More info on LIFECYCLE online:
lifecycle-project.eu

Report on DNA methylation changes as measures of early-life exposure

LifeCycle report D8.5

Authors:

Sylvain Sebert (UOULU)

Priyanka Choudhary (UOULU)

Justiina Ronkainen (UOULU)

Version 2.0

Delivery date: Month 66

LIFECYCLE Project
Coordination and leadership:

Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands

Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

Document Information

Grant Agreement No.	**733206**
Project Title	Early-life stressors and LifeCycle health (The LifeCycle Project)
Project Start Date	01 January 2017
Work package title	WP8 – DNA methylation and gene expression in life course health trajectories
Related task(s)	Task 8.5
Lead Organisation	University of Oulu
Submission date	16 September 2022
Dissemination Level	Public

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

Table of Contents

List of Figures

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

List of Tables

List of Abbreviations

AA: Age Acceleration
ALSPAC: Avon Longitudinal Study of Parents and Children
AUC: Area Under the Curve
BiB: Born in Bradford
BMI: Body Mass Index
CHOP: EU Childhood Obesity Programme
CI: Confidence Interval
CpG: Cystosine-phosphate-guanine
DNA: Deoxyribonucleic acid
DNAm: DNA methylation
DNAmAge: DNA methylation age
EDEN: Etude des Déterminants pré et post natals précoces du développement psychomoteur et de la santé de l'Enfant (Study on the pre- and early postnatal determinants of child health and development)
EWAS: Epigenome-wide Association Study
FAIR: Findability, accessibility, interoperability, and reusability
GECKO: Groningen Expert Center for Kids with Obesity
GEN R: Generation R
HELIX: Human Early Life Exposome
INMA: Infancia y Medio Ambiente (Environment and Childhood)
MoBa: Norwegian Mother, Father and Child Cohort Study
N: Sample size
NFBC: Northern Finland Birth Cohort
PRS: Poligenic Risk Score
RAINE: Raine Study
RHEA: Mother Child Cohort in Crete
RNA: Ribonucleic Acid
SD: Standard Deviation
SE: Satandard Error
SEM: Structural Equation Modelling
SEP: Socio-economic Position

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

LIFECYCLE Project
Coordination and leadership:

Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands

Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

Executive summary

Description of deliverable: The LifeCycle project has consistently observed evidence that changes in early-life stressors, such as cigarette smoking, maternal and paternal obesity, low socio-economic factors or psychosocial stress, are associated with later health of children. The present report presents the work done in LifeCycle on '*DNA methylation changes as measures of early-life* exposure'. The work reported here had two objectives: we have (i) explored if DNA methylation patterns can be used as markers of early-life stressors and (ii) further developed, harmonized and made epigenetic scores for smoking and biological age available in the LifeCycle catalogue to test reproducibility and their associations with early-life stressors and long-term outcomes.

Findings: Findings can be structured in three main areas:

1) Harmonization and inclusion of DNA methylation scores in the LifeCycle catalogue: The DNA methylation scores for maternal smoking have been harmonized and can be calculated in DataShield to allow federated analyses and support re-use of FAIR data. The on-going projects are currently including multiple cohorts in DataShield and will provide the first evidence from large-scale meta-analyses of DNA methylation scores. We have developed a Bioconductor package that allows computation of several existing DNA methylation adult/childhood and gestational age clocks.

2) Application of DNA methylation scores in evaluating the role of early-life stressors on lifecourse health: The analyses exemplify the predictive impact of DNA methylation and the usefulness of DNA-methylation based scores to study and understand the relationship between early life-stressors and exposure and health during the lifecourse. Specifically, we found that quitting maternal smoking before the third trimester of pregnancy, and possibly lowering smoking dose, may prevent differential DNA methylation in the newborns at CpGs associated with sustained smoking. Thirteen CpGs were identified to be most suitable for inferring smoking versus non-smoking status from blood. Numerous signatures specific to newborns along with many shared between newborns and adults were identified, which were enriched in xenobiotic metabolism pathways.

3) DNA methylation age: Preliminary analysis showed maternal determinants of DNA methylation age estimates. Furthermore, another study exemplified an evidence-based approach to show persistent influences of paternal prenatal adversity mediated through DNA methylation age estimates in the offspring on their psycho-cardiometabolic co-morbidities in adolescence.

Next steps for research and policy: Some key articles arising from this team work have been published and the protocols are open-sourced. Some of the on-going projects are currently including multiple cohorts anlyzed via DataSHIELD to support robust evidence (or absence thereof) from large-scale meta-analyses of the utility of DNA methylation scores and DNA methylation age estimates to objectively measure early stress and their early health outcomes. *Policy context:* The identification of finite set of DNA methylation markers allows

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

a more accurate inference of smoking habit and prenatal exposure from blood, which we envision becoming useful in public health as well as in medical and forensic applications. The detection though DNA methylation may become a very safe (a blood drop can suffice) and cost effective measure of early stressors to reduce errors of recall associated with questionnaires and interviews.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

1. Introduction

Work package 8 of the LifeCycle Project focuses on using the DNA methylation and RNA expression data to define possible biological pathways or molecular signatures underlying the exposure to early-life stressors. The **specific objective of Task 8.5** refers to WP8 objective 3: *To determine whether DNA methylation could be used as an accurate measure of early-life stressors.*

As it is described in the DoA, this task assesses if DNA methylation sites or patterns that are related to early-life stressors can accurately predict those exposed and can explain long term associations. This paradigm can be exemplified by two of our published works by Rauschert *et al* (1) and Parmar *et al* (2), previously reported in D8.3 (for review see also Rauschert et al 2020 (3)).

2. Description of progress and results

Below, we present the progress and results for this deliverable. There are multiple completed and ongoing projects under this task. For each completed project, we present a short description of the work performed. Following that, we briefly describe the ongoing projects

2.1 Harmonization and inclusion of DNA methylation scores in the LifeCycle catalogue

2.1.1 DNA methylation smoking scores

Partner(s) involved: UWA (lead), ERASMUS, ISGLOBAL, UNIVBRIS, BTHFT, UMCG, UOC, INSERM, UOULU, LMU

Summary of findings: The smoking score was developed with machine learning analysis by the UWA and UOULU partners to enable the quantification of DNA methylation present in children for CpGs associated with *in utero* exposure to maternal smoking. The DNAsmoke package was developed using net elastic regression algorithms to provide an easy method of calculating the score for LifeCycle partners with available 450K or EPIC array DNA methylation data. A Shiny app is used to perform the validation of the score (**Figure 1**). The work is available as an open access publication (1).

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

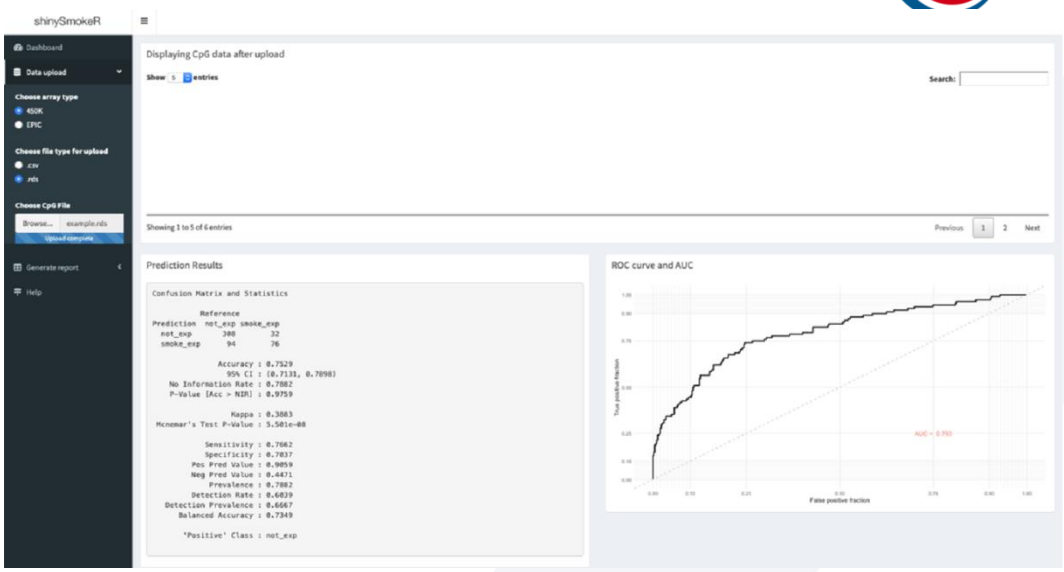

Figure 1. Example of Shiny App output to measure the epigenetic materal smoking score

The smoking score, with supporting variables, was included in the release of the outcome dictionary version 1.3 in Q2 2021. The majority of our participating partners have completed the score calculations and are in the process of loading the data into their respective servers (**Table 1**).

Table 1. Status of the smoking score calculations by participating partners

No	Partners	N
1	*ALSPAC*	Application submitted
2	*BiB*	Application submitted
3	*CHOP*	Approved, ongoing
4	*EDEN*	Approved, ongoing
5	*GECKO*	Approved, ongoing
6	*GEN R*	1,380
7	*INMA*	381
8	*HELIX*	Approved, ongoing
9	*NFBC 1986*	546
10	*RAINE Study*	995

2.1.2 Biological age scores; *methylclock*

Partner(s) involved: ISGLOBAL (lead), UOULU (lead), ERASMUS, UWA

The work described below has been led by partners UOULU and ISGLOBAL. The Bioconductor package used to generate the methylclocks is available as an open access publication (4).

Summary of findings: Ageing is a biological and psychosocial process related to diseases and mortality. It correlates with changes in DNA methylation (DNAm) in all human tissues. Therefore, epigenetic markers can be used to estimate biological age using DNAm profiling across tissues.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

We developed a Bioconductor package that allows computation of several existing DNAm adult/childhood and gestational age clocks. Functions to visualize the DNAm age prediction versus chronological age and the correlation between DNAm clocks are also available as well as other features, such as missing data imputation of cell type estimates, that are required for DNAm age clocks.
Availability and implementation: https://github.com/isglobal-brge/methylclock.
package

2.1.3 Biological age scores in DataShield; DS*methylclock*

Participating LifeCycle partners: UOULU (lead), ERASMUS, ISGLOBAL, UNIVBRIS, BTHFT, UMCG, UOC, NIPH, INSERM, LMU, UWA
Summary of findings: Following this work, we further implemented the dsUploadMethyl function in DataSHIELD, which calculates the DNAm age estimates using the methylclock package and uploads the estimates to DataSHIELD. The function creates and uploads non-repeated DNAm gestational age estimates based on four clocks (Knight(5), Bohlin(6), Mayne(7) and Lee(8)) and yearly repeated DNAmAge estimates based on seven clocks (Horvath(9), Hannum(10), Levine(11), BNN(12), skinHorvath(13), PedBE(14) and TL(15)). In addition, three age acceleration measures are provided for each clock: the raw difference between chronological age and epigenetic age, the residuals obtained after regressing chronological age on epigenetic age without and with adjustment for estimated cell counts. Currently, the following partners are involved: UOULU (lead), ERASMUS, ISGLOBAL, UNIVBRIS, UWA. We expect to include UNITO, UOC, INSERM, NIPH, BTHFT, UMCG, LMU in the future (**Figure 2**).

	Birth				Childhood									Adolescence				Adulthood	
Cohort	Cord blood	Pla-centa	Saliva	...	4yr	5yr	6yr	7yr	8yr	9yr	10yr	11yr	...	15yr	16yr	17yr	...	31yr	46yr
Generation R	1390						490			465									
INMA	385	168			220														
HELIX*							1190												
ALSPAC	900							950						950					
ALSPAC-G2	1000																		
GECKO	255																		
RHEA	100																		
MOBA	1753																		
NFBC1966																		850	850
NFBC1986															530				
CHOP						380						400							
RAINE																1200			
EDEN	164	668																	
BIB	1000																		
NINFEA			139																
TOTAL	6947	836	139		4095									2680				1700	

* HELIX includes samples from INMA, RHEA, EDEN, MoBa, KAUNAS (not primarily included in LifeCycle) and BIB, in addition to the samples mentioned at those specific cohorts.

Figure 2. Availability of DNA methylation data in the LifeCycle cohorts

2.2 Application of DNA methylation scoring in evaluating the role of early life stressors on lifecourse health

Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification (3)

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

Partner(s) involved: UWA

Summary of findings: Machine learning is a sub-field of artificial intelligence, which utilises large data sets to make predictions for future events. Although most algorithms used in machine learning were developed as far back as the 1950s, the advent of big data in combination with dramatically increased computing power has spurred renewed interest in this technology over the last two decades. Within the medical field, machine learning is promising in the development of assistive clinical tools for detection of e.g. cancers and prediction of disease. Recent advances in deep learning technologies, a subdiscipline of machine learning that requires less user input but more data and processing power, has provided even greater promise in assisting physicians to achieve accurate diagnoses (**Figure 3**). Within the fields of genetics and its sub-field epigenetics, both prime examples of complex data, machine learning methods are on the rise, as the field of personalised medicine is aiming for treatment of the individual based on their genetic and epigenetic profiles (**Figure 4**). We now have an ever-growing number of reported epigenetic alterations in disease, and this offers a chance to increase sensitivity and specificity of future diagnostics and therapies. Currently, there are limited studies using machine learning applied to epigenetics. They pertain to a wide variety of disease states and have used mostly supervised machine learning methods.

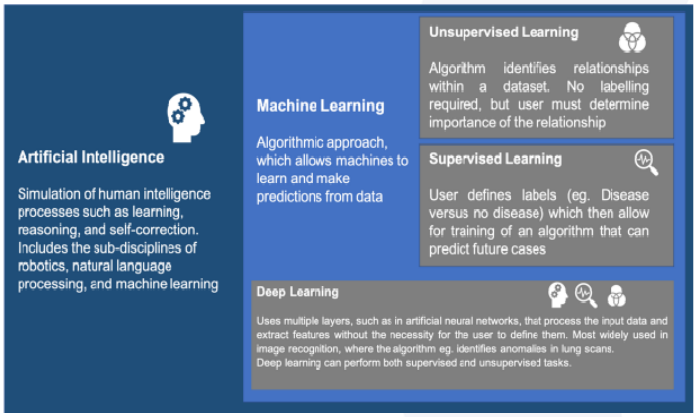

Figure 3. Overview of the field of Artificial Intelligence and its subfield of machine learning

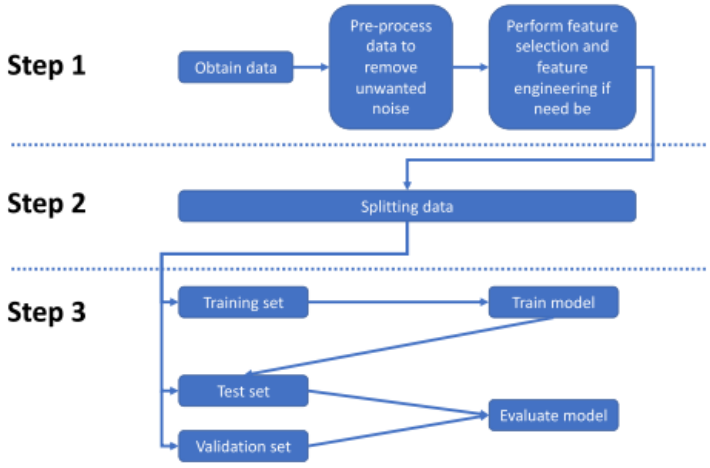

Figure 4. Workflow for applying a machine learning algorithm

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

2.2.1 Timing and dose-specific associations of prenatal smoke exposure with newborn DNA methylation (16)

Partner(s) involved: ERASMUS

Summary of findings: We examined critical periods and doses of maternal smoking during pregnancy in relation to newborn DNA methylation, and associations of paternal smoking with newborn DNA methylation.

This study was embedded in the Generation R Study. We assessed parental smoking during pregnancy using questionnaires. We analyzed associations of prenatal smoke exposure with newborn DNA methylation at 5915 known maternal smoking-related cytosine-phosphate-guanine sites (CpGs) in 1261 newborns using linear regression. Associations with false discovery rate-corrected p-values < .05 were taken forward.

Sustained maternal smoking was associated with newborn DNA methylation at 1391 CpGs, compared with never smoking. Neither quitting smoking early in pregnancy nor former smoking were associated with DNA methylation, compared with never smoking. Among sustained smokers, smoking ≥5, compared with <5, cigarettes/day was associated with DNA methylation at seven CpGs (**Figure 5**). Paternal smoking was not associated with DNA methylation, independent of maternal smoking status.

Our results suggest that CpGs associated with sustained maternal smoking are not associated with maternal smoking earlier in pregnancy or with paternal smoking. Some of these CpGs show dose-response relationships with sustained maternal smoking. The third trimester may comprise a critical period for associations of smoking with newborn DNA methylation, or sustained smoking may reflect higher cumulative doses. Alternatively, maternal smoking limited to early pregnancy and paternal smoking may be associated with DNA methylation at specific other CpGs not studied here.

Our results suggest that quitting maternal smoking before the third trimester of pregnancy, and possibly lowering smoking dose, may prevent differential DNA methylation in the newborns at CpGs associated with sustained smoking. If the relevance of DNA methylation for clinical outcomes is established, these results may help in counseling parents-to-be about quitting smoking.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

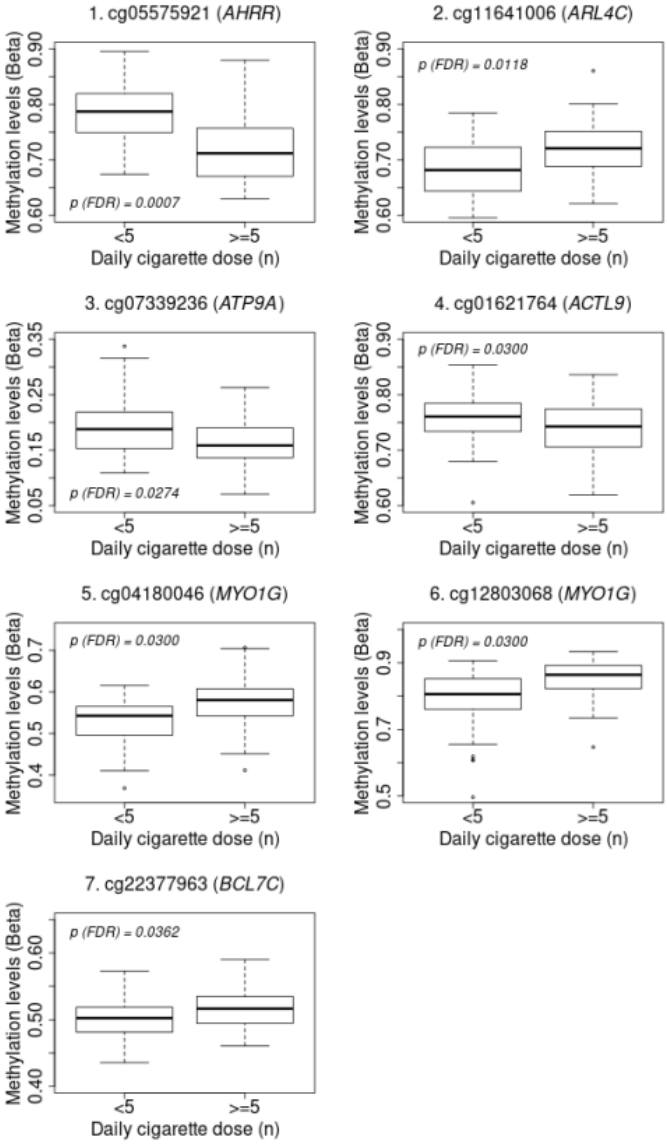

Figure 5. DNA methylation levels in cord blood at the seven CpGs associated with maternal smoking dose

2.2.2 Validated inference of smoking habits from blood with a finite DNA methylation marker set (17)

Partner(s) involved: ERASMUS

Summary of findings: Inferring a person's smoking habit and history from blood is relevant for complementing or replacing self-reports in epidemiological and public health research, and for forensic applications. However, a finite DNA methylation marker set and a validated statistical model based on a large dataset are not yet available. Employing 14 epigenome-wide association studies for marker discovery, and using data from six population-based cohorts (N=3764) for model building, we identified 13 CpGs most suitable for inferring smoking versus non-smoking status from blood with a cumulative Area Under the Curve (AUC) of 0.901 (**Figure 6**). Internal fivefold cross-validation yielded an average AUC of 0.897±0.137, while external model validation in an independent population-based cohort

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

(N=1608) achieved an AUC of 0.911. These 13 CpGs also provided accurate inference of current (average $AUC_{crossvalidation}$ 0.925±0.021, $AUC_{externalvalidation}$ 0.914), former (0.766±0.023, 0.699) and never smoking (0.830±0.019, 0.781) status, allowed inferring pack-years in current smokers (10 pack-years 0.800±0.068, 0.796; 15 pack-years 0.767±0.102, 0.752) and inferring smoking cessation time in former smokers (5 years 0.774±0.024, 0.760; 10 years 0.766±0.033, 0.764; 15 years 0.767±0.020, 0.754). Model application to children revealed highly accurate inference of the true non-smoking status (6 years of age: accuracy 0.994, N=355; 10 years: 0.994, N=309), suggesting prenatal and passive smoking exposure having no impact on model applications in adults. The finite set of DNA methylation markers allows accurate inference of smoking habit, with comparable accuracy as plasma cotinine use, and smoking history from blood, which we envision becoming useful in epidemiology and public health research, and in medical and forensic applications.

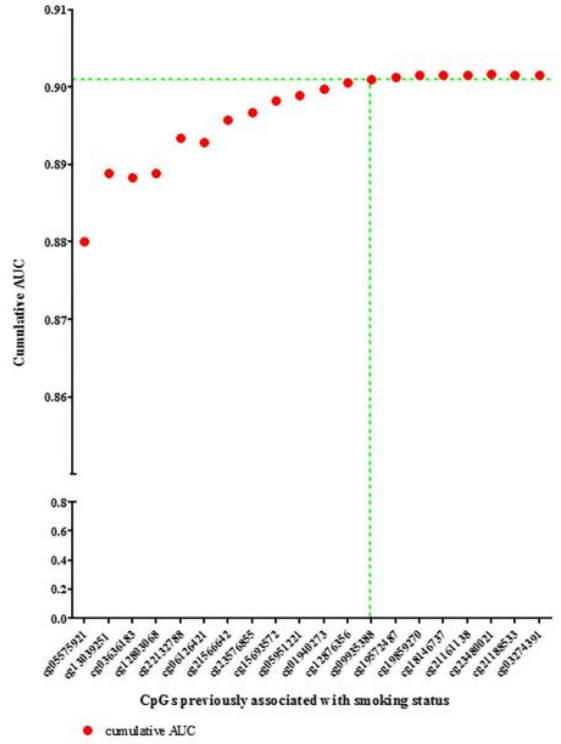

Figure 6. Cumulative AUC profile for smoking habit inference from blood based on the top 20 CpGs.
The 20 CpGs were selected from previous EWASs on smoking habits and were tested in the model-building set (N=3764). Presented is the cumulative contribution of each of the selected 20 CpGs to the model-based smoking habit inference, shown as the AUC plotted against the number of CpGs included in the binary logistic regression model. In the model selection process, first all CpGs were included, and using backward elimination procedures, those with the lowest z-statistic per model were removed one by one. After 13 CpGs, the AUC plateaus; therefore, and by considering the results from Chi squared testing, these 13 CpGs were used for further analyses.

2.2.3 Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking (18)

Partner(s) involved: ERASMUS, ISGLOBAL, UNIVBRIS, UMCG, NIPH, INSERM
Summary of findings: Cigarette smoking influences DNA methylation genome-wide, in newborns from pregnancy exposure and in adults from personal smoking. Whether a

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu
This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

unique methylation signature exists for *in utero* exposure in newborns is unknown. We separately meta-analyzed newborn blood DNA methylation (assessed using Illumina450k Beadchip), in relation to sustained maternal smoking during pregnancy (9 cohorts, 5648 newborns, 897 exposed) and adult blood methylation and personal smoking (16 cohorts, 15907 participants, 2433 current smokers).
Comparing meta-analyses, we identified numerous signatures specific to newborns along with many shared between newborns and adults. Unique smoking-associated genes in newborns were enriched in xenobiotic metabolism pathways (**Figure 7**). Our findings may provide insights into specific health impacts of prenatal exposure on offspring.

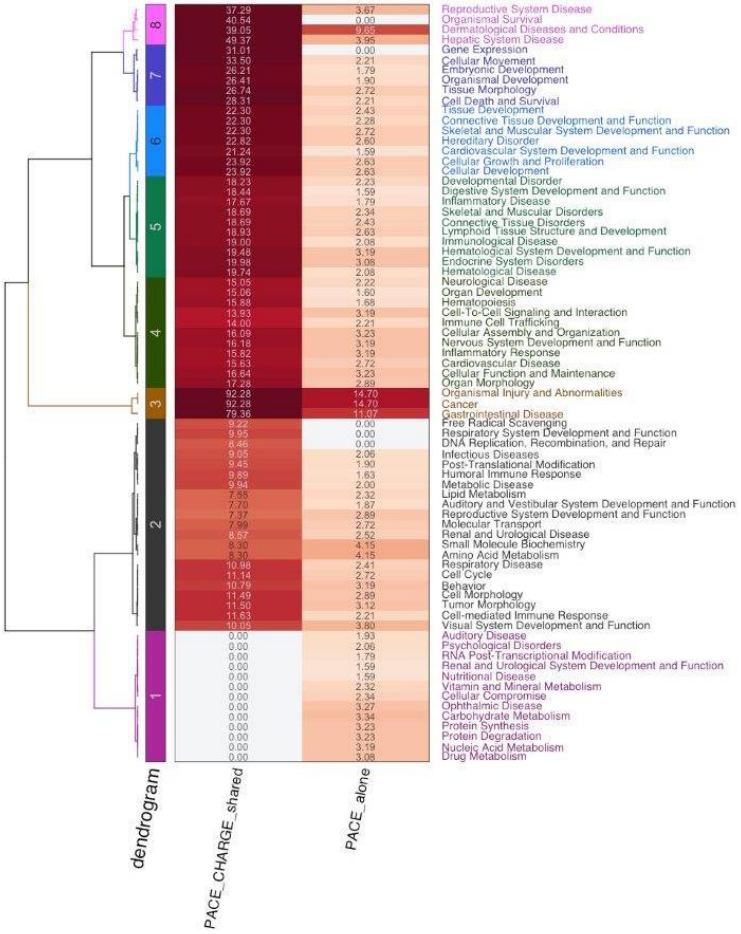

Figure 7. Heatmap of the biological pathways, significant at p-value cutoff of 0.05 in at least one of the two enrichment tests shown, functions enriched with newborn-specific genes, or functions enriched with genes shared between newborns and adults. For each pathway, the color coding is done to show the level of significance (based on p-values). Darker shades indicate higher level of significance.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

2.2.4 Determination of saliva epigenetic age in infancy, and its association with parental socio-economic characteristics and pregnancy outcomes (19)

Partner(s) involved: UNITO

Summary of findings: Epigenetic age acceleration (AA) has been associated with adverse environmental exposures and many chronic conditions. We estimated, in the NINFEA birth cohort, infant saliva epigenetic age, and investigated whether parental socio-economic position (SEP) and pregnancy outcomes are associated with infant epigenetic AA. A total of 139 saliva samples collected at on average 10.8 (range 7–17) months were used to estimate Horvath's DNA methylation age. Epigenetic AA was defined as the residual from a linear regression of epigenetic age on chronological age. Linear regression models were used to test the associations of parental SEP and pregnancy outcomes with saliva epigenetic AA. A moderate positive association was found between DNA methylation age and chronological age, with the median absolute difference of 6.8 months (standard deviation [SD] 3.9). The evidence of the association between the indicators of low SEP and epigenetic AA was weak; infants born to unemployed mothers or with low education had on average 1 month higher epigenetic age than infants of mothers with high education and employment (coefficient 0.78 months, 95% confidence intervals [CIs]: −0.79 to 2.34 for low/medium education; 0.96, 95% CI: −1.81 to 3.73 for unemployment). There was no evidence for association of gestational age, birthweight or caesarean section with infant epigenetic AA (**Figure 8**). Using the Horvath's method, DNA methylation age can be fairly accurately predicted from saliva samples already in the first months of life. This study did not reveal clear associations between either pregnancy outcomes or parental socio-economic characteristics and infant saliva epigenetic AA.

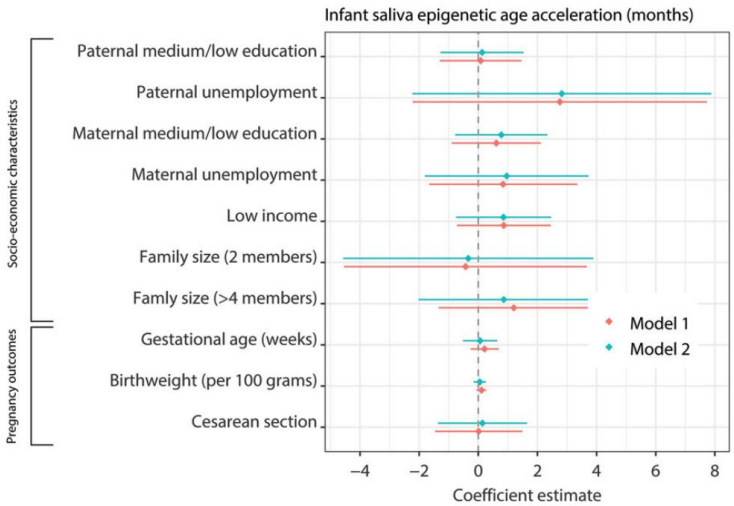

Figure 8. Coefficient estimates with 95% CIs for the association of parental and familial socio-economic characteristics and pregnancy outcomes with infant saliva epigenetic age acceleration (months).
Reference groups: parental high education (university or higher), parental employment, high income (ranked to ≥ 3rd quintile), 3–4 family members in the household, vaginal delivery. Model 1 adjustment (red): child's sex, technical batch (chip), estimated saliva cell count types and child wheezing as a selection factor; Model 2 adjustment (blue): as Model 1 and additional adjustment for maternal age and parity in analyses of socio-economic characteristics, and for maternal age, parity, maternal education and maternal pre-pregnancy BMI in analyses of pregnancy outcomes. Gestational age and birthweight were mutually adjusted in Model 2.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

2.2.5 The effect of timing and cessation of maternal smoking upon the DNA methylation score

Partner(s) involved: UWA (lead), ERASMUS, ISGLOBAL, UMCG, UOULU

Summary of findings: In this project, the aim is to assess if there is an association between the DNA methylation smoking score and the timing of in-utero nicotine exposure during pregnancy. The outcome of this project will help to advise mothers as to when and whether the cessation of smoking is likely to prevent long term detriment to unborn child's DNA.

Status: A preliminary review of accessible partner data has shown promising results to date. As additional partners finalise the smoking score variables and upload the data (see above and **Figure 9**), a more robust assessment of trimester exposure to smoking will be available.

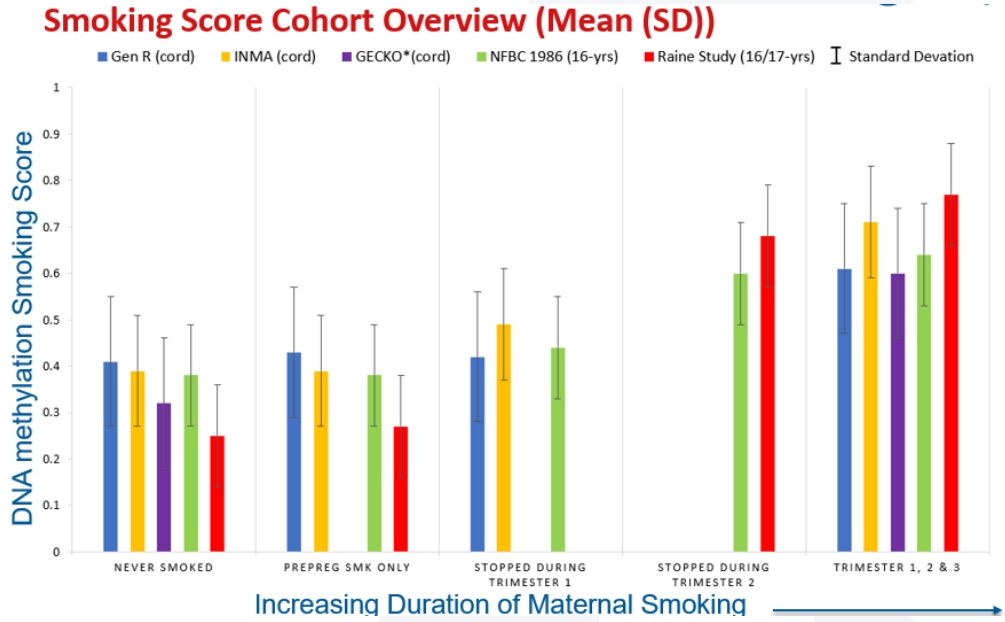

Figure 9. Interim summary of the mean smoking score values by trimester exposure categories.

2.3 DNA methylation age

2.3.1 Maternal and paternal determinants of offspring DNA methylation age

Partner(s) involved: UOULU (lead), ERASMUS, ISGLOBAL, UWA

Summary of findings: The epigenetic age algorithms computed from DNA methylation arrays could represent biomarkers of the ageing process associated with the risk of chronic diseases. Variation in DNA methylation age may also mediate the relationship between early exposures and the risk of later-life diseases. This study analyses the associations of maternal and paternal exposures with DNA methylation age at birth and in childhood to detect possible parental determinants of epigenetic ageing in the offspring. Objective is to evaluate the associations of maternal and paternal smoking, BMI and age at birth with offspring DNA methylation age at birth and at various later ages.

Status: Preliminary analyses have been conducted within INMA, HELIX, Generation R and NFBC1986 and they show associations mainly in maternal exposures and some interesting

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

differences between DNA methylation age estimates (**Figure 10**). Next steps are to generate and upload the DNA methylation estimates to DataSHIELD for all the cohorts and analyse them in DataSHIELD.

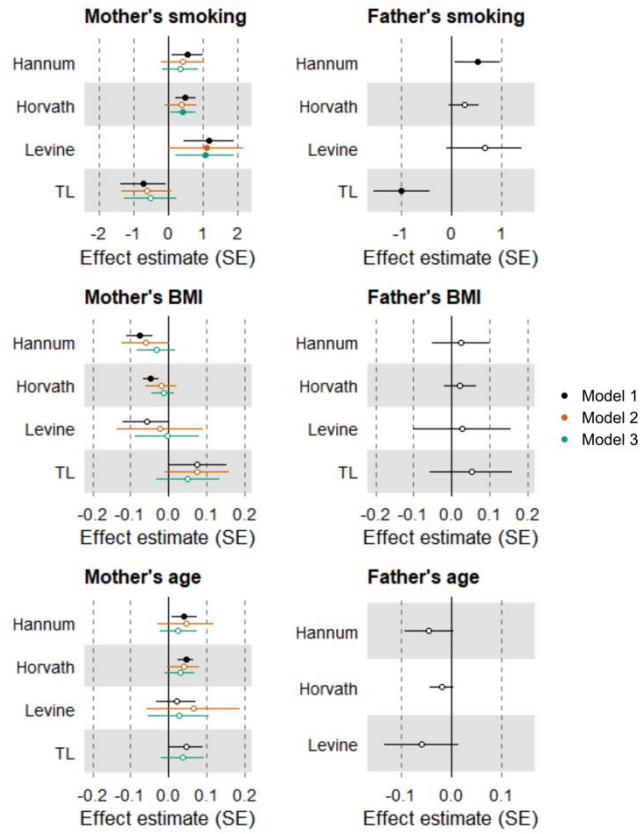

Figure 10. Preliminary meta-analysis of association between parental exposures and selected DNA methylation age estimates in the offspring.
Effect estimates are from meta-analysis of three cohorts participating to the preliminary analysis: HELIX at 8 years of age (N=1144), Generation R at 10 years of age (N=461) and NFBC1986 at 16 years of age (N=508). Model 1 is adjusted for sex and ethnicity, model 2 is adjusted for sex, ethnicity and other parent's exposure and model 3 is adjusted for sex, ethnicity and all other parental exposures presented in the figure. All paternal exposures were not available for preliminary analysis in HELIX and father's age was not available for NFBC1986 so only model 1 is shown for father's exposures. TL, DNA methylation estimate for telomere length.

2.3.2 Do DNA methylation scores mediate the association between maternal stressors and adolescent mental and cardio-metabolic co-morbidity? A path analysis study in NFBC1986 and the Raine study

Partner(s) involved: UOULU (lead), UWA

Summary of findings: Understanding the biological mechanisms that determine co-morbidity patterns in adolescence is important as it may act as hub for distal health outcomes. The aim of the study is to explore latent patterns of prenatal adversity and adolescent psycho-cardiometabolic intermediary traits and further understand the linkage between them *via* epigenetic biomarkers. We used data of mother-child pairs at pregnancy and adolescence at 16-17yr of age from two prospective cohorts: Northern Finland Birth Cohort 1986 (NFBC1986) from Finland and Raine Study from Australia. Exploratory and

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu
This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

confirmatory factor analysis was applied to generate two different latent factor structures: a) paternal adversity using measures from pregnancy and, b) adolescence co-morbidities using psycho-cardiometabolic intermediary traits. Furthermore, two epigenetic scores were included: 1) PRS generated epigenetic smoking score from maternal smoking during pregnancy in our previous study and 2) DNA methylation clock markers (PhenoAge and Telomere Length). Firstly, the relationships between all the measures were assessed using a correlation matrix. Secondly, structural equation modelling (SEM) was used to investigate pathways from paternal adversity to adolescent co-morbidity factors, accounting for the mediating effect of epigenetic markers.

Similar factor structure and correlation patterns were observed for paternal adversity *in utero* and adolescent comorbidities between both cohorts ($P<0.001$). We derived three latent factors for paternal adversity named as: Maternal BMI, Maternal lifestyle and Maternal SOP (socio-obstetric profile) and four latent factors for adolescent co-morbidities named as: BMI, Insulin Triglycerides, Blood Pressure and Mental traits. In the SEM pathways, stronger direct effects of the maternal BMI (NFBC: ß: 0.27; RAINE: ß: 0.39) and SOP (ß: -0.11) factors were observed on adolescent multimorbidity. The indirect effect of the paternal adversity factors through epigenetic markers was mediated by the PhenoAge methyl clock (NFBC: ß: 0.04; RAINE: ß: 0.14), showing consistent influences in both cohorts ($P<0.001$, **Figure 11**).

The present study exemplifies an evidence-based approach validated in two cohorts to show persistent influences of paternal adversity mediated through epigenetic changes in the offspring on their co-morbidities in adolescence.

Status: Manuscript submitted for publication.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

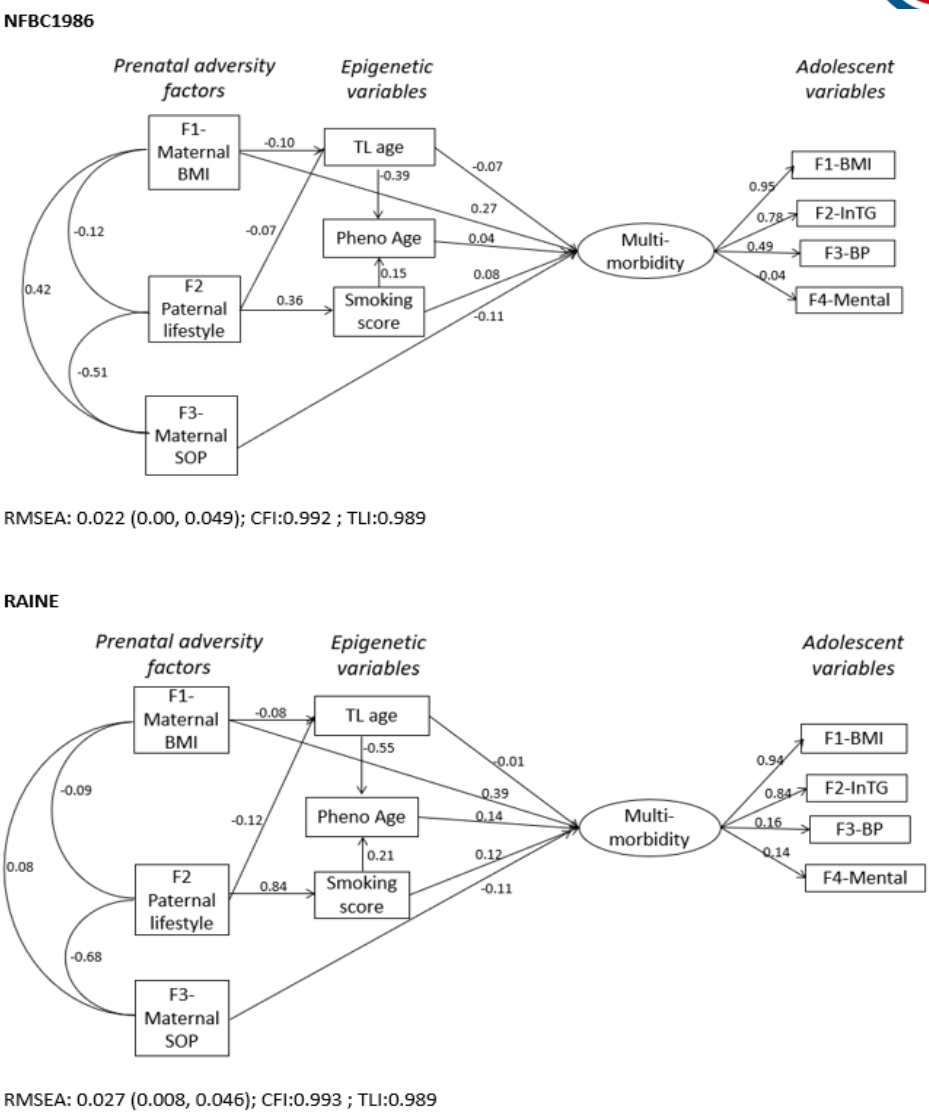

Figure 1. Structural Equation Model (SEM) pathways in NFBC1986 and RAINE study describing the association between maternal stressors and adolescent co-morbidities.

F1-4: Factor; BMI: Body Mass Index; BP: Blood Pressure; InTg: Insulin Triglyceride; SOP: Socio-obstetric Profile; and TL: Telomere Length.

3. Conclusion

These analyses as part of Task 8.5 exemplify the predictive impact of DNA methylation and the usefulness of DNA methylation based scores to study and understand the relationship between early life-stressors and exposure and health during the life-course. Importantly, the DNA methylation scores for maternal smoking have been harmonized and can be calculated in DataShield to allow federated analyses and support re-use of FAIR data. Some key articles arising from this team work have been published and the protocols are open-sourced. The on-going projects are currently including multiple cohorts in DataShield and will soon provide the first evidence from large-scale meta-analyses of DNA methylation scores.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

4. Contribution of partners

The contribution of partner for each subproject is detailed below. Lead partners were responsible for developing the proposals and the analytical plan and generating the scripts for creating the scores. The partners contributed by generating the scores to add to the LifeCycle catalogue or to provide aggregated summary statistics for meta-analysis. In case of published articles, all partners reviewed and agreed upon the publication of the final version of the manuscripts. The report was jointly prepared by the leaders of the respective projects.

- **UOULU**: Leader of the deliverable 8.5. Led the following studies: 'Biological age scores; methylclock package,' 'Maternal and paternal determinants of offspring DNA methylation age' and, 'Do DNA methylation scores mediate the association between maternal stressors and adolescent mental and cardio-metabolic co-morbidity? A path analysis study in NFBC1986 and the Raine study'. Participated in the study on DNA methylation smoking score.
- **UWA**: Led the following studies: 'DNA methylation smoking scores', 'Machine learning and clinical epigenetics', and 'The effect of timing and cessation of maternal smoking on DNA methylation score'. Participated in the following studies: 'Maternal and paternal determinants of offspring DNA methylation age' and, 'Do DNA methylation scores mediate the association between maternal stressors and adolescent mental and cardio-metabolic co-morbidity? A path analysis study in NFBC1986 and the Raine study'.
- **ERASMUS**: Led the following studies: 'Timing- and Dose-Specific Associations of Prenatal Smoke Exposure With Newborn DNA Methylation', ' Validated inference of smoking habits from blood with a finite DNA methylation marker set'. Participated in the following studies: DNA methylation smoking scores, Biological age scores; methylclock, 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking', 'The effect of timing and cessation of maternal smoking upon the DNA methylation score', and 'Maternal and paternal determinants of offspring DNA methylation age'.
- **UNITO**: Led the study on 'Determination of saliva epigenetic age in infancy, and its association with parental socio-economic characteristics and pregnancy outcomes'.
- **ISGLOBAL**: Participated in the following studies: 'DNA methylation smoking scores', 'Biological age scores; methylclock', 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking', 'The effect of timing and cessation of maternal smoking upon the DNA methylation score', and 'Maternal and paternal determinants of offspring DNA methylation age'.
- **UMCG**: Participated in the following studies: The effect of timing and cessation of maternal smoking upon the DNA methylation score' and 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking.'
- **UNIVBRIS**: Participated in 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking.'

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

- **NIPH**: Participated in the 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking'.
- **INSERM**: Participated in the 'Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking'.

5. Deviations from the original plan

This deliverable has been fulfilled fully in line with the original plan as state in the Grant Agreement.

6. Dissemination activities

Findings generated under task 8.5 have been published in peer-reviewed journals and can be accessed online. Findings have been presented as oral and poster presentation at international conferences and meetings. Further publications generated from the data and work in progress will be deposited in relevant repositories in accordance with the principles of open science.

7. References

1. Rauschert S, Melton PE, Heiskala A, Karhunen V, Burdge G, Craig JM, et al. Machine learning-based DNA methylation score for fetal exposure to maternal smoking: development and validation in samples collected from adolescents and adults. Environ Health Perspect. 2020;128(9):097003.
2. Parmar P, Lowry E, Cugliari G, Suderman M, Wilson R, Karhunen V, et al. Association of maternal prenatal smoking GFI1-locus and cardio-metabolic phenotypes in 18,212 adults. EBioMedicine. 2018;38:206-16.
3. Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clin Epigenetics. 2020;12(1):51-4.
4. Pelegí-Sisó D, de Prado P, Ronkainen J, Bustamante M, González JR. methylclock: a Bioconductor package to estimate DNA methylation age. Bioinformatics. 2021;37(12):1759-1760.
5. Knight AK, Craig JM, Theda C, Baekvad-Hansen M, Bybjerg-Grauholm J, Hansen C, et al. An Epigenetic Clock for Gestational Age at Birth Based on Blood Methylation Data. Genome Biol. 2016;17(1):206.
6. Bohlin J, Håberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, et al. Prediction of Gestational Age Based on Genome-Wide Differentially Methylated Regions. Genome Biol. 2016;17(1):207.
7. Mayne BT, Leemaqz SY, Smith AK, Breen J, Roberts CT, BiancoMiotto T. Accelerated Placental Aging in Early Onset Preeclampsia Pregnancies Identified by Dna Methylation. Epigenomics. 2017;9 (3):279–89.
8. Lee Y, Choufani S, Weksberg R, Wilson SL, Yuan V, Burt A, et al. Placental Epigenetic Clocks: Estimating Gestational Age Using Placental Dna Methylation Levels. Aging (Albany NY). 2019;11(12):4238.
9. Horvath S. DNA Methylation Age of Human Tissues and Cell Types. Genome Biol. 2013;14(10):3156.

LIFECYCLE Project Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).

10. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Molecular Cell. 2013;49(2):359–67.
11. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An Epigenetic Biomarker of Aging for Lifespan and Healthspan. Aging (Albany NY). 2018;10(4):573.
12. Alfonso G, Gonzalez JR. Bayesian neural networks for the optimisation of biological clocks in humans. bioRxiv. 2020.
13. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic Clock for Skin and Blood Cells Applied to Hutchinson Gilford Progeria Syndrome and Ex Vivo Studies. Aging (Albany NY). 2018;10(7):1758.
14. McEwen LM, O'Donnell KJ, McGill MG, Edgar RD, Jones MJ, MacIsaac JL, et al. The PedBE Clock Accurately Estimates Dna Methylation Age in Pediatric Buccal Cells. Proc Natl Acad Sci USA. 2020;117(38):23329-35.
15. Lu AT, Seeboth A, Tsai PC, Sun D, Quach A, Reiner AP, et al. DNA methylation-based estimator of telomere length. Aging (Albany NY). 2019;11(16):5895-923.
16. Monasso GS, Jaddoe VWV, de Jongste JC, Duijts L, Felix JF. Timing- and Dose-Specific Associations of Prenatal Smoke Exposure With Newborn DNA Methylation. Nicotine Tob Res. 2020;22(10):1917-22.
17. Maas SCE, Vidaki A, Wilson R, Teumer A, Liu F, Van Meurs JBJ, et al. Validated inference of smoking habits from blood with a finite DNA methylation marker set. Eur J Epidemiol. 2019;34(11):1055-74.
18. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics. 2019;11(13):1487-1500.
19. Popovic M, Fiano V, Isaevska E, Moccia C, Trevisan M, Rusconi F, et al. Determination of saliva epigenetic age in infancy, and its association with parental socio-economic characteristics and pregnancy outcomes. J Dev Orig Health Dis. 2021;12(2):319-27.

LIFECYCLE Project
Coordination and leadership:
Erasmus MC
Generation R (Na-2918)
PO box 2040
3000 CA Rotterdam
The Netherlands
Phone: +31 (0)10 704 3405
Email: lifecycle@erasmusmc.nl
Web: lifecycle-project.eu

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle).