TailorAge: A Cell-Specific Epigenetic Clock Framework for Profiling Human Biological Aging at Single-Cell Resolution

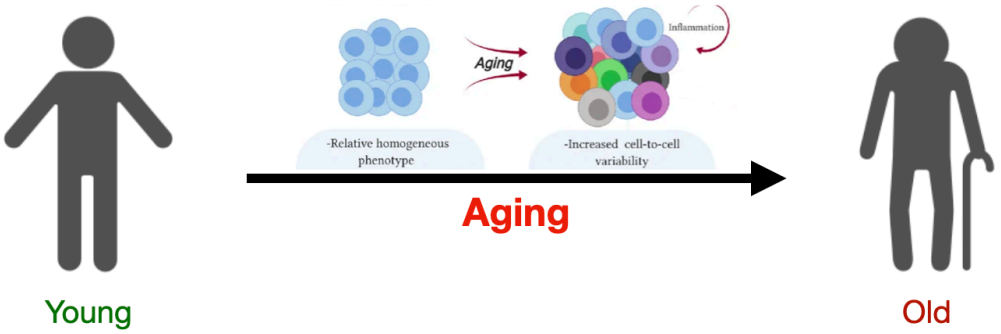
Abstract / Executive Summary

Biological aging is fundamentally driven by progressive cellular heterogeneity and somatic epigenomic drift. Although DNA methylation (DNAm) clocks serve as the gold standard for biological age estimation, existing frameworks rely predominantly on bulk tissue profiles, masking cell-to-cell variability and cell-type-specific aging trajectories. Extending epigenetic clocks to single-cell DNA methylation (scDNAm) data remains challenging due to extreme data sparsity (>95% missingness) and non-overlapping CpG coverage across individual cells.

To overcome this, we developed **TailorAge**, a novel computational framework that dynamically constructs cell-specific epigenetic clocks by aligning bulk reference models to the sparse CpG coverage profile of each single cell. Applied to a benchmark human scDNAm dataset ($n = 291$ cells across 8 donors), TailorAge achieved a strong, statistically significant correlation between predicted single-cell biological age and donor chronological age ($R = 0.79, p < 0.001$). These findings demonstrate that TailorAge effectively stratifies cells across human aging trajectories while capturing intrinsic intra-donor single-cell heterogeneity.

1. Introduction & Research Rationale

Aging is a complex physiological process characterized by loss of cellular homeostasis, programmatic alterations, and stochastic epigenetic noise (Figure 1). Because tissues age mosaicly—with distinct cellular lineages displaying non-uniform rates of functional decline—evaluating biological age at single-cell resolution is critical for deciphering the mechanisms of aging and validating rejuvenation therapies.

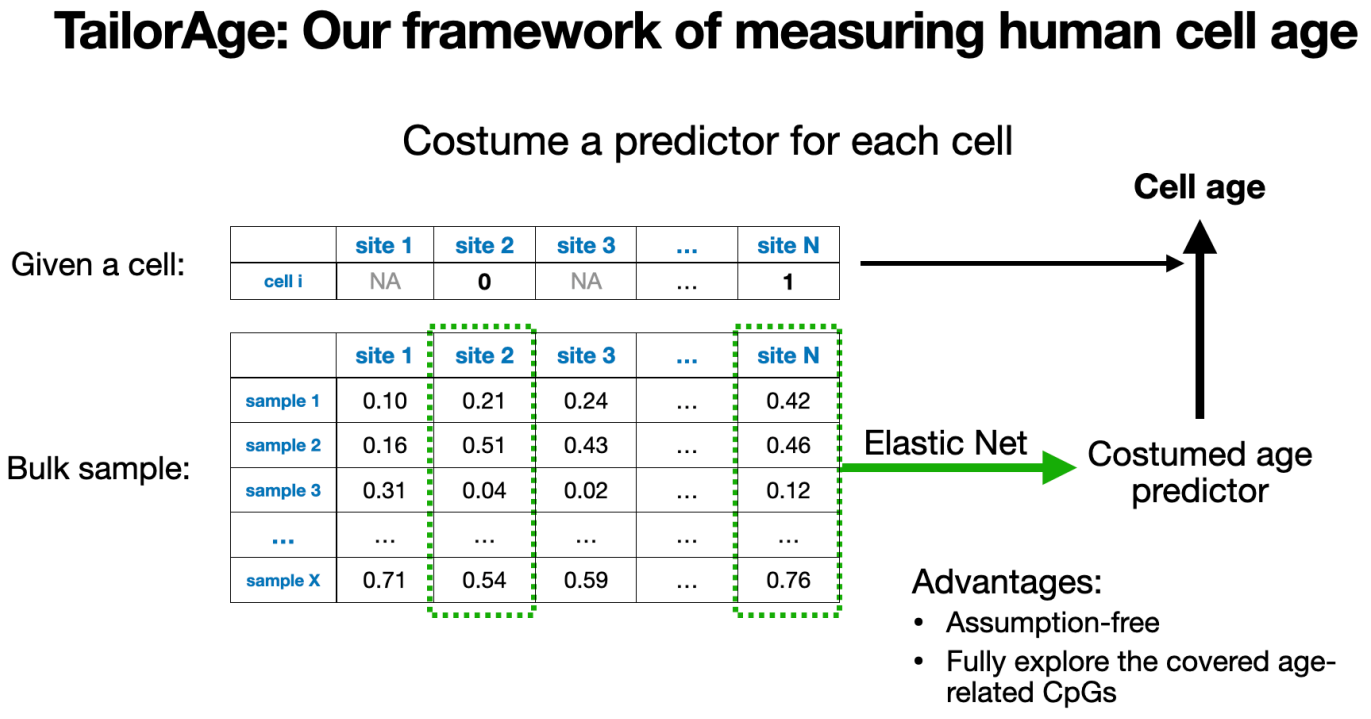


(Figure 1)

DNA methylation-based epigenetic clocks represent the most accurate molecular biomarkers of chronological and biological age. Standard clocks rely on unified matrices of high-density CpG sites derived from bulk tissue samples. However, profiling single-cell methylomes yields severe data sparsity, with over 95% of genomic CpG loci uncaptured per cell due to technical dropouts. Furthermore, the subset of captured CpGs varies randomly from cell to cell, rendering fixed-feature bulk regression models inapplicable. Consequently, robust single-cell epigenetic clocks explicitly calibrated for human scDNAm data remain a critical unfulfilled need in geroscience.

2. Methodology: The TailorAge Framework

To resolve feature non-overlap caused by random sparsity, TailorAge implements an adaptive, cell-specific training strategy that leverages bulk reference data without relying on artificial value imputation or cell-pooling (Figure 2).



(Figure 2)

For a given target single cell i :

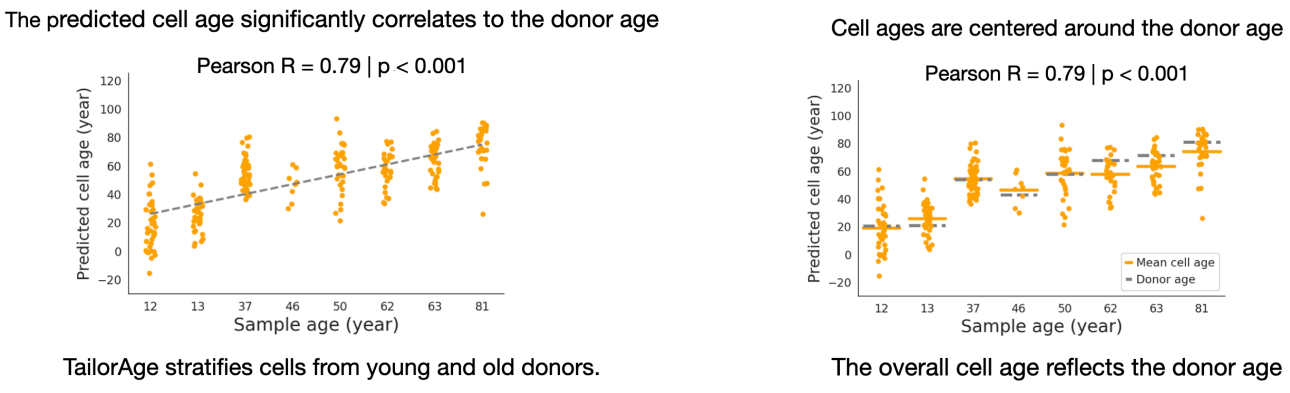
1. **Coverage Intersection:** The subset of genomic CpG sites successfully covered in cell i (denoted as $\mathcal{S}_i$) is identified.
2. **Sub-matrix Feature Extraction:** The reference bulk DNAm training dataset—comprising known chronological/BLUP ages—is filtered to retain only the CpG features present in $\mathcal{S}_i$.
3. **Cell-Specific Model Fitting:** An Elastic Net penalized regression model is trained on this feature-restricted bulk dataset. The joint L_1 (Lasso) and L_2 (Ridge) regularization holistically selects the most informative, non-redundant age-associated CpGs from $\mathcal{S}_i$:
4. $\hat{\beta}^{(i)} = \arg\min_{\beta} \left(\frac{1}{2N} \|y_{\text{bulk}} - X_{\text{bulk},\mathcal{S}_i}\beta\|_2^2 + \lambda \left[\alpha\|\beta\|_1 + \frac{1-\alpha}{2}\|\beta\|_2^2 \right] \right)$
5. **Single-Cell Age Prediction:** The trained model parameters $\hat{\beta}^{(i)}$ are applied directly to the methylation state of cell i to compute its single-cell biological age.

By tailoring the feature space per cell, TailorAge maximizes the extraction of age-informative signal from the unique genomic coordinates captured in every single-cell assay.

3. Results & Discussion

To evaluate the performance of TailorAge, we curated bulk training profiles from GEO and benchmarked the framework on human scDNAm data spanning 8 donors across a broad age range (12 to 81 years). Following stringent quality control filtering ($\geq 3.6 \times 10^6$ CpGs covered per cell), $n = 291$ high-quality cells were retained for analysis.

- **Macro-Scale Age Stratification:** Predicted single-cell biological ages exhibited a strong, highly significant linear correlation with donor chronological age ($R = 0.79, p < 0.001$; Figure 3). The central tendencies of predicted cell ages within each donor closely aligned with donor age, confirming that TailorAge reliably captures population-level biological aging trajectories.



(Figure 3)

- **Micro-Scale Epigenetic Dispersion:** Within individual donors, single-cell age predictions displayed prominent dispersion. Rather than reflecting prediction failure, this intra-sample variance captures underlying cell-to-cell epigenetic heterogeneity and stochastic drift that are routinely masked in bulk ensemble averages.

4. Future Directions & Project Roadmap

While these preliminary results demonstrate the feasibility of human single-cell epigenetic age quantification, further validation is required to disentangle biological signal from technical artifacts:

1. **Quantification of Technical Artifacts:** Deconstruct single-cell age variance using variance decomposition models to quantify the proportion of intra-donor variance attributable to technical coverage noise versus genuine biological heterogeneity.
2. **Experimental & Cross-Omics Validation:** Validate predicted single-cell biological ages against orthogonal single-cell modalities (e.g., scRNA-seq expression signatures, somatic mutation burden) and benchmarked *in vitro* or *in vivo* cellular senescent models.
3. **Methodological Benchmarking:** Evaluate predictive accuracy, noise resilience, and feature stability against established murine single-cell clocks (e.g., `scAge`) adapted for human datasets.

5. Reference

Bonder MJ, Clark SJ, Krueger F, Luo S, Agostinho de Sousa J, Hashtroud AM, Stubbs TM, Stark AK, Rulands S, Stegle O, Reik W, von Meyenn F. scEpiAge: an age predictor highlighting single-cell ageing heterogeneity in mouse blood. Nat Commun. 2024 Aug 31;15(1):7567. doi: 10.1038/s41467-024-51833-5. PMID: 39217176; PMCID: PMC11366017.

Elyahu Y, Monsonego A. Thymus involution sets the clock of the aging T-cell landscape: Implications for declined immunity and tissue repair. Ageing Res Rev. 2021 Jan;65:101231. doi: 10.1016/j.arr.2020.101231. Epub 2020 Nov 25. PMID: 33248315.

Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet. 2025 May;26(5):350-368. doi: 10.1038/s41576-024-00807-w. Epub 2025 Jan 13. PMID: 39806006.

Trapp A, Kerepesi C, Gladyshev VN. Profiling epigenetic age in single cells. Nat Aging. 2021 Dec;1(12):1189-1201. doi: 10.1038/s43587-021-00134-3. Epub 2021 Dec 9. PMID: 36211119; PMCID: PMC9536112.