**Epigenome-wide meta-analysis of the associations of vitamin B12 concentrations in pregnancy and in newborns with newborn DNA methylation – Plan for lookup in childhood**

09 November 2020

**Contact details**

For questions and/or joining this meta-analysis: please email Giulietta Monasso ([g.monasso@erasmusmc.nl](mailto:g.monasso@erasmusmc.nl)) or Janine Felix ([j.felix@erasmusmc.nl](mailto:j.felix@erasmusmc.nl))

**BACKGROUND**

We previously conducted a meta-analysis of epigenome-wide association studies investigating associations of vitamin B12 concentrations during fetal development and cord blood DNA methylation. This meta-analysis resulted in 109 FDR-significant CpGs in relation to maternal vitamin B12 concentrations during pregnancy and 7 FDR-significant CpGs in relation to newborn vitamin B12 concentrations. (**Table 1**). To follow-up these findings, we would like to ask you to perform a look-up of the findings in children and/or adolescents.

**EXPOSURES**

***Maternal total vitamin B12 concentrations (mat.b12)***

Maternal total vitamin B12 concentration during pregnancy (mat.b12) in serum or plasma.

***Newborn total vitamin B12 concentrations (newborn.b12)***

Cord blood total vitamin B12 concentrations (newborn.b12), measured in serum or plasma.

Please analyze concentrations of these exposures continuously per Standard Deviation (SD) increase (Z-scores). Z-scores for each individual can be calculated as follows:

Z-score = (B12 concentration in the individual – mean B12 concentration in study population) / SD of B12 concentration in study population

If you need any help with calculating Z-scores, please email [g.monasso@erasmusmc.nl](mailto:g.monasso@erasmusmc.nl) and we can provide you with the R code.

**Exclusions**

* Exclude mothers and newborns with total vitamin B12 concentrations outside +/- 5 standard deviations from the mean in your dataset.
* Exclude all twins.
* For mothers with multiple children (non-twin siblings), include only one child per mother, based on completeness of data and, if equal, randomly. Please describe in the readme file how many siblings were excluded.

**Look-up in childhood/adolescence (whole blood)**

For the childhood/adolescence analyses, please prepare a phenotype file with the exposure (maternal or cord blood vitamin B12 in Z-scores) and the following covariates (In *red Italic* font are those covariates that are new or changed, as compared to the phenotype file of the main analysis):

* *Child age at blood sampling (child.age, continuous in years)*
* *Estimated cell types: Use the “Houseman” reference set for cell type estimation in the ‘’FlowSorted.Blood.Combined.450K’’ or “FlowSorted.Blood.Combined.EPIC’’ Bioconductor package for cell type correction* (1)*. Include the following cell types: CD8T, CD4T, NK, Bcell, Mono, Gran. If you need a script for estimating these, please contact us.*
* Maternal age (mat.age): Age at conception. Continuous, in years.
* Maternal socio-economic status (mat.ses): Use the definition most suited in your cohort. Preferably, a three category ordinal variable (1 = lower, 2 = middle, 3 = higher). If you need to categorize mat.ses differently, please inform us.
* Maternal pre-pregnancy BMI (mat.bmi): Continuous in kg/m2.
* Maternal smoking during pregnancy (mat.smoking): binary numeric variable, 0 = no smoking or 1st trimester only (quit before 2nd trimester); 1 = sustained smoking.
* Parity (parity): binary numeric variable, 0 = nulliparous; 1 = multiparous.
* Child sex (sex): binary numeric variable, 0 = boy; 1 = girl.
* Batch (batch): Please adjust for batch effects by including the most important covariate(s), such as plate, in the models. Alternatively, you can use a batch correction method such as ComBat. Please indicate in the README file how you adjusted for batch.
* Gestational age at vitamin B12 measurement (gest.age.sampling). Continuous in weeks [**only** in analyses of maternal B12, this covariate will **not** be included in analyses of newborn B12].
* Selection factors: optional covariate, include if relevant for your study and describe in the readme file. For example, case-control status if your study was designed as a case-control study.
* Ethnicity: If your study includes more than one major ethnic group (e.g. African, Asian, European or Hispanic) please analyze these groups separately.

Please name all covariates exactly as asked (name defined between brackets for each variable), in order to prevent problems with running the R code.

For the childhood look-up, please run one or both of the following robust linear regression models, using the syntax provided, depending on which exposure is available in your cohort:

**Maternal main model (childhood analysis):** DNA methylation at each of the 109 FDR-significant CpGs of maternal vitamin B12 meta-analysis ~ mat.b12 + gest.age.sampling + mat.age + mat.ses + mat.bmi + mat.smoking + sex + batch +Houseman cell types + child age at blood sampling (+selection factors);

**Newborn main model (childhood analysis):** DNA methylation at each of the 7 FDR-significant CpGs of newborn vitamin B12 meta-analysis ~ newborn.b12 + mat.age + mat.ses + mat.bmi + mat.smoking + sex + batch +Houseman cell types + child age at blood sampling (+selection factors);

**R CODE**

We have prepared a script for the lookup analysis. As before, you can delete the parts of the script that contains code for an analysis you are not able to run (maternal vs newborn analysis). For the script, you need as input:

1. A phenotype file (**Table 2**): a tab delimited file (.txt) or data frame. Please note the following:
   1. Missing values for any of the covariates should be denoted by NA, do not leave any cells blank. The provided R code will exclude these cases for you.
   2. The script will automatically make factors of the following covariates: batch, mat.ses, mat.smoking, sex.
   3. If you adjust for batch effect by including more than one covariate, or if you use a batch correction method, you have to adapt the script and phenotype file accordingly. If you need any help with this, please email [g.monasso@erasmusmc.nl](mailto:g.monasso@erasmusmc.nl).
   4. If you have added any selection factor, please add this variable as an extra column to the phenotype file. You also have to change the R script accordingly. If you need any help with this, please email g.monasso@erasmusmc.nl.
2. A matrix of beta values (each column representing a newborn and each column a probe on the array). From this matrix, the script will select the CPGs of interest (the hits from the main model).

**OUTPUT DATA FILE FORMAT**

In the first part of the R script, please specify your cohort name and the date of analysis:

*## Please only change cohort and analysis.date below ##*

*cohort <- "GENR" #change to name of your cohort/study*

*analysis.date <- "20200914" # change to the date (yyyymmdd) at which you perform the analyses*

**README FILE**

When you send the results, please include a readme file with information on:

* At which age(s) your cohort has DNA methylation data available in childhood and/or adolescents.

**DATA EXCHANGE**

Please let us know when you are ready to upload your results ([g.monasso@erasmusmc.nl](mailto:g.monasso@erasmusmc.nl)). We will send you a link to upload the files.

**TIMELINE (tentative)**

* **December 7**: Deadline for uploading results. Any questions can be sent via email to [g.monasso@erasmusmc.nl](mailto:g.monasso@erasmusmc.nl) or [j.felix@erasmusmc.nl](mailto:j.felix@erasmusmc.nl).

**REFERENCES**

1. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics. 2014;30(10):1431-9.

Table 1: 109 and 7 CpGs that were significantly associated with maternal vitamin B12 concentrations during pregnancy and newborn cord blood vitamin B12 concentrations, respectively (FDR <≤0.05 and I2 ≤50%).

Maternal B12

cg26393629

cg19529709

cg04206517

cg08935125

cg27539527

cg15730180

cg02105458

cg17900015

cg03043822

cg04634427

cg09807524

cg19474546

cg10678190

cg05312960

cg26536593

cg03829137

cg13174253

cg25843439

cg14605520

cg27403609

cg01743593

cg22809920

cg17234513

cg20581874

cg14776195

cg10340210

cg09317502

cg25327343

cg09520393

cg24554151

cg03327325

cg27353899

cg01613965

cg16964673

cg00534274

cg05665581

cg00200803

cg00292513

cg15908975

cg08849628

cg25396728

cg21341928

cg19083407

cg12302982

cg05282518

cg09367432

cg07296387

cg19149785

cg06468072

cg11421509

cg03873392

cg14120919

cg05371552

cg07594247

cg27181142

cg17445212

cg21482265

cg12889195

cg07646362

cg02679336

cg27343456

cg06882571

cg21540359

cg02614024

cg15612221

cg22157494

cg23504719

cg01809217

cg13315047

cg23638640

cg06324373

cg08097631

cg01988325

cg22730830

cg15015426

cg01925498

cg11763394

cg17465112

cg19480274

cg19479373

cg05086444

cg05779786

cg15971980

cg21218093

cg11258452

cg03689146

cg01273232

cg06178315

cg17270081

cg04579415

cg08761659

cg06671706

cg05010260

cg16991589

cg07091678

cg17780447

cg00074818

cg08930881

cg13605615

cg26434090

cg18914258

cg01232511

cg15574301

cg01330954

cg21610815

cg07280731

cg00466136

cg05146852

cg04453501

Newborn B12

cg13863764

cg00658405

cg08243619

cg24371425

cg02615136

cg04096723

cg09573658

Table 2: phenotype file format (tab delimited (.dat))

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sample.id | mat.b12 |  | newborn.b12 | batch | bcell | mono | cd4t | cd8t | gran | nk | gest.age.sampling | mat.age | mat.ses | mat.bmi | mat.smoking | Parity | sex | child.age |
| 1234559 | 5.5555 |  | 5.5555 | 4 | 0.1340 | 0.3992 | 0.4201 | 0.1104 | 0.1100 | 0.011 | 80.4032 | 31.3013 | 2 | 20.4544 | 1 | 1 | 0 | 6.3942 |