

Age Acceleration and Mortality Risk Constitute Distinct Dimensions of Organ Aging

Rong Na

narong.hs@gmail.com

The University of Hong Kong <https://orcid.org/0000-0001-7470-5108>

Chang Shi

The University of Hong Kong

Xiaohao Ruan

Shanghai Jiao Tong University School of Medicine

Ruofan Shi

Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine <https://orcid.org/0009-0005-1001-390X>

Ruochen Ma

The University of Hong Kong

Zixin Liang

School of Public Health (Shenzhen), Sun Yat-sen University

Yao Sun

The University of Hong Kong

Chi Yao

The University of Hong Kong

Xiangchuan Xu

The University of Hong Kong

Xue Li

The University of Hong Kong <https://orcid.org/0000-0003-4836-7808>

Haitao Chen

School of Public Health (Shenzhen), Sun Yat-sen University

Article

Keywords:

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10881849/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

Organ biological aging is highly heterogeneous, yet existing computational clocks typically model it as a single-rate phenomenon. It remains unclear whether chronological age acceleration and mortality risk are driven by shared or distinct biological programs. Here, we established multi-omics, organ-specific aging clocks for 14 organs across 409,206 UK Biobank participants, integrating nuclear magnetic resonance (NMR) metabolomics, Olink proteomics, and clinical phenotypes. By evaluating age acceleration against both organ-specific mortality and all-cause mortality, we demonstrate that age acceleration and mortality risk are divergent dimensions of organ biology with separable phenotypic, molecular, and genetic architectures. Feature-ablation genome-wide association studies revealed three molecular layers underlying these signatures including NMR-dependent, plasma-protein-direct and clinical-biochemistry-mediated, demonstrating that single-platform designs systematically miss complementary biology. This divergence is highlighted by a prostate paradox, where all-cause mortality training entirely obscures the incident prostate cancer signal that the organ-specific mortality clock powerfully recovers. Finally, applying polygenic scores in a non-overlapping validation cohort of 383,227 individuals alongside bidirectional Mendelian randomization and multi-tier gene prioritization, we prioritized 13 candidate genes underlying these distinct aging dimensions. Our findings establish that multi-omics integration combined with organ-specific training captures vital components of organ aging invisible to conventional single-objective clocks. Crucially, this multi-objective framework enables organ-resolved comorbidity risk assessments that traditional competing-risk models obscure, providing a highly translatable foundation for targeted clinical surveillance and actionable, organ-specific interventions.

Introduction

Aging is the primary risk factor for chronic disease and mortality, driven by progressive molecular and cellular deterioration (López-Otín et al. 2013, 2023). Yet organs deteriorate at heterogeneous rates and through different molecular mechanisms, producing diverse disease trajectories that whole-organism aging measures cannot resolve. Biological aging clocks, computational models that estimate biological age from molecular biomarkers, have emerged to quantify this variation. These models have progressed from first-generation chronological-age clocks based on DNA methylation (Horvath 2013; Hannum et al. 2013) to second-generation mortality-trained whole-body clocks (PhenoAge (Levine et al. 2018), GrimAge (Lu et al. 2019), DunedinPACE (Belsky et al. 2022)), and most recently, toward organ-specific resolution (Oh et al. 2023). Studies including our previous work have shown that whole-body biological aging is significantly associated with multiple chronic diseases, such as cancers, cardiovascular diseases, neurodegenerative diseases (Zhan et al. 2025).

While plasma-based organ-specific aging clocks have advanced rapidly across populations (Wang et al. 2026) and long-term diseases prediction frameworks (Kivimäki et al. 2025), three critical limitations remain. First, existing molecular organ clocks have relied on single omics platform (epigenetic, proteomics or metabolomics) (Oh et al. 2023, 2025; Anagnostakis et al. 2025). This single-platform

approach misses complementary biology carried by biochemically distinct molecular layers. For example, epigenetic (such as methylation) clock may reflect the expression of the gene to downstream proteins, while intracellular biology changes such as mitochondrial functions may not be fully evaluated which is equally critical for organ and cellular senescence. Second, although recent organ-aging clocks have introduced mortality training (Goeminne et al. 2025; Peng et al. 2026), existing designs rely exclusively on all-cause mortality. None have yet trained clocks on organ-specific endpoints. Consequently, organs whose primary disease burden contributes little to overall mortality, such as the prostate, are systematically diluted under all-cause training, obscuring aging signals that organ-specific training would expose. Third, prior organ-aging genome-wide association studies (GWAS) have each been built on a single aging objective, including proteomic age acceleration (Zhu et al. 2026) or NMR-based metabolome age (Anagnostakis et al. 2025), comprehensive GWAS for multi-omic aging clock has never been reported. Consequently, whether the age acceleration and the risk of mortality arise from shared or distinct genetic architectures were poorly understood.

Here we address these three gaps by constructing multi-omics organ-specific aging clocks that integrate multiple genetic and molecular levels, including Olink proteomics (Sun et al. 2023) to reflect gene expression related to genetic and epigenetic regulation, NMR metabolomics (Julkunen et al. 2023) to evaluate intracellular biochemical processes, and multiple phenotypes (Levine et al. 2018) to illustrate the consequent changes of the body across 14 organs. By evaluating three training objectives in parallel, including age acceleration (AA), organ-specific mortality (OS-Mort) and all-cause mortality (AC-Mort), in UK Biobank participants, we ask whether an organ carries one aging signal or several, and follow the answer from clock scores through incident disease to genetic architecture.

By decoupling chronological age acceleration from organ-specific mortality risk, this approach overcomes the competing-risk dilution inherent in all-cause models. Ultimately, this establishes a translatable framework capable of detecting localized organ deterioration and actionable biological targets long before systemic failure occurs.

Results

Performance and Concordance of Aging Clocks

We built multi-omics aging clocks for fourteen organs in the UK Biobank (Sudlow et al. 2015). Clock training drew on the 28,754 participants with both NMR metabolomic (Julkunen et al. 2023) and Olink proteomic (Sun et al. 2023) measurements, of whom 15,262 (53.1%) were women; the median baseline age was 58 years (interquartile range 51 to 64, range 39 to 70). Over 354,539 person-years of follow-up (median 12.7 years per participant, interquartile range 12.0 to 13.4), 2,728 all-cause deaths occurred, or 7.7 per 1,000 person-years (Supplementary Data 1). Participants were recruited in 2006–2010 at 22 UK centres; UK Biobank volunteers are healthier than the general population, so absolute rates are not population-representative. Genetic analyses rested on the 409,206 European-ancestry participants with genotype data: the 25,979 who also carried complete multi-omics measurements formed the discovery

cohort for genome-wide association, and the remaining 383,227 formed a genotype-only validation cohort used for polygenic scoring (Extended Data Fig. 1).

For each organ we assembled input features by a three-tier rule: Olink proteins were assigned to organs by Human Protein Atlas tissue-enrichment annotation; all 233 NMR metabolites were supplied to every organ, with regularisation left to select the organ-relevant ones; and a common set of clinical-biochemistry and hormone markers was added. The three tiers combined gave 254 to 433 input features per organ. On this single feature set we trained three clocks per organ with different objectives. Age-acceleration (AA) clocks predict chronological age by elastic-net regression, with age acceleration taken as the residual of predicted on chronological age. Organ-specific mortality (OS-Mort) clocks predict time to death from the organ's own causes by elastic-net Cox regression, with the organ-to-cause mapping pre-specified by ICD-10 codes (full mapping in Supplementary Data 1). All-cause mortality (AC-Mort) clocks predict time to death from any cause under the same Cox setting. All three scores were age-calibrated within the training folds (Methods): correlations with age ran from 0.22 to 0.82 before calibration and fell to zero afterwards ($|r| < 0.001$), so differences among the three scores are no longer driven by a shared age component. The three objectives share one source population (Methods). All scores are out-of-fold predictions from five-fold cross-validation, avoiding in-sample fitting bias. Training-fold versus out-of-fold agreement is shown in Fig. 2f. The 14 organs $\times$ 3 objectives yield 42 organ aging scores, which are the basis of every analysis that follows.

AA clocks reached out-of-fold R^2 values ranging from 0.37 (muscle) to 0.67 (brain), with mean absolute errors of 3.8 to 5.3 years (Fig. 1a; per-organ predicted-age scatters in Fig. 2c). Concordance indices for OS-Mort varied widely, from 0.70 (intestine) to 0.91 (kidney). In contrast AC-Mort clocks were nearly organ-invariant, clustering tightly between 0.74 (breast) and 0.79 (brain), representing a quarter of the OS-Mort variance spread (Fig. 1b). Because the 14 AC-Mort clocks were established to predict a shared systemic outcome (death from any cause), their performance converges despite relying on distinct, organ-specific features. Conversely, OS-Mort clocks predict organ-specific causes of death, inherently reflecting each organ's unique biology and event frequency.

At the individual level, the three clock types diverged considerably. Cross-organ score correlations were lowest for the OS-Mort clocks (mean $r = 0.50$), highest for the systemic AC-Mort clocks (0.80) and intermediate for AA (0.56; Fig. 1c). Within a given organ, AA and OS-Mort scores exhibited the weakest concordance among the three pairings (mean $r = 0.41$; Fig. 1d). Furthermore, the extreme agers identified by each metric (score $z > 1.5$) overlapped minimally, yielding a Jaccard index of only 0.12 to 0.27 (Fig. 1e). Sex further exacerbated the separation: across the 12 mixed-sex organs, both mortality clocks consistently yielded higher scores for males (mean $|d| = 0.41$ and 0.39), corroborating the well-known survival disadvantage. By contrast, AA clocks showed negligible sex difference (mean $|d| = 0.07$), and trended lower for males in 11 of 12 organs (Fig. 1f).

Crucially, the mortality clocks successfully stratified survival across every organ. All 14 OS-Mort clocks significantly separated survival curves (log-rank $P < 10^{-4}$ in thirteen organs; Fig. 2a), and every age-

adjusted hazard ratio (HR) comparing the top and bottom quartiles was significant. The median HR was 5.99, with the kidney proving the most robust (HR = 23.85, 95% CI: 14.20 to 40.05). AC-Mort clocks achieved similar stratification for all-cause death (median HR = 4.67; Fig. 2b). Per-organ HRs are shown in Fig. 2d and Supplementary Data 1. Male-versus-female all-cause survival and full sex effect sizes for all three clock types are provided in Fig. 2e. Ultimately, applying 3 distinct training objectives to a single feature set yields scores that rank individuals differently, demonstrating that age acceleration and mortality risk represent distinct biological trajectories rather than dual readouts of a single underlying quantity.

Organ-specific Training Isolates Organ-matched Disease Signals

We subsequently tested each clock against 30 incident diseases using Cox regression (top versus middle score quintile, adjusted for age and sex), defining incident disease as the first occurrence during follow-up after excluding prevalent baseline diagnoses. The strongest associations were exclusively derived from OS-Mort clocks, for instance, brain to Alzheimer's disease (HR = 8.23, 95% CI: 5.22 to 12.98), pancreas to type 2 diabetes (HR = 7.66, 95% CI: 6.60 to 8.89), lung to chronic obstructive pulmonary disease (HR = 7.45, 95% CI: 6.14 to 9.03) and kidney to chronic kidney disease (HR = 7.18, 95% CI: 6.09 to 8.46; Fig. 3a). These strongest associations are all organ-matched: throughout, a disease is matched to an organ when it pathologically involves primarily that organ. Overall, the OS-Mort clock was the strongest of the three clock types in 17 of the 23 pre-specified organ-matched disease pairs (full pair table, matching rationale, and all 1,260 associations are shown in Supplementary Data 2).

While AC-Mort clocks concentrated disease events into the top quintile at a comparable magnitude (median 68% across pairs, versus 65% for organ-specific and 57% for age acceleration; Fig. 3b) and generated more significant associations overall, the two mortality objectives demonstrated distinct domains of efficacy. The OS-Mort clock outperformed the others in 17 of the 23 organ-matched pairs (Fig. 3c), whereas the all-cause performance surplus was driven almost entirely by non-matched pairs (Fig. 3d). This indicates that all-cause training captures a generic systemic signal of failing health, while organ-specific training isolates an organ's intrinsic pathology. The six exceptions largely involved cardiovascular diseases, which are themselves leading causes of death, or rarely fatal conditions such as fractures. AA clocks were generally weaker on these pairs, as their training objective is not intrinsically disease-linked.

The prostate clock provides a concrete mechanistic illustration of this divergence. Its organ-specific clock significantly predicted incident prostate cancer (HR = 2.62, 95% CI: 2.09 to 3.29), whereas the all-cause clock yielded a null association (HR = 0.84, 95% CI: 0.68 to 1.03) and the AA clock was only marginally significant (HR = 1.28, 95% CI: 1.03 to 1.59, $P = 0.03$; Extended Data Fig. 2a). Because prostate cancer accounts for merely 8.4% of cohort deaths, all-cause training dilutes the signal roughly twelve-fold (Extended Data Fig. 2b). This dilution flattens the quintile separation of cumulative incidence (Extended Data Fig. 2c) and demotes the prostate-specific antigen KLK3 from first to 30th in feature importance, shifting weight toward systemic markers such as GDF15 (Extended Data Fig. 2d).

Conversely, benign prostatic hyperplasia, an age-progressive but non-lethal condition, associated most strongly with the AA clock (HR = 1.73, 95% CI: 1.47 to 2.03), followed by the organ-specific (HR = 1.59, 95% CI: 1.35 to 1.86) and all-cause (HR = 1.29, 95% CI: 1.10 to 1.50) clocks. The distribution of the organ-specific advantage over all-cause models across the 23 matched pairs, analyzed as a function of disease fatality, is detailed in Extended Data Fig. 2e. These findings confirm that all-cause training inherently dilutes organ-specific disease signals during model fitting (a loss that organ-specific training prevents) and that this dilution is not an analytical artefact of competing-risk adjustments.

Distinct Genetic Architectures and Molecular Layers

Genome-wide association studies across all 42 organ aging phenotypes (11,236 to 23,878 participants per phenotype, adjusted for age, sex, and ten principal components) identified 372 phenotype-locus associations. Merging lead SNPs by position ($\pm$ 500-kb windows) collapsed these into 172 independent loci (Fig. 4a). Stratified by objective, AA involved 105 independent loci, OS-Mort 86, and AC-Mort 63. Notably, 114 loci were significant under a single objective exclusively (60 for AA, 39 for OS-Mort, 15 for AC-Mort), 34 spanned two objectives, and 24 were shared across all three (Supplementary Data 3).

The genetic correlation structure mirrored these phenotypic divergences. Cross-organ genetic correlations were lower for AA and OS-Mort clocks (mean r_g = 0.54 and 0.40) and higher for AC-Mort clocks (0.70), aligning with the score-level gradient and exhibiting SNP heritability of 8% to 26% (Fig. 4b). The decoupling of age acceleration from mortality risk persisted at the genetic level within individual organs, the AA-to-OS-Mort pairing displayed the weakest and most variable genetic correlations (r_g 0.21 to 0.65, with adipose and liver tissues being the most decoupled). In contrast, the two mortality clocks remained genetically concordant in 13 of 14 organs (r_g 0.54 to 0.96), with the prostate serving as the sole exception (0.28; Fig. 4c). Test statistics showed no evidence of confounding inflation (genomic-control $\lambda \leq 1.074$, LDSC intercepts consistent with unity; Methods). Fine-mapping across the 227 genome-wide-significant regions yielded 388 variants with a posterior inclusion probability (PIP) > 0.5, and 263 with PIP > 0.95. For instance, the APOE ϵ 4-defining variant rs429358 was definitively resolved as the single causal candidate for the organorgan-specific signal at its locus (PIP = 1.0; Supplementary Data 4).

Because most GWAS loci reside in non-coding regions, interpreting the 172 loci required mapping them to effector genes via a three-tier framework (transcriptome-wide association, colocalization, and summary-based Mendelian randomization across GTEx tissues). Integrating these tiers with LD-refined plasma protein-QTL colocalization yielded 13 high-confidence candidate genes (Fig. 5; Extended Data Fig. 3a–c). These prioritized genes segregate into two primary biological programs such as fatty-acid and lipoprotein metabolism (the *FADS* cluster and *APOE*), and secreted markers of organ dysfunction (*CPS1*, *CFH*, *NPPB*, *GDF15*, *HAVCR1*, *SFTPD*, *SERPINA1*, and *PLA2G2A*).

Polygenic Scores and Causal Disease Links

To determine whether the genetic component of organ aging reproduces disease associations independently of measured omics data, we applied four complementary genotype-only approaches (Fig. 7a). Genetic correlations recapitulated the established pattern where organ-matched pairs exhibited higher genetic correlations with disease than non-matched pairs exclusively under organ-organ-specific training (Extended Data Fig. 4a,b). Polygenic scores successfully stratified incident disease risk in the 383,227 non-overlapping validation participants. Most notably, individuals in the top quintile of the brain OS-Mort score exhibited 3.25 times the Alzheimer's risk of the middle quintile (95% CI: 2.91 to 3.64) and 3.87 times that of the bottom quintile (Fig. 7b,d). In absolute terms, the top 1% of the brain score conferred a 4.0% Alzheimer's prevalence, roughly five-fold higher than the 0.8% population average (Fig. 7c). This stratification was objective-specific, as the corresponding AA and AC-Mort scores failed to stratify Alzheimer's risk entirely (Fig. 7d). A dose-response relationship appeared only for the organ-specific score (Fig. 7e); the matched organ ranked first among all 42 scores, and only the brain association hinged on a single locus, APOE (Extended Data Fig. 4c,d).

Mendelian randomization (MR) provided robust causal support. Internal two-sample MR identified 22 Bonferroni-significant organ-to-disease associations, 14 of which were risk-increasing (Fig. 7g). Bidirectional FinnGen MR revealed 67 forward significant pairs aligning seamlessly with internal results (Fig. 7f). Sensitivity analyses showed no evidence of directional pleiotropy, weak instruments or single-variant dominance (Extended Data Fig. 5a–d). Modifiable-exposure MR mapped modifiable exposures specifically to the mortality axis, for instance, higher body-mass index significantly elevated the mortality-clock scores of 13 organs, while vigorous physical activity attenuated scores in 11, with neither exerting a significant effect on the AA clocks (Extended Data Fig. 5e,f). Integrating these four independent lines of evidence revealed four associations that achieved simultaneous significance across polygenic scoring, internal MR, and FinnGen MR including brain (OS-Mort) to Alzheimer's disease, kidney (OS-Mort) to gout, kidney (AA) to hypertension, and muscle (OS-Mort) to Alzheimer's disease. Ultimately, from multi-omics scores to genetic variants to incident disease, convergent methodologies robustly validate these specific organ-disease architectures.

Discussion

In this study, we developed multi-omics, organ-specific aging clocks for 14 organs, training each on three objectives (chronological-age acceleration, organ-specific mortality and all-cause mortality) from one shared multi-omics feature set. This comparative framework reveals two fundamental separations in the biology of aging. First, an organ's age acceleration and its mortality risk constitute distinct, uncoupled dimensions: they rank individuals differently, are driven by divergent molecular signatures, and map to entirely separate genetic architectures. Crucially, we demonstrate that organ-specific training recovers intrinsic, organ-matched disease signals that are heavily diluted by conventional all-cause mortality models. Second, this approach successfully delineates systemic aging signals from those restricted to a single organ which is a resolution achievable only through organ-specific, multi-objective modeling..

Prior computational aging models have largely relied on single training objectives. While first-generation clocks captured molecular variations tracking chronological age with limited connection to disease homeostasis, subsequent models like GrimAge and PhenoAge were explicitly trained on mortality and lifespan proxies, substantially improving the prediction of morbidity (Lu et al. 2019; Levine et al. 2018). However, recent organ-specific clocks have continued to rely on either chronological age or all-cause mortality (Oh et al. 2023, 2025; Goeminne et al. 2025; Zhu et al. 2026; Anagnostakis et al. 2025). Because they do not parse specific causes of death, these models cannot resolve whether age acceleration and mortality risk diverge. Our organ-specific objective exposes localized organ biology that all-cause training obscures. For example, because prostate cancer accounts for a small fraction of overall cohort mortality, all-cause training effectively abolishes its disease signal; organ-specific training, however, recovers it entirely. Furthermore, unlike previous reported clocks using only single dimension of the profiling, the signatures we established integrate across multiple biological layers: : Olink-profiled secreted proteins reflecting genetic and epigenetic regulation, metabolic variations indexing intracellular shifts, and clinical biomarkers reflecting downstream phenotypic morphology. To our knowledge, this represents the first multi-dimensional, organ-specific clocks to simultaneously predict incident diseases and disease-specific mortality.

These phenotypic separations are deeply related to the distinct genetic architecture, rendering them immune to the confounding that can affect predictive modeling.. Genome-wide association analyses of the clock scores resolve the age-acceleration and mortality signals into near-orthogonal axes. Across independent loci, these two axes are essentially uncorrelated; most loci load exclusively on a single axis, and their anchors are molecularly distinct, with lipid-metabolism variants at the *FADS* locus driving age acceleration, and *APOE* dominating mortality risk. The distinction between systemic and organ-restricted aging is similarly formal rather than merely descriptive. For most age-acceleration loci, the effect size in a single organ differs significantly from the rest (via Cochran's Q heterogeneity), confirming that organ-restricted signals reflect genuine biological heterogeneity rather than arbitrary significance thresholds. Because genetic instruments are fixed at conception and remain unaffected by reverse causation or preclinical disease, the genetic decoupling serves as our strongest evidence that these two axes index fundamentally distinct trajectories of organ biology.

Beyond mapping the fundamental biology of organ aging, this dual-axis framework establishes a concrete path to clinical translation. Conventional all-cause mortality models inherently suffer from competing-risk dilution, obscuring early warning signs of highly specific, yet locally fatal, organ pathologies, a limitation perfectly illustrated by the prostate paradox. By evaluating all 14 organs independently under organ-specific objectives, our approach recovers per-organ comorbidity risk assessments that would otherwise remain hidden. This capability paves the way for targeted clinical surveillance, allowing for the identification of specific organs accelerating toward disease prior to systemic failure. The robust replication of these risk profiles using polygenic scores ensures that this framework can be widely applied for population-level risk stratification even in the absence of deep multi-omics phenotyping. Furthermore, the prioritised secreted markers offer immediate translational value as actionable clinical targets. For instance, monitoring GDF15 for systemic cellular stress, NPPB

for localized cardiac dysfunction, and HAVCR1 for targeted kidney injury provides concrete molecular entry points for organ-directed monitoring and future evaluation of these proteins as therapeutic targets (Han et al. 2002; Alderson et al. 2016; Hartiala et al. 2016), providing.

Several limitations remain. First, our discovery cohort is of predominantly European ancestry. Validation across different ancestries is necessary. Second, organ attribution relies on plasma proxies. While proteins are anchored by stringent Human Protein Atlas tissue-enrichment threshold and metabolites are selected by regularized regression, neither was directly measured in situ; future, tissue-level or imaging-anchored studies are needed to confirm these attributions. Finally, these clocks rely on cross-sectional plasma sampling; the breast clock, which predicts systemic disease rather than tissue-confined breast cancer biology, illustrates this ceiling. Longitudinal multi-omics resampling and multimodal imaging represent natural next steps.

In conclusion, this multi-omics evaluation of 14 organ-specific clocks demonstrates that chronological age acceleration and mortality risk are not merely manifestations of a single aging rate, but rather divergent, genetically uncoupled dimensions of organ biology. By establishing that these two axes possess entirely separable phenotypic, molecular, and genetic architectures, we expose the critical blind spots of conventional, single-objective. Furthermore, our organ-specific training framework resolves the dilution inherent in all-cause mortality models, powerfully recovering localized disease signals that would otherwise remain invisible. Grounded by causal inference and robust polygenic scoring that isolate distinct candidate genes, this dual-axis approach provides a comprehensive, multi-layered molecular atlas of human aging. Ultimately, uncoupling age acceleration from mortality risk paves the way for precise, organ-resolved risk stratification and the targeted interception of age-related diseases long before systemic failure occurs.

Methods

UK Biobank cohort

We used data from the UK Biobank (UKB) (Sudlow et al. 2015), a prospective cohort study of approximately 500,000 participants aged 40–69 at recruitment (2006–2010). We restricted analyses to participants of European ancestry (self-reported White British, confirmed by genetic principal component analysis), yielding 409,206 individuals with available genotype data. The discovery cohort (N = 25,979) comprised the intersection of participants with NMR metabolomics (Nightingale Health platform, 233 features), Olink Explore proteomics (3,244 proteins), and genotype data. The independent validation cohort (N = 383,227) included all remaining genotyped European-ancestry participants with no sample overlap. All analyses were conducted under UKB Application Number 66813.

Organ-specific feature assignment

We assigned biomarkers to 14 organs (adipose, artery, bone, brain, breast, heart, immune, intestine, kidney, liver, lung, muscle, pancreas, prostate) using a knowledge-driven strategy that pairs each feature type with a method-appropriate evidence source. **Proteomic features.** Olink proteins were assigned by Human Protein Atlas (HPA)(Uhlén et al. 2015) tissue-specific RNA expression at ≥ 4 -fold enrichment, following prior plasma-based organ-aging clocks(Oh et al. 2023, 2025; Wang et al. 2026; Goeminne et al. 2025; Zhu et al. 2026); immune-system proteins additionally included lymphoid- or bone-marrow-enriched proteins. **NMR metabolomic features.** Because plasma metabolites circulate systemically and lack a natural organ of origin, we did not assign metabolites to individual organs. Instead, the full panel of 233 NMR metabolites was provided to every organ clock, and organ-relevant metabolites were selected empirically by the L1 penalty during model training. Organ specificity in the metabolomic layer therefore derives from data-driven regularisation rather than a prior mapping, whereas tissue anchoring is provided by the HPA-assigned proteomic features. **Clinical biochemistry features.** The nine PhenoAge biomarkers(Levine et al. 2018) (albumin, alkaline phosphatase, creatinine, C-reactive protein, glucose, lymphocyte percentage, mean cell volume, red cell distribution width, white blood cell count) were applied uniformly across all 14 organs as a systemic-aging baseline. To these we added 20 organ-specific hormones and biochemistry markers (testosterone, IGF-1, vitamin D, SHBG, phosphate, HbA1c, urate, oestradiol, HDL/total cholesterol, haemoglobin, ALT, total and direct bilirubin, PSA, urea, and blood-cell differential counts) selected on the basis of established organ-specific physiology and clinical relevance (from UK Biobank Field IDs not already covered by Olink, NMR or PhenoAge). The combined three-tier panel yielded 254–433 input features per organ (mean 294 features; full per-organ assignment in Supplementary Data 1). Chronological age was not among the input features; it enters only as the prediction target of the AA clocks and in the age-calibration of all scores. To test whether the three platforms are complementary rather than redundant, each clock was retrained under eight feature scopes in the same sample and under the same protocol: the full panel, each platform alone, each pair of platforms, and a control matched to the full panel on feature count. Median performance across the 14 organs ranked full (AA R^2 0.461; OS-Mort C-index 0.810) above NMR plus protein (0.441; 0.808), protein alone (0.323; 0.783), NMR alone (0.257; 0.756) and clinical biochemistry alone (0.106; 0.744), with the feature-count-matched control at 0.323 and 0.783. The full panel was therefore best under every metric, but its margin over the strongest two-platform combination was small (0.020 in R^2 and 0.002 in C-index), so the platforms are treated throughout as complementary rather than individually indispensable.

Aging clock training

Age acceleration (AA) clocks. For each organ, we trained ElasticNet regression(Zou and Hastie 2005) models to predict chronological age from organ-specific biomarkers. All available participants were used for training, with no disease-based exclusion applied, matching the mortality clocks so that the age-acceleration and mortality comparisons draw on the same source population (N = 13,492 to 28,754 per organ for age acceleration; 13,456 to 28,676 for the mortality clocks, which additionally require mortality-registry linkage). Between 36 and 78 participants per organ (at most 0.3%) were absent from the linkage file and therefore entered the age-acceleration clocks only; their age and sex distributions matched those

of retained participants. Per-organ sample sizes for all three objectives are given in Supplementary Data 1. Hyperparameters were selected once per organ on the full training set by ElasticNetCV (internal 5-fold), searching the L1 ratio over {0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99} and the penalty α over 50 values log-spaced in $[10^{-4}, 10]$. We then generated out-of-fold (OOF) predictions by 5-fold cross-validation: in each fold the model (with the fixed selected α and L1 ratio) was refit on four folds and applied to the held-out fold, so every participant's score comes from a model fitted without that participant's own observations. Because α and the L1 ratio were selected once on the full sample, these scores are out-of-fold with respect to model fitting but not with respect to hyperparameter selection. Re-selecting both hyperparameters independently inside every outer fold changed AA R^2 by at most 0.001 and the mortality-clock concordance index by at most 0.036 (median – 0.0004 across the 42 phenotypes; 3 of 42 exceeded 0.02 in absolute value), so the reported performance does not depend on this shortcut. Sex was included as a covariate for all organs except prostate (male-only) and breast (female-only). Models were implemented using scikit-learn (v1.4.1) under Python 3.10. All downstream analyses use these OOF predictions.

Mortality clocks. Organ-specific mortality (OS-Mort) clocks were trained to predict time-to-death from organ-relevant causes (ICD-10 codes mapped to each organ); all-cause mortality (AC-Mort) clocks used time-to-death from any cause. Deaths and their underlying causes were ascertained through the national death registries linked by UK Biobank (ICD-10 underlying cause of death). Training used all available samples (not restricted to healthy individuals). We fitted L1-penalised (LASSO) Cox proportional-hazards models implemented as a custom PyTorch model (Cox partial-likelihood loss plus an L1 penalty, optimised by Adam); the penalty strength α (30 values log-spaced in $[10^{-4}, 1]$) and epoch count were selected by 5-fold inner cross-validation maximising the concordance index (scikit-survival). As for the AA clocks, per-individual scores were generated by 5-fold OOF prediction with the fixed selected hyperparameters. Concordance was evaluated on the OOF scores using both Harrell's and Uno's C (medians across organs 0.733 and 0.705). Because deaths from other causes compete with the organ-specific endpoint, and outnumber it by 1.5-fold to 93-fold depending on the organ, we also fitted Fine-Gray subdistribution hazard models: subdistribution hazard ratios ran 0.83 to 0.95 times the corresponding cause-specific estimates (median 0.91), and cumulative incidence functions were used in place of naive Kaplan-Meier estimates, which overstated 5- and 10-year risk in the top quintile by a median of 2.3% and 6.5% (maximum 8.5%). The proportional-hazards assumption was tested by scaled Schoenfeld residuals and was violated for 9 of the 14 OS-Mort clocks; in 11 of 14 the hazard ratio attenuated with follow-up time, the direction expected if part of the early signal reflects preclinical disease. Time-stratified and time-dependent estimates are in Supplementary Data 2. For survival stratification (Fig. 2a,b), participants were split into quartiles of each out-of-fold mortality score, and Kaplan-Meier curves with age-adjusted Cox models compared the top against the bottom quartile (Q4 vs Q1).

Clock score calibration

To remove the dependence of predicted age on chronological age, we computed calibrated age acceleration as the residual from a linear regression of predicted age on chronological age, fitted on the training set and applied to all samples(Oh et al. 2023):

$$\text{Age acceleration} = \text{Predicted age} - (\beta_0 + \beta_1 \times \text{Age})$$

After calibration, the Pearson correlation between age acceleration and chronological age was effectively zero for all 14 organs ($|r| < 0.007$; Supplementary Note 1.1). For mortality clocks, we applied an analogous calibration: the mortality score was first transformed to a mortality age acceleration (predicted mortality risk residualised against the Cox baseline), then a linear regression of this mortality age acceleration on chronological age was fitted on the training set and the residuals were used as the calibrated mortality score. This ensures that downstream associations reflect organ-specific mortality risk deviation rather than age-correlated variation.

Disease association analysis

We assessed associations between organ aging phenotypes and 30 incident diseases (ICD-10 defined, diagnosed after the baseline assessment) using Cox proportional hazards regression. For each organ–disease pair, participants were stratified into quintiles of clock score. We compared the top quintile (Q5, fastest aging) to the middle quintile (Q3, population average), with time-to-event defined as years from assessment to first diagnosis (or censoring). This design measures excess risk relative to average aging rather than comparing extreme groups, providing more stable estimates and a clinically interpretable reference. Models were adjusted for age at recruitment and sex (sex omitted for sex-restricted diseases: prostate cancer, BPH in males; breast cancer in females). Hazard ratios (HR) with 95% confidence intervals were reported, both per standard deviation of the score and for the quintile contrast. Bonferroni correction was applied within the family being tested: $P < 0.05/30$ for the ten curated pairs displayed in Fig. 3a and $P < 0.05/1,260$ for the full 42 phenotype $\times$ 30 disease matrix. Four sensitivity analyses were run for the organ-matched pairs. Participants with the disease already recorded at baseline were excluded; events occurring within the first two and five years of follow-up were additionally excluded to test for reverse causation from preclinical disease; competing mortality was handled by Fine-Gray subdistribution hazards alongside the cause-specific Cox models; and each clock was refit after removing the diagnostic or near-diagnostic biomarkers of its matched disease. Across the eleven organ-matched pairs for which a diagnostic marker exists, ten retained 73 to 119% of the per-standard-deviation effect after that removal, whereas prostate cancer retained 13% (HR 1.57 to 1.07), identifying KLK3 (PSA) as the dominant contributor to that particular association. The two- and five-year washouts attenuated the median per-standard-deviation HR from 1.81 to 1.67 and 1.63, and Fine-Gray estimates tracked the cause-specific estimates closely (median subdistribution HR 1.68, 1.57 and 1.56). Full per-pair results are in Supplementary Data 2.

Hazard-ratio cutoff for the pairwise scatter (Fig. 3d).

For the pairwise clock-comparison panels in Fig. 3d, organ–disease pairs were dichotomised as “predictive” versus “non-predictive” at $HR \geq 2$ (Q5 vs Q3). Because no established HR cutoff exists for the Q5-vs-Q3 framework in the organ-aging or polygenic-risk-score literature, we anchored this threshold via a data-driven Gaussian mixture model on $\log(HR)$ across the 1,251 of 1,260 organ–clock–disease pairs with a valid HR (9 pairs had too few events in Q3 or Q5 to estimate an HR). Bayesian information criterion preferred a two-component fit (BIC = 1,498 for $k = 2$ versus 1,715 for $k = 1$ and 1,520 for $k = 3$), yielding a null/weak component (mean HR = 1.44, 74% of pairs) and a signal/strong component (mean HR = 2.99, 26% of pairs) with density crossover at HR = 2.43. We round to $HR \geq 2$ as a clean threshold near this empirical break-point, which also coincides with the conventional “strong association” benchmark for clinical predictors (Steyerberg 2009) and lies within the magnitude range used by polygenic-risk-score studies for high-risk classification (Khera et al. 2018; Mavaddat et al. 2019).

Genome-wide association studies

GWAS were performed for all 42 organ aging phenotypes using PLINK2 (Chang et al. 2015) linear regression ($-\text{glm}$) on 8.19 million autosomal SNPs in the 25,979-sample discovery cohort; after phenotype and covariate completeness filtering the per-phenotype analysis sample size ranged from 11,236 (prostate, male only) to 23,878. Covariates included age, sex, and the first 10 genetic principal components. SNPs were restricted to minor allele frequency > 0.01 and minor allele count ≥ 50 ; beyond the central UK Biobank imputation quality control (Bycroft et al. 2018), no additional call-rate or Hardy-Weinberg filters were applied. The genome-wide significance threshold was $P < 5 \times 10^{-8}$. Independent loci were identified by distance-based LD clumping (± 500 kb). Test-statistic inflation was minimal across the 42 phenotypes (genomic-control λ 1.014 to 1.074) and LD-score-regression intercepts were consistent with unity (0.997 to 1.018, standard errors 0.006 to 0.008), indicating that the modest inflation present reflects polygenicity rather than cryptic relatedness or residual population stratification. Because a clock score is a weighted combination of measured plasma analytes, some of its genetic signal is expected to be the cis-QTL of an input protein. We quantified this directly: across the 42 phenotypes a median of 9.1% of lead SNPs (range 0 to 33.9%; 6 phenotypes with none) fell within the cis window of an Olink protein that the same clock uses as an input, exceeding 25% only for immune age acceleration, immune organ-specific mortality and liver organ-specific mortality. All organ aging phenotypes were generated using OOF 5-fold cross-validation predictions to eliminate in-sample fit bias.

Statistical fine-mapping.

Lead SNPs (at $P < 5 \times 10^{-6}$) were merged into ± 500 kb regions per clock type, and each region was fine-mapped with FINEMAP v1.4.2 (Benner et al. 2016) in shotgun-stochastic-search mode ($-\text{sss}$, $-\text{n-causal-snp}$ 5) using in-sample LD from the discovery genotypes. We report SNPs with posterior inclusion probability (PIP) > 0.5 and > 0.95 (Supplementary Data 4).

LD score regression

SNP-based heritability (h^2) and pairwise genetic correlations (r_g) were estimated using LD score regression (LDSC)(Bulik-Sullivan et al. 2015) with pre-computed European LD scores from the 1000 Genomes Project. GWAS summary statistics were filtered to HapMap3 SNPs (approximately 1.2 million variants). Disease genetic correlations were computed against FinnGen Release 10 GWAS summary statistics for 29 diseases (the 30th, pancreatic cancer, was excluded because its disease-GWAS SNP-heritability did not converge to a non-zero estimate); defining a test family of 1,218 organ-disease-clock combinations (14 organs x 29 diseases x 3 clock types). The genomic coordinate reference was GRCh37.

Polygenic score construction and validation

PGS were constructed using clumping and thresholding (C + T). Lead SNPs were selected from the discovery GWAS at $P < 5 \times 10^{-8}$ with distance-based clumping (± 500 kb), yielding 372 lead-SNP-phenotype pairs across the 42 phenotypes. Individual-level PGS were computed as the sum of effect-size-weighted allele dosages using PLINK2 –score. PGS were validated in the 383,227-sample independent cohort (no overlap with discovery). Each PGS was stratified into quintiles, and associations with 30 diseases were assessed by logistic regression of incident disease on quintile membership (top versus middle quintile, Q5 vs Q3, as the primary contrast and top versus bottom, Q5 vs Q1, as a secondary contrast), adjusted for age, sex and 10 genetic PCs, with Benjamini–Hochberg FDR correction across all 1,260 tests. As a complementary extreme-tail comparison we fitted logistic models for top-versus-bottom-decile membership (D10 vs D1), and assessed dose-response by per-decile logistic odds ratios with a spline-versus-linear likelihood-ratio test (Supplementary Data 7). Cumulative disease incidence by PGS quintile was computed across age strata. For Bonferroni-significant PGS-disease pairs, the quintile association was repeated after removing the strongest lead SNP and all variants within ± 500 kb of it (leave-top-locus-out; Extended Data Fig. 4d).

Mendelian randomisation

Internal two-sample MR. We used genome-wide significant ($P < 5 \times 10^{-8}$) lead SNPs (distance-clumped within ± 500 kb) from the 25,979-sample discovery GWAS as instruments, and disease GWAS from the independent 383,227-sample validation cohort (PLINK2 logistic regression) as outcomes. The primary method was inverse-variance weighted (IVW) regression assuming multiplicative random effects. Sensitivity analyses included MR-Egger regression (to test for directional pleiotropy via the intercept), weighted median estimation (robust to up to 50% invalid instruments), and heterogeneity testing (Cochran's Q); instrument strength was summarised by the mean per-SNP F -statistic $(\beta/SE)^2$. Instrument strength was benchmarked against the conventional $F > 10$ threshold (Burgess and Thompson 2011). MR estimation relies on the three core instrumental-variable assumptions (relevance, independence from confounders, and exclusion restriction), and reporting follows the STROBE-MR checklist (Skrivankova et al. 2021). Pairs with fewer than 3 instruments were excluded. Bonferroni correction was applied across the 1,140 exposure-outcome pairs that met this instrument threshold. Robustness of the

Bonferroni-significant pairs was assessed by leave-one-out analysis across instruments (Extended Data Fig. 5d).

External MR with FinnGen. Forward MR (organ aging → disease) and reverse MR (disease → organ aging) used FinnGen Release 10 GWAS as the outcome and the exposure data, respectively. The instrument protocol was identical to the internal MR: $P < 5 \times 10^{-8}$ leads, distance-based clumping (± 500 kb), IVW as the primary estimator with MR-Egger and weighted-median as sensitivity analyses, and Bonferroni correction within each direction (1,110 forward and 1,050 reverse pairs).

Modifiable-exposure MR. To test genetic associations of modifiable exposures on organ aging, we used published European-ancestry GWAS for five traits (body-mass index, hand grip strength, sleep duration, coffee intake and vigorous physical activity) as exposures and the 14 organ aging phenotypes as outcomes. Instruments were genome-wide-significant SNPs ($P < 5 \times 10^{-8}$) distance-clumped within ± 500 kb, and effects were estimated by IVW with MR-Egger and weighted-median sensitivity analyses; Bonferroni correction was applied across the 210 exposure–organ tests.

Gene prioritisation

We applied a three-tier framework to prioritise candidate genes at GWAS loci.

TWAS. Transcriptome-wide association studies were performed using pre-computed GTEx v8 gene expression prediction models (GTEx Consortium 2020) across 49 tissues. For each organ, only eQTL models from the corresponding tissue were used. TWAS was implemented as a MASHR-weighted, S-PrediXcan-style summation of eQTL weights over GWAS z-statistics, not with the FUSION software. Bonferroni correction was applied within each organ.

Colocalisation. For TWAS-significant genes, we performed Bayesian colocalisation using an in-house Python implementation of the Wakefield approximate Bayes factor framework of Giambartolomei et al. (algorithmically equivalent to coloc.abf; prior probabilities $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$, prior effect-size standard deviation 0.2 for quantitative traits), taking GTEx v8 significant variant-gene pairs in the organ-matched tissue as the eQTL dataset. Genes with a posterior probability of a shared causal variant (PP.H4) > 0.9 were retained.

SMR. Summary-based Mendelian randomisation was performed with the SMR software (v1.3.1) against GTEx v8 cis-eQTL BESD summary datasets in the organ-matched tissue, together with the accompanying HEIDI heterogeneity test. A gene was called SMR-validated when the SMR P value passed a Bonferroni threshold within its phenotype (0.05 divided by the number of tested probes) and HEIDI was not rejected (HEIDI $P > 0.05$), the latter separating a shared causal variant from linkage. Genes passing all three tiers (TWAS, COLOC and SMR) were designated triple-validated candidate genes.

cis-pQTL validation of plasma-secretory flagship genes. Flagship biomarkers are plasma-secreted proteins whose causal variant is best captured by a protein quantitative trait locus rather than a tissue eQTL, making GTEx-eQTL-based validation indirect. For these we instead tested concordance between

the organ-aging GWAS lead SNP of each gene and a literature cis-pQTL lead for the same protein (from Sun et al. 2018 INTERVAL, Sun et al. 2023 UKB-PPP, Ferkingstad et al. 2021 deCODE and Pietzner et al. 2021), computing pairwise linkage disequilibrium in the 1000 Genomes Phase 3 European panel (Ensembl REST API). A gene was deemed cis-pQTL-validated when the organ-aging lead and the literature cis-pQTL lead were in strong LD ($r^2 \geq 0.8$, i.e. very likely the same causal variant or haplotype); nine such genes also retained a genome-wide-significant lead SNP in the all-NMR GWAS and are reported here. Six of these nine proteins (*NPPB*, *GDF15*, *HAVCR1*, *SFTPD*, *SERPINA1* and *PLA2G2A*) are themselves features of the clocks they associate with, so their cis-pQTL concordance is circular and is labelled as such in Fig. 5; only *APOE*, *CPS1* and *CFH* are independent of the clock feature set. The triple-validated set (4 genes, all at the *FADS* locus) and the cis-pQTL-validated set (9) are disjoint, giving 13 prioritised candidate genes.

Gene Ontology and pathway over-representation analysis of the prioritised gene set was performed with g:Profiler (gprofiler-official Python client; sources GO:BP, GO:MF, GO:CC and KEGG; g:SCS multiple-testing correction at threshold 0.05). These results are descriptive, were not used to select candidate genes, and are provided in Supplementary Data 9.

Three-layer (L1/L2/L3) classification from feature-ablation GWAS. To assign each gene $\times$ organ $\times$ clock pair to a molecular layer, we re-ran the OOF GWAS pipeline at three feature scopes: *Original* (full multi-omics: NMR + Olink + clinical biochemistry), *non-NMR* (NMR features removed; protein + clinical retained), and *protein-only* (Olink proteins alone). For each pair we recorded the minimum GWAS P -value within a ± 500 kb window around the gene body centroid under each scope: P_o (Original), P_a (non-NMR), P_b (protein-only). With $P_{hit} = 5 \times 10^{-8}$ (GW-significance) and $P_{lost} = 5 \times 10^{-6}$ (signal-loss threshold), each pair with $P_o < P_{hit}$ was classified by signal survival across scopes; the same rules apply uniformly.

A pair is NMR-dependent (L1) when both non-NMR scopes lose the signal ($P_a > P_{lost}$ and $P_b > P_{lost}$); protein-direct (L2) when the signal survives the strict protein-only scope ($P_b < P_{hit}$); clinical-biochemistry-mediated (L3) when the non-NMR scope retains genome-wide significance but the protein-only scope collapses ($P_a < P_{hit}$ but $P_b > P_{lost}$); and mixed L2/L3 when the protein-only scope weakens the signal without abolishing it ($P_a < P_{hit}$ and $P_b \in [P_{hit}, P_{lost}]$). Pairs satisfying none of these rules remain unclassified (rare edge cases).

Pairs with $P_o \geq P_{hit}$ (no original GW-sig hit) are marked as “no GW-sig” in Fig. 5 with a saltire cross, indicating that layer assignment is not applicable. Each gene’s *primary* layer is the layer of its strongest (lowest- P_o) organ $\times$ clock pair, which we use for sorting and colour-coding in Fig. 5 and Fig. 6a. Full per-pair classifications and ablation P -values are deposited in Supplementary Data 6.

Gene-label attribution convention

A gene-level minimum P within a fixed ± 500 kb window can assign the same GWAS lead SNP to several neighbouring genes wherever linkage disequilibrium is high; the main offenders are the major

histocompatibility complex (MHC, chromosome 6:25–34 Mb), the APOE/TOMM40 region (chromosome 19:45 Mb) and the FADS1/FADS2/FADS3 cluster (chromosome 11:61.5 Mb). We therefore applied a standardised five-step attribution audit before committing any gene symbol to Manhattan, locus-zoom or gene-matrix plot annotations.

Step 1: Neighbour enumeration. For each candidate gene label, all triple-validated (TWAS, COLOC and SMR, COLOC PP.H4 > 0.8) neighbours within ± 500 kb of the gene body centroid were enumerated.

Step 2: Same-family classification. Triple neighbours were classified as same- or different-family relative to the candidate against six curated family / LD-block sets: *FADS* cluster (*FADS1/2/3*, *TMEM258*, *FEN1*, *MYRF*); HLA class I (*HLA-A/B/C/E/F/G*, *MICA*, *MICB*); HLA class II (*HLA-DR/DQ/DP* paralogues); MHC central (*TCF19*, *CCHCR1*, *PSORS1C1-2*, *POU5F1*, *C4A/B*, *CYP21A2*, *STK19*, *NOTCH4*, *CFB*, *C2*); *KLK* cluster (*KLK1–KLK15*); and *APOE* region (*APOE*, *APOC1/2/4*, *TOMM40*, *NECTIN2*).

Step 3: MHC hard block. Genes within chromosome 6:25–34 Mb were forced to locus-level labels (“MHC class I” for position < 32 Mb, “MHC central” otherwise) regardless of triple-neighbour count, reflecting the intrinsic unresolvability of gene-level attribution at the MHC.

Step 4: Olink-uniqueness rescue. Candidates with remaining different-family triple neighbours were rescued to gene-level labelling if and only if the candidate was the unique Olink-measurable gene at the locus (i.e., all triple neighbours absent from the 3,244-protein Olink Explore panel used in our clock training).

Step 5: Signal-dominance rescue. Candidates not rescued by Step 4 were rescued if the candidate was in the Olink panel and its best Original-scope GWAS *P* across all 42 phenotypes was at least 10^5 -fold stronger than that of every Olink-measurable triple neighbour.

Genes surviving Steps 1–3 with zero triple neighbours, or rescued at Step 4 or 5, kept their gene symbol (Tiers A and A*). Genes whose triple neighbours were all same-family received a family-cluster label (Tier B, e.g. “FADS cluster”). Genes with unrescued different-family neighbours, or genes in the MHC region, received chromosomal-position labels (Tier C, e.g. “chr8:16 Mb”, “MHC class I”). The audit covered the 69 gene labels on main-text and Extended Data Manhattan, scatter and gene-matrix plots: 44 were classified Tier A, 5 Tier A* (including GDF15, rescued by $\sim 10^{170}$ -fold signal dominance over the adjacent Olink-measured neighbour COMP), 2 Tier B (FADS1/2 → “FADS cluster”) and 18 Tier C. The complete per-gene audit table and the audit script are deposited as Supplementary Data 9.

Cross-organ tail-aggregate score for the two-axis decomposition

For the two-axis gene landscape in Fig. 6a we computed, for each gene *g* and each clock type (AA, OS-Mort), a per-gene cross-organ tail-aggregate score

$$S_g = (1/K) \sum_{o=1}^K \max(0, z_{g,o} - z_\tau)$$

where $K = 14$ is the number of organ-aging GWAS, $z_{g,o} = \sqrt{(2 \ln 10 \cdot (-\log_{10} P_{g,o}))}$ is the absolute one-degree-of-freedom Z-score derived from the gene's strongest LD-clumped lead SNP in organ o ($z_{g,o} = 0$ if no genome-wide-significant lead is annotated to g in that organ), and $z_\tau = \sqrt{(2 \ln 10 \cdot 7.30)} \approx 5.82$ is the Z corresponding to the GWAS threshold $P < 5 \times 10^{-8}$ used for LD clumping. The two clock-specific scores are denoted Z_{AA} and Z_{CS} on the axes of Fig. 6a.

This score resembles the truncated product method of Zaykin et al. (Zaykin et al. 2002) in that below-threshold organs contribute zero, but it operates on z-scores with a soft-thresholding rectifier rather than on P-values, and it is not equivalent to Stouffer's combined test. The rectifier $\max(0, z_{g,o} - z_\tau)$ makes the score reflect both how many organs reach genome-wide significance and how strong those signals are, while ignoring sub-threshold noise. Because the 14 organ-aging GWAS share the 25,979-sample discovery cohort and the phenotypes are correlated within each clock type (mean cross-organ phenotypic $r = 0.39$ for AA and $r = 0.47$ for OS-Mort, Fig. 1c), the null distribution of S_g is not standard normal; we therefore use the score only as a descriptive ranking statistic for quadrant assignment, never as a calibrated test.

Statistical analysis

All statistical analyses were performed in Python 3.10 with NumPy (v1.26), pandas (v2.3), scikit-learn (v1.6), scikit-survival (v0.25), PyTorch (v2.7), statsmodels (v0.14.5), lifelines (v0.30.0; disease-association and decile Cox models) and SciPy (v1.15). Gene prioritisation additionally used the standalone SMR binary (v1.3.1) and the g:Profiler Python client; no R packages were used. Multiple testing was corrected using Bonferroni correction (GWAS, MR, disease associations) or Benjamini-Hochberg FDR (PGS validation). Two-sided tests were used throughout unless otherwise specified.

Use of large language models

Large language models (Claude, Anthropic) were used to assist with language editing and analysis-code development, under full author supervision. All study design, analyses, results and conclusions were conceived, verified and approved by the authors, who take complete responsibility for the content.

Declarations

Acknowledgements

This research was conducted using UK Biobank data under Application Number 66813.

Data availability

UK Biobank data are available through the UK Biobank Access Management System (<https://www.ukbiobank.ac.uk>). FinnGen summary statistics are available at <https://r10.finnngen.fi>. GTEx data are available at <https://gtexportal.org>.

Code availability

All analysis code is available at [GitHub repository URL upon publication].

References

1. Alderson HV, Ritchie JP, Pagano S et al (2016) The Associations of Blood Kidney Injury Molecule-1 and Neutrophil Gelatinase-Associated Lipocalin with Progression from CKD to ESRD. *Clin J Am Soc Nephrol* 11(12):2141–2149. <https://doi.org/10.2215/CJN.02670316>
2. Anagnostakis F, Ko S, Saadatinia M et al (2025) Multi-Organ Metabolome Biological Age Implicates Cardiometabolic Conditions and Mortality Risk. *Nat Commun* 16(1):4871. <https://doi.org/10.1038/s41467-025-59964-z>
3. Belsky DW, Avshalom Caspi DL, Corcoran et al (2022) DunedinPACE, a DNA Methylation Biomarker of the Pace of Aging. *eLife* 11:e73420. <https://doi.org/10.7554/eLife.73420>
4. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016) Bioinformatics 32(10):1493–1501. <https://doi.org/10.1093/bioinformatics/btw018>. FINEMAP: Efficient Variable Selection Using Summary Data from Genome-Wide Association Studies
5. Bulik-Sullivan, Brendan K, Po-Ru Loh, Hilary K, Finucane et al (2015) LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nat Genet* 47(3):291–295. <https://doi.org/10.1038/ng.3211>
6. Burgess S, Thompson SG (2011) Avoiding Bias from Weak Instruments in Mendelian Randomization Studies. *Int J Epidemiol* 40(3):755–764. <https://doi.org/10.1093/ije/dyr036>
7. Bycroft C, Freeman C, Petkova D et al (2018) The UK Biobank Resource with Deep Phenotyping and Genomic Data. *Nature* 562:203–209. <https://doi.org/10.1038/s41586-018-0579-z>
8. Chang CC, Carson C, Chow, Laurent CAM, Tellier S, Vattikuti SM, Purcell, James JL (2015) Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* 4:7. <https://doi.org/10.1186/s13742-015-0047-8>
9. Ferkingstad E, Sulem P, Atlason BA et al (2021) Large-Scale Integration of the Plasma Proteome with Genetics and Disease. *Nat Genet* 53(12):1712–1721. <https://doi.org/10.1038/s41588-021-00978-w>
10. Goeminne LJE, Vladimirova A, Eames A et al (2025) Plasma Protein-Based Organ-Specific Aging and Mortality Models Unveil Diseases as Accelerated Aging of Organismal Systems. *Cell Metabol* 37(1):205–222. <https://doi.org/10.1016/j.cmet.2024.10.005>
11. GTEx Consortium (2020) The GTEx Consortium Atlas of Genetic Regulatory Effects Across Human Tissues. *Science* 369(6509):1318–1330. <https://doi.org/10.1126/science.aaz1776>

12. Han WK, Bailly V, Abichandani R, Thadhani R, and Joseph V. Bonventre (2002) Kidney Injury Molecule-1 (KIM-1): A Novel Biomarker for Human Renal Proximal Tubule Injury. *Kidney Int* 62(1):237–244. <https://doi.org/10.1046/j.1523-1755.2002.00433.x>
13. Hannum G, Guinney J, Zhao L et al (2013) Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* 49(2):359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
14. Hartiala JA, Wilson WH, Tang Z, Wang et al (2016) Genome-Wide Association Study and Targeted Metabolomics Identifies Sex-Specific Association of CPS1 with Coronary Artery Disease. *Nat Commun* 7:10558. <https://doi.org/10.1038/ncomms10558>
15. Horvath S (2013) DNA Methylation Age of Human Tissues and Cell Types. *Genome Biol* 14(10):R115. <https://doi.org/10.1186/gb-2013-14-10-r115>
16. Julkunen H, Cichońska A, Tiainen M et al (2023) Atlas of Plasma NMR Biomarkers for Health and Disease in 118,461 Individuals from the UK Biobank. *Nat Commun* 14:604. <https://doi.org/10.1038/s41467-023-36231-7>
17. Khera AV, Mark Chaffin, Krishna G, Aragam et al (2018) Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations. *Nat Genet* 50(9):1219–1224. <https://doi.org/10.1038/s41588-018-0183-z>
18. Kivimäki M, Frank P, Pentti J et al (2025) Proteomic Organ-Specific Ageing Signatures and 20-Year Risk of Age-Related Diseases: The Whitehall II Observational Cohort Study. *Lancet Digit Health* 7(3):e195–204. <https://doi.org/10.1016/j.landig.2025.01.006>
19. Levine ME, Ake T, Lu A, Quach et al (2018) An Epigenetic Biomarker of Aging for Lifespan and Healthspan. *Aging* 10(4):573–591. <https://doi.org/10.18632/aging.101414>
20. López-Otín C, Blasco MA, Partridge L, Serrano M, and Guido Kroemer (2013) The Hallmarks of Aging. *Cell* 153(6):1194–1217. <https://doi.org/10.1016/j.cell.2013.05.039>
21. López-Otín C, Blasco MA, Partridge L, Serrano M, and Guido Kroemer (2023) Hallmarks of Aging: An Expanding Universe. *Cell* 186(2):243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
22. Lu AT, Austin Quach JG, Wilson et al (2019) DNA Methylation GrimAge Strongly Predicts Lifespan and Healthspan. *Aging* 11(2):303–327. <https://doi.org/10.18632/aging.101684>
23. Mavaddat N, Michailidou K, Dennis J et al (2019) Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 104(1):21–34. <https://doi.org/10.1016/j.ajhg.2018.11.002>
24. Oh H, Se-Hwee et al (2025) Plasma Proteomics Links Brain and Immune System Aging with Healthspan and Longevity. *Nat Med* 31:2703–2711. <https://doi.org/10.1038/s41591-025-03798-1>
25. Oh HS-H, Rutledge J, Nachun D et al (2023) Organ Aging Signatures in the Plasma Proteome Track Health and Disease. *Nature* 624:164–172. <https://doi.org/10.1038/s41586-023-06802-1>
26. Peng L, Xie R, Holleccek B et al (2026) Development of Age- and Sex-Specific Metabolomics-Based Biological Ageing Clocks for 10-Year Mortality Prediction. *Adv Sci* 13(1):e10189. <https://doi.org/10.1002/advs.202510189>

27. Pietzner M, Wheeler E, Carrasco-Zanini J et al (2021) Mapping the Proteo-Genomic Convergence of Human Diseases. *Science* 374(6569):eabj1541. <https://doi.org/10.1126/science.abj1541>
28. Steyerberg EW (2009) *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer. <https://doi.org/10.1007/978-0-387-77244-8>
29. Skrivankova VW, Rebecca C, Richmond BAR, Woolf et al (2021) Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* 326(16):1614–1621. <https://doi.org/10.1001/jama.2021.18236>
30. Sudlow C, Gallacher J, Allen N et al (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* 12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
31. Sun BB, Joseph C, Maranville JE, Peters et al (2018) Genomic Atlas of the Human Plasma Proteome. *Nature* 558(7708):73–79. <https://doi.org/10.1038/s41586-018-0175-2>
32. Sun BB, Joshua Chiou M, Traylor et al (2023) Plasma Proteomic Associations with Genetics and Health in the UK Biobank. *Nature* 622:329–338. <https://doi.org/10.1038/s41586-023-06592-6>
33. Uhlén M, Fagerberg L, Björn M, Hallström et al (2015) Tissue-Based Map of the Human Proteome. *Science* 347(6220):1260419. <https://doi.org/10.1126/science.1260419>
34. Wang Y et al (2026) Organ-Specific Proteomic Aging Clocks Predict Disease and Longevity Across Diverse Populations. *Nat Aging* 6:162–180. <https://doi.org/10.1038/s43587-025-01016-8>
35. Zaykin DV, Lev A, Zhivotovsky, Peter H, Westfall, Bruce SW (2002) Truncated Product Method for Combining P-Values. *Genet Epidemiol* 22(2):170–185. <https://doi.org/10.1002/gepi.0042>
36. Zhan Y, Shi R, Ruan X et al (2025) Profile of Biological Aging in First Primary Cancers: A Pan-Cancer Analysis of Two Large-Scale Cohorts from the UK and Hong Kong. *BMC Med* 23:570. <https://doi.org/10.1186/s12916-025-04409-z>
37. Zhu R-J et al (2026) Revealing the Genetic Architectures Underlying Organ-Specific Aging Based on Proteomic Data. *Nat Commun* 17:528. <https://doi.org/10.1038/s41467-025-67223-4>
38. Zou H, and Trevor Hastie (2005) Regularization and Variable Selection via the Elastic Net. *J Royal Stat Society: Ser B (Statistical Methodology)* 67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Figures

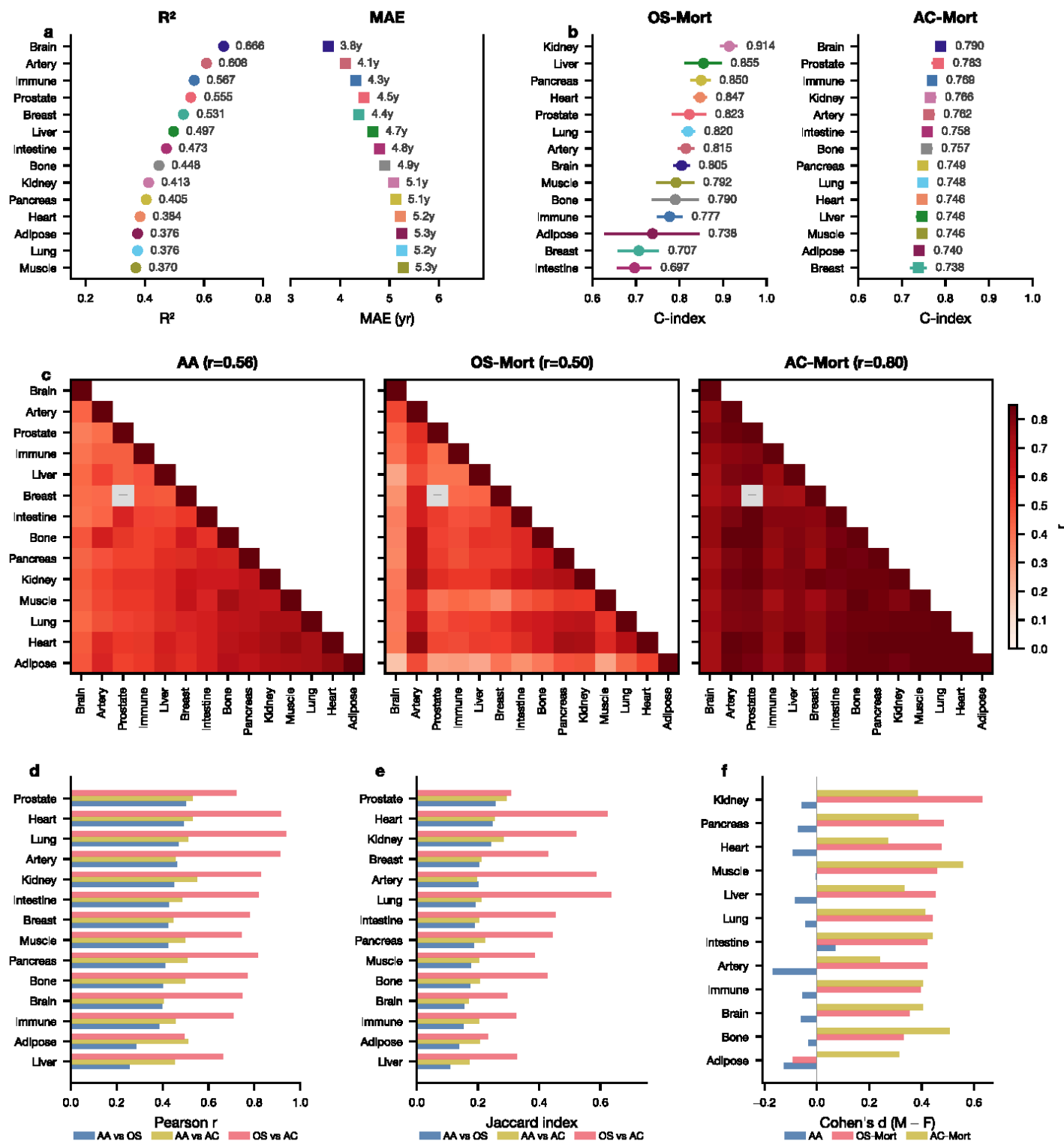


Figure 1

Construction and performance of multi-omics organ aging clocks. (a) Out-of-fold R² and mean absolute error of each organ's age-acceleration clock; bars, bootstrap 95% CI. (b) Concordance index of each organ's organ-specific and all-cause mortality clocks; bars as in a. Organ-specific performance spreads across organs while all-cause packs into a narrow band. (c) Cross-organ score-correlation matrices (Pearson r , lower triangle) for the three clock types, computed from out-of-fold predictions; the

prostate and breast clocks are sex-exclusive and share no individuals, so their correlation is not estimable (grey cells). (d) Within-organ pairwise correlations among the three scores, one row per organ. (e) Per-organ Jaccard overlap of the extreme-ager sets ($z > 1.5$) flagged by each clock-type pair. (f) Cohen's d of male versus female calibrated scores for the 12 mixed-sex organs (positive, male higher); age-acceleration clocks show almost no sex difference and mostly negative signs, while both mortality clocks are consistently male-high. Per-organ values for d to f are in Supplementary Data 1.

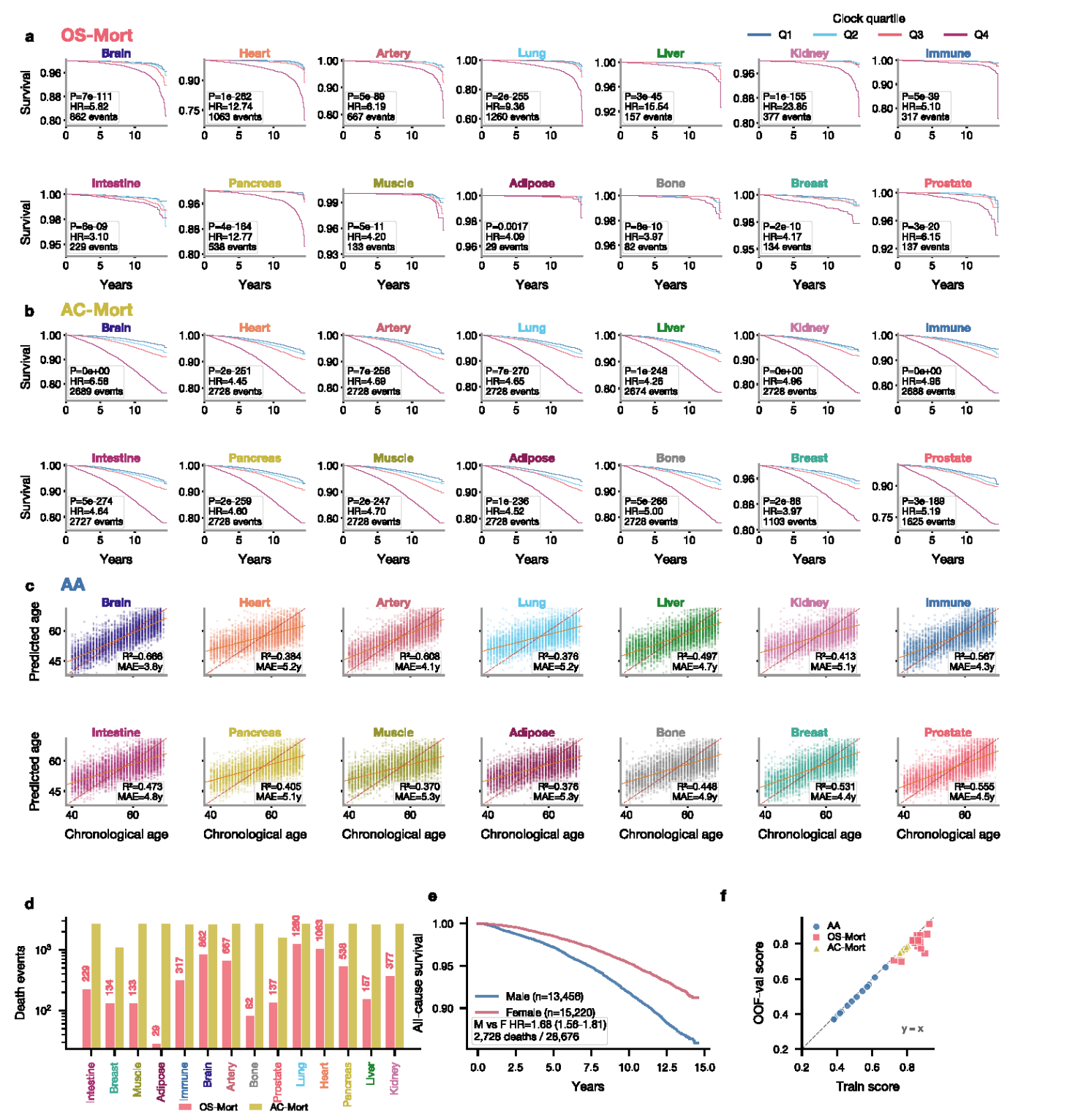


Figure 2

Clock performance and sex differences across 14 organs and three training objectives. (a) Kaplan–Meier survival curves for each organ-specific mortality clock by score quartile (Q1–Q4; outcome, organ-specific death); annotations give the log-rank P, the age-adjusted Q4/Q1 Cox hazard ratio and the death-event count; curves and hazard ratios use out-of-fold predictions. (b) As in a for all-cause mortality clocks, with all-cause death as the outcome. (c) Predicted versus chronological age for each age-acceleration clock (out-of-fold; subsampled for rendering, metrics on the full cohort); red dashed identity line, orange linear fit, annotated R^2 and mean absolute error. (d) Observed death events per mortality clock and organ (log scale); organ-specific counts span a roughly 45-fold range. (e) Male versus female all-cause survival in the pooled clock cohort (Kaplan–Meier); the annotation gives the age-adjusted male-versus-female Cox hazard ratio (HR = 1.68, 95% CI: 1.56 to 1.81) and the death-event count. The male survival disadvantage is the backdrop to the male-high mortality-clock scores in Fig. 1f. (f) Training versus out-of-fold validation scores for all 42 clocks; the dashed $y = x$ line marks the no-overfitting reference.

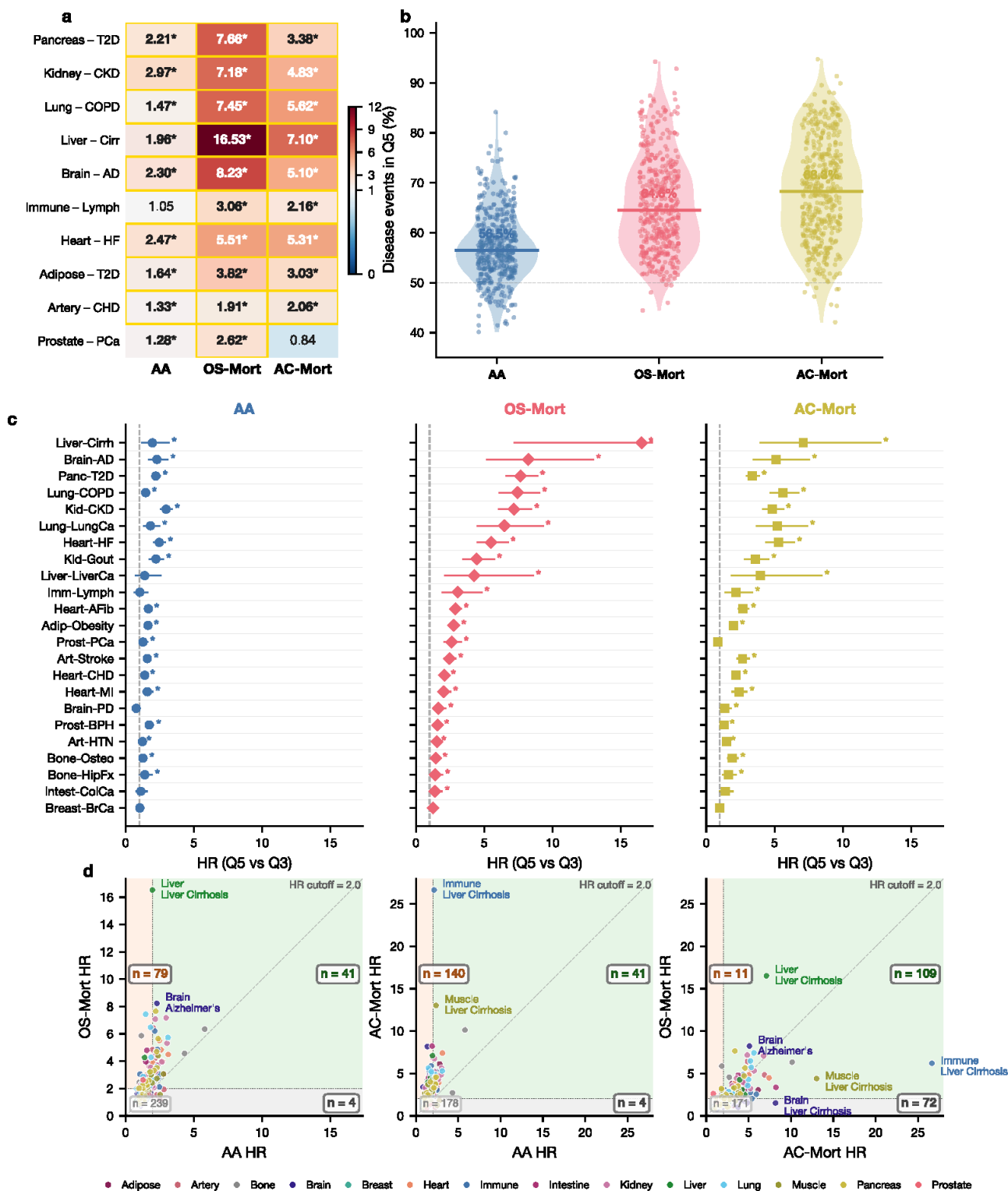


Figure 3

Disease prediction by organ aging clocks. (a) Highlight heatmap of Cox hazard ratios for 10 organ-disease pairs across the three clock types (top versus middle quintile, adjusted for age and sex); asterisk beside an estimate, $P < 0.05$; gold border, Bonferroni-significant ($P < 0.05/30$); full matrix in Supplementary Data 2. The 23 organ-matched pairs are pre-specified by primary pathological involvement. (b) Disease-event concentration in the top quintile relative to the middle quintile (50%, no

enrichment), across organ-disease pairs with at least 20 events. (c) Forest plot of hazard ratios across the three clock types for the 23 organ-matched pairs; asterisk, $P < 0.05$. (d) Pairwise hazard-ratio scatter (organ-specific versus age-acceleration, all-cause versus age-acceleration, organ-specific versus all-cause) with 2x2 counts of pairs reaching $HR \geq 2$ (cutoff from a Gaussian-mixture break-point on log HR; Methods); each dot one organ-disease pair, coloured by organ.

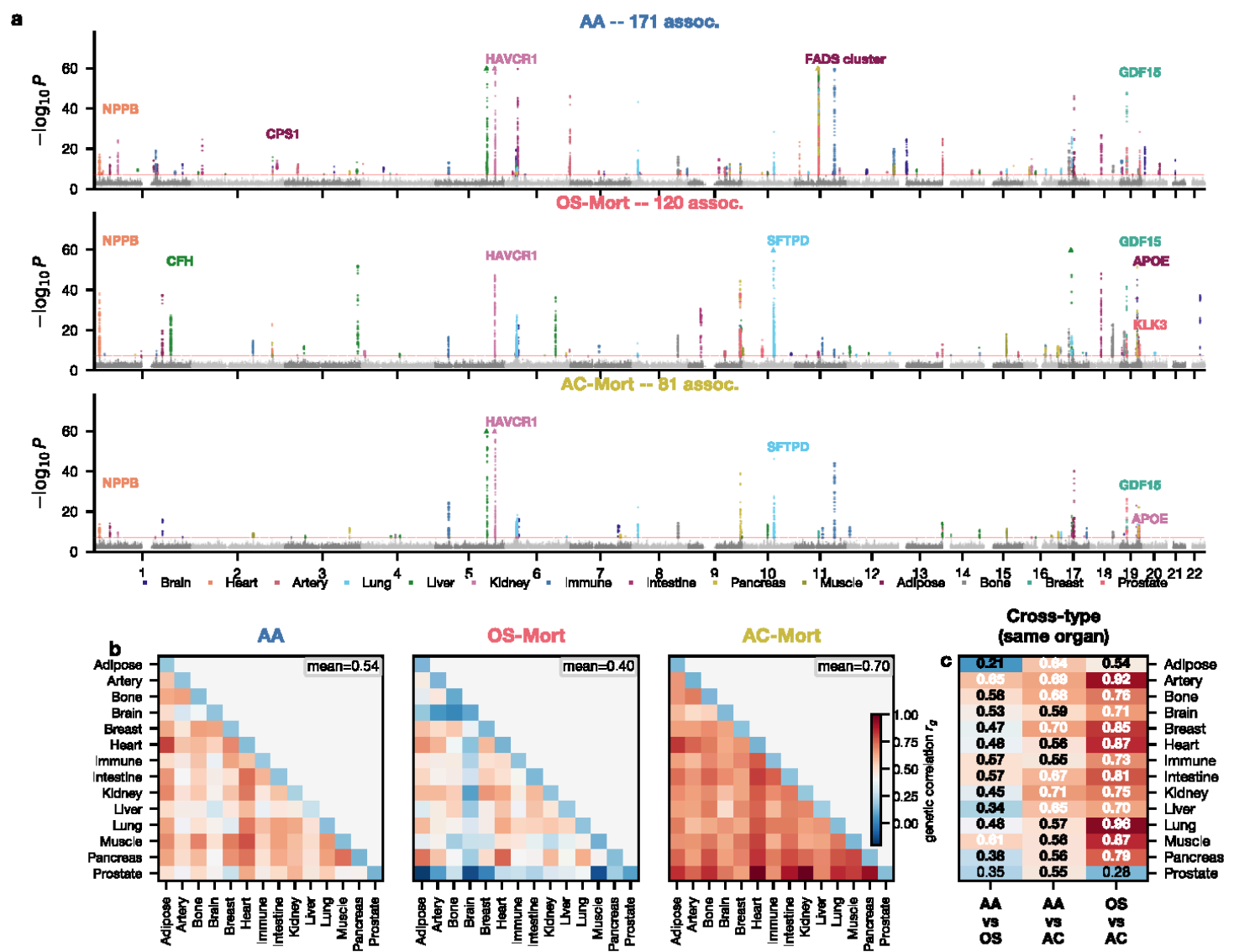


Figure 4

Genetic architecture of organ aging. (a) Stacked Manhattan plots, one row per clock type (age acceleration / organorgan-specific / all-cause); only a curated subset of loci is labelled (manuscript-discussed genes and the strongest peaks per row; full list in Supplementary Data 3, labelling convention in Methods). (b) Cross-organ genetic-correlation matrices (r_g , lower triangle) with SNP heritability (h^2 , diagonal) per clock type, estimated by LDSC. (c) Within-organ cross-type genetic correlations (r_g between age-acceleration, organorgan-specific and all-cause scores, one row per organ).

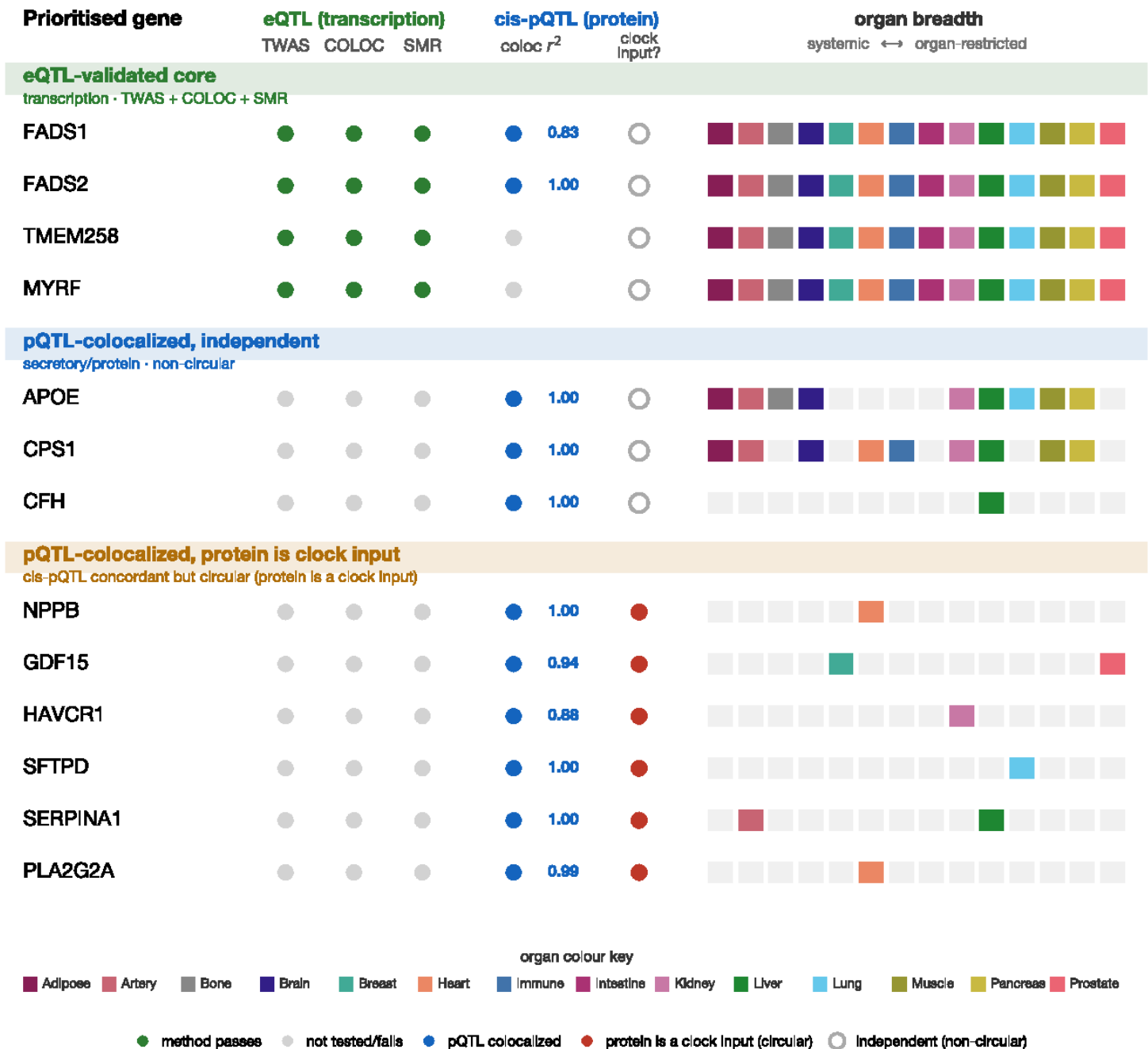


Figure 5

Gene prioritisation identifies organ-associated candidate genes. Evidence matrix for the 13 candidate genes, grouped into three confidence tiers: an eQTL-validated core (TWAS, COLOC and SMR; the *FADS* cluster), pQTL-colocalized and independent (*APOE*, *CPS1*, *CFH*), and pQTL-colocalized genes whose protein is itself a clock input and therefore circular (*NPPB*, *GDF15*, *HAVCR1*, *SFTPD*, *SERPINA1*, *PLA2G2A*). Per gene: filled dots mark eQTL support (TWAS, COLOC, SMR), the cis-pQTL colocalization r^2 is annotated, a red dot flags circularity, and a 14-organ breadth strip runs from systemic to organ-restricted (per-gene layer assignment in Supplementary Data 6; full catalogue in Supplementary Data 9).

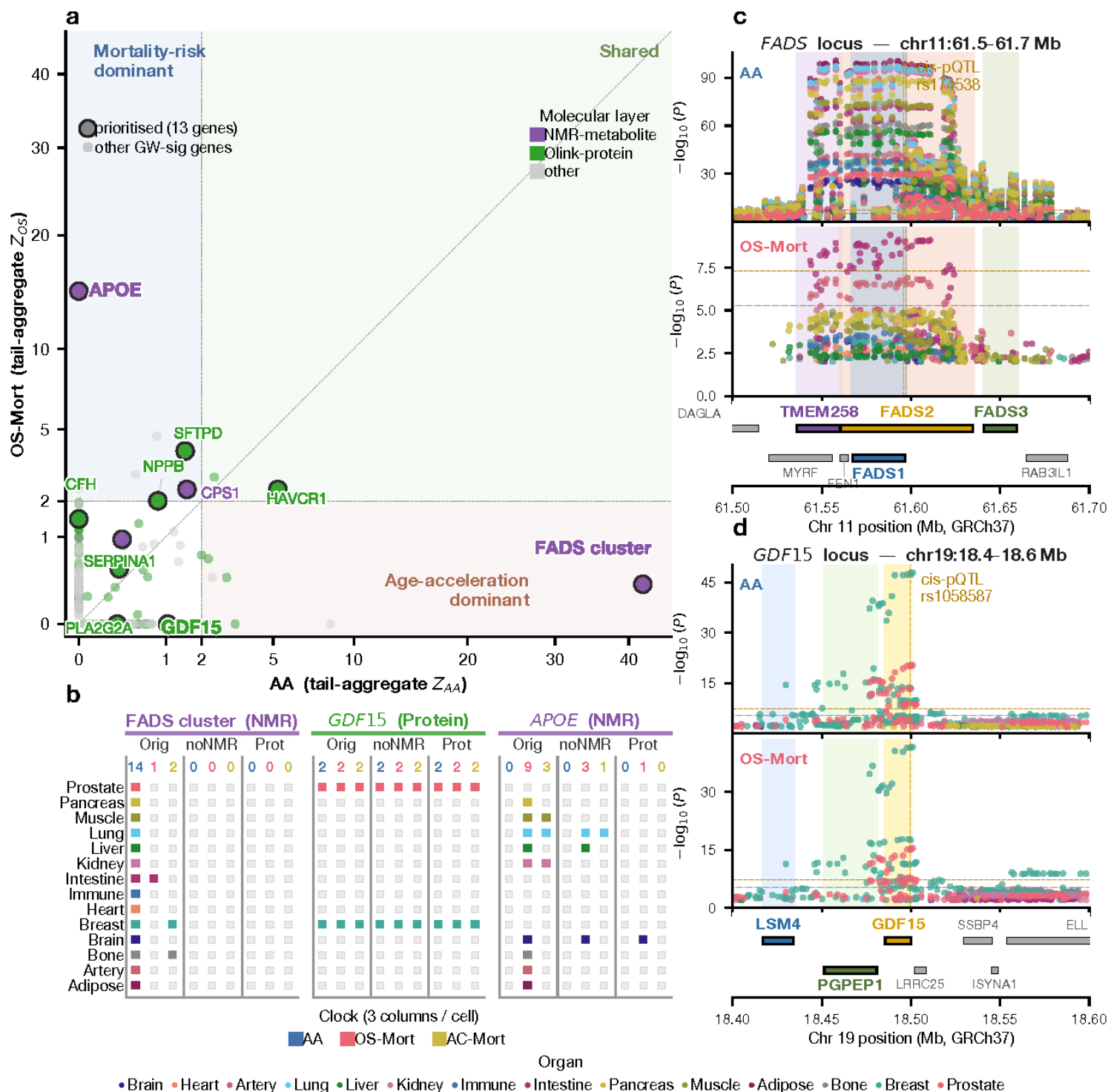


Figure 6

Two-axis decomposition of the prioritised gene set and the two-layer molecular architecture behind it.

(a) Per-gene cross-organ tail-aggregate z-statistic: age acceleration (x, ZAA) versus organorgan-specific mortality (y, ZCS), accumulated across 14 organs (Methods); both axes are square-root-scaled with raw z tick labels. Fill colour marks the primary molecular layer (purple, NMR metabolite; green, Olink protein; grey, other); the 13 prioritised candidates are drawn as large dark-edged markers and all other genome-wide-significant genes as small dots (per-gene QTL evidence tiers in Fig. 5); quadrant tints mark age-

acceleration-dominant, mortality-dominant and shared sets. *FADS1*/*FADS2* are collapsed to “*FADS*cluster”. (b) Three-gene × three-scope feature-ablation grid (rows *FADS* cluster / *GDF15* / *APOE*; columns original / non-NMR / protein-only); each cell is a 14-organ × 3-clock sub-matrix, a filled square marks genome-wide significance ($P < 5 \times 10^{-8}$) within ± 500 kb of the gene, with significant-organ counts annotated per clock. (c) *FADS* locus zoom (chr11:61.5–61.7 Mb): age-acceleration and organorgan-specific Manhattans coloured by organ, gene track below, literature cis-pQTL leads marked. (d) *GDF15* locus zoom (chr19:18.4–18.6 Mb), same layout.

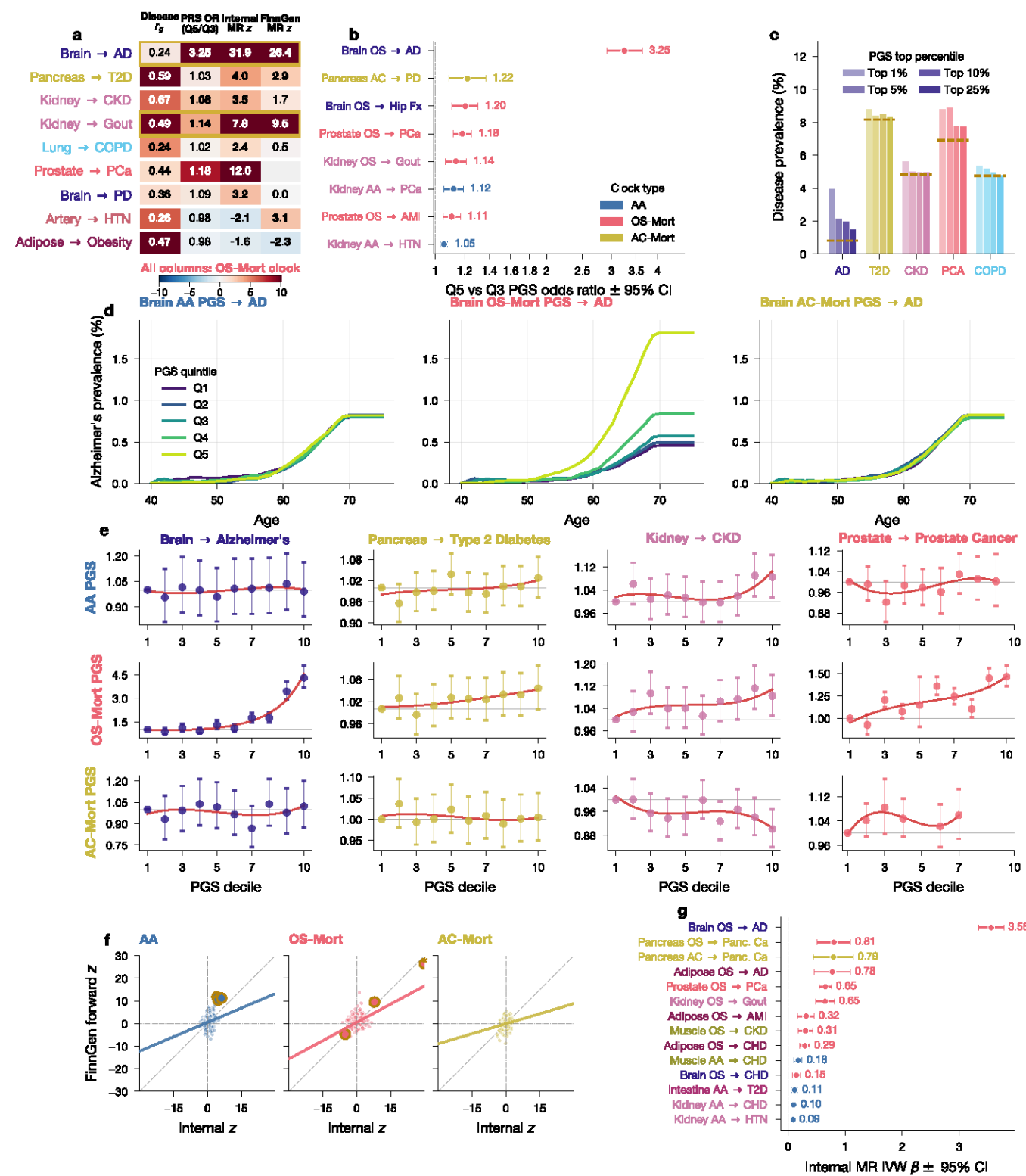


Figure 7

Genetic evidence links organ aging to disease risk. (a) Integrated evidence for the core organ-matched disease pairs (rows) across four independent lines (columns): disease genetic correlation, polygenic-score odds ratio (Q5/Q3), internal MR z and FinnGen forward MR z, all for the organ-specific clock; cell colour, signed $-\log_{10}P$; bold, $P < 0.05$; gold border, the three disease-prediction lines all significant and concordant (full matrices in Supplementary Data 5, 7 and 8). (b) All eight FDR-significant risk-increasing polygenic-score-disease associations (Q5 versus Q3 odds ratio $\pm$ 95% CI); dot colour, clock type; row-label colour, organ; the 13 FDR-significant inverse associations are listed in Supplementary Data 7. (c) Absolute disease prevalence in the top 1/5/10/25% of the organ-specific polygenic score for five organ-matched pairs (gold dashed line, population average; prostate cancer, males only). (d) Cumulative Alzheimer's incidence by brain polygenic-score quintile under each of the three training objectives. (e) Non-linear polygenic-score dose-response: three score types $\times$ four organ-matched diseases, logistic odds ratio per decile ($\pm$ 95% CI) with spline smoothers; per-s.d. odds ratios and model tests in Supplementary Data 7. (f) Cross-cohort MR replication: per-pair IVW z (internal UKB versus FinnGen forward), faceted by clock type; gold border, Bonferroni-significant on both sides. (g) All 14 Bonferroni-significant risk-increasing internal-MR associations (IVW $\beta \pm$ 95% CI; colour, clock type); the eight inverse associations are listed in Supplementary Data 8.

Supplementary Files

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

- [supplementary.zip](#)
- [ExtendedDataFig4.pdf](#)
- [ExtendedDataFig5.pdf](#)
- [ExtendedDataFig1.pdf](#)
- [ExtendedData.docx](#)
- [ExtendedDataFig3.pdf](#)
- [ExtendedDataFig2.pdf](#)