



ProBAge – Probabilistic Inference of Epigenetic Age Acceleration

Emma J. Yang*, Jan K. Dabrowski*, Samuel J. C. Crofts, Robert F. Hillary, Daniel J. Simpson, Daniel L. McCartney, Riccardo E. Marioni, Kristina Kirschner, Eric Latorre-Crespo+, Tamir Chandra+

Abstract

While epigenetic clocks excel at predicting chronological and biological age, they lack mechanistic understanding. We propose a novel probabilistic model that:

- Separates **age acceleration** (the rate of methylation transitions) from **bias** (overall methylation levels).
- Shows **improved associations** with known health-influencing factors such as smoking and alcohol consumption comparing to previous clocks.
- Identified **novel genetic loci** potentially influencing ageing rates.

Predicting biological age using methylation data

Epigenetic clocks leverage DNA methylation patterns to estimate an individual's cellular age, distinct from chronological age.

This innovation led to the concept of '**epigenetic age acceleration**', reflecting the deviation between the model prediction and an individual's chronological age within a cohort.

The widespread adoption of epigenetic clocks stems from their ability to capture the combined influence of diverse environmental and disease factors on cellular ageing in a single metric.

Limitations of current epigenetic clocks

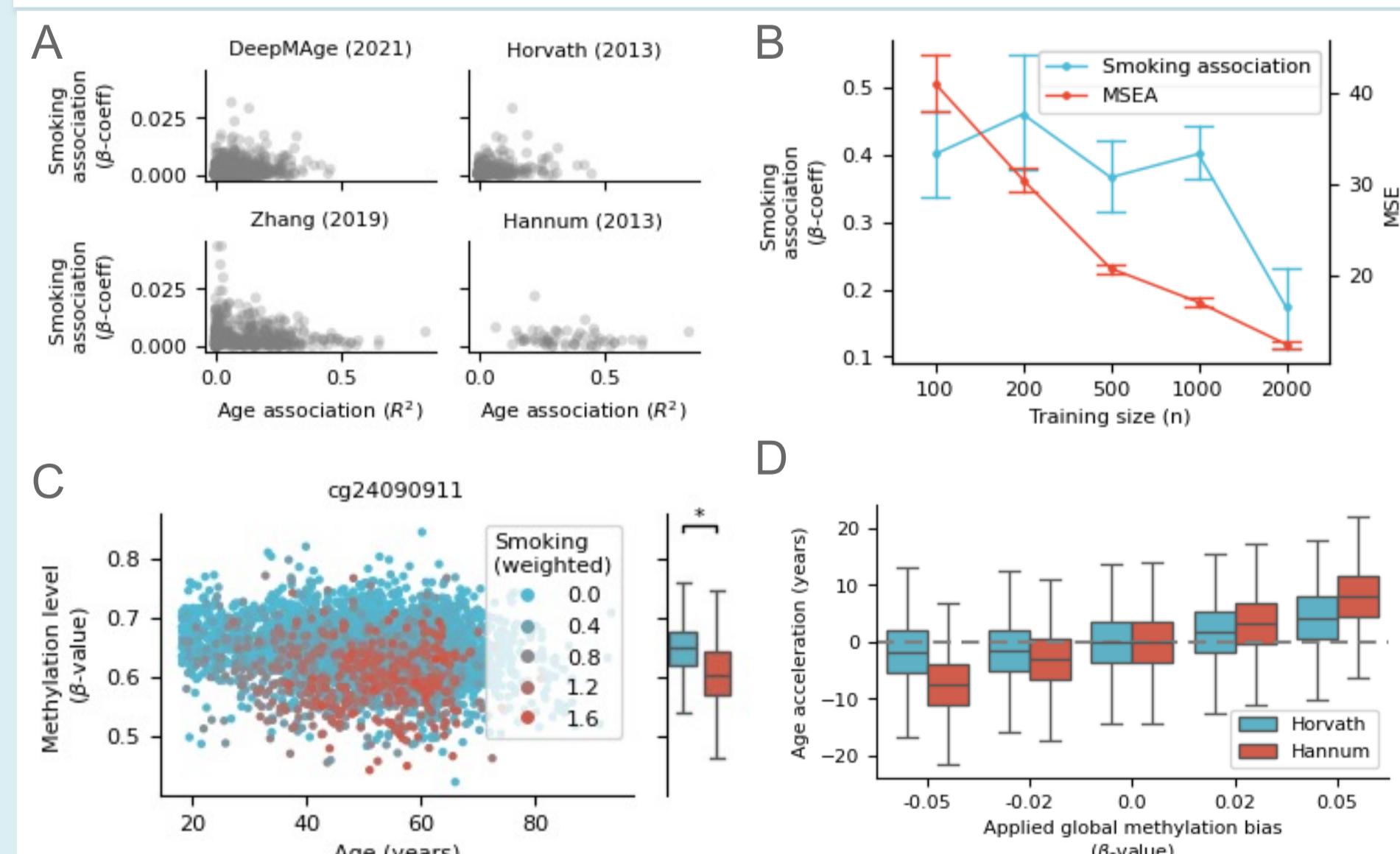


Figure 1. A) Comparison between the association of methylation levels with smoking and age on each CpG site included in various epigenetic clocks. B) Acceleration obtained from bootstrapped lasso linear regressions trained on chronological age for a range of training cohort sizes. C) Methylation beta values vs age for a single CpG that is included in the epigenetic clock of Zhang et al.¹ D) Impact of global offsets on the inferred accelerations using Horvath, Skin and Blood² and Hannum³ clocks.

Overfitting: Large datasets can capture noise, not true ageing.

Global Methylation Batch Effect: Can lead to over/underestimation of epigenetic acceleration.

Age Variance: May misinterpret natural ageing patterns, lead to artificially inflated epigenetic age acceleration in older subjects.

Hence, age acceleration estimates may fail to capture robust disease and lifestyle influences.

ProBAge: Acceleration and Bias

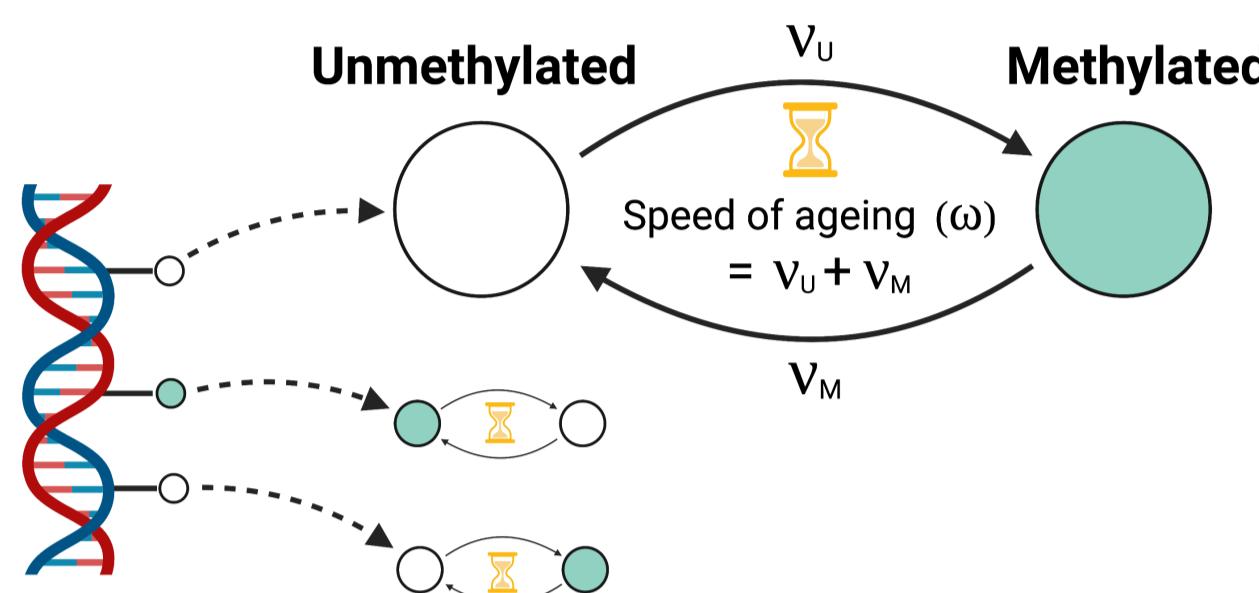
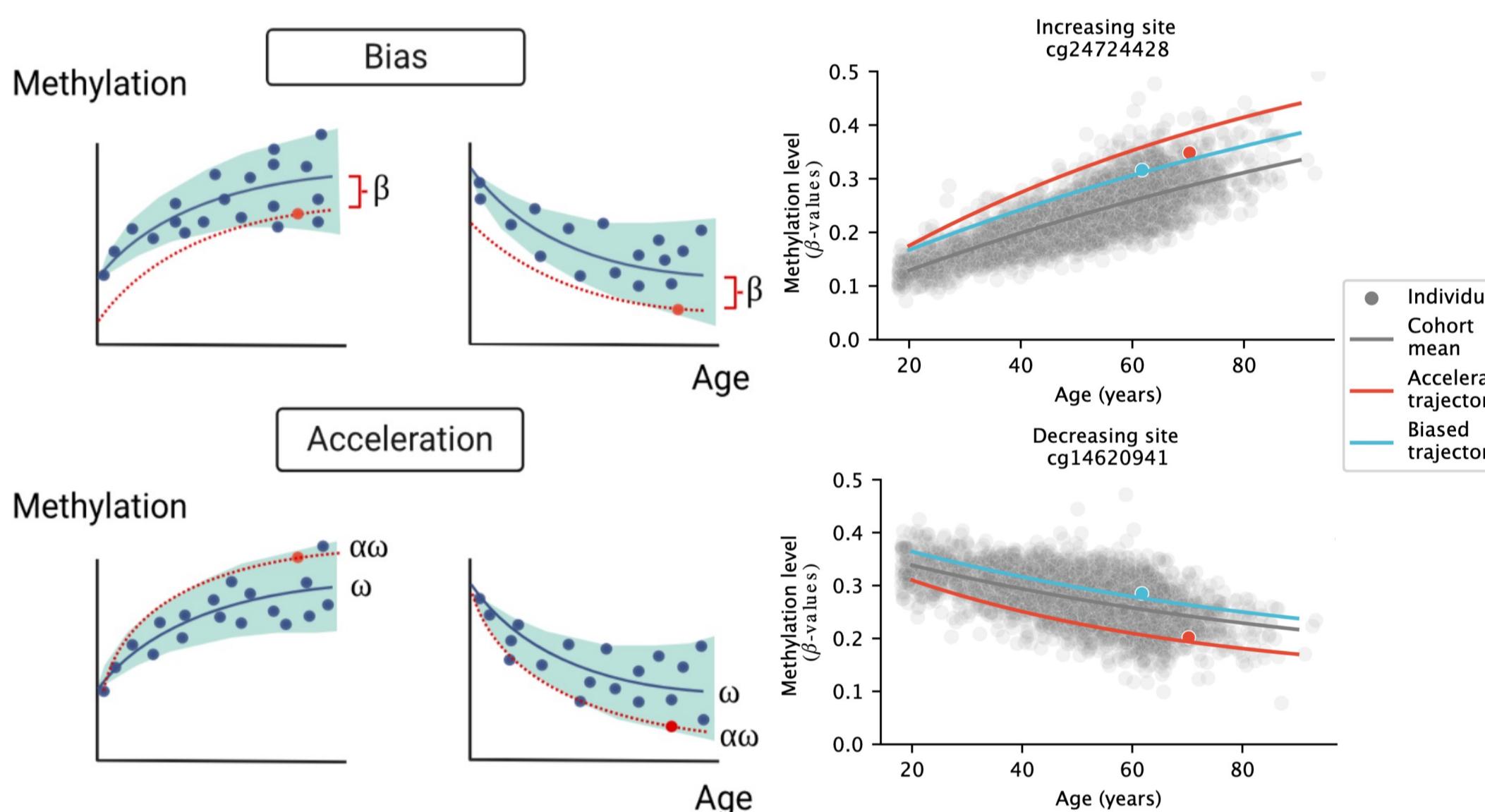


Figure 2. Schematic showing the biological model's underlying stochastic process. For a single cell, multiple CpG sites can over time either methylate or demethylate with rates V_u and V_m respectively. The speed of ageing (ω) is equal to $V_u + V_m$.

Two individual-level parameters:

- Bias β :** global changes in methylation levels
- Acceleration α :** an increase in the speed of methylation changes, ω



Acceleration is associated with lifestyle factors

We tested the association of age acceleration and bias with disease phenotypes and lifestyle factors. As well as a comprehensive performance comparison between our model and the four most widely used epigenetic clocks: Horvath, Hannum (first generation), PhenoAge, and GrimAge (second generation/composite).

The following analyses were conducted on the combined subsets of Generation Scotland (n=15900).

GWAS and downstream analysis of acceleration α

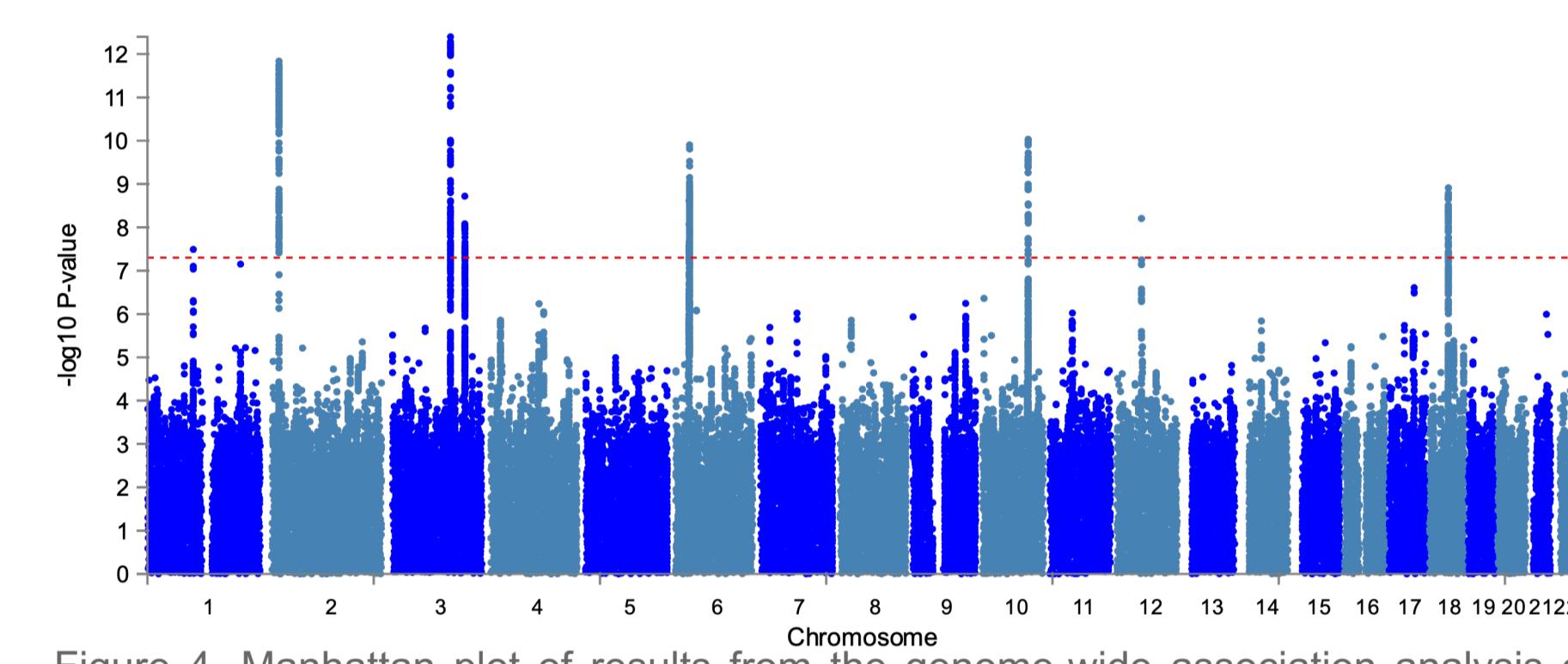


Figure 4. Manhattan plot of results from the genome-wide association analysis of age acceleration.

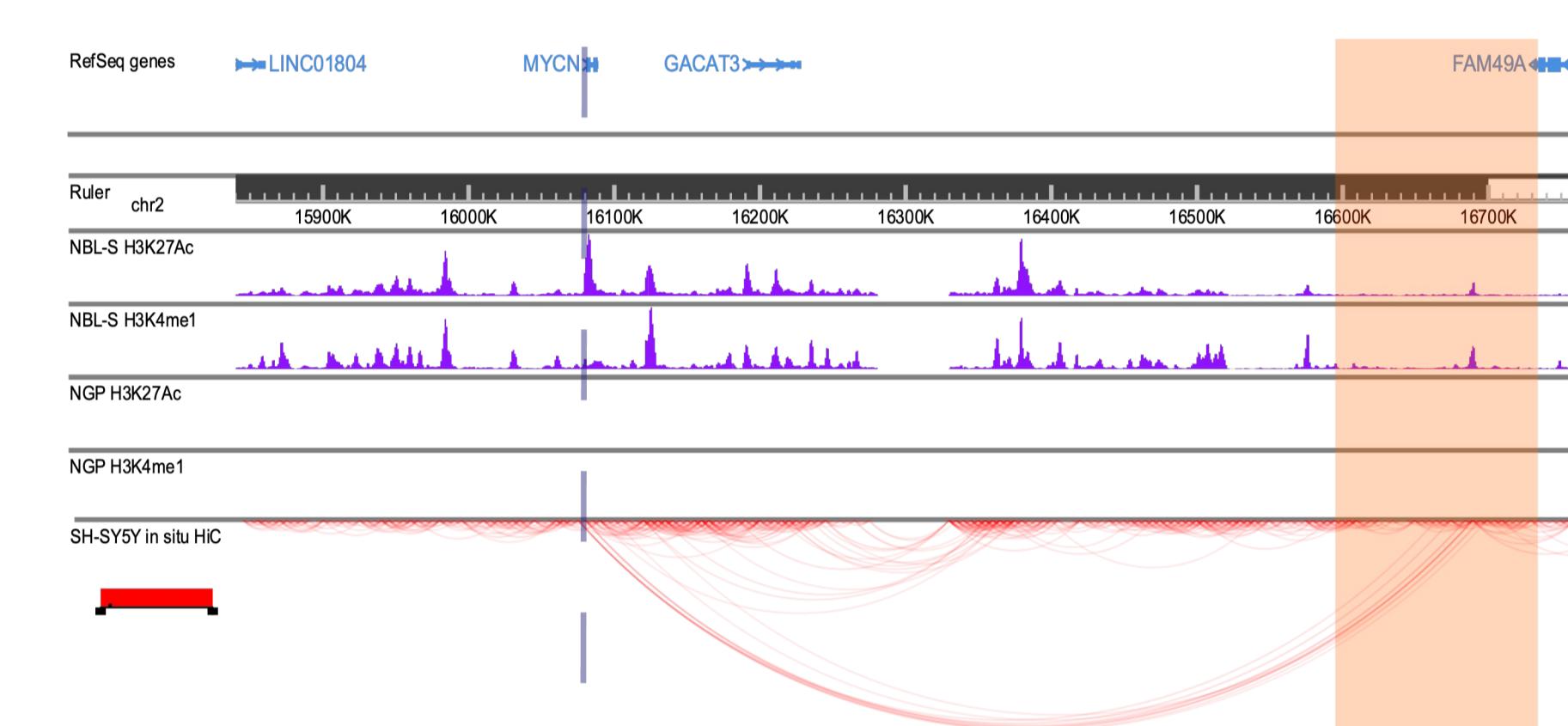


Figure 5. A zoomed-out view of chromatin interactions and chromatin features in and around the genomic risk loci chr2:15841100-17142797 (highlighted in orange). Demonstrates interaction of the genomic risk loci with the MYCN gene locus (dash line).

Position	Chr1	Chr2	Chr3	Chr10	Chr12	Chr18
Lead SNP	rs2893468	rs11096574	rs2492286	rs2001703	rs74808447	rs3740078
Acceleration clocks	N	PhenoAge	Horvath	Horvath	N	PhenoAge + GrimAge
Mitochondria related	N	N	N	N	N	Y
Telomere related	N	N	N	N	Y	N
Chronic kidney disease	Y	N	N	N	Y	N
Other age-related diseases	N	N	N	Y	Y	N
Severe COVID-19	N	N	Y	Y	N	N
Alcohol consumption	N	Y	N	Y	N	N

Table 1. Presence of GWASCatalog trait associations of any SNPs in LD ($R^2 > 0.1$) with the lead SNP in the six genome-wide significant genomic loci.

Discussion

Association Analysis: The acceleration measure shows significant associations with various phenotypes and diseases, including diabetes and COPD. Increased acceleration also correlates with higher mortality risk.

Bias as a Confounder: Bias affects epigenetic age acceleration and, if unaccounted for, can lead to inaccurate associations, as seen with alcohol use in first-generation and composite clocks.

GWAS Findings: Our GWAS associations align more closely with those from the composite clock PhenoAge rather than first-generation clocks, highlighting the biological relevance of our model.

Acknowledgements

The authors thank all of the participants of the Generation Scotland: Scottish Family Health Study as well as study team members for their previous and ongoing contribution to this study.

Reference:

- Zhang, Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 11, 54 (2019).
- Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384 (2018).
- Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49, 359–367 (2013).