Integrating Epigenetic Clocks via Deep Neural Networks

Yahui Zhang and Peter Song (pxsong@umich.edu)

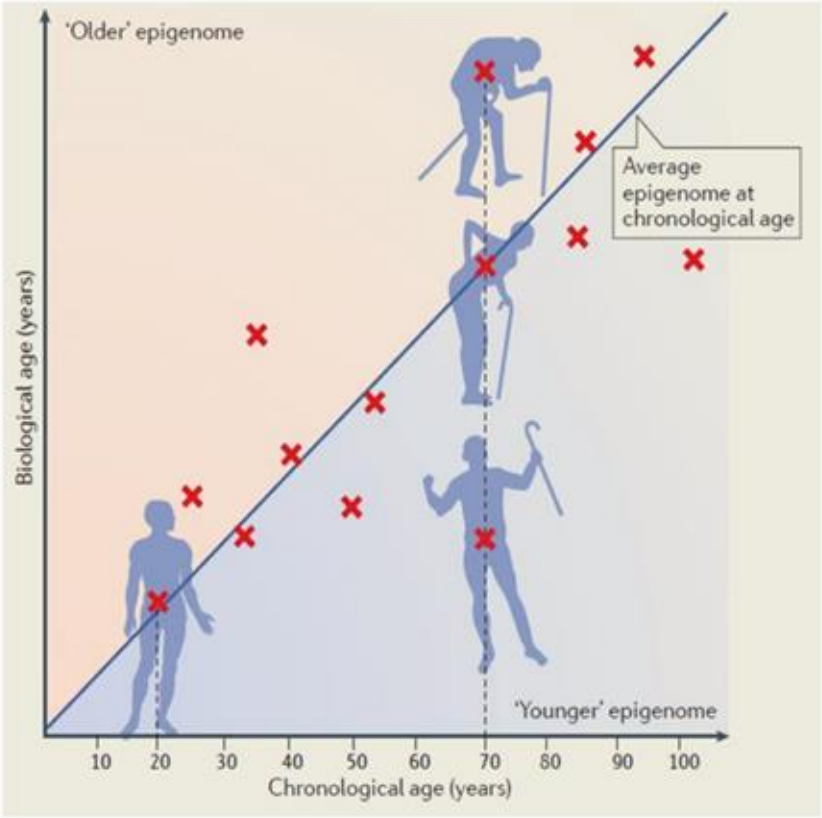
School of Public Health, University of Michigan

June 24, 2026

The Institute for Mathematical and Statistical Innovation

NIH R01ES033656, NIH R24ES028502

Aging is invertible & heterogeneous



Scientific Questions

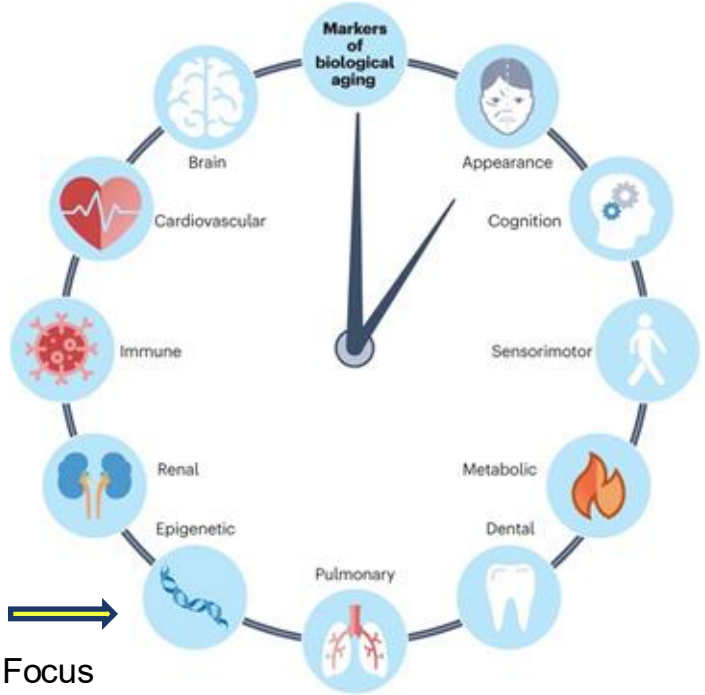
- What causes aging?
- How to quantify aging and aging variability/heterogeneity?
- How to slow down aging?
- How to have successful aging?

Chronological age vs biological age

Chronological age

Time since birth: Ticks away with the course of time and is based on the day you were born

Biological age

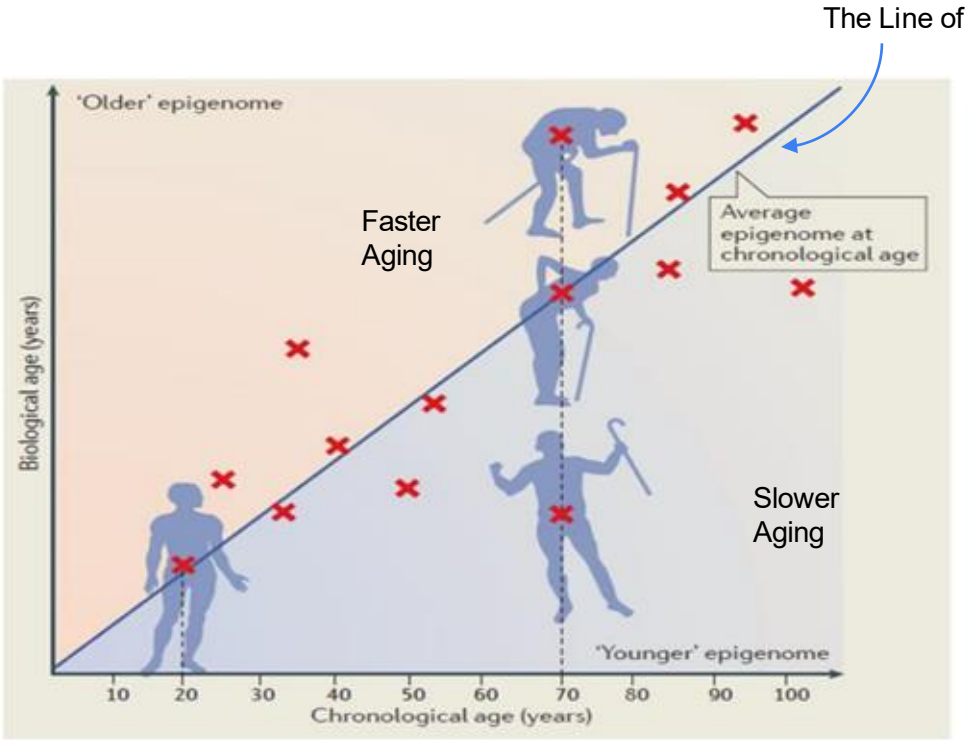


Adapted from M. L. Elliott et al. *Nat. Aging* **1**, 295–308 (2021), Springer Nature

What is Age Acceleration?

Fact: Individuals with the same chronological age can age at very different biological rates.

BA < CA ⇒ **slower aging** vs BA > CA ⇒ **faster aging**



Discovery of Telomere Shrinkage

The Nobel Prize in Physiology or Medicine 2009

Photo: U. Montan

Elizabeth H. Blackburn

Prize share: 1/3

Photo: U. Montan

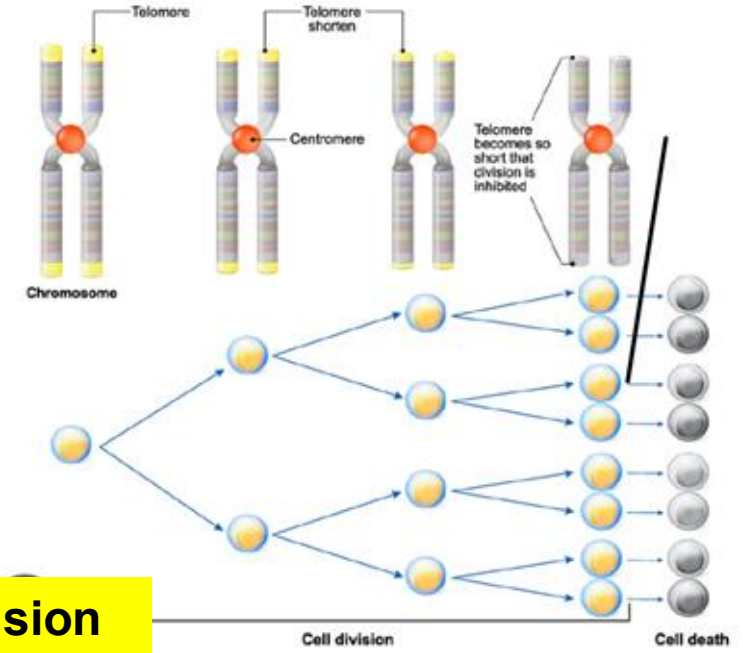
Carol W. Greider

Prize share: 1/3

Photo: U. Montan

Jack W. Szostak

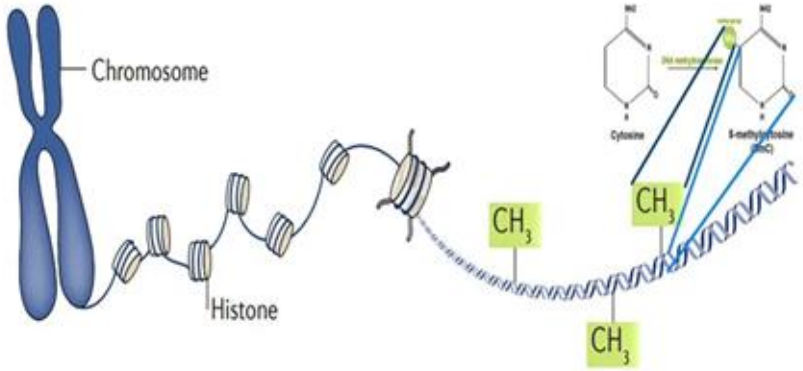
Prize share: 1/3



Telomere shrinks in the cycle of cell division

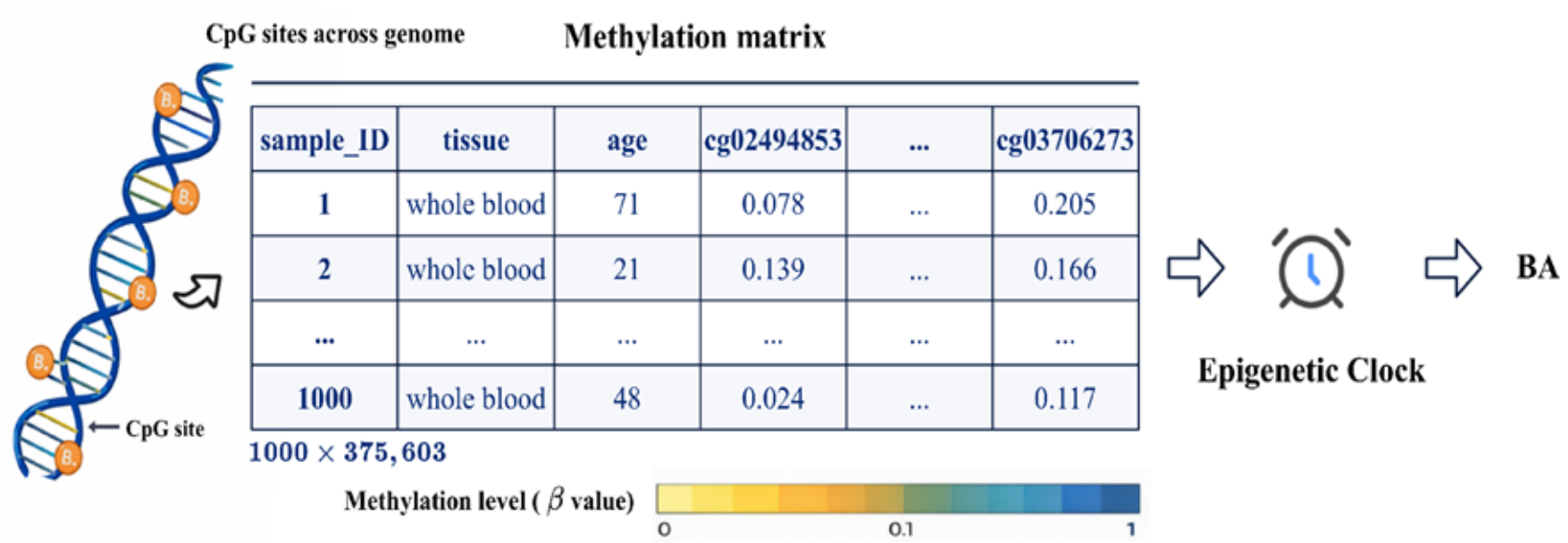
What is epigenetic clock?

- Biological clock measures biological age via different biomarkers in **multiple organ systems.**
- **Epigenetic clock:** a specific type of biological clock, relies on changes in DNA methylation (DNAm) as a marker of cellular evolution and reproduction, which may be influenced by social and environmental determinants.



Quantify Biological Age and Age Acceleration

- **DNA methylation:** Illumina Array 450K or EPIC (850K) measures 450 or 850 CpG sites for one person, resulting in **high-dimensional features** $X \in \mathbb{R}^p$.
- **Epigenetic clock:** age prediction model, $BA = g(X) + \varepsilon$ via the elastic-net penalized regression.



sample_ID	tissue	age	cg02494853	...	cg03706273
1	whole blood	71	0.078	...	0.205
2	whole blood	21	0.139	...	0.166
...	...	...	...	...	...
1000	whole blood	48	0.024	...	0.117

Hundreds of Epigenetic Clocks Have Been Proposed

Clock	Reference	Tissue	# CpGs	Method	Description
Horvath	Horvath (2013)	Multi-tissue	353	Elastic net	Chronological age
Hannum	Hannum et al. (2013)	Whole blood	71	Elastic net	Chronological age
Zhang (Mortality)	Zhang et al. (2017)	Peripheral blood	10	Elastic net	Mortality risk
PhenoAge	Levine et al. (2018)	Blood	513	Elastic net	Phenotypic age
Skin & Blood	Horvath et al. (2018)	Skin, blood	391	Elastic net	Chronological age
GrimAge	Lu et al. (2019)	Blood	1,030	Elastic net	Mortality risk
DNAmTL	Lu et al. (2019)	Blood	140	Elastic net	Telomere length proxy
PedBE	McEwen et al. (2020)	Buccal	94	Elastic net	Pediatric age
Zhang (EN)	Zhang et al. (2019)	Blood + saliva	514	Elastic net	Chronological age
DNAm IC	Fuentealba et al. (2025)	Blood	91	Elastic net	Intrinsic capacity

Each clock is trained on DNAm from different cell types/tissues/cohorts, resulting in feature subsets for a prediction model.

Original training datasets are not available but selected CpG sites and loading coefficients are published.

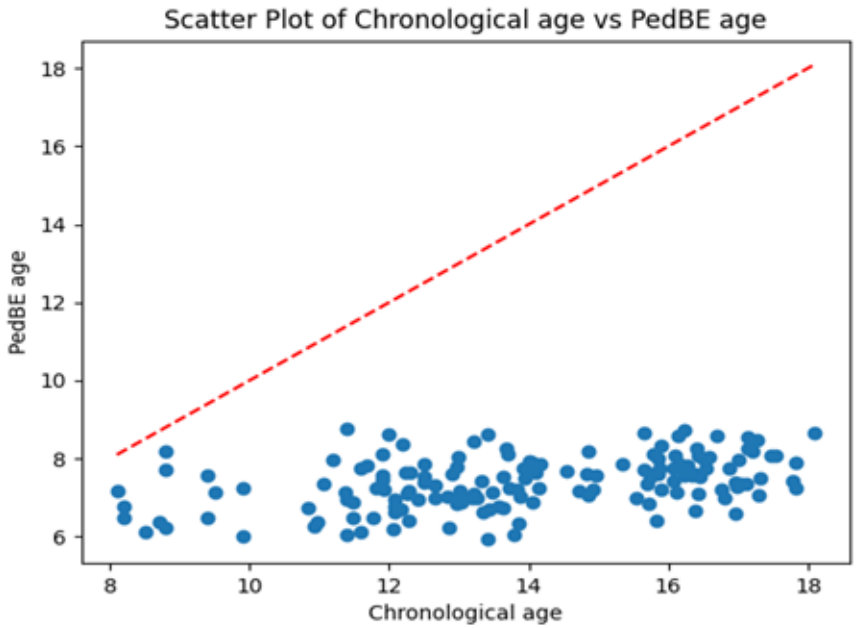
Challenge: Too many clocks (prediction models) established with various samples and cohorts!

Question: Which one(s) should be used? Fishing?!

An Example: PedBE clock (McEwen et al., 2020)

ELEMENT Cohort:

- **601 children in adolescent period**
- **8-18 years**

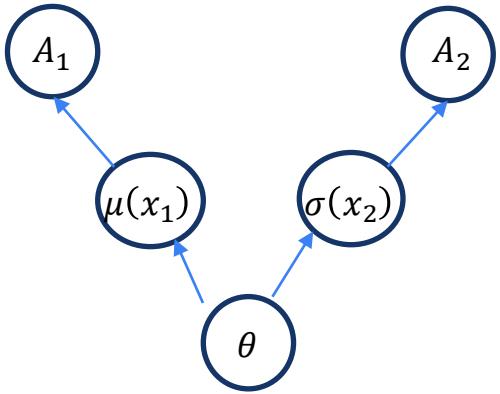


- **Lack of transportability**
- Need a “better” clock that is transportable, and little dependent on used training data
- Perform certain calibration on a clock towards target data under investigation
- Balance transportability and adaptiveness

Meta-Clock: Transportability and Adaptiveness

Objective: Integrate existing clocks with new target data to yield a more accurate, robust, adaptive and transportable "meta-clock".

Consider two clocks:



$$\hat{\theta} = \mathrm{g}(\mu(x_1), \sigma(x_2)]$$

Pool multiple data sources

Transportability

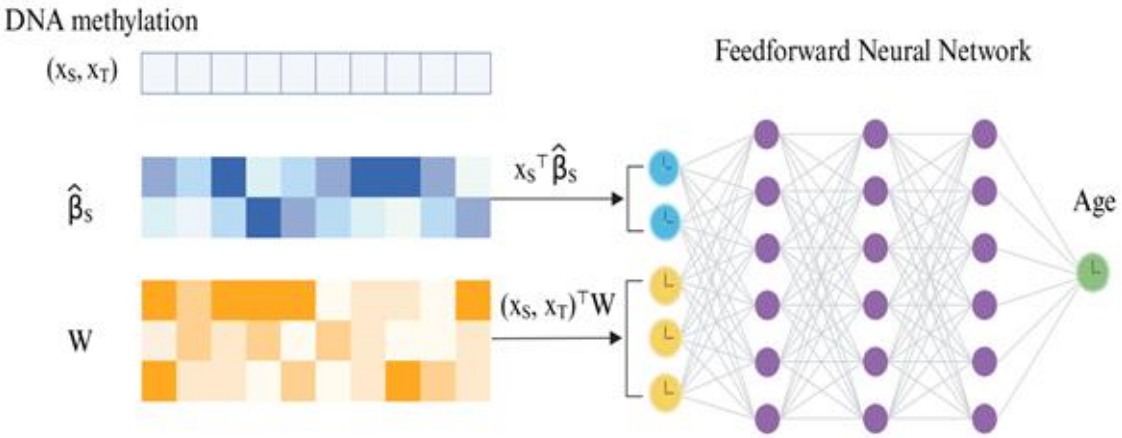
- M existing clocks are chosen to derive a meta-clock
- Pool distinct CpG sites from individual clocks to form a set of CpG sites, denoted by $\mathcal{P}_S$

Adaptiveness

- Perform univariate screening by to select top hits with highest absolute Pearson correlations between chronological age and CpG sites.
- Form another set of CpG sites, denoted by $\mathcal{P}_T$

- Combine the two sets of CpG sites: $\mathcal{P}_s \cup \mathcal{P}_T$
- $\mathcal{A} = \{A^{(m)}, m=1,\ldots, M\}$ is the set of predicted ages from the M existing clocks for the target data.

Neural Network Machinery



$X_S^T \widehat{\beta_S}$ is the set of predicted ages $A^{(m)}$ by existing clocks

$(X_S, X_T)^T W$ is the set of predicted ages by local clocks, **where L1-norm penalty is imposed on W for sparsity**

- Retain the structure of the established and well-validated models
- Generates new local clocks using new DNAm data with dimensionality reduction
- Let NN think and output integration, allowing for interplays across different clocks
- Enable multi-channel convolutions of existing and new clocks, superior over the conventional model averaging.

Feedforward ReLU network

- Let $L \in \mathbb{N}$ and width vector $\boldsymbol{d} = (d_0, \ldots, d_{L+1})$ with $d_0 = q + k$ and $d_{L+1} = 1$.
- A deep ReLU network $g : \mathbb{R}^{d_0} \to \mathbb{R}$ is defined as

$$g(\mathbf{x}) = \mathcal{L}_{L+1} \circ \bar{\sigma} \circ \mathcal{L}_L \circ \cdots \circ \bar{\sigma} \circ \mathcal{L}_1(\mathbf{x}),$$

where $\mathcal{L}_\ell(\boldsymbol{x}) = \boldsymbol{U}_\ell \boldsymbol{x} + \boldsymbol{b}_\ell$ is an affine transformation with weight matrix $\boldsymbol{U}_\ell \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$ and bias vector $\boldsymbol{b}_\ell \in \mathbb{R}^{d_\ell}$, $\ell = 1, \ldots, L+1$, and $\bar{\sigma}$ denotes the element-wise ReLU activation function $\bar{\sigma}(\cdot) = \max\{\cdot, 0\}$.

FFN-regular family

- The class of the bounded ReLU family with upper bound $B_1 > 0$:

$$\mathcal{G}(L, \boldsymbol{d}, B_1) = \left\{ g : \ \|\mathbf{U}_\ell\|_\infty \le B_1, \ \|\mathbf{b}_\ell\|_\infty \le B_1, \ \ell = 1, \ldots, L+1 \right\}.$$

- A common specification is $\mathcal{G}(L, q + k, 1, N, B_1)$, with hidden layers of width N.

Meta-clock via integration of predictions

Meta-epigenetic clock (with existing clocks)

- Input for subject i:

$$[\,\mathbf{A}_i,\ \mathbf{z}_i\,] \in \mathbb{R}^{q+k}, \qquad \mathbf{z}_i = \begin{pmatrix} \mathbf{x}_{S,i} \\ \mathbf{x}_{T,i} \end{pmatrix}^{\top} \mathbf{W}.$$

transportability

Adaptiveness

- We jointly estimate the sparse projection $\mathbf{W}$ and the deep ReLU predictor g by minimizing the penalized least-squares risk:

$$\min_{\mathbf{W},\, g \in \mathcal{G}(L,q+k,1,N,B_1)} \frac{1}{n} \sum_{i=1}^{n} \Big\{ y_i - g\big([\mathbf{A}_i,\ \mathbf{z}_i]\big) \Big\}^2 + \lambda \, \|\mathbf{W}\|_1,$$

Optimally mix the two by NN

where $\|\mathbf{W}\|_1 = \sum_{j=1}^{p_S+p_T} \sum_{h=1}^{k} |W_{j,h}|$, and $\lambda \geq 0$ controls the sparsity level.

- The resulting *meta-epigenetic clock* is

$$\widehat{m}_{\text{meta}}(\mathbf{x}) = \widehat{g}\Big(\big[\mathbf{A}_i,\ (\mathbf{x}_{S,i}, \mathbf{x}_{T,i})^{\top} \widehat{\mathbf{W}}\big] \Big).$$

Local clocks with no transportability

Locally adaptive clock (without existing clocks)

- To assess the contribution of estimated coefficients from existing clocks, we remove $\mathbf{A}_i$ from the input:

$$\widehat{m}_{\text{local}}(\mathbf{x}) = \widehat{g}_{\text{local}}\left((\mathbf{x}_{S,i}, \mathbf{x}_{T,i})^\top \widehat{\mathbf{W}}_{\text{local}}\right),$$

with $\widehat{\mathbf{W}}_{\text{local}}$ and $\widehat{g}_{\text{local}}$ obtained by solving the same penalized problem restricted to DNAm inputs.

Theoretical Guarantees

- Loss: $\ell(y, f(x)) = \{y - f(x)\}^2$
- Population risk: R(f) = $E_{(y,X)}\{\ell(y, f(x))\}$
- Empirical risk: $\mathbb{R}(f) = \frac{1}{n}\Sigma_{i=1}^{n}\ell(y_i, f(x_i))$
- Estimand: $f^*(x) = E(y|X = x)$
- Predictor: f(x) = g($\psi_W^A(x)$) $= g \circ \psi_W^A(x)$ with $\psi_W^A(x) = [A(x), x^T W]$ and g$\in \mathcal{G}$
- Class of regular composite functions
- $\mathcal{F} = \mathcal{F}(L, d, B_1, B_2) =$

$$\{g \circ \psi_W^A(x), g \in \mathcal{G}\ (L, d, B_1), \|A\|_\infty \leq B_2, \|W\|_1 \leq B_2, \|W\|_\infty \leq B_2\}$$

- Uniform empirical covering number $\mathcal{N}_{2n}(\delta, \mathcal{H})$ = $sup_x\ \mathcal{N}_{2n}(\delta, \|.\|_\infty, \mathcal{H}|x)$

Main Theorem

Assumption: Model: y = $f^*(x) + \varepsilon$, ε is sub-Gaussian.

Let $(\hat{g}, \widehat{W})$ be the estimator optimizing the following objective:

$$min_{W,g\in\mathcal{F}} \frac{1}{n}\Sigma_{i=1}^{n}\{y_i - g([A_i, x_i{}^T W])\}^2 + \lambda \|W\|_1.$$

Suppose λ satisfies $\lambda \simeq \frac{1}{n}\log \mathcal{N}_{2n}(\delta, \mathcal{F})$. We have the following bound:

$$\mathrm{E}\{\mathrm{R}(\hat{g} \circ \psi_{\widehat{W}}^{A}(x)) - R(f^*)\} \leq C[\inf_{f\in\mathcal{F}}\{R(f) - R(f^*)\} + \frac{1}{n}\log \mathcal{N}_{2n}(\delta, \mathcal{F})]$$

where C is a constant independent of n.

Hölder function class

- Suppose f^* from a Hölder function class $\mathcal{H}_\beta([0,1]^d, B_o)$.

$$\mathrm{E}\{\mathrm{R}(\hat{g} \circ \psi_{\hat{W}}^{A}(x)) - R(f^*)\} \leq C_1[(NL)^{-2\beta/d} + \frac{\log n}{n} PL \log P]$$

P: the total number of parameters

L: Depth

N: Width

- Suppose $\mathrm{P} \simeq n^{\frac{d}{2\beta+d}}$, $\mathrm{L} \simeq \log n$, $\mathrm{N} \simeq \frac{n^{\frac{d}{2\beta+d}}}{\log n}$

$$\mathrm{E}\{\mathrm{R}(\hat{g} \circ \psi_{\hat{W}}^{A}(x)) - R(f^*)\} \leq C_2 n^{-2\beta/(2\beta+d)} (\log n)^{\kappa}, \qquad \kappa > 0.$$

Competing Models: Training with Transportability of Existing Clocks

Model	Name	Info from existing clocks	Target data	Model Training
S1	Off-the-shelf	Existing model/clock	No	No
S2	Target-only	None	Top-hit CpGs	EN or NN
S3	Feature-augmented	Selected CpGs in Clocks	Top-hit CpGs	EN or NN
S4	Coefficient-augmented	Coefficients (Loadings) in Clocks	Top-hit CpGs	Transfer Learning with EN
S5	Prediction-augmented	Predicted ages by Clocks	Top-hit CpGs	EN or NN

Application

- 601 children from Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) cohort aged 8-18 who experience fast growth and development (height, sexual maturation, cognitive functions), a critical time window for biological aging.

- **Outcome data:** Demographic data, including chronological age

- **DNAm array data:** 850K DNA methylation data
  - Apply univariate pre-screening by Pearson correlation to select the top-hits of 500 CpG sites.
  - Yield 1,095 unique CpGs in the combination.

- **Data Split:** $N_{\text{train}} = 481$, $N_{\text{valid}} = 60$, and $N_{\text{test}} = 60$, repeat 20 times

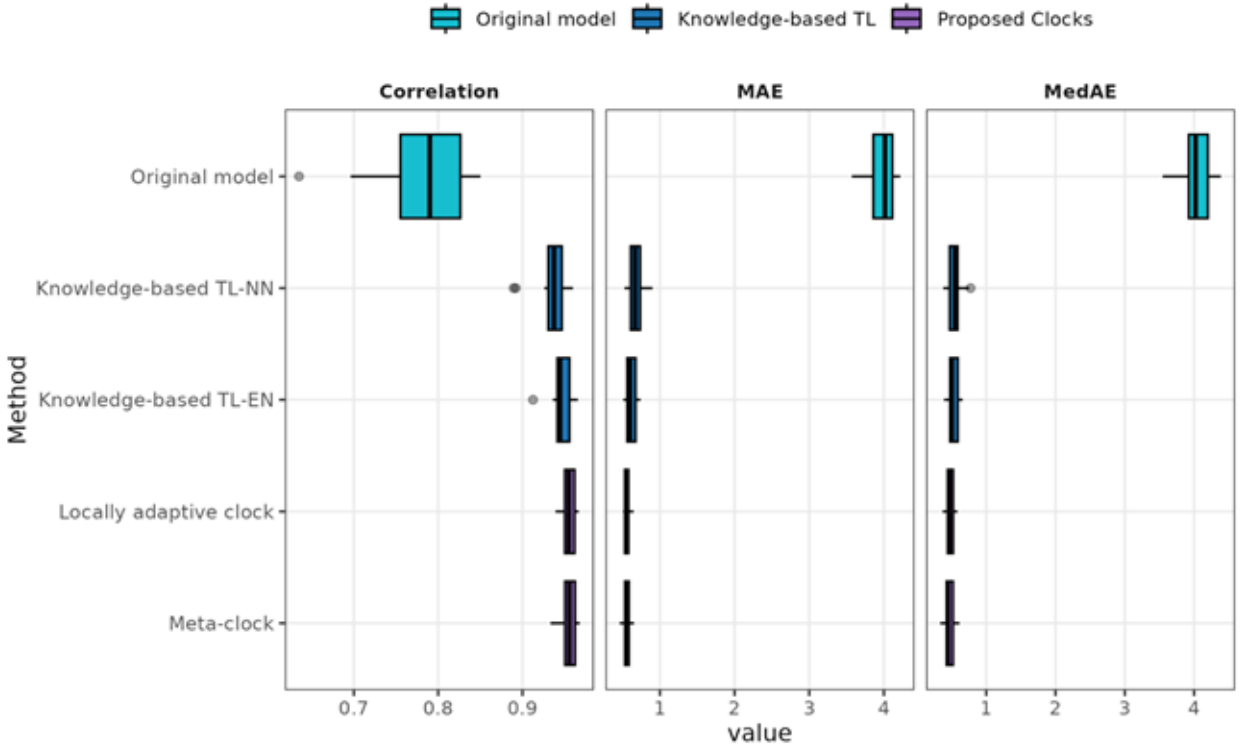
Selection of existing clocks to create a meta-clock

- Two clocks with correlations of 0.8 or bigger between the predicted and chronological age

- Horvath's Skin & Blood clock: trained on DNAm from dermal and hematopoietic tissues and widely applied to whole-blood DNAm.
  - It has the **best** predictive performance among many existing clocks and serves as the off-the-shelf model

- PAYA clock: Developed for adolescents and young adults

Results: Comparison with the best off-the-shelf clock



Mean absolute error (MAE), median absolute error (MedAE), and correlation between predicted and observed ages.

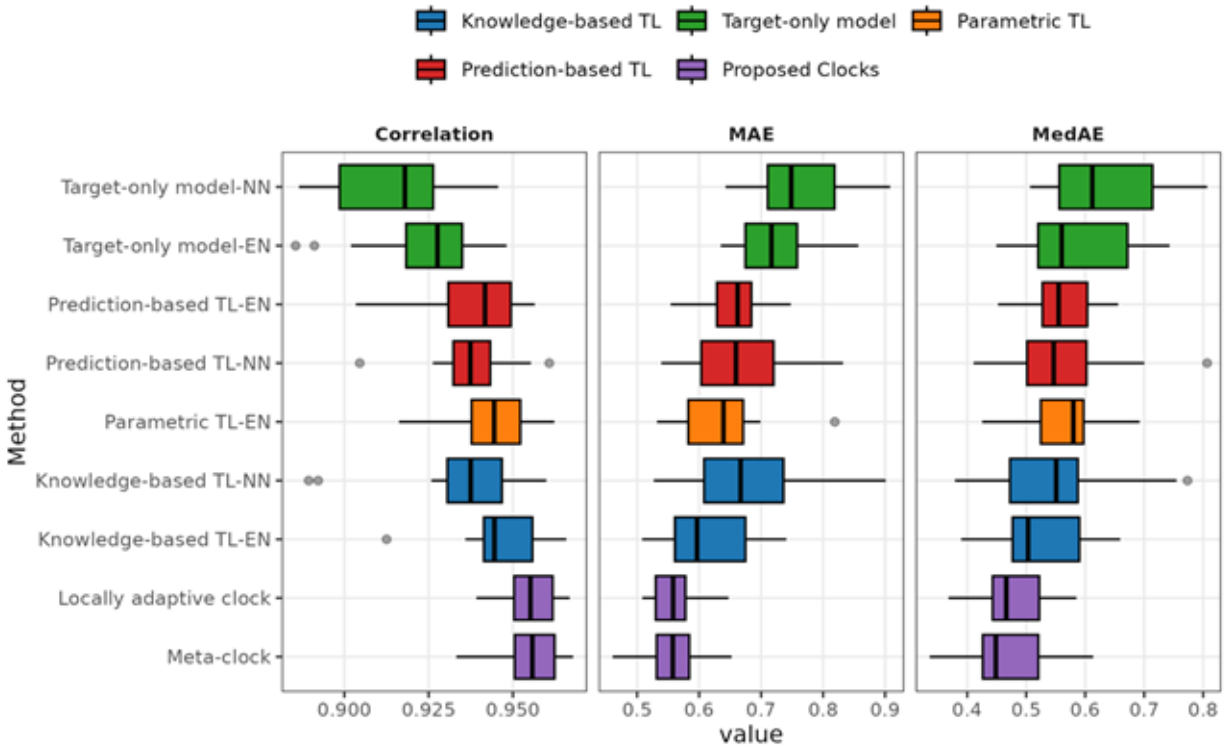
S1: the Horvath Clock

S3: Combining features

Local clock: NN with local features

Meta-clock: NN

Results



Focus on:

- With or with transportability
- EN vs NN

S2: training with local features

S5: combining predicted ages
EN vs NN

S4: EN with TL on coefficients

S3: combining features
EN vs NN

Local clock: NN with local features

Meta-clock: NN

Mean absolute error (MAE), median absolute error (MedAE), and correlation between predicted and observed ages on the test data.

Take-home messages

Incorporating information from existing clocks helps improve prediction.

Information from exiting clocks is more important than the choice of model type between EN (linear) and NN (nonlinear).

NN-based integration framework allows to identify new biological features relative to existing ones, leading to new insights and knowledge in the understanding of epigenetic aging mechanism.

Learning from the NN Model: Derived local clocks

- Low redundancy
  - Average Jaccard similarity between latent nodes = 0.10 (maximum = 0.16), indicating minimal feature overlap.

- Near-orthogonality among latent dimensions
  - Mean absolute off-diagonal cosine similarity = 0.12 (maximum = 0.26), suggesting that the latent representations are nearly orthogonal and capture distinct feature subspaces.

Biological Relevance

The latent nodes (local clocks) recapitulates the dual architecture of epigenetic aging:

Promoter/Polycomb-linked sites tend to gain methylation with age, whereas distal regulatory regions tend to lose it.

Table 2: Characterizations of five new local clocks, their influence on age, biological plausibility, as well as their representative CpG sites, including IDs and coefficient signs.

Local clock	Size (CpGs)	Influence on Age	Biological Plausibility	Leading CpG's ID (sign)
1	80	**Age-HYPER; Pro-age activation** (↑ with age)	Promoter/CGI CpGs (like Polycomb/PRC2 targets) frequently hypermethylate with age, giving a pro-age signal.	cg04105250 (–), cg00481951 (+), cg19785559 (+), cg01256539 (–), cg09173378 (–)
2	94	**Age-HYPER; Pro-age activation** (↑ with age)	Similar profile to Node 1; promoter/CGI CpGs typically hypermethylate with age, yielding large positive contributions.	cg09364988 (–), cg01949324 (–), cg09278098 (–), cg21572722 (+), cg10612237 (+)
3	82	**Mixed; Composite**	Mixture of + (CGI-like) and – (open-sea-like) CpGs.	cg17238334 (–), cg03771840 (+), cg21572722 (+), cg04055490 (–), cg07268926 (+)
4	101	**Age-HYPO; Anti-age activation** (↓ with age)	Open-sea/enhancer CpGs tend to lose methylation with age (global hypomethylation / enhancer drift).	cg12460140 (–), cg21055197 (–), cg03565081 (+), cg10959651 (–), cg24348981 (+)
5	85	**Weak/neutral; Minor**	Small, mixed-sign weights without a clear age trend; likely neutral/background relative to 1–4.	cg13108341 (+), cg25345365 (+), cg27503801 (+), cg05093811 (–), cg02369313 (+)

Concluding Remarks

- **Importance of prior info:** Integrate information from existing clocks help build robust biological age clocks, the meta-clock is shown consistently as a top-performer across diverse simulation settings and real-world empirical studies.

- **Yield of local clocks:** New local clocks are yielded that provide new biological insights on the relationship between DNAm and aging beyond what existing clocks cover.

- **Small-sample regimes:** When n is not large, simple transfer strategies, such as feature pooling and elastic net, are recommended as there is no clear gain of prediction accuracy from the choice of nonlinearity.

- **Methodological flexibility:** The meta-clock framework is applicable to second-generation clocks with other aging-related outcomes such as healthspan and mortality.

- **Latent dimension selection:** Predictive accuracy improves as the number of latent nodes/clocks (k) increases and then plateaus. A future work on the choice of k is needed.

Thank for Your Attention!

Questions?

Elastic-net penalized regression for an epigenetic clock

Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ be a dataset of n independently and identically distributed (i.i.d.) observations, where $\boldsymbol{x}_i \in \mathbb{R}^p$ is a p-dimensional vector containing the DNAm levels at p different CpG sites and $y_i \in \mathbb{R}_+$ is chronological age.

The regression is fitted via an elastic-net (EN) penalty by minimizing the penalized least-squares objective:

$$(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}) = \arg\min_{\beta_0, \boldsymbol{\beta}} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda_{EN} \left\{\alpha_{EN} \|\boldsymbol{\beta}\|_1 + (1 - \alpha_{EN}) \|\boldsymbol{\beta}\|_2^2 / 2\right\}.$$

- $\alpha_{EN} \in [0, 1]$: mixing parameter controlling the balance between ℓ_1 (lasso) and ℓ_2 (ridge) penalties
- $\lambda_{EN} \geq 0$: global regularization parameter that governs the overall sparsity in the estimated coefficients
- $\mathcal{L} = \text{supp}(\widehat{\boldsymbol{\beta}})$: the set of CpG sites with detected signals (the support of the estimated coefficients)
- $A = \hat{\beta}_0 + \boldsymbol{x}_{\mathcal{L}}^\top \hat{\boldsymbol{\beta}}_{\mathcal{L}}$: the (predicted) epigenetic age