*Environ Health Perspect*

DOI: 10.1289/EHP4522



## Supplemental Material

# Prenatal Particulate Air Pollution and DNA Methylation in Newborns: An Epigenome-Wide Meta-Analysis

Olena Gruzieva, Cheng-Jian Xu, Paul Yousefi, Caroline Relton, Simon Kebede Merid, Carrie V. Breton, Lu Gao, Heather E. Volk, Jason I. Feinberg, Christine Ladd-Acosta, Kelly Bakulski, Charles Auffray, Nathanaël Lemonnier, Michelle Plusquin, Akram Ghantous, Zdenko Herceg, Tim S. Nawrot, Costanza Pizzi, Lorenzo Richiardi, Franca Rusconi, Paolo Vineis, Manolis Kogevinas, Janine F. Felix, Liesbeth Duijts, Herman T. den Dekker, Vincent W. V. Jaddoe, José L. Ruiz, Mariona Bustamante, Josep Maria Antó, Jordi Sunyer, Martine Vrijheid, Kristine B. Gutzkow, Regina Grazuleviciene, Carles Hernandez-Ferrer, Isabella Annesi-Maesano, Johanna Lepeule, Jean Bousquet, Anna Bergström, Inger Kull, Cilla Söderhäll, Juha Kere, the Biobank-based Integrative Omics Studies (BIOS) Consortium, Ulrike Gehring, Bert Brunekreef, Allan C. Just, Rosalind J. Wright, Cheng Peng, Diane R. Gold, Itai Kloog, Dawn L. DeMeo, Göran Pershagen, Gerard H. Koppelman, Stephanie J. London, Andrea A. Baccarelli, and Erik Melén

## Table of Contents

**Table S6.** Correlation analysis of DNA methylation and gene transcription levels in *cis* using the BIOS data (n=3,075).

**Table S7.** Distribution of expression levels of genes associated with CpG methylation in response to maternal $PM_{10}$ or $PM_{2.5}$ exposure in newborn children of the EARLI cohort.

**Table S8.** Distribution of gene expression levels of genes associated with CpG methylation in response to maternal $PM_{10}$ or $PM_{2.5}$ exposure in 16-year-old children of the BAMSE cohort.

**Table S9.** Top 10 significant gene ontology terms derived from pathway analysis based on CpGs significantly associated with prenatal $PM_{10}$ exposure in the discovery meta-analysis at a cut-off of $P<10^{-5}$ using missMethyl method.

**Figure S1.** Quantile-quantile plot for epigenome-wide meta-analysis of the association between A) prenatal $PM_{10}$ (n=1,949) and B) prenatal $PM_{2.5}$ exposure (n=1,551) and cord blood DNA methylation.

**Figure S2.** Forest plots for six epigenome-wide significant associations between prenatal $PM_{10}$ exposure and cord blood DNA methylation.

**Excel Table S1.** Statistically significant DMRs (FDR $p<0.01$) in relation to prenatal $PM_{10}$ exposure identified by the DMRcate method.

**Excel Table S2.** Statistically significant DMRs (FDR $p<0.01$) in relation to prenatal $PM_{10}$ exposure identified by the comb_p method.

**Excel Table S3.** Statistically significant DMRs (FDR $p<0.01$) in relation to prenatal $PM_{2.5}$ exposure identified by the DMRcate method.

**Excel Table S4.** Statistically significant DMRs (FDR $p<0.01$) in relation to prenatal $PM_{2.5}$ exposure identified by the comb_p method.

**Additional File-** Excel Document

**References**

## Materials and Methods

### ALSPAC

*Study population*

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a population-based birth cohort drawn from the South West of England. 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were recruited, and detailed information on the mothers and their offspring has since been regularly collected (Boyd et al. 2013; Fraser et al. 2013). The study website contains additional details on the population and all the data that are available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Written informed consent has been obtained for all ALSPAC participants. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. The Accessible Resource for Integrated Epigenomic Studies (ARIES, http://www.ariesepigenomics.org.uk/) is a sub-sample of ALSPAC with DNA methylation measurement repeated at multiple time points (Relton et al. 2015). The resource contains longitudinal DNA methylation data on approximately 1000 mother-child pairs.

*DNA methylation measurement*

Following DNA extraction, samples were bisulfite converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA). Following conversion genome-wide methylation was measured using the Illumina HumanMethylation450 BeadChip. The arrays were scanned using an Illumina iScan, with initial quality review using GenomeStudio. During the data generation process a wide range of batch variables were recorded in a purpose-built laboratory information management system (LIMS). The LIMS also reported quality control (QC) metrics from the standard control probes on the 450k BeadChip for each sample. Samples failing QC were excluded from further analysis and the assay repeated. In total, there were 1127, 1086 and 1073 DNA methylation profiles generated from cord, peripheral blood at age 7 and peripheral blood at age 15-17, respectively. Of these, 914, 980 and 981 profiles passed quality control. Quality control included genotype and sex reporting checks, as well as measurement quality assessment using standard quality control plots for the Illumina Infinium HumanMethylation450 BeadChip. Quality control and normalization was completed using the meffil R package (Min et al. 2018) in R version 3.2.0. The Functional Normalization algorithm (Fortin, Labbe, et al. 2014) as implemented in the meffil R package (Min et al.

2018) was applied to normalize between microarrays. The top 10 control probe principal components were used in the normalization.
In all EWAS, batch effects were handled by including 10 surrogate variables calculated from the normalized DNA methylation data using surrogate variable analysis (Leek et al. 2007).

*Air pollution exposure assessment*
Exposure to particulate matter ≤10 μm ($PM_{10}$) was modeled using dispersion modeling of annual average exposures for the period 1990−2008 based on daily total $PM_{10}$ assessed at maternal residential addresses (including address changes) (Plusquin et al. 2018). For exposure during pregnancy trimesters we used local (ADMS-Urban) and regional/longrange (NAME-III) air pollution models. Total pregnancy $PM1_0$ exposure was then computed as a time-weighted average across all trimesters. We accounted for changes in address in all periods using a bespoke algorithm developed at Imperial College London: Algorithm for Generating Address-History and Exposures (ALGAE; https://smallareahealthstatisticsunit.github. io/algae/index.html).

**BAMSE EpiGene**

In the BAMSE cohort we had an independent methylation dataset (in addition to the BAMSE MeDALL dataset; see below).(Gruzieva et al. 2017) Epigenome-wide DNA methylation was measured in 472 Caucasian children, using DNA extracted from blood samples collected at the age of 8 years. An aliquot (500 ng) of DNA per sample underwent bisulfite conversion using the EZ-96 DNA Methylation kit (Zymo Research Corporation, Irvine, USA). Samples were plated onto 96-well plates in randomized order. Samples were processed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA).

Quality control of analysed samples was performed using standardized criteria. Samples were excluded in case of sample call rate <99%, colour balance >3, low staining efficiency, poor extension efficiency, poor hybridization performance, low stripping efficiency after extension and poor bisulfite conversion. We also applied multidimensional scaling (MDS) plot to evaluate gender outliers based on chromosome X data, that produced two separated clusters for male and female. We omitted 5 samples that do not belong to the distinct cluster. Furthermore, we applied median intensity plot for methylated and unmethylated intensity by using the minfi R package (3 samples below the 10.5 cutoff were excluded). All above led to

exclusion of 8 samples. We implemented "DASEN" recommended from wateRmelon package to do signal correction and normalization.(Pidsley et al. 2013)

**The Children's Health Study (CHS)**

*Study population*

The CHS is a population-based prospective cohort study from age 5 onwards in Southern California, which has been described in detail elsewhere (McConnell et al. 2006). The study protocol was approved by the University of Southern California Institutional Review Board and informed, written consent and assent were provided by the parents and children respectively. A total of 5341 children were recruited, all of whom were born between 1995 and 1997 and are currently being followed until age 18.

*DNA methylation measurement*

Based on the availability of newborn bloodspots archived by the state of California, a subset of 273 children was selected for a sub-study in which epigenome-wide DNA methylation was assessed in newborn bloodspots using the Infinium HumanMethylation450 BeadChip (HM450). Laboratory personnel performing DNA methylation analysis were blinded to study subject information. DNA was extracted from whole blood cells using the QiaAmp DNA blood kit (Qiagen Inc, Valencia, CA) and stored at -80 degrees Celsius. 700-1000ng of genomic DNA from each sample was treated with bisulfite using the EZ-96 DNA Methylation Kit™ (Zymo Research, Irvine, CA, USA), according to the manufacturer's recommended protocol and eluted in 18 μl. The results were compiled for each locus and were reported as beta (β) values. A normal-exponential background correction with dye bias correction was applied to the raw intensities at the array level to reduce background noise (Triche et al. 2013). We then normalized each sample's methylation values to have the same quantiles to address sample to sample variability.(Touleimat and Tost 2012) CpG loci on the HM450 array were removed from analyses if they were on the X and Y chromosomes, or if they contained SNPs, deletions, repeats, or if they have more than 10% missing values, leaving a total of 384,310 probes for analysis. Seven cord blood cell sub-populations were estimated using the regression calibration approach algorithm described by Bakulski *et al.* that was implemented in minfi.(Jaffe and Irizarry 2014; Bakulski et al. 2016). Estimated cord blood cell subpopulations (CD8+ T-lymphocytes, CD4+ T-lymphocytes, natural killer cells, B-lymphocytes, monocytes, granulocytes and nucleated red blood cells) were subsequently included as linear predictors in regression models.

*Air pollution exposure assessment*

The CHS air quality monitoring data (Gauderman et al. 2007; Peters, Avol, Navidi, et al. 1999; Peters, Avol, Gauderman, et al. 1999) and the US EPA air Quality System (AQS) were used to assign estimates of prenatal air pollution exposures for $PM_{2.5}$ and $PM_{10}$, based on residential address reported on the birth certificate and at the time of the baseline questionnaire. Addresses were geocoded using TeleAtlas Inc.'s Address Point Geocoding Services. Station-specific air quality data were spatially interpolated to each birth residence using inverse-distance-squared weighting. The data from up to four air quality measurement stations were included in each interpolation. Due to the regional nature of $PM_{10}$ and $PM_{2.5}$ concentrations, a maximum interpolation radius of 50 km was used for all pollutants. However, when a residence was located within 5 km of one or more stations with valid observations, the interpolation was based solely on the nearby values. When multiple addresses were reported, the average concentrations were time-weighted to account for portion of year spent at each address. Individuals with incomplete residential histories were excluded from analyses.

**EARLI**

*Study population*

The Early Autism Risk Longitudinal Investigation (EARLI) is an enriched risk prospective pregnancy cohort to study autism etiology.(Newschaffer et al. 2012) The EARLI study was reviewed and approved by Human Subjects Institutional Review Boards (IRBs) from each of the four study sites (Johns Hopkins University, Drexel University, University of California Davis, and Kaiser Permanente Northern California). This longitudinal study recruited mothers of confirmed ASD child who were early in a subsequent pregnancy or were trying to become pregnant. There were 232 mothers with a subsequent sibling born through this study, between November 2009 and March 2012.

*DNA methylation measurement*

Biospecimens including cord blood and placenta were collected and archived at 213 births. Cord blood DNA was extracted using the DNA Midi kit (Qiagen, Valencia, CA) and samples were bisulfite treated and cleaned using the EZ DNA methylation gold kit (Zymo Research, Irvine, CA). DNA was plated randomly and assayed on the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA) at the Johns Hopkins SNP Center, a shared lab and informatics operation with the Center for Inherited Disease Research (Johns Hopkins

University). Methylation control gradients and between-plate repeated tissue controls were used.

We used the minfi library (version 1.18.2) in R (version 3.3) to process raw Illumina image files into noob background corrected methylation values.(Aryee et al. 2014; Triche 2014) Probes with failed detection P-value (>0.05) in >10% of samples were removed (n=508). Samples with discordant methylation predicted sex and observed sex were removed (n=2) as were samples that appeared as outliers on the first principal component of methylation data across the genome prior to normalization (n=2). We adjusted normalized data for batch effects using ComBat in the sva package (version 3.9.1).(Leek and Storey 2007) Estimation of six different white blood cell types (CD8+ T and CD4+ T lymphocytes, CD56+ natural killer cells, CD19+ B cells, CD14+ monocytes, and granulocytes) by Houseman method(Houseman et al. 2012) was performed using the default implementation of the *estimateCellCounts* function in the minfi package.(Jaffe and Irizarry 2014)

After pre-processing, DNAm data was available for 166 cord blood samples across 455,698 probes on the Illumina Infinium HumanMethylation450 BeadChip in EARLI.

*Gene expression measurement*

Umbilical cord blood samples were collected after delivery at each clinical lab site in PaxGene 2.5mL collection tubes (Fisher, Cat # BD762165) and then shipped same-day delivery at ambient temperature to Johns Hopkins Biological Repository (JHBR) in Baltimore, Maryland, where they were incubated at -20°C for 24 hours before transfer to long-term storage at -80°C prior to processing.

RNA was isolated at JHBR on a QIAgen QIAsymphony automated workstation using the PAXgene Blood RNA kit as per manufacturer's instructions. RNA yields were quantified by NanoDrop and quality was assessed and Agilent BioAnalyser, respectively. For the RNA samples that met our quality control measures (RIN $\geq$ 7, A260/230 $\geq$ 1.5), biotin-labeled cRNA was generated and hybridized to Human Gene 2.0 ST Affymetrix microarray chips by the lab staff at the Johns Hopkins University SNP Center.

A standardized Affymetrix protocol was followed for all washing, staining and scanning procedures. Raw fluorescence data (in Affymetrix CEL file format) with one perfect match and one mismatched probe in each set were analyzed using oligo package in R. Linear regression was then performed on normalized expression values using the same models as the methylation analyses.

*Air pollution exposure assessment*

As part of EARLI, maternal residential histories, reflecting addresses and dates of residence, during pregnancy were collected. Addresses were standardized and geo-coded using the TeleAtlas US_Geo_2 database and software (Tele Atlas, Inc., Boston, CA, www.geocoded.com). Air quality assignments for $PM_{2.5}$ and $PM_{10}$ were derived from the US EPA's Air Quality System (AQS) data (www.epa.gov/ttn/airs/airsaqs). Pregnancy averages were computed based on the monitor data for that period. We used inverse distance-squared weighting to assign PM measures from up to four closest stations located within 50 km of each participant residence. However, if one or more stations were located within 5 km of a residence then only data from the stations within 5 km were used for the interpolation. A total of 166 participants with PM data and cord DNAm were included in the analysis.

**EXPOsOMICS project**

Within the EXPOsOMICS collaborative European project, a combination of three population-based birth cohorts ENVIRonmental influence ON AGEing in early life (ENVIR*ON*AGE), Rhea and Piccolipiu was established to conduct DNA methylation analyses (Vineis et al. 2017). The phenotypic variables were harmonized across the three cohorts, and their biospecimen were semi-randomized on the DNA methylation arrays such that the latter would incorporate proportional representations of the three cohorts and that batch effects do not completely confound with biological covariates of interest. In this work, the air pollution analysis was performed separately for each cohort.

*DNA methylation measurement*

Aliquots of cord blood samples (collected and frozen at birth at -80°C) were shipped on dry ice to the Epigenetics Group at the International Agency for Research on Cancer (IARC), Lyon, France, where DNA was extracted (QIAamp 96 DNA Blood Kit, Qiagen 51161), quantified (Quant-iT PicoGreen dsDNA Assay Kit, Molecular Probes P7589) and bisulfite converted (600 ng of DNA using EZ-96 DNA Methylation kit, Zymo Research D5004). DNA methylation was measured at 485 577 CpGs using Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA). The arrays were designed such that batch effects (eg. sample position and intra- and inter-variability in arrays and chips) do not completely confound with biological covariates of interest. This design allows the retention of biological variation even after correction of technical variation.

Raw intensity (.idat) files were handled in R using the minfi-package to calculate the methylation level at each CpG as the beta-value (β=intensity of the methylated allele (M)/(intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100)), and the data were exported for quality control and processing. Methylation features were filtered from cross-reactive probes and low-quality probes (probes having bead counts < 3 in at least 5% of samples). Data quality was further assessed using box plots for the distribution of methylated and unmethylated signals, and multidimensional scaling plots and unsupervised clustering were used to check for sample outliers and potential gender mismatches, which were removed from the analysis. Also samples having >1% of CpG sites with a detection P-value >0.05 were removed. The remaining dataset was normalized using the funnorm normalization of the minfi package.

*Air pollution exposure assessment in ENVIRONAGE cohort*

The regional background levels of $PM_{10}$ and $PM_{2.5}$ were modelled for each mother's residential address using a high resolution spatial temporal interpolation method that combines data from land cover obtained from satellite images (Corine land cover data set) with measures from monitoring stations (n=34) and a dispersion model.(Maiheu et al.) Validation statistics of the model showed that the spatial explained variance ($R^2$) for annual mean $PM_{10}$ was more than 0.80.

Air pollution exposure assessment in Rhea and Piccolipiu is described below.

**The Generation R Study**

The Generation R Study is a population-based prospective cohort study from fetal life onwards in Rotterdam, the Netherlands, which has been previously described in detail.(Kooijman et al. 2016) Assessments in pregnant women consisted of physical examinations, fetal ultrasounds, biological samples, and questionnaires.(Kruithof et al. 2014) All children were born between April 2002 and January 2006. The study has been approved by the Medical Ethical Committee of the Erasmus MC, University Medical Center Rotterdam and written consent was obtained for all participating mothers and children. For the current study, data was available for 809 Caucasian mothers and their children with information on $PM_{10}$ exposure during pregnancy and DNA methylation at birth.

*DNA methylation measurements*

DNA was extracted from cord blood samples of 979 Caucasian children. Using the EZ-96 DNA Methylation kit (Shallow-well, Zymo Research Corporation, Irvine, USA), 500 ng DNA per sample underwent bisulfite conversion. Samples were transferred onto 96-well plates in a random order. Samples were processed with Illumina's Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA). Quality control of analyzed samples was performed using standardized criteria. Samples were excluded due to sample call rate <99% (n=7) or poor bisulfite conversion (n=1). In addition, 2 samples were excluded because of a gender mismatch and 1 sample because of a retracted informed consent, leaving a total of 969 samples in the statistical analysis. Probes with a single nucleotide polymorphism in the single base extension site with a frequency of >1% in the GoNLv4 reference panel were excluded, as were probes with non- optimal binding (non-mapping or mapping multiple times to either the normal or the 8 bisulphite-converted genome), resulting in the exclusion of 49,564 probes, leaving a total of 436,013 probes in the analysis. Data were normalized with DASES normalization using a pipeline adapted from that developed by Touleimat and Tost.(Touleimat and Tost 2012) DASES normalization includes background adjustment, between-array normalization applied to type I and type II probes separately, and dye bias correction applied to type I and type II probes separately. DASES is based on the DASEN method, but adds the dye bias correction, which is not included in DASEN.(Pidsley et al. 2013) Beta-values were calculated for all CpG sites.

**The Human Early-Life Exposome (HELIX)**

The HELIX study represents a collaborative project across six established and ongoing longitudinal population-based birth cohort studies in six European countries (EDEN-France, RHEA-Greece, KANC-Lithuania, MoBa-Norway, INMA-Spain, and BIB-United Kingdom). The aim of HELIX is to measure and describe multiple environmental exposures during early life (pregnancy and childhood) in a prospective exposome cohort and associate these exposures with molecular omics signatures and child health outcomes. In the HELIX subcohort of 1301 mother-child pairs, omics signatures and child health outcomes were measured at age 6-11 years as described elsewhere.(Vrijheid et al. 2014) The study was approved by the Ethical Committees of each participating centre and written consent was obtained from parents. This particular study was conducted in children participating in the HELIX subcohort from BIB, EDEN, KANC, and MOBA. RHEA and INMA children were excluded as they were already included in other parts of this study.

*DNA methylation measurement*

DNA was obtained from buffy coat collected in EDTA tubes at age 7-9y. Briefly, DNA was extracted using the Chemagen kit (Perkin Elmer) in batches of 12 samples. Samples were extracted by cohort. DNA concentration was determined in a NanoDrop 1000 UV-Vis Spectrophotometer (ThermoScientific) and with Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies).

DNA methylation was assessed with the Infinium HumanMethylation450 beadchip from Illumina, following manufacturer's protocol. Briefly, 700 ng of DNA were bisulfite-converted using the EZ 96-DNA methylation kit following the manufacturer's standard protocol, and DNA methylation measured using the Infinium protocol. A HapMap sample was included in each plate. In addition, 24 HELIX inter-plate duplicates were included. Samples were randomized taking into account cohort and sex.

DNA methylation data ware preprocessed using the minfi package.(Aryee et al. 2014) We increased the stringency of the detection p-value threshold to $10^{-16}$, and probes not reaching a 98% call rate were excluded (Lehne et al. 2015). Two samples were filtered due to overall quality: one had a call rate <98% and the other did not pass quality control parameters of the MethylAid package (van Iterson et al. 2014). Then, data was normalized with the functional normalization method, which also includes Noob background subtraction and dye-bias correction (Triche 2014). After that, several quality control checks were performed. First, we checked sex consistency using the shinyMethyl package and two samples were excluded.(Fortin, Fertig, et al. 2014) Genetic consistency of duplicates was checked with the 450k genotypes. In addition, genetic consistency was evaluated in those samples that had GWAS data and two of them were excluded. Finally, duplicated samples and HapMap samples were removed as well as control probes, probes designed to detect SNPs and probes designed to measure methylation levels at non-CpG sites.

**INfancia y Medio Ambiente (INMA)**

The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project is a network of birth cohorts in Spain that aims to study the role of environmental pollutants in air, water and diet during pregnancy and early childhood in relation to child growth and development[24]. Mothers were enrolled at week 12 of pregnancy from 1997 to 2008 in seven regions of Spain (Flix, Granada, Menorca, Asturias, Gipuzkoa, Sabadell and Valencia). The cohort consisted of 3,768 children at birth. During the follow-up visits information on environmental exposures

and health outcomes (reproductive, growth and obesity, lung function, allergies and neurodevelopment) were assessed through questionnaires, biomarker measurements, clinical data, and physical exploration. The study website contains details of the design and data available in INMA project (http://www.proyectoinma.org/). The study was approved by the Ethical Committees of each participating centre and written consent was obtained from parents. Present study was conducted in children from INMA Sabadell subcohort.

*DNA methylation measurement*

Cord blood was extracted using the Chemagen kit (Perkin Elmer). DNA concentration was determined by NanoDrop spectrophotometer (Thermo Scientific) and with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies).

Methylation data was produced in two different laboratories as part of two different projects: in the Genome Analysis Facility of the University Medical Center Groningen (UMCG) in Holland, and in the Bellvitge Biomedical Research Institute (IDIBELL, Barcelona). Both laboratories used the recommended Illumina protocol for the Infinium HumanMethylation450 beadchip. Briefly, 500 ng of DNA was bisulfite-converted using the EZ 96-DNA methylation kit following the manufacturer's standard protocol, and DNA methylation measured using the Illumina Infinium HumanMethylation450 beadchip.

DNA methylation data were preprocessed using the minfi package.(Aryee et al. 2014) A series of steps were completed for quality control and data analysis. The first step was low quality sample removal. First, 2 samples with bad overall quality or with low detection p-value according to the output of the MethylAid package were removed (van Iterson et al. 2014). Then, we removed 3 samples whose sex was wrongly predicted using shinyMethyl.(Fortin, Fertig, and Hansen 2014) Following guidelines of Lehne work, (Lehne et al. 2015) we increased the stringency of the detection p-value threshold to $10^{-16}$ and we filtered 18 samples with a call rate lower than 98%. The second step was normalizing data with functional normalization. Correlation between SNP in replicates samples was checked and probes not measuring SNPs were discarded. 7,136 probes with a call rate lower than 95% were also removed. ComBat was applied to remove batch effect.(Johnson et al. 2007)

**MeDALL**

MeDALL (Mechanisms of the Development of ALLergy) is a collaborative project supported by the European Union under the Health Cooperation Work Programme of the 7th Framework programme (grant agreement number 261357).(Bousquet et al. 2011) MeDALL epigenetics studies include four birth cohorts: EDEN, INMA, BAMSE and PIAMA.

**MeDALL - PIAMA**

For the PIAMA birth cohort study, pregnant women were recruited in 1996-1997 during their second trimester of pregnancy from a series of communities in the North, West, and Centre of The Netherlands as described elsewhere.(Wijga et al. 2014) Non-allergic pregnant women were invited to participate in a "natural history" study arm. Pregnant women identified as allergic through a validated screening questionnaire were primarily allocated to an intervention arm with a random subset allocated to the natural history arm. The intervention involved the use of mite-impermeable mattress and pillow covers. The study started with 3,963 newborns. Information on the children's health, socio-demographic and lifestyle factors as well as residential characteristics was collected by questionnaire annually until age 8 years and then at ages 11 and 14 years.

**MeDALL – BAMSE**

BAMSE is a prospective population-based cohort study of children recruited at birth and followed during childhood and adolescence. Details of the study design, inclusion criteria, enrolment and data collection are described elsewhere (Wickman et al. 2002; Thacher et al. 2016). In short, 4,089 children born between 1994 and 1996 in four municipalities of Stockholm County were enrolled. At baseline, when the infant was approximately 2 months of age, parents completed a questionnaire that assessed residential characteristics, as well as socioeconomic and lifestyle factors. When children were 1, 2, 4, 8, 12 and 16 years, the parents completed questionnaires focusing on children's symptoms related to wheezing and allergic diseases, as well as various exposures. The survey response rates were 96%, 94%, 91%, 84%, 82% and 78%, respectively. Furthermore, blood was obtained at ages 4, 8 and 16 years from 2,605 (63.7%), 2,470 (60.4%) and 2,547 (62.2%) children, respectively. The baseline and follow-up studies were approved by the Regional Ethical Review Board, Karolinska Institutet, Stockholm, Sweden, and the parents of all participating children provided informed consent.

*MeDALL DNA methylation measurements*

In the MeDALL study, peripheral blood samples were collected from all consenting cohort participants, and DNA from peripheral and cord blood samples was extracted using the QIAamp blood kit (Qiagen, Inc, Valencia, CA) or equivalent protocols, followed by a

precipitation-based concentration using GlycoBlue (Ambion, Austin, Tex). DNA concentration was determined by Nanodrop measurement and picogreen quantification. 500 ng of DNA were bisulfite-converted using the EZ 96-DNA methylation kit following the manufacturer's standard protocol, and DNA methylation measured using the Illumina Infinium HumanMethylation450 beadchip (Illumina, Inc., San Diego, USA). DNA methylation data were preprocessed using the Minfi package.(Aryee et al. 2014)

In quality control, samples that did not provide significant methylation signals in more than 10% of probes (detection p-value=0.01) were regarded as bad quality samples and were directly removed. In addition, samples were excluded in case of low staining efficiency, low single base extension efficiency, low stripping efficiency of DNA from probes after single base extension, poor hybridization performance, poor bisulfite conversion and high negative control probe staining. Moreover, we used 65 SNP probes to check for concordances between paired DNA samples from the sample individual and assessed the methylation distribution of X-chromosome to verify gender. Paired samples which show Pearson correlation coefficient <0.9 were regarded as sample mixed ups and were excluded from the study. Furthermore, we excluded probes on sex chromosomes, probes that mapped on multi-loci, the 65 random SNPs assay, and probes that contained SNP(s) at the target CpG sites with a minor allele frequency >10% . A series of steps were completed for quality control and data analysis. First, we implemented sample filtering to remove bad quality and mixed up samples. Second, we filtered out the probes to remove the CpG sites which are not mapped to unique location on the genome and CpGs containing single nucleotide polymorphisms (SNPs) at the target site. Third, we implemented "DASEN" to perform signal correction and normalization.(Pidsley et al. 2013)

To remove bias in methylation profiles unrelated to underlying biological processes, we implemented a correction procedures based on 613 negative control probes presented in 450K arrays since these negative control probes are supposed to not relate to biological variation. Finally, we implemented principal component analysis (PCA) on control probes data, then, we performed 10000 permutation for controls probes data and selected principal components with p-value defined as to get the p-value of (number of var(random pc) >var(pc)) / (number of permutations) $<10^{-4}$. The methylation data for each CpG are thus the residuals from a linear model fitting incorporating the significant 5 PCs.

*BAMSE gene expression measurements*

Data on mRNA gene expression were available in 244 children aged 16 years BAMSE cohorts.(Gruzieva et al. 2017) Whole blood was collected in PAXGene tubes and RNA was extracted using PAXgene Blood RNA kit (QIAGEN, Courtaboeuf, France). Quantity of extracted RNA and quality assessment were performed with Dropsense96 (Trinean, Gentbrugge, Belgium) and Tapestation (Agilent, Les Ulis, France) instruments, respectively, discarding 2 samples. RNA of highest quality was selected for amplification, labeling and hybridization on Affymetrix HTA 2.0 Genechips using Affymetrix IVT kit (Affymetrix, Inc. USA) at the European Institute for Systems Biology and Medicine in Lyon. Data were then processed at the probesets level for RMA normalization using Expression Console Software from Affymetrix v1.4. Expression transcripts were annotated using version 35 of Affymetrix annotation. Automated cell count was obtained by flow cytometry performed at the Karolinska University Laboratory in Stockholm, Sweden.

*Air pollution exposure assessment in Rhea, Piccolipiu, MeDALL, INMA, HELIX, Generation R and BAMSE EpiGene studies*

The procedures for measurements and LUR modeling have been extensively described elsewhere.(Eeftens et al. 2012; Pedersen et al. 2013) In short, 20 sampling sites for particles and other agents were selected in each study area to characterize the spatial distribution of the cohort addresses, including regional background, urban background, and traffic sites. Measurements were performed at each site 3 times during 2 weeks in the cold, warm, and intermediate seasons, and the results were averaged to estimate the annual average. LUR models for $PM_{10}$ and $PM_{2.5}$ were developed based on measured annual average concentrations by using a range of Geographic Information System–derived predictor variables selected through a supervised stepwise procedure. Modeling was done locally at each center according to a common exposure assessment manual (http://www.escapeproject.eu/manuals/) following harmonized procedures regarding air pollutants measurements, development of land use regression models, and validation.(Beelen et al. 2013) For the present analyses, total $PM_{10}$ and $PM_{2.5}$ exposure levels averaged throughout entire pregnancy was used. Data from routine monitoring stations were used to temporally adjust the LUR estimates to the periods corresponding to each individual pregnancy. The current exposure was estimated by assignment of modeled annual average $PM_{10}$ and $PM_{2.5}$ concentrations to the current addresses at the time when the blood samples were collected.

**PRISM**

*Study population*

The PRogramming of Intergenerational Stress Mechanisms (PRISM) study is a prospective pregnancy cohort of mother-child pairs originally designed to examine how perinatal stress influences respiratory health in children. Procedures were approved by the Institutional Review Boards at the Brigham and Women's Hospital (BWH) and the Icahn School of Medicine at Mount Sinai. Beth Israel Deaconess Medical Center (BIDMC) relied on BWH for review and oversight of the protocol. Written consent was obtained from all participants. Women were recruited from prenatal clinics during the first or second trimester (<28 weeks gestation) from the Beth Israel Deaconess Medical Center (BIDMC) and the East Boston Neighborhood Health Center in Boston Massachusetts, USA, from March 2011 to August 2012. Recruitment sites were chosen to ensure desired heterogeneity in sociodemographic and racial/ethnic characteristics. Eligibility criteria included: (i) English- or Spanish-speaking; (ii) age ≥18 years at enrollment; and (iii) singleton pregnancy.

*DNA methylation measurement*

Cord blood DNA was isolated using Qiagen Tissue DNA extraction kits (Qiagen, Valencia CA) and quantified using an Implen Nanophotometer Pearl (Westlake Village, CA). 500ng of DNA was bisulfite-treated using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Orange, CA) and analyzed by the Illumina Infinium HumanMethylation450 array. Samples were arranged on chips and plates with a stratified randomization followed by statistical checks for balance on birthweight z-score, gestational age, sex, and city of collection. The presence of failed arrays or outliers was checked with detection p-values (all samples passed with detection p-values <0.05 in >99% of probes) and through visualization of principal components analysis (PCA). Potential batch effects were further assessed with five pairs of technical replicates per tissue that were arranged across chips and plates. Sample identity was checked via imputed sex and agreement of genotype with paired tissues. Probes with detection p-values >0.05 or beadcount <3 in >1% of samples were dropped using the wateRmelon pfilter function. Data were preprocessed using background correction (Triche et al. 2013), dye bias and probe type adjustment (Teschendorff et al. 2013). BMIQ (Beta Mixture Quantile dilation) intra-sample normalization was applied to all probes to adjust the methylation values of Infinium II probes into a statistical distribution characteristic of Infinium I probes.

*Air pollution exposure assessment*

We estimated $PM_{2.5}$ concentration at each participant's residential address using a hybrid land use regression and satellite-based model. In brief, we used MODIS (moderate resolution imaging spectroradiometer) satellite-derived aerosol optical depth (AOD) measurements to predict daily $PM_{2.5}$ concentration levels at a 1km * 1km spatial resolution. Ground level predictions at EPA monitoring stations were made calibrating the daily AOD with mixed effects models including land use regression and meteorological variables. A smoothing stage interpolated missing estimates using seasonal generalized additive models and local ground stations measurements to capture temporal dynamics. A detailed description of the $PM_{2.5}$ prediction model has been published elsewhere (Kloog et al. 2014). We calculated exposure using the average daily predicted $PM_{2.5}$ concentration at the 1km * 1km grid cell centroid closest to the residential address over the 1-year period prior to delivery.

**Project Viva**

*Study population*

Project Viva enrolled 2670 pregnant women during 1999-2002 at their initial obstetrical visit at Atrius Harvard Vanguard Medical Associates in Boston Massachusetts (2128 pregnant women had live births). Detailed protocols have been published previously.(Oken et al. 2015) Three hundred and twenty-seven mother-child pairs had complete information on cord blood DNA methylation and annual $PM_{2.5}$ measures before delivery. Written informed consent was obtained for all children in both studies. All study protocols were reviewed and approved by the Institutional Review Boards of the participating institutions.

*DNA methylation measurement*

In Project Viva, we extracted cord blood and peripheral blood DNA using Qiagen Puregene Kit (Valencia, CA) and bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA). We randomized samples by chips and plates and generated DNA methylation data using the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA). In the sample quality control step, we removed samples that were technical replicates, samples with low quality, and samples with genotype or sex mismatches. We removed low quality probes (based on detection p-values >0.05), probes on sex chromosomes, SNP probes, probes within 10bp of a known SNP with minor allele frequency ≥ 1%, non-CpG probes, as well as non-specific and cross-reactive probes. We applied the normal-exponential out-of-band ("noob") method for background correction and dye bias adjustment, and we normalized our data using Beta Mixture Quantile Dilation (BMIQ) - a model based intra-array

normalization strategy to reduce bias of type 2 probes, while eliminates type 1 probe enrichment bias.(Teschendorff et al. 2013) We reported methylation as beta values, which were expressed as the percentage of methylated (M) cytosine residuals over the sum of methylated (M) and un-methylated (U) cytosine residuals at the 5 carbon position (i.e., beta value = [M/(M+U)]*100%).

*Air pollution exposure assessment*

We estimated $PM_{2.5}$ concentration at each participant's residential address using a hybrid land use regression and satellite-based model. In brief, we used MODIS (moderate resolution imaging spectroradiometer) satellite-derived aerosol optical depth (AOD) measurements to predict daily $PM_{2.5}$ concentration levels at a 1km spatial resolution. Daily AOD was calibrated using ground $PM_{2.5}$ measurements, land use regression and meteorological variables. A detailed description of the $PM_{2.5}$ prediction model has been described elsewhere (Kloog et al. 2012). We calculated the average of $PM_{2.5}$ concentration at residential address 1-year prior to delivery.

**Statistical analyses**

**Table S1. Input parameters for DMR calling.**

| Input parameter | Value | Description |
|---|---|---|
| **Comb-p algorithm** | | |
| dist | 1000 | Maximum distance to search for adjacent peaks |
| seed | 0.05 | A value must be at least this large/small in order to seed a region |
| region-filter-p | 0.01 | Maximum adjusted region-level p-value to be reported in final output |
| region-filter-n | 2 | Require at least this many probes for a region to be reported in final output |
| **DMRcate algorithm** | | |
| lambda | 1000 | Gaussian kernel bandwidth for smoothed-function estimation. Gaps ≥ lambda between significant CpG sites will be in separate DMRs |
| **C** | 2 | Scaling factor for bandwidth. Gaussian kernel is calculated where lambda/C = sigma. Empirical testing shows that, for 450k data when lambda = 1000, near-optimal prediction of sequencing-derived DMRs is obtained when C is approximately 2. |
| **Pcutoff** | 0.01 | P-value cutoff to determine DMRs. |
| **min.cpgs** | 2 | Minimum number of consecutive CpGs constituting a DMR |

**Table S2. Characteristics of cohorts included in the discovery EWAS meta-analysis.**

| Characteristic | Newborns | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **INMA (Spain)** | **Generation R (the Netherlands)** | **CHS (California, USA)** | **EARLI (USA)** | **PRISM (Massachusetts, USA)** | **Project ViVa (Massachusetts, USA)** | **ENVIRONAGE (Belgium)** | **Rhea (Crete)** | **Piccolipiu (Italy)** |
| | (n=355) | (n=809) | (n=227 in $PM_{10}$; n=175 in $PM_{2.5}$) | (n=166) | (n=136) | (n=327) | (n=194) | (n=99) | (n=98) |
| $PM_{2.5}$ during pregnancy, $\mu g/m^3$: Median (IQR) | 15.1 (2.1) | NA | 26.3 (5) | 9.9 (2.3) | 8.1 (0.8) | 12.1 (1.2) | 11.9 (1.9) | 14.3 (2.1) | 30.6 (0.9) |
| $PM_{10}$ during pregnancy, $\mu g/m^3$: Median (IQR) | 27.5 (4.7) | 29.1 (5.9) | 40.7 (14.0) | 17.7 (3.2) | NA | NA | 17.2 (2.8) | 35.6 (2.3) | 48.5 (6.2) |
| Male sex (%) | 197 (51.0) | 427 (52.8) | 93 (41.0)/69 (39.4) | 85 (51.2) | 75 (54.7) | 177 (54.0) | 98 (50.5) | 53 (53.5) | 53 (54.1) |
| Maternal smoking during pregnancy (%) | 48 (17.1) | 200 (24.7) | 14 (6.2) / 9 (5.1) | 0 | 31 (22.6) | 34 (10.0) | 24 (12.4) | 20 (10.3) | 21 (21.4) |

IQR= interquartile range (the difference between the $75^{th}$ and $25^{th}$ percentiles); NA = not available

**Table S3. Study-specific lambda values.**

| Cohort | No. of CpGs included in the analysis | Lambda (λ) | |
|---|---|---|---|
| | | Models with $PM_{10}$ exposure | Models with $PM_{2.5}$ exposure |
| INMA | 476 946 | 1.11 | 1.39 |
| Generation R | 436 013 | 1.39 | NA |
| CHS | 384 310 | 1.01 | 1.07 |
| EARLI | 455 698 | 1.08 | 1.68 |
| PRISM | 484 360 | NA | 1.21 |
| Project Viva | 462 155 | NA | 1.35 |
| ENVIR*ON*AGE | 468 782 | 1.17 | 1.12 |
| RHEA | 468 782 | 0.93 | 0.97 |
| Piccolipiu | 468 782 | 1.10 | 0.94 |

The genomic inflation factor lambda was calculated by dividing the median of observed chi-square statistics by the median of theoretical chi-square statistics with 1 degree of freedom.
NA = not available

**Table S4. DMRs in relation to prenatal $PM_{2.5}$ exposure that overlap between DMRcate and comb-p methods.**

| DMRcate | | | | | | | Comb -p | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Chr** | **Start** | **End** | **No. of probes** | **Max βFC[a]** | **P value[b]** | **Gene[c]** | **Start** | **End** | **No. of probes** | **P value[d]** |
| 6 | 29794503 | 29796439 | 31 | -1.28E-02 | 1.03E-17 | *HLA-G* | 29795350 | 29795596 | 9 | 7.26E-12 |
| 11 | 2321937 | 2323459 | 26 | 8.96E-04 | 8.84E-10 | *TSPAN32; C11orf21* | 2322286 | 2322936 | 16 | 1.68E-06 |
| 6 | 4135575 | 4136564 | 16 | -3.99E-04 | 1.11E-08 | *ECI2* | 4135575 | 4136041 | 12 | 1.02E-06 |
| 6 | 30227145 | 30228431 | 42 | 1.96E-03 | 2.48E-06 | *HLA-L* | 30227583 | 30228255 | 28 | 3.39E-03 |
| 4 | 1342394 | 1343700 | 8 | -1.06E-03 | 2.49E-05 | *UVSSA* | 1343273 | 1343376 | 3 | 8.88E-03 |
| 7 | 1177635 | 1178254 | 11 | 4.14E-04 | 6.29E-05 | *C7orf50* | 1177794 | 1178138 | 9 | 8.24E-04 |
| 1 | 162759914 | 162760659 | 6 | -3.12E-03 | 1.24E-04 | *HSD17B7* | 162760199 | 162760585 | 4 | 3.12E-05 |
| 15 | 83680418 | 83681092 | 9 | -1.90E-03 | 1.90E-04 | *C15orf40* | 83680832 | 83681093 | 3 | 6.53E-03 |
| 10 | 26986312 | 26986590 | 4 | -4.69E-04 | 1.21E-03 | *PDSS1* | 26986312 | 26986591 | 4 | 5.58E-03 |
| 1 | 186344125 | 186344998 | 14 | -3.73E-04 | 1.42E-03 | *C1orf27; TPR* | 186344558 | 186344652 | 5 | 7.93E-03 |
| 9 | 124132919 | 124133094 | 3 | -2.63E-03 | 1.70E-03 | *STOM* | 124132919 | 124133095 | 3 | 3.03E-03 |
| 1 | 160011013 | 160011295 | 2 | -7.01E-04 | 1.98E-03 | *KCNJ10* | 160011013 | 160011296 | 2 | 4.82E-03 |
| 12 | 47219626 | 47220092 | 12 | 1.99E-03 | 3.20E-03 | *SLC38A4* | 47219626 | 47219959 | 5 | 3.09E-03 |
| 16 | 712147 | 712872 | 5 | -5.59E-04 | 3.43E-03 | *WDR90* | 704976 | 705050 | 2 | 8.13E-03 |
| 10 | 134000009 | 134000410 | 9 | -3.52E-04 | 4.74E-03 | *DPYSL4* | 134000124 | 134000411 | 7 | 5.63E-03 |

[a] Fold change in DNA methylation β-value
[b] Minimum FDR P-value for the region
[c] Annotated gene(s) in the region
[d] Sidak P-value

**Table S5. Statistically significant CpGs (FDR $p<0.05$) associated with IQR increases in prenatal $PM_{10}$ (5.6 µg/m$^3$) exposure and DNA methylation in newborns (Discovery meta-analysis), and replication analyses in children (age 7–9 years) using concurrent $PM_{10}$ exposure at the time of biosampling**.

| | | | | Discovery: newborns[a] (n=1,949) | | Replication: age 7-9 years | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | BAMSE EpiGene + MeDALL (n=829) | HELIX (n=456) | ALSPAC (n=814) |
| **Chr** | **Position[b]** | **CpG** | **Gene[c]** | β (P-value) | **Direction[d]** | β (P-value) | β (P-value) | β (P-value) |
| 5 | 180670110 | cg15082635 | *GNB2L1; SNORD96A* | 0.001 (8.29E-08) | ↓↑↓↑↑↑↑ | <0.0001 (0.89) | 0.0001 (0.74) | 0.0005 (0.59) |
| 17 | 9559558 | cg20340716 | *USP43* | -0.002 (1.50E-07) | ↓↑↓↓↓↓↓ | 0.0006 (0.33) | 0.0009 (0.71) | -0.0075 (0.13) |
| 4 | 89744363 | cg00905156 | *FAM13A* | 0.001 (3.55E-07) | ↓↑X↑↑↑↑ | 0.0017 (0.08) | <0.0001 (0.95) | 0.0009 (0.31) |
| 3 | 133524572 | cg24127244 | *SRPRB* | 0.001 (7.33E-07) | ↓↑↑↑↑↑↑ | 0.0006 (0.47) | -0.0004 (0.45) | -0.0012 (0.14) |
| 6 | 32165893 | cg06849931 | *NOTCH4* | -0.001 (1.72E-06) | ↓↓↓↑↑↓↓ | 0.0011 (0.25) | -0.0008 (0.56) | -0.0004 (0.92) |
| 5 | 131563610 | cg18640183 | *P4HA2* | 0.001 (1.80E-06) | ↑↑↑↑↑↑↑ | -0.0006 (0.29) | -0.0005 (0.51) | 0.0008 (0.39) |

Abbreviations: CHR:chromosome; β, coefficient for methylation with an IQR increase in prenatal $PM_{10}$ exposure;

[a] Discovery meta-analysis does not include the PRISM or Project ViVa cohorts due to missing prenatal $PM_{10}$ data

[b] Chromosomal position based on NCBI human reference genome assembly Build 37

[c] UCSC annotated gene

[d] Direction of methylation for each cohort included in the analysis (INMA, Generation R, CHS, ENVIRONAGE, RHEA, Piccolipiu, EARLI): , ↑ = increased methylation , ↓ = decreased methylation, X = not available

**Table S6. Correlation analysis of DNA methylation and gene transcription levels in *cis* using the BIOS data (n=3,075).**

| CpG site | Gene name | N | Z-score | P value | Direction[a] | FDR P value | hgnc_symbol |
|---|---|---|---|---|---|---|---|
| cg22038738 | ENSG00000070718 | 3075 | 8.979 | 2.73E-19 | ++++ | 2.19E-18 | *AP3M2* |
| cg20340716 | ENSG00000154914 | 3075 | 7.754 | 8.88E-15 | ++++ | 2.67E-14 | *USP43* |
| cg23270359 | ENSG00000037757 | 3075 | -7.303 | 2.82E-13 | ---- | 3.95E-12 | *MRI1* |
| cg00905156 | ENSG00000138640 | 3075 | -6.302 | 2.94E-10 | ---- | 1.18E-09 | *FAM13A* |
| cg01011943 | ENSG00000231412 | 3075 | 4.651 | 3.31E-06 | ++-+ | 9.93E-06 | *NA* |
| cg22038738 | ENSG00000104368 | 3075 | 3.862 | 0.000113 | ++++ | 4.50E-04 | *PLAT* |
| cg24709511 | ENSG00000157916 | 3075 | -3.943 | 8.05E-05 | ---- | 1.37E-03 | *RER1* |
| cg15082635 | ENSG00000248275 | 3075 | -3.537 | 0.000405 | --+- | 4.05E-03 | *TRIM52-AS1* |
| cg06846669 | ENSG00000255833 | 3075 | -3.276 | 0.001051 | ---- | 4.20E-03 | *TIFAB* |
| cg24709511 | ENSG00000224387 | 2182 | -3.151 | 0.001625 | ---X | 1.38E-02 | *NA* |
| cg24709511 | ENSG00000157933 | 3075 | 2.925 | 0.003447 | -+++ | 1.95E-02 | *SKI* |
| cg16253537 | ENSG00000164398 | 3075 | 2.674 | 0.007487 | ++++ | 2.25E-02 | *ACSL6* |

[a] Direction of expression for each cohort included in the analysis (Lifelines DEEP, The Rotterdam Study, Leiden Longevity Study, The Netherlands Twin study): , ↑ = increased expression , ↓ = decreased expression, X = not available

**Table S7. Distribution of expression levels of genes associated with CpG methylation in response to maternal $PM_{10}$ or $PM_{2.5}$ exposure in newborn children of the EARLI cohort.**

| Chr | Gene | ProbeID | Mean | Median | IQR | Min | Max |
|---|---|---|---|---|---|---|---|
| 4 | *FAM13A* | 16977925 | 6.87 | 6.86 | 0.39 | 5.71 | 7.54 |
| 6 | *NOTCH4* | 17017814 | 4.92 | 4.92 | 0.28 | 4.21 | 5.97 |
| 6 | *NOTCH4* | 17027038 | 4.84 | 4.85 | 0.24 | 4.29 | 5.21 |
| 6 | *NOTCH4* | 17029639 | 4.92 | 4.89 | 0.19 | 4.49 | 5.75 |
| 6 | *NOTCH4* | 17034630 | 4.89 | 4.88 | 0.18 | 4.49 | 5.73 |
| 6 | *NOTCH4* | 17037128 | 4.93 | 4.92 | 0.19 | 4.24 | 5.78 |
| 6 | *NOTCH4* | 17039839 | 4.90 | 4.91 | 0.24 | 4.22 | 5.75 |
| 6 | *NOTCH4* | 17042335 | 4.86 | 4.85 | 0.20 | 4.19 | 5.83 |
| 5 | *SNORD96A* | 17119456 | 5.46 | 5.50 | 0.36 | 3.98 | 5.92 |
| 5 | *P4HA2* | 16999712 | 4.47 | 4.48 | 0.21 | 3.96 | 5.19 |
| 17 | *USP43* | 16831046 | 4.08 | 4.08 | 0.31 | 3.43 | 4.77 |
| 3 | *SRPRB* | 16945907 | 5.81 | 5.81 | 0.27 | 4.90 | 6.27 |
| 7 | *C7orf50* | 17054312 | 4.61 | 4.61 | 0.18 | 4.09 | 5.07 |
| 19 | *ZNF606* | 16876074 | 6.89 | 6.91 | 0.20 | 6.34 | 7.36 |
| 19 | *PSG5* | 16872926 | 3.38 | 3.31 | 0.44 | 2.53 | 5.53 |
| 1 | *ZNF695* | 16701484 | 3.41 | 3.40 | 0.25 | 3.00 | 3.98 |
| 10 | *MKX* | 16712773 | 3.25 | 3.24 | 0.19 | 2.83 | 3.66 |
| 2 | *CAPN10* | 16893222 | 5.78 | 5.79 | 0.23 | 5.19 | 6.24 |
| 8 | *COL22A1* | 17081580 | 4.51 | 4.51 | 0.20 | 4.05 | 4.88 |
| 12 | *ZNF705A* | 16747907 | 2.68 | 2.65 | 0.35 | 2.08 | 3.65 |
| 5 | *FNIP1* | 16999631 | 8.96 | 8.99 | 0.34 | 7.50 | 9.56 |
| 7 | *PLXNA4* | 17063005 | 5.08 | 5.09 | 0.35 | 4.50 | 5.58 |
| 13 | *TMCO3* | 16776883 | 6.43 | 6.44 | 0.32 | 5.26 | 7.54 |
| 8 | *PLAT* | 17076726 | 4.61 | 4.58 | 0.23 | 4.06 | 5.50 |
| 1 | *VANGL2* | 16672635 | 4.51 | 4.50 | 0.29 | 4.01 | 5.56 |
| 19 | *MRI1* | 16858849 | 6.95 | 6.96 | 0.31 | 4.93 | 7.59 |

Chr=chromosome; IQR=Interquartile range; Min=lowest value; Max=largest value

**Table S8. Distribution of gene expression levels of genes associated with CpG methylation in response to maternal $PM_{10}$ or $PM_{2.5}$ exposure in 16-year-old children of the BAMSE cohort.**

| Chr | Gene | ProbeID | Mean | Median | IQR | Min | Max |
|---|---|---|---|---|---|---|---|
| 1 | *VANGL2* | TC01001369.hg.1 | 4.80 | 4.80 | 0.17 | 4.44 | 5.33 |
| 1 | *ZNF695* | TC01006392.hg.1 | 3.00 | 2.99 | 0.13 | 2.76 | 3.22 |
| 2 | *CAPN10* | TC02005015.hg.1 | 5.38 | 5.37 | 0.12 | 5.12 | 5.80 |
| 3 | *SRPRB* | TC03000725.hg.1 | 4.61 | 4.61 | 0.15 | 4.25 | 4.93 |
| 4 | *FAM13A* | TC04001380.hg.1 | 3.98 | 3.97 | 0.16 | 3.66 | 4.40 |
| 6 | *NOTCH4* | TC06001564.hg.1 | 4.95 | 4.95 | 0.16 | 4.62 | 5.59 |
| 7 | *C7orf50* | TC07001077.hg.1 | 4.97 | 4.97 | 0.15 | 4.58 | 5.31 |
| 7 | *PLXNA4* | TC07001877.hg.1 | 4.42 | 4.42 | 0.16 | 4.11 | 4.98 |
| 8 | *PLAT* | TC08001175.hg.1 | 4.48 | 4.48 | 0.18 | 4.18 | 5.13 |
| 8 | *COL22A1* | TC08001675.hg.1 | 4.82 | 4.81 | 0.20 | 4.39 | 5.63 |
| 10 | *MKX* | TC10001133.hg.1 | 2.74 | 2.74 | 0.19 | 2.39 | 3.28 |
| 13 | *TMCO3* | TC13000425.hg.1 | 5.35 | 5.34 | 0.14 | 5.06 | 5.76 |
| 17 | *USP43* | TC17000146.hg.1 | 4.21 | 4.21 | 0.15 | 3.91 | 4.81 |
| 19 | *MRI1* | TC19000239.hg.1 | 5.78 | 5.77 | 0.17 | 5.30 | 6.13 |
| 19 | *PSG5* | TC19001582.hg.1 | 2.34 | 2.33 | 0.17 | 2.02 | 2.79 |
| 19 | *ZNF606* | TC19001910.hg.1 | 5.04 | 5.04 | 0.10 | 4.73 | 5.23 |
| 6 | *NOTCH4* | TC6_apd_hap1000098.hg.1 | 4.88 | 4.86 | 0.17 | 4.45 | 5.56 |
| 6 | *NOTCH4* | TC6_cox_hap2000190.hg.1 | 4.96 | 4.95 | 0.17 | 4.63 | 5.61 |
| 6 | *NOTCH4* | TC6_mann_hap4000155.hg.1 | 4.96 | 4.96 | 0.17 | 4.65 | 5.63 |
| 6 | *NOTCH4* | TC6_mcf_hap5000165.hg.1 | 4.94 | 4.93 | 0.17 | 4.63 | 5.60 |
| 6 | *NOTCH4* | TC6_qbl_hap6000179.hg.1 | 4.93 | 4.93 | 0.16 | 4.59 | 5.60 |
| 6 | *NOTCH4* | TC6_ssto_hap7000159.hg.1 | 4.93 | 4.92 | 0.17 | 4.58 | 5.57 |

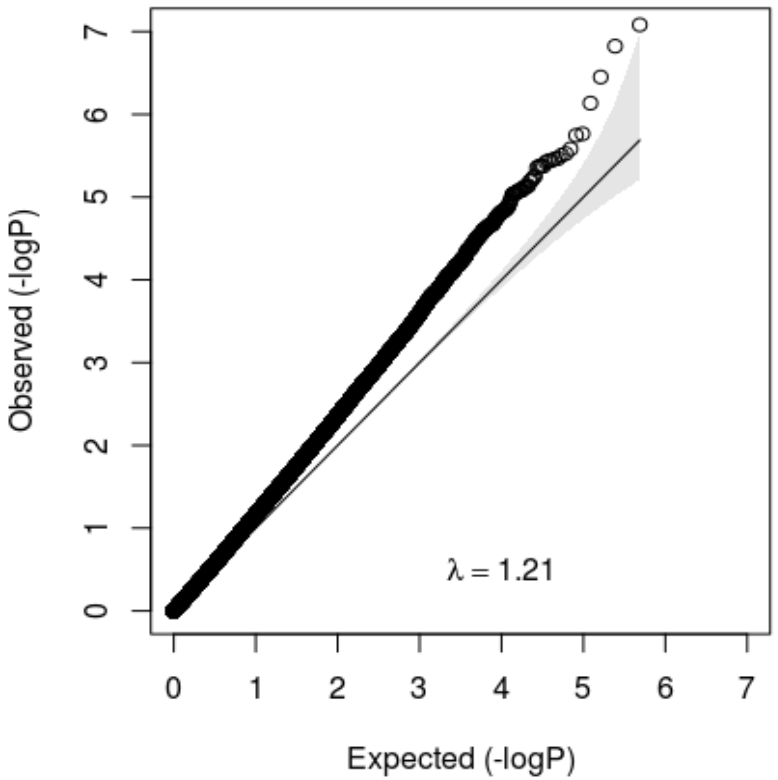
Chr=chromosome; IQR=Interquartile range; Min=lowest value; Max=largest value.

**Table S9. Top 10 significant gene ontology terms derived from pathway analysis based on CpGs significantly associated with prenatal $PM_{10}$ exposure in the discovery meta-analysis at a cut-off of $P<10^{-5}$ using missMethyl method.**

| Term | N | DE | P value | FDR P value |
|---|---|---|---|---|
| regulation of GTPase activity | 450 | 8 | 4.67E-06 | 0.11 |
| positive regulation of GTPase activity | 380 | 7 | 1.30E-05 | 0.15 |
| enzyme regulator activity | 981 | 8 | 3.50E-04 | 1.00 |
| regulation of hydrolase activity | 1215 | 9 | 3.52E-04 | 1.00 |
| enzyme activator activity | 496 | 6 | 3.61E-04 | 1.00 |
| GTPase regulator activity | 286 | 5 | 3.82E-04 | 1.00 |
| positive regulation of hydrolase activity | 729 | 7 | 5.11E-04 | 1.00 |
| negative regulation of Wnt signaling pathway | 158 | 4 | 5.44E-04 | 1.00 |
| nucleoside-triphosphatase regulator activity | 330 | 5 | 6.04E-04 | 1.00 |
| regulation of GTP cyclohydrolase I activity | 1 | 1 | 6.58E-04 | 1.00 |

N: Number of genes in the GO term; DE: number of genes that are differentially methylated

**A) prenatal $PM_{10}$**




**B) prenatal $PM_{2.5}$**

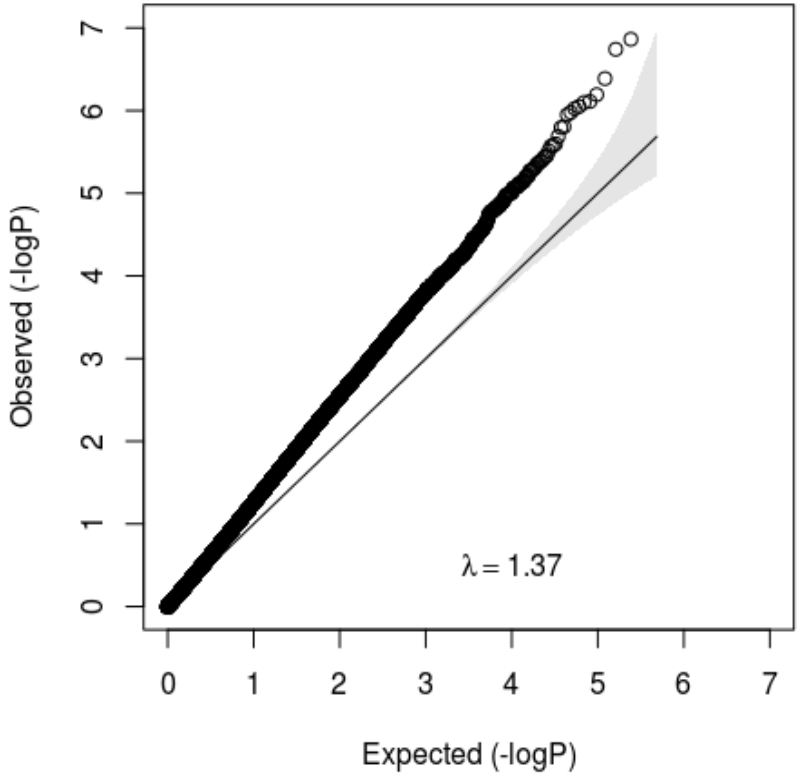
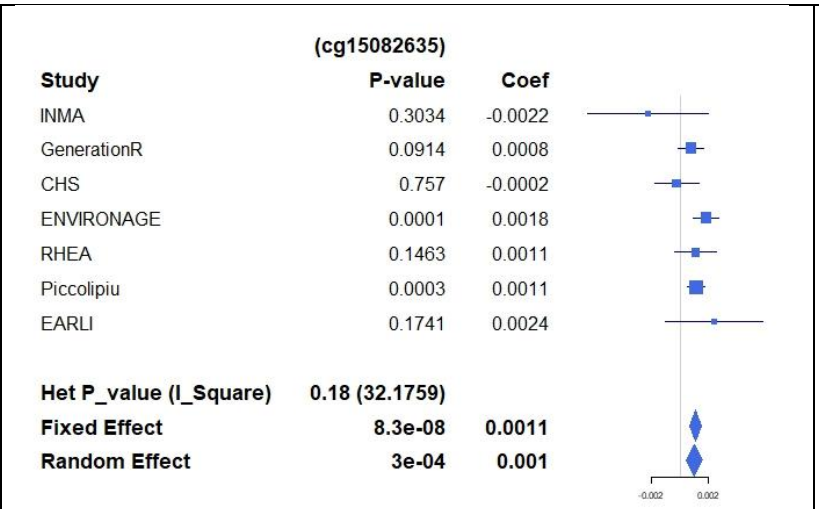
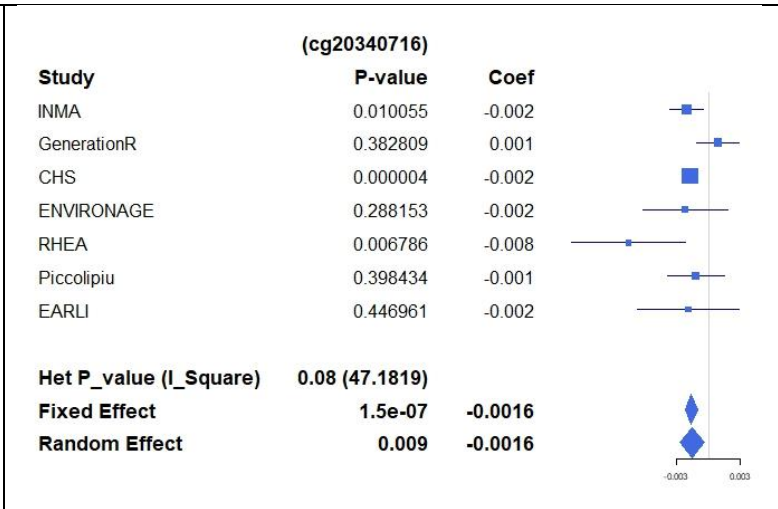
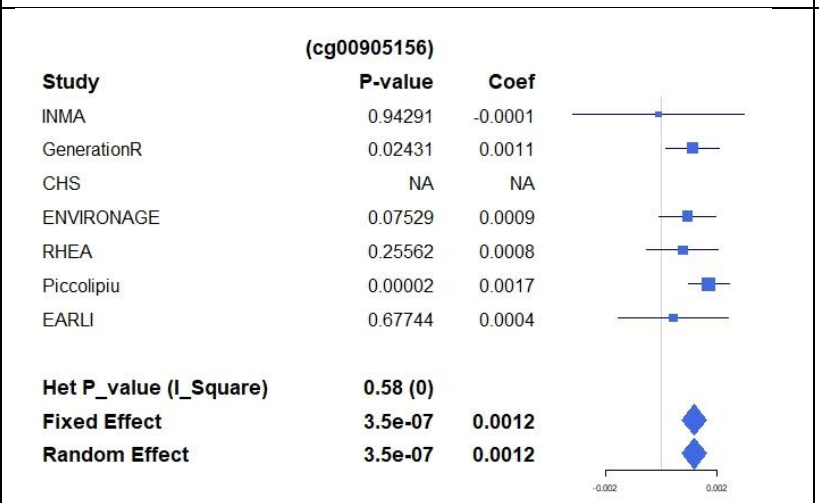


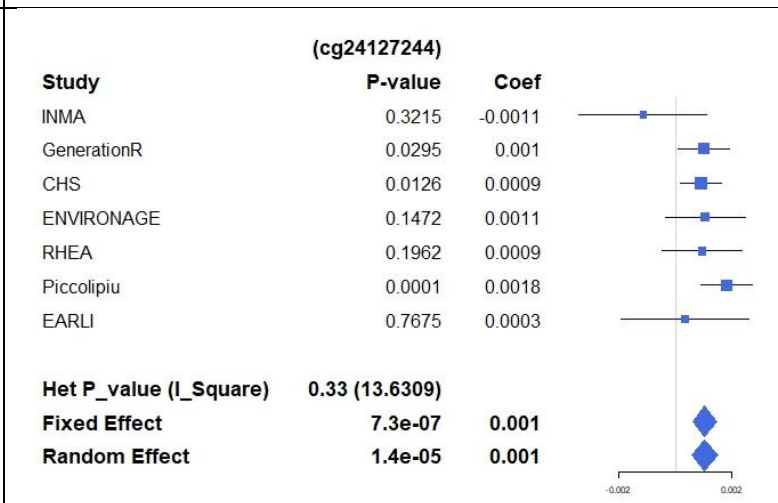

**Figure S1. Quantile-quantile plot for epigenome-wide meta-analysis of the association between A) prenatal $PM_{10}$ (n=1,949) and B) prenatal $PM_{2.5}$ exposure (n=1,551) and cord blood DNA methylation.**

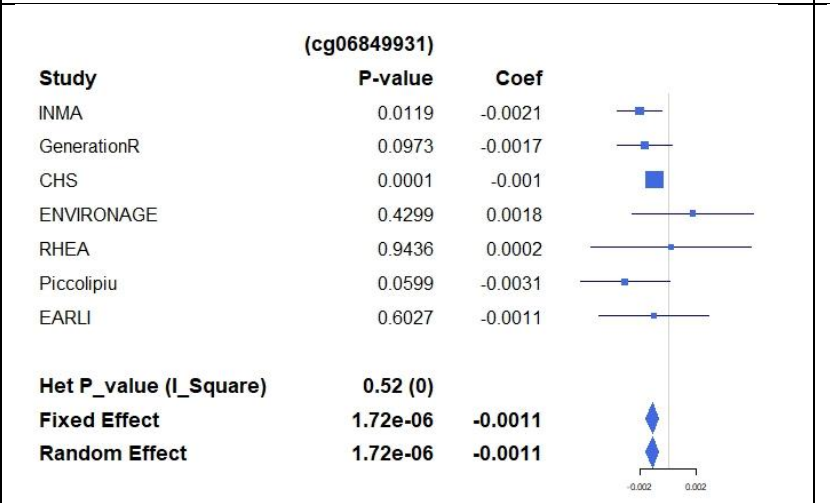
Black circles represent the observed -log10 transformed p-values, grey bands represent a 95% pointwise confidence region; the straight black line indicates the distribution of p-values expected under the null hypothesis; inflation factor lambda (λ) estimates the deviation of a distribution of observed p-values from expected distribution under the null hypothesis.



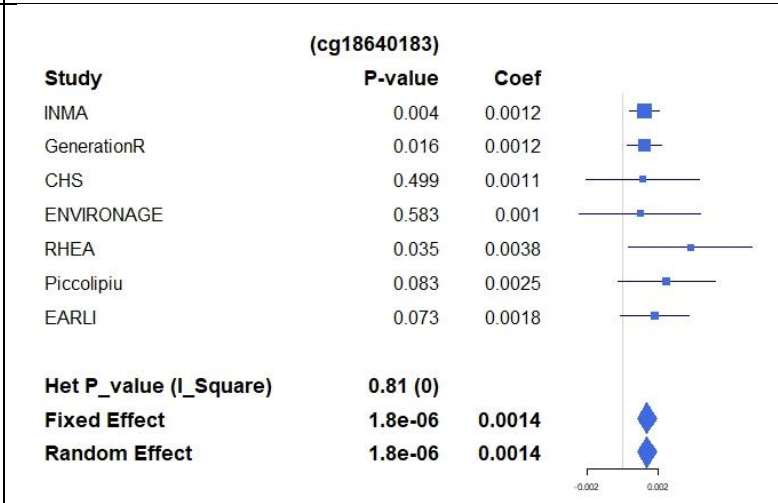




**Figure S2. Forest plots for six epigenome-wide significant associations between prenatal $PM_{10}$ exposure and cord blood DNA methylation.**

## References


Aryee, M. J., A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, et al. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30: 1363-69, PMID: 24478339. https://doi.org/10.1093/bioinformatics/btu049.

Bakulski, K. M., J. I. Feinberg, S. V. Andrews, J. Yang, S. Brown, S. L. McKenney, et al. 2016. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics 11: 354-62, PMID: WOS:000377274700004. https://doi.org/10.1080/15592294.2016.1161875.

Beelen, R., G. Hoek, D. Vienneau, M. Eeftens, K. Dimakopoulou, X. Pedeli, et al. 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. Atmospheric Environment 72: 10-23, PMID: ISI:000318262000002. https://doi.org/DOI 10.1016/j.atmosenv.2013.02.037.

Bousquet, J., J. Anto, C. Auffray, M. Akdis, A. Cambon-Thomsen, T. Keil, et al. 2011. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. Allergy 66: 596-604, PMID: 21261657. https://doi.org/10.1111/j.1398-9995.2010.02534.x.

Boyd, A., J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, et al. 2013. Cohort Profile: The 'Children of the 90s'-the index offspring of the Avon Longitudinal Study of Parents and Children. International Journal of Epidemiology 42: 111-27, PMID: 22507743. https://doi.org/10.1093/ije/dys064.

Eeftens, M., R. Beelen, K. de Hoogh, T. Bellander, G. Cesaroni, M. Cirach, et al. 2012. Development of Land Use Regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environ Sci Technol 46: 11195-205, PMID: 22963366. https://doi.org/10.1021/es301948k.

Fortin, J. P., E. Fertig, and K. Hansen. 2014. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Res 3: 175, PMID: 25285208. https://doi.org/10.12688/f1000research.4680.2.

Fortin, J. P., A. Labbe, M. Lemire, B. W. Zanke, T. J. Hudson, E. J. Fertig, et al. 2014. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biology 15, PMID: 25599564. https://doi.org/10.1186/s13059-014-0503-2.

Fraser, A., C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. D. Smith, et al. 2013. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology 42: 97-110, PMID: 22507742. https://doi.org/10.1093/ije/dys066.

Gauderman, W. J., H. Vora, R. McConnell, K. Berhane, F. Gilliland, D. Thomas, et al. 2007. Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. Lancet 369: 571-7, PMID: 17307103. https://doi.org/10.1016/S0140-6736(07)60037-3.

Gruzieva, O., C. J. Xu, C. V. Breton, I. Annesi-Maesano, J. M. Anto, C. Auffray, et al. 2017. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. Environ Health Perspect 125: 104-10, PMID: 27448387. https://doi.org/10.1289/EHP36.

Houseman, E. A., W. P. Accomando, D. C. Koestler, B. C. Christensen, C. J. Marsit, H. H. Nelson, et al. 2012. DNA methylation arrays as surrogate measures of cell mixture

distribution. BMC Bioinformatics 13: 86, PMID: 22568884. https://doi.org/10.1186/1471-2105-13-86.

Jaffe, A. E., and R. A. Irizarry. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 15: R31, PMID: 24495553. https://doi.org/10.1186/gb-2014-15-2-r31.

Johnson, W. E., C. Li, and A. Rabinovic. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118-27, PMID: WOS:000242715400008. https://doi.org/10.1093/biostatistics/kxj037.

Kloog, I., A. A. Chudnovsky, A. C. Just, F. Nordio, P. Koutrakis, B. A. Coull, et al. 2014. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. Atmospheric Environment 95: 581-90, PMID: 28966552. https://doi.org/10.1016/j.atmosenv.2014.07.014.

Kloog, I., S. J. Melly, W. L. Ridgway, B. A. Coull, and J. Schwartz. 2012. Using new satellite based exposure methods to study the association between pregnancy pm(2.5) exposure, premature birth and birth weight in Massachusetts. Environmental Health 11, PMID: 22709681. https://doi.org/10.1186/1476-069x-11-40.

Kooijman, M. N., C. J. Kruithof, C. M. van Duijn, L. Duijts, O. H. Franco, IJzendoorn M. H. van, et al. 2016. The Generation R Study: design and cohort update 2017. Eur J Epidemiol 31: 1243-64, PMID: 28070760. https://doi.org/10.1007/s10654-016-0224-9.

Kruithof, C. J., M. N. Kooijman, C. M. van Duijn, O. H. Franco, J. C. de Jongste, C. C. Klaver, et al. 2014. The Generation R Study: Biobank update 2015. Eur J Epidemiol 29: 911-27, PMID: 25527369. https://doi.org/10.1007/s10654-014-9980-6.

Leek, J. T., and J. D. Storey. 2007. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 3: 1724-35, PMID: 17907809. https://doi.org/10.1371/journal.pgen.0030161.

Leek, Jeffrey T., John D. Storey, L Kruglyak, ME Weinblatt, and A SN. 2007. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet 3: e161, PMID. https://doi.org/10.1371/journal.pgen.0030161.

Lehne, B., A. W. Drong, M. Loh, W. H. Zhang, W. R. Scott, S. T. Tan, et al. 2015. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biology 16, PMID: 25853392. https://doi.org/10.1186/s13059-015-0600-x.

Maiheu, B. V., B.;, P.; Viaene, K.; De Ridder, D.; Lauwaet, N.; Smeets, F.; Deutsch, et al. Identifying the Best Available Large-Scale Concentration Maps for Air Quality in Belgium. . Study Commissioned by the Flemish Environment (MIRA); Mol, Belgium, 2013., PMID.

McConnell, R., K. Berhane, L. Yao, M. Jerrett, F. Lurmann, F. Gilliland, et al. 2006. Traffic, susceptibility, and childhood asthma. Environmental Health Perspectives 114: 766-72, PMID: 16675435. https://doi.org/10.1289/ehp.8594.

Min, J. L., G. Hemani, G. Davey Smith, C. Relton, and M. Suderman. 2018. Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics, PMID: 29931280. https://doi.org/10.1093/bioinformatics/bty476.

Newschaffer, C. J., L. A. Croen, M. D. Fallin, I. Hertz-Picciotto, D. V. Nguyen, N. L. Lee, et al. 2012. Infant siblings and the investigation of autism risk factors. J Neurodev Disord 4: 7, PMID: 22958474. https://doi.org/10.1186/1866-1955-4-7.

Oken, E., A. A. Baccarelli, D. R. Gold, K. P. Kleinman, A. A. Litonjua, D. De Meo, et al. 2015. Cohort profile: project viva. Int J Epidemiol 44: 37-48, PMID: 24639442. https://doi.org/10.1093/ije/dyu008.

Pedersen, M., L. Giorgis-Allemand, C. Bernard, I. Aguilera, A. M. Andersen, F. Ballester, et al. 2013. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir Med 1: 695-704, PMID: 24429273. https://doi.org/10.1016/S2213-2600(13)70192-9.

Peters, J. M., E. Avol, W. J. Gauderman, W. S. Linn, W. Navidi, S. J. London, et al. 1999. A study of twelve Southern California communities with differing levels and types of air pollution. II. Effects on pulmonary function. Am J Respir Crit Care Med 159: 768-75, PMID: 10051249. doi: http://www.ncbi.nlm.nih.gov/pubmed/10051249

Peters, J. M., E. Avol, W. Navidi, S. J. London, W. J. Gauderman, F. Lurmann, et al. 1999. A study of twelve Southern California communities with differing levels and types of air pollution. I. Prevalence of respiratory morbidity. Am J Respir Crit Care Med 159: 760-7, PMID: 10051248. https://doi.org/10.1164/ajrccm.159.3.9804143.

Pidsley, R., Y. Wong CC, M. Volta, K. Lunnon, J. Mill, and L. C. Schalkwyk. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14: 293, PMID: 23631413. https://doi.org/10.1186/1471-2164-14-293.

Plusquin, M., M. Chadeau-Hyam, A. Ghantous, R. Alfano, M. Bustamante, L. Chatzi, et al. 2018. DNA Methylome Marks of Exposure to Particulate Matter at Three Time Points in Early Life. Environ Sci Technol, PMID: 29597345. https://doi.org/10.1021/acs.est.7b06447.

Relton, Caroline L, Tom Gaunt, Wendy McArdle, Karen Ho, Aparna Duggirala, Hashem Shihab, et al. 2015. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). Int J Epidemiol 44: 1181-90, PMID. https://doi.org/10.1093/ije/dyv072.

Teschendorff, A. E., F. Marabita, M. Lechner, T. Bartlett, J. Tegner, D. Gomez-Cabrero, et al. 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29: 189-96, PMID: 23175756. https://doi.org/10.1093/bioinformatics/bts680.

Thacher, J. D., O. Gruzieva, G. Pershagen, A. Neuman, M. van Hage, M. Wickman, et al. 2016. Parental smoking and development of allergic sensitization from birth to adolescence. Allergy 71: 239-48, PMID: 26475651. https://doi.org/10.1111/all.12792.

Touleimat, N., and J. Tost. 2012. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 4: 325-41, PMID: 22690668. https://doi.org/10.2217/epi.12.21.

Triche, T and Jr. 2014. IlluminaHumanMethylation450k.db: Illumina Human Methylation 450k annotation data. R package version 2.0.9., PMID.

Triche, T. J., Jr., D. J. Weisenberger, D. Van Den Berg, P. W. Laird, and K. D. Siegmund. 2013. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 41: e90, PMID: 23476028. https://doi.org/10.1093/nar/gkt090.

van Iterson, M., E. W. Tobi, R. C. Slieker, W. den Hollander, R. Luijk, P. E. Slagboom, et al. 2014. MethylAid: visual and interactive quality control of large Illumina 450k datasets. Bioinformatics 30: 3435-37, PMID: 25147358. https://doi.org/10.1093/bioinformatics/btu566.

Wickman, M., I. Kull, G. Pershagen, and S. L. Nordvall. 2002. The BAMSE Project: presentation of a prospective longitudinal birth cohort study. Pediatric Allergy and Immunology 13: 11-13, PMID: 12688617. https://doi.org/10.1034/j.1399-3038.13.s.15.10.x.

Wijga, A. H., M. Kerkhof, U. Gehring, J. C. de Jongste, D. S. Postma, R. C. Aalberse, et al. 2014. Cohort profile: the prevention and incidence of asthma and mite allergy

(PIAMA) birth cohort. Int J Epidemiol 43: 527-35, PMID: 23315435. https://doi.org/10.1093/ije/dys231.

Vineis, P., M. Chadeau-Hyam, H. Gmuender, J. Gulliver, Z. Herceg, J. Kleinjans, et al. 2017. The exposome in practice: Design of the EXPOsOMICS project. Int J Hyg Environ Health 220: 142-51, PMID: 27576363. https://doi.org/10.1016/j.ijheh.2016.08.001.

Vrijheid, M., R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, et al. 2014. The human early-life exposome (HELIX): project rationale and design. Environ Health Perspect 122: 535-44, PMID: 24610234. https://doi.org/10.1289/ehp.1307204.