Machine learning reveals tissue-agnostic and region-specific isoform aging markers in the human hippocampus

Xingyi Chen[1], Beril Erdogdu[2], Mihaela Pertea[2], Stephanie C. Hicks[2,3]

[1]Johns Hopkins University, Department of Applied Mathematics & Statistics, Whiting School of Engineering, Baltimore, MD
[2]Johns Hopkins University, Department of Biomedical Engineering, Whiting School of Engineering, Baltimore, MD
[3]Johns Hopkins University, Department of Biostatistics, Bloomberg School of Public Health, Baltimore, MD

Background

- Biological aging is associated with widespread transcriptional changes in the human brain.
- Previous work in dorsolateral prefrontal cortex (DLPFC) shows that isoform switching is predictive for biological age.
- Most studies focus on gene-level expression and regulation, overlooking isoform-level regulation.
- It remains unclear whether isoform-based features associated with aging generalize across brain regions.

Study Overview

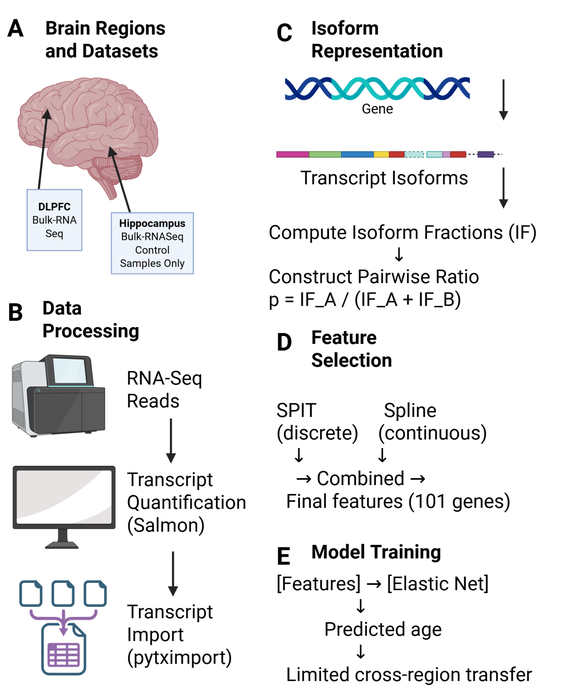


Figure 1: Overview of the isoform-based age prediction framework.
(A) Bulk RNA-seq from DLPFC and hippocampus.
(B) Transcript quantification (Salmon + pytximport).
(C) Isoform fractions and pairwise ratio features.
(D) Feature selection via SPIT (switches) and splines (trajectories).
(E) Elastic Net models for age prediction and cross-region evaluation.

DLPFC → Hippocampus Transfer Fails

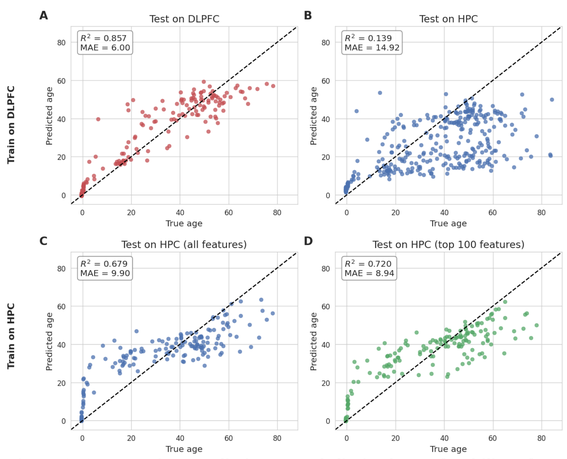


Figure 2. Cross-region prediction reveals limited transferability of DLPFC-derived isoform aging models.
(A) Models trained and tested on DLPFC show strong predictive performance.
(B) Direct application to hippocampus results in substantial performance loss and regression toward intermediate ages.
(C-D) Retraining on hippocampus using DLPFC-derived features partially improves performance, but predictions remain biased and do not recover full accuracy.

Feature Selection Strategy

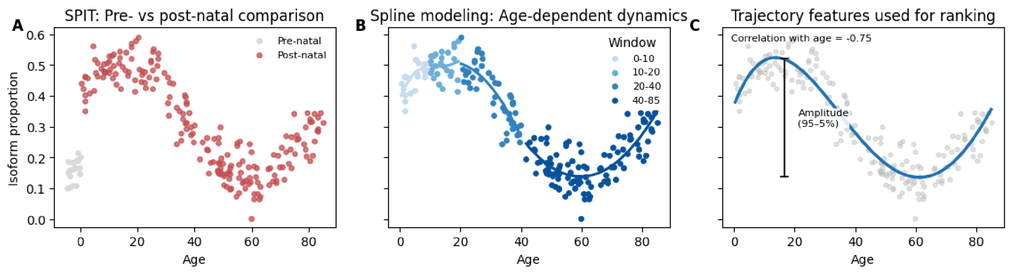
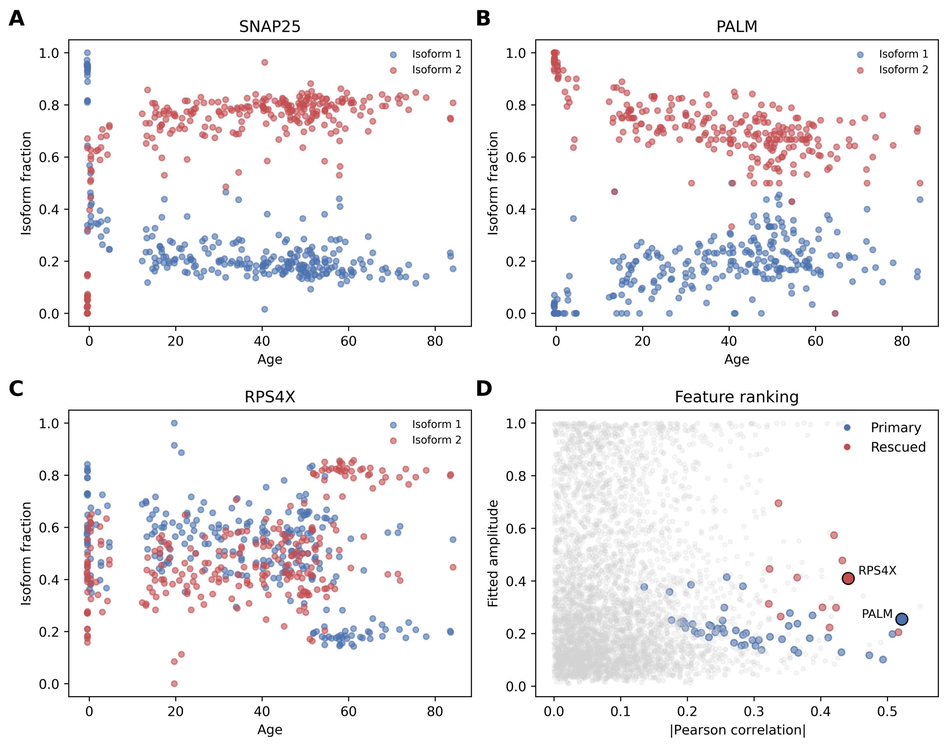


Figure 3: Conceptual schematic of complementary feature selection strategies for age-associated isoform dynamics.
(A) SPIT identifies discrete differential transcript usage by contrasting pre- and post-natal samples, capturing switch-like isoform changes.
(B) Spline-based modeling (postnatal samples only) captures continuous and potentially nonlinear isoform trajectories across age by fitting smooth functions within age windows.
(C) From spline fits, trajectory-derived features - such as correlation with age and robust amplitude (95^{th} - 5^{th} percentile difference) - are extracted to quantify the strength of age-associated dynamics for downstream ranking.
Together, these approaches capture complementary classes of isoform behavior, including abrupt developmental switches and gradual or localized postnatal changes.



Distinct isoform dynamics - switch-like, gradual, and localized - motivate complementary feature selection.
Figure 4: Distinct classes of age-associated isoform dynamics in the hippocampus.
(A) SNAP25 shows a canonical pre- to post-natal isoform switch identified by SPIT.
(B) PALM exhibits a gradual postnatal trajectory captured by spline-based modeling.
(C) RPS4X displays localized late-life shifts recovered by the rescue procedure.
(D) Feature ranking based on correlation and trajectory amplitude highlights both primary and rescued candidates.
Together, these results demonstrate that isoform dynamics span switch-like, continuous, and localized patterns.

Feature Overlap across Methods and Regions

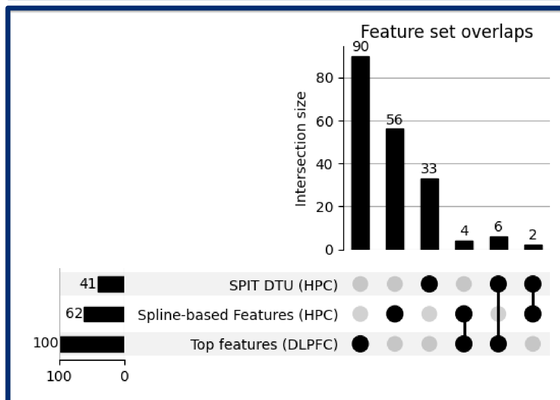
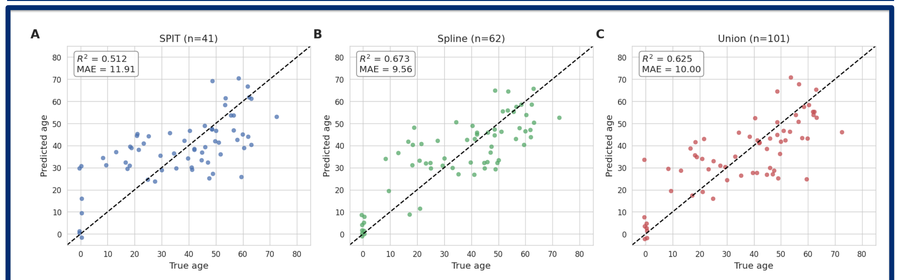


Figure 5: Limited overlap between feature sets highlights distinct isoform aging signals.
UpSet plot showing overlap between SPIT-selected, spline-derived, and DLPFC-derived features. Most features are unique to a single set, indicating that different methods and regions capture complementary isoform dynamics.

Spline-derived features improve hippocampus age prediction



Spline features outperform SPIT, no improvement when combining feature sets.
Figure 6: Predictive performance of hippocampus-derived isoform feature sets.
(A) Models using SPIT-selected features show moderate performance.
(B) Spline-derived features achieve the strongest predictive accuracy.
(C) Combining feature sets does not further improve performance.
Together, these results indicate that continuous isoform dynamics dominate age prediction in the hippocampus.
n denotes the number of genes (features) included in each model.

Hippocampus-derived isoform features show limited transferability to DLPFC

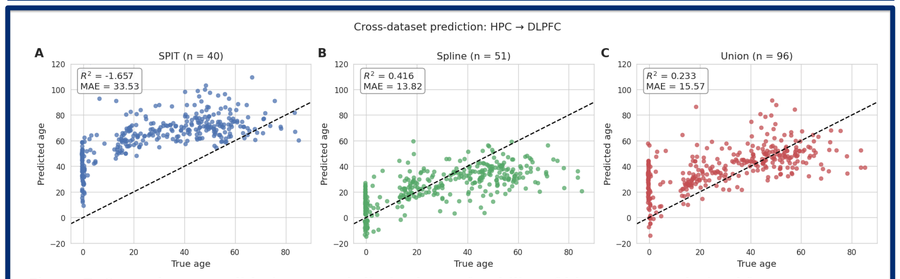


Figure 7: Cross-dataset validation reveals limited transferability of hippocampus-derived features.
(A) Models trained on hippocampus using SPIT-selected features perform poorly when applied to DLPFC.
(B) Spline-derived features show partial recovery but remain biased and compressed, with significant regression to the mean effect.
(C) Combining feature sets does not improve performance.
These results indicate that isoform aging signals are largely region-specific.

Conclusions

- Isoform dynamics in the human hippocampus exhibit distinct patterns (switch-like, gradual, localized), motivating complementary feature selection strategies.
- Spline-derived features capture continuous postnatal isoform variation and improve age prediction in the hippocampus.
- Isoform-derived features show limited bidirectional transferability between hippocampus and DLPFC, indicating strong region-specific aging signal.

Future Work

- Perform biological interpretation of isoform features to understand region-specific regulatory mechanisms in hippocampus and DLPFC.
- Evaluate feature robustness across independent datasets and brain regions to assess generalizability.
- Explore alternative modeling frameworks (e.g. non-linear methods) to capture complex isoform-age relationships.

References

Erdogdu B, Ji HJ, Rudnick ZC, Pertea M, Salzberg SL. *Isoswitching drives the aging process in human brains.* bioRxiv, 2025.
Collado-Torres L et al. *Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia.* Neuron, 2019.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017.
Kuehl M, Wong MN, Wanner N, Bonn S, Puelles VG. *Gene count estimation with pytximport enables reproducible analysis of bulk RNA sequencing data in Python.* Bioinformatics, 2024.

Acknowledgements

Funding for this study was provided by NIH/NIGMS R35GM150671 (SCH). We would like to acknowledge the families of the brain donors who made this research possible. The authors declare that they have no conflict of interest. We thank the maintainers of the Joint High Performance Computing Exchange (JHPCE) compute cluster at Johns Hopkins Bloomberg School of Public Health for providing essential computing resources. We also thank Hicks lab for their constructive comments and suggestions. Figures were made with the help of Biorender.