Analysis Plan: Prenatal maternal BMI and placental DNA methylation

Proposed by: Nora Fernandez-Jimenez, Mariona Bustamante, Marie-France Hivert and Jose Ramon Bilbao.

Date: 2018.05.10

Background

Maternal obesity is associated with increased risk of maternal and perinatal morbidity, such as gestational diabetes, preeclampsia and birth complications (PMID: 27743975). Additionally, it is known that offspring of mothers with a high body mass index (BMI) are at a higher risk of developing obesity-related and non-related diseases than the general population, also in later life (PMID: 23007319). However, whether an early epigenetic reprogramming in utero is involved in the abovementioned outcomes needs to be ascertained. In this context, PACE consortium recently addressed that maternal BMI is robustly associated with DNA methylation changes in newborn blood, although part of these seemed to be explained by genetic/lifestyle factors rather than intrauterine mechanisms (PMID: 29016858).

Objective

The overall aim is to investigate the association of prenatal maternal BMI with epigenome-wide placental DNA methylation. Functional enrichment analysis will be performed and results will be compared with cord blood DNA methylation alterations in order to identify tissue specific patterns.

Cohorts

We identified several cohorts in PACE that could participate in the analysis (N>1,500 total participants with placenta samples):

- AQUA (Asking Questions about Allcohol in pregnancy)
- EDEN (Etudes des Déterminants pré et postnatals précoces du dévelopment et de la santé de l'Enfant)
- Gen3G (Genetics of Glucose regulation in Gestation and Growth)
- INMA (Infancia y Medio Ambiente)
- NHBCS (New Hampshire birth cohort)
- RICHS (Rhode Island Child Health study)
- EPIC (Epigenetics of Placenta In Complications of pregnancy)

Exclusion criteria

- non-singleton births
- DNA methylation assessed only in the maternal side of the placenta

Exposure

Maternal pre-pregnancy BMI either self-reported or measured. If pre-pregnancy BMI is unavailable, BMI in early pregnancy (1st trimester) can be used. Please double check values >= +/- 5 SD from the mean in your dataset to make sure they are not data entry errors. BMI will be analysed in 2 ways: continuously and categorically (binary). For the latter analysis, we will use the following categories (WHO categories): underweight (<18.5 kg/m2), overweight (25.0-29.9 kg/m2), obese (>= 30 kg/m2) and overweight or obese (>=25.0 kg/m2), compared to the normal range (18.5-24.9 kg/m2). Therefore, the five exposures of interest are:

1. Maternal prepregnancy BMI (kg/m2) - continuous

2. Maternal prepregnancy obesity – binary (versus BMI in the normal range only)

3. Maternal prepregnancy overweight and obese combined into one category – binary (versus BMI in the normal range only)

If you have <=10 participants in either group for any of the categorical exposures, please do not run that particular analysis.

If you have multiple ethnicities in your population, please 1st run all analyses adjusting for ethnicities (as instructed below), and 2nd run analyses using only European descent individuals.

Methylation data

Placental DNA methylation from fetal site assessed with the Illumina Infinium450k BeadChip methylation (EWAS) data. Studies with EPIC 850k chip data are also invited to participate, but only sites present on the 450k chip will be utilized. We are providing a code for the quality control and normalization of the methylation data. QC will include dye-bias adjustment, background correction, functional normalization (FunNorm), and beta-mixture quantile (BMIQ) normalization. We also provide code for investigating whether technical batch effects exist and how to adjust with combat if needed, but each cohort can make the decision individually.

Covariates

- Maternal age: Continuous

- Parity: Preferred categorization is into 2 groups: 0 and >=1.

- Maternal education: Preferred categorization is into 3 groups: primary, secondary and university. It is used as an indicator of socioeconomic status.

- Maternal smoking status: Preferred classification is into three groups: 1. No smoking in pregnancy, 2. Smoking, but stopped in early pregnancy, 3. Smoking throughout pregnancy. If you want to use a different categorization, please contact the meta-analysis center to discuss.

- Ancestry: if you have GWAS data in your cohort, please correct for ethnicity by using GWAS-PCAs; if GWAS is not available, use self-reported ethnicity as a categorical

covariate (categories with few participants can be combined or either excluded). Please run the models again with children with White European ancestry alone, to be able to compare the results.

- Selection factors (optional): If your study design includes an enrichment of cases of some condition, please include the case-control variable.

If you want to use different categorization of the above covariates or other covariates, please contact the meta-analysis center to discuss.

Models and analyses

Models:

Run the models for which you have data (all the models are not needed for participation).

- **Main model**

  A1. Methylation= maternal BMI (continuous) + maternal age + parity + maternal education + maternal smoking (+ selection) (+ ancestry)

  B1. Methylation= maternal obesity (binary) + maternal age + parity + maternal education + maternal smoking (+ selection) (+ ancestry)

  C1. Methylation= maternal overweight & obesity (binary) + maternal age + parity + maternal education + maternal smoking (+ selection) (+ ancestry)

- **Additionally adjusted model for cellular heterogeneity**

  A2. Methylation= maternal BMI (continuous) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection) (+ ancestry)

  B2. Methylation= maternal obesity (binary) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection) (+ ancestry)

  C2. Methylation maternal overweight & obesity (binary) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection) (+ ancestry)

- **Children with White European ancestry alone (if applicable)**

  A1'. Methylation= maternal BMI (continuous) + maternal age + parity + maternal education + maternal smoking (+ selection)

  B1'. Methylation= maternal obesity (binary) + maternal age + parity + maternal education + maternal smoking (+ selection)

C1'. Methylation= maternal overweight & obesity (binary) + maternal age + parity + maternal education + maternal smoking (+ selection)

A2'. Methylation= maternal BMI (continuous) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection)

B2'. Methylation= maternal obesity (binary) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection)

C2'. Methylation maternal overweight & obesity (binary) + maternal age + parity + maternal education + maternal smoking + Houseman free reference estimations (+ selection)

Statistical analyses:

The code for the full analysis is provided in R. It is divided in four steps:

- **1_Load_IDATs_QC_FunNorm.R:**
  - It reads IDAT files
  - It performs quality control based on minfi (Aryee et al.): sample quality, sex validation, and Principal Component Analysis.
  - It performs dye bias adjustment, background correction, and functional normalization method (minfi).
  - Generates plots to evaluate normalization.
- **2_Perform_BMIQ_Adjustments.R**:
  - It performs beta-mixture quantile normalization (BMIQ) for correcting probe design bias (Teschendorf et al.).
- **3_BatchEffects_Corrections.R** (optional):
  - It performs technical batch effect correction with Combat (Johnson et al). Apply it if needed in your data and under your criteria.
- **4_RefFree_Adjusted_Models.R**:
  - It filters probes based on Chen et al.
    - Probes in sexual chromosomes.
    - Cross-hybridizing probes.
    - Probes with SNPs at CpG site, extension site, and within 10 bp from the extension site with an average MAF>0.01.
  - It estimates cellular heterogeneity with the RefFree method (Houseman et al).
    - After estimating Omega, examine OutlierScreening plots to ensure no extreme outliers are being generated. If so, reduce Kchoose value by 1, re-run Step 2, Part A and re-examine outlier plots. Reduce Kchoose until extreme outliers are gone.
  - It tests the association between estimated cellular components and main variables.
  - It gets rid of extreme outliers (Gaphunter).

- It performs association analysis between prenatal maternal BMI and each CpG site individually with robust linear regression modelling (rlm() function in R).
- It formats results and performs QQ plot and lambda estimation.

Data upload

To upload data, please contact nora.fernandez@ehu.eus.

Please include the following items in your uploaded file:

- A short description of the placenta collection, DNA extraction, and 450k processing in the laboratory.
- Information on the distribution of exposures and covariates.
- Output of the fourth step of the code (**4_RefFree_Adjusted_Models.R)**:
  - Cellular heterogeneity:
    - "PACE_ STUDYNAME _Omega_OutlierScreening.pdf"
    - "PACE_ STUDYNAME _Omega_heatmap.pdf"
    - "PACE_ STUDYNAME _Variability_By_CellMix_CpGs.txt"
  - Association between cellular heterogeneity and main variables:
    - "PACE_ STUDYNAME _CatCovariate_CellMix_Associations.csv"
    - "PACE_ STUDYNAME _ContCovariate_CellMix_Associations.csv"
  - Results for models A1, B1, C1, A2, B2, C2 & A1', B1', C1', A2', B2', C2':
    - "PACE_STUDYNAME_Model_date.txt"
    - "PACE_STUDYNAME_Model_QQ_Plot.pdf"
    - "PACE_STUDYNAME_Model_Lambdas.txt."
- Any additional information about the uploaded files.

Timeline

- Comments to the protocol. Complete **Table T1**. DONE.
- EWAS results + descriptive of the cohort (**Table T2**): 27th July 2018.
- Secondary analysis: to be determined after EWAS results.

Secondary analysis

Secondary analysis might be required for the **top CpG signals** in all or in particular cohorts. We will contact the cohorts after the first results.

- **Sensitivity analysis 450k vs EPIC:** We will check consistency of results between 450k and EPIC. We will look at CpGs surrounding top signals and present only in the EPIC array.
- **Negative control analysis:** The negative control design is a method to deal with residual confounding by comparing the association of interest with that of another related association but for which there is no biologically plausible mechanism for causation. In this project, **paternal BMI** could be used as a negative control.
- **Correlation between placental methylation and gene expression.**

- Association of the placental DNA methylation changes with **reproductive outcomes** (including birth weight, birth length and gestational age).
- **Interaction with genetic variation:** To test whether the association between maternal BMI and placental DNA methylation is influenced by genetic variation, we will perform a **cis mQTL analysis** with the SNPs surrounding (<5 Kb) the most significant differentially methylated hits for which we have genotype data. Depending on the alterations observed, trans mQTL analysis could also be performed.