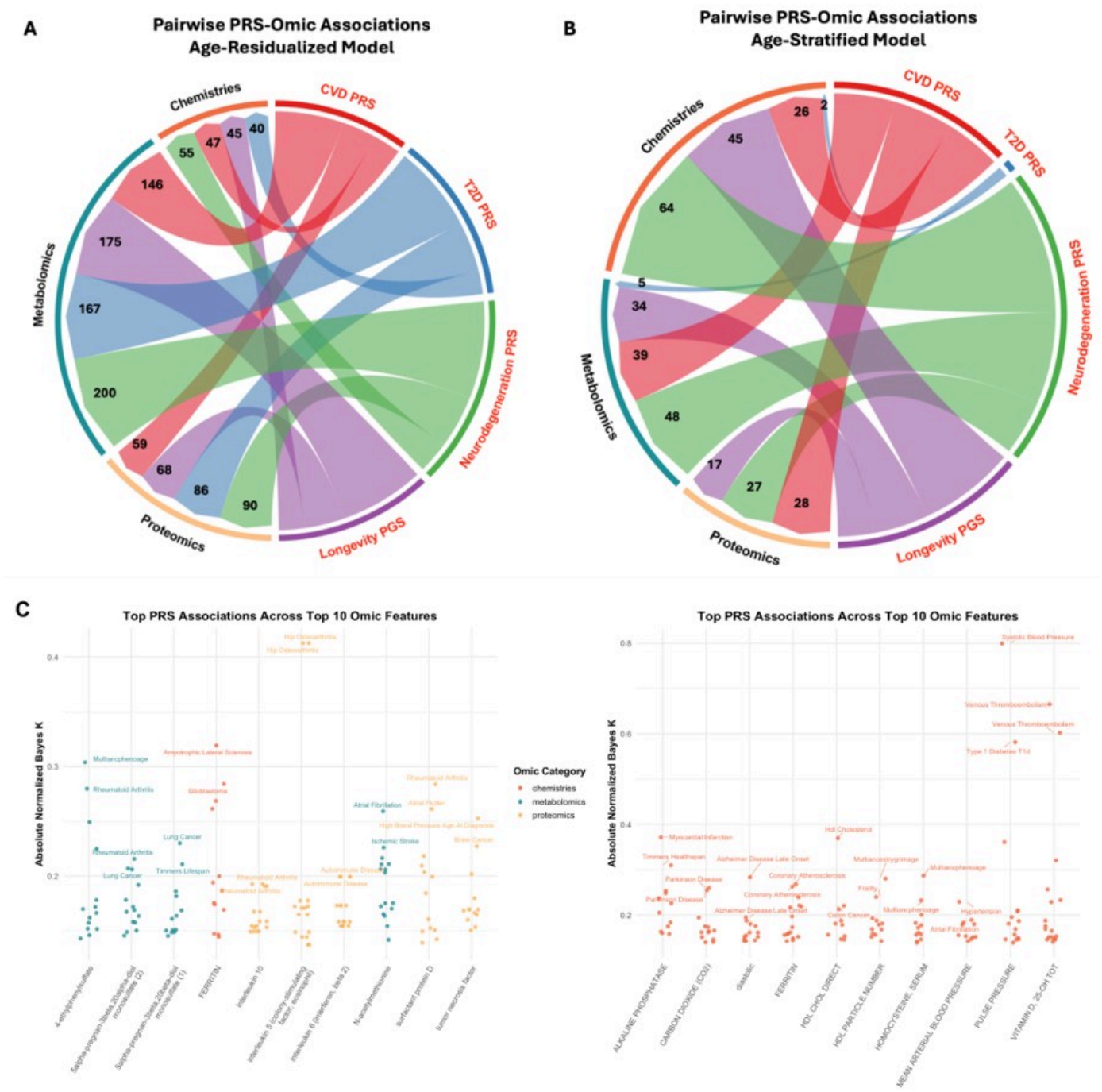


Figure 2

Chronological age reshapes detectable PRS–omics associations.

(A) Age-residualized VAE model and (B) age-stratified model chord diagrams showing significant associations between polygenic risk scores (PRS) for cardiovascular disease (CVD), type 2 diabetes (T2D), neurodegeneration (Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis), and longevity with metabolomic, proteomic, and clinical chemistry features. Supplying age reduced total associations 3.5-fold (1,178 vs. 335), enriched for chemistry traits, and produced a 42-fold drop in T2D PRS links. (C) The strongest associations illustrate this shift: the age-residualized model highlights

diverse metabolomic and proteomic signals, while the age-stratified model is dominated by clinical chemistry traits such as cholesterol fractions, blood pressure metrics, and alkaline phosphatase. Together, these findings indicate that residualized models capture broad age-independent molecular signatures, whereas age-stratified models emphasize narrower, age-conditioned effects.

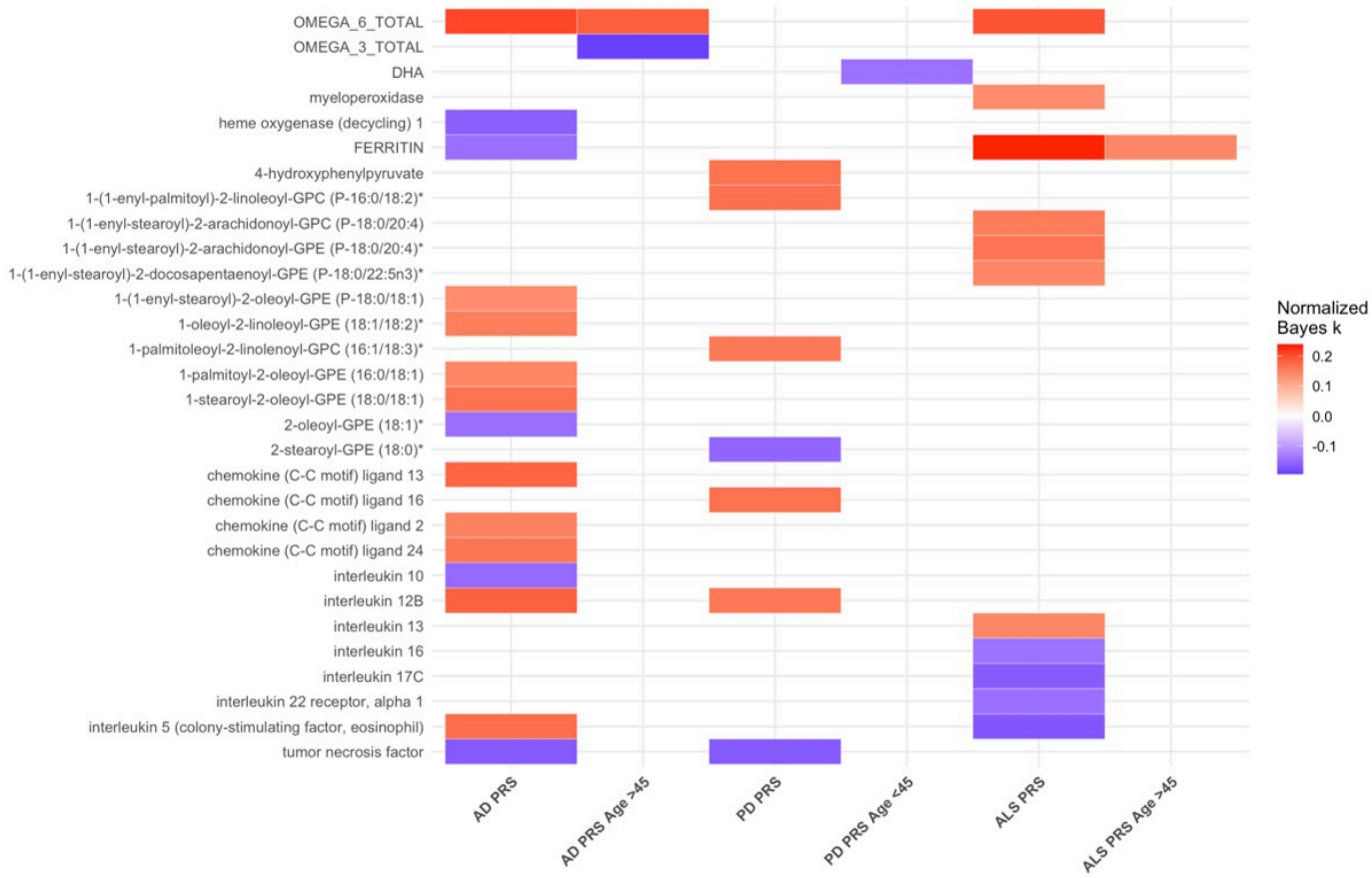


Figure 3

Shared molecular themes across neurodegenerative PRS.

Heatmap of associations between polygenic risk scores (PRS) for Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS) and omic features spanning lipid metabolism, oxidative stress, iron handling, and inflammation. Across two independent AD PRS and one each for PD and ALS, we observed convergent shifts in ether-linked GPE/GPC plasmalogens, implicating membrane lipid remodeling as a shared vulnerability pathway. Additional signals included oxidative stress markers (e.g., myeloperoxidase, heme oxygenase), iron-related biology (ferritin), pro-inflammatory cytokines (e.g., TNF, IL-12B), and a consistent imbalance of ω-6/ω-3 long-chain polyunsaturated fatty acids. Together, these findings suggest common perturbations in plasmalogen homeostasis, oxidative/ferroptotic stress, and immune signaling that may underlie genetic susceptibility to neurodegeneration.

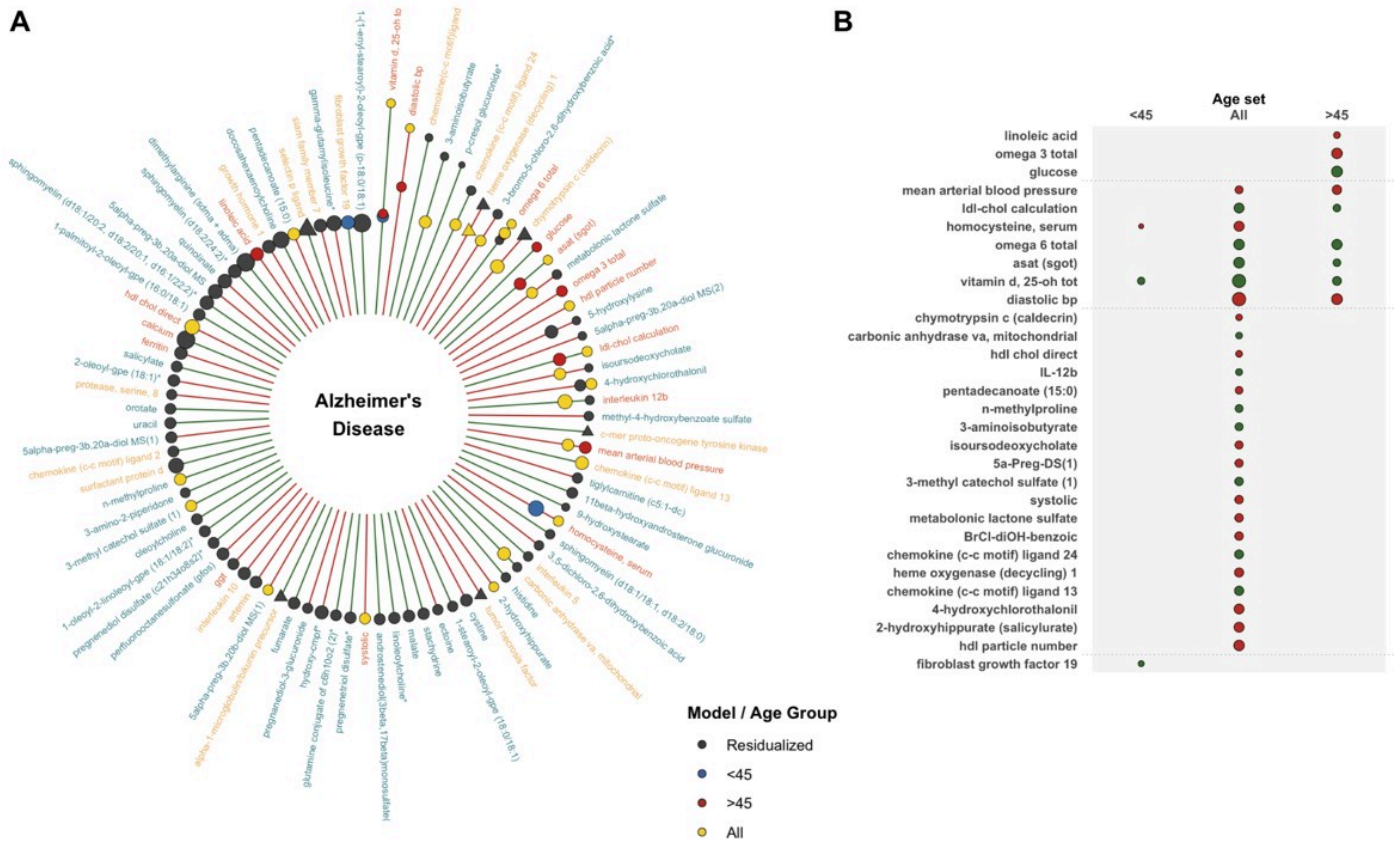


Figure 4

Alzheimer's disease PRS–omics associations highlight antioxidant defense, vascular physiology, and age-related adaptations.

(A) Circular lollipop plot of significant AD PRS–omics associations meeting FDR and replicability criteria (see Methods). ⍰- no associated QTL; △ - eQTL; red stem - negative association; green stem - positive association; stem length indicates effect size and data point size is inversely proportional to the confidence interval (CI) (note that no CI overlapped zero effect size). Dark gray data points depict the residualized model, whereas yellow, red, and blue data points represent the results from the age-included model (see legend). (B) Serum features that associated across age strata in the age-included VAE model, showing age-conditioned manifestation of polygenic AD risk (red - negative association; green - positive association; data point size is proportional to effect size).

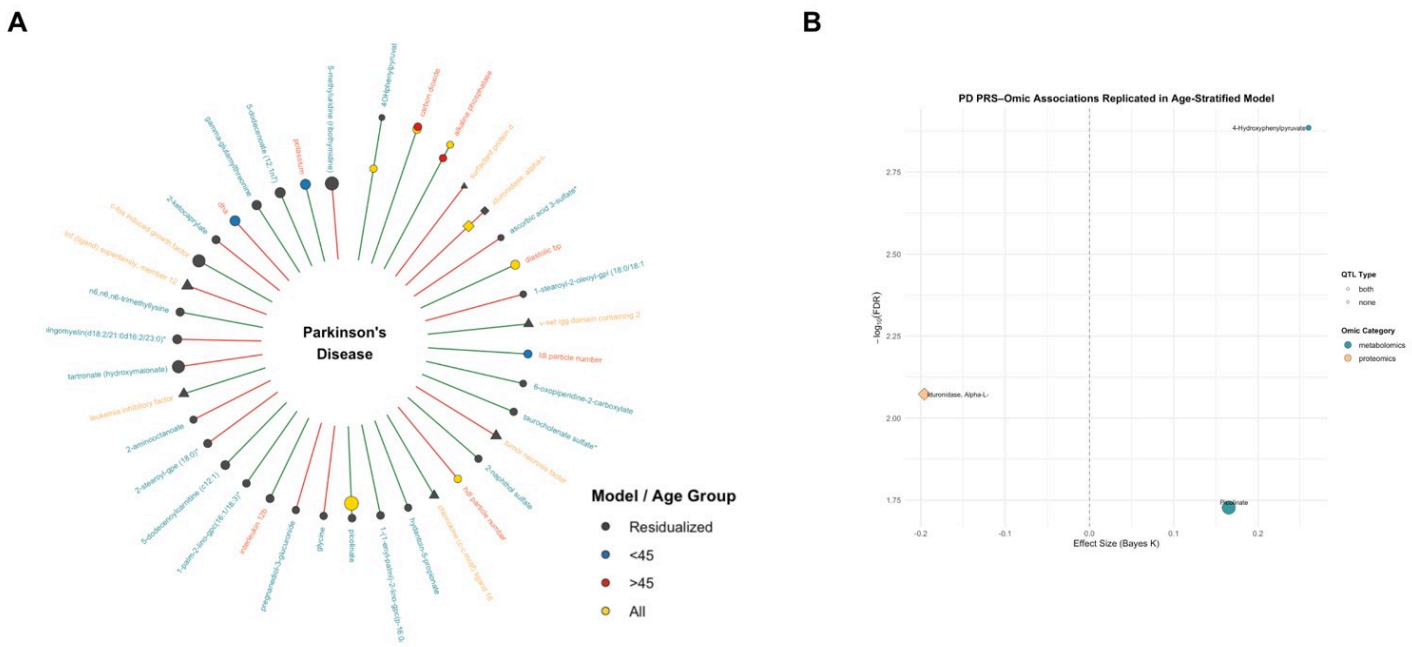


Figure 5

Parkinson's disease PRS–omics associations identify IDUA and tyrosine catabolism as consistent signals.

(A) Circular lollipop plot of significant Parkinson's disease (PD) PRS–omics associations passing FDR and replicability thresholds, annotated by QTL colocalization status and model framework. (B) Replicated associations in the age-stratified model. Three age-invariant features were consistently detected across both modeling approaches: α-L-iduronidase (IDUA), 4-hydroxyphenylpyruvate (4-HPP), and picolinate. IDUA, supported by eQTL and pQTL evidence, was consistently downregulated, suggesting a prodromal reduction in lysosomal function among genetically at-risk individuals; two-sample Mendelian randomization confirmed a likely causal relationship between IDUA and PD risk (see Results). Together, these results highlight impaired lysosomal biology and altered tyrosine metabolism as core biochemical signatures of PD genetic liability. - no associated QTL; △- eQTL; - both eQTL and pQTL; red stem - negative association; green stem - positive association; stem length indicates effect size and data point size is inversely proportional to the confidence interval (CI) (note that no CI overlapped zero effect size). Dark gray data points depict the residualized model, whereas yellow, red, and blue data points represent the results from the age-included model (see legend).

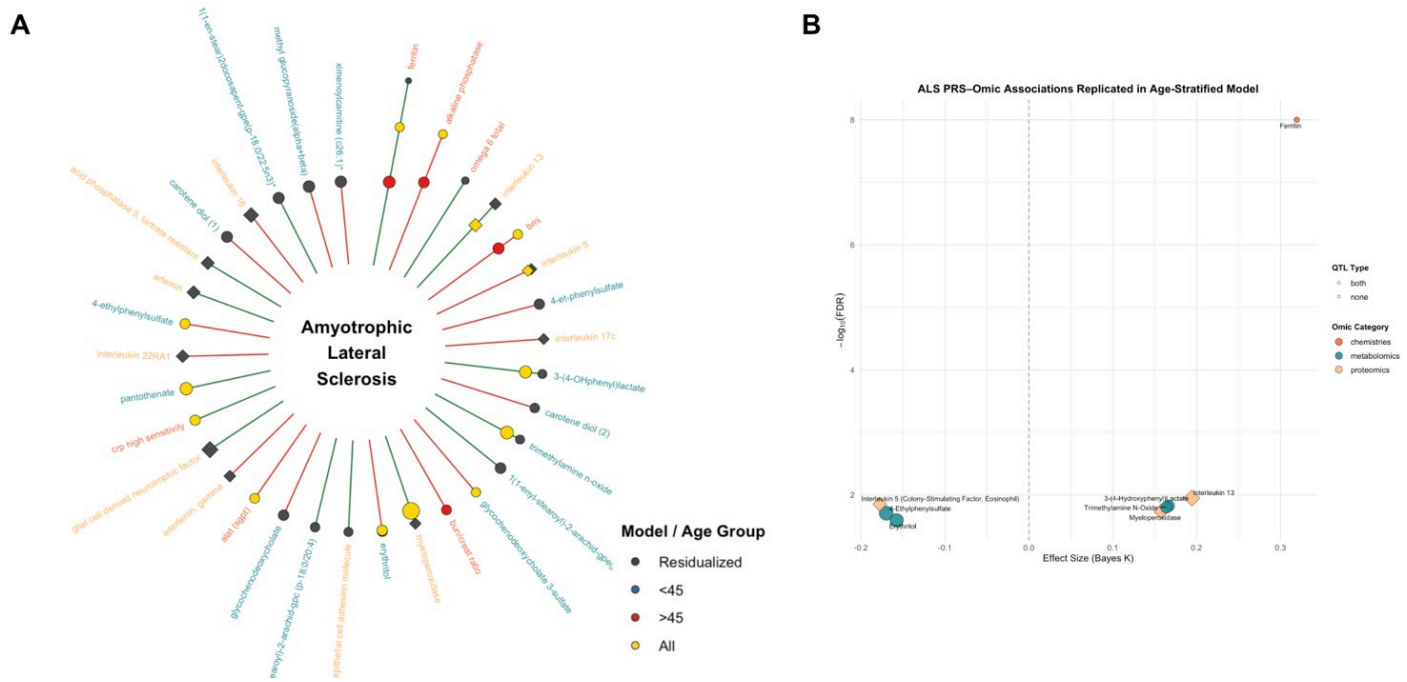


Figure 6

Amyotrophic lateral sclerosis PRS–omics associations implicate iron metabolism, oxidative stress, and inflammatory signaling.

(A) Circular lollipop plot of significant ALS PRS–omics associations passing FDR and replicability thresholds, annotated by omic category, age-model framework, and QTL colocalization status. (B) Replicated associations in the age-stratified model. The ALS genetic risk landscape was dominated by iron-linked oxidative stress, systemic inflammation, and energy–lipid remodeling. Ferritin, MPO, and hsCRP—previously reported as elevated in ALS pathology—persisted across both modeling frameworks. □- no associated QTL; △ - eQTL; □ - both eQTL and pQTL; red stem - negative association; green stem - positive association; stem length indicates effect size and data point size is inversely proportional to the confidence interval (CI) (note that no CI overlapped zero effect size). Dark gray data points depict the residualized model, whereas yellow, red, and blue data points represent the results from the age-included model (see legend).

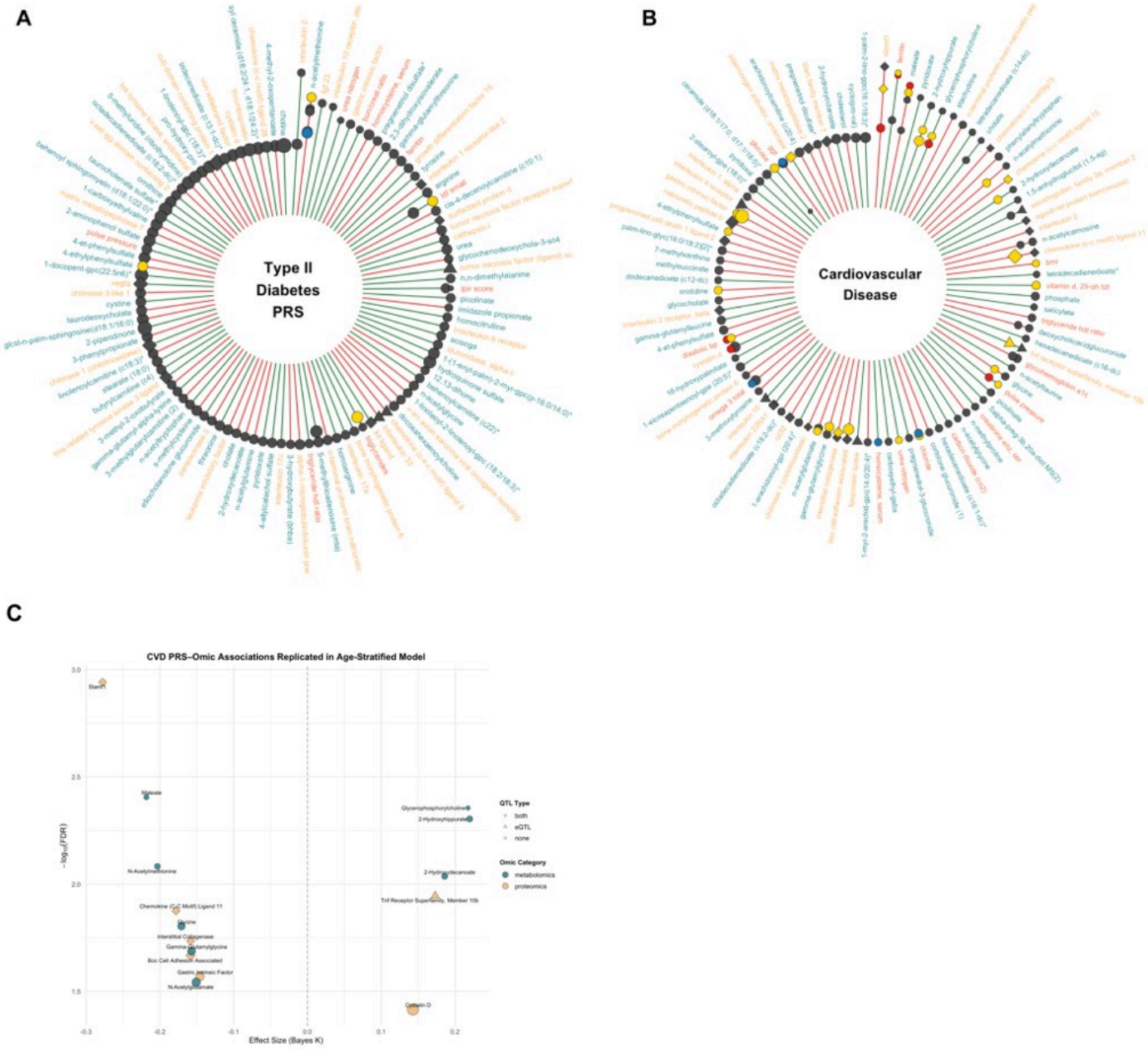


Figure 7

Cardiometabolic PRS–omics associations reveal age-conditioned T2D signals and immune-driven CVD features.

(A) Circular lollipop plots of type 2 diabetes (T2D) PRS–omics associations. Across four T2D PRS, the age-residualized VAE detected broad metabolic signatures—lipotoxic lipids and depletion of insulin-sensitizing species—that were largely lost (>95%) when age was modeled, indicating strong age-dependence of T2D genetic risk manifestation. (B) Circular plot and volcano plot of cardiovascular disease (CVD) PRS–omics associations.The most pronounced was depletion of SLAMF1, supported by cis-eQTL and pQTL evidence. (C) Replicated CVD associations in the age-stratified model highlighted core prodromal features including reduced SLAMF1, adhesion proteins, and amino acid conjugates such

as N-acetylmethionine (NAM). ⊟- no associated QTL; △ - eQTL; ⊠ - both eQTL and pQTL; red stem - negative association; green stem - positive association; stem length indicates effect size and data point size is inversely proportional to the confidence interval (CI) (note that no CI overlapped zero effect size). Dark gray data points depict the residualized model, whereas yellow, red, and blue data points represent the results from the age-included model (see legend).

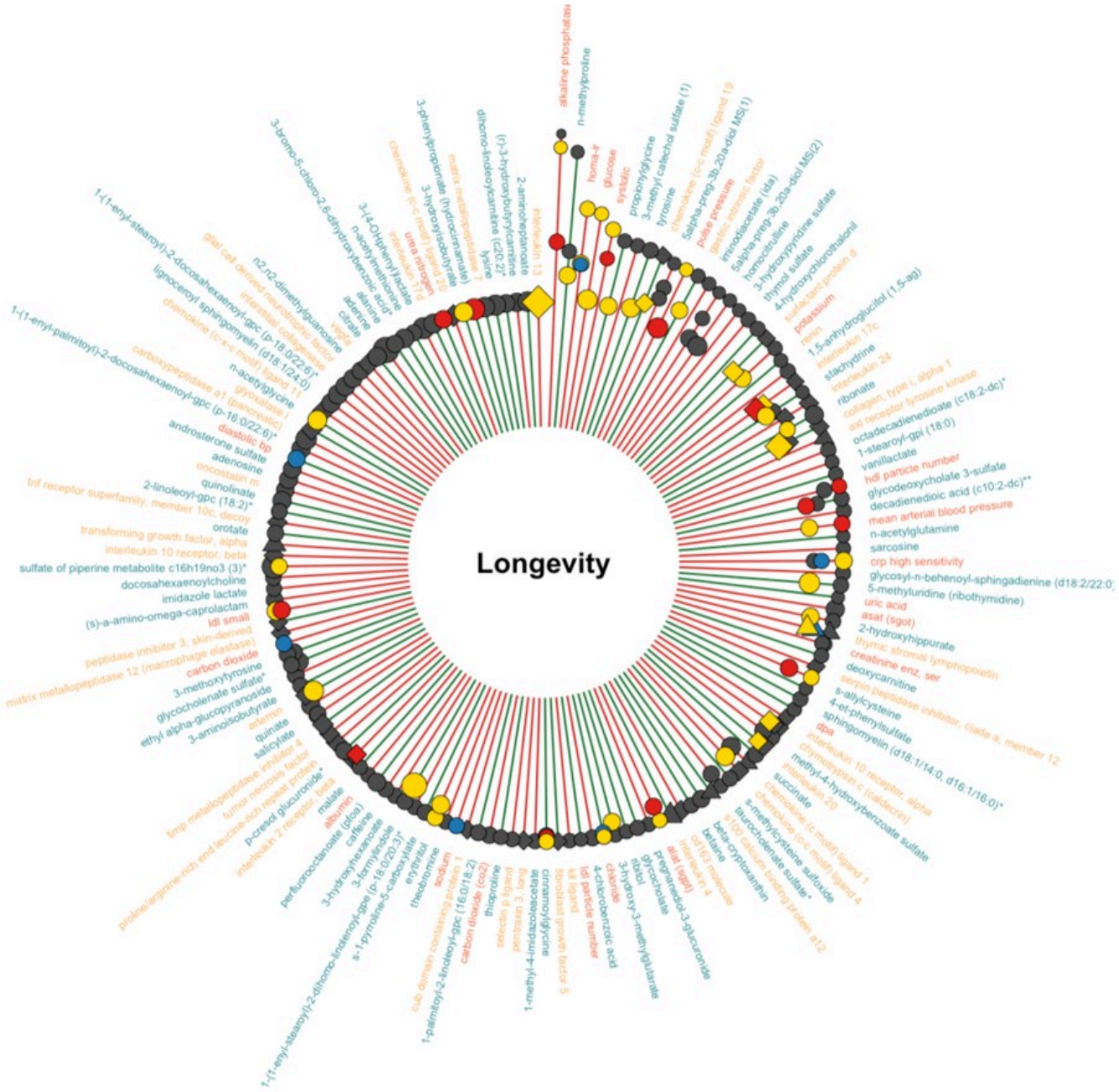


Figure 8

Longevity PRS–omics associations define an anti-inflammatory, metabolic resilience signature. Circular and lollipop plots of significant longevity PRS associations, annotated by QTL colocalization and age-model framework. Across five longevity-related PGS, higher genetic predisposition to longevity was associated with lower systemic inflammation, lower liver enzymes, improved glycemic control, and

favorable lipid features, including higher long-chain ω-3 fatty acids. □ - no associated QTL; △ - eQTL; □ - both eQTL and pQTL; red stem - negative association; green stem - positive association; stem length indicates effect size and data point size is inversely proportional to the confidence interval (CI) (note that no CI overlapped zero effect size). Dark gray data points depict the residualized model, whereas yellow, red, and blue data points represent the results from the age-included model (see legend).

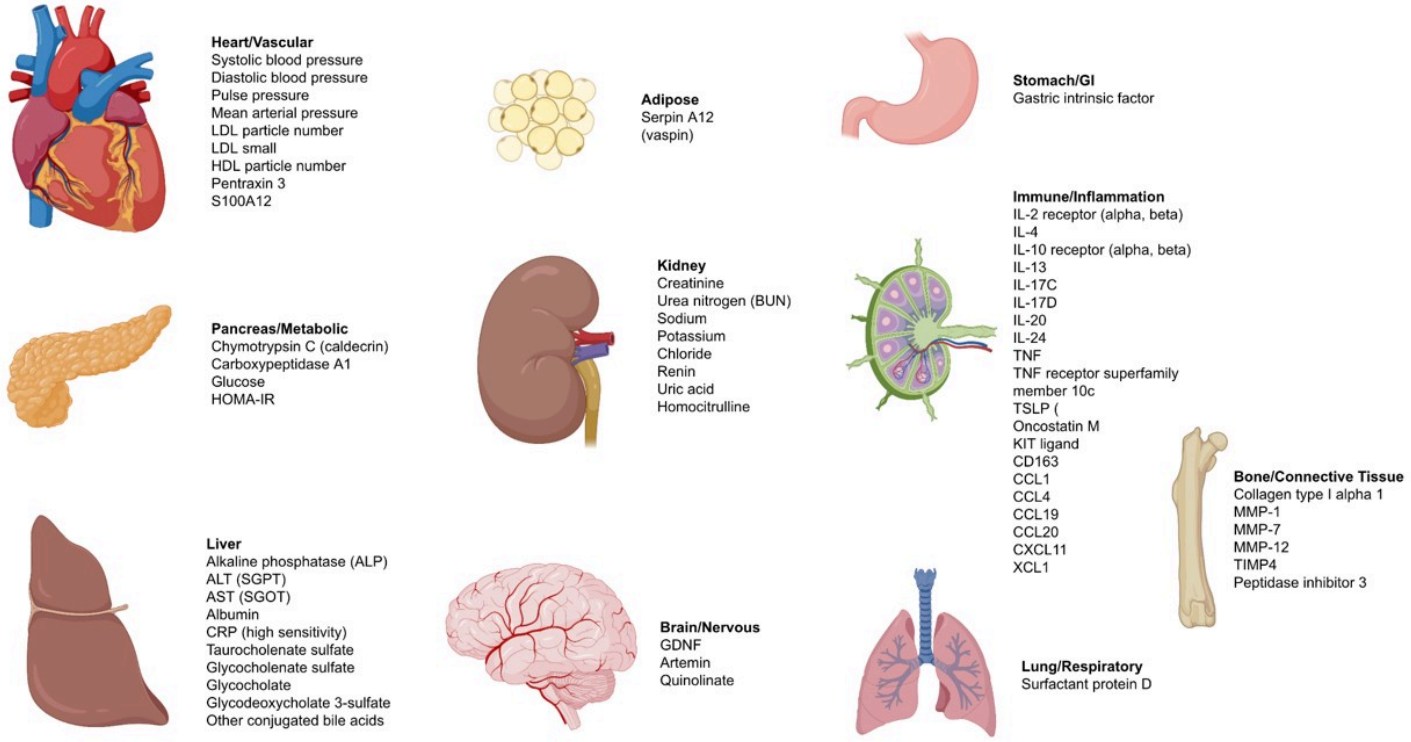


Figure 9

Longevity PRS associations map to organ-level protective profiles.

Schematic overview of longevity PRS–omics associations categorized by organ system. Higher longevity PRS were linked to reduced hepatocellular and systemic inflammation, lower arterial stiffness and blood pressure (heart/vascular), improved renal and electrolyte balance (kidney), more favorable adipose and glycemic signaling (pancreas/adipose), and neurotrophic support (brain/nervous, gastric intrinsic factor).

Supplementary Files

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

- SupplementaryTables2.xlsx