Modeling Tissue-Specific Aging from Transcriptomic Data

Advances in molecular biology and machine learning have revealed organ-specific aging patterns. Oh et al. (2023) [1]used plasma proteomics to predict organ-specific biological ages, introducing age-gaps to quantify deviations from chronological age and model disease risk. However, plasma measurements capture only proteins released into circulation. To directly assess tissue-resident aging signatures, we leveraged tissue-specific transcriptomic data from actual organ samples to predict tissue-specific age.

We used the Adult GTEx v10 dataset [2], which provides transcript per million (TPM) and read count data across 54 tissues from 948 subjects, focusing on 12 tissues with adequate sample-size and strong correlation with mortality. Metadata include age, sex, and circumstance of death characterized by the Hardy Scale [3]. Since GTEx age data are binned, we refined labels by shifting bin midpoints toward sample means using exhaustive search. For each tissue, TPM data were split into 80-20 train-test set via age-stratified sampling, reserving Hardy Scale 1 samples for test set—representing sudden deaths—as outliers. Gene subsets were selected using three strategies: (1) genes with Pearson correlation >0.2 with age, (2) tissue-specific differentially expressed genes identified via optimal $\log_2$(fold-change) between age groups, and (3) tissue-enriched genes (expressed $\geq$4 times higher in one tissue than others).

Linear models using selected genes, age, and sex were trained with 20x bootstrapping. Nonlinear patterns were captured using a shallow neural network (two fully connected layers with batch normalization and ReLU). Predicted tissue ages were regressed against chronological ages to yield adjusted estimates $\hat{y}$ and age-gaps were defined as *(predicted age - $\hat{y}$)*. Given the limited dataset size, we employed leave-p-out cross-validation ($p = 5\%$) to ensure that the entire dataset was used for evaluation. For each tissue, age-gap distributions were modeled as Gaussian, and subjects were classified as **extreme negative agers** ($< \mu_{gap} - \frac{\sigma_{gap}}{2}$), **extreme positive agers** ($> \mu_{gap} + \frac{\sigma_{gap}}{2}$) or **average agers** (within $\pm\frac{\sigma_{gap}}{5}$). To assess clinical relevance, we compared conditional probabilities of prolonged disease (*dthhrdy* $\in \{3,4\}$) and unnatural deaths (*dthhrdy* = 1) across aging groups, where *dthhrdy* denotes a subject's Hardy Scale [3]. The overall analysis pipeline is illustrated in Figure 1.

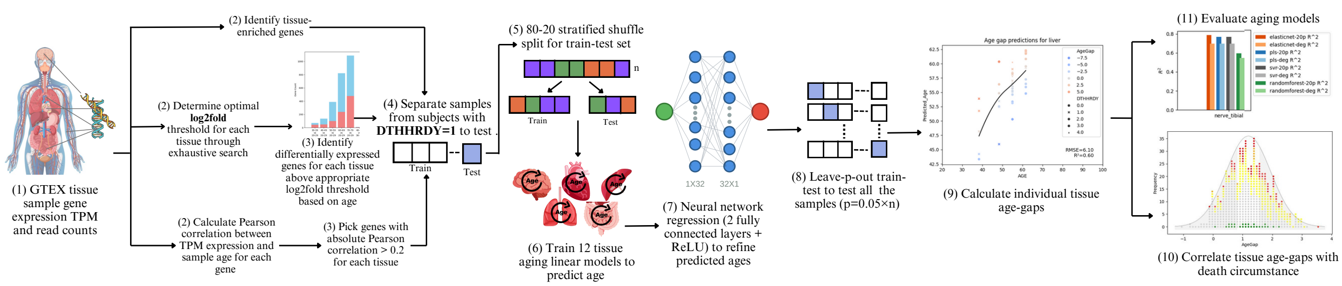

Figure 1: *Overview of the analysis pipeline: dataset preprocessing, feature selection, model training, age-gap computation, subject classification, and clinical relevance analysis.*

Across all feature selection strategies and models, correlation-based gene selection with elastic net achieved the best performance (RMSE = 6.44, R^2 across 1,285 test samples), while PLS regression without cross-validation ran the fastest. Analysis of age-gap distributions from 657 subjects revealed a low average pairwise correlation across tissues (Pearson *r* = 0.20), with extreme aging ($> |2\sigma|$) observed in 26% of all subjects—suggesting mostly organ-specific aging with only 1% showing multi-organ aging (extreme aging in 3+ tissues). Conditional probability analysis of Hardy Scale ratings indicated that *extreme positive agers* were nearly twice as likely to die from illness-related causes (*dthhrdy* = 3 or 4) compared to *extreme negative agers*. While *extreme negative agers* were more likely to have died from unnatural causes (*dthhrdy*=1), suggesting that accelerated biological aging aligns with disease-related mortality.

[1] Oh SH, Rutledge J, Nachun D, et al. Organ aging signatures in the plasma proteome track health and disease. Nature. 2023;624(7990):164-172.

[2] Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-585.

[3] National Center for Biotechnology Information. Hardy Scale Variable - phs000424.v4.p1. 2025.