More info on LIFECYCLE online:
lifecycle-project.eu

Report on the DNA methylation loci that mediate the relationships of early-life stressors with asthma and chronic obstructive respiratory disease

LifeCycle report D5.4

Authors:

Johanna L. Nader (NIPH)

Chris Thio (UMCG)

Jennifer R. Harris (NIPH)

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	WP5 - Early-life stressors and respiratory health life course trajectories
Related task(s)	Task 5.4 - Identify DNA methylation loci that mediate the relationships of early-life stressors with asthma and chronic obstructive respiratory disease
Lead Organisation	NIPH
Submission date	19 September 2022
Dissemination Level	Public

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALSPAC: Avon Longitudinal Study of Parents and Children
BiB: Born in Bradford
BMI: body mass index
Bp: base pair
CHOP: EU Childhood Obesity Programme
COPD: chronic obstructive pulmonary disease
DALYs: disability-adjusted life years
DMR: differentially methylated region
DNA: deoxyribonucleic acid
DNAm: DNA methylation data
EDEN: Etude des Déterminants du développement et de la santé de l'Enfant
EWAS: epigenome-wide association studies
FEF: forced expiratory flow
FEV: forced expiratory volume
FVC: forced vital capacity
GECKO: Groningen Expert Center for Kids with Obesity Drenthe cohort
Generation R: the Generation R Study
GWAS: genome-wide association studies
INMA: Infancia y Medio Ambiente
Kbp: Kilobase pair
LRTI: lower respiratory tract infection
Mbp: Megabase pair
MeDALL: Mechanisms of the Development of ALLergy
MoBa: The Norwegian Mother, Father and Child Cohort Study
NFBC1966/1986: Northern Finland Birth Cohorts
PACE: Pregnancy And Childhood Epigenetics
Rhea: Rhea Mother & Child Cohort Study
SNP: single nucleotide polymorphism
The Raine Study: The Western Australian Pregnancy Cohort Study
URTI: upper respiratory tract infection
WHO: World Health Organization

Executive summary

The aim of task 5.4 in the LifeCycle project is to identify deoxyribonucleic acid (DNA) methylation loci that mediate causal relationships between exposures and stressors in early life with later respiratory health and disease. In collaboration with the Pregnancy And Childhood Epigenetics (PACE) Consortium and WP8 in LifeCycle (DNA methylation and gene expression in life course health trajectories), this task brings together epigenetic and phenotypic data from multiple pregnancy and child cohorts around Europe and Australia, in order to perform epigenome-wide association studies (EWAS) meta-analyses to elucidate novel and existing early-life markers of respiratory disease in childhood, adolescence, and adulthood. This report outlines work done in the LifeCycle research consortium to map the intersection of available DNA methylation data and respiratory health data in the nine cohorts participating in this task and describes a novel mendelian randomization study on the effects of DNA methylation related to birthweight on respiratory disease outcomes. We also present work conducted under this deliverable that fulfilled though the participation of LifeCycle in parallel PACE studies that investigate associations between epigenetic markers and respiratory outcomes. Together these studies provide important insights into the potential mediating role of epigenetic modification between early exposures and respiratory health throughout later in life. This report outlines the potential epidemiological, diagnostic, and clinical relevance of these as well as future studies that make use of the unique epigenetic data resource made possible through the LifeCycle consortium.

1. Introduction

Respiratory diseases are one of the leading causes of morbidity and mortality around the world, accounting for more than 112 million disability-adjusted life years (DALYs) in the latest global health report (1). The two most common and debilitating of these are asthma and chronic obstructive pulmonary disease (COPD), causing 22.8 and 81.6 million DALYs annually, respectively. In children, asthma is the leading cause of chronic non-communicable disease in childhood, and more than 1000 deaths per day due to asthma have been recorded around the world since the early 21st century (2). COPD is the third leading cause of death in adults, with more than 3 million deaths reported each year between 1990 and 2010 (3). Given the severe impact these respiratory diseases impose on human health, identifying important etiological features that drive risk and prognosis is a key priority to informing and improving global public health.

Asthma and COPD prevalence have been rising rapidly in recent years (4), and scientific efforts to elucidate their environmental, biological, and genetic origins are increasing as well. These studies span demographically, ethnically, and geographically diverse populations, and include epidemiological, clinical, molecular, and sociological features of disease risk, progression, and prognosis. Some of the most important well-established environmental risk factors that have been identified and described for asthma include air pollution (5), exposure to tobacco smoke (6), and work environment (7). Causal associations between asthma and psychosocial conditions have also been established, including chronic stress (8), anxiety (9), depression (10), and parental common mental disorders (11). Important biological risk factors include pre-existing atopic conditions (12), the gastrointestinal microbiome (13), obesity (14), and prematurity (15). In addition to important environmental and biological risk factors, the most significant contributor to asthma risk stems from genetic heritability (16). Heritability estimates as high as 95% have been reported from human studies (17), and hundreds of asthma-related candidate genes have been identified following genome-wide association studies (GWAS) (18).

Like asthma, a broad range of diverse risk factors have also been well-established for COPD, many of which are overlapping with asthma and other respiratory conditions. The single greatest risk factor for COPD is smoking tobacco, with more than two-thirds of global COPD cases linked to current or former smoking (19). Simultaneously, this highlights that approximately one-third of COPD cases occur due to other environmental, biological, and genetic factors in non-smokers. Work exposure is one of the most important secondary causes of COPD, particularly exposure to biomass fuels, which simultaneously implicates socioeconomic status and geographic location as relevant risk factors as well (20). COPD is also a well-known complication in chronic viral and bacterial infections, such as the human immunodeficiency virus (21) and tuberculosis (22). Genetic factors play an important role, with several genomic regions identified that significantly impact susceptibility and outcome in COPD patients (23). Another important consideration in understanding COPD risk and prognosis is its common co-occurrence with asthma (24), which means that important risk factors for either disease become implicitly intertwined.

In recent years, causal research has shifted beyond traditional epidemiological factors to focus on epigenetic origins and drivers of disease, the process whereby environmental exposures can affect gene activity and function. Emerging research shows that several differentially methylated genetic regions are linked to genes associated with COPD and asthma (25, 26). These studies demonstrate that epigenetics may play an important part in influencing disease risk in cases where obvious risk factors are otherwise. Previous meta-analyses of data from pregnancy and birth cohorts identified several significant asthma-related loci that are differentially-methylated in asthma cases versus controls (27), and further reported that differential methylation in blood is largely replicated in other tissues, such as respiratory epithelium (28). DNA methylation of specific immune-related genes such as receptor tyrosine kinases and hexokinases has also been linked to significant increases in asthma risk throughout childhood (29) and adolescence (30). This demonstrates the usefulness of these types of epigenetic association studies in identifying genetic loci that can be used as diagnostic markers of disease.

In addition to association analyses, more recent research has begun to focus on methylation as a mediating factor between early life exposures and health outcomes. Parental smoking for example has been associated with differential methylation of genes that are linked to childhood asthma (31, 32). Exposure to air pollution has similarly been found to increase methylation at genetic loci that are associated with asthma and other respiratory health outcomes (33). The mediating effect of epigenetic changes between genetic variants and COPD in adulthood has also been demonstrated (34, 35). This presents new opportunities to address lifestyle factors that affect later health via epigenic modification. Early prenatal intervention studies have proven successful at mitigating the mediating effects of DNA methylation (36). An important component to building on this emerging area of epigenetic research is the inclusion of more diverse populations of larger sample sizes, as well as extending the variety of phenotypic exposures and respiratory outcomes that can be analysed. This makes it possible to identify more reliable epigenetic mediators and exposure factors that can be used to design and develop novel approaches to respiratory disease prevention and treatment.

The LifeCycle consortium helps to advance this agenda by bringing together pregnancy and child cohorts from Europe and Australia to study how exposures in early life affect health and wellbeing throughout the life cycle. One of the main benefits of large scientific consortia is the ability to increase sample size, and thus statistical power, in situations where available data are limited in individual cohorts. This reflects the situation with LifeCycle cohorts, where collaboration is essential for epigenetic studies of respiratory outcomes. Several LifeCycle cohorts have collected DNA methylation data on cohort participants at birth, during childhood and adolescence, and during adulthood, in addition to a wide range of retrospective and prospective data on physical and mental health. This provides opportunities to perform meta-analyses to study the potential mediating effect of epigenetic changes on the relationship between early life stressors and respiratory health outcomes throughout the lifecycle. Identification of novel epigenetic markers that may mediate the risk and progression of respiratory diseases like asthma and COPD provides important

insights into issues of aetiology and help identify potential directions for targeting novel intervention strategies and treatment.

2. DNA methylation and respiratory health data in the LifeCycle project

Nine LifeCycle cohorts with DNA methylation and phenotypic data on respiratory health participated in Task 5.4: Avon Longitudinal Study of Parents and Children (ALSPAC/ALSPAC G2) at the University of Bristol (UNIVBRIS), United Kingdom; EU Childhood Obesity Programme (CHOP, Germany/Italy/Spain/Poland/Belgium) at udwig-Maximilians-Universität (LMU), Germany; Groningen Expert Center for Kids with Obesity Drenthe cohort (GECKO Drenthe cohort), University Medical Center Groningen (UMCG), The Netherlands; the Generation R Study (Generation R/Generation R Next), Erasmus University Medical Center (ERASMUS), The Netherlands; Infancia y Medio Ambiente (INMA), Barcelona Institute for Global Health (ISGlobal), Spain; The Norwegian Mother, Father and Child Cohort Study (MoBa), Norwegian Institute of Public Health (NIPH), Norway; Northern Finland Birth Cohorts (NFBC1966 and NFBC1986), University of Oulu (UOULU), Finland; and The Raine Study (The Western Australian Pregnancy Cohort Study), University of Western Australia (UWA), Australia.

A first step was to map the available DNA methylation data (DNAm) initially involved illustrating the number of participants per cohort for which epigenetic data are available at birth, childhood (0-12 years), adolescence (12-18 years) and adulthood (18+ years). This is essential for project planning and revealed that most (six) of the nice participating cohorts contain DNA methylation data at birth and/or childhood, three had DNA methylation data in adolescence, and only a single cohort (NFBC1966) has data in adulthood (**Table 1**; **Figure 1**). Three of the cohorts additionally have DNA methylation data measured longitudinally in at least two separate age groups (ALSPAC, Generation R and INMA), with a total of 10,688 sets of methylation data measurements available across the nine cohorts involved in Task 5.4 of the LifeCycle project. Many of these cohorts continue to develop their epigenetic programs of research and are expanding the sample sizes with epigenetic data. This is paving the way for continued collaborations between these groups.

Table 1. DNA methylation data (DNAm) across cohorts and age groups in Task 5.4 of the LifeCycle Project

Cohort	DNAm at birth (N)	DNAm childhood (N)	DNAm adolescence (N)	DNAm adulthood (N)
ALSPAC	900	950	950	
CHOP		521		
GECKO	255			
Generation R	1381	613		
INMA	385	201		
MOBA	1753			
NFBC1966				807
NFBC1986			593	
RAINE			1192	

Extensive harmonization of respiratory data was performed under Task 5.1 in the LifeCycle project, and spanned a range of symptoms, conditions, and diseases (for detailed harmonization protocol, visit https://lifecycle-project.eu/for-scientists/guides-manuals/). The main respiratory outcomes encompassed some of the most frequent public health problems, and included conditions such as asthma, wheezing, lower/upper respiratory tract infections, and COPD. In addition to providing general overviews of past and current medical histories of these conditions in offspring and their parents, harmonization produced repeated measures of respiratory outcomes such as wheezing, lower and upper respiratory tract infections, and asthma at yearly intervals from birth until adulthood (0-18 years).

To demonstrate the utility of the combined data resource provided through Task 5.4 for projects aiming to study the mediating effects of DNA methylation on respiratory outcomes, an overview of the intersection between epigenetic and harmonized respiratory data was developed. This is an important component to understanding the potential range of cases across outcomes and ages that are available to study exposure and outcome relationships.

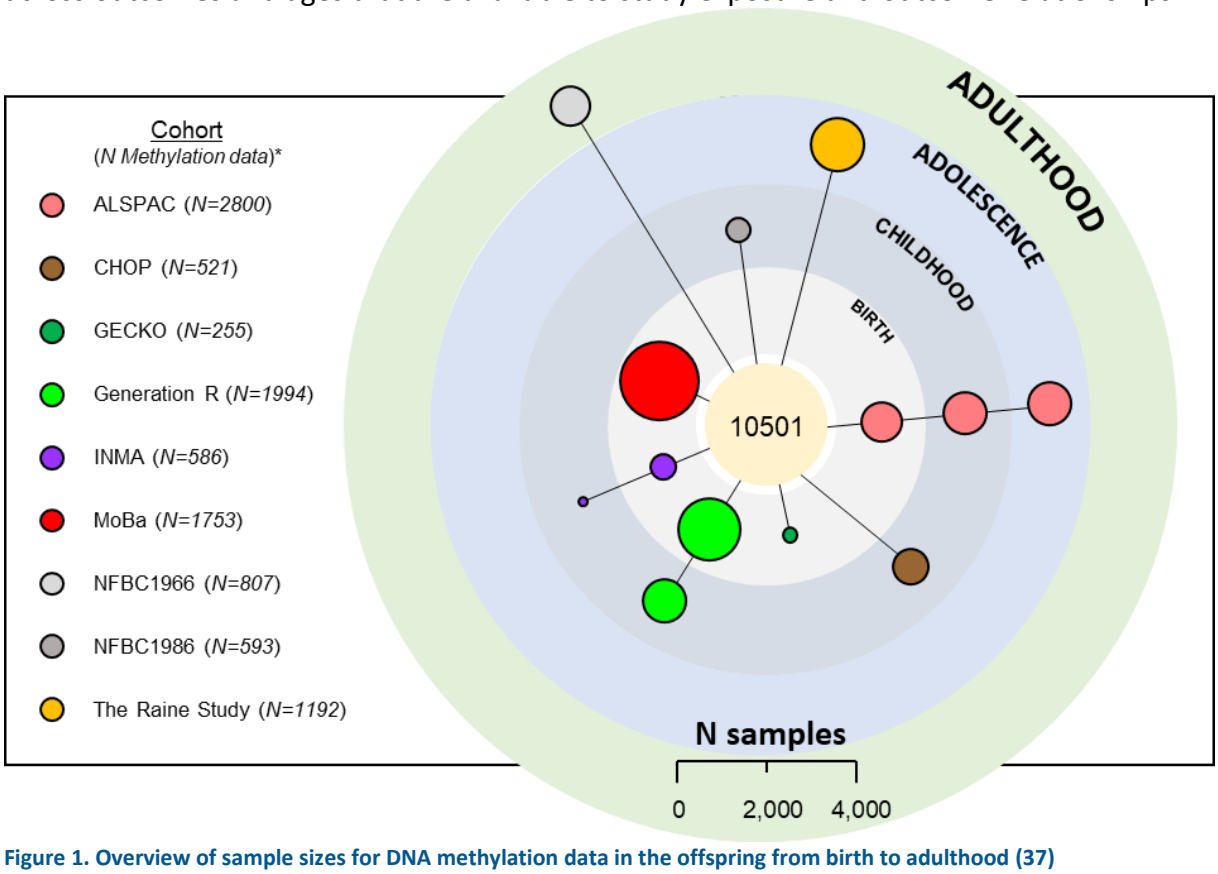

Figure 1. Overview of sample sizes for DNA methylation data in the offspring from birth to adulthood (37)
Comprises cohorts participating in Task 5.4 of the LifeCycle project as of January 2022. Circle sizes are proportionate to the DNA methylation sample sizes as indicated in the scale at the bottom of the figure.

This is a valuable overview that has not been compiled previously and is essential for the planning of collaborative international projects. Most of the cohorts participating in Task 5.4 have participants with DNA methylation data that also comprise cases with current or past histories of asthma, wheezing, inhalant allergies, and upper/lower respiratory tract infections (**Figures 2-7**; https://catalogue.lifecycle-project.eu/). However, almost all cohorts have certain gaps or limitations in terms of case numbers or relevant ages for which data are

available. Generation R, for example, only has lower respiratory tract infection (LRTI) cases with DNAm data from ages 0-6 years, while ALSPAC contains these respective data for ages 6-12 years (**Figure 6**); individually, these cohorts are limited to studying epigenetic associations in early and late childhood, respectively, but combined they achieve coverage of the entire childhood period. The scope of scientific questions that can be asked and addressed, as well as the statistical power that can be achieved through sample size, are thus both increased and improved.

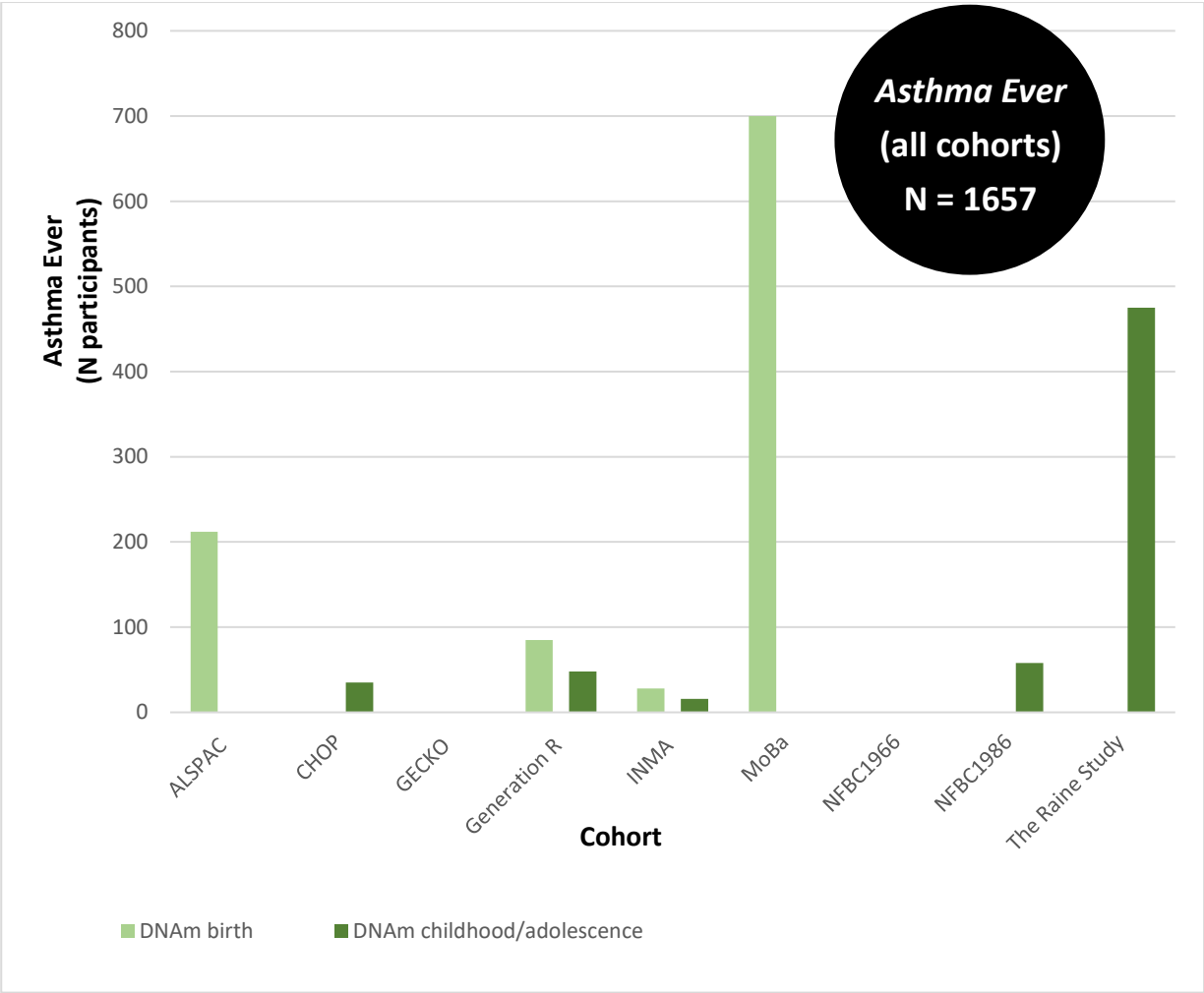

Figure 2. Number of participants with DNA methylation data (DNAm) and asthma ever data in cohorts participating in Task 5.4 of the LifeCycle project. Note: numbers represent asthma ever cases, i.e. participants with a reported history of asthma

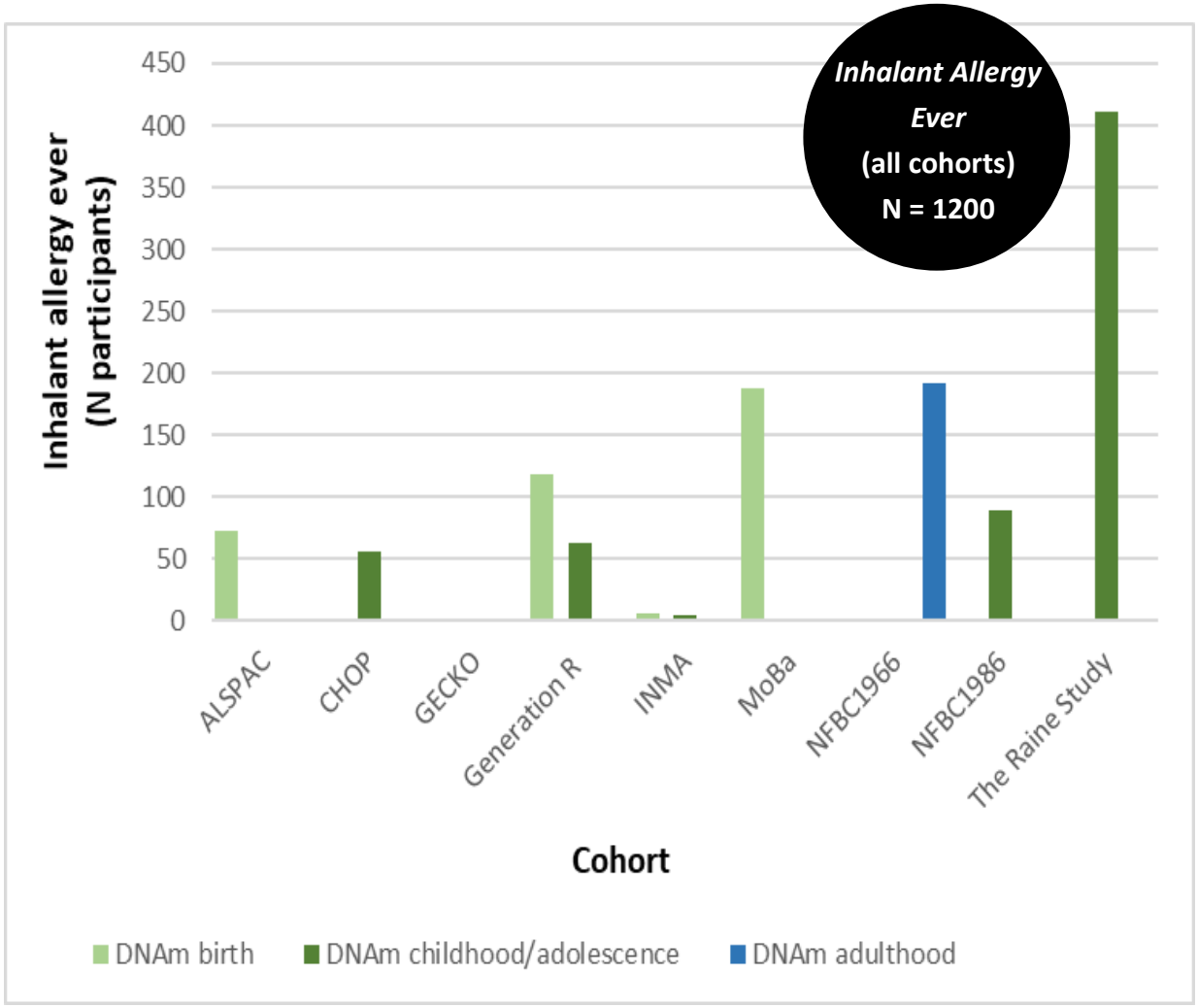

Figure 3. Number of participants with DNA methylation data (DNAm) and inhalant allergy ever data in cohorts participating in Task 5.4 of the LifeCycle project. Note: numbers represent inhalant allergy ever cases i.e. participants with a reported history of inhalant allergies

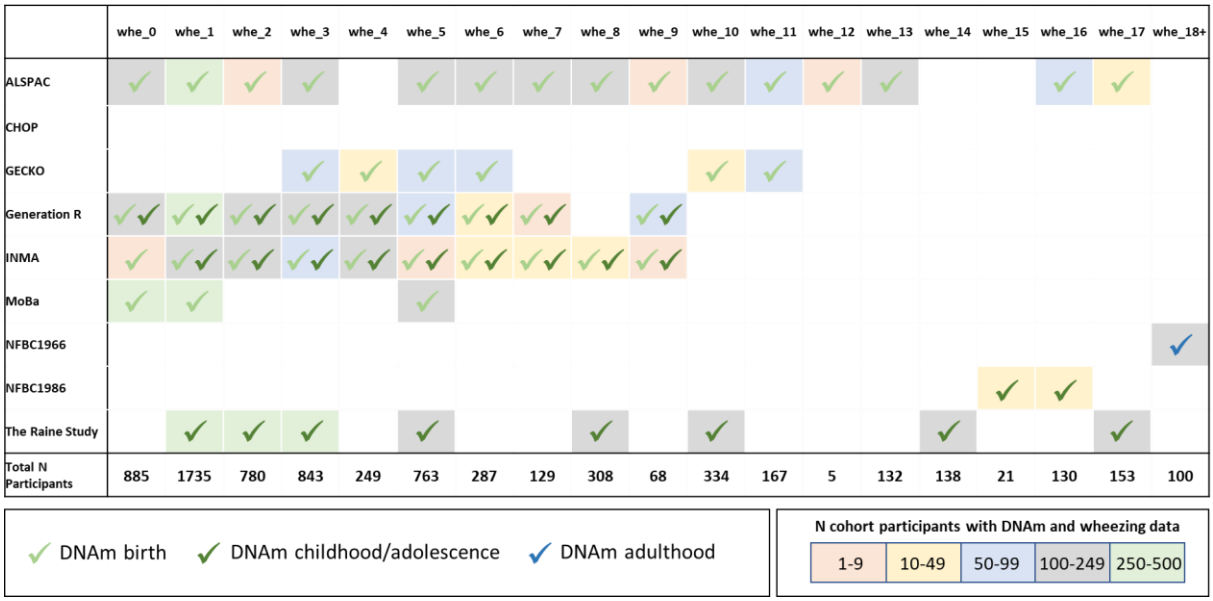

Figure 4. Overview of yearly-repeated measures of wheezing in cohort participants with DNA methylation data (DNAm) in Task 5.4 of the LifeCycle project. Note: numbers represent wheezing cases

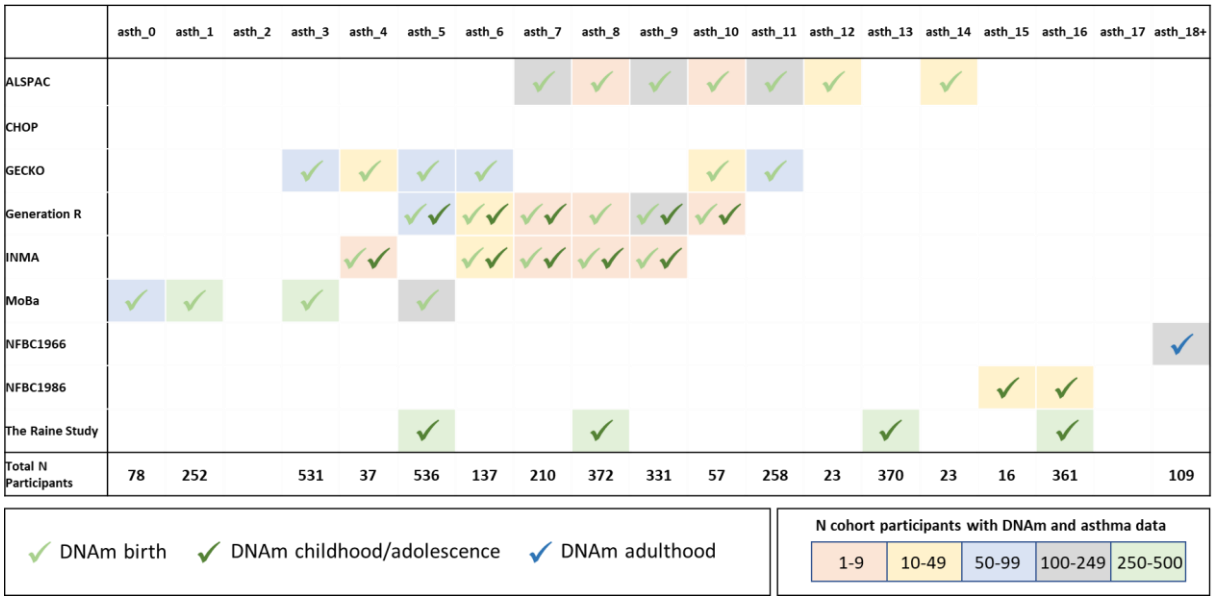

Figure 5. Overview of yearly-repeated measures of asthma in cohort participants with DNA methylation data (DNAm) in Task 5.4 of the LifeCycle project. Note: numbers represent asthma cases

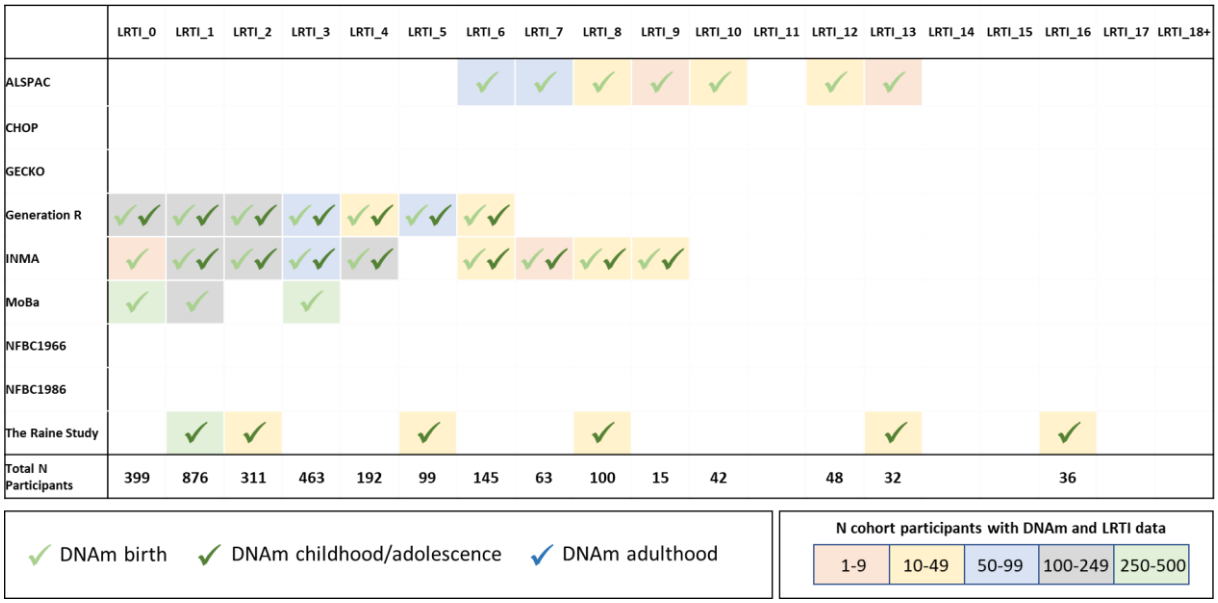

Figure 6. Overview of yearly-repeated measures of lower respiratory tract infections (LRTI) in cohort participants with DNA methylation data (DNAm) in Task 5.4 of the LifeCycle project. Note: numbers represent LRTI cases

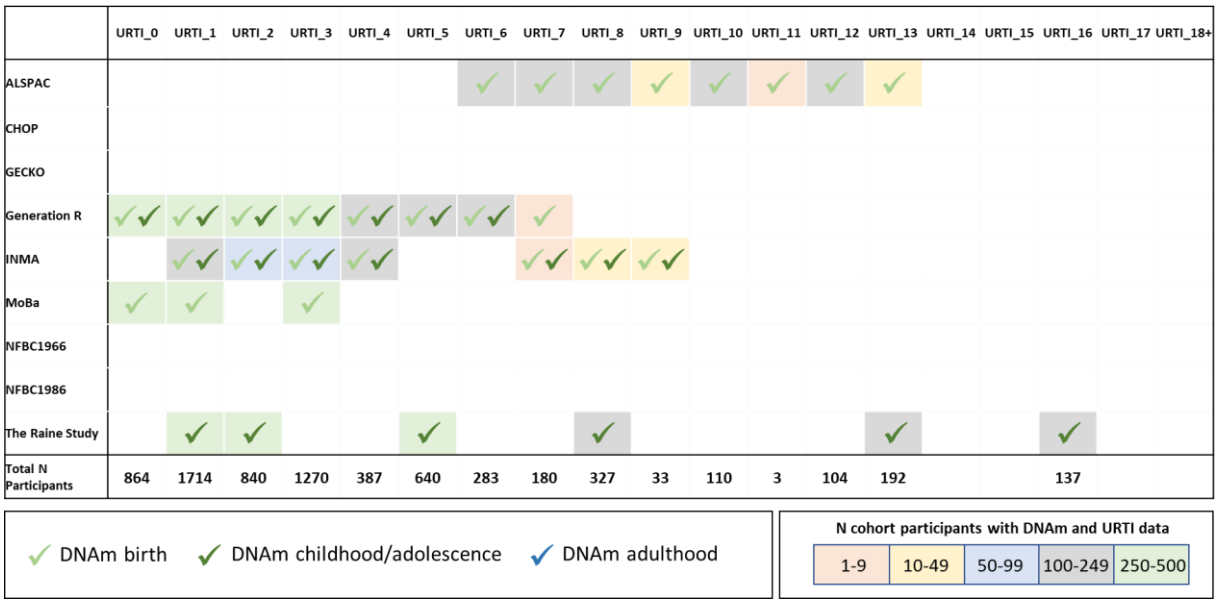

Figure 7. Overview of yearly-repeated measures of upper respiratory tract infections (URTI) in cohort participants with DNA methylation data (DNAm) in Task 5.4 of the LifeCycle project. Note: numbers represent URTI cases

3. Development of ExpressionSet with DNA methylation data for DataSHIELD analyses

Most of the analyses in LifeCycle are planned and conducted through DataSHIELD (38, 39), a publicly available analysis platform that allows meta-analyses of multiple datasets without users ever gaining physical access to the data. Recently, the analytical functionality of this platform was extended to enable use of DataSHIELD for performing epigenome-wide association studies (EWAS). This novel approach makes it possible to update versions of the data as new methylation data become available and perform EWAS remotely across

geographic borders and analyze sensitive data in an anonymized and non-disclosive manner (**Figure 8**). All of the participating cohorts in this task have installed DataSHIELD and have it operational as of March, 2022.

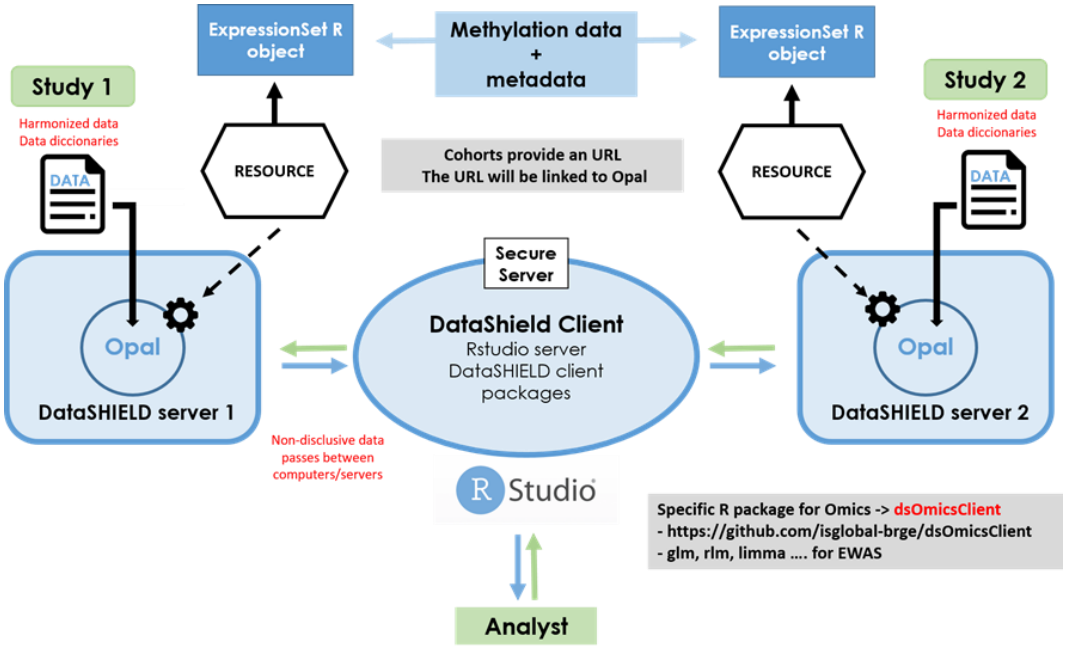

Figure 8. Representation of epigenome-wide association analysis (EWAS) through DataSHIELD (Bustamente et al. ExpressionSet with methylation data protocol)

To complete the work for Task 5.4 we coordinated aspects of this task with other LifeCycle teams and WPs where synergy of effort was essential. The ExpressionSet with DNA methylation data was performed by cohorts participating in Task 5.4 and other tasks in the LifeCycle project, following a protocol developed by Bustamente et al. in LifeCycle WP8. These R objects comprise a single matrix containing DNA methylation data, with samples in columns and CpG sites in rows, and two annotated data frames. The first of these frames provides metadata for the samples, including descriptive variables required for EWAS analyses such as participant age, sex, ethnic background, birth weight, cell type proportions, and so on. The second data frame contains information on the specific CpGs and probes within the ExpressionSet. For more information on the specific protocol, see Appendix I.

4. Results

Several LifeCycle studies have been conducted under task 5.4 as summarized below. These include multi-cohort and individual cohort studies that have focused on DNA methylation and asthma in childhood, DNA methylation and COPD, and DNA methylation as a mediating factor.

4.1 DNA methylation in childhood asthma

i. DNA methylation in childhood asthma: an epigenome-wide meta-analysis (27)

Participating LifeCycle cohorts: BTHFT (BiB), INSERM (EDEN), ISGlobal (INMA), UoC (RHEA)
Summary of findings: Asthma is a debilitating respiratory condition affecting more than 300 million people around the world and is a disease that is widely reported to be influenced by

multiple environmental and genetic factors. Epigenetic modification at CpG sites linked to asthma has previously been established, but the robustness of these associations and our understanding of age-specific patterns remain limited. In this study, DNA methylation data from preschool and school-aged children participating in four cohorts within the Mechanisms of the Development of ALLergy (MeDALL) project were used to investigate and address these limitations, further using data from children and adults in 9 European cohorts for replication and validation purposes. To examine potential age-specific effects, discovery analysis was performed on data from 207 child asthma cases aged 4-5 years old (610 controls), and 185 child asthma cases aged 8 years old (546 controls), using Illumina Infinium Human Methylation 450 BeadChips (450K). Selected CpG sites were subsequently validated in the replication cohorts using an independent technology (iPlex), focusing on the most significant sites in the two discovery age groups. Cell-specificity was further investigated, using epithelial and eosinophil cell samples and comparing these to DNA extracted from whole blood. Affymetrix Human Transcriptome Array 2.0 Genechips were used to study associations between single nucleotide polymorphisms (SNPs) and CpG methylation, and significant hits were further assessed and annotated to perform functional enrichment analysis and cell-type specificity.

Discovery analyses identified 27 CpG sites following quality control that were subject to replication analyses. Fourteen of these were found to significantly associate with asthma in all cohorts and age groups included in the analyses (**Figure 9**). Annotation of genes that localized to significant CpG sites found that cg01901579 in gene *DICER1* was most significant. All 14 replicated sites revealed lower methylation levels in asthma patients compared to controls, and a second replication study using Canadian families found that all 14 CpG sites retained a significant association with asthma. Methylation at these sites did not predict asthma in four cohorts for the 3-4 year age group. In addition to identifying 14 CpG sites that are robustly associated with asthma in childhood, this study found that methylation of these sites was lower in asthma cases from early childhood to adolescence. Asthma-related CpG sites in eosinophils had significantly reduced methylation, which implies that lower levels of methylation in whole blood is due to reduced methylation in this type of cell. Annotation of transcriptional profiles associate these 14 asthma-linked CpG sites to activation of important immune-regulatory cells, such as natural killer cells and CD8-positive T cells.

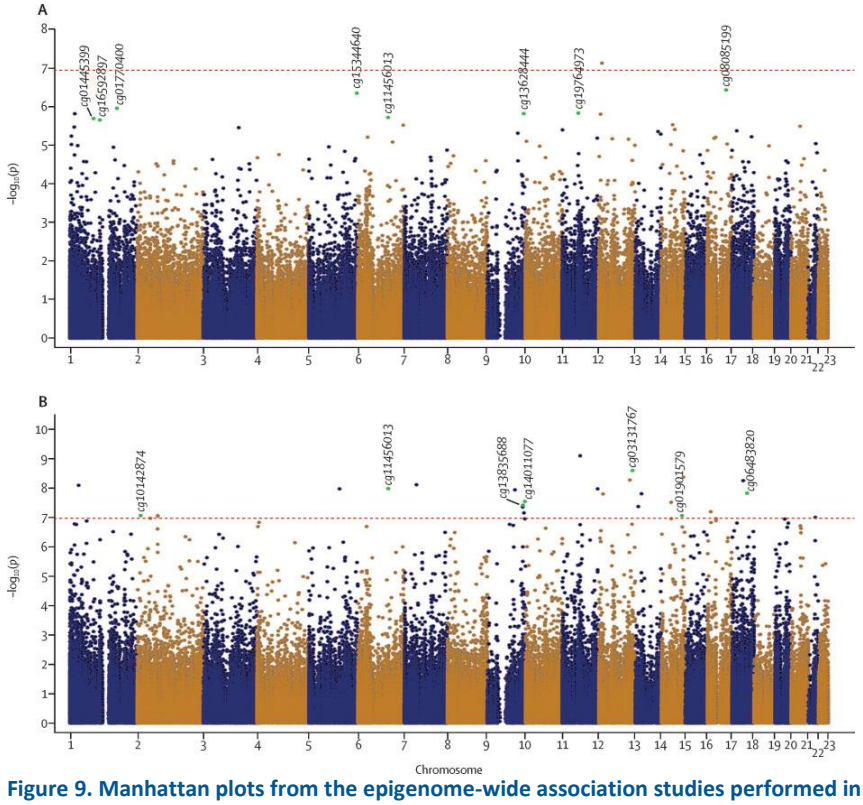

Figure 9. Manhattan plots from the epigenome-wide association studies performed in four European birth cohorts of childhood asthma (discovery) (figure reproduced from (27)). 439 306 CpG sites were tested for association with childhood asthma. The red dotted horizontal line represents the Bonferroni-corrected threshold ($p<1{\cdot}14 \times 10{-}7$) of genome-wide significance. All 14 replicated CpG sites are marked in green dot and annotated with CpG site name. (A) Results for the EWAS analysis in children aged 4–5 years (n=817) from the four European cohorts BAMSE, EDEN, INMA and PIAMA. (B) Results for the EWAS meta-analysis in children aged 8 years (n=731) from BAMSE and PIAMA.

ii. Epigenome-wide meta-analysis of DNA methylation and childhood asthma (28)

Participating LifeCycle cohorts: UNIVBRIS (ALSPAC), LMU (CHOP), INSERM (EDEN), ERASMUS (Generation R), NIPH (MoBa), UWA (The Raine Study)

Summary of findings: As the most common chronic respiratory condition in childhood, asthma is an important global public health concern. Despite this, biological mechanisms and drivers of disease risk, progression and prognosis are not well-established. Genetic and epigenetic variation are believed to underly many of the observed differences in epidemiological trends and disease manifestations, and DNA methylation has previously been linked to CpG sites within relevant asthma-associated genetic loci. Despite this, previous studies are limited by their focus on epigenetic data from childhood, rather than including other age groups such as newborns, which might increase the statistical ability to identify significant CpG sites associated with asthma. This study used DNA methylation data collected at or near birth from participants in 8 cohorts, as well as DNA methylation data collected during childhood from participants in 9 cohorts, to study prospective and cross-sectional associations with asthma, respectively. Methylation was measured in DNA from blood using the Illumina450K platform, and the primary outcome was defined as asthma in school-aged children (>5 years). CpG sites and differentially methylated regions (DMRs) that were significant were localized to the most proximal genetic loci (within 10 Mbp), and associations with differential expression patterns were also investigated. Finally, functional annotations and cell-type specificity were examined using available epigenomic data tools and comparisons with existing literature.

Associations between methylation at birth and childhood asthma found 9 significant CpG sites and 35 significant DMRs through meta-analysis of the 8 participating cohorts (**Figure 10**). There was no overlap between DMRs and CpGs, and the 9 sites that were identified were novel (not previously reported in existing literature). Associations between methylation in childhood and concurrent asthma revealed 179 significant CpGs and 36 significant DMRs via meta-analysis of the 9 participating cohorts (**Figure 11**), and 31 CpGs were localized to within a single DMR. Several of the 179 CpG sites were replicated in DNA methylation data from epithelial cells and eosinophils in childhood from past studies. Annotation of putative gene function revealed that all the significant CpGs identified in newborn DNA methylation analyses are located within proximity to a transcription factor binding site, which may indicate transcriptional activity as a possible mechanistic driver. The CpGs linked to childhood DNA methylation were largely localized to regions containing promoters and CpG islands, and significant enrichment for several histone markers in lung tissue and blood was further identified. Gene expression was associated with almost all of the CpGs identified in childhood DNA methylation datasets, and with nearly half of the CpGs identified in newborn DNA methylation data, and pathway enrichment analyses found that many genes were enriched for immune-regulatory pathways in asthma patients. This study identified several novel CpGs and DMRs that are found to be differentially methylated in asthma cases. Far more sites were identified through meta-analysis using childhood DNA methylation data compared to newborn DNA methylation data, which could have implications for the mechanisms that drive disease risk and development. Many of these novel CpGs appear to be localized to genetic loci that encode for proteins that influence immune function and response in asthma, and several of these have already been identified as putative targets for drug treatments. Our understanding of asthma pathogenesis may benefit from further investigation of the CpGs and associated genetic loci identified in this study.

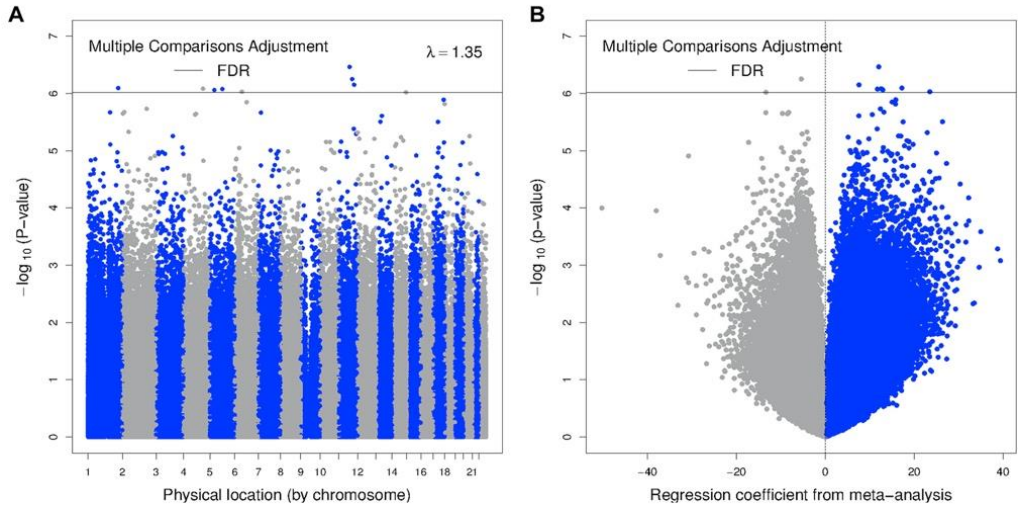

Figure 10. Meta-analysis of asthma in relation to newborn methylation: A, Manhattan plot; B, volcano plot. The model is adjusted for covariates and cell types (figure reproduced from (28))

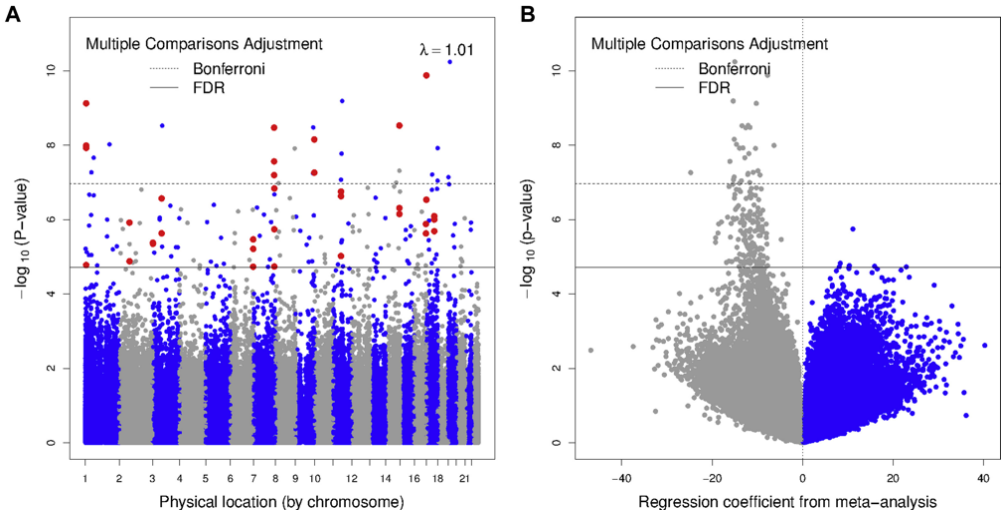

Figure 11. Meta-analysis of asthma in relation to childhood methylation: A, Manhattan plot; B, volcano plot. The model is adjusted for covariates and cell types. CpGs corresponding to more than 1 gene with significant CpGs (FDR < 0.05) are highlighted in red (figure reproduced from (28))

4.2 Epigenetics asthma and wheeze

iii. Epigenome-wide association study of asthma and wheeze in childhood and adolescence (40)

Participating LifeCycle cohorts: UNIVBRIS (ALSPAC)

Summary of findings: Asthma is one of the most common public health concerns around the world, and prevalence among children is increasing due to rapid changes in lifestyle and the environment. Several environmental factors and other exposures are known to affect asthma risk and progression, but among these epigenetic mechanisms remain the least studied and understood. Although a number of studies have reported on associations between DNA methylation and chronic inflammatory conditions such as asthma, causal relationships have not thoroughly been elucidated, and confounding due to factors such as cell type has not directly been addressed. The current study investigates asthma and wheeze outcomes in relation to DNA methylation measured at birth, childhood, and adolescence, and specifically employs a range of analytical approaches to be able to examine causality of relationships in more detail. The study sample comprised 149 current asthma cases at 7.5 years of age (632 controls), and 184 current asthma cases at 16.5 years of age (427 controls), out of the 1000 mother-child pairs for whom DNA methylation data were available in the ALSPAC cohort. A Mendelian Randomization strategy was employed in order to study causality, as this analytical approach implies unbiased effects of causal relationships identified in exposure-outcome analyses.

There was a strong positive association identified for current asthma at 7.5 years of age and eosinophil and B cell counts, and a strong negative association observed for neutrophils. Eosinophil estimates were similarly identified to be positively associated with current asthma at 16 years of age, but no significant associations were observed for this age group with B cells or neutrophils. 302 CpG sites were identified that were associated with asthma at 7.5 years of age, and that retained significance after adjusting for granulocyte cell types. The most significant hit localized to gene *ZFPM1*, a zinc finger protein whose primary

putative annotation is transcription factor binding. However, all associations were attenuated after separately adjusting for eosinophil and neutrophil counts. For the 16.5-year age group, two hits were still significant after separate adjustment of granulocyte cell types (**Figure 12**), which were individually mapped to *AP2A2* and *IL5RA*. However, these associations were only observed for current asthma, and were not replicated in models using ever asthma. When longitudinal models were used, there were 4 CpG sites that were significant for DNA methylation at 7.5 years of age and current asthma at 16 years of age, but these associations similarly attenuated following separate adjustment for eosinophil and neutrophil counts. A total of 192 genes were identified from the 302 CpG sites that were associated with asthma at 7.5 years of age, and some of the annotated pathways included eosinophil migration, movement of cellular components, and interleukin production. KEGG pathway enrichment analyses revealed that these genes were enriched for asthma and metabolism of amino/nucleotide sugar. Mendelian randomization analyses identified some evidence of a causal effect on DNA methylation by current asthma at 7.5 years of age, but none of the calculated estimates were significant at the defined threshold of $P = 3.4 \times 10^{-4}$. The current study demonstrated some important associations between DNA methylation in peripheral blood and current asthma. In addition, significant evidence was uncovered that strongly suggests these observations are influenced and driven by higher cell counts (especially eosinophils) during childhood, an important consideration for future research due to the important role eosinophils play in asthma pathology.

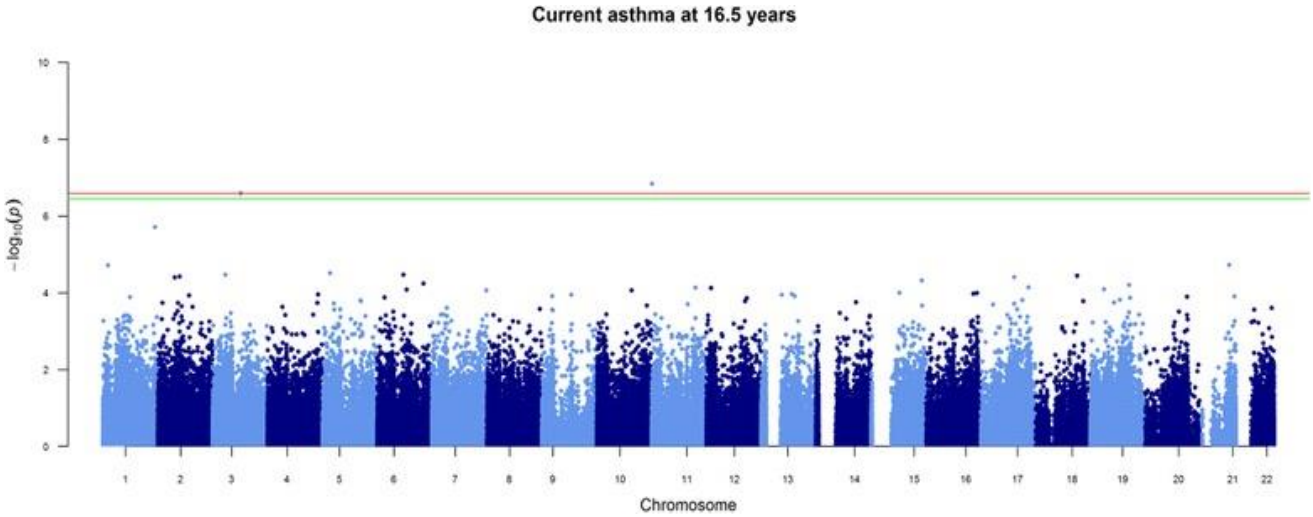

Figure 12. Manhattan plot of current asthma at 16.5 years adjusted for detailed cells. FDR significance cut-off line in green and Bonferroni cut-off line in red (figure reproduced from (40))

4.3 Methylation of ILRL1 and asthma

iv. Genetic regulation of IL1RL1 methylation and IL1RL1-a protein levels in asthma (41)

Participating LifeCycle cohorts: ISGlobal (INMA)

Summary of findings: Genetics are believed to drive a significant proportion of heritability in chronic respiratory conditions such as asthma. Several genome-wide association studies (GWAS) have been performed to date, revealing important genetic loci that consistently associate with differential risk and pathology. Of these, one of the most frequently identified genes is interleukin-1 receptor–like 1 (*IL1RL1*), which is known to affect time to onset of

several atopic traits and asthma, encoding a receptor superfamily found on the surface of immune-related cells in lung tissue. Although the function of *IL1RL1*-encoded proteins and their relationship with asthma are quite well-described, the potential role of DNA methylation in regulating gene expression and thus protein levels remain poorly understood. The current study used data from four European pregnancy and birth cohort studies to investigate relationships between methylation, single nucleotide polymorphisms (SNPs) and serum protein levels of IL1RL1.

The genetic region of study comprised an area on chromosome 2 that covered 200 kbp upstream to 200 kbp downstream of *IL1RL1* (chr2:102,927,962–102,968,497). Marginally significant results were obtained that identified several SNPs within *IL1RL1* or regions proximal to this gene that associated with asthma at 4 years. A single CpG site was found the be significantly associated with asthma, located in the distal promoter region of *IL1RL1*, but significance was not retained following multiple testing and after blood cell composition was corrected for. To examine whether IL1RL1-a levels are regulated by SNPs via DNA methylation, causal inference testing was performed. These causal analyses revealed independent associations between SNPs and protein levels, as well as between SNPs and DNA methylation, thus implying that DNA methylation is not a mediating factor in the relationship between *IL1RL1* SNPs and IL1RL1-a levels. Contrary to findings from previous reports, this study did not find any associations between IL1RL1-a levels and child or adult asthma. Although no mediating effects for DNA methylation could be identified, analyses revealed that SNPs in *IL1RL1* are highly significantly associated with DNA methylation and expression levels for IL1RL1-a, but that neither of these are associated with asthma in childhood or adulthood.

4.4 DNA methylation at birth and later respiratory function

v. Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course (42)

Participating LifeCycle cohorts: UNIVBRIS (ALSPAC), ERASMUS (Generation R), ISGlobal (INMA)

Summary of findings: Chronic obstructive pulmonary disease (COPD) and asthma are two of the most common and debilitating chronic public health conditions, affecting millions of people around the world. Compromised lung function through airway obstruction is a characteristic trait of both diseases. Lung function in childhood can predict the risk for adverse lung function and thus conditions such as asthma and COPD in later life. Although genetic differences explain some portion of observed variation in disease onset and pathology, other mechanisms such as epigenetic factors may influence the risk for chronic respiratory conditions. These could exert effects either directly or by mediating the association between genetic predisposition or early environmental exposures and later disease outcomes. In order to investigate the role of epigenetic modification in asthma and COPD, this study explored associations between differentially methylated regions (DMRs) in cord blood at birth with three diagnostic measures of lung function in childhood: forced expiratory volume (FEV), FEV1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC (FEF75%). Significant DMRs were consequently evaluated for their associations with asthma, lung function, and COPD throughout life.

A total of 59 DMRs were found to be significantly associated with the three lung function measures included in the present study, comprising 22, 15 and 22 DMRs respectively (**Figure 13**). Most of these were associated with higher lung function measures (67%), while the remainder were associated with reduced lung function. 18 of the identified DMRs were associated with asthma in childhood, 20 DMRs were associated with lung function (11 in adolescence, 9 in adulthood), and 9 DMRs were significantly associated with COPD in adulthood. Several of the DMRs housed genes that previously have been implicated in adverse lung development and respiratory diseases or were within or proximal to described regulatory elements involved in gene expression. This study used data from multiple pregnancy and birth cohorts, facilitating meta-analyses with larger sample sizes from Europe and the United States and thus increasing statistical power and the epidemiological diversity of included populations. This study identified several novel DMRs that may help explain how epigenetic differences at birth can affect gene function and expression related to lung development, and thus influences risk, progression and prognosis of common obstructive respiratory conditions throughout the life course.

4.5 A closer look at the relationships between DNA methylation, obesity and asthma

vi. BMI trajectory in childhood is associated with asthma incidence at young adulthood mediated by DNA methylation (43)

Participating LifeCycle partners: UNIVBRIS (ALSPAC)

Summary of findings: Obesity and asthma comprise two of the most debilitating conditions in childhood, and several have reported that obesity significantly increases the risk for asthma through a dose-response relationship. Obesity and specifically body mass index (BMI) are believed to drive changes in DNA methylation in adipose tissue and blood. Given that epigenetic mechanisms have been implicated in asthma risk and pathology, and the direct association between BMI and asthma development, this study aimed to investigate potential mediating effects of DNA methylation on the causal relationship between obesity and asthma. The analyses studied trajectories of BMI at multiple timepoints during childhood (1, 2, 4, and 10 years), asthma diagnosis at around 18 years, and the mediating effects of DNA methylation measured at 10 years of age. BMI trajectories were available for a total of 577 girls and 602 boys.

Basic logistic regression analyses revealed that higher BMI trajectories were significantly associated with asthma incidence at 18 years of age, and these associations held across sex comparisons and persisted after adjustment for various covariates and confounders. Sex-stratified EWAS revealed a total of 212 CpGs and 159 CpGs that were associated with BMI trajectory groups in girls and boys, respectively. These CpGs were subsequently screened for significant associations with asthma at 18 years, which identified six hits in females and nine

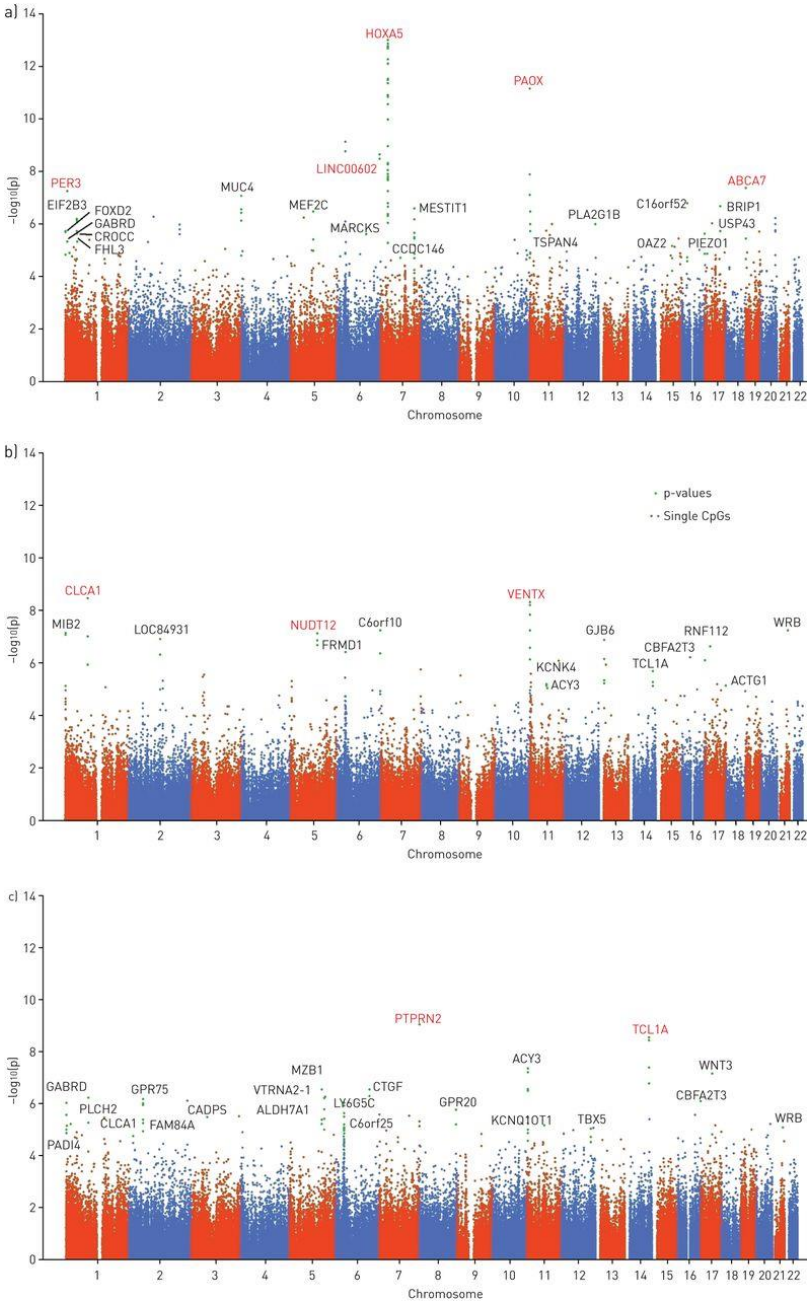

Figure 13. Manhattan plots of associations of CpGs located in differentially methylated regions (DMRs) with childhood lung function outcomes. Green dots represent p-values from associations of CpGs located in DMRs at birth with childhood forced expiratory volume in 1 s (FEV1), FEV1/forced vital capacity (FVC) ratio and forced expiratory flow at 75% of FVC (FEF75%) (figure reproduced from (42))

in males following adjustment for confounders. Path analysis was used to investigate the potential mediating effects of these specific CpGs on BMI trajectories in childhood, revealing three hits in females and one hit in males (**Figure 14**). Indirect and direct mediating effects were statistically significant for two of these CpGs, both in females. Direct effects in this context describes the observation of high BMI trajectories in childhood being significantly associated with higher incidence of asthma at 18 years, and that these effects were attenuated by DNA methylation at the identified CpG sites. For the remaining two CpGs only

indirect effects were found to be statistically significant. Three of the four CpGs also mapped to proximal genetic loci, and expression data revealed a negative correlation between DNA methylation and gene expression levels. In summary, this study identified four CpGs that may mediate causal effects of adverse BMI throughout childhood on asthma in late adolescence. Most of the genes that contained these CpGs have previously been linked to BMI/obesity and/or asthma, lending credibility to the potential significance of their mediating role, and expression data of proximal genes indicate that mediation may occur via up- or down-regulation of nearby genetic loci.

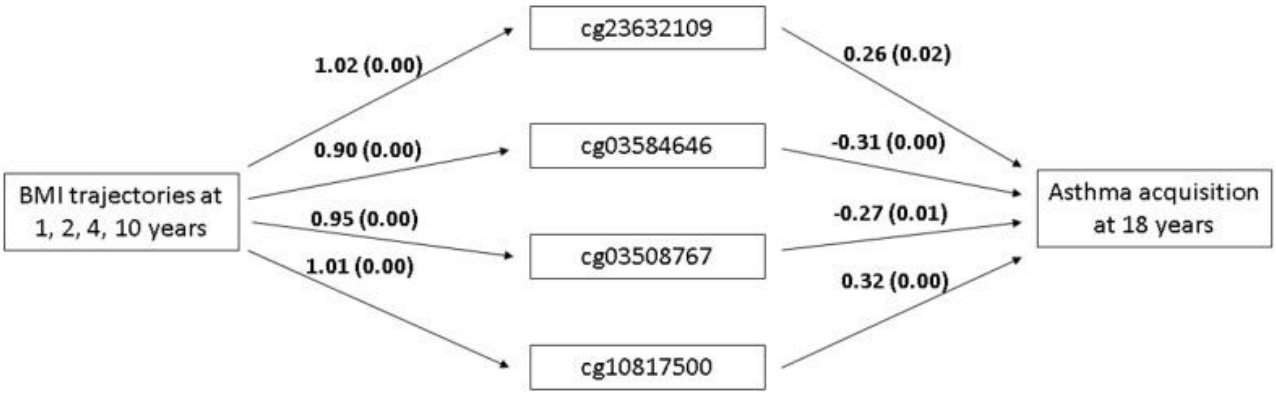

Figure 14. Indirect effects of childhood BMI trajectory on adulthood asthma incidence via DNAm at four CpGs. The figure shows the estimates (and p-values) of direct effects at each path, based on which indirect effects of BMI trajectory were inferred (figure reproduced from (43))

vii. Causal effects of DNAm related to birthweight on adult disease outcomes: Phenome wide Mendelian randomisation study

Participating LifeCycle partners: UMCG, NIPH (lead for 5.4), UNIBRIS, OULU.

Summary of findings: Birthweight is a marker thought to represent intrauterine environmental quality and is associated with a wide range of outcomes during the life-course. In this project, we examined causal effects of DNA methylation (DNAm) at 914 known birthweight-related cytosine-phosphate-guanine sites (CpGs) on a range of adult outcomes within, but not restricted to, the interest field of Lifecyle (i.e.cardiometabolic [work package 4.4], respiratory [work package 5.4], and mental health outcomes [work package 6.4]). To this end, we performed a phenome-wide Mendelian randomization analysis (MR) using summary statistics data from large-scale genome-wide association studies (GWAS). Furthermore, to explore potential causal pathways (e.g.birthweight > CpG > outcome, or CpG > birthweight > outcome), we performed bidirectional MR of prioritized CpGs and birthweight.

MR exploits the random assignment and independent assortment of common single nucleotide polymorphisms (SNPs) as a natural experiment; using SNPs as instrumental variables theoretically minimizes confounding bias and MR thus yields causal estimates of the association between exposure and outcome. For each CpG – outcome combination, we extracted SNPs with genome-wide significant associations ($p<5 \times 10^{-8}$) with DNAm at this specific CpG from data from the Genetics of DNA Methylation (GoDMC) Consortium (N up to

~28,000). For 757 CpG sites, we were able to identify a sufficient number of suitable SNPs (i.e. at least 1 independent SNP with strong associations with DNAm). Next, we extracted these SNPs from publicly available GWAS on 1803 disease endpoints performed by the FinnGen project (round 5 FinnGenn analyses, N up to ~300,000), including 81 GWAS on respiratory disease endpoints categorized by asthma, COPD, comorbidities of asthma and COPD, interstitial lung disease, comorbidities of interstitial lung disease, and general respiratory disease. If the original SNP was not available, we attempted to identify proxy SNPs in high linkage disequilibrium with the original SNP. To limit dependence between SNPs, we performed clumping based on LD r^2=0.2 within 1 MB. We excluded palindromic SNPs with ambiguous minor allele frequency (>0.42). This procedure resulted in 1 to 166 SNPs per CpG site (median: 7 ; interquartile range: 3 to 15 SNPs). For the MR analyses, we calculated Wald ratios (i.e. SNP-outcome effect divided by SNP-exposure effect) to estimate causal effects per SNP. In case more than 1 SNP per CpG site was available, we pooled single-SNP Wald ratios using inverse-variance weighted meta-analysis. Several MR sensitivity analyses were performed to assess robustness of the results to violations of MR-specific assumptions regarding horizontal pleiotropy (i.e. MR Egger, median and mode based MR). We controlled for the false discovery rate (FDR) by adjusting the significance threshold according to Benjamini-Hochberg for each outcome separately. All MR analyses were performed using the *TwoSampleMR* R package.

We identified 2066 significant CpG-outcome pairs in the respiratory domain. Of all examined 757 CpGs, 468 (61.8%) had a significant effect on at least 1 outcome (range 1 to 30**). Figure 15** shows the proportion of birthweight-related CpG sites at which DNAm showed significant (FDR<0.05) effects on respiratory outcomes. Enrichment for asthma-related endpoints was found as various definitions of asthma consistently ranked high with regards to proportion of significant CpGs. **Figure 16** shows the effects of 19 selected CpG sites with significant effects in at least 15 outcomes. Clustering of significant MR estimates can again be observed for asthma and various definitions thereof, with consistent effects of the selected CpGs across definitions. In **Figures 17-18**, we present data on a sample outcome, namely Asthma/COPD defined by KELA (Finland Social Insurance), as it ranked highest among respiratory outcomes with regards to proportion of significant CpGs. For this outcome, there were significant effects of 63 CpG sites, of which cg12655416 was the top hit. In **Figure 18**, we show MR sensitivity analyses for this CpG. Higher levels of DNAm at cg12655416 was found to reduce odds of asthma outcomes, robust to violations of MR assumptions regarding pleiotropy. cg12655416 maps to the *ORMDL3* region, a region previously identified in relation with (childhood) asthma outcomes. *ORMDL3* encodes sphingolipid biosynthesis regulator 3, which has a general function in protein binding. It is ubiquitously expressed, notably also in both adipose and lung tissue. Furthermore, secondary MR of cg12655416 showed consistent and significant positive effects on continuous lung function outcomes in non-FinnGen GWAS data (peak expiratory flow, FEV1/FVC, FEV1, but not FVC, data not shown), i.e. higher levels of DNAm at this CpG is suggested to cause better lung function. cg12655416 and its associated gene, *ORMDL3,* are thus potential targets for intervention and risk stratification for asthma. Next, we explored potential causal pathways. In the original EWAS on birthweight, higher levels of DNAm at cg12655416 was associated with higher birthweight. However, using 36 SNPs for birthweight, we found a null causal effect of birthweight-proxied

exposure on DNAm levels at cg12655416 (data not shown). In reversed MR, using 36 SNPs for cg12655416, we found a suggestive positive causal effect on DNAm on birthweight but this was not corroborated by additional MR sensitivity analyses (data not shown). This specific finding is therefore not consistent with the hypothesis that DNAm at this CpG site mediates the relation between a more beneficial intrauterine environment (as proxied by higher birthweight) and reduced risk of asthma. It is however consistent with a model in which DNAm at this CpG precedes both birthweight and later life asthma (e.g. BW as a mediator of the CpG > asthma relation, other relations also possible). Additional work is warranted to establish the best fitting causal model.

Our analysis has yielded a wealth of potentially important sets of CpG sites that warrant follow-up work and verification steps. We are currently conducting thorough sensitivity analyses, colocalization with eQTL and GWAS data, multivariable MR with blood cell types, bioinformatics annotation, and observational mediation analysis.

In conclusion, we investigated causal effects of 757 birthweight related CpG sites on 1803 disease endpoints in FinnGen, generating data on >1.35 $x10^6$ CpG-outcome pairs. In the respiratory domain, this amounts to 60,560 investigated CpG-outcome pairs, of which 2066 pairs show promise for follow up-work and verification.

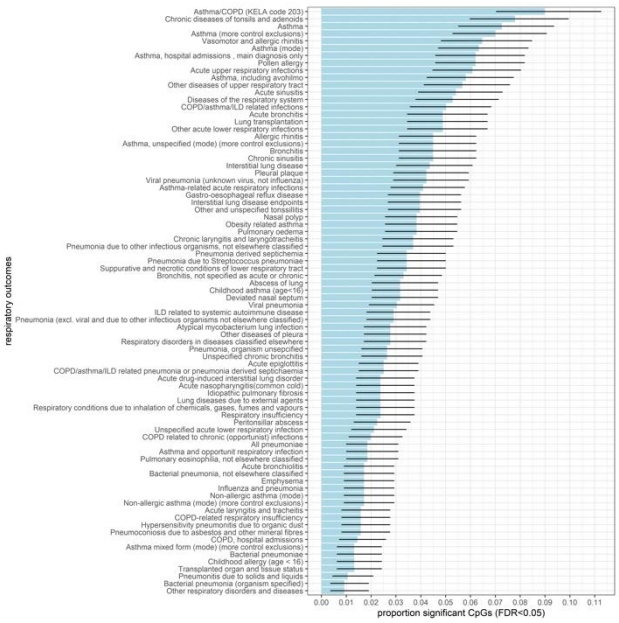

Figure 15. Enrichment of birthweight-related DNAm with MR estimated causal effects on 81 respiratory outcomes. Bars represent the proportion of significant (FDR<0.05) CpG sites, error bars represent binomial-distribution derived 95%CI

Figure 16. MR estimated causal effects (transformed to T-values) on 81 respiratory outcomes of 19 selected CpG sites with >15 significant findings. Red indicates positive effect of DNAm on outcome, blue indicates negative effect of DNAm on outcome. Black dot indicates significance (FDR<0.05)

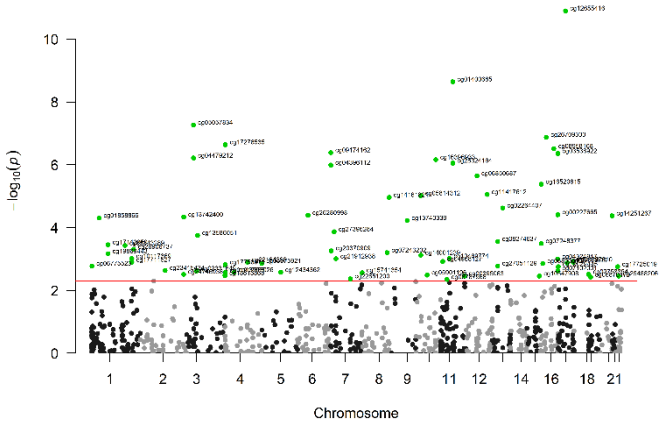

Figure 17. Manhattan plot for MR estimates of DNAm on Asthma/COPD (by KELA definition, Finnish Social Insurance). Each dot represents one of 757 birthweight-related CpG sites. The Y-axis shows -log10(p-value), the X-axis shows position. The red line indicates the significance threshold (outcome-specific FDR<0.05)

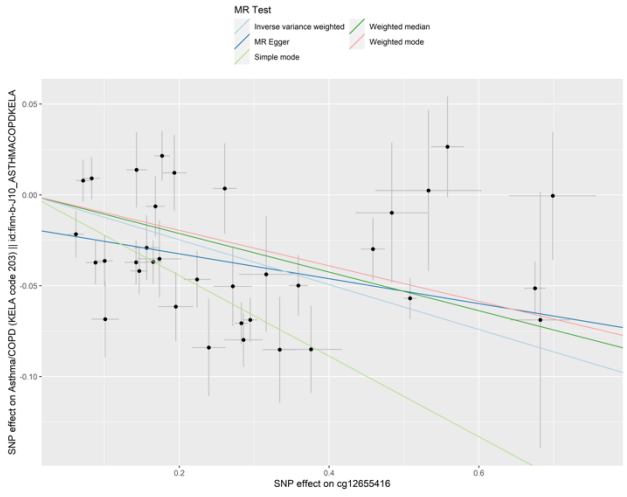

Figure 18. MR scatterplot for the causal effect of DNAm at cg12655416 on Asthma/COPD (by KELA definition, Finnish Social Insurance). Each dot represents a single SNP used as instrument for DNAm at cg12655416. The Y-axis shows SNP effects on Asthma/COPD (log-odds scale), the X-axis shows SNP effects on DNAm (in standard deviations)

5. Conclusion

The LifeCycle Task 5.4 focused on the identification of DNA methylation loci that mediate the relationships of early-life stressors with respiratory health, including asthma and chronic obstructive respiratory disease. We have used epigenome-wide data on DNA methylation and conducted meta-analyses to identify relevant epigenetic markers associated with respiratory outcomes. Where relevant, we have collaborated with the Pregnancy and Childhood Epigenetics (PACE) consortium to increase our statistical power, strengthen the science, and build on ongoing work.

Work in this task was also extended through novel analyses that exploited publicly available GWAS results and DNA methylation data to investigate causal effects of birthweight-related CpG sites for which DNAm is significantly associated with respiratory outcomes.

The collection of LifeCycle published studies and analyses comprising this task helps to elucidate pathways and mechanisms involved in the development of respiratory disease. Findings based on DNA methylation among newborns suggest that epigenetic status at birth can have life-long effects on respiratory health and disease and the discovery of novel loci that are differentially methylated in newborns should be pursued to determine their value as biomarkers of asthma risk among children. Evidence from multiple studies also pointed to the activation of eosinophils and cytotoxic T-cells in childhood asthma. Moreover, these studies provide insights into complex genetic-epigenetic interplay between SNPs and methylation sites that affect asthma related genes, such as the Interleukin-1 receptor–like 1 (*IL1RL1*).

Investigation of the causal effects of birthweight related CpG sites on disease have identified more than 2,000 CpG outcome pairs that warrant follow-up for understanding methylation effects on respiratory health. Together, these findings have provided important new insights into the role of DNA methylation in mediating the relationships between early-life stressors and the development of respiratory disease.

6. Contribution of partners

The deliverable report for Task 5.4 was prepared by NIPH and UMCG and reviewed by all participating partners. All partners provided additional data and input for overviews mapping DNA methylation data across respiratory traits.

- **UNIVBRIS:** Completed harmonization of respiratory variables, preparing ExpressionSets for EWAS via DataSHIELD, and participated in the following EWAS: DNA methylation and childhood asthma, EWAS of wheeze/asthma in childhood and adolescence, DNA methylation and lifetime asthma/COPD risk, mediating effects of DNA methylation on the association between BMI and asthma, causal effects of DNA methylation on the association between birthweight and respiratory outcomes

- **LMU:** Completed harmonization of respiratory variables and participated in the following EWAS: DNA methylation and childhood asthma

- **UMCG:** Completed harmonization of respiratory variables and lead the analytical work to investigate causal effects of DNA methylation on the association between birthweight and respiratory outcomes

- **ERASMUS:** Completed harmonization of respiratory variables, preparing ExpressionSets for EWAS via DataSHIELD, and participated in the following EWAS: DNA methylation and childhood asthma, DNA methylation and lifetime asthma/COPD risk

- **ISGLOBAL:** Completed harmonization of respiratory variables, preparing ExpressionSets for EWAS via DataSHIELD, and participated in the following EWAS: EWAS of childhood asthma, methylation of interleukin-1 receptor–like 1 in asthma, DNA methylation and lifetime asthma/COPD risk

- **NIPH:** Completed harmonization of respiratory variables, preparing ExpressionSets for EWAS via DataSHIELD, and participated in the following EWAS: DNA methylation and childhood asthma, causal effects of DNA methylation on the association between birthweight and respiratory outcomes

- **UOULU:** Completed harmonization of respiratory variables and participated in the following EWAS: causal effects of DNA methylation on the association between birthweight and respiratory outcomes

- **UWA:** Completed harmonization of respiratory variables and participated in the following EWAS: DNA methylation and childhood asthma

7. Deviations from original plan

This deliverable was delayed from month 60 to month 63 due to the pandemic, version 2.0 was submitted in month 66. As described above, we extended our original analytical plan using a novel phenome-wide Mendelian randomization study that exploited methylation data available through the Genetics of DNA Methylation (GoDMC) Consortium (see section 5) to study asthma. Otherwise, there no deviations from our original, proposed plan.

8. Dissemination activities

Findings generated under task 5.4 have been published in peer-review journals and can be accessed online. This report is public. Further publications generated from these data will be deposited in relevant repositories in accordance with principles of open-science.

9. References

1. World Health Organization FoIRS. The Global Impact of Respiratory Disease. Sheffield; 2017.
2. Asher MI, Rutter CE, Bissell K, Chiang CY, El Sony A, Ellwood E, et al. Worldwide trends in the burden of asthma symptoms in school-aged children: Global Asthma Network Phase I cross-sectional study. Lancet. 2021;398(10311):1569-80.
3. Burney PG, Patel J, Newson R, Minelli C, Naghavi M. Global and regional trends in COPD mortality, 1990-2010. Eur Respir J. 2015;45(5):1239-47.
4. Eder W, Ege MJ, von Mutius E. The asthma epidemic. N Engl J Med. 2006;355(21):2226-35.
5. Guarnieri M, Balmes JR. Outdoor air pollution and asthma. Lancet. 2014;383(9928):1581-92.
6. Burke H, Leonardi-Bee J, Hashim A, Pine-Abata H, Chen Y, Cook DG, et al. Prenatal and passive smoke exposure and incidence of asthma and wheeze: systematic review and meta-analysis. Pediatrics. 2012;129(4):735-44.
7. Toskala E, Kennedy DW. Asthma risk factors. Int Forum Allergy Rhinol. 2015;5 Suppl 1:S11-6.
8. Rosenberg SL, Miller GE, Brehm JM, Celedon JC. Stress and asthma: novel insights on genetic, epigenetic, and immunologic mechanisms. J Allergy Clin Immunol. 2014;134(5):1009-15.
9. Powell H, McCaffery K, Murphy VE, Hensley MJ, Clifton VL, Giles W, et al. Psychosocial outcomes are related to asthma control and quality of life in pregnant women with asthma. J Asthma. 2011;48(10):1032-40.
10. Akula M, Kulikova A, Khan DA, Brown ES. The relationship between asthma and depression in a community-based sample. J Asthma. 2018;55(12):1271-7.
11. da Costa A, Dos Santos LM, Feitosa CA, de Miranda CT. Maternal common mental disorder over time and asthma control: The role of social support. Pediatr Allergy Immunol. 2020;31(6):628-35.
12. Deng X, Yang M, Wang S, Wang Q, Pang B, Wang K, et al. Factors Associated With Childhood Asthma and Wheeze in Chinese Preschool-Aged Children. Front Med (Lausanne). 2021;8:742581.
13. Ege MJ, Mayer M, Normand AC, Genuneit J, Cookson WO, Braun-Fahrlander C, et al. Exposure to environmental microorganisms and childhood asthma. N Engl J Med. 2011;364(8):701-9.
14. Weiss ST. Obesity: insight into the origins of asthma. Nat Immunol. 2005;6(6):537-9.
15. Castro-Rodriguez JA, Forno E, Rodriguez-Martinez CE, Celedon JC. Risk and Protective Factors for Childhood Asthma: What Is the Evidence? J Allergy Clin Immunol Pract. 2016;4(6):1111-22.
16. Stern J, Pier J, Litonjua AA. Asthma epidemiology and risk factors. Semin Immunopathol. 2020;42(1):5-15.
17. Willis-Owen SAG, Cookson WOC, Moffatt MF. The Genetics and Genomics of Asthma. Annu Rev Genomics Hum Genet. 2018;19:223-46.
18. Vercelli D. Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol. 2008;8(3):169-82.
19. Lopez-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2016;21(1):14-23.
20. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. Lancet. 2015;385(9971):899-909.
21. Lalloo UG, Pillay S, Mngqibisa R, Abdool-Gaffar S, Ambaram A. HIV and COPD: a conspiracy of risk factors. Respirology. 2016;21(7):1166-72.
22. Rosenberg SR, Kalhan R, Mannino DM. Epidemiology of Chronic Obstructive Pulmonary Disease: Prevalence, Morbidity, Mortality, and Risk Factors. Semin Respir Crit Care Med. 2015;36(4):457-69.
23. Silverman EK. Genetics of COPD. Annu Rev Physiol. 2020;82:413-31.
24. Cosío BG, Dacal D, Pérez de Llano L. Asthma-COPD overlap: identification and optimal treatment. Ther Adv Respir Dis. 2018;12:1753466618805662.
25. Edris A, den Dekker HT, Melén E, Lahousse L. Epigenome-wide association studies in asthma: A systematic review. Clin Exp Allergy. 2019;49(7):953-68.
26. de Vries M, Nedeljkovic I, van der Plaat DA, Zhernakova A, Lahousse L, Brusselle GG, et al. DNA methylation is associated with lung function in never smokers. Respir Res. 2019;20(1):268.

27. Xu CJ, Söderhäll C, Bustamante M, Baïz N, Gruzieva O, Gehring U, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. Lancet Respir Med. 2018;6(5):379-88.
28. Reese SE, Xu CJ, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. J Allergy Clin Immunol. 2019;143(6):2062-74.
29. Gao L, Millstein J, Siegmund KD, Dubeau L, Maguire R, Gilliland FD, et al. Epigenetic regulation of AXL and risk of childhood asthma symptoms. Clin Epigenetics. 2017;9:121.
30. Everson TM, Zhang H, Lockett GA, Kaushal A, Forthofer M, Ewart SL, et al. Epigenome-wide association study of asthma and wheeze characterizes loci within HK1. Allergy Asthma Clin Immunol. 2019;15:43.
31. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680-96.
32. Gao L, Liu X, Millstein J, Siegmund KD, Dubeau L, Maguire RL, et al. Self-reported prenatal tobacco smoke exposure, AXL gene-body methylation, and childhood asthma phenotypes. Clin Epigenetics. 2018;10(1):98.
33. Prunicki M, Stell L, Dinakarpandian D, de Planell-Saguer M, Lucas RW, Hammond SK, et al. Exposure to NO(2), CO, and PM(2.5) is linked to regional DNA methylation differences in asthma. Clin Epigenetics. 2018;10:2.
34. Nedeljkovic I, Carnero-Montoro E, Lahousse L, van der Plaat DA, de Jong K, Vonk JM, et al. Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics. Eur J Hum Genet. 2018;26(5):709-22.
35. Nedeljkovic I, Lahousse L, Carnero-Montoro E, Faiz A, Vonk JM, de Jong K, et al. COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression. Hum Mol Genet. 2018;27(2):396-405.
36. Shorey-Kendrick LE, McEvoy CT, Ferguson B, Burchard J, Park BS, Gao L, et al. Vitamin C Prevents Offspring DNA Methylation Changes Associated with Maternal Smoking in Pregnancy. Am J Respir Crit Care Med. 2017;196(6):745-55.
37. Nader JL, López-Vicente M, Julvez J, Guxens M, Cadman T, Elhakeem A, et al. Cohort description: Measures of early-life behaviour and later psychopathology in the LifeCycle Project - EU Child Cohort Network. J Epidemiol. 2021.
38. Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, et al. DataSHIELD: resolving a conflict in contemporary bioscience--performing a pooled analysis of individual-level data without sharing the data. Int J Epidemiol. 2010;39(5):1372-82.
39. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol. 2014;43(6):1929-44.
40. Arathimos R, Suderman M, Sharp GC, Burrows K, Granell R, Tilling K, et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. Clin Epigenetics. 2017;9:112.
41. Dijk FN, Xu C, Melén E, Carsin AE, Kumar A, Nolte IM, et al. Genetic regulation of IL1RL1 methylation and IL1RL1-a protein levels in asthma. Eur Respir J. 2018;51(3).
42. den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, et al. Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. Eur Respir J. 2019;53(4).
43. Rathod R, Zhang H, Karmaus W, Ewart S, Kadalayil L, Relton C, et al. BMI trajectory in childhood is associated with asthma incidence at young adulthood mediated by DNA methylation. Allergy Asthma Clin Immunol. 2021;17(1):77.

Appendix 1

LifeCycle and ATHLETE projects - SOP:
How to create an ExpressionSet with methylation data for analysis through DataSHIELD

Developed by: Sofia Aguilar, Janine Felix, Martine Vrijheid, Juan Ramon González, Mariona Bustamante

Version: v1
Date: 2021.04.16

Workflow of EWAS through DataSHIELD in LifeCycle and ATHLETE projects

In LifeCycle and ATHLETE, differently from PACE consortium, EWAS will not be run by the cohorts but by the leading team through DataSHIELD. DataSHIELD is an infrastructure and series of R packages that enable remote and non-disclosive analysis of sensitive research data (https://www.datashield.ac.uk/).

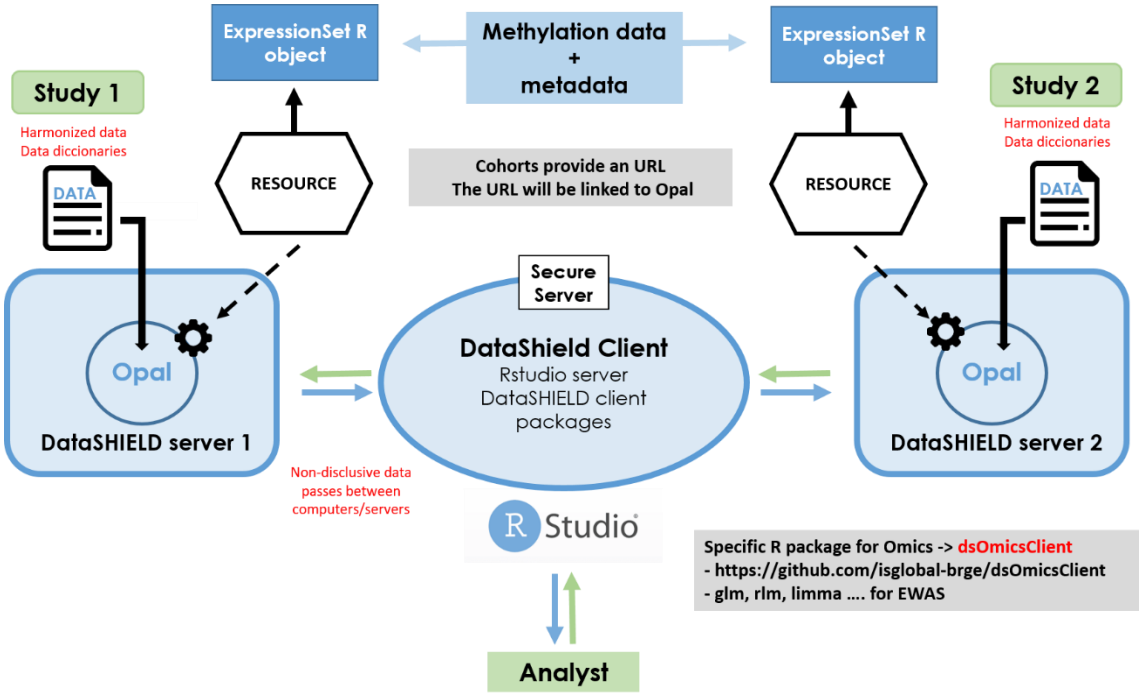

Figure 1. Representation of epigenome-wide association analysis (EWAS) through DataSHIELD.

The workflow will include tasks done by the cohorts and others done by the leaders.
To be done by the cohorts:

1) **Prepare DNA methylation and associated metadata**
   Cohorts will have to prepare the DNA methylation and associated metadata in a specific format named ExpressionSet and link it as a resource to an Opal server. This ExpressionSet has only to be created once. See below for more details.
2) **Sign a Data Access Agreement**

Cohorts and the leading teams will sign a Data Access Agreement (DAA) to allow access the data needed for the analysis, both the ExpressionSet and the other data already harmonized in LifeCycle/ATHLETE.

To be done by the leading team:

1) **Merge datasets and calculate descriptive**
   First, the leading team will merge the methylation data with other variables (exposures, phenotypic traits, covariates), and after selecting complete cases, will calculate descriptive by cohort. This will be done through DataShield and the dsBaseClient package/tool (1,2).
2) **EWAS**
   Then, the leading team will run EWAS for each of the models in each cohort. This will be done through DataShield and the dsOmicsClient package/tool (3). Summarized results will be saved locally at the leading team servers for the next steps.
2 **Meta-analyses**
   After that, the leading team will combine the results from each cohort through meta-analyse, using their preferred method/tool.
3 **Sensitivity analyses**
   Several sensitivity analyses are possible: by array, by ancestry, by region, leave-one-out, etc.
4 **Biological interpretation of findings**
   Finally, the leading team will compare the results across models and top CpGs will be annotated and several types of functional enrichment analyses will be conducted.
5 **Manuscript writing**
   The leading team will prepare a first draft of the manuscript which will be sent to all co-authors.

LifeCycle and ATHLETE recommend being at least two co-leading teams in each project:
1) To take advantage of the expertize of each team.
2) To decrease errors during the analyses.
3) To allow leadership by all partners in their topics of interest.

We recommend splitting the work as follows:
1) **Analysis plan:** to be designed by all co-leaders
2) **EWAS:** each co-leader runs the EWAS of ½ of the cohorts with around 10-20% of the cohorts being run by both teams.
3) **Meta-analyses:** to be run in parallel by the two co-leaders.
4) **Sensitivity analysis and downstream functional enrichment analysis:** to be split among the co-leaders.
5) **Manuscript writing:** to be split among the co-leaders.

What is an ExpressionSet?

An ExpressionSet is an R object designed to combine several different sources of information into a single convenient structure. It contains:

- A matrix with the DNA methylation data (CpGs in rows and samples in columns).
- A dataframe, specifically an AnnotadedDataFrame, with metadata about samples (samples in rows and variables in columns).

NOTE! Column names of the methylation matrix have to be the same row names of the metadata.

- An annotation data frame describing the probes/CpGs included in the ExpressionSet.

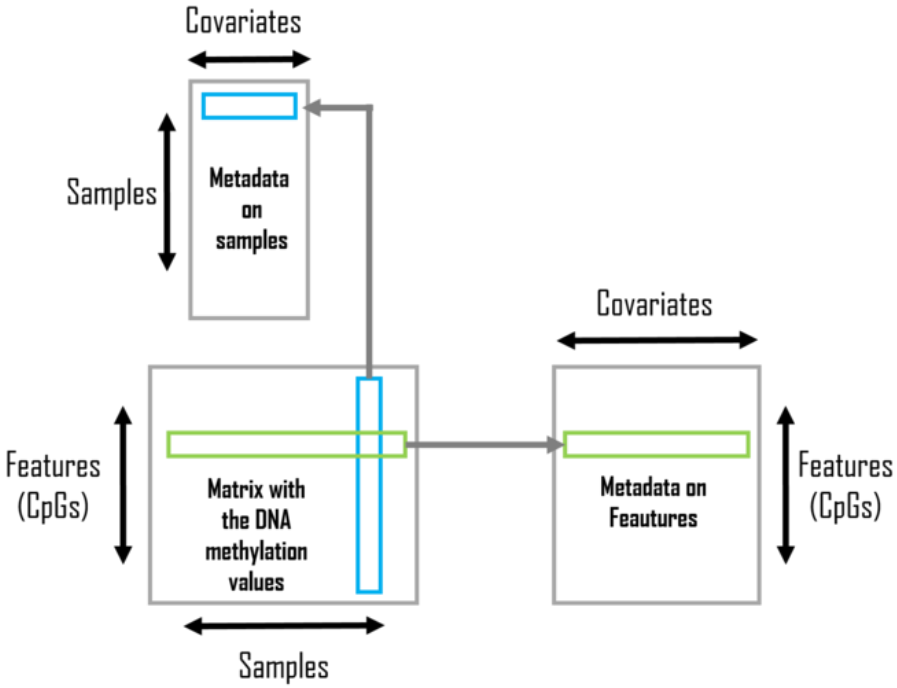

Figure 2. Representation of the structure an ExpressionSet

How many ExpressionSets have to be created?

The cohorts will have to create an ExpressionSet for each:

- **Subject type:**
  - Child
  - Mother
  - Father
- **Tissue:**
  - Blood
  - Placenta
  - Saliva
  - Buccal epithelial cells (BEC)
  - Nasal epithelial cells (NEC)
- **Age:**
  - For mothers during pregnancy, mean weeks of pregnancy of the visit: 12 weeks (12w), 20w, 32w, etc.
  - For children, at birth: 0 years (0y)
  - For children, mean age of the postnatal visit: 2y, 4y, 7y, etc.
  - For parents, during postnatal visits: Adult
- **Array:**
  - 450K
  - EPIC

ExpressionSets will have to be named as:
Cohort_Methyl_Subject_Tissue_Age_Array_Date.Rdata
(ie. INMA_Methyl_Child_Blood_0y_450K_20210215.Rdata).

The ExpressionSets only have to be created once, as the same ExpressionSets will be used in different projects of LifeCycle and ATHLETE (after signing the proper Data Access Agreements). Only in the case where we decide to apply a cross-cohort harmonize pipeline for the quality control and normalization of the methylation data, the cohorts will have to create new ExpresisonSets.
Besides the ExpressionSet, the cohorts will have to prepare a document named **Cohort_Methyl_Subject_Array_Blood_Age_Methods_Date.doc** (ie. INMA_Methyl_Child_Blood_0y_450K_Methods_20210513.doc, attached as an example), giving methodological details about the DNA methylation data. It should include the following information:

- Sample collection and DNA extraction (age, blood fractioning, placental region, extraction method, etc.)
- DNA methylation acquisition (array, laboratory, randomization, etc.)
- QC and normalization and batch control (sample QC, probes QC, normalization and batch correction, etc.).
- Additional information of the metadata in the ExpressionSet. Please give us details on which variables should be used to adjust the models for (batch_methyl, cohort_methyl, sel_methyl, gwas_pcs, etc), and their definition.

Which type of methylation data has to be included in the ExpressionSet?
DNA methylation assessed with the Illumina Infinium 450K or EPIC arrays and expressed as beta values (0 = completely un-methylated, 1 = completely methylated).
Each cohort can use their preferred normalization method and quality control pipeline. At this point, include rather than exclude probes. Exclusion of probes will be done by leaders of each project.

Which metadata has to be included in the ExpressionSet?
The metadata file of the ExpressionSet should include the following variables, with the names and levels/units indicated:

- **ID1 (id_methyl):** ID used in the methylation dataset. It is the row name of the metadata and has to be consistent with the column name of the methylation matrix.
- **ID2 (id):** ID used in the cohort (it might be or not the same ID as the methylation dataset). It will be used to link the DNA methylation data with the other data harmonized in LifeCycle and ATHLETE (EU Child Cohort Variable Catalogue).
- **Sex (sex_methyl):** Sex of the participant [1=Male, 2=Female].
- **Age (age_methyl):** Age of each participant when DNA methylation was assessed [continuous, years].
- **Ancestry (anc_methyl):** Ancestry of the participant. Major ethnic groups [1=European, 2=African or African American, 3=East Asian, 4=South Asian, 5=Native or admixed American, 6=Other]. If you cannot create a variable with the suggested levels, then add

the most appropriate variable for your cohort study and describe it in the Methods file. If possible, use genetic data to define ancestry. If genetic data is missing for a subset of subjects, then complement it with self-reported ancestry. If genetic data is not available or it has many missing values, use self-reported ancestry.

- **Gestational age (ga_methyl):** Gestational age of each participant [continuous, days]. Only for cord blood DNA methylation.
- **Birth weight (bw_methyl):** Weight of the participant at birth [continuous, grams]. Only for cord blood DNA methylation.
- **Child/adolescent body mass index (BMI) z-score (zbmi_methyl)**: Body mass index (BMI) z-score calculated at the age when blood DNA methylation was assessed, defined following the WHO standard curves for children (4,5) [continuous, z-score]. Only for samples from children or adolescents (from 2 to 18 years old). If you need help to calculate it, contact us.
- **Adult body mass index (BMI) (bmi_methyl):** Body mass index (BMI) at the age when blood DNA methylation was assessed. Only for samples from adult participants (>18 years old). For maternal samples obtained during pregnancy use maternal pre-pregnancy BMI or maternal early-pregnancy BMI (up to 16th week of gestation). Indicate it in the Methods file.
- **Cell type proportions:** Blood cell type proportions calculated from different reference panels. Note that depending on the type of biological samples and age, the reference panel and algorithm will be different. We are providing the code to calculate them, if not done yet in your study.

  A) Cord blood cell type proportions: Cord blood cell proportions should be estimated using the Gervin and Salas reference panel (6), the IDOL algorithm (7) for selection of 517 CpGs (for 450K and EPIC arrays) (8) and the constrained projection-quadratic programming algorithm by Houseman (6) for deconvolution of 7 main blood cell types.
    - Variable names: CD8T, CD4T, NK, Bcell, Mono, Gran, nRBC
    - Code: LC-ATH_Deconv_CordBlood_Gervin_Code_v1_20210513.R(attached to this SOP)

  B) Child and adult blood cell type proportions: Child and adult blood cell proportions should be estimated using two reference panels:

  B.1) The Reinius reference panel (9) with the pickCompProbes method (minfi) for CpG selection, and the Houseman algorithm (10) for deconvolution of 6 main blood cell types (450K array):
    - Variable names: CD4T_H, CD8T_H, NK_H, Bcell_H, Mono_H, Gran_H
    - Code: LC-ATH_Deconv_AdultBlood_Houseman_Code_v1_20210513.R (attached to this SOP)

  B.2) The Salas reference panel (8) with the IDOL algorithm (7) for CpG selection (450 CpGs for EPIC and 350 CpGs for 450K) and the Houseman algorithm (10) for deconvolution of 6 main blood:
    - Variable names: CD4T_S, CD8T_S, NK_S, Bcell_S, Mono_S, Neu_S
    - Code: LC-ATH_Deconv_AdultBlood_Salas_Code_v1_20210513.R (attached to this SOP)

NOTE: We would like you to include in the ExpressionSet cell type proportions calculated with both the Reinius and Salas reference panels naming them as we explained above. In this way, each leading group will be able to select one or the other for their EWAS.

- **Technical batch variable (batch_methyl):** Optional. Models will be adjusted for the batch variable suggested by the cohort, unless batch effect has already been corrected during the quality control through methods such as ComBat. Batch variable will be specific to each ExpressionSet created at each age.
- **Cohort specific variable (cohort_methyl):** Optional. If your cohort includes several centres or subcohorts, add this variable in the ExpressionSet.
- **Selection factor (sel_methyl)**: Optional. If your study population oversampled on a condition, then you should add this variable in the ExpressionSet (e.g., if your study population is from a case-control study of asthma, then asthma status should be included in the ExpresionSet).
- **GWAS PCs (gwas_pc1, gwas_pc2, etc.)**: Optional. Models will be run by major ancestry groups (e.g. European). However, if your cohort might still have population substratification within major ancestry groups, then models will be adjusted for GWAS PCs. Whether this is needed and, if so, the number of GWAS PCs to be included as covariates has to be determined by the cohort.

NOTES:

- ID1 (id_methyl). It is essential! We will use it to link methylation data with metadata.
- ID2 (id). It is essential! We will use it to link methylation data and metadata with other harmonized variables from LifeCycle and/or ATHLETE.
- It is really important that you name and level the variables as indicates in this SOP. Any deviation in the name or definition should be indicated in the Methods file.
- If your cohort has twins or siblings, you should keep just one of them in the ExpressionSet. Select them randomly. Removing this sample from the metadata will be enough to exclude this sample from the ExpressionSet.
- Samples with missing values in the covariates (ei. with missing values in birth weight, ancestry, etc.) do NOT have to be filtered out from the ExpressionSet. Each project will have a different set of covariates and will decide how to deal with missing values.

Which is the code to create an ExpressionSet with methylation data?

A tutorial and R code to create the ExpressionSets with methylation data are attached to this SOP. They also contain the code to link the ExpressionSet as a resource to your Opal server.

LC-ATH_ExprSet_MethBlood_Tutorial_v2_20210623.html
LC-ATH_Code_Methyl-ExpressionSet_v2_20210623.R

NOTE: If you have questions, please contact sofia.aguilar@isglobal.org.

References to Appendix

1. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol. 2014;43(6):1929-44.
2. datashield/dsBaseClient: DataSHIELD client side base functions. Available at: https://github.com/datashield/dsBaseClient. (Accessed: 5th February 2021)

3. GitHub - isglobal-brge/dsOmicsClient: dsOmic client site base functions. Available at: https://github.com/isglobal-brge/dsOmicsClient. (Accessed: 15th October 2020)
4. WHO | Application tools. WHO (2018).
5. WHO | WHO Anthro Survey Analyser and other tools. WHO (2019).
6. Gervin K, Salas LA, Bakulski KM, van Zelm MC, Koestler DC, Wiencke JK, et al. Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. Clin Epigenetics. 2019;11(1):125.
7. Koestler DC, Jones MJ, Usset J, Christensen BC, Butler RA, Kobor MS, et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinformatics. 2016;17:120.
8. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. 2018;19(1):64.
9. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7(7):e41361.
10. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.