DNA methylation in epidemiological research

27 September 2022 - University of Padua Summer School

Gianluca Severi, PhD

DISIA, University of Florence, Italy

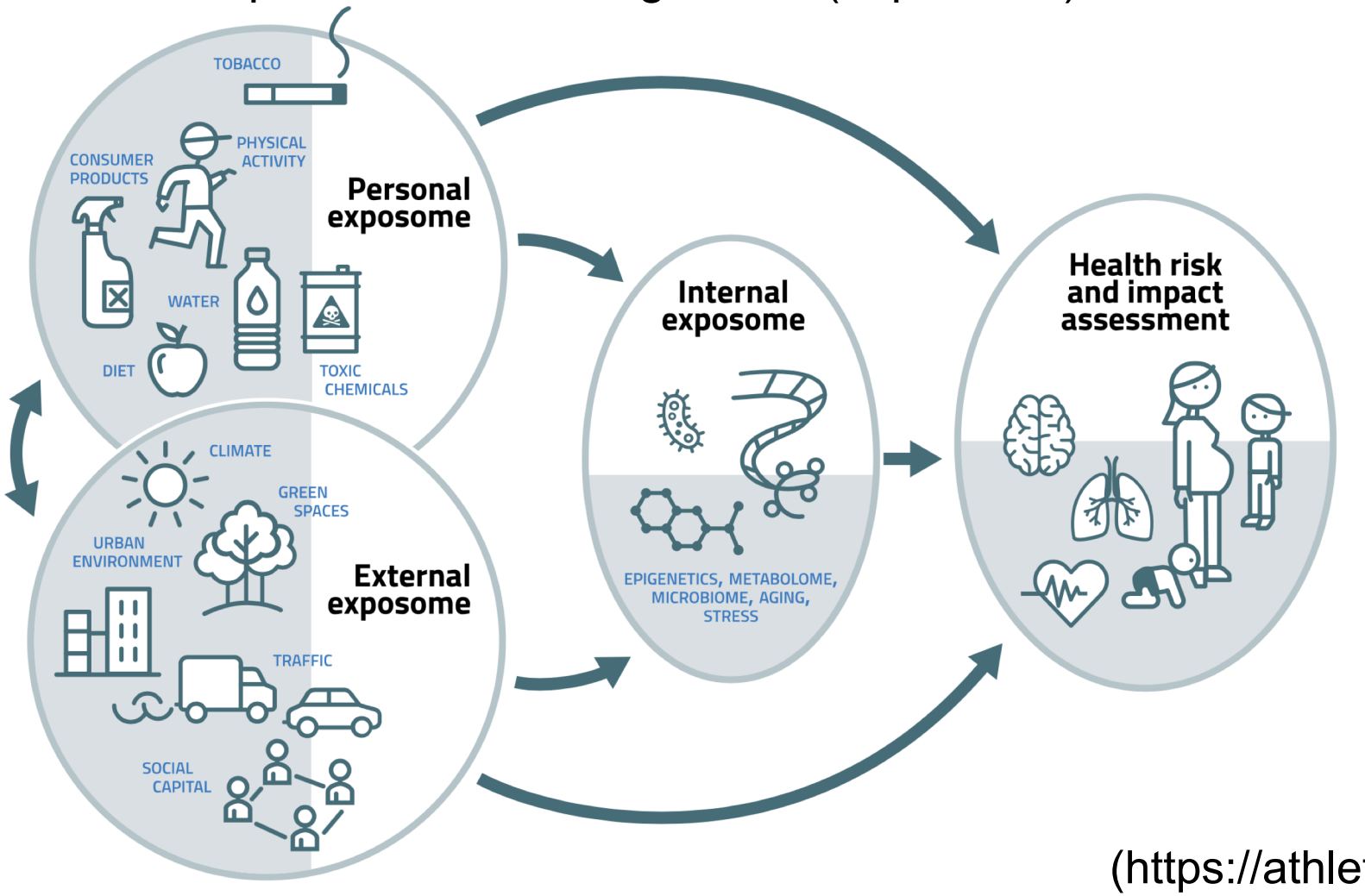
«Exposome, Heredity, Cancer and Health», Inserm CESP U1018, Villejuif, France

Outline

- Introduction :
  - what is epigenetics and DNA methylation
  - why DNA methylation is useful in epidemiology
  - how to measure DNA methylation
- DNA methylation in epidemiological research
  - EWAS (early life exposures, tobacco smoking, cancer risk)
  - «Epigenetic clocks», methylation-based scores of exposure (e.g. tobacco)

DNA methylation, exposures and health

"Omic" analyses and in particular analyses of the methylome are important tools to explores the multiple factors affecting health (exposome)



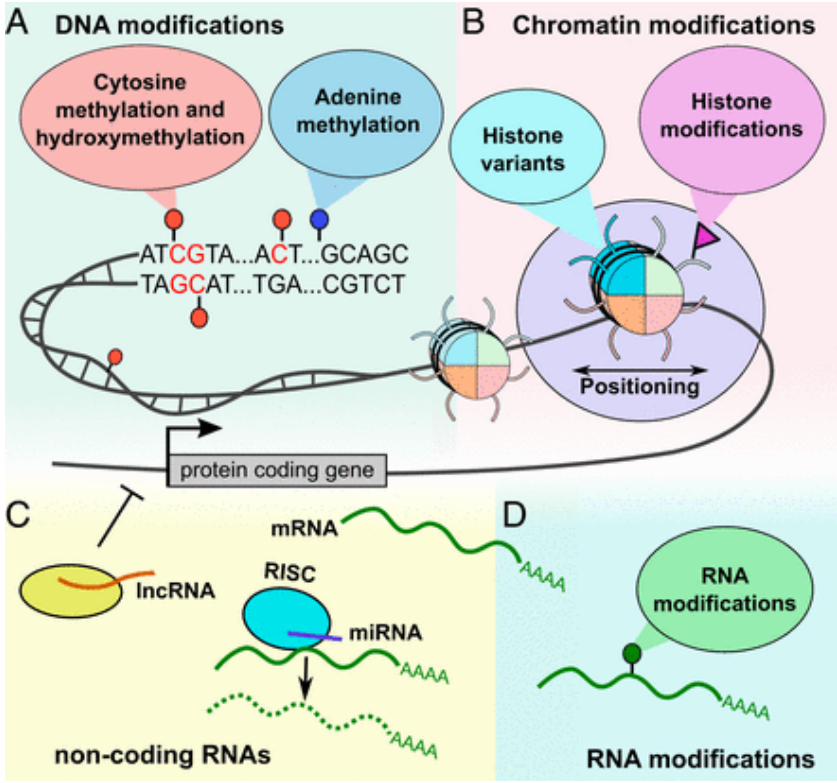
(https://athleteproject.eu)

Epigenetics

Epigenetics refers to the study of variations in phenotypes (e.g. gene expression) that occur **without alterations in the DNA sequence** (*epigenetics=«in addition to genetics»*).

Epigenetic mechanisms include :

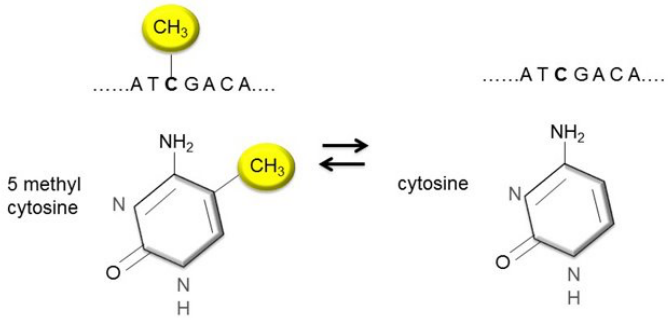
- DNA methylation
- Chromatin folding / remodeling
- Histone modifications
- RNA-based mechanisms



(Aristizabal et al. PNAS 2019)

DNA methylation

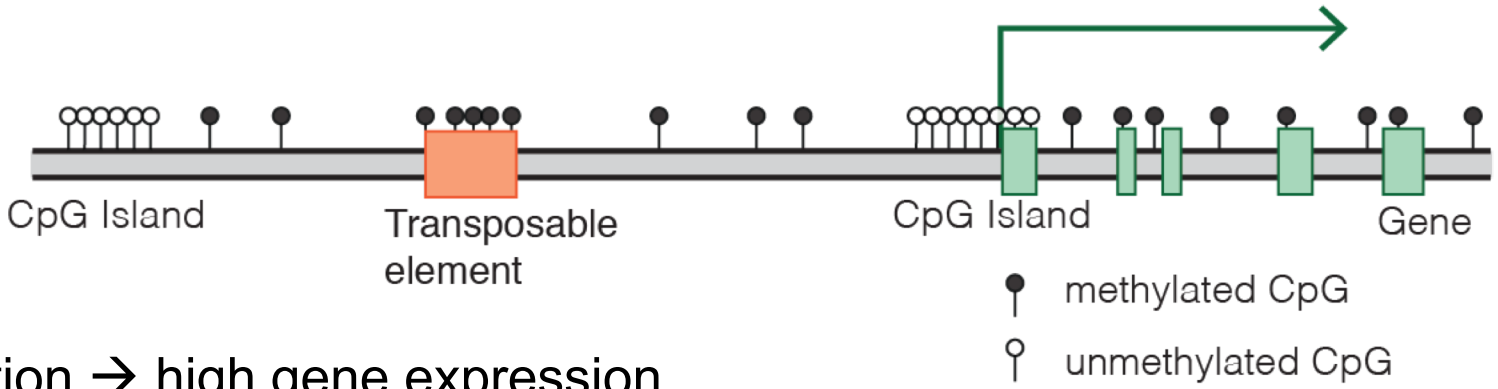
- DNA methylation (DNAm) is a process first discovered the 1940s by which a methyl group is added to the DN molecule (mostly at cytosines in CpG sites).



- In humans 70% to 80% of cytosines in CpG sites are methylated depending on the cell types and tissue.

- In the 1980s studies demonstrated that DNA methylation was involved in gene regulation and cell differentiation.

The role of DNA methylation

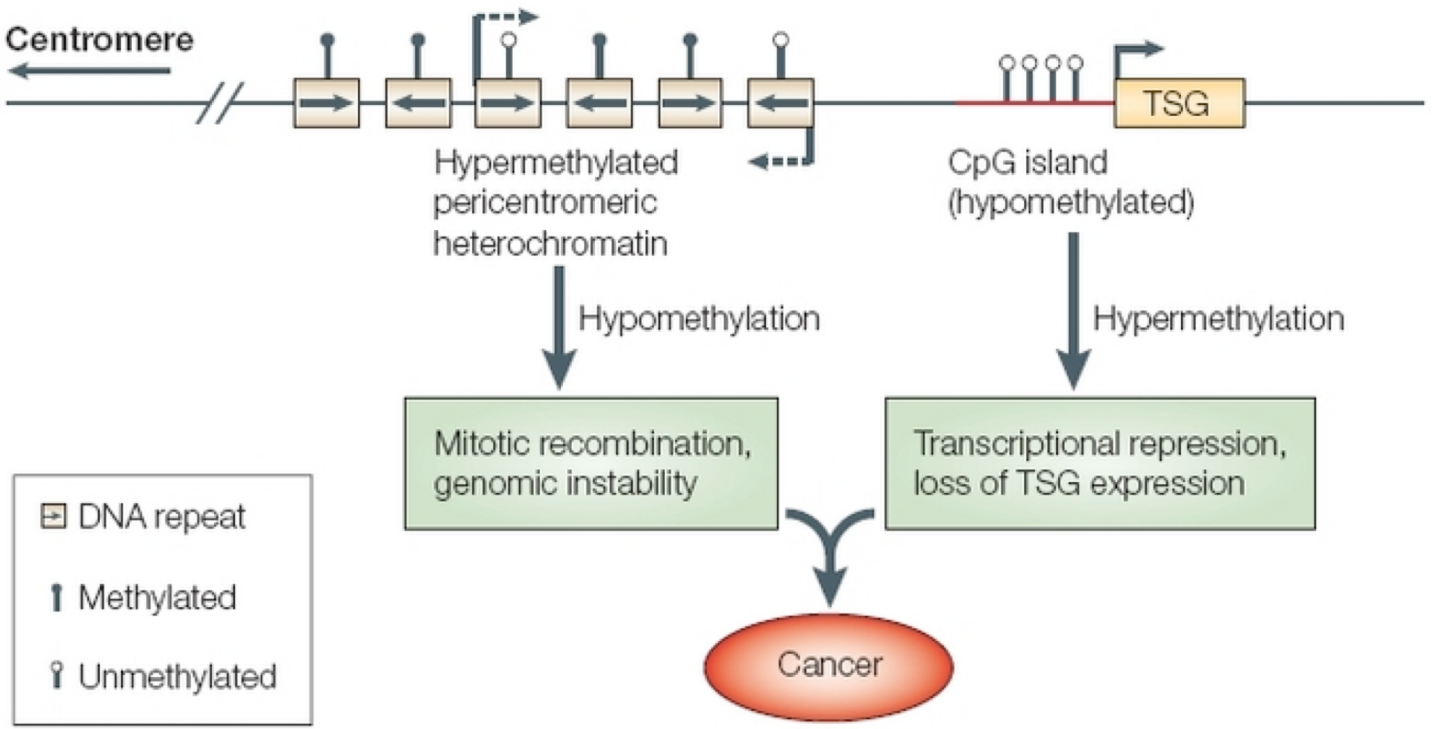


- Low promoter methylation → high gene expression
- Elevated gene body methylation → high gene expression
- Transcriptional inactivation of foreign DNA elements
- X-chromosome inactivation
- Gene imprinting (e.g. paternally expressed IGF2, DNAm silence maternal allele)

DNA methylation in cancer

DNA methylation is vastly altered in cancer with two clear characteristics that emerge:

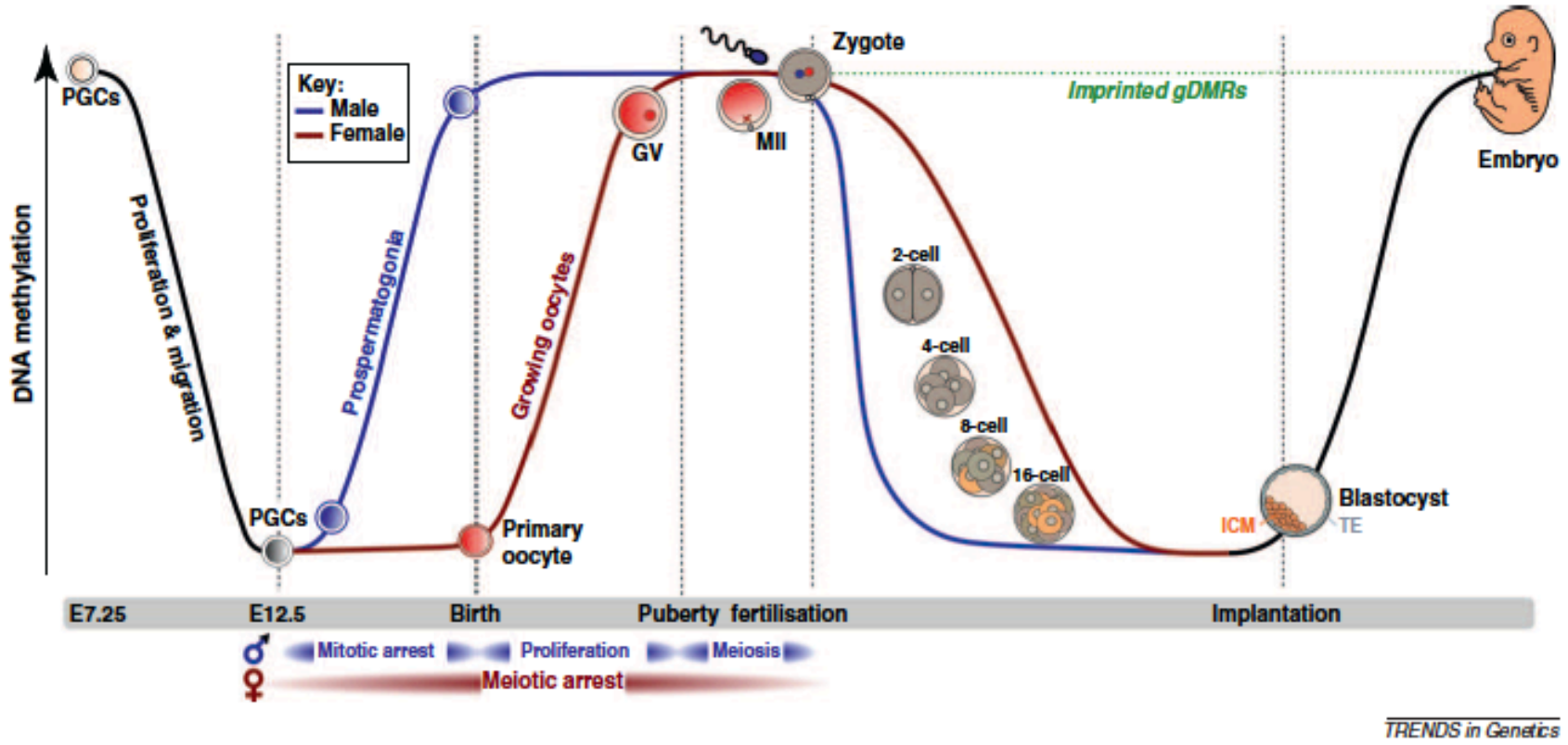
- "global" hypomethylation leading to genomic instability
- hypermethylation of promoters of tumour suppressor genes



DNA methylation is dynamic

During gametogenesis DNA methylation is erased (but probably not completely),

Before implantation, the genome-wide removal of methylation in the embryo allows to acquire an epigenetic profile coherent with pluripotency.

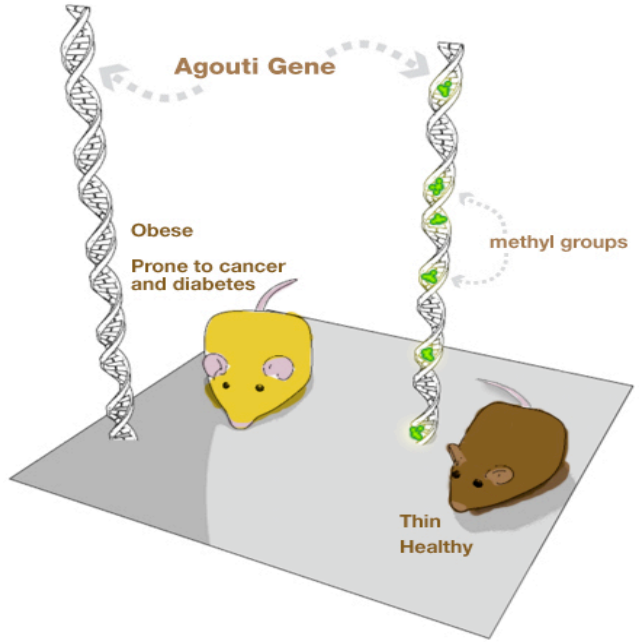


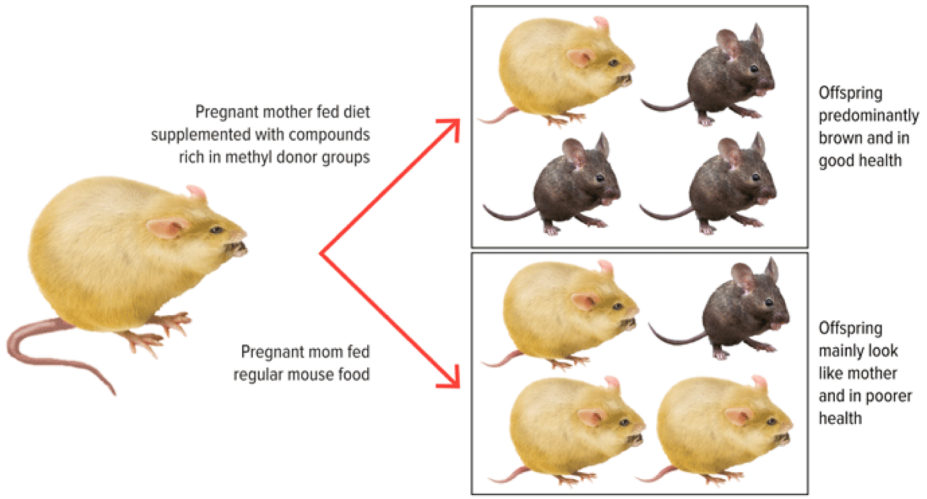
(Smallwood SA and Kelsey Trends in Genet 2012)

DNA methylation may be heritable and reversible

Genetically identical mice display **phenotypic differences due to variations in DNA methylation** of the (e.g. A^{vy} mice, methylation at retrotransposon at agouti gene site).

The phenotype is **transmitted to offspring in a non-Mendelian fashion**. The transmission is rather due to incomplete erasure of DNA methylation marks during gametogenesis.





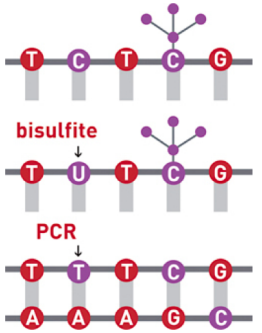
Dietary supplementation with methylating molecules (e.g. methionine) in pregnancy affects phenotype in successive generations (e.g. *A^{vy}* mice offspring predominantly brown, methylated retrotransposon at agouti gene site)

http://learn.genetics.utah.edu/content/epigenetics/nutrition/

Adapted by Bill Day from Waterland, RA., Jirtle, RL. Molecular and Cellular Biology. 2003

Methods for DNA methylation analysis

Bisulfite conversion of DNA : unmethylated cytosines converted into uracil

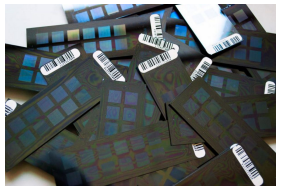


Research question

No prior candidate

Genome-wide studies (EWAS)

- **Bisulphite sequencing (NGS)**
- **MeDIP (immunoprecipitation)**
- **Illumina arrays (27k, 450k, 850k)**



Candidate sites

DNA methylation at target regions

- **Pyrosequencing**
- **Bisulphite sequencing**
- **BeadXpress**

DNA methylation measures from Illumina array data

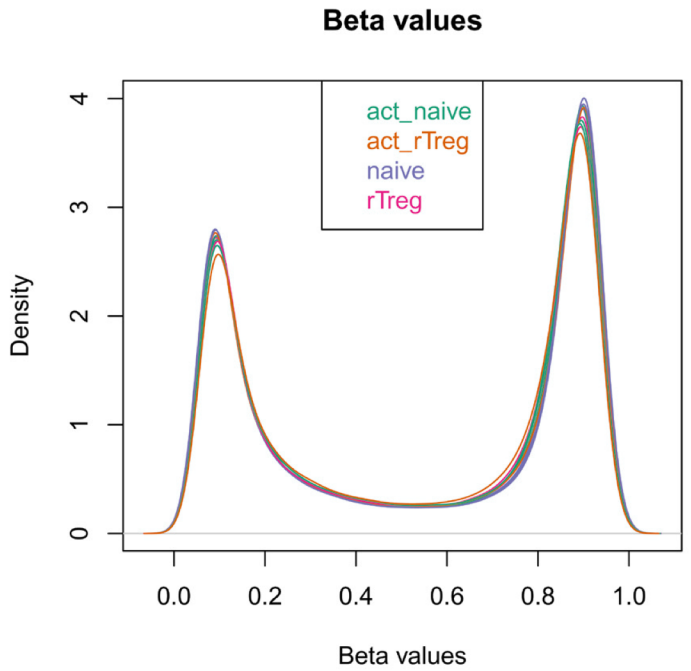
I_M = Intensity of methylated signal, I_U = Intensity of unmethylated signal

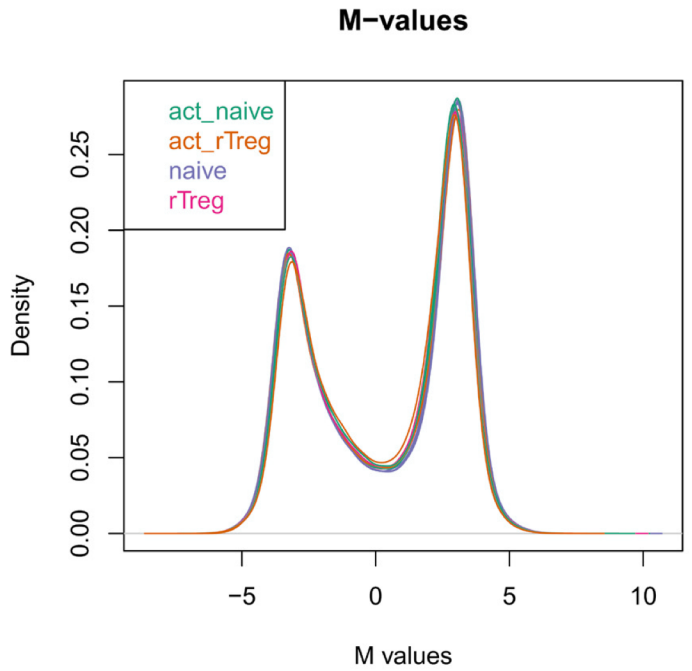
Beta value:

$$\beta = \frac{I_M}{I_M + I_U + 100}$$

M value:

$$M = \log_2\left(\frac{\beta}{1-\beta}\right)$$





DNA methylation depends on the cell of origin

DNA Methylation of DNA blood is an average of methylation states across different cell types.

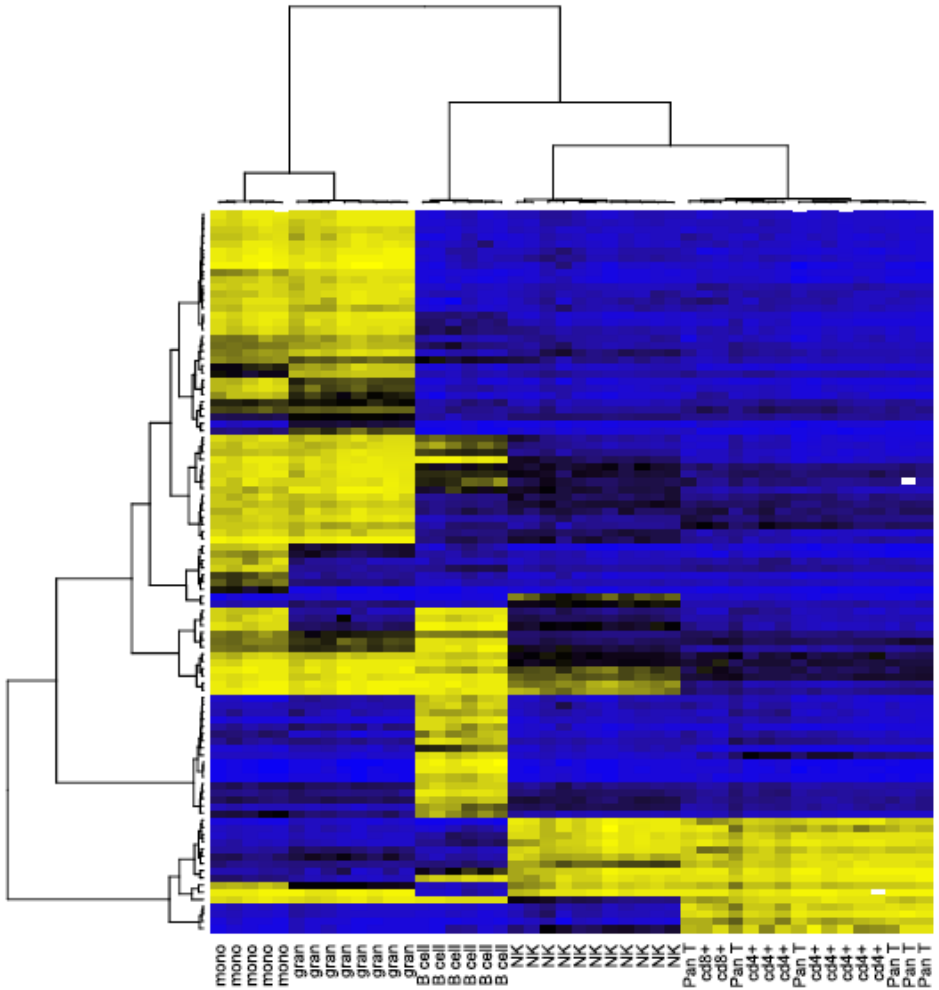


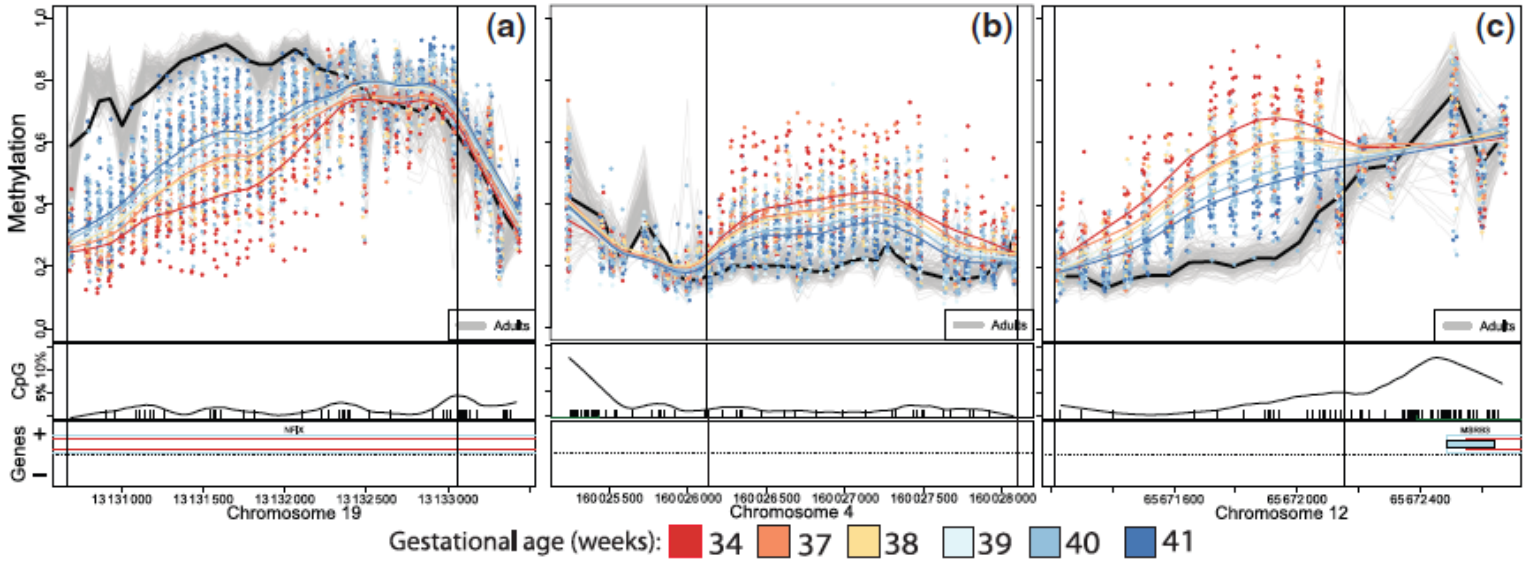
Figure 1 Clustering heatmap for external validation white blood cell data (S_0). Yellow = unmethylated ($Y_{hj} = 0$), black = partially methylated ($Y_{hj} = 0.5$), blue = methylated ($Y_{hj} = 1$).

- Distinct methylation profiles for each specific cell type
- Various methods developed to estimate **cell type distribution in blood** from DNA methylation (to be used to adjust analyses

Houseman et al. BMC Bioinformatics 2012

DNA methylation and gestational age

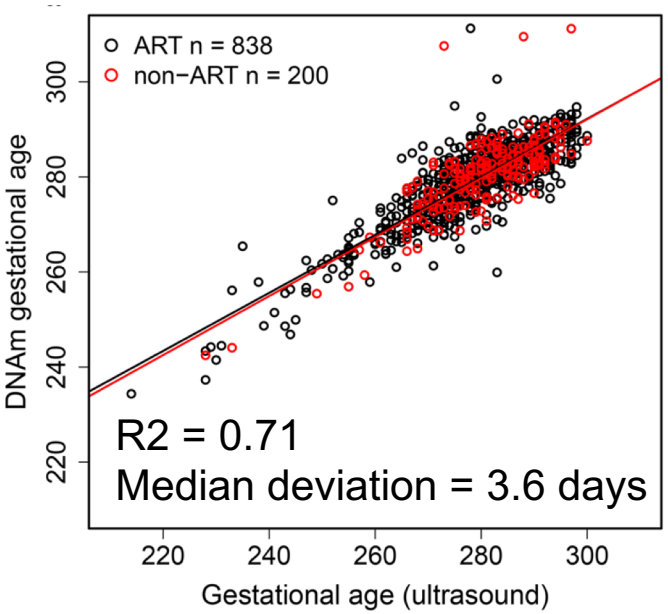
DNA methylation levels in newborn blood are associated with gestational age



Blood DNA methylation in newborns with increasing gestational age approaching levels in adults

(Lee et al. Int J Epidemiol 2012)

«Epigenetic clock» based on 176 CpGs shows strong correlation with gestational age based on ultrasound *(Haftorn et al. Clin Epigenet 2021)*



Prenatal exposure to famine and persistent methylation changes

- Individuals prenatally exposed to the Dutch Hunger Winter in 1944-1945 (N=60) compared with same-sex unexposed siblings.

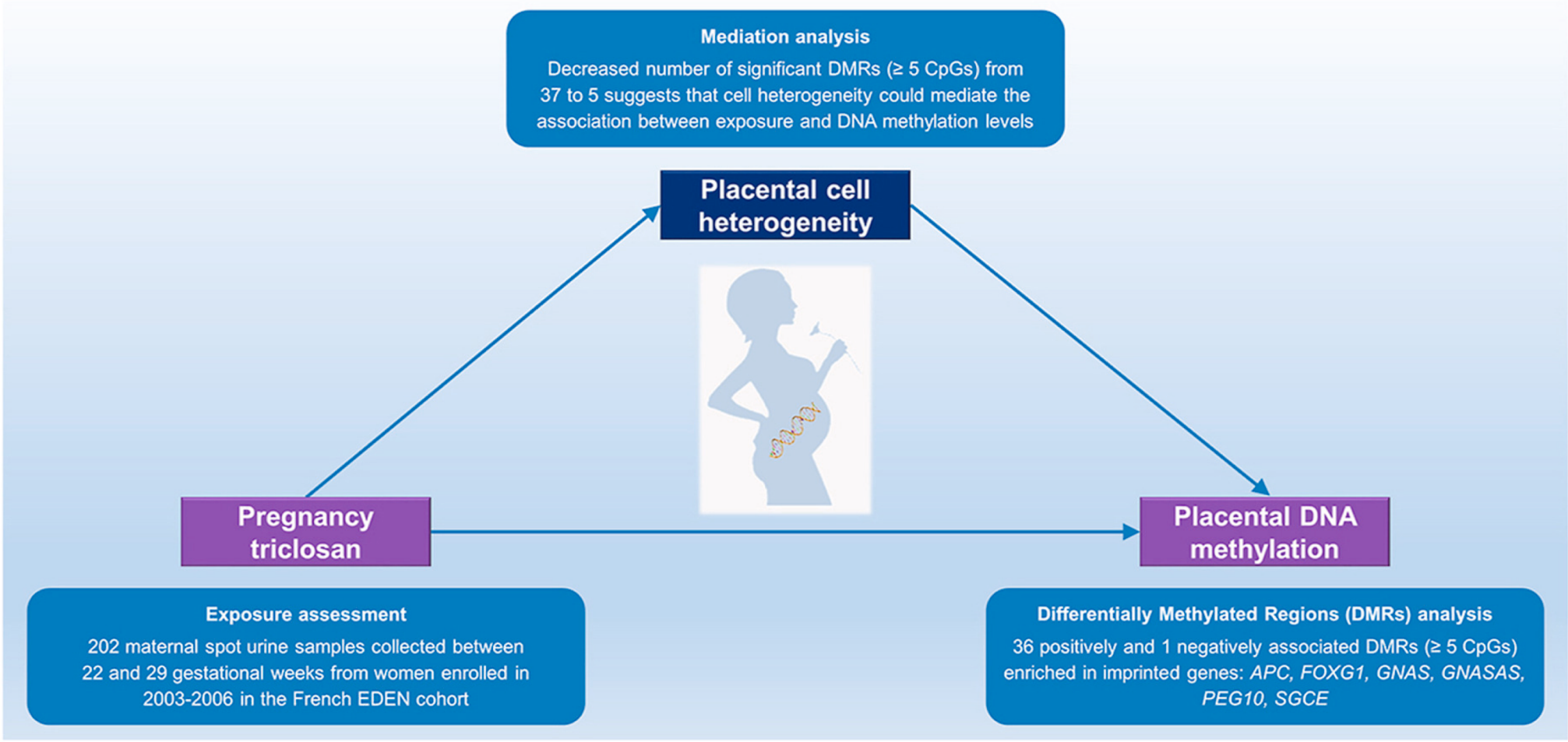
- Hypomethylation of the maternally imprinted *IGF2* gene in the exposed individuals leading to bi-allelic expression. Association specific to **periconceptional exposure to famine** (Heijmans et al. PNAS 2008).

Genome-wide methylation analysis (bisulphite sequencing) *(Tobi et al. Nat Commun 2014)* :

- Widespread DNA methylation changes associated with the famine (181 DMR regions)
- DNA methylation changes in regions linked to growth and metabolism

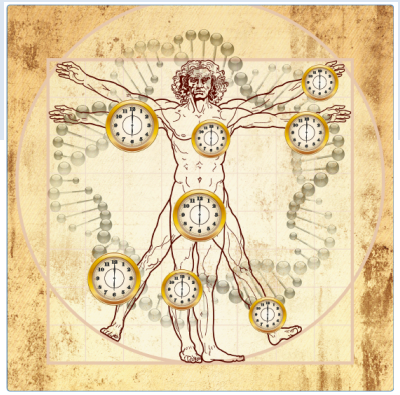
Exposure to synthetic phenols and placental DNA methylation

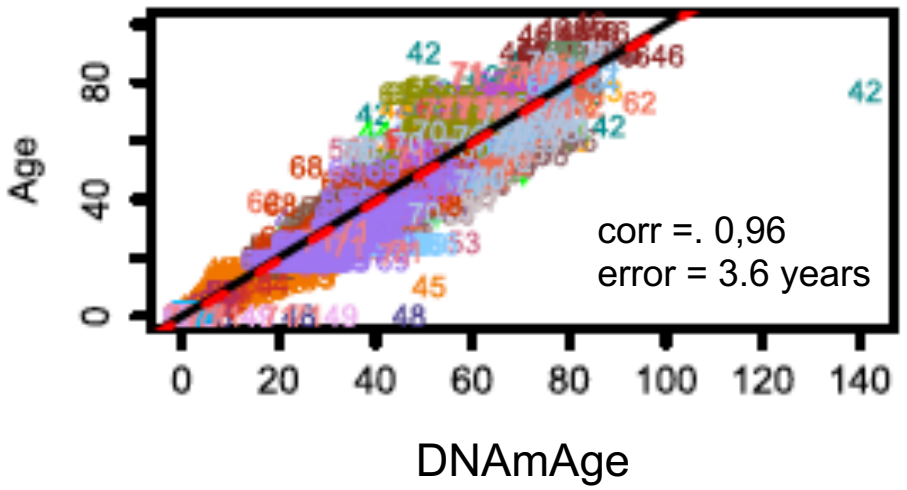
In utero exposure to chemical pollutants may alter placental DNA methylation (or cell heterogeneity in placenta). Limitation : cross-sectional design



Jedynak et al. Env Pollut 2021

«Epigenetic clocks»



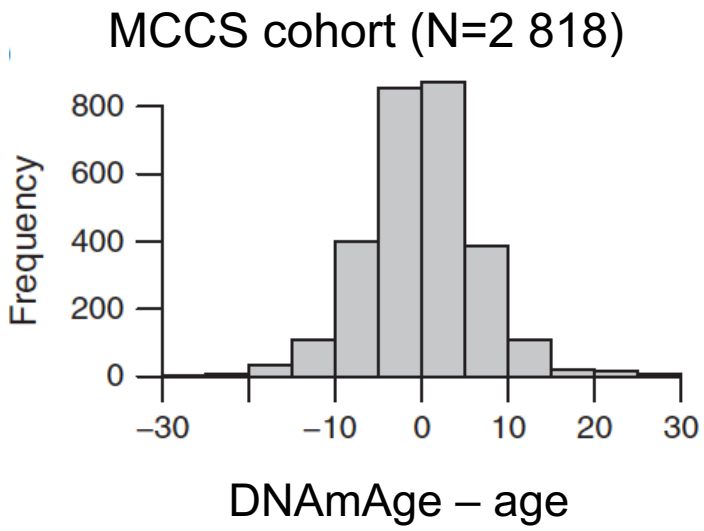


DNA methylation-based scores («epigenetic clocks» or DNAmAge) strongly predict age across different tissue and at different stages of life *(Horvath Genome Biol 2013; Hannum et al. Mol Cell 2013)*.

Increasing «age acceleration» (DNAmAge – age) associated with :

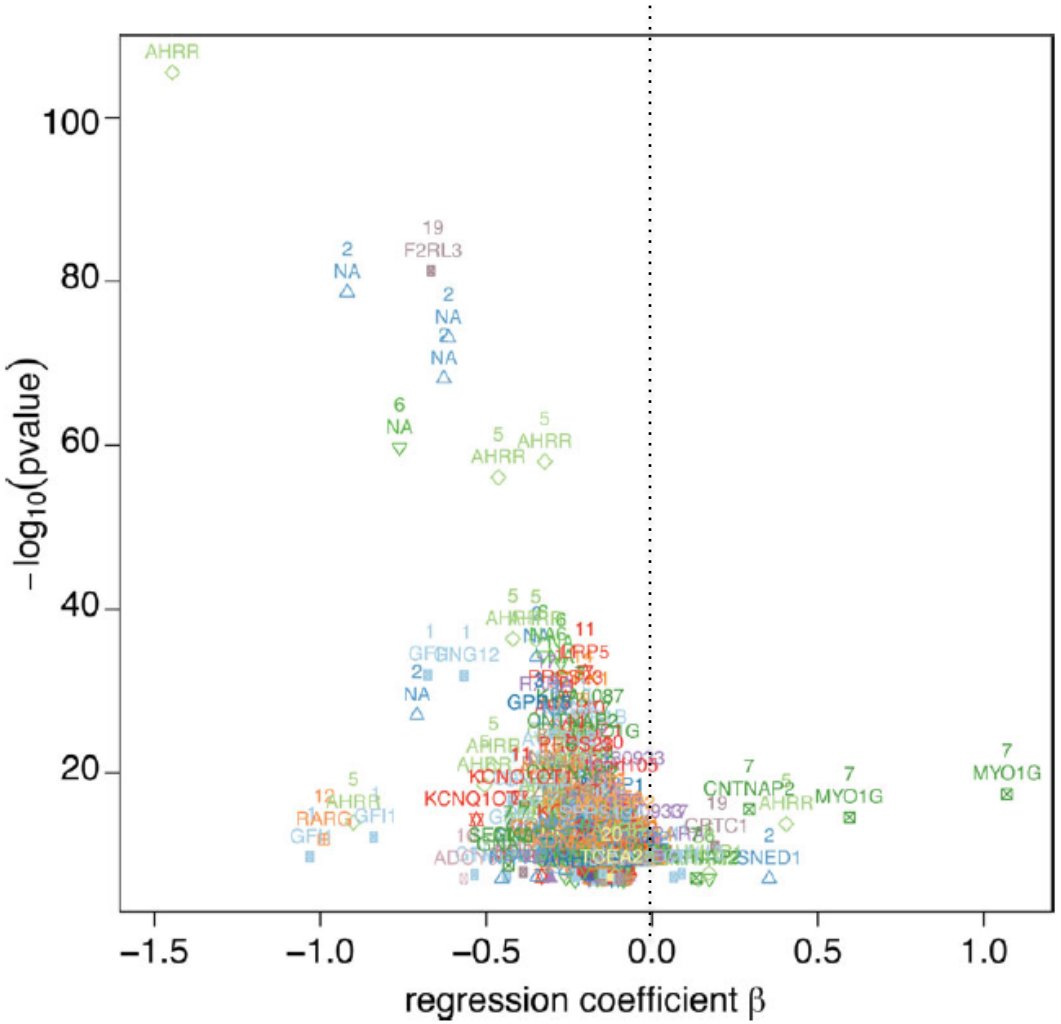
- Male gender
- Tobacco smoking
- Obesity
- Mortality risk and risk of some cancer types (e.g. kidney cancer)

(Dugué et al. Int J Cancer 2018; Fransquet et al. Clin Epigenet 2019)



DNA methylation and tobacco smoking

- Strong differences in methylation, mainly hypomethylation, in DNA from blood between current and never smokers
- The strongest associations are found in CpGs in the genes *AHRR and MYOG1*



EPIC-Italy and NOWAC cohorts N=745
Illumina Infinium 450k methylation array
Current smokers vs never smokers

Guida et. al. Hum Mol. Genet 2015;
Dugué et al. Am J Epidemiol 2018

Tobacco-related DNA methylation changes : functional relevance

***AHRR* gene – Aryl Hydrocarbon Receptor Repressor**

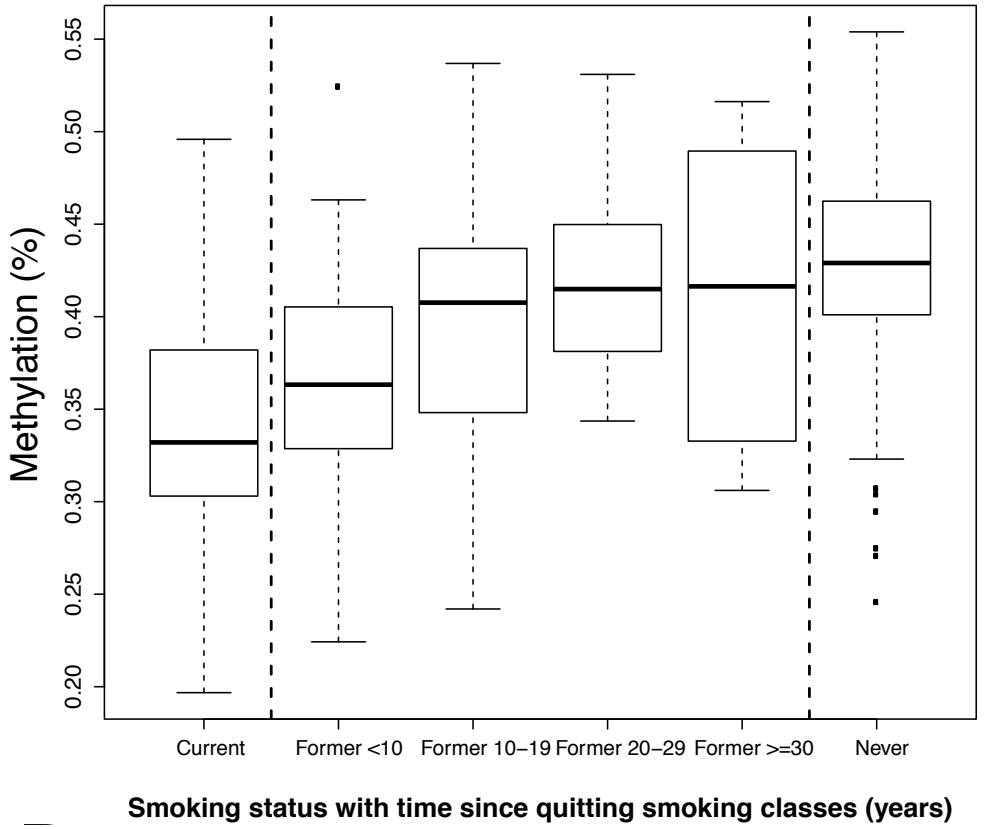
- Role in the detoxification of dioxin and polycyclic aromatic hydrocarbons contained in tobacco smoke
- Around 40% of variance in methylation at this site explained by smoking
- Hypomethylation at this site associated with increased expression

***MYO1G* gene – Myosin immunoglobulin**

- Expression restricted to hematopoietic cells, abundant in T and B lymphocytes

DNA methylation and tobacco smoking

cg05951221 in *AHRR*



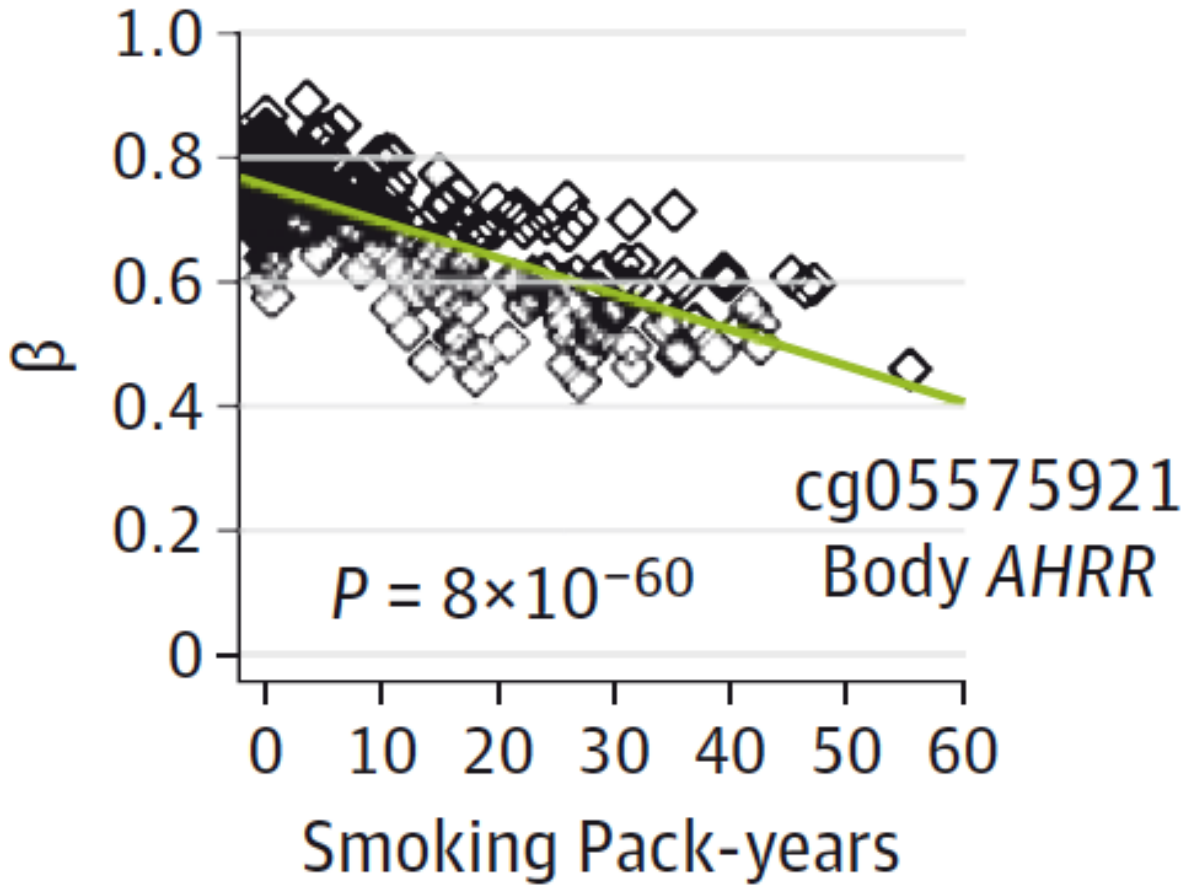
EPIC-Italy and NOWAC cohorts N=745
Illumina Infinium 450k methylation array

- At some CpG sites methylation in former smokers tend to approach the levels in never smokers with increasing time since quitting increases.
- For some other CpG sites, methylation levels remain different from levels in never smokers even after decades since quitting.

Guida et al. Hum Mol Genet 2015 ; Fasanelli et al. Nat Comm 2015

DNA methylation (in saliva) and tobacco smoking

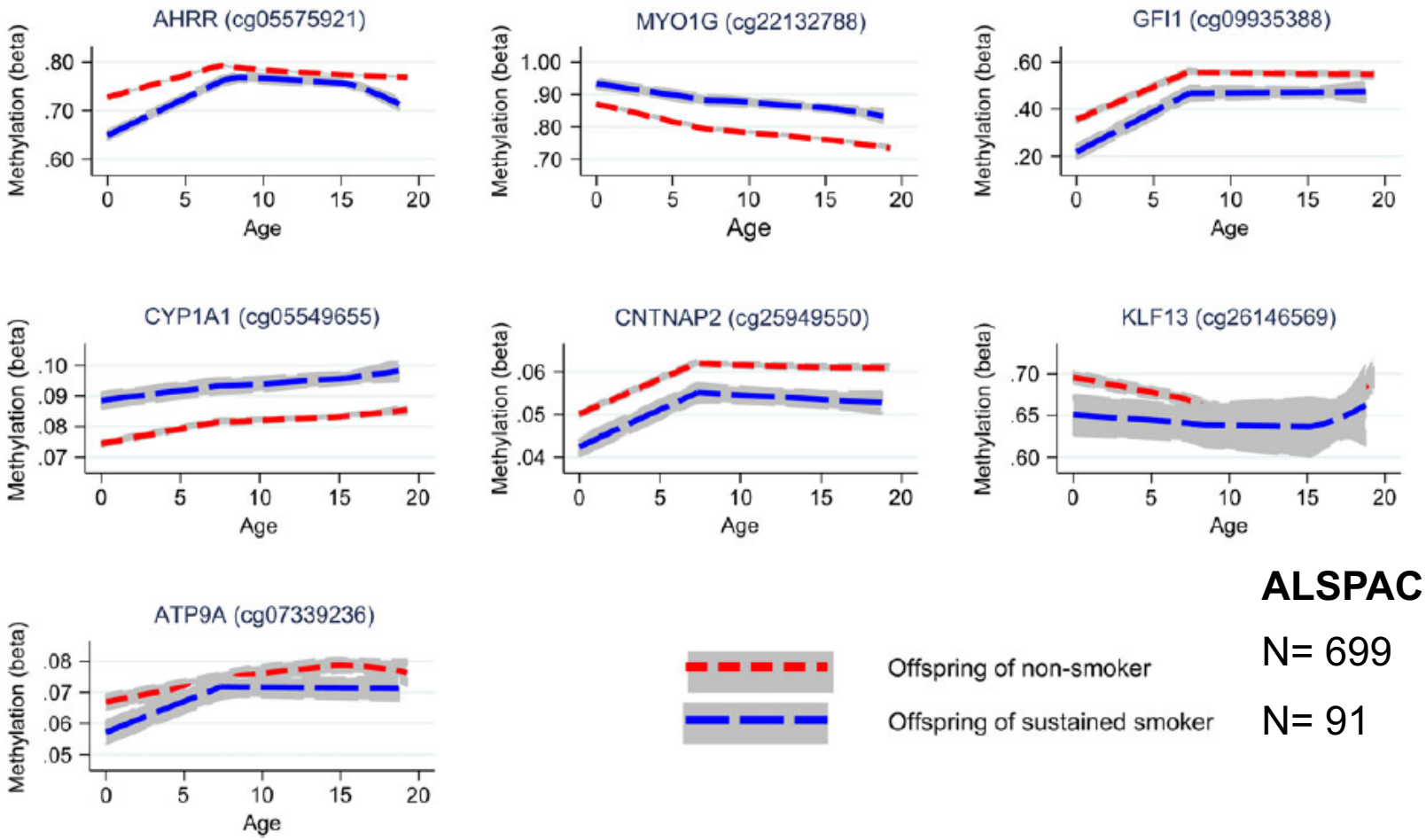
Tobacco related-DNA methylation alterations similar to those found in DNA from blood are observed in DNA from saliva



Teschendorff et al. JAMA Oncol 2015

Tobacco-related DNA methylation changes in early life

Tobacco-related DNA methylation changes identified also in offspring of sustained smokers. Such alterations appear to be **persistent across the first years of life into adolescence**



ALSPAC cohort

N= 699

N= 91

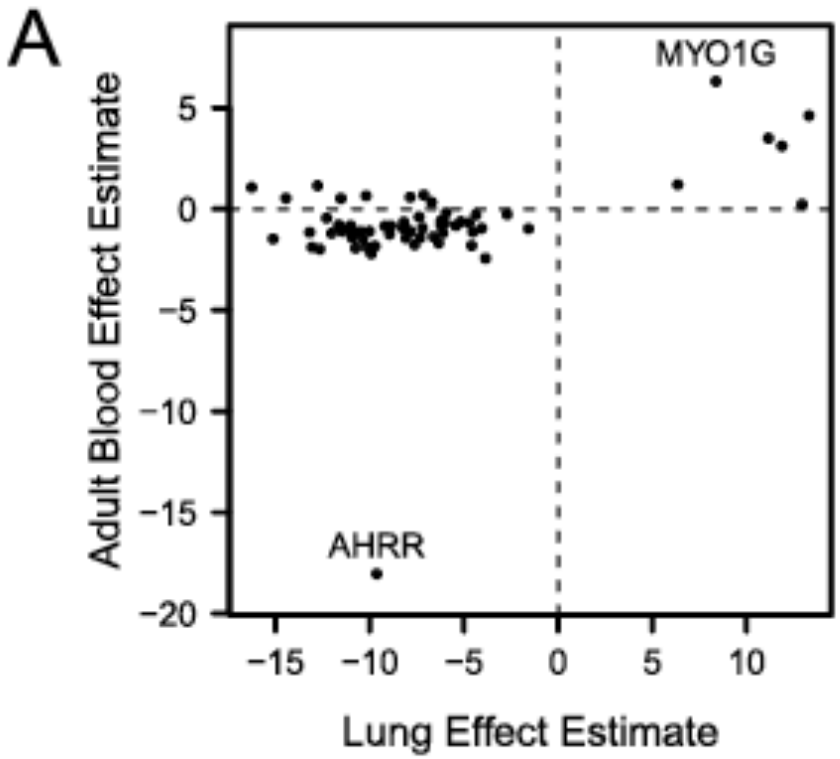
Richmond et al. Hum Mol Genet 2015

Tobacco-related DNA methylation changes in lung cancer

Differences in % DNA methylation by smoking status in lung cancer (N=390)

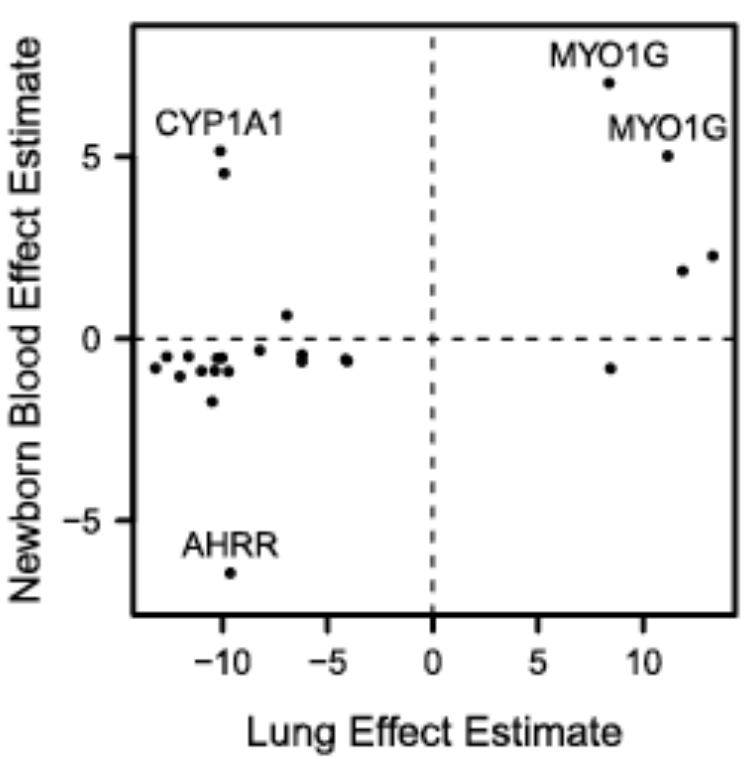
Adults (current – never smokers; $P<10^{-4}$)

blood and lung cancer



Newborn (mother smoking yes – no)

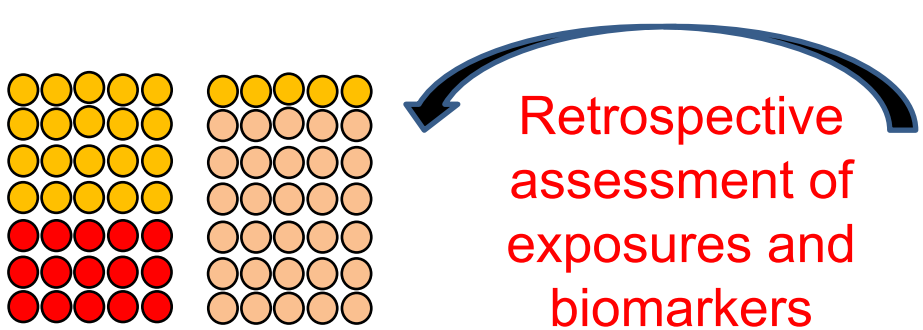
blood vs lung cancer DNA

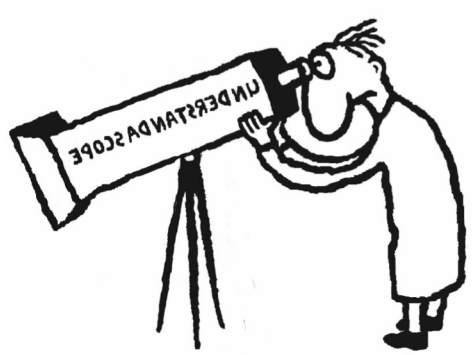


(Bakulski et al. Sci Reports 2019)

Study design : prospective cohort vs case-control study

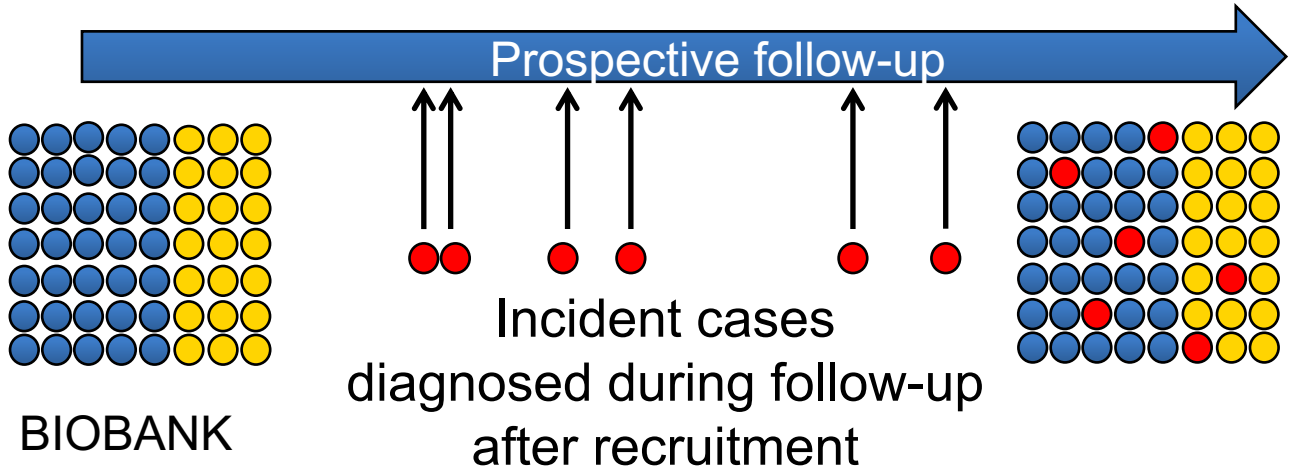
- **Retrospective case-control study**





EXPOSED

- **Prospective cohort studies**



DNA methylation and lung cancer risk

Pooled analysis of 5 case-control studies nested within prospective cohorts (e.g. EPIC-Italy and MCCS, total N pairs = 981) adjusted for self-reported smoking *(Baglietto et al. Int J Cancer 2017):*

- DNA methylation at 6 CpGs associated with lung cancer risk
- Pooled ORs for 1 SD increase in DNAm 0.50 to 0.74 ($10^{-17} < P < 5x10^{-7}$)
- Gain in risk prediction accuracy (overall +3% in AUC; +11% for former smokers)

Methylation scores of smoking associated with lung cancer risk independently of self-reported smoking *(Dugué et al. medRxiv 2021.02.08.21251370)* :

- RR per 1 SD increase in the methylation score = 1.68 (95% CI 1.29 to 2.19)
- Gain in risk prediction accuracy +7% in AUC

DNA methylation and breast cancer risk

Pooled analysis of case-control studies on breast cancer nested within 4 prospective cohorts (MCCS, EPIC-IARC, EPIC-Italy, PLCO, total N pairs = 1 655)

Methylation-based measures	Pooled OR [95%CI]	*P*
Epigenetic aging		
AA-Horvath[b]	**1.02 [0.95–1.10]**	0.59
IEAA-Horvath[b]	**1.03 [0.96–1.11]**	0.40
EEAA[b]	**1.02 [0.93–1.11]**	0.67
AA-Hannum[b]	**1.03 [0.95–1.12]**	0.43
IEAA-Hannum[b]	**1.04 [0.96–1.12]**	0.34
PhenoAge	**1.01 [0.94–1.09]**	0.75
GrimAge	**1.03 [0.94–1.12]**	0.53
Lifestyle-related factors		
BMI methylation score	**1.10 [1.02–1.18]**	0.01
Adjusted BMI score[c]	**1.09 [1.01–1.17]**	0.02
Smoking methylation score	**1.04 [0.97–1.12]**	0.25
Adjusted smoking score[c]	**1.04 [0.97–1.12]**	0.29
Alcohol methylation score	**1.00 [0.93–1.07]**	0.91
Adjusted alcohol score[c]	**0.97 [0.90–1.04]**	0.42

(Dugué et al. Breast. Cancer Res 2022)

Conclusions

- A decade after the first genome-wide association studies of DNA methylation, “epigenetic epidemiology” is still growing as a powerful tool to understand the aetiology of chronic diseases and in particular the role of environmental exposures

- The dynamic (and tissue-specific) nature of DNA represents an excellent opportunity for epidemiological research but requires careful study design and presents challenges.

- Caution : association is not causation ! Need for further evaluation of observed associations (e.g. mendelian randomisation, functional studies …)

Resources on DNA methylation research in epidemiology

- http://www.bristol.ac.uk/integrative-epidemiology/research/epigenetics/
- The EWAS Catalog (https://mrcieu.github.io/software/ewascatalog/)
- Studies on DNA methylation, education and healthy ageing and health inequalities can be found in the Lifepath project (https://www.lifepathproject.eu/)
- https://horvath.genetics.ucla.edu/html/dnamage/