HALLMARKAGE: BINARIZED HALLMARK-AWARE TRANSCRIPTOMIC CLOCKS TO DISCOVER AGING MECHANISMS

Saleem A. Al Dajani[1,2,3*], **Elvira Kinzina**[2,3], **Juhyung Jung**[2,3], **Owen Mulledy**[2,3], **Gonzalo Cardenal-Antolin**[2,3], **Patrick T. Paine**[2,3], **Bohan Zhang**[2], **Dmitrii Glubokov**[2], **Alexander Tyshkovskiy**[2], **John R. Williams**[1], **Jesse R. Poganik**[2], **Vadim N. Gladyshev**[2], **Jonathan S. Gootenberg**[3], **Omar O. Abudayyeh**[2]

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology
[2]Brigham and Women's Hospital, Mass General Brigham, Harvard Medical School
[3]Beth Israel Deaconess Medical Center, Beth Israel Lahey Health, Harvard Medical School

ABSTRACT

Aging is frequently described through mechanistic "hallmarks," sets of pathways that differ across studies and outline separate aging mechanisms. Such classification may help interpret aging clock scores and generalize across cell-types, cohorts, and sequencing technologies. We introduce HallmarkAge, a framework that converts various hallmark gene sets into supervised binary hallmark state labels via score thresholding and trains probabilistic hallmark state classifiers from gene expression. We employ seven gene-set frameworks, producing a compact and interpretable hallmark representation that supports hallmark-space embedding, clustering, and hallmark-aware age prediction. We evaluate HallmarkAge on a single-cell and single-nucleus RNA-sequencing human skeletal muscle aging atlas across multiple cellular subsets. Hallmark-space reveals structured heterogeneity across aging states and clustering produces mechanistically readable programs with coherent marker genes. We train aging clocks using hallmark-derived features, selecting models under donor-level splits and aggregating predictions to donor-level estimates. Residual analysis highlights previously undescribed aging signals, nominating targets for perturbation screening and validation.

1 INTRODUCTION

Aging clocks built from molecular measurements can predict chronological age and correlate with morbidity and mortality risk. While epigenetic clocks, such as the foundational multi-tissue clock developed by Horvath, have demonstrated high accuracy using DNA methylation (Horvath, 2013), transcriptomic clocks offer a complementary approach that can capture dynamic, cell-state-specific and gene module-level changes (Tyshkovskiy et al., 2024; Ying et al., 2024). However, transcriptomic clocks are often difficult to interpret mechanistically and generalize poorly across cohorts. More recently, the advent of single-cell RNA-sequencing has enabled the development of cell-type-specific aging clocks, providing unprecedented resolution into the heterogeneity of aging (Buckley et al., 2023), yet features of these clocks can still be difficult to frame in the context of aging mechanisms. Large-scale single-cell atlases have also characterized tissue-specific aging signatures across multiple organs (Consortium, 2020; Gao et al., 2024). In parallel, aging biology has converged on frameworks such as the Hallmarks of Aging (López-Otín et al., 2013; 2023), the Open Genes database of human genes associated with aging and longevity (Rafikova et al., 2024), and the Hallmarks of Stem Cell Aging (Rando et al., 2025), which provide structured hypotheses about aging mechanisms. The hallmarks frameworks are useful heuristics rather than mechanistic definitions of aging: because many hallmarks are interdependent and context-dependent, their boundaries overlap, and the categories conflate drivers with responses and consequences (Gladyshev, 2012; 2013; Gladyshev et al., 2021).

A key gap is that hallmark frameworks are largely descriptive, whereas many aging clocks are high-dimensional regressors with limited mechanistic grounding. We propose HallmarkAge, which converts hallmark biology into supervised state variables: it scores hallmark gene sets per cell/sample, thresholds them into binary on/off labels, trains probabilistic hallmark state classifiers, and uses hallmark-derived representations as interpretable features for age prediction and mechanistic profiling. Both coefficients and residuals are informative: coefficients quantify hallmark contributions, while reproducible residual structure may reflect missing mechanisms, motivating a coefficient- and residual-driven loop for hallmark validation and discovery and nominating novel rejuvenation targets.

*Corresponding authors: sdajani@mit.edu, vgladyshev@rics.bwh.harvard.edu, jgootenb@bidmc.harvard.edu, oabudayyeh@bwh.harvard.edu

Contributions.

1. We introduce HallmarkAge, which converts hallmark gene sets into binary pseudo-labels via score thresholding and trains probabilistic hallmark state classifiers from gene expression.
2. We show that hallmark-derived representations support interpretable hallmark-space embeddings and clustering, yielding mechanistically readable human skeletal muscle aging programs.
3. We train and compare baseline (gene-PCA), hallmark-only, and hybrid Ridge regression clocks under donor-level splits, and report both per-cell predictions and aggregated per-donor estimates.

2 Datasets and Methods

We evaluate HallmarkAge on a human skeletal muscle single-cell and single-nucleus RNA-sequencing aging atlas assembled from multiple subsets within a unified donor cohort (Tab. 1). The dataset spans 10x Genomics single-cell and single-nucleus profiling of donor ages from 15 to 75. We focus on skeletal muscle because age-associated decline in muscle function is a major driver of frailty and systemic aging phenotypes (Perez et al., 2022), positioning muscle as a testbed for whether hallmark-derived states capture quantitative decline mechanisms rather than descriptive correlates (Schlessinger & Van Zant, 2001). Transcriptomic aging clocks for skeletal muscle have identified gene signatures associated with muscle aging (Ali et al., 2025), and multi-modal atlases have characterized the cellular landscape of muscle aging (Lai et al., 2024). Like MultiTIMER's interpretable structure (Jung et al., 2023) and the module-based design of tAge and MetroAge (Tyshkovskiy et al., 2024; Ying et al., 2024), HallmarkAge is designed to be modular and tissue-agnostic by anchoring representations to universal hallmark biology rather than tissue-specific gene signatures, so our framework can be readily applied across tissues and datasets while capturing tissue-specific aging dynamics through hallmark score distributions.

Table 1: Human skeletal muscle single-cell and single-nucleus 10x Genomics RNA-sequencing aging atlas (donor-level splits to avoid leakage; donors overlap across subsets). Ages set at bin midpoints.

Subset	Donors	Old	Young	Cells
Cells to nuclei mapping (skeletal muscle)	17	99,240	83,921	183,161
Myonuclei to myofiber mapping	17	47,839	39,683	87,522
Myonuclei (2024 release)	13	37,721	36,905	74,626
Fibroblasts and Schwann cells	12	10,672	9,939	20,611
Immune cells	12	14,337	5,984	20,321
Blood vessel endothelial cells and vascular smooth muscle cells	12	8,021	11,296	19,317
Muscle stem cells	12	6,465	11,063	17,528
Total (Kedlian et al., 2024)	17	224,295	198,791	423,086

HallmarkAge converts hallmark gene sets into (i) binary hallmark states and (ii) hallmark-aware clock components. This enables mechanistic age prediction and residual-driven analysis for candidate missing programs. We used 7 hallmark gene-set frameworks (Appendix A and L), comprising 86 hallmarks, with per-hallmark gene lists ranging from 1–444 genes (median 12), and 43/86 hallmarks had at least 3 genes detected and were used for modeling. For each hallmark H with gene set G_H, we compute a per-cell hallmark score $S_H(x) \in \mathbb{R}$ from expression x using standard gene-set scoring, and collect scores into $S(x) = [S_{H_1}(x), \ldots, S_{H_K}(x)] \in \mathbb{R}^K$. To obtain supervision without curated perturbation labels, we define pseudo-labels $y_H(x) \in \{0, 1\}$ by thresholding scores: $y_H(x) = 1$ if $S_H(x) > Q_{0.8}(S_H)$ and $y_H(x) = 0$ if $S_H(x) < Q_{0.2}(S_H)$, excluding intermediates. For each hallmark H, we train a probabilistic classifier $f_H(x) = \Pr(y_H = 1 \mid x)$, yielding a mechanism vector $f(x) = [f_{H_1}(x), \ldots, f_{H_K}(x)] \in [0, 1]^K$. We train and evaluate models using two feature spaces: gene-PCA features, where highly variable genes are selected and PCA is computed to obtain $\mathrm{PCA}(x) \in \mathbb{R}^d$, and hallmark features, where hallmark scores $S(x)$ are used directly. We then store hallmark activation probabilities $f(x)$ as a hallmark-level representation for analysis. A schematic overview of the full pipeline is provided in Fig. S1, and the hallmark gene-set frameworks used are summarized in Tab. S1 and Tab. S6-S12.

3 Results

Hallmark-space embeddings provide a compact view of heterogeneous aging states across cell types and datasets. Cells are represented by their hallmark score vectors $S(x)$ and embedded using UMAP,

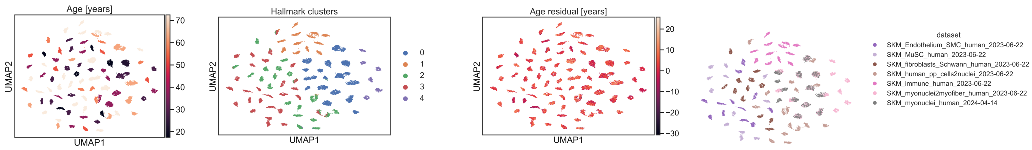

Figure 1: Hallmark-space UMAP visualizations. Hallmark-space embeddings provide a compact, interpretable representation of heterogeneous aging modes.

colored by age, cluster identity, and dataset of origin (Fig. 1). The embedding reveals structured, continuous gradients rather than a single aging axis, with multiple coexisting hallmark programs active within the same tissue. This motivates downstream clustering and hallmark-level analysis. Additional UMAP visualizations across multiple hallmark frameworks are provided in Fig. S3-S8.

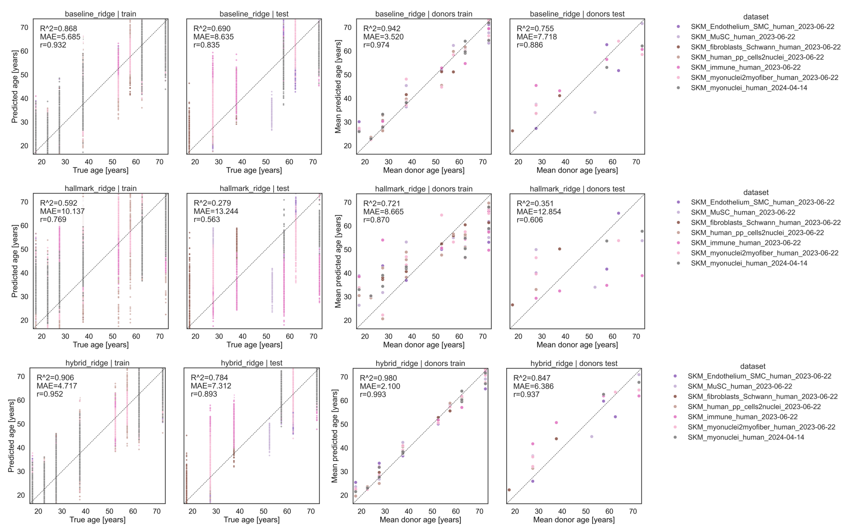

Figure 2: Age prediction performance across feature spaces. We select the best-performing clock by held-out MAE and report both per-metacell and per-donor performance.

We next evaluated whether hallmark-derived representations support age prediction. Ridge regression models were trained using gene-PCA features (baseline), hallmark features, or a hybrid concatenation, with train/test splits performed at the donor level to prevent leakage. Figure 2 shows predicted versus chronological age at both the cell and donor-aggregated levels. Hybrid models achieve the lowest held-out MAE, while hallmark-only models retain predictive signal with substantially reduced dimensionality. We infer chronological age a from gene-PCA features $X_{\text{gene}} = \mathrm{PCA}(x)$, hallmark features $X_{\text{hall}} = H(x)$, or a hybrid concatenation $X_{\text{hyb}} = [X_{\text{gene}}; X_{\text{hall}}]$, using Ridge regression $\hat{a}(x) = g_{\text{Ridge}}(X)$. We split train/test by donor, evaluate baseline, hallmark-only, and hybrid Ridge models using MAE and R^2 on held-out donors (Fig. 2), select the best model by MAE, refit it on all cells, and store predictions $\hat{a}_{\text{cell}}(x)$ and residuals $r(x) = a - \hat{a}_{\text{cell}}(x)$. For donor d with cells C_d, we compute $\hat{a}_{\text{donor}}(d) = \frac{1}{|C_d|}\sum_{x \in C_d} \hat{a}_{\text{cell}}(x)$, reducing within-donor variance (Fig. 2). Under donor-level evaluation, the HallmarkAge hybrid clock achieves a test MAE of 6.39 years and $R^2 = 0.85$ on held-out donors (Appendix K), while retaining mechanistic interpretability in a low-dimensional hallmark space. This outperforms SkeletAge (Ali et al., 2025), a bulk RNA-seq skeletal muscle clock with a validation MAE 6.62 years and R^2 of 0.79. Full in-distribution and out-of-distribution results are reported in Appendices K.1 and K.2. We further evaluate generalization across multiple tissues and cell types by combining skeletal muscle atlases with Tabula Sapiens, spanning 1.4M cells across 28 organs/tissue and 63 donors (Appendix Table S13). The hybrid model retains strong performance (r_{test} of 0.8 and R^2_{test} of 0.62) across this heterogeneous setting (Fig. S16).

Figure 3 summarizes hallmark activity across clusters identified in hallmark space. Clusters exhibit mixed hallmark signatures rather than single-pathway dominance, enabling mechanistic labeling of aging programs used for residual stratification. Clustering in hallmark feature space yields coherent and interpretable programs. For each cluster, we summarize the mean residual, residual class, top hallmarks, aging mechanism, and representative marker genes (Tab. 2).

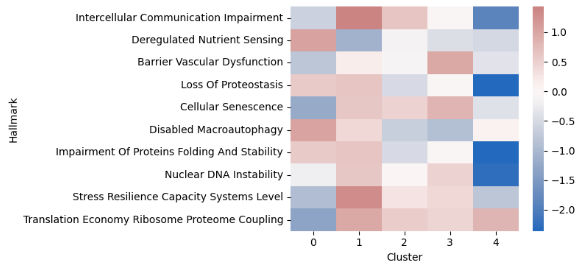

Figure 3: Hallmark activity across clusters. Clusters correspond mixed hallmark aging modes.

Table 2: Residual-stratified hallmark clusters in hallmark-space embedding for the human skeletal muscle aging atlas (Kedlian et al., 2024). Positive mean age residuals indicate candidate protective (anti-aging) programs, while negative residuals denote actionable pro-aging states.

Cluster	Color	Mean residual	Residual class	Top hallmarks	Mechanism label	Representative marker genes
0	Blue	+0.070	Protective (anti-aging)	Deregulated nutrient sensing; disabled macroautophagy; histone modification changes; transcriptional alterations; chromatin remodeling; stem cell exhaustion	Epigenetic–metabolic rigidity with partial stress compensation	ZBTB20; FOXO1; HDAC9; PPARGC1A; RBFOX1; TTN; CAMK2D; ERBB4; JAK2
1	Orange	−0.021	Actionable (pro-aging)	Sterile inflammation; altered intercellular communication; proteostasis impairment; inflammaging; transcriptional alterations	Innate immune inflammaging / stress response	IL1B; IL10; NLRP3; TYROBP; CXCL8; LYZ; CSF3R; NFKBIA; TNF
2	Green	−0.252	Actionable (pro-aging)	Extracellular matrix remodeling; fibrosis; IGF/INS dysregulation; oxidative stress; senescence/SASP	Fibrotic ECM remodeling and stromal aging	COL1A1; COL3A1; DCN; FBN1; CXCL12; IGFBP3; MMP2; CTGF; LOXL1
3	Red	+0.171	Protective (anti-aging)	Stem cell self-renewal; quiescence depth; transcriptional plasticity; stress resilience; ECM signaling	Adaptive progenitor plasticity / stress-resilient renewal	JUN; JUNB; KLF2; CDKN1A; IGFBP5; SPARCL1; CLDN5; ATF3
4	Purple	−0.193	Actionable (pro-aging)	Translation economy dysregulation; macroautophagy suppression; proteostasis decline; mitochondrial impairment	Energetic–proteostatic collapse program	CKM; COX7A1; MYH7; ATP2A1; DES; CRYAB; FBXO32; PDK4

4 Discussion

Hallmark-space structure captures heterogeneous aging modes. Projecting cells into hallmark feature space yields a compact, interpretable representation in which age-associated structure and distinct cellular states are visible without requiring gene-level inspection. Hallmark-space UMAPs organize the combined dataset into coherent regions of hallmark activity (Fig. 1), supporting the view that aging hallmarks are not well-described by single axes but by mixtures of hallmark programs. Hallmark cluster occupancy vary with different populations expressing overlapping but non-identical hallmark mixtures (Appendix E). Extended visualizations across alternative hallmark definitions further support this structure (Fig. S3-S8). Cluster markers are used to prioritize *in silico* perturbations that would shift HallmarkAge to validate rejuvenation targets in follow-up screens (Appendix I). These perturbation directions can be mapped to known anti-aging intervention strategies targeting specific hallmarks (Fig. S15).

Hallmark-derived clocks enable interpretable age prediction. While recent work has combined machine learning with explainable AI methods to interpret biological age predictions post-hoc (Qiu et al., 2023), HallmarkAge achieves interpretability by construction, anchoring representations directly to established hallmark biology rather than deriving explanations from black-box models. We trained ridge regression clocks using gene-PCA features (baseline), hallmark-derived features, and a hybrid concatenation of both. Under donor-level train/test splits, hybrid models achieved the lowest held-out MAE (Fig. 2), while hallmark-only models retained predictive signal with substantially reduced dimensionality. Because hallmark-derived features correspond to predefined biological programs, learned coefficient structure is directly interpretable (Appendix J) and residual patterns can be localized to hallmark-space programs (Fig. 1; Tab. 2). This behavior remains consistent when evaluated across diverse tissues and cell types (Fig. S16; Tab. S13).

Data-driven hallmark clusters are mechanistically readable. Clustering in hallmark feature space yields mechanistically interpretable programs composed of coherent mixtures of aging hallmarks rather than single-pathway states (Tab. 2). As shown in Figures 1 and 3, hallmarks related to chromatin remodeling, transcriptional plasticity, nutrient sensing, and stress resilience lie close in hallmark space and co-occur in protective clusters (0, 3), whereas inflammatory signaling, fibrotic ECM remodeling, proteostasis collapse, and mitochondrial impairment form distinct but biologically consistent pro-aging programs (clusters 1, 2, 4). These programs occupy distinct but partially overlapping regions of hallmark space, consistent with aging as a continuous, multi-dimensional process rather than a linear trajectory (Fig. 1). Not all canonical hallmarks are equally engaged in skeletal muscle: stem cell exhaustion, regenerative decline, and altered intercellular communication receive near-zero coefficients in hallmark-based and hybrid clocks (Appendix J), consistent with muscle being a largely post-mitotic tissue where aging is dominated by metabolic, proteostatic, inflammatory, and stromal programs rather than active stem cell turnover.

Linking hallmark states to proteins via transcriptome–proteome coupling. To connect hallmark states to molecular outputs beyond transcriptomics, we map genes (Moqri et al., 2026) to proteins (Karra et al., 2025; Amon et al., 2019) to quantify RNA–protein coupling (Triana et al., 2021; Furtwängler et al., 2024; Bandyopadhyay et al., 2024). In kidney aging data, transcriptome coverage is near-complete (22,254 mRNA-measured symbols; 99.7% of the union), whereas proteome coverage is lower (6,515 protein-measured symbols; 29.2%), with overlap for only 6,451 genes (28.9%) (Takemon et al., 2021). This linkage supports mechanistic interpretation of age- and mortality-associated global protein increases (Goeminne et al., 2025) by stratifying genes into RNA/protein response quadrants (Takemon et al., 2021), separating coupled upregulation from protein accumulation and proteostasis impairment, and diagnosing transcript–protein decoupling (Ruoss et al., 2024; Komic et al., 2025) and age-dependent functional decline (Hennrich et al., 2018). Complementary methylation-based skeletal muscle clocks achieve high accuracy (Voisin et al., 2020), motivating multi-omic integration of transcriptomic and epigenetic aging signatures for improved prediction and mechanistic insight.

Relationship to existing single-cell aging clocks. Several approaches have been developed to estimate biological age at single-cell resolution. SCALE provides tissue-specific aging measures using gene expression profiles (Mao et al., 2023), while CellBiAge uses binarization strategies similar to our approach for age group classification (Yu et al., 2023). Unlike these methods, HallmarkAge explicitly anchors its representations to established hallmark biology, enabling direct mechanistic interpretation of aging states rather than purely data-driven age estimation.

Limitations and confounding considerations. While hallmark-derived representations improve interpretability, hallmark scores and inferred structure can reflect confounding factors such as batch effects, technology differences, or shifts in cell-state composition. We partially mitigate evaluation by splitting train/test sets by donor and by evaluating on a multi-cell-type combined dataset. There is currently no experimental ground truth for rejuvenation or pro-aging effects in this setting; thus, any future perturbation analyses would be evaluated only relative to unperturbed baselines, interpreted as model-dependent hypotheses, and would require experimental validation to establish biological causality. *In silico* perturbation analyses are constrained by current virtual cell models, which support only single-gene perturbations and cannot represent combinatorial or non-linear interactions, missing coordinated, program-level effects. Hallmark states cannot yet be validated via transcriptome–proteome coupling as the data analyzed here lack simultaneous, cell-matched transcriptomic and proteomic measurements; gene-to-protein mapping remains a translational hypothesis.

5 Conclusion

We introduced HallmarkAge, a hallmark-aware framework for interpretable aging clock engineering in single-cell transcriptomic data. HallmarkAge converts hallmark gene sets into binary pseudo-labels and trains probabilistic hallmark state classifiers, producing a compact mechanism representation that enables hallmark-space visualization and clustering (Fig. 1; Table 2). On an aging atlas of human skeletal muscle single-cell and single-nucleus combined dataset (Tab. 1), HallmarkAge supports age prediction using hallmark-derived features and hybrid gene–hallmark feature sets, with model selection under donor-level splits and stable donor-level aggregation (Fig. 2). HallmarkAge provides a practical bridge between hallmark biology and predictive modeling, improving interpretability while remaining compatible with single-cell workflows. Residual analysis nominates rejuvenation targets for perturbation screening and validation (Tab. 2), linking machine learning representations to wet-lab experiments for target discovery.

Author Contributions

S.A.A., J.S.G., and O.O.A. conceived and designed the study. S.A.A. curated and consolidated the datasets, performed the primary computational analysis, and drafted the manuscript. G.C.A. provided parts of code and compared results with parallel pipelines. P.T.P. and B.Z. contributed to the methodology regarding hallmark gene set selection. D.G. shared a parallel pipeline code for comparison and A.T. contributed to data interpretation and code optimization. E.K., J.J., and O.M. participated in technical and conceptual discussions. J.R.W., J.R.P., and V.N.G. provided critical advising and supervision. J.S.G. and O.O.A. provided overall project supervision. The manuscript was revised and edited by S.A.A., V.N.G., J.R.P., J.S.G., and O.O.A.

Acknowledgments

This work was supported by grants from the NIA and the Hevolution Foundation. S.A.A. was funded by the MIT Department of Civil and Environmental Engineering (CEE) under the Friesecke (1961) Fellowship Fund. J.S.G. and O.O.A. are supported by NIH grants R01-EB031957, R01-AG074932, and R01-GM148745; G. Harold & Leila Y. Mathers Charitable Foundation; Rett Syndrome Research Trust; The Gordon and Betty Moore Foundation; Impetus Grants; Cystic Fibrosis Foundation Pioneer Grant; Google Deepmind; Sanofi; Yosemite; Michelson Foundation; Hevolution Foundation; American Federation for Aging Research; Pivotal Life Sciences; and the MGB Gene and Cell Therapy Institute.

References

Muhammad Ali, Fei Li, and Manpreet Katari. SkeletAge: Transcriptomics-based Aging Clock Identifies 26 New Targets in Skeletal Muscle Aging. *bioRxiv*, 2025.

Sabine Amon, Fabienne Meier-Abt, Ludovic C Gillet, Slavica Dimitrieva, Alexandre PA Theocharides, Markus G Manz, and Ruedi Aebersold. Sensitive quantitative proteomics of human hematopoietic stem and progenitor cells by data-independent acquisition mass spectrometry. *Molecular & Cellular Proteomics*, 18(7):1454–1467, 2019.

Shovik Bandyopadhyay, Michael P Duffy, Kyung Jin Ahn, Jonathan H Sussman, Minxing Pang, David Smith, Gwendolyn Duncan, Iris Zhang, Jeffrey Huang, Yulieh Lin, et al. Mapping the cellular biogeography of human bone marrow niches using single-cell transcriptomics and proteomic imaging. *Cell*, 187(12):3120–3140, 2024.

Marianne Böni-Schnetzler, Hélène Méreau, Leila Rachid, Sophia J Wiedemann, Friederike Schulze, Kelly Trimigliozzi, Daniel T Meier, and Marc Y Donath. IL-1beta promotes the age-associated decline of beta cell function. *iScience*, 24(11), 2021.

Matthew T Buckley, Eric D Sun, Benson M George, Ling Liu, Nicholas Schaum, Lucy Xu, Jaime M Reyes, Margaret A Goodell, Irving L Weissman, Tony Wyss-Coray, et al. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nature Aging*, 3(1): 121–137, 2023.

The Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590–595, 2020.

Benjamin Furtwängler, Nil Üresin, Sabrina Richter, Mikkel Bruhn Schuster, Despoina Barmpouri, Henrietta Holze, Anne Wenzel, Kirsten Grønbæk, Kim Theilgaard-Mönch, Fabian J Theis, et al. Mapping the human hematopoietic stem and progenitor cell hierarchy through integrated single-cell proteomics and transcriptomics. *bioRxiv*, pp. 2024–07, 2024.

Shihong Max Gao, Yanyan Qi, Qinghao Zhang, Youchen Guan, Yi-Tang Lee, Lang Ding, Lihua Wang, Aaron S Mohammed, Hongjie Li, Yusi Fu, et al. Aging atlas reveals cell-type-specific effects of pro-longevity strategies. *Nature Aging*, 4(7):998–1013, 2024.

Vadim N Gladyshev. On the cause of aging and control of lifespan: Heterogeneity leads to inevitable damage accumulation, causing aging; control of damage composition and rate of accumulation define lifespan. *Bioessays*, 34(11):925–929, 2012.

Vadim N Gladyshev. The origin of aging: Imperfectness-driven non-random damage defines the aging process and control of lifespan. *Trends in Genetics*, 29(9):506–512, 2013.

Vadim N Gladyshev, Stephen B Kritchevsky, Steven G Clarke, Ana Maria Cuervo, Oliver Fiehn, João Pedro de Magalhães, Theresa Mau, Michal Maes, Robert L Moritz, Laura J Niedernhofer, et al. Molecular damage in aging. *Nature Aging*, 1(12):1096–1106, 2021.

Ludger JE Goeminne, Anastasiya Vladimirova, Alec Eames, Alexander Tyshkovskiy, M Austin Argentieri, Kejun Ying, Mahdi Moqri, and Vadim N Gladyshev. Plasma protein-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems. *Cell Metabolism*, 37(1):205–222, 2025.

Marco L Hennrich, Natalie Romanov, Patrick Horn, Samira Jaeger, Volker Eckstein, Violetta Steeples, Fei Ye, Ximing Ding, Laura Poisa-Beiro, Mang Ching Lai, et al. Cell-specific proteome analyses of human bone marrow reveal molecular features of age-dependent functional decline. *Nature Communications*, 9(1):4004, 2018.

Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10):3156, 2013.

Hongli Jiao, Brian E Walczak, Ming-Song Lee, Madeleine E Lemieux, and Wan-Ju Li. GATA6 regulates aging of human mesenchymal stem/stromal cells. *Stem Cells*, 39(1):62–77, 2021.

Chen Jin, Xizhe Wang, Jiping Yang, Seungsoo Kim, Adam D Hudgins, Amir Gamliel, Mingzhuo Pei, Daniela Contreras, Melody Devos, Qinghua Guo, et al. Molecular and genetic insights into human ovarian aging from single-nuclei multi-omics analyses. *Nature Aging*, 5(2):275–290, 2025.

Sascha Jung, Javier Arcos Hodar, and Antonio Del Sol. Measuring biological age using a functionally interpretable multi-tissue RNA clock. *Aging Cell*, 22(5):e13799, 2023.

Emin Onur Karakaslar, Neerja Katiyar, Muneer Hasham, Ahrim Youn, Siddhartha Sharma, Cheng-han Chung, Radu Marches, Ron Korstanje, Jacques Banchereau, and Duygu Ucar. Transcriptional activation of Jun and Fos members of the AP-1 complex is a conserved signature of immune aging that contributes to inflammaging. *Aging Cell*, 22(4):e13792, 2023.

Laila Karra, Anna-Marie Finger, Lauren Shechtman, Milana Krush, Rita Meng-Yao Huang, Morgan Prinz, Iliana Tennvooren, Kriti Bahl, Lisiena Hysienaj, Paulina G Gonzalez, et al. Single-cell proteomics characterization of bone marrow hematopoiesis with distinct Ras pathway lesions. *Blood Advances*, 9(19):4996–5008, 2025.

Veronika R Kedlian, Yaning Wang, Tianliang Liu, Xiaoping Chen, Liam Bolt, Catherine Tudor, Zhuojian Shen, Eirini S Fasouli, Elena Prigmore, Vitalii Kleshchevnikov, et al. Human skeletal muscle aging atlas. *Nature Aging*, 4(5):727–744, 2024.

Hana Komic, Tessa Schmachtel, Catia Simoes, Marius Külp, Weijia Yu, Adrien Jolly, Malin S Nilsson, Carmen Gonzalez, Felipe Prosper, Halvard Bonig, et al. Continuous map of early hematopoietic stem cell differentiation across human lifetime. *Nature Communications*, 16(1):2287, 2025.

Yiwei Lai, Ignacio Ramírez-Pardo, Joan Isern, Juan An, Eusebio Perdiguero, Antonio L Serrano, Jinxiu Li, Esther García-Domínguez, Jessica Segalés, Pengcheng Guo, et al. Multimodal cell atlas of the ageing human skeletal muscle. *Nature*, 629(8010):154–164, 2024.

José Alberto López-Domínguez, Sandra Rodríguez-López, Ulises Ahumada-Castro, Pierre-Yves Desprez, Maria Konovalenko, Remi-Martin Laberge, César Cárdenas, José Manuel Villalba, and Judith Campisi. Cdkn1a transcript variant 2 is a marker of aging and cellular senescence. *Aging (Albany NY)*, 13(10):13380, 2021.

Carlos López-Otín, Maria A Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–1217, 2013.

Carlos López-Otín, Maria A Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. Hallmarks of aging: An expanding universe. *Cell*, 186(2):243–278, 2023.

Shulin Mao, Jiayu Su, Longteng Wang, Xiaochen Bo, Cheng Li, and Hebing Chen. A transcriptome-based single-cell biological age model and resource for tissue-specific aging measures. *Genome Research*, 33(8):1381–1394, 2023.

Fabiola Marín-Aguilar, Ana V Lechuga-Vieco, Elísabet Alcocer-Gómez, Beatriz Castejón-Vega, Javier Lucas, Carlos Garrido, Alejandro Peralta-Garcia, Antonio J Pérez-Pulido, Alfonso Varela-López, José L Quiles, et al. NLRP3 inflammasome suppression improves longevity and prevents cardiac aging in male mice. *Aging Cell*, 19(1):e13050, 2020.

Mahdi Moqri, Kejun Ying, Jesse R Poganik, Chiara Herzog, Qingwen Chen, Mehrnoosh Emamifar, Alexander Tyshkovskiy, Alec Eames, Jure Mur, Dmitrii Glubokov, et al. Integrative epigenetics and transcriptomics identify aging genes in human blood. *Nature Communications*, 2026.

Matthew D Park, Jessica Le Berichel, Pauline Hamon, C Matthias Wilk, Meriem Belabed, Nader Yatim, Alexis Saffon, Jesse Boumelha, Chiara Falcomatà, Alexander Tepper, et al. Hematopoietic aging promotes cancer by fueling IL-1α–driven emergency myelopoiesis. *Science*, 386(6720): eadn0327, 2024.

Kevin Perez, Serban Ciotlos, Julia McGirr, Chandani Limbad, Ryosuke Doi, Joshua P Nederveen, Mats I Nilsson, Daniel A Winer, William Evans, Mark Tarnopolsky, et al. Single nuclei profiling identifies cell specific markers of skeletal muscle aging, frailty, and senescence. *Aging (Albany NY)*, 14(23):9393, 2022.

Wei Qiu, Hugh Chen, Matt Kaeberlein, and Su-In Lee. ExplaiNAble BioLogical Age (ENABL Age): an artificial intelligence framework for interpretable biological age. *The Lancet Healthy longevity*, 4(12):e711–e723, 2023.

Ekaterina Rafikova, Nikolay Nemirovich-Danchenko, Anna Ogmen, Anna Parfenenkova, Anastasiia Velikanova, Stanislav Tikhonov, Leonid Peshkin, Konstantin Rafikov, Olga Spiridonova, Yulia Belova, et al. Open Genes—a new comprehensive database of human genes associated with aging and longevity. *Nucleic Acids Research*, 52(D1):D950–D962, 2024.

Thomas A Rando, Anne Brunet, and Margaret A Goodell. Hallmarks of stem cell aging. *Cell Stem Cell*, 2025.

Severin Ruoss, Chanond A Nasamran, Scott T Ball, Jeffrey L Chen, Kenneth N Halter, Kelly A Bruno, Thomas C Whisenant, Jesal N Parekh, Shanelle N Dorn, Mary C Esparza, et al. Comparative single-cell transcriptional and proteomic atlas of clinical-grade injectable mesenchymal source tissues. *Science Advances*, 10(28):eadn2831, 2024.

David Schlessinger and Gary Van Zant. Does functional depletion of stem cells drive aging? *Mechanisms of Ageing and Development*, 122(14):1537–1553, 2001.

Yuka Takemon, Joel M Chick, Isabela Gerdes Gyuricza, Daniel A Skelly, Olivier Devuyst, Steven P Gygi, Gary A Churchill, and Ron Korstanje. Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. *eLife*, 10:e62585, 2021.

Rumiana Tenchov, Janet M Sasso, Xinmei Wang, and Qiongqiong Angela Zhou. Antiaging Strategies and Remedies: A Landscape of Research Progress and Promise. *ACS Chemical Neuroscience*, 15 (3):408–446, 2024.

Sergio Triana, Dominik Vonficht, Lea Jopp-Saile, Simon Raffel, Raphael Lutz, Daniel Leonce, Magdalena Antes, Pablo Hernández-Malmierca, Diana Ordoñez-Rueda, Beáta Ramasz, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nature Immunology*, 22(12):1577–1589, 2021.

Alexander Tyshkovskiy, Daria Kholdina, Kejun Ying, Maria Davitadze, Adrian Molière, Yoshiyasu Tongu, Tomoko Kasahara, Leonid M Kats, Anastasiya Vladimirova, Alibek Moldakozhayev, et al. Transcriptomic hallmarks of mortality reveal universal and specific mechanisms of aging, chronic disease, and rejuvenation. *BioRxiv*, pp. 2024–07, 2024.

Stef F Verlinden. The genetic advantage of healthy centenarians: unraveling the central role of NLRP3 in exceptional healthspan. *Frontiers in Aging*, 5:1452453, 2024.

Sarah Voisin, Nicholas R Harvey, Larisa M Haupt, Lyn R Griffiths, Kevin J Ashton, Vernon G Coffey, Thomas M Doering, Jamie-Lee M Thompson, Christian Benedict, Jonathan Cedernaes, et al. An epigenetic clock for human skeletal muscle. *Journal of Cachexia, Sarcopenia and Muscle*, 11(4): 887–898, 2020.

Anissa A Widjaja, Wei-Wen Lim, Sivakumar Viswanathan, Sonia Chothani, Ben Corden, Cibi Mary Dasan, Joyce Wei Ting Goh, Radiance Lim, Brijesh K Singh, Jessie Tan, et al. Inhibition of IL-11 signalling extends mammalian healthspan and lifespan. *Nature*, 632(8023):157–165, 2024.

Meng Wu and Shixuan Wang. FOXP1 is a gatekeeper of cellular senescence with ovarian aging, 2024.

Meng Wu, Weicheng Tang, Ying Chen, Liru Xue, Jun Dai, Yan Li, Xiaoran Zhu, Chuqing Wu, Jiaqiang Xiong, Jinjin Zhang, et al. Spatiotemporal transcriptomic changes of human ovarian aging and the regulatory role of FOXP1. *Nature Aging*, 4(4):527–545, 2024.

Kejun Ying, Alexander Tyshkovskiy, Qingwen Chen, Eric Latorre-Crespo, Bohan Zhang, Hanna Liu, Benyamin Matei-Dediu, Jesse R Poganik, Mahdi Moqri, Kristina Kirschne, et al. High-dimensional Ageome Representations of Biological Aging across Functional Modules. *bioRxiv*, 2024.

Doudou Yu, Manlin Li, Guanjie Linghu, Yihuan Hu, Kaitlyn H Hajdarovic, An Wang, Ritambhara Singh, and Ashley E Webb. CellBiAge: Improved single-cell age classification using data binarization. *Cell Reports*, 42(12), 2023.

A Hallmark Gene Sets Used for Classifier Construction

Table S1 summarizes the hallmark gene-set frameworks and derived feature spaces used in this work to construct binary hallmark classifiers and downstream age models.

Table S1: Hallmark gene sets used in this work for binary hallmark classifiers and age modeling. For the explicit gene lists used in scoring and classifier construction (by gene set), see Appendix L.

Framework / feature set	Contents (used as classifier targets or features)	Reference
Hallmarks of Aging (2013)	9 hallmark categories (canonical 2013 hallmarks) used to define gene sets and hallmark ON/OFF states.	López-Otín et al. (2013)
Hallmarks of Aging (2023)	12 hallmark categories (expanded 2023 hallmarks including disabled macroautophagy, chronic inflammation, and dysbiosis).	López-Otín et al. (2023)
Hallmarks of Stem Cell Aging (2025)	5 stem-cell-specific hallmarks (depth of quiescence, self-renewal propensity, fate of progeny, resilience, population heterogeneity).	Rando et al. (2025)
OpenGenes (mechanism-level)	Curated mapping of human genes to aging mechanisms (22 mechanisms; 2405 genes) used to expand and validate hallmark modules.	Rafikova et al. (2024)
OpenGenes summary (collapsed)	Collapsed OpenGenes mechanisms into higher-level axes (e.g., DNA damage/repair, epigenetic/transcriptional regulation, proteostasis/autophagy, inflammation/communication) for compact modeling.	Derived from OpenGenes Rafikova et al. (2024)
Overlap (intersection sets)	Gene sets defined by overlap between frameworks (e.g., shared genes across HoA 2023, Stem Cell Aging 2025, and OpenGenes) used as high-confidence "core" programs.	This work (computed)
LLM-unified hallmark sets	Unified hallmark taxonomy and gene sets produced by harmonizing labels across frameworks into a consistent classifier target space.	This work (LLM-assisted)

B Implementation summary

Feature construction. Gene-level features are constructed by selecting HVGs and computing PCA. Hallmark features are computed by gene-set scoring and used for both visualization and regression.

Clock training and selection. We train Ridge regression clocks on gene-PCA features, hallmark features, and their concatenation. When donor identifiers are available, we split train/test by donor to prevent leakage. We select the best model by held-out MAE, refit it on all cells, and store per-cell predictions as $\hat{a}_{\mathrm{cell}}(x)$.

Per-donor aggregation. Donor-level predictions are computed by averaging per-cell predictions within each donor.

Residual analysis. Residuals $r(x) = a - \hat{a}_{\mathrm{cell}}(x)$ are visualized across hallmark-space embeddings and compared across clusters to identify structured unexplained signal.

C Dataset availability

All single-cell datasets used to construct the human skeletal muscle aging atlas single-cell and single-nucleus combined dataset are publicly available. Source links are provided below:

- Human skeletal muscle aging atlas (data availability) (Kedlian et al., 2024): `https://www.nature.com/articles/s43587-024-00613-3#data-availability`

- Muscle Ageing Cell Atlas portal (Kedlian et al., 2024): https://www.muscleageingcellatlas.org/
- GEO accession GSE167186 (Kedlian et al., 2024): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167186
- GEO accession GSE143704 (Kedlian et al., 2024): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143704
- Repository (DOI) (Kedlian et al., 2024): https://doi.org/10.7272/Q65X273X
- Multimodal muscle aging atlas (Lai et al., 2024): https://db.cngb.org/cdcp/hlma/download
- Tabula Sapiens (Consortium, 2020): https://figshare.com/articles/dataset/Tabula_Sapiens_v2/27921984

D HallmarkAge Pipeline: End-to-End Scoring, Classification, Clock Modeling, and Discovery

This appendix summarizes the full pipeline implemented in the accompanying notebook. The pipeline operationalizes hallmark biology in single-cell RNA-seq by (i) scoring predefined hallmark gene sets, (ii) converting continuous hallmark activity into discrete ON/OFF training targets, (iii) training supervised hallmark classifiers that output calibrated hallmark probabilities, (iv) aggregating hallmark probabilities to the sample level to build hallmark-informed aging clocks, and (v) using residual signal for candidate novel hallmark and longevity target discovery. Pseudobulking methodology is described in detail in Appendix F.

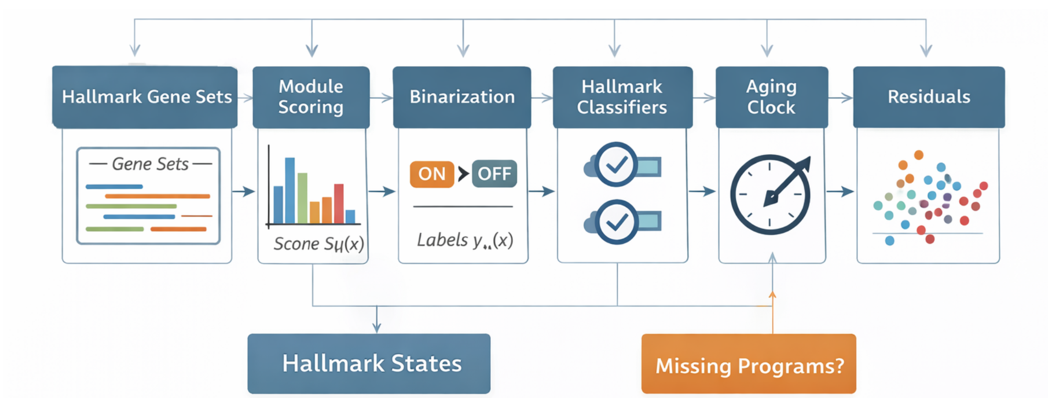

Figure S1: HallmarkAge overview. Hallmark gene sets are scored and converted into binary state labels, used to train probabilistic hallmark classifiers. Hallmark representations support interpretable embeddings and age prediction, and residual analysis can highlight candidate missing programs.

D.1 Inputs and required metadata

The pipeline operates on a log-normalized single-cell `AnnData` object `adata` (using `adata.X` or a specified expression `layer`). Cell-level metadata are stored in `adata.obs`. At minimum, the notebook requires:

- `sample_id`: donor/sample identifier used to aggregate cell-level predictions.
- `age`: chronological age (numeric) for each sample/donor (propagated to cells if needed).

D.2 Hallmark gene set frameworks

The notebook supports multiple hallmark gene set frameworks (Appendix L), including canonical hallmark collections (e.g., Hallmarks of Aging 2013/2023), stem-cell aging hallmarks, OpenGenes mechanism-level mappings, and unified/overlap-derived hallmark sets computed in this work. Each framework defines a set of hallmark gene lists $\mathcal{G} = \{G_h\}_{h=1}^{H}$.

D.3 Step 1: Hallmark scoring (continuous activity)

For each hallmark gene set G_h, the notebook computes a per-cell hallmark activity score using Scanpy gene-set scoring (`scanpy.tl.score_genes`):

$$S_h(i) = \mathrm{Score}(x_i; G_h), \tag{1}$$

where x_i is the expression profile of cell i. Scores are stored in `adata.obs` as `score__<framework>__<hallmark>`.

D.4 Step 2: Quantile-based binarization into hallmark ON/OFF states

To train supervised hallmark classifiers, the notebook converts continuous scores into binary labels using quantile rules. The default label is based on the top 20% of scores:

$$y_h^{(q80)}(i) = \mathbb{I}\big(S_h(i) > q_{0.8}\big), \tag{2}$$

where $q_{0.8}$ is the 80th percentile of $\{S_h(i)\}$ across cells. These labels are stored as `label_q80__<framework>__<hallmark>`.

To increase label purity, the notebook additionally supports an "extreme-only" scheme using both 20th and 80th percentiles:

$$y_h^{(\mathrm{ext})}(i) = \begin{cases} 1 & S_h(i) > q_{0.8}, \\ 0 & S_h(i) < q_{0.2}, \\ \mathrm{NaN} & \text{otherwise}, \end{cases} \tag{3}$$

with a corresponding training mask $m_h(i) = \mathbb{I}(y_h^{(\mathrm{ext})}(i) \text{ is not NaN})$ to exclude ambiguous mid-range cells. The notebook also optionally attempts a two-component GMM binarization on $S_h(i)$ (when the fit succeeds).

D.5 Step 3: Feature representation

Hallmark classifiers are trained on low-dimensional gene-expression features, typically PCA embeddings derived from the expression matrix:

$$X_{\mathrm{gene}} \in \mathbb{R}^{N \times d}, \tag{4}$$

where N is the number of cells and d is the embedding dimension.

D.6 Step 4: Hallmark classifier training (cell-level)

For each hallmark h, a binary classifier f_h is trained to predict hallmark ON/OFF labels from gene-derived features:

$$\hat{p}_h(i) = f_h\big(X_{\mathrm{gene}}(i)\big) \approx \mathbb{P}(y_h(i) = 1 \mid X_{\mathrm{gene}}(i)). \tag{5}$$

The notebook evaluates standard model families (including logistic regression and gradient boosting) and reports classification performance (e.g., AUROC and average precision). Predicted probabilities $\hat{p}_h(i) \in [0, 1]$ are retained as continuous hallmark activation estimates.

D.7 Step 5: Sample-level aggregation of hallmark probabilities

To build donor/sample-level representations, the notebook aggregates cell-level hallmark probabilities across all cells belonging to a sample s:

$$\bar{p}_h(s) = \mathrm{mean}\big(\{\hat{p}_h(i) : i \in s\}\big), \tag{6}$$

yielding a sample-level hallmark probability vector $\bar{\mathbf{p}}(s) = [\bar{p}_1(s), \ldots, \bar{p}_H(s)]$.

D.8 STEP 6: HALLMARK-INFORMED AGING CLOCKS

The notebook fits regression models to predict chronological age from sample-level representations. It supports: (i) a baseline gene-space clock $g_{\mathrm{gene}}(X_{\mathrm{gene}}(s))$, (ii) a hallmark clock $g_{\mathrm{hall}}(\bar{\mathbf{p}}(s))$ using hallmark probabilities, and (iii) a hybrid clock combining gene and hallmark features. Performance is reported using MAE and R^2 with predicted-vs.-true age visualizations.

D.9 STEP 7: RESIDUAL-BASED DISCOVERY OF MISSING AGING SIGNAL

To identify aging-associated signal not captured by known hallmark sets, the notebook computes residuals from hallmark-based age models:

$$r(s) = \mathrm{Age}(s) - \widehat{\mathrm{Age}}(s), \tag{7}$$

and prioritizes genes or programs correlated with $r(s)$ as candidates for novel hallmarks or longevity targets, emphasizing robustness and reproducibility.

D.10 STEP 8: DATA-DRIVEN CLUSTER PROGRAMS FROM MARKER GENES

In addition to predefined hallmark gene sets, the notebook derives cluster-level transcriptional programs by computing marker genes for cluster labels available in `adata.obs` (via Scanpy `rank_genes_groups`). These marker-defined programs provide an additional discovery layer and can be used as auxiliary targets or features in age models.

D.11 OUTPUTS

The notebook saves hallmark scores, multiple label variants (quantile, extreme-only, optional GMM), training masks, predicted hallmark probabilities (cell- and sample-level), age model predictions and evaluation metrics, and cluster marker gene summaries for downstream analysis and reporting.

E UMAP VISUALIZATION OF HALLMARK AND AGING PROGRAM SCORES

To assess how aging hallmarks and related biological programs vary across the cellular manifold, we projected per-cell hallmark scores onto a shared UMAP embedding. Scores were computed using multiple gene set sources and abstraction levels, including OpenGenes (full TSV and summary), LLM-unified aging programs, canonical Hallmarks of Aging (2013, 2023), and stem-cell–specific aging dimensions. All panels use the same UMAP coordinates for direct visual comparison.

Across all scoring frameworks, hallmark signals exhibit structured but overlapping gradients on the UMAP rather than sharp cluster boundaries. This supports a view of aging as a continuous, multi-dimensional process with partially decoupled biological programs operating within the same cellular populations.

E.1 OVERLAPPING HALLMARKS SCORES

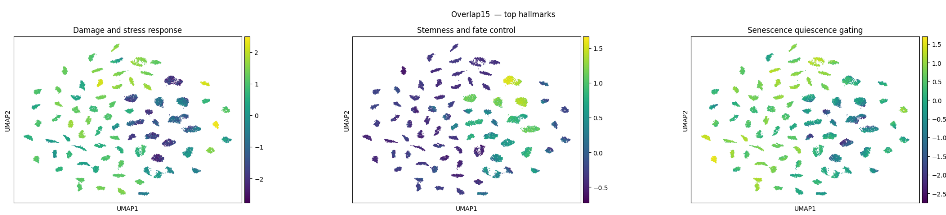

Figure S2: UMAP projection of overlapping 15 hallmark genes scores between HoA2023, HoSCA2025, and OpenGenes, representing aggregated pathway-level signals for damage and stress response, stemness an fate control, metabolic and quiescence control, epigenetic aging, and senescence/quiescence gating.

E.2 OpenGenes Summary Scores

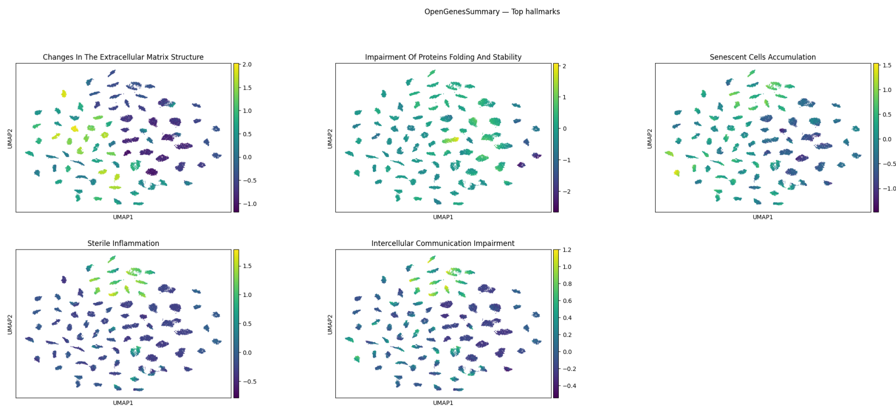

Figure S3: UMAP projection of OpenGenes Summary hallmark scores, representing aggregated pathway-level signals for canonical aging mechanisms including IGF/mTOR signaling, telomere attrition, chromatin remodeling, mitochondrial instability, and extracellular matrix alterations.

E.3 OpenGenes Full Gene Set Scores

Figure S4: UMAP projection of per-cell scores for OpenGenes Full TSV aging-related gene sets, including stem cell exhaustion, autophagy impairment, telomere attrition, epigenetic alterations, proteostasis decline, mitochondrial dysfunction, inflammatory signaling, and related pathways.

E.4 LLM-Unified Aging Programs

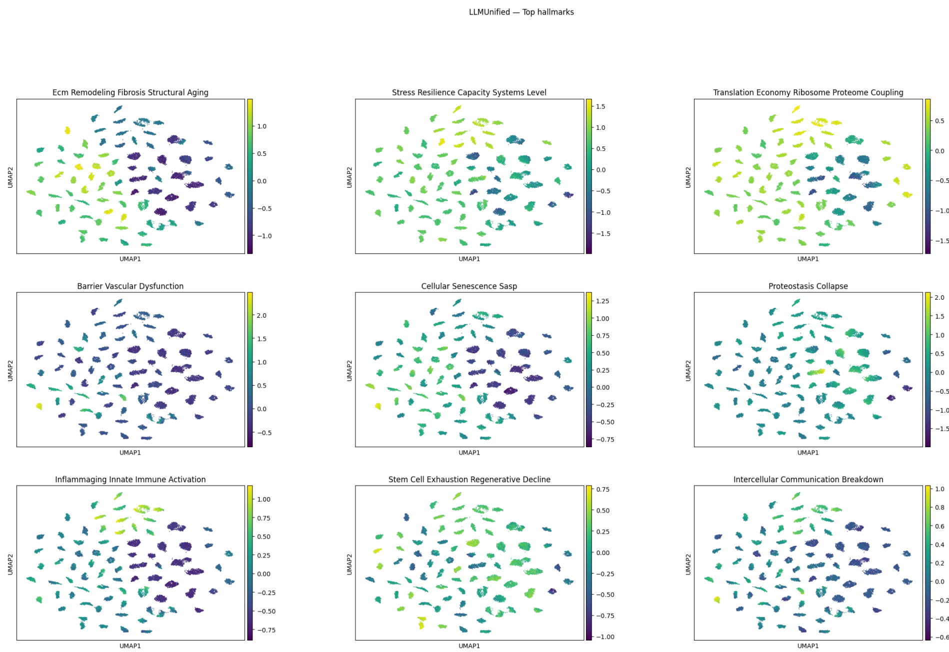

Figure S5: UMAP visualization of LLM-unified aging program scores, integrating genome integrity failure, epigenetic drift, proteostasis collapse, autophagy–lysosome impairment, mitochondrial stress, nutrient sensing dysregulation, inflammaging, cellular senescence, and regenerative decline.

E.5 Stem Cell–Specific Aging Dimensions

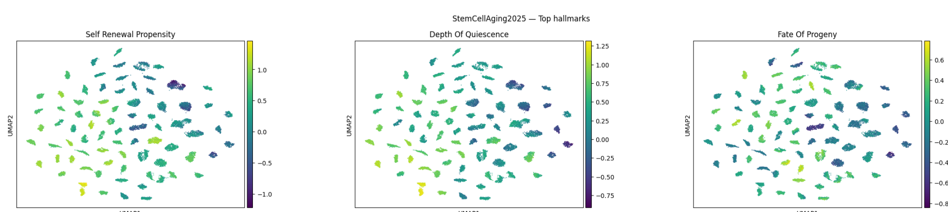

Figure S6: Stem cell–specific aging scores projected onto the shared UMAP, capturing resilience, population heterogeneity, quiescence depth, self-renewal propensity, fate bias, and transcriptional dysregulation.

E.6 Hallmarks of Aging (2023)

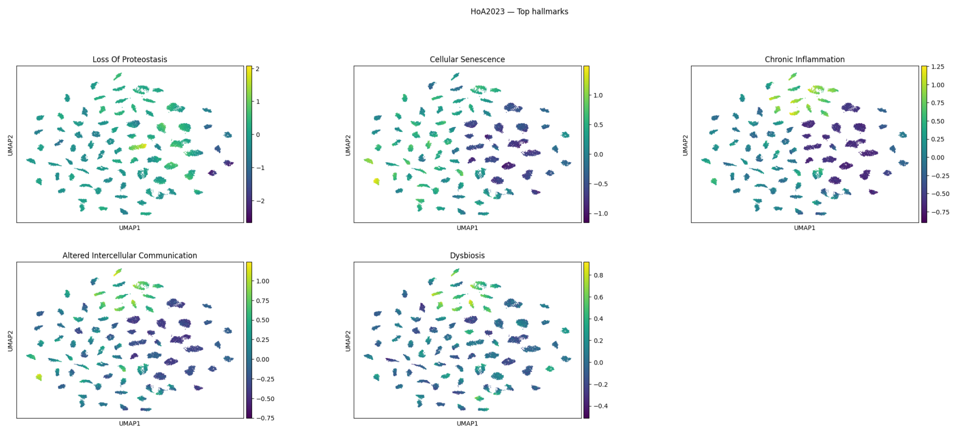

Figure S7: UMAP visualization of Hallmarks of Aging (2023) scores, extending canonical hallmarks to include dysbiosis, chronic inflammation, nutrient sensing deregulation, and updated definitions of proteostasis and mitochondrial decline.

E.7 Hallmarks of Aging (2013)

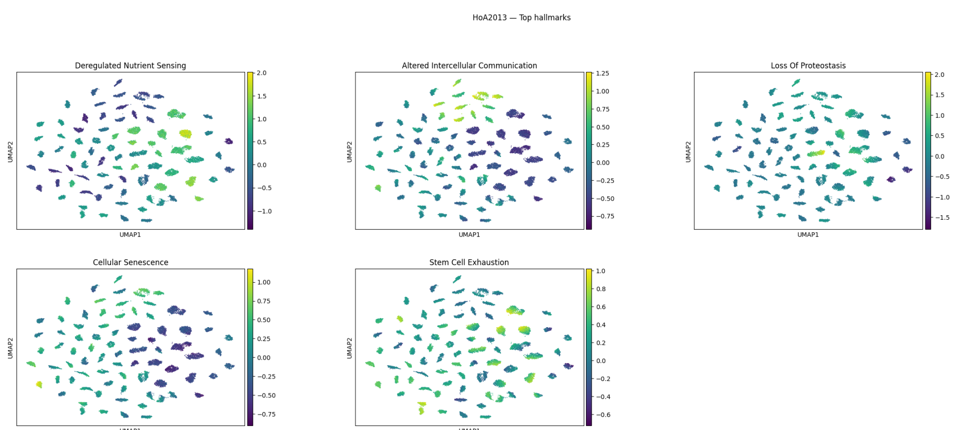

Figure S8: UMAP visualization of Hallmarks of Aging (2013) scores, including genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication.

F PSEUDOBULK (METACELL) CONSTRUCTION

Let $X \in \mathbb{N}^{N \times G}$ denote the single-cell count matrix with N cells and G genes, and let $d(i)$ and $a(i)$ denote the donor identity and age associated with cell i, respectively. To reduce technical noise while preserving donor-level biological signal, we constructed pseudobulk profiles via metacell aggregation.

Highly variable genes (HVGs) were selected on the *single-cell data from training donors only* using a Seurat v3–style variance model as implemented in Scanpy. In this work, batches correspond to datasets, and HVG selection was performed in a batch-aware manner to account for cross-dataset technical effects. Specifically, in `select_hvgs` we ran `scanpy.pp.highly_variable_genes(..., flavor="seurat_v3", layer="counts", batch_key="batch")` after `normalize_total` followed by `log1p`. Providing `batch_key` prompts Scanpy to fit the mean–variance trend separately within each dataset and then aggregate dispersion across datasets, yielding HVGs that are consistently variable across datasets rather than driven by a single dataset. This reduces the selection of dataset-specific HVGs and preferentially retains genes exhibiting reproducible variability across multiple datasets.

Let $H \subset \{1, \ldots, G\}$ denote the resulting set of intersecting HVGs across datasets. All subsequent aggregation was performed after restricting X to H to avoid introducing non-informative features.

For each donor d, metacells were constructed by randomly sampling k cells $\{i_1, \ldots, i_k\}$ such that $d(i_j) = d$ and summing their raw counts to form a pseudobulk expression vector

$$\tilde{x}_{d,m} = \sum_{j=1}^{k} X_{i_j,H},$$

where m indexes metacells for donor d. Sampling was performed with replacement when fewer than k cells were available. This procedure was repeated to generate a fixed number of metacells per donor, yielding multiple pseudobulk observations per individual.

Each metacell was assigned an age

$$\tilde{a}_{d,m} = \frac{1}{k} \sum_{j=1}^{k} a(i_j),$$

corresponding to the mean age of its constituent cells. Following aggregation, pseudobulk count matrices were library-size normalized, log-transformed, and scaled prior to downstream modeling. HVG sets computed on the training fold were reused unchanged for validation and test folds to ensure strict separation between training and evaluation data.

G Scope, Hallmark Definitions, and Modeling Choices

This appendix clarifies the biological scope, hallmark definitions, modeling strategy, and intended interpretation of HallmarkAge. The objective of this section is to describe the analyses performed, the construction and evaluation of the models, and the interpretation of the resulting findings in the context of existing transcriptomic aging clocks.

G.1 Hallmark frameworks and biological scope

HallmarkAge does not assume a single, definitive set of aging hallmarks. Instead, it operates over multiple published hallmark and mechanism frameworks, treating them as structured biological priors rather than mechanistic ground truth. In this work, we analyze canonical Hallmarks of Aging (2013), the expanded Hallmarks of Aging (2023), Hallmarks of Stem Cell Aging (2025), and the OpenGenes mechanism-level annotation of aging-related genes, along with derived collapsed and overlap-based abstractions (Appendix L).

All empirical results presented in the main text are restricted to a single biological system: human skeletal muscle cells. While the combined dataset spans multiple donors and cell-types, it does not span multiple tissues. Accordingly, HallmarkAge is evaluated here as a single-tissue, multi-cell-type framework, and no claims are made regarding cross-tissue generalization.

G.2 Hallmark state construction and classifiers

For each hallmark within a given framework, we compute continuous per-cell gene-set activity scores using standard gene-set scoring. These continuous scores are then converted into high-confidence binary pseudo-labels using quantile-based thresholding, with only extreme-score cells used for supervision. This binarization is employed solely to obtain contrastive training labels and does not imply that hallmark activity is biologically discrete.

Using these pseudo-labels, we train one probabilistic classifier per hallmark. Each classifier predicts the probability that a given cell exhibits high activity for that hallmark, yielding a vector of hallmark activation probabilities for every cell. These probabilities form a low-dimensional, interpretable representation of cellular aging state that is used for visualization, clustering, and downstream modeling. The number of genes contributing to each hallmark varies by framework and reflects the original literature sources; HallmarkAge does not redefine or manually curate gene sets beyond explicitly labeled derived abstractions.

G.3 Age modeling and prediction targets

All aging clocks in this work are trained to predict chronological age only. Chronological age is used as a calibration signal rather than a proxy for health, mortality, or functional decline. We do not train or evaluate models for lifespan, mortality risk, or clinical phenotypes.

Age prediction models are constructed using three feature representations: (i) gene-level PCA features, (ii) hallmark-derived features, and (iii) a hybrid concatenation of gene and hallmark features. Ridge regression is used throughout for consistency and interpretability. When donor identifiers are available, train/test splits are performed at the donor level to avoid information leakage across cells from the same individual. Model selection is based on held-out mean absolute error, and final models are refit on all cells to produce per-cell predictions that are subsequently aggregated to the donor level by averaging.

Hallmark-only clocks are not designed to outperform high-dimensional gene-level clocks in predictive accuracy. Their primary purpose is to provide mechanistically interpretable structure and to enable analysis of aging-related variation in a biologically grounded coordinate system. Hybrid models retain predictive performance while providing hallmark interpretability.

G.4 Interpretation of hallmark structure and residuals

Hallmark-derived embeddings and clusters reveal structured but overlapping aging states, consistent with aging as a continuous, multi-dimensional process rather than a collection of independent pathways. Individual clusters typically reflect mixtures of hallmark activities rather than single dominant mechanisms, and their interpretation is correspondingly descriptive rather than causal.

After fitting the selected age model, we analyze residuals between predicted and chronological age. Residual structure is treated as informative only when it is reproducible, structured, and non-redundant with existing hallmark representations. Such residuals are used to nominate candidate missing programs or perturbation targets, but they are not interpreted as direct measures of rejuvenation, healthspan, or biological age. Residual-driven hypotheses are explicitly framed as model-dependent and require independent experimental validation.

G.5 Relationship to existing transcriptomic aging clocks

HallmarkAge is intended to complement, not replace, existing transcriptomic aging clocks, including module-based and gene-level models trained for chronological age or mortality. Unlike purely data-driven clocks, HallmarkAge anchors its representation to published biological frameworks, enabling direct interpretation in terms of known aging processes while making explicit the limitations and incompleteness of those frameworks.

We emphasize that hallmark frameworks themselves are heuristic and have been criticized for oversimplifying aging as a set of separable drivers. Consistent with this view, HallmarkAge does not assume independence or causal primacy of individual hallmarks. Instead, it uses hallmark structure as an organizing scaffold for interpretable modeling, while explicitly examining mixed states and residual signal that may reflect systemic or unmodeled aspects of aging.

H Distribution of Hallmark Gene-Set Scores

Figure S9 shows the empirical distributions of continuous hallmark gene-set scores for a representative subset of ten hallmarks spanning the Hallmarks of Aging (2013) and Hallmarks of Aging (2023) frameworks. Scores were computed at the single-cell level using standard gene-set scoring (Scanpy `score_genes`).

Each violin plot summarizes the full score distribution across all cells, with dashed vertical lines indicating the median and interquartile range. These distributions illustrate several properties relevant to the HallmarkAge pipeline. First, hallmark scores exhibit broad, overlapping, and often unimodal distributions rather than clearly separable modes, supporting our treatment of hallmark activity as a continuous latent state rather than a discrete class. Second, the relative spread and skewness differ across hallmarks, reflecting heterogeneity in pathway activity and gene-set construction rather than a universal scale.

These score distributions motivate the quantile-based pseudo-labeling strategy used in HallmarkAge, in which only extreme score regions (e.g., top and bottom quantiles) are used to define high-confidence ON/OFF training labels, while intermediate cells are excluded from classifier supervision. The figure therefore serves as a diagnostic visualization validating that (i) extreme-score subsets exist for each hallmark and (ii) a single global cutoff would be inappropriate. Binarization is used solely to obtain high-purity contrastive supervision for classifier training, not to assert discrete biological states; continuous hallmark activity is preserved for downstream analysis.

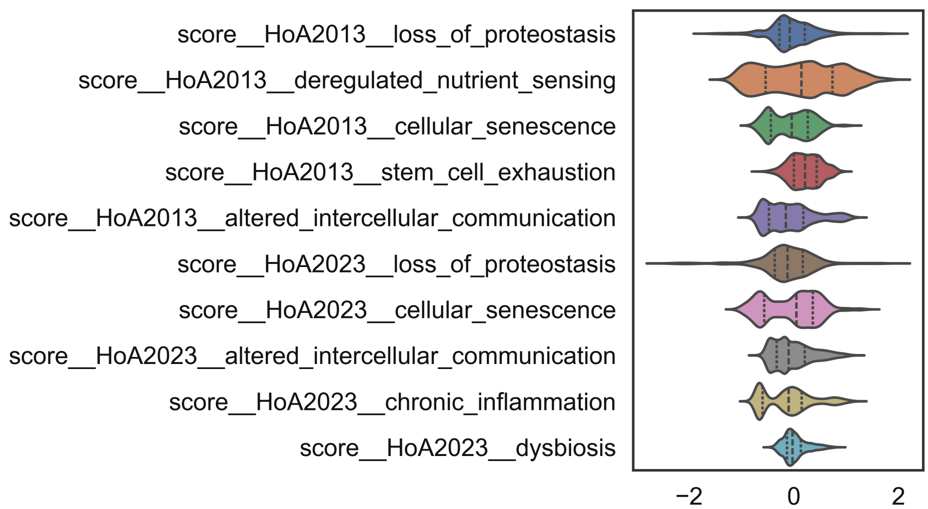

Figure S9: Score distributions for selected hallmark gene sets. Violin plots of single-cell hallmark gene-set score distributions for a subset of ten hallmarks spanning Hallmarks of Aging (2013) and (2023). Dashed vertical lines indicate the median and interquartile range. Broad, overlapping distributions motivate quantile-based pseudo-labeling, where supervision is derived from extreme-score subsets and mid-range cells are excluded.

I GENE-LEVEL EVIDENCE SUPPORTING RESIDUAL-ASSOCIATED AGING PROGRAMS

Table S2: Gene-level correspondence restricted to genes that overlap between residual-associated cluster markers (Table 2) and prior experimental aging literature. Direction indicates the association reported in the literature, while recommended perturbation specifies the intervention expected to improve aging-related outcomes.

Gene	Predicted class	Literature effect	Dir.	Evidence type	Perturbation	References
FOXP1	Anti-aging (C0)	Anti-aging (context-specific)	↑	Stem cell aging (MSC and skeletal aging models)	Activate / preserve (stem cells)	(Wu & Wang, 2024; Wu et al., 2024; Jiao et al., 2021)
IL1B	Anti-aging (C4)	Pro-aging	↑	Inflammaging and frailty (mouse, human)	Inhibit (chronic signaling)	(Böni-Schnetzler et al., 2021; Park et al., 2024; Widjaja et al., 2024)
NLRP3	Pro-aging (C1)	Pro-aging	↓	Healthspan improvement; limited lifespan evidence	Inhibit	(Marín-Aguilar et al., 2020; Verlinden, 2024)
CDKN1A (p21)	Pro-aging (C2)	Pro-aging	↑	Cellular senescence accumulation	Reduce / clear senescent cells	(López-Domínguez et al., 2021; Jin et al., 2025)
JUN / JUNB	Pro-aging (C2)	Pro-aging	↑	Stress-response and aging tissues (transcriptomic)	Inhibit (chronic activation)	(Karakaslar et al., 2023)

Integrating the residual-stratified hallmark states (Table 2) with prior gene-level precedent (Table S2) indicates that only a small subset of residual-associated genes currently demonstrate direct experimental support in aging models. Among genes that explicitly overlap between residual-associated cluster markers and the curated gene-level validation set, FOXP1 emerges as the clearest example of a context-specific anti-aging factor, where preservation of expression within mesenchymal and skeletal stem cell compartments maintains stem cell identity and delays tissue aging. This supports the interpretation that certain residual-young states may reflect preservation of progenitor competence rather than generalized activation of youth-associated pathways.

In contrast, the majority of overlapping genes are consistently reported as pro-aging when chronically active, with beneficial effects arising primarily from their suppression rather than sustained expression. These include inflammatory and innate immune regulators such as IL1B and NLRP3, whose persistent activation drives inflammaging, frailty, and tissue dysfunction, as well as stress- and senescence-associated transcriptional regulators such as CDKN1A (p21) and JUN/JUNB. Experimental evidence across genetic and pharmacologic studies indicates that attenuation of these programs improves healthspan and, in some cases, partially reverses age-associated phenotypes.

The recurrence of these genes within residual-associated clusters therefore supports an interpretation in which hallmark-space embeddings capture aging state signatures rather than direct single-gene longevity drivers. Residual-young or protective clusters may reflect transient stress-adaptive, immune-modulatory, or progenitor-associated states that correlate with lower predicted biological age, even when the underlying genes function as pro-aging factors under chronic activation.

More broadly, many gene categories enriched within residual-young clusters—including epigenetic modifiers, interferon-stimulated genes, and proliferation-associated programs—currently lack direct experimental evidence demonstrating anti-aging effects when activated. As such, the residual analysis primarily prioritizes pathway-level aging programs and candidate intervention directions, rather than definitive gene-level rejuvenation targets.

J COEFFICIENT STRUCTURE AND REGULARIZATION DIAGNOSTICS

This appendix reports coefficient-level diagnostics for the Ridge regression age models used throughout this work. We analyze (i) the most positively and negatively weighted features, and (ii) features with near-zero coefficients, across baseline gene-level, hallmark-only, and hybrid gene–hallmark models. These analyses are intended as diagnostic and hypothesis-generating tools rather than mechanistic claims, illustrating how regularization redistributes explanatory weight across correlated biological representations and how learned structure can be leveraged for perturbation prioritization and exclusion.

J.1 GENE-LEVEL COEFFICIENT STRUCTURE IN BASELINE AND HYBRID MODELS

We first examine the largest-magnitude gene coefficients learned by the baseline Ridge model trained on gene-PCA features and by the hybrid Ridge model trained on concatenated gene and hallmark features.

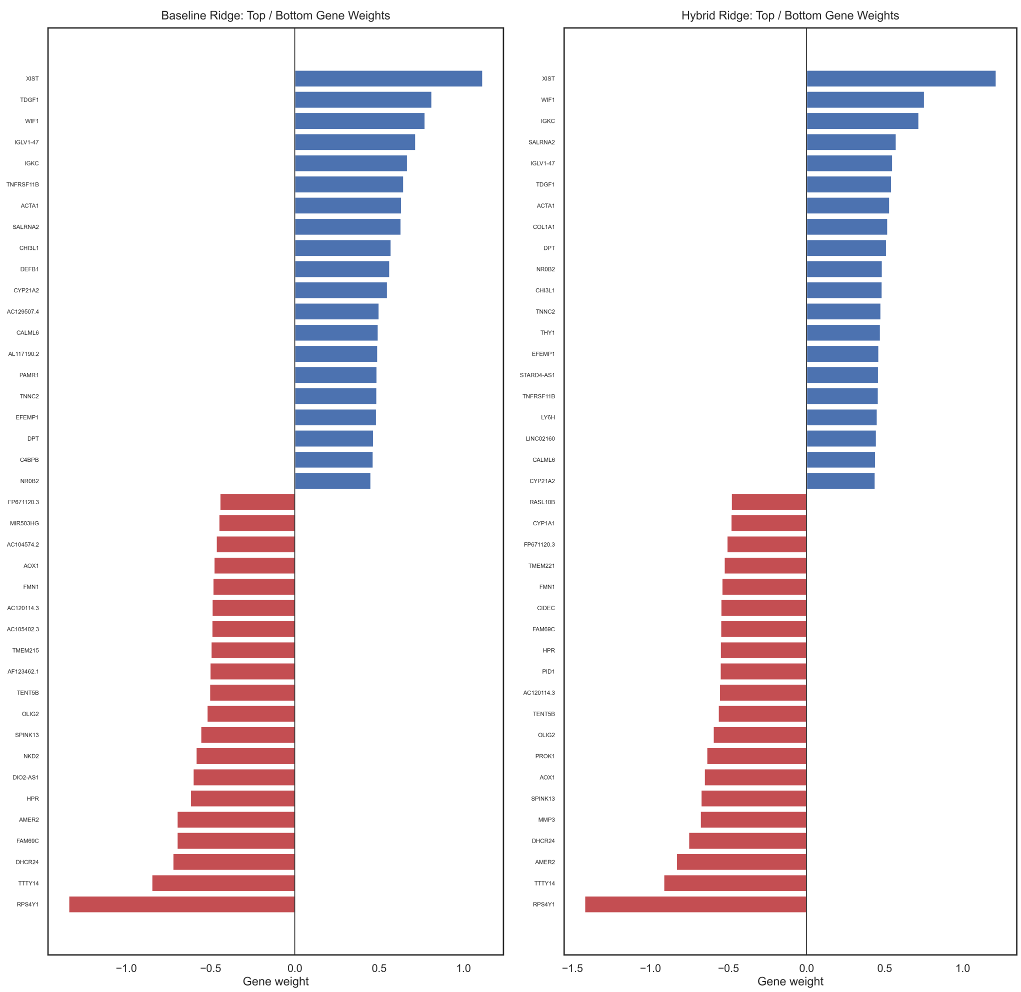

Figure S10: Top and bottom gene coefficients in baseline and hybrid Ridge models. Horizontal bar plots show the genes with the largest positive (blue) and negative (red) coefficients. In the baseline model, individual genes absorb a large fraction of the predictive signal. In the hybrid model, gene-level coefficients are attenuated as part of the age signal is reallocated to hallmark-derived features.

Several observations motivate the use of hallmark-derived representations. In the baseline model, individual genes absorb substantial predictive weight, reflecting correlated expression structure, donor composition effects, and cell-state variation. In the hybrid model, many gene-level coefficients are reduced in magnitude, consistent with hallmark features capturing shared biological variance and reducing reliance on individual genes as proxies for broader aging programs.

Importantly, large-magnitude gene coefficients are not interpreted as causal drivers of aging. Instead, they identify genes whose expression patterns act as sensitive markers of aging-associated states under regularization. As such, these genes provide a tractable and model-informed entry point for downstream *in silico* perturbation analyses (e.g., simulated knockdown or overexpression within learned representations) and experimental validation, where perturbations can be evaluated for their ability to shift hallmark activation profiles or age-prediction residuals.

J.2 Near-zero hallmark coefficients as redundancy and exclusion diagnostics

To assess redundancy, identifiability, and population-specific relevance among hallmark features, we examine hallmarks whose absolute coefficients fall below fixed fractions of the maximum coefficient magnitude. We report thresholds at 1%, 5%, and 10% of the maximum absolute coefficient, following standard regularization diagnostics.

J.2.1 Hybrid Ridge (1% threshold)

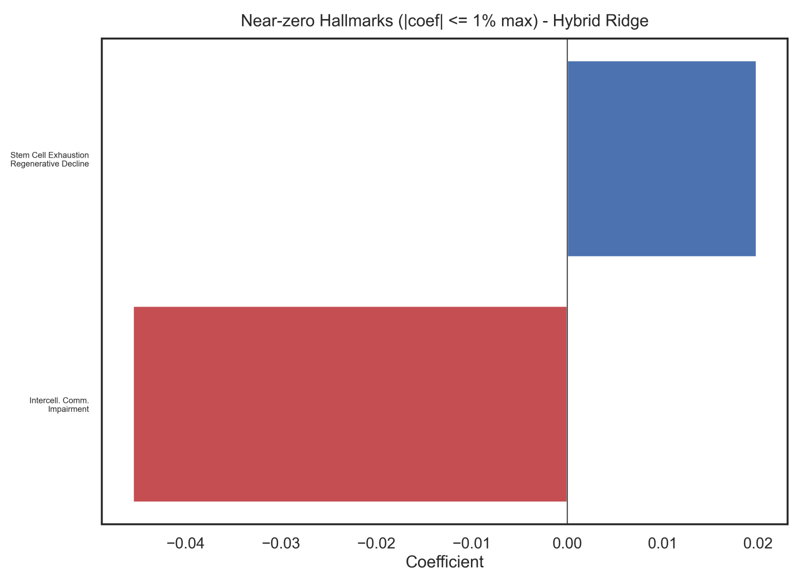

Figure S11: Near-zero hallmark coefficients ($|\beta| \leq 1\%$ of max) in the hybrid Ridge model. Only a small subset of hallmarks exhibit coefficients close to zero, indicating that most hallmark abstractions contribute non-negligible predictive signal when combined with gene-level information.

At this strict threshold, only a limited number of hallmarks are effectively unused by the hybrid model, suggesting that most canonical aging programs retain some relevance after accounting for correlated gene-level structure.

J.2.2 Hallmark-only vs hybrid Ridge (5% threshold)

The hallmark-only model suppresses a greater number of hallmarks, indicating higher redundancy within hallmark space alone. In contrast, the hybrid model distributes weight more smoothly across hallmarks, suggesting complementary rather than competitive use of gene- and hallmark-level information.

J.2.3 Hallmark-only vs hybrid Ridge (10% threshold)

Across thresholds, hallmarks with near-zero coefficients are interpreted not as universally irrelevant aging mechanisms, but as programs that contribute little independent predictive signal in this tissue

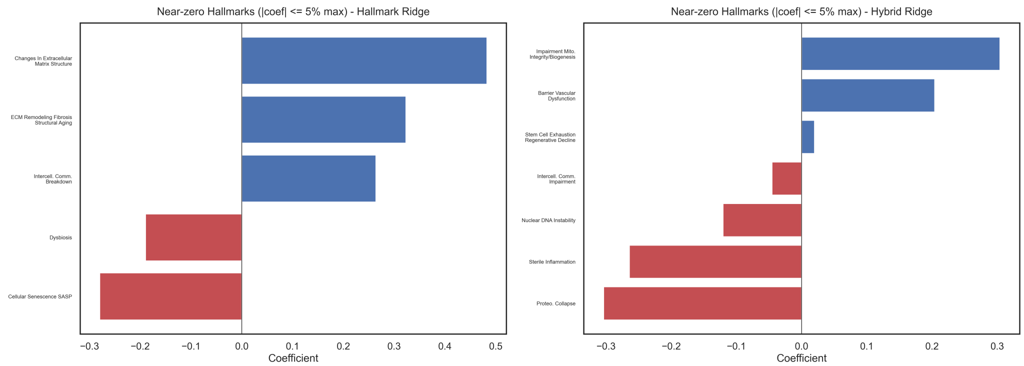

Figure S12: Near-zero hallmark coefficients ($|\beta| \leq 5\%$ of max). Comparison between hallmark-only and hybrid Ridge models. The hallmark-only model exhibits a larger fraction of near-zero coefficients, reflecting stronger competition among correlated hallmark features when gene-level context is absent.

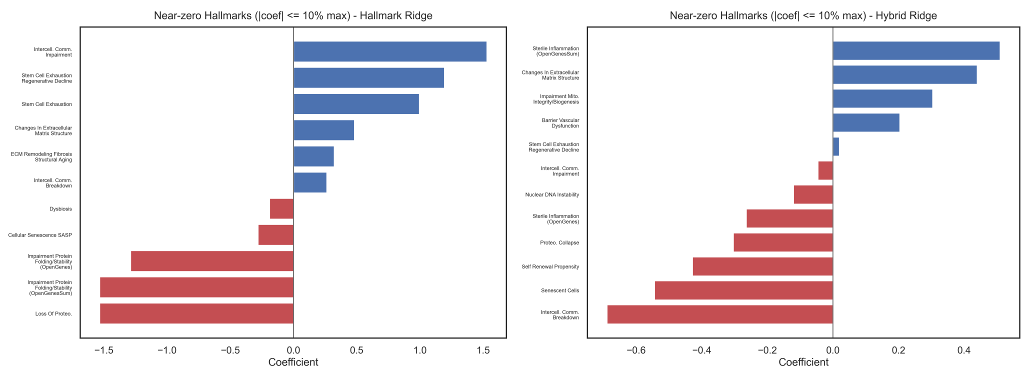

Figure S13: Near-zero hallmark coefficients ($|\beta| \leq 10\%$ of max). At a more permissive threshold, hallmark-only models suppress multiple overlapping aging dimensions, while hybrid models retain broader hallmark engagement.

and cohort. These may correspond to mechanisms that are weakly active in skeletal muscle, redundant with other hallmarks, or downstream consequences rather than limiting constraints on aging rate in this context.

J.3 Top and bottom hallmark coefficients

We finally report the most positively and negatively weighted hallmarks in hallmark-only and hybrid Ridge models.

Coefficient sign and magnitude reflect predictive utility under regularization and correlated feature structure, not direct pro- or anti-aging effects. These diagnostics therefore motivate downstream residual-based analyses rather than single-feature interpretations.

J.4 Perturbation nomination and heuristic validation

Taken together, coefficient structure enables two complementary uses. First, genes with consistently large positive or negative coefficients nominate candidate entry points for *in silico* perturbation screening and experimental validation. Perturbations can be evaluated based on their ability to shift learned aging representations or reduce age-prediction residuals, rather than on single-gene effects alone.

Second, hallmarks with near-zero coefficients provide empirical evidence that certain canonical aging mechanisms are not limiting factors for this population. This selective engagement supports the

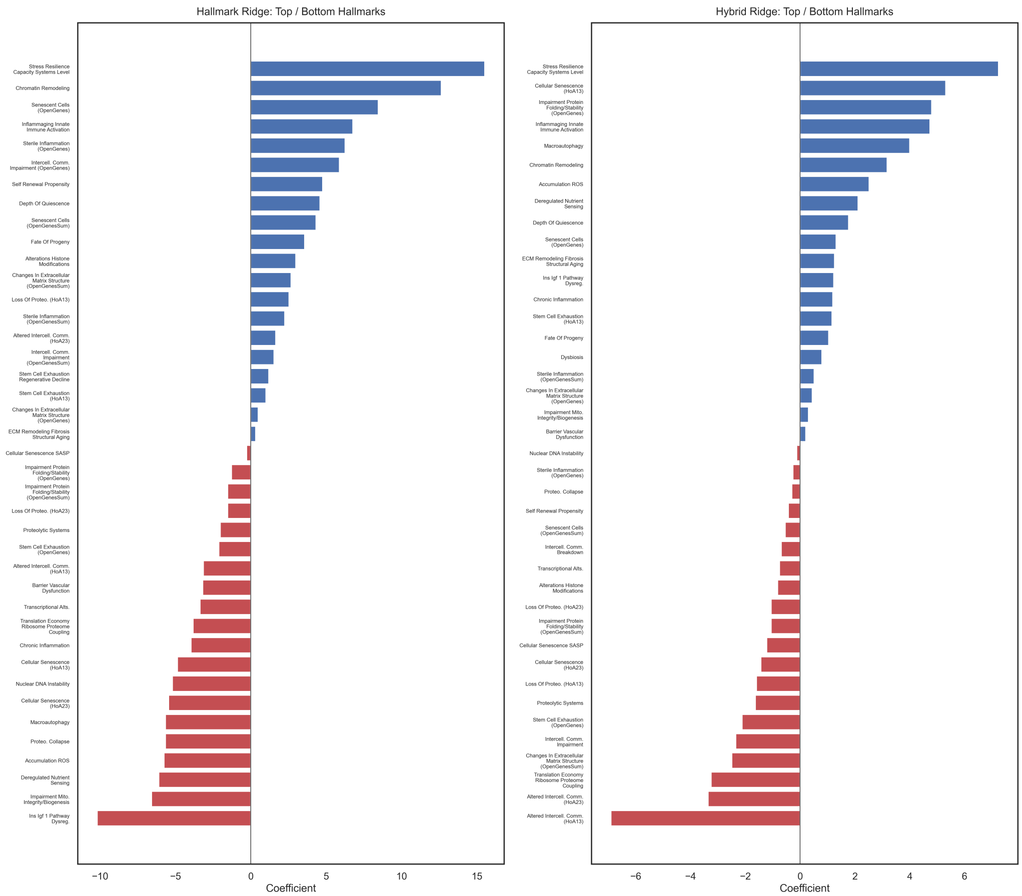

Figure S14: Top and bottom hallmark coefficients. Stress resilience, chromatin remodeling, and senescence-related programs receive the largest positive weights, while proteostasis collapse, inflammatory signaling, and intercellular communication breakdown receive large negative weights. Differences between hallmark-only and hybrid models reflect redistribution of predictive signal rather than changes in qualitative directionality.

view that hallmark taxonomies function as heuristic organizing frameworks rather than mechanistic definitions of aging. Many hallmarks are interdependent and context-dependent; their boundaries overlap, and categories often conflate primary drivers with compensatory responses and downstream consequences. Treating hallmarks as independent causal axes therefore risks biasing interpretation toward single-factor models at the expense of system-level constraints (Gladyshev, 2012; 2013; Gladyshev et al., 2021).

Summary. Coefficient diagnostics demonstrate that hallmark-derived features meaningfully contribute to age prediction, reduce reliance on individual genes, and distribute predictive signal across biologically interpretable axes. Crucially, coefficient structure is leveraged for prioritization and exclusion: large-magnitude gene coefficients nominate candidates for perturbation-based validation, while near-zero hallmark coefficients identify aging mechanisms that are weakly informative or redundant in this population. This framing reinforces HallmarkAge as a structured, falsifiable interface between descriptive aging frameworks and data-driven modeling, rather than a claim of causal sufficiency.

K Age Prediction Performance Tables

This appendix reports full quantitative performance metrics for all age prediction models evaluated in this work. We report mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and Pearson correlation (r), evaluated at both the single-cell and aggregated donor levels where applicable. All models are Ridge regressions trained on baseline gene-PCA features, hallmark-derived features, or their hybrid concatenation.

K.1 Model in-distribution training and testing performance across learning representations

Table S3 reports age prediction performance under donor-level train/test splits, which prevent information leakage across cells from the same individual. Metrics are reported separately for per-metacell predictions and donor-aggregated predictions. We emphasize that hallmark-only clocks are not intended to outperform gene-level baselines, but to provide mechanistically interpretable structure and residual signals for discovery.

Table S3: Age prediction performance under donor-level train/test splits.

Model	Level	Split	MAE	RMSE	R^2	**Pearson** r	n
Baseline Ridge	Metacell	Train	5.68	7.16	0.87	0.93	15,200
Baseline Ridge	Metacell	Test	8.63	10.51	0.69	0.83	3,800
Baseline Ridge	Donor	Train	3.52	4.73	0.94	0.97	76
Baseline Ridge	Donor	Test	7.72	9.35	0.75	0.89	19
Hallmark Ridge	Metacell	Train	10.14	12.58	0.59	0.77	15,200
Hallmark Ridge	Metacell	Test	13.24	16.03	0.28	0.56	3,800
Hallmark Ridge	Donor	Train	8.66	10.41	0.72	0.87	76
Hallmark Ridge	Donor	Test	12.85	15.21	0.35	0.61	19
Hybrid Ridge	Metacell	Train	4.72	6.03	0.91	0.95	15,200
Hybrid Ridge	Metacell	Test	7.31	8.78	0.78	0.89	3,800
Hybrid Ridge	Donor	Train	2.10	2.80	0.98	0.99	76
Hybrid Ridge	Donor	Test	6.39	7.39	0.85	0.94	19

K.2 Model out-of-distribution generalization performance in leave-one-batch-out (LOBO) and leave-one-dataset-out (LODO) evaluations

To assess out-of-distribution (OOD) robustness to dataset-specific shifts, we evaluate models under two complementary cross-validation settings. First, we perform leave-one-batch-out (LOBO) where batches correspond to the source of transcript (cells, nuclei, myonuclei), testing sensitivity to acquisition and processing differences (Table S4). Second, we perform leave-one-dataset-out (LODO) where the held-out set corresponds to a specific cell type or cell-type group, testing generalization across biological composition shifts (Table S5). Across both settings, we observe that OOD evaluation can substantially degrade performance—particularly for hallmark-only clocks—highlighting the value of hybrid representations for robust generalization across heterogeneous single-cell datasets.

Table S4: Leave-one-batch-out (LOBO) age prediction results stratified by transcript source (cells, nuclei, myonuclei). Each model is trained on all remaining transcript sources and evaluated on the held-out source.

Held-out batch	Model	MAE	RMSE	R^2	Pearson r	n_{train}	n_{test}
SKM_myonuclei_human_2024-04-14	Baseline Ridge	4.92	6.26	0.90	0.95	16,400	2,600
SKM_myonuclei_human_2024-04-14	Hallmark Ridge	27.64	32.50	-1.66	0.36	16,400	2,600
SKM_myonuclei_human_2024-04-14	Hybrid Ridge	13.84	15.29	0.41	0.94	16,400	2,600
cells	Baseline Ridge	20.01	25.48	-0.69	0.27	7,800	14,400
cells	Hallmark Ridge	22.97	29.96	-1.33	0.13	7,800	14,400
cells	Hybrid Ridge	19.32	24.50	-0.56	0.24	7,800	14,400
nuclei	Baseline Ridge	5.70	7.51	0.86	0.93	17,000	5,200
nuclei	Hallmark Ridge	43.54	49.53	-5.18	0.39	17,000	5,200
nuclei	Hybrid Ridge	11.24	13.23	0.56	0.89	17,000	5,200

Table S5: Leave-one-dataset-out (LODO) age prediction results where each held-out dataset corresponds to a cell type or cell-type group. Models are trained on all remaining datasets and evaluated on the held-out cell-type group.

Held-out dataset	**Model**	**MAE**	**RMSE**	$\mathbf{R^2}$	**Pearson r**	$\mathbf{n_{train}}$	$\mathbf{n_{test}}$
SKM_human_pp_cells2nuclei_2023-06-22	Baseline Ridge	5.64	8.17	0.82	0.93	3,400	3,400
SKM_human_pp_cells2nuclei_2023-06-22	Hallmark Ridge	10.36	13.06	0.54	0.82	3,400	3,400
SKM_human_pp_cells2nuclei_2023-06-22	Hybrid Ridge	5.61	8.07	0.82	0.94	3,400	3,400
SKM_immune_human_2023-06-22	Baseline Ridge	12.92	16.74	0.27	0.90	3,400	2,400
SKM_immune_human_2023-06-22	Hallmark Ridge	15.76	18.07	0.15	0.39	3,400	2,400
SKM_immune_human_2023-06-22	Hybrid Ridge	11.17	14.92	0.42	0.91	3,400	2,400
SKM_fibroblasts_Schwann_human_2023-06-22	Baseline Ridge	13.80	16.85	0.26	0.85	3,400	2,400
SKM_fibroblasts_Schwann_human_2023-06-22	Hallmark Ridge	15.54	17.91	0.17	0.66	3,400	2,400
SKM_fibroblasts_Schwann_human_2023-06-22	Hybrid Ridge	12.99	15.59	0.37	0.84	3,400	2,400
SKM_myonuclei2myofiber_human_2023-06-22	Baseline Ridge	10.35	15.49	0.35	0.70	3,400	3,400
SKM_myonuclei2myofiber_human_2023-06-22	Hallmark Ridge	14.60	20.54	-0.14	0.41	3,400	3,400
SKM_myonuclei2myofiber_human_2023-06-22	Hybrid Ridge	11.81	15.80	0.33	0.66	3,400	3,400
SKM_Endothelium_SMC_human_2023-06-22	Baseline Ridge	21.19	25.08	-0.64	0.66	3,400	2,400
SKM_Endothelium_SMC_human_2023-06-22	Hallmark Ridge	18.08	21.90	-0.25	0.28	3,400	2,400
SKM_Endothelium_SMC_human_2023-06-22	Hybrid Ridge	16.72	19.11	0.05	0.67	3,400	2,400
SKM_MuSC_human_2023-06-22	Baseline Ridge	22.78	25.99	-0.76	0.72	3,400	2,400
SKM_MuSC_human_2023-06-22	Hallmark Ridge	17.17	20.69	-0.11	0.41	3,400	2,400
SKM_MuSC_human_2023-06-22	Hybrid Ridge	20.64	23.69	-0.46	0.78	3,400	2,400
SKM_myonuclei_human_2024-04-14	Baseline Ridge	5.24	7.16	0.87	0.95	3,400	2,600
SKM_myonuclei_human_2024-04-14	Hallmark Ridge	22.56	26.79	-0.81	0.49	3,400	2,600
SKM_myonuclei_human_2024-04-14	Hybrid Ridge	9.68	11.13	0.69	0.94	3,400	2,600

L Hallmark Gene Set Frameworks and Cross-Framework Overlap

This appendix documents the biological gene-set frameworks used to construct binary hallmark classifiers, define hallmark-space representations, and contextualize residual structure in the HallmarkAge pipeline. We include canonical Hallmarks of Aging (2013), the expanded Hallmarks of Aging (2023), the Hallmarks of Stem Cell Aging (2025), the OpenGenes mechanism-level annotation, and a restricted overlap set of genes shared across frameworks. These frameworks are treated as structured biological priors rather than mechanistic ground truth.

L.1 Hallmarks of Aging (2013)

Table S6: Canonical Hallmarks of Aging (2013) and representative gene sets.

Hallmark	Representative gene set
Genomic instability	TP53, ATM, ATR, CHEK1/2, BRCA1/2, RAD51, PARP1, H2AFX, MRE11, XRCC5/6, LIG4
Telomere attrition	TERT, TERC, DKC1, TERF1/2, POT1, TINF2, RTEL1, WRAP53
Epigenetic alterations	DNMT1/3A/3B, TET1/2/3, EZH2, SUZ12, EED, HDAC1/2, SIRT1/6, KDM6A, SETD2
Loss of proteostasis	HSPA1A, HSP90AA1, DNAJB1, HSF1, BAG3, UBB, SQSTM1, VCP, PSMB5
Deregulated nutrient sensing	IGF1R, INSR, MTOR, AKT1, FOXO1/3, PRKAA1, SIRT1/6
Mitochondrial dysfunction	PPARGC1A, TFAM, POLG, NDUFS1, SDHA, COX4I1, SOD2, PINK1, PRKN
Cellular senescence	CDKN2A, CDKN1A, TP53, RB1, LMNB1, IL6, CXCL8, SERPINE1, TGFB1
Stem cell exhaustion	FOXO3, TERT, BMI1, MYC, CDKN2A, TP53, NOTCH1, WNT3A
Altered intercellular communication	TNF, IL1B, IL6, TGFB1, IFNG, NFKB1, STAT1/3, JAK1, SOCS3

Source: López-Otín et al. (2013).

L.2 Hallmarks of Aging (2023)

Table S7: Expanded Hallmarks of Aging (2023) and representative gene sets.

Hallmark (2023)	Representative gene set
Genomic instability	TP53, ATM, ATR, CHEK1/2, BRCA1/2, RAD51, PARP1, H2AFX, MRE11, NBN, XRCC5/6, LIG4
Telomere attrition	TERT, TERC, DKC1, TERF1/2, POT1, TINF2, ACD, RTEL1, WRAP53
Epigenetic alterations	DNMT1, DNMT3A, TET2, EZH2, HDAC1, SIRT1/6, KDM6A, SETD2, KMT2D
Loss of proteostasis	HSPA1A, HSP90AA1, DNAJB1, HSF1, BAG3, UBB, SQSTM1, VCP, PSMB5
Disabled macroautophagy	BECN1, ATG5/7/12, MAP1LC3B, SQSTM1, ULK1/2, TFEB, LAMP1/2
Deregulated nutrient sensing	IGF1R, INSR, MTOR, RPTOR, AKT1, FOXO3, PRKAA1, SIRT1/6
Mitochondrial dysfunction	PPARGC1A, TFAM, NDUFS1, SDHA, COX4I1, ATP5F1A, SOD2, PINK1, PRKN, MFN2
Cellular senescence	CDKN2A, CDKN1A, TP53, SERPINE1, IL6, CXCL8, CCL2, MMP3, TGFB1
Stem cell exhaustion	SOX2, BMI1, LGR5, PROM1, FOXO3, CDKN2A, TP53, NOTCH1, WNT3A
Altered intercellular communication	NFKB1, RELA, STAT3, IL6, IL1B, TNF, TGFB1, ICAM1, VCAM1, CXCL10
Chronic inflammation	NFKB1, RELA, IL1B, TNF, IL6, IFNG, NLRP3, CASP1, IL18
Dysbiosis	TLR4, TLR2, MYD88, NFKB1, IL1B, TNF, REG3A, DEFB1, MUC2

Source: López-Otín et al. (2023).

L.3 HALLMARKS OF STEM CELL AGING (2025)

Table S8: Hallmarks of Stem Cell Aging (2025) and representative gene sets.

Stem cell aging dimension	Representative gene set
Depth of quiescence	CDKN1A/B, RB1, BTG1/2, HES1, ID1; MYC, CCND1, MKI67; CPT1A, ACADM; MTOR, RPTOR
Self-renewal propensity	SOX2, BMI1, MYC, FOXO3, HMGA2, DNMT3A, TET2, TP53, NOTCH1, WNT3A
Fate of progeny	WNT3A, CTNNB1, AXIN2; NOTCH1, HES1; TGFB1, COL1A1/2; SPI1, CSF1R
Resilience	TP53, ATM, ATR, H2AFX; BAX, BCL2; SOD2, GPX4; NOTCH1
Population heterogeneity	DNMT3A, TET2, TP53, ASXL1, JAK2, SF3B1, SRSF2

Source: Rando et al. (2025).

L.4 OPENGENES AGING MECHANISMS

Table S9: OpenGenes mechanism-level annotation of top aging-related genes.

Mechanism	# of genes	Representative genes
Transcriptional alterations	444	TP53, MYC, FOXO3, STAT3, NFKB1
Intercellular communication impairment	204	IL6, TNF, IL1B, TGFB1
Nuclear DNA instability	201	TP53, ATM, ATR, CHEK1, H2AFX
Proteolytic system degradation	193	PSMB5, VCP, SQSTM1, UBB
ROS accumulation	165	SOD2, GPX4, GCLC
Sterile inflammation	131	NFKB1, RELA, IL1B, NLRP3
Mitochondrial integrity and biogenesis	110	PPARGC1A, TFAM, PINK1, PRKN
Stem cell exhaustion	49	SOX2, BMI1, FOXO3

Source: OpenGenes database (Rafikova et al., 2024).

L.5 OpenGenes Summary Mechanism Gene Sets

In addition to the fine-grained OpenGenes Full TSV annotations (Table S9), we constructed a reduced *OpenGenes Summary* gene-set framework that aggregates closely related mechanisms into higher-level biological programs. This summary representation is designed to (i) reduce redundancy across overlapping OpenGenes categories, (ii) stabilize classifier training in small-sample regimes, and (iii) align OpenGenes mechanisms with canonical aging hallmarks and system-level failure modes.

The OpenGenes Summary gene sets are not intended to be exhaustive, but instead to represent robust, interpretable mechanism-level signals that recur across multiple aging studies and tissues.

Table S10: OpenGenes Summary mechanism-level gene sets used for hallmark classifier construction.

Summary mechanism	Representative gene set
Nuclear DNA instability	TP53, ATM, ATR, CHEK1, CHEK2, H2AFX, BRCA1, BRCA2, RAD51
Alterations in DNA methylation	DNMT1, DNMT3A, DNMT3B, TET2, TET3, EZH2, SETD2
Chromatin remodeling	EZH2, SUZ12, EED, KDM6A, HDAC1, HDAC2, SMARCA4
Impairment of proteostasis	HSPA1A, HSP90AA1, DNAJB1, BAG3, SQSTM1, VCP, PSMB5
Disabled macroautophagy	BECN1, ATG5, ATG7, MAP1LC3B, ULK1, TFEB, LAMP1
Accumulation of reactive oxygen species	SOD2, GPX4, GCLC, PRDX1, CAT
Mitochondrial integrity and biogenesis impairment	PPARGC1A, TFAM, PINK1, PRKN, MFN2, NDUFS1
Intercellular communication impairment	IL6, TNF, IL1B, TGFB1, STAT3, NFKB1
Chronic inflammation / inflammaging	NFKB1, RELA, IL6, IL1B, TNF, NLRP3, CASP1
ECM remodeling and fibrosis	TGFB1, COL1A1, COL1A2, FN1, ACTA2, MMP2
Stem cell exhaustion	SOX2, BMI1, FOXO3, NOTCH1, WNT3A
Nutrient sensing dysregulation	IGF1R, INSR, MTOR, RPTOR, AKT1, PRKAA1, FOXO3

Source: Curated summary of OpenGenes mechanisms derived from the OpenGenes database (Rafikova et al., 2024).

L.6 LLM-UNIFIED HALLMARKS OF AGING

The LLM-derived "unified" hallmark schema here is used as a biological prior for hallmark classifiers. The intent is to (i) reduce redundancy across canonical hallmark taxonomies, (ii) emphasize classifier-stable signals observable in single-cell and multimodal assays, and (iii) represent aging state as mutually exclusive high-level failure modes spanning damage, maintenance, signaling, and regeneration.

Table S11: LLM-unified hallmarks of aging and representative classifier genes, measurement modes, and longevity angles.

Unified hallmark	**Captures**	**Strong classifier genes (ex.)**	**Best measured**	**Longevity angle**
Genome integrity failure	DNA damage, replication stress, faulty repair	TP53, ATM, ATR, CHEK1/2, H2AFX, BRCA1/2, RAD51	RNA+ATAC	DDR tuning
Epigenetic drift / chromatin remodeling	Chromatin state drift $\rightarrow$ transcriptional noise	DNMT1/3A, TET2, EZH2, HDAC1/2, KDM6A	RNA+ATAC	Epigenetic reset
Proteostasis collapse	Misfolding, UPS/ER decline	HSPA1A, HSP90AA1, BAG3, SQSTM1, VCP, PSMB5	RNA / proteome	Chaperone & UPS
Autophagy / lysosome impairment	Reduced cellular recycling	BECN1, ATG5/7, MAP1LC3B, TFEB, LAMP1/2	RNA / proteome	Autophagy boost
Mitochondrial dysfunction + ROS	ETC imbalance, oxidative stress	PPARGC1A, TFAM, SOD2, GPX4, PINK1, PRKN	RNA / proteome	Mitophagy & redox
Nutrient sensing mistuning	mTOR/IGF vs maintenance imbalance	MTOR, RPTOR, AKT1, IGF1R, PRKAA1, FOXO3	RNA / proteome	mTOR/AMPK
Inflammaging / innate activation	Chronic NFκB / IFN tone	NFKB1, RELA, IL1B, TNF, IL6, NLRP3	Plasma + RNA	Anti-inflammatory
Cellular senescence + SASP	Growth arrest + secretory load	CDKN2A, CDKN1A, IL6, CXCL8, TGFB1	RNA / plasma	Senolytics
Stem cell exhaustion	Loss of renewal, lineage bias	SOX2, BMI1, FOXO3, NOTCH1, WNT3A	scRNA	Stemness restore
Intercellular communication breakdown	Maladaptive tissue/niche signaling	IL6, TNF, TGFB1, JAK1/2, ICAM1	Plasma + RNA	Signal recalibration
ECM remodeling / fibrosis	Matrix stiffening, scarring	TGFB1, COL1A1/2, FN1, ACTA2	Proteome	Anti-fibrotic
Stress resilience capacity	Recovery from genotoxic/metabolic stress	ATM, TP53, SOD2, NFE2L2, ATG7	Multi-omic	Resilience hubs
Clonal selection / mosaicism	Expansion of advantaged clones	DNMT3A, TET2, ASXL1, JAK2, TP53	DNA-seq / scRNA	Clone control
Barrier + vascular dysfunction	Endothelial transport failure	ICAM1, VCAM1, VWF, KDR, NOS3	Plasma	Vascular repair
Translation economy	Ribosome–proteome coupling shifts	EIF4EBP1, RPS6, EEF2, NCL	RNA / proteome	Translation tuning

Mutually exclusive, classifier-stable hallmarks spanning molecular damage, maintenance capacity, signaling, and regenerative state — optimized for single-cell and multimodal aging clocks.

L.7 Overlapping Core Aging Genes Across Hallmark Gene Sets

Table S12: Fifteen genes shared across Hallmarks of Aging (2023), Hallmarks of Stem Cell Aging (2025), and OpenGenes, grouped into biological modules and mapped to their originating hallmark gene sets.

Module	Gene	HoA 2023 hallmark(s)	SCA 2025 hallmark(s)
Damage & stress response	ATM	Genomic instability	Resilience
	ATR	Genomic instability	Resilience
	TP53	Genomic instability; cellular senescence; stem cell exhaustion	Self-renewal propensity; resilience; population heterogeneity
	SOD2	Mitochondrial dysfunction	Resilience
Stemness & fate control	SOX2	Stem cell exhaustion	Self-renewal propensity
	BMI1	Stem cell exhaustion	Self-renewal propensity
	NOTCH1	Stem cell exhaustion	Self-renewal propensity; fate of progeny; resilience
	WNT3A	Stem cell exhaustion	Self-renewal propensity; fate of progeny
	LGR5	Stem cell exhaustion	Fate of progeny
Metabolic / quiescence control	MTOR	Deregulated nutrient sensing	Depth of quiescence
	RPTOR	Deregulated nutrient sensing	Depth of quiescence
	FOXO3	Deregulated nutrient sensing; stem cell exhaustion	Self-renewal propensity
Epigenetic aging	EZH2	Epigenetic alterations	Population heterogeneity
Senescence / quiescence gating	CDKN1A	Cellular senescence	Depth of quiescence
	TGFB1	Cellular senescence; altered intercellular communication	Fate of progeny

Sources: López-Otín et al. (2023); Rando et al. (2025); Rafikova et al. (2024).

M Mapping hallmarks to anti-aging interventions

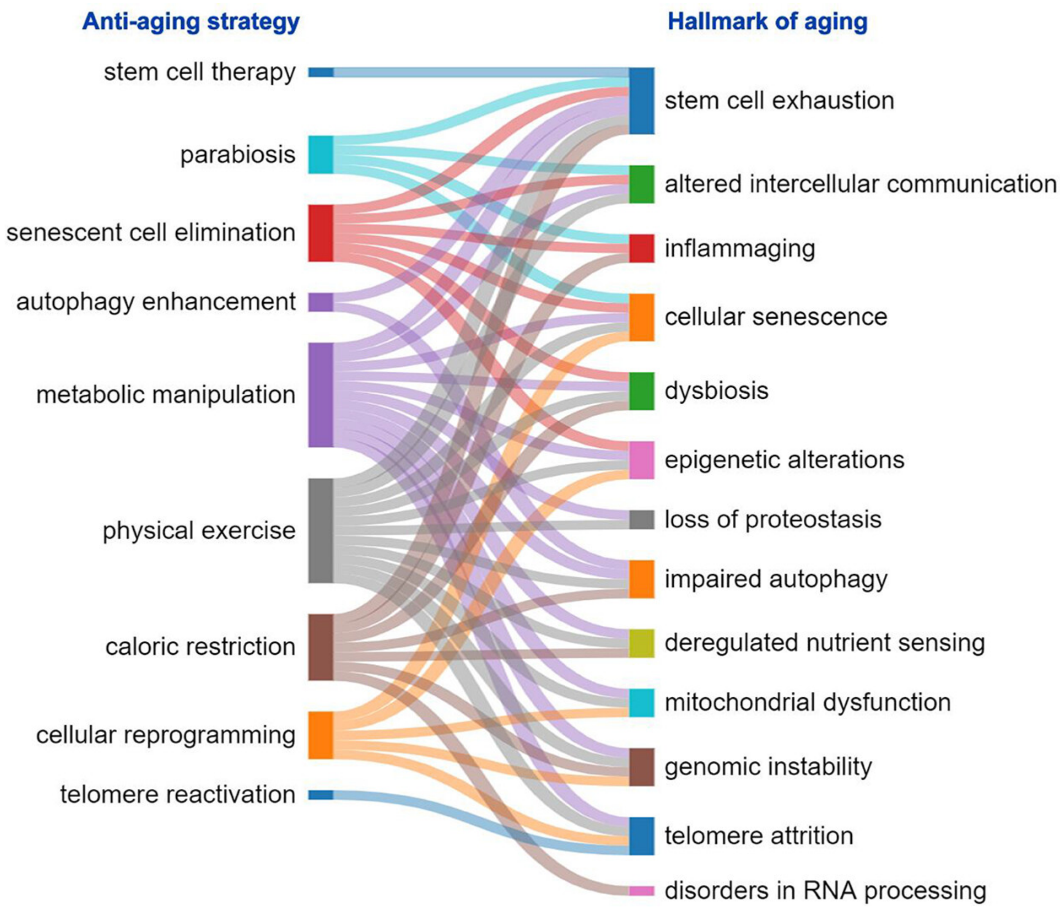

Figure S15: Hallmarks can be selectively targeted through specific anti-aging interventions, including combinatorial therapeutic strategies (Tenchov et al., 2024).

N SECONDARY EVALUATION ON MULTIPLE CELL-TYPES

Table S13: Summary statistics for datasets used in this study across major skeletal muscle cell aging atlases and Tabula Sapiens. The combined dataset contains 1,399,647 cells across 29,400 genes. Data were aggregated into 24,800 pseudobulk metacells using 4,000 highly variable genes (HVGs). Mean ages across datasets is weighted by number of cells per dataset. Results shown in Fig. S16.

Dataset	n_{cells}	n_{donors}	Minimum age	Maximum age	Mean age
FAP (OMIX004308-03) (Lai et al., 2024)	23,504	22	15	92	75.02
Endothelium & SMC (Kedlian et al., 2024)	19,317	12	17.5	72.5	42.01
Muscle Stem Cells (MuSC) (Kedlian et al., 2024)	17,528	12	17.5	72.5	42.65
Fibroblasts & Schwann Cells (Kedlian et al., 2024)	20,611	12	17.5	72.5	46.79
Immune Cells (Kedlian et al., 2024)	20,321	12	17.5	72.5	57.67
Myonuclei (Myofiber-associated) (Kedlian et al., 2024)	87,522	17	17.5	72.5	50.25
Myonuclei (Kedlian et al., 2024)	74,626	13	17.5	72.5	49.05
Tabula Sapiens (Kedlian et al., 2024)	1,136,218	24	22	74	55.36
Total	1,399,647	63	15	92	55.33

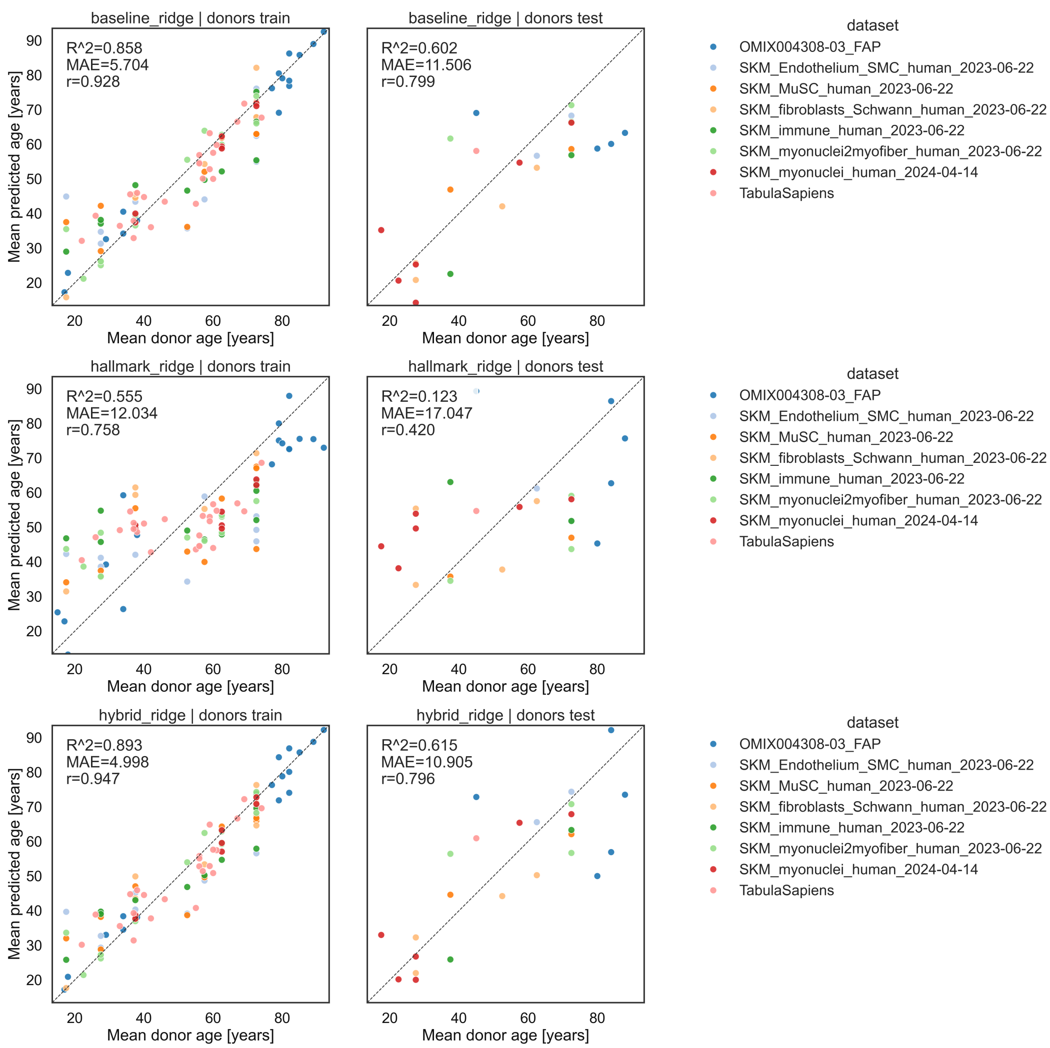

Figure S16: Age prediction performance across feature spaces combining two muscle aging atlases (Kedlian et al., 2024; Lai et al., 2024) and Tabula Sapiens (Consortium, 2020). Hybrid transcriptomic-hallmark model achieves an r of 0.80 and an R^2 of 0.62 across different cell types from 28 organs, each including various tissue types, collected from 63 unique donors as described in Tab. S13.