DNA methylation array data and Ageing Clock

SubGM Genomics

DNA methylation

- DNA methylation is a well-studied epigenetic mechanism in plants and animals. Methylation involves the covalent transfer of a methyl group to the C-5 position of the cytosine ring on a DNA strand.

- DNA methylation has been shown to play a role in normal development, genomic imprinting, X-chromosome inactivation, chromosome stability, many diseases, especially cancer.

- DNA methylation is regulated by DNA methyltransferases. In humans there are four - DNMT1, DNMT2, DNMT3a, and DNMT3b - which serve a variety of functions.

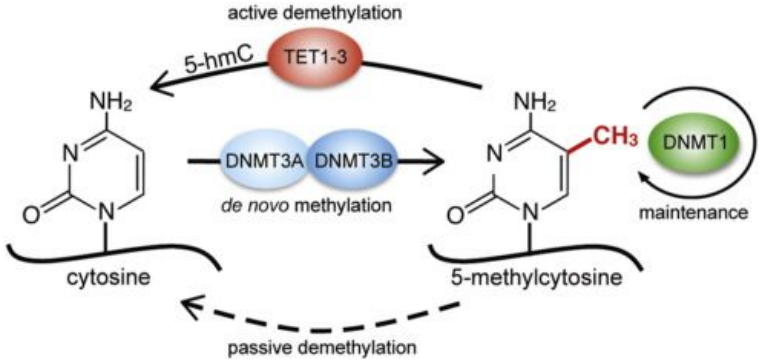
- It is a reversible modification that can occur actively or passively. Passive demethylation occurs when there is a lack of maintenance machinery, such as DNMT1.

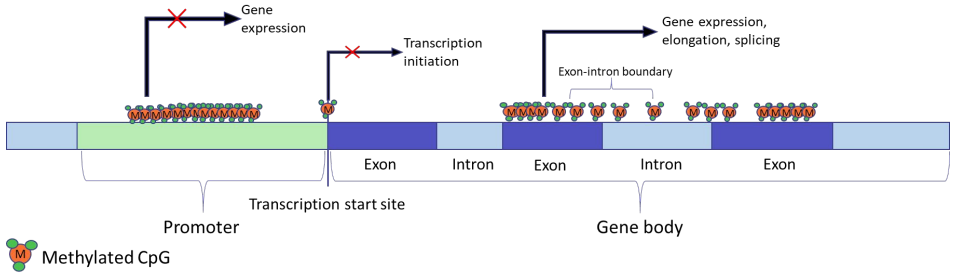
- The effects of DNA methylation vary depending upon the position in the genome at which it occurs.

- Epigenome-wide association studies (EWAS) are a commonly used approach for detecting epigenetic changes.

- Whole-genome bisulphite sequencing (WGBS), reduced representation bisulfite sequencing (RRBS), array-based technologies.

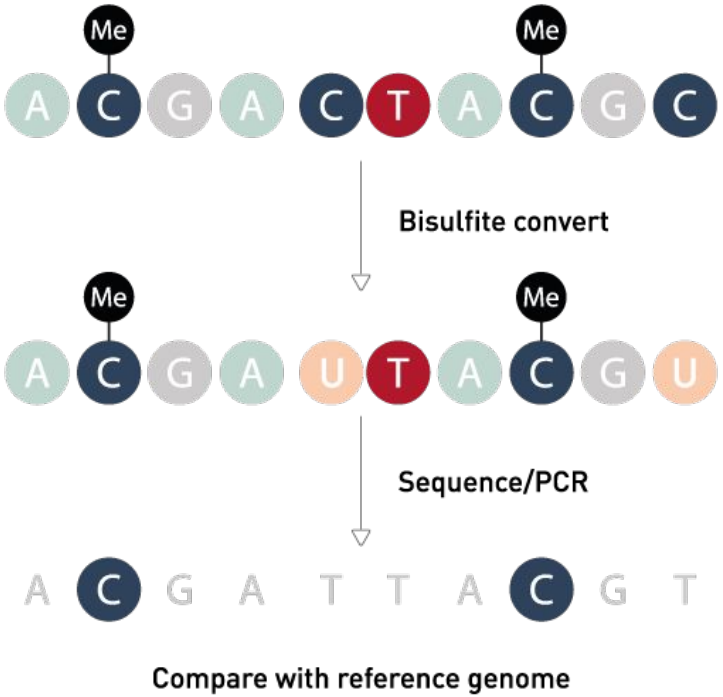
- In the field of microarrays, Illumina Infiniumis the market leader with the EPIC array, the popular 450K and the previous 27K arrays.





Aquino EM, Benton MC, Haupt LM, Sutherland HG, Griffiths LR, Lea RA. Current Understanding of DNA Methylation and Age-related Disease. OBM Genetics 2018;2(2):016; doi:10.21926/obm.genet.1802016.

Bisulphite conversion

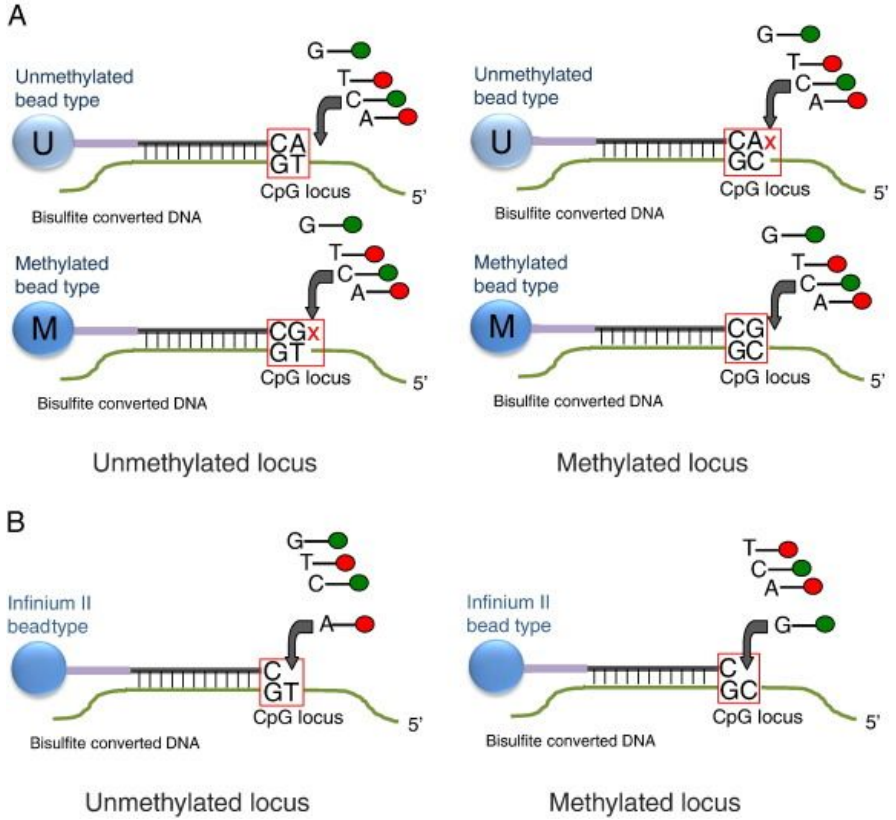


Marina Bibikova et al. High density DNA methylation array with single CpG site resolution. Genomics (2011)

Infinium Methylation Assays

Infinium I assay
(Type I probes)

Infinium II assay
(Type II probes)



Marina Bibikova et al. High density DNA methylation array with single CpG site resolution. Genomics (2011)

Methods to measure methylation levels

The methylation level is then estimated based on the measured intensities of this pair of probes. To date, two methods have been proposed to measure the methylation level:

- The Beta-value is the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities).
- M-value is calculated as the log2 ratio of the intensities of methylated probe versus unmethylated.
- The relationship between Beta-value and M-value is a logistic function.

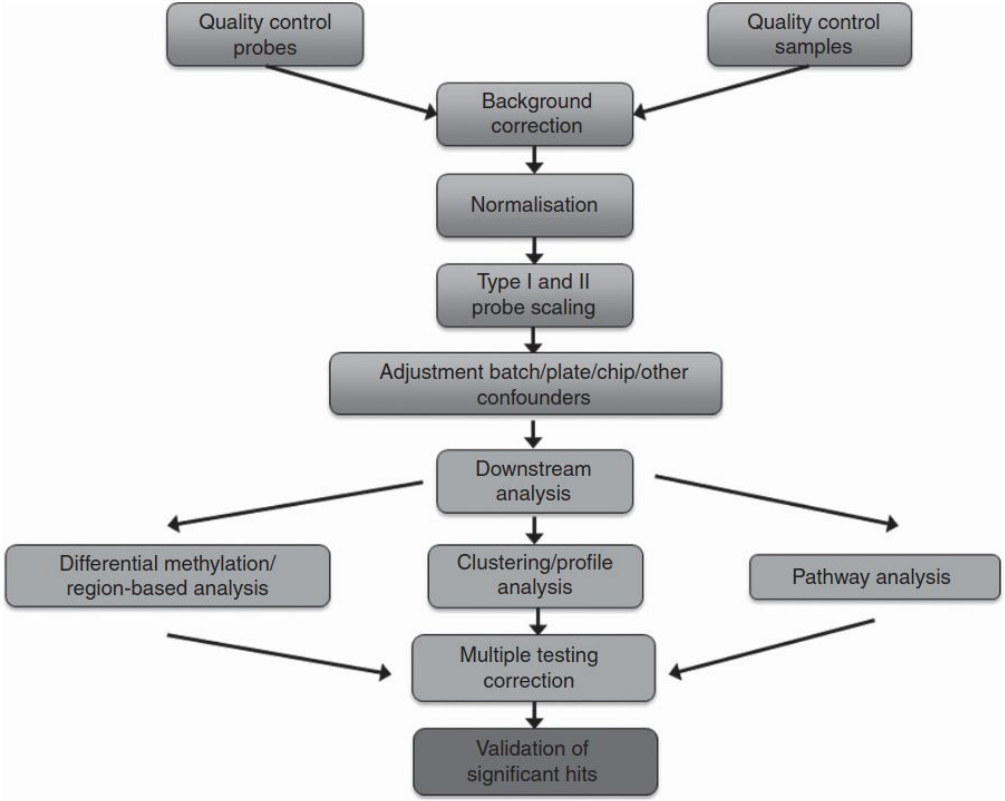
$$Beta_i = \frac{\max(y_{i,methy}, 0)}{\max(y_{i,unmethy}, 0) + \max(y_{i,methy}, 0) + \alpha}$$

$$M_i = \log_2 \left(\frac{\max(y_{i,methy}, 0) + \alpha}{\max(y_{i,unmethy}, 0) + \alpha} \right)$$

$$Beta_i = \frac{2^{M_i}}{2^{M_i} + 1}; M_i = \log_2 \left(\frac{Beta_i}{1 - Beta_i} \right)$$

Pan Du et al., Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010

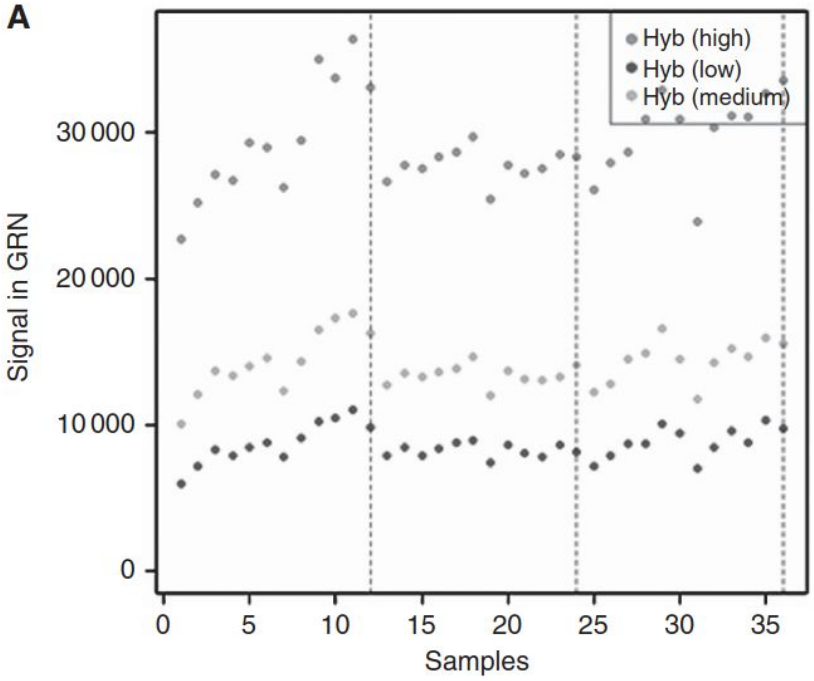
Methylation data processing and analysis pipeline



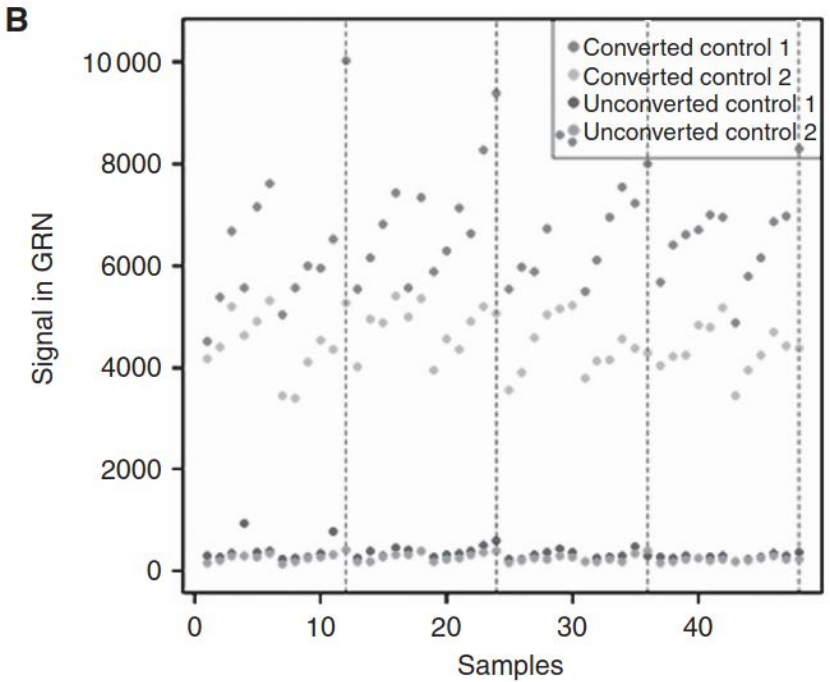
C S Wilhelm-Benartzi et al.Review of processing and analysis methods for DNA methylation array data. British Journal of Cancer (2013)

Quality Control of Samples

Hybridisation quality control plot in the green channel



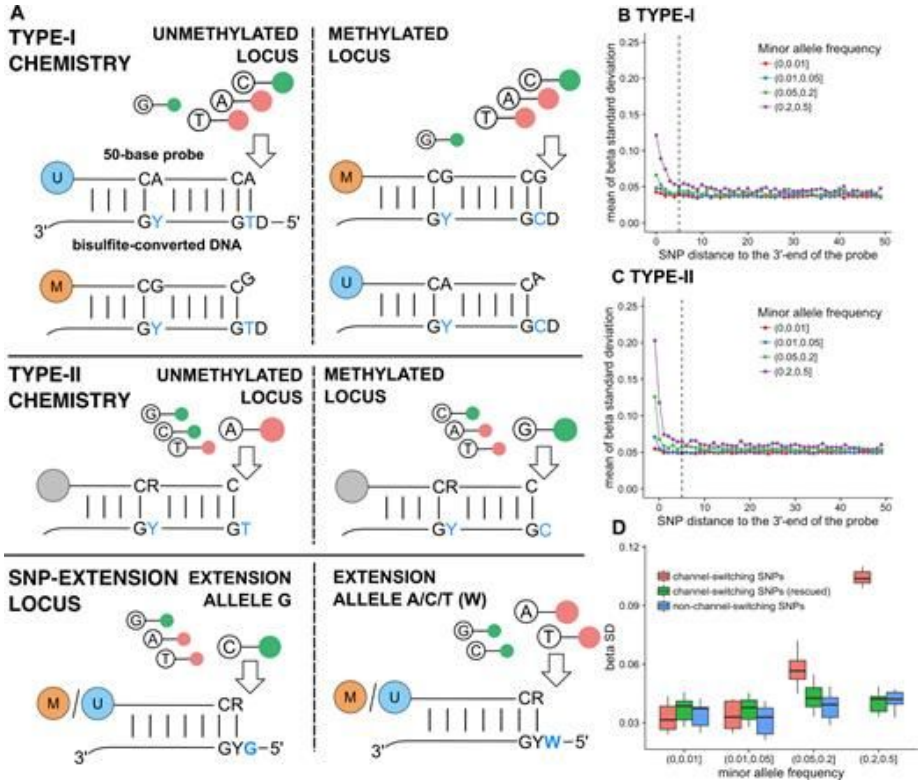
Bisulphite conversion quality control plot in the green channel



C S Wilhelm-Benartzi et al.Review of processing and analysis methods for DNA methylation array data. British Journal of Cancer (2013)

Quality Control of Probes

- Illumina methylation data is usually obtained in the form of Intensity Data (IDAT) Files.
- Detection p-value below a threshold.
- SNP-related probes.
- Probes with non-unique mapping.
- Probes with off-target hybridization due to partial overlap with non-unique elements.
- Other filtering: probes with <3 beads in at least 5% of samples per probe, probes located in chromosome X and Y.



C S Wilhelm-Benartzi et al.Review of processing and analysis methods for DNA methylation array data. British Journal of Cancer (2013)

Normalization

Normalisation concerns the removal of sources of experimental artefacts, random noise and technical and systematic variation caused by microarray technology, which, if left unaddressed, has the potential to mask true biological differences.

(1) Between-array normalisation, removing technical artefacts between samples on different arrays. Batch effect evaluation (SVD) and removal if required (ComBat method, Surrogate variable analysis).

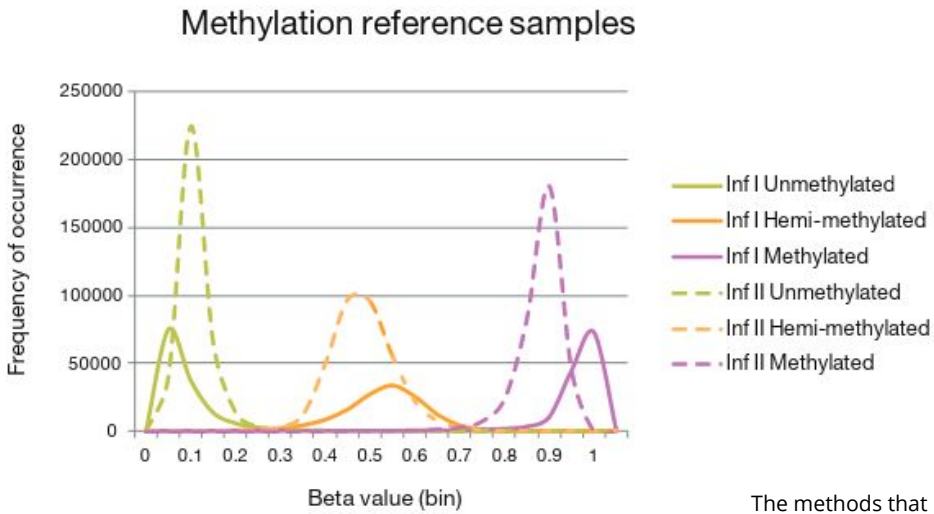
(2) Within-array normalisation, correcting for intensity-related dye biases.

- Quantile normalization.
- Locally weighted scatterplot smoothing (LOESS) normalisation.
- Subset quantile normalisation.
- Weighted P-spline intensity-dependent normalisation.

C S Wilhelm-Benartzi et al.Review of processing and analysis methods for DNA methylation array data. British Journal of Cancer (2013)

Type I and Type II probe scaling

- Type-I and type-II probes, with different hybridization chemistries → different distributions.
- 450K array → 72% use the Infinium type II and the remainder use the Infinium type I primer extension assay.



Normalization methods

- Peak-based correction (PBC).
- Subset-quantile within array normalization (SWAN).
- Beta Mixture Quantile dilation (BMIQ).

The methods that I used are highlighted.
Benchmarking of the methods: Ting Wang. A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. Epigenetics. 2015.
https://www.bioconductor.org/packages/release/bioc/vignettes/ChAMP/inst/doc/ChAMP.html

Downstream Analysis

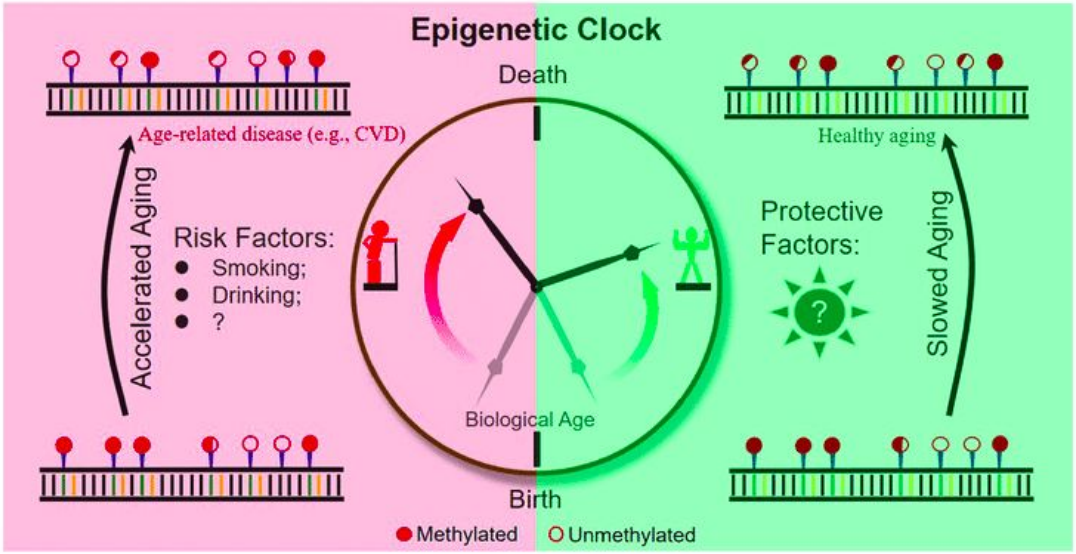
- Differential methylated probes → limma and eBayes methods (the same as RNA-seq).
- Differentially Methylated Regions (DMRs) that are extended segments of the genome. Several approaches to find DMRs:
  - Bumphunter
  - ProbeLasso
  - DMRcate
- Differential Variability (library missMethyl).
- Gene Set Enrichment Analysis (GSEA). Issue: different genes have different numbers of CpGs contained inside.
  - Gometh
  - ebayes

The methods that I used are highlighted.
https://www.bioconductor.org/packages/release/bioc/vignettes/ChAMP/inst/doc/ChAMP.html

Epigenetic Clock

Epigenetic age estimators are mathematical algorithms that use values assigned to the methylation state of specific CpGs in the genome to estimate the age of a person or biological sample.

This estimated age (or epigenetic age) is not only a reflection of **chronological age** but also of the **biological age** of the DNA source.



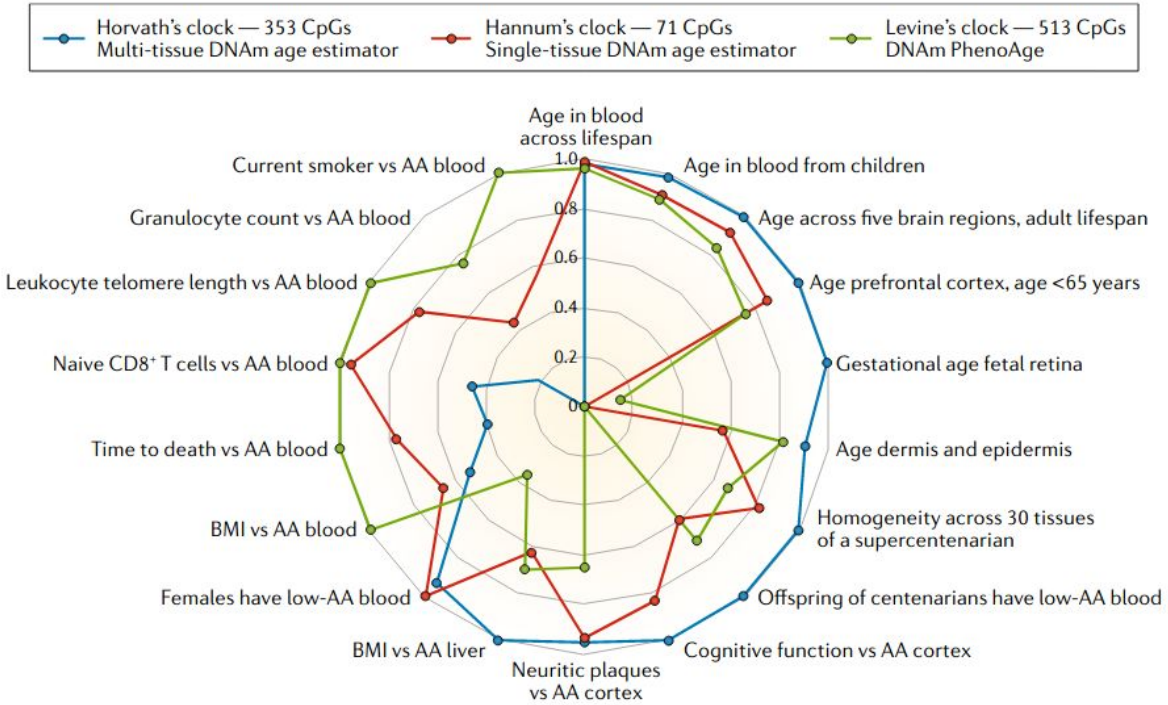
Two interpretations for the "epigenetic clock":

- Age estimators.
- Innate biological processes.

Steve Horvath et al. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature (2018)

Epigenetic Clock(s)

Comparison of three DNA methylation-based biomarkers of ageing



Steve Horvath et al. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature (2018)