Accurate age prediction from blood using a small set of DNA methylation sites

Miri Varshavsky[1], Gil Harari[1,2], Bracha Lea Ohana[2], Benjamin Glaser[3], Yuval Dor[2], Ruth Shemer[2], and Tommy Kaplan[1,2]

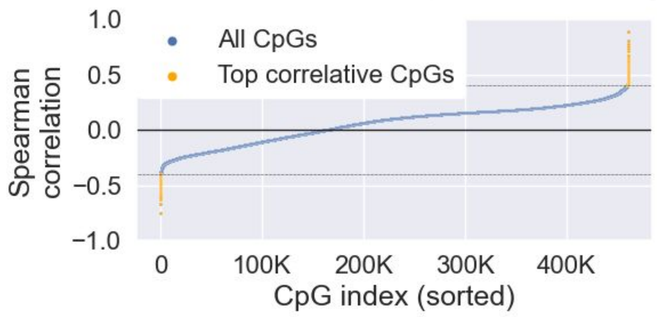
[1] School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel
[2] Faculty of Medicine, The Hebrew University of Jerusalem, Israel
[3] Dept. of Endocrinology and Metabolism, Hadassah Medical Center, Jerusalem, Israel

1 Introduction

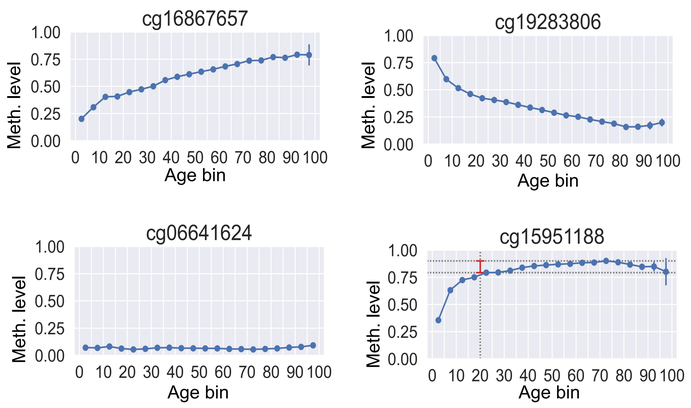
- Age prediction from methylation sheds light on the aging process, with indications of poor health and predicted lifespan
- We present **GP-age**, a non-linear epigenetic clock which uses only 30 CpG sites, applicable to simple blood tests
- Large cohort of 11,910 blood-derived methylomes
- Median error (MedAE) of ~2 years on held-out data

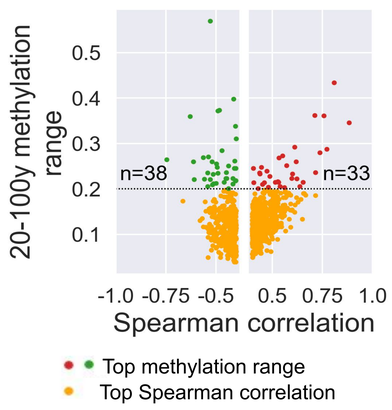
2 Selection of 30 CpG sites

Correlation of Illumina 450K CpG sites with age



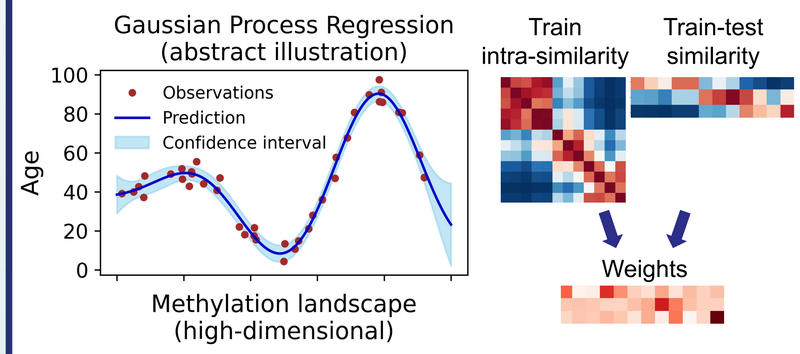
Methylation range of Illumina 450K CpG sites





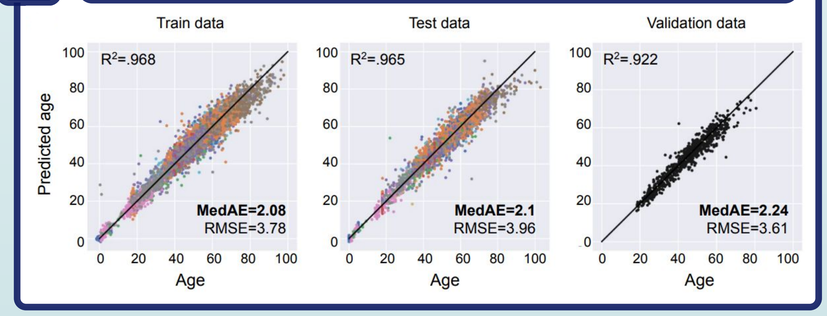
Spectral clustering (k=30), and selection of most correlative CpG site from each cluster

3 Gaussian Process Regression

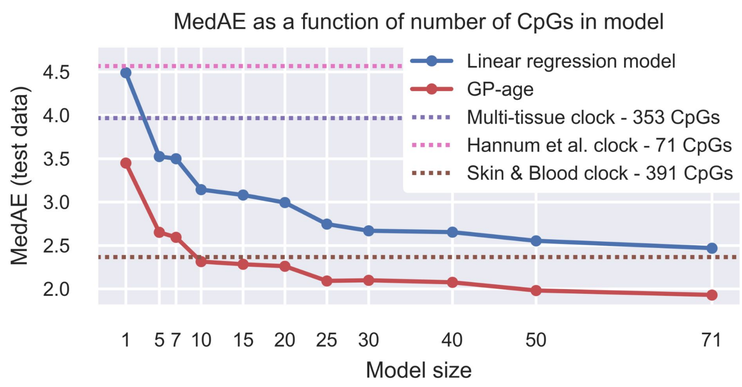


Per sample, the prediction of age is a normal distribution centered around a weighted average* of similar train-set methylomes.

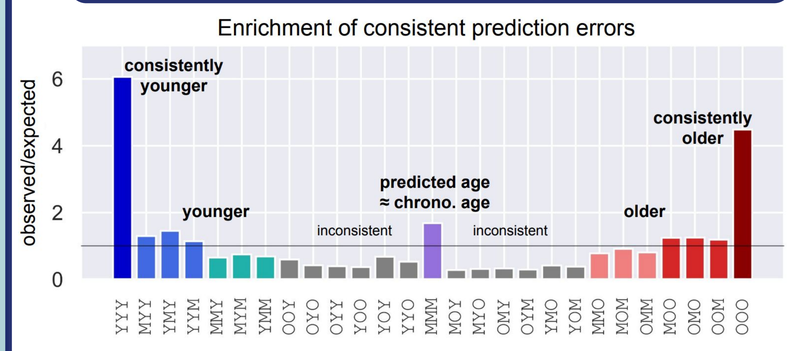
4 GP-age predicts well on held-out data



5 GP-age outperforms state-of-the art epigenetic clocks with only 30 CpG sites



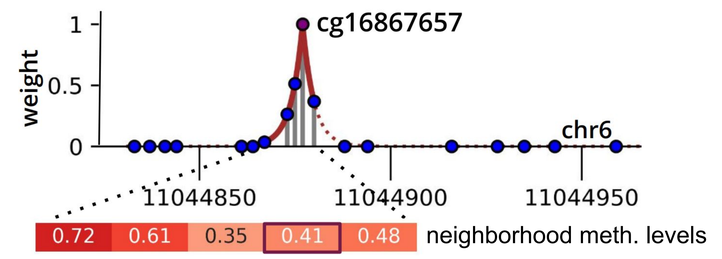
6 Consistent errors across independent GPR models suggest biological variance



Types of prediction errors of 3 independent models:
Y=younger, M=match, O=older than real age

7 From array to sequencing data

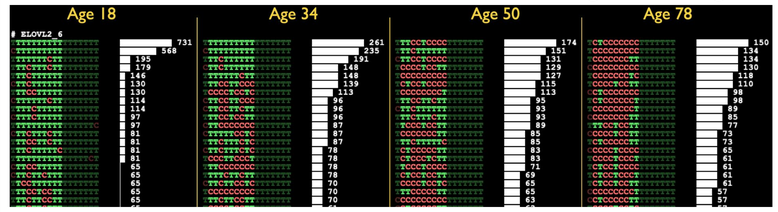
Estimation of methylation levels with Laplace kernel over methylation levels of neighboring CpG sites.



Dataset	Error (years)	GP-age (30 CpGs)	GP-age w/ neigh. (30 CpGs)	Multi-tissue clock[1]	Skin& Blood clock[2]	Hannum et al. clock[3]
Jensen et al. (Buffy coat, n=7)	MedAE (RMSE)	3.00 (6.10)	**1.75 (3.06)**	10.73 (11.37)	14.56 (12.16)	9.43 (11.40)
Fox-Fisher et al. (WBC, n=23)	MedAE (RMSE)	3.55 (4.92)	**2.29 (4.16)**	7.07 (10.44)	14.41 (14.46)	4.65 (6.67)

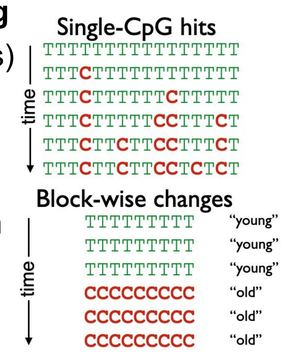
8 Single-molecule views of methylation

Meth. patterns at the ELOVL2 locus change with age



Molecular/cellular views of aging

- Aging as a stochastic (continuous) process of single CpG changes
- Aging as a binary switch between "young" and "old" states, at the cellular level



9 Summary

- Using only 30 CpG sites, GP-age provides an accessible and accurate alternative for epigenetic clocks
- Coordinated prediction errors from independent models suggest a biological signal
- GP-age is applicable on mid-coverage sequencing data, by incorporating neighborhood information
- Paves the way for cheap, targeted PCR-based, age estimation from blood

[1] Horvath, DNA methylation age of human tissues and cell types. *Genome Biology*, 2013
[2] Horvath et al., Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging*, 2018
[3] Hannum et al., Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 2013

miri.varshavsky@mail.huji.ac.il
tommy.kaplan@mail.huji.ac.il