Data source and ethics: public mouse liver same-nucleus multiome resource GSE328206 (training) and external mouse datasets GSE327303, E-MTAB-12579, TMS 10x and TMS ss2, plus GSE302188 (batch–age crossed), GSE225576 (bulk liver) and GSE262290 (cross-run); no new animals were used. Animal cohort: 40 C57BL/6JN mice, five ages, eight per age. RNA and ATAC profiles were processed into donor-by-gene and donor-by-peak pseudobulk matrices; RNA and ATAC nuclei counts are reported after the stated QC and are not interchangeable across filtering versions. Clocks used variance-filtered features and outer leave-one-donor-out validation with inner tuning; the fast benchmark used a different feature budget. External transfer used gene-ID mapping and frozen or target-domain standardisation; no target labels were used for tuning in frozen mode. Batch identifiability used full-dummy rank, nullity and condition number, with non-redundant alias checks; within-batch and batch-specific slopes were fitted separately. Composition was analysed at the donor level with descriptive conditional regressions and variance partitioning, not causal mediation. Gene-level and gene-peak tests are separate families (defined below). Protein TMT channels are nested with age; ATAC and cross-species analyses are reported with their confounds. Spatial donors were non-overlapping with the bulk and scRNA cohorts (n=4 mice, four Stereo-seq sections), collected in an independent study and processed on a different platform (Stereo-seq bin200); no spatial section was used in clock training. All code and source tables are archived (see Code availability); analyses were run on processed matrices derived from the public raw data.

Gene-set definitions and mapping chain. The primary clock used the top 3,000 variance genes as ElasticNet input. Of these, 310 had non-zero coefficients (180 positive, 130 negative; Fig. 3f) and form the clock signature. The 300-gene gene-peak family is a separate set defined by the 300 smallest gene-peak permutation P values; it is distinct from the 310 clock genes. For the 300-gene family, 3 replicates x 1,000 permutations gave 3,000 nulls per gene (effective 3,001); the 10-gene escalation used 3 x 10,000 permutations, giving 30,000 nulls per gene (effective 30,001). The P value is (1 + #{null >= observed})/(1 + n_null). Of the 310 signature genes, 216 have a human symbol, 213 mapped to an Ensembl human gene ID, 206 had an MGI 1:1 human–mouse pair, 124 fell in the LINCS universe, and 122 (69 up, 53 down) passed the final orientation/membership filter. Per-gene mapping and loss tables are in the source data.

Pharmacological integration and null calibration

Data sources. The liver-ageing signature was the non-zero P0 hepatocyte-clock predictor (ElasticNet; 310 genes, 180 up, 130 down), with WGCNA modules and batch-half sign stability. The drug–target map had 15 compounds and 175 pairs (148 human genes); compound signatures were LINCS L1000 consensus Up/Down sets (5,425 compounds; 9,525-gene universe). GTEx liver reach used donor-level mean log2 TPM and linear age β (n=214). Clinical candidates came from a 24-row table using only tier, design and endpoint columns [45]. GSE280382 AgedLT liver formed the intervention challenge set.

Orthologue mapping and null calibration. Mouse genes were mapped to human symbols via MGI HOM_MouseHumanSequence [46]; all sets were compared in human space. Drug–

module overlap used an upper-tail hypergeometric test (M=3,853 human-mapped network symbols from 6,016 mouse genes) with BH across 315 pairs. Clock overlap used 10,000 size-matched draws from the same background, separately for up and down sets. Hub enrichment used the degree-matched null for the GTEx 426-gene signature.

LINCS reversal. Query was 69 clock-up and 53 clock-down symbols; strict Z = min(zA,zB) − max(zC,zD), with 1,000 query-set permutations (seed 20261112), BH q and a family-maximum FWER null. Geroprotector enrichment was the AUC of strict Z for 15 labelled compounds.

GTEx and intervention scoring. Liver expression was mean log2 TPM and liver age association the linear β with its P; strict Z ≥1.96 was an uncorrected screen and permutation q was reported separately. In GSE280382, only AgedLT liver samples were used (young n=8, aged n=9, aged+exenatide n=8; all batch 3). Counts were log1p(CPM), gene-wise z-scored, and scored with the frozen clock linear predictor standardised by training mean and scale. Module scores were the mean z of measured genes. Contrasts used Welch t-tests with Welch–Satterthwaite 95% CI and BH within contrast; cells were not replicates.

Constraint analyses

Human liver projection. GTEx liver bulk TPM (n=226 donors; age, sex, RIN, ischaemic time and centre as covariates) and HLiCA hepatocyte donor pseudobulk (48 health-tier-filtered donors) were projected with the 216 human symbols of the frozen signature. Expression was log2(TPM+1) for GTEx and log2(CPM+1) for HLiCA, gene-wise z-scored; the score was mean z of up genes minus mean z of down genes. GTEx models adjusted for age, sex, RIN, ischaemic time and centre; HLiCA for age, sex and study. Spearman correlations and 1,000 expression-matched permutations were reported; the analysis is rank-based.

Human failure modes. GTEx strata were sex, centre, RIN quartile and ischaemic-time quartile; HLiCA strata were study, sex and hepatocyte cell-number quartile (≥5 donors and ≥3 ages). Gene-level direction was conserved when the human age β matched the mouse clock coefficient in both datasets, partial in one, divergent in neither; module enrichment used Fisher exact tests.

Ortholog retention. Human–mouse orthologs came from MGI HOM_MouseHumanSequence (18,784 1:1 pairs). Sets were the 310-gene signature, the 2,690 zero-coefficient clock background, a housekeeping proxy and RPS/RPL symbols. Fractions carry Wilson 95% CIs; Fisher tests were BH-adjusted across 12 comparisons. Macaque 1:1 orthologs and protein identity came from Ensembl Compara. dN/dS was not computed (Compara returned null; transcript lookup too slow).

Reach map and detectability. Units were donor × cell type, never cells (345 primary units; hepatocyte restricted to subtype all). Human targets were mapped to the most highly expressed mouse ortholog. Hepatocyte expression required mean >1 logCPM and detection >50% of units; enrichment required ≥2× the highest non-hepatocyte mean.

Detectability used the frozen clock score in SD units of the 25-sample distribution; MDE was the exact noncentral-t two-sample value (n1=8, ratio 9/8, α=0.05, power 0.8). The null used 20,000 label permutations × 21 modules (420,000 contrasts; seed 20261112). GLP1R was summarised under whole-hepatocyte (46 units) and zone-resolved (138 units) policies; zone units per donor were never treated as independent.

Methods: SASP, functional enrichment and multi-omic validation

SASP gene set and scoring. The primary SASP set contained 96 curated genes (95 measured in hepatocyte pseudobulk), combining canonical senescence-associated secretory genes with the SenMayo core. The published SenMayo set (118 unique mouse genes) was used as an external redefinition. Scores were the mean z-score of detected genes within each cohort; cell-type scores were computed per cell and aggregated to donor x cell-type units. Age associations used donor-level Spearman correlation with BH correction; external bulk cohorts used batch- or run-partial Spearman with 10,000 age-within-batch permutations. GSE302188 (83 liver samples, five ages, three fully crossed batches), GSE262290 (52 samples, five ages, two runs), GSE327303 (six donors, 13 vs 78 weeks) and the independent cohorts GSE209823, GSE221122 and GSE181797 were analysed with the same scoring pipeline. M06–SASP shared-gene sensitivity removed the 12 genes present in both sets and recomputed the module and SASP scores in GSE302188 and GSE262290; Pearson, Spearman and age-adjusted partial correlations were compared with the full module.

M06 and cell-type deconvolution. WGCNA modules were defined previously; M06 contains 403 genes. Module eigengenes were the first PC of z-scored module genes (mean z as fallback). M06–SASP correlations used donor-level Spearman with Williams tests for dependent correlations. Cell-type fractions in bulk liver were estimated by NNLS on log-CPM profiles using an eight-cell-type reference from the 295,364-nucleus snRNA atlas; marker-score correlations and atlas M06 density were used as orthogonal checks. Composition-adjusted partial correlations controlled age, batch and all estimated fractions.

Functional enrichment and multi-omic layers. Enrichr REST was queried for GO Biological Process 2023, MSigDB Hallmark 2020, KEGG 2021 Human and Reactome 2022. Hypergeometric tests used two backgrounds (all 30,279 expressed genes and the 6,016 WGCNA-mapped genes); 10,000 expression-decile-matched permutations calibrated enrichment; split-half analyses compared genes in the top and bottom 50% of SASP correlation; GSEA used weighted Kolmogorov–Smirnov with 1,000 permutations. Proteome, lipidome, methylation, spatial (COMMOT/scDiffCom) and ATAC analyses are described in the source reports; all cross-layer comparisons are gene- or module-level because the cohorts are not sample-matched.

Methods: formal decision and identifiability analyses

Six-gate encoding. Gates g1–g6 were expression, liver age association, module overlap, LINCS reversal, clinical endpoint and safety; states were pass, partial, fail or unknown. Encodings were optimistic (1.00, 0.50, 0), neutral (0.80, 0.40, 0.05) and conservative (0.50, 0.25, 0.02); unknown took p0=0.28 (12/43). The joint score was a declared-probability

noisy-AND P=Πp_g (conditional independence given class); no likelihood or posterior was computed. Veto rules were none, any-fail, k=2, k=3 and a weighted product; sensitivity crossed encodings × rules and p0 ∈ {0.15, 0.45}. The counterfactual set gates 1–4 to pass or best observed state; ceiling=p5×p6. Reach was p_reach=1/(1+e^{−(log2TPM−1)/0.8}), with ≥0.80 reachable.

Binary gate tree and stress. All 64 pass/fail patterns were enumerated under the three encodings; the all-pass and one-fail contrasts quantify gate necessity; no resampling. Stress simulations used the measured 40-donor pseudobulk and frozen clock (top-3,000 variance genes, ElasticNet α=0.01, l1_ratio=0.5). Refit-holdout fitted n=4–30 donors (100 draws); the design grid crossed young-heavy/balanced/old-heavy compositions with n=4–30 (60 draws/cell); age imbalance used 200 draws. The positive control implanted β×SD×z(age) into 200 genes against 800 null genes. Nonlinear batch added a stationary RBF Gaussian-process field over age to 1,000 null genes under balanced or confounded splits. Platform noise was cohort donor residual SD divided by internal LODO residual SD.

Identifiability. X was intercept, five age indicators, two batch indicators and sex (n=40). Rank was computed by SVD; null vectors were age indicators summing to the intercept and batch_01 = ages 13+81+121.5, batch_02 = ages 53+105. Supports are disjoint, so a batch-adjusted age contrast is not estimable non-parametrically (rank 6/9, nullity 3). DoWhy/EconML cannot identify age net of batch without additional assumptions.

Spatial zonation and statistics. GSE328448 Stereo-seq bin200 sections were scored for the continuous zonation axis (Layer1, central vein; Layer9, portal vein) using published liver-zonation markers [6–8,12]. M06, myeloid-marker, hepatocyte and endothelial scores were mean log-normalized expression over measured genes. Within each section, I computed Spearman rho with the zonation score, 6-nearest-neighbour Moran's I and bivariate spatial lags; null distributions used 1,000 coordinate/label permutations and 500 value permutations. Donor-level tests used one-sample t-tests across four sections (n=4), not spot counts; M06 was residualized on myeloid score, nCount, nFeature, hepatocyte and endothelial scores where stated. Disease-zonation context and spatial-method scope are discussed in refs [39,47].

Drug-module completeness and power. All 15×21=315 pairs used the upper-tail hypergeometric null with BH. LINCS signature dependence used Kendall tau and Jaccard overlap of top 20/50/100/500 strict-Z sets. Synthetic effects 0.24–2.0 SD were added to aged-control scores (n=9; SD 0.343); 20,000 Monte Carlo Welch tests per design and noncentral-t power were computed. This is a design calculation on the score scale.

Peak-to-gene, TF, MOFA, cell-level. The precomputed motif–peak–gene table (1,642,253 edges) was summarized by cell type, contrast and module; concordance was the fraction of same-sign dAccess/dExpr edges. TF activity was the maximum absolute age effect (descriptive). MOFA2 used RNA/ATAC logCPM, top 2,000 variable features per view and eight factors. Cell-level donor-block variance used 179,655 hepatocytes; a MixedLM was

attempted but the donor random-effect variance hit the boundary, so the donor-block partition is reported.

Additional cohorts and zone models. Guilliams human liver h5ad was aggregated to donor pseudobulk and scored with the 216-symbol signature; random-effects meta used Fisher-z and DerSimonian–Laird. Open Targets/ChEMBL tractability was queried for 12 targets. scCODA [48] used donor×cell-type counts (hepatocyte reference; continuous age; NUTS 10,000/5,000); fixed-composition reconstruction used observed or fixed young (13-week) weights; real fgsea [49] used fgseaMultilevel with 1,000 permutations. External GSE302188/GSE262290 and brain GSE212336 scoring followed the same frozen-signature rules; cross-region brain transfer applied each region's fitted SelectKBest–StandardScaler–RidgeCV pipeline and masked the in-sample diagonal. ATAC peaks were processed with Signac [50].

Cross-dataset pathway direction. Full-transcriptome age rankings were computed as Spearman rho (GSE302188, n=83) or age effect (GSE262290, n=52) per gene, and fgseaMultilevel was run with the same mouse Hallmark sets (minSize 15, maxSize 500, eps=0, 1,000 permutations). For the 18 discovery-significant pathways, NES signs and Spearman NES were compared across cohorts. Clock-gene direction was the sign of the frozen clock coefficient versus external age rho; M14 genes used the discovery full-ranking rho. Binomial tests assessed concordance, and within-batch and within-run label permutations (1,000 draws, gene covariance preserved) gave P=0.002 (GSE302188) and P=0.008 (GSE262290); Mup11 was evaluated separately because it is not in the 310-gene signature.

Methods: donor hierarchy, trajectories and stability

Donor-level variance partition, trajectory comparison and stability selection. Cell-type pseudobulk units were scored with the frozen clock; donor variance share was estimated from a donor-block OLS model with age, sex and cell type. No random-effects model was fitted. Trajectories were compared with linear, quadratic, B-spline (df=3), plateau, segmented and sex-interaction models using AIC/BIC and leave-one-donor-out RMSE; five ages limit spline and segmented inference. Stability selection resampled donors or age strata 200 times, refit ElasticNet on the frozen 3,000-gene budget, and recorded selection frequency and sign stability for the 310 signature genes. Module concordance used the signature module assignments and human gene-level age effects.

Methods: donor-level zone, composition and programme audits

Software and reproducibility: analyses used Python 3.12 (numpy 2.1, pandas 3.0, scikit-learn 1.9, statsmodels 0.15, semopy, matplotlib 3.8), scCODA 0.1.9 with TensorFlow 2.19, and R 4.6.1 with fgsea. Random seeds were 20261112 (LINCS/permutations), 20261113 (fgsea/GSE262290) and 0 (brain CV). Processed tables and code are archived at Zenodo (10.5281/zenodo.24381654 and 10.5281/zenodo.24381798).