

Cohort Profile: Epigenetics in Pregnancy (EPIREG) – population-based sample of European and South Asian pregnant women living in Norway with epigenome-wide DNA methylation (850k) in peripheral blood leukocytes

Nicolas Fragoso-Bargas^{1,2}, Julia O. Opsahl², Nadezhda Kiryushchenko^{1,3}, Yvonne Böttcher^{2,4,5}, Sindre Lee-Ødegård², Elisabeth Qvigstad^{1,2}, Kåre Rønn Richardsen⁶, Christin W. Waage⁷, Line Sletner^{2,8}, Anne Karen Jenum⁷, Rashmi B. Prasad⁹, Leif C. Groop⁹, Gunn-Helen Moen^{2,10,11,12}, Kåre I. Birkeland^{1,2}, Christine Sommer¹.

¹ Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Norway

² Institute of Clinical Medicine, University of Oslo, Norway

³ Department of Bioscience, University of Oslo, Norway

⁴ Department of Clinical Molecular Biology, Akershus University Hospital, Lørenskog, Norway

⁵ Helmholtz-Institute for Metabolic, Adiposity and Vascular Research, Germany

⁶ Department of Physiotherapy, Faculty of Health Sciences, Oslo Metropolitan University, Norway

⁷ General Practice Research Unit (AFE), Department of General Practice, Institute of Health and Society, University of Oslo, Oslo, Norway

⁸ Department of Pediatric and Adolescents Medicine, Akershus University Hospital, Lørenskog, Norway

⁹ Lund University Diabetes Centre, Malmö, Sweden.

¹⁰ The University of Queensland Diamantina Institute, The University of Queensland, Woolloongabba, QLD 4102, Australia.

¹¹ K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.

¹² Population Health Science, Bristol Medical School, University of Bristol, UK.

Corresponding author: Christine Sommer, Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Postbox 4959 Nydalen, N-0424 Oslo, Norway, phone: +47 22 89 46 28, e-mail: christine.sommer@medisin.uio.no

ABSTRACT

Purpose: Pregnancy is a valuable model to study the association between DNA methylation and several cardiometabolic traits, due to its direct potential to influence mother's and child's health. Epigenetics in Pregnancy (EPIPREG) is a population-based sample with the aim to study associations between DNA-methylation in pregnancy and cardiometabolic traits in South Asian and European pregnant women and their offspring.

Participants: In EPIPREG we have quantified epigenome-wide DNA methylation in maternal peripheral blood leukocytes in gestational week 28 ± 1 in Europeans (n=312) and South Asians (n=168) that participated in the population-based cohort STORK Groruddalen, in Norway. DNA methylation was measured with Infinium MethylationEPIC Kit (850k sites), with technical validation of four CpG sites using bisulphite pyrosequencing in a subset (n=30). The sample is well characterized with few missing data on e.g. genotype, universal screening for gestational diabetes, objectively measured physical activity, bioelectrical impedance, anthropometrics, biochemical measurements, and a biobank with maternal serum and plasma, urine, placenta tissue. In offspring, we have repeated ultrasounds during pregnancy, cord blood, and anthropometrics up to 4 years of age.

Results to date: We have quantified DNA methylation in peripheral blood leukocytes in nearly all eligible women from the STORK Groruddalen study, to minimize the risk of selection bias. Genetic principal components distinctly separated Europeans and South Asian women, which fully corresponded with the self-reported ethnicity. Technical validation of 4 CpG sites from the methylation bead chip showed high concordance with bisulfite pyrosequencing ($R=0.98$, $p<0.001$).

Future plans: We plan to study associations between DNA methylation and cardiometabolic traits and outcomes. We hope to identify cohorts with similar data to replicate our findings, collaborate on joint efforts such as meta-analysis, and serve as a replication cohort for other studies.

Strengths and limitations of this study

- Epigenome-wide DNA methylation data in maternal peripheral blood leukocytes in gestational week 28±1 in 312 Europeans and 168 South Asians living in Norway
- EPIPREG's population-based design and comprehensive phenotyping allows for studies of a wide range of phenotypic traits, exposures and outcomes in relation to DNA methylation
- The inclusion of women with both European and South Asian ethnic background enables interesting studies into the role of DNA methylation in ethnic disparities in health.
- The wide range of collected phenotypes, exposures and outcomes makes the EPIPREG sample well suited to serve as a replication cohort for other cohorts with DNA methylation data
- The EPIPREG sample has limited statistical power for epigenome-wide association studies, and we are interested in collaborative efforts such as meta-analysis of several cohorts

INTRODUCTION

Studies of epigenetic marks have in recent years gained increased interest in the context of human diseases (1). Such studies may enhance our biological understanding of the aetiology of several diseases, increase our understanding of detrimental or protective mechanisms, or for prognosis and risk prediction (2). One of the most studied epigenetic mechanisms is DNA methylation, which plays an important role in normal development, chromatin organization and gene expression (3). Several studies have indicated that DNA methylation is associated with cardiovascular risk factors such as body mass index (BMI)(4-7), gestational diabetes (GDM) (8), type 2 diabetes (T2D) (9-11), lipid levels (12, 13), hypertension (14, 15), smoking (16, 17) and alcohol intake (18, 19), suggesting that cardiometabolic diseases have an epigenetic component.

Although scarcely studied, pregnant women provide a unique opportunity to study the association between blood DNA methylation and several phenotypes related to glucose homeostasis and cardiovascular traits. This is because pregnancy has been proposed as a stress test for metabolism in several organs (20), including the pancreatic beta-cells, since insulin resistance increase naturally in all pregnancies (21). In the third trimester of pregnancy, this insulin resistance in many women reaches a level similar to that observed in type 2 diabetes, requiring the beta-cell to increase its insulin secretion considerably to compensate (22). Similarly, pregnancy-induced hypertension is associated with increased risk of future cardiovascular disease (23).

In the Epigenetics in Pregnancy (EPIPREG) sample, we have quantified epigenome-wide DNA-methylation in peripheral blood leukocytes in women of European and South Asian origin attending the well-characterized, multi-ethnic and population-based STORK Groruddalen (STORK G) study (24). The population based design and inclusion of a significant number of women with European and South Asian ethnicity allows us to study a wide range of phenotypes. The aim of EPIPREG is to discover novel associations between DNA-methylation in pregnancy and cardiometabolic related traits in South Asian and European pregnant women and their offspring which may have potential for prevention and treatment.

COHORT DESCRIPTION

Study population

EPIPREG (n=480) is a sub-study of the larger STORK G, which is a population-based cohort of 823 healthy women with different ethnic origin (European, South Asian, African, Middle Eastern and South American) attending three public Child Health Clinics for antenatal care in the multi-ethnic area of Groruddalen, Oslo, Norway, 2008-2010 (24). Briefly, women were eligible if they: 1) Lived in the study districts; 2) planned to give birth at one of two study hospitals; 3) were <20 weeks pregnant; 4) could communicate in Norwegian or any of the eight translated languages; and 5) were able to give an informed consent. Women with pre-gestational diabetes, or in need of intensive hospital follow-up during pregnancy, were excluded. The participation rate was 74%, varying from 63.9% to 82.6% across ethnic groups (24).

Ethical approval

The STORK G study including the collection of data on genetics and epigenetics is approved by the Norwegian Regional Committee for Medical Health Research Ethics South East (ref.number 2015/1035). We obtained written informed consent from all participants before any study-related procedure.

Data collection

Questionnaire data and Anthropometrics

In STORK G, interviewer-administered questionnaires were completed in gestational week 15±3 (visit 1), 28±2 (visit 2), and 12±2 (visit 3) weeks postpartum (24, 25). Ethnic origin was defined by either the individual's country of birth or their mother's country birth, if the last was born outside Europe (26). We have detailed data on parity, pre-pregnant BMI, smoking status, alcohol intake, education, marital status and diet (24, 25, 27). At all the three visits, we measured maternal height, body weight, fat mass with bioelectrical impedance (Tanita-Weight BC-418 MA), skinfold thickness at three sites (Holtain T/W Skinfold Caliper, Holtain Ltd., Crymych)(28) and systolic and diastolic blood pressure (Omron HEM-7000-E M6 Comfort) (29).

Universal screening for gestational diabetes

All women underwent a 75 g oral glucose tolerance test at gestational week 28 \pm 2. Fasting and 2-hour glucose were analysed with a point-of-care instrument (HemoCue, Angelholm, Sweden). Women were diagnosed with gestational diabetes based on the WHO 1999 criteria (fasting glucose \geq 7.0 mmol/l or 2-hour glucose \geq 7.8 mmol/l) (30). Furthermore, in retrospect and exclusively for research purposes, we also classified the samples using the WHO 2013 criteria (fasting glucose \geq 5.1-6.9 mmol/l or 2-hour glucose \geq 8.5-11 mmol/l) (31).

Laboratory data

Venous blood was drawn at the three visits into tubes with ethylenediaminetetraacetic acid (EDTA). Subsequently, the samples were aliquoted and biobanked or subject to routine laboratory analyses that were performed continuously during the study period. Fasting glucose, total Cholesterol, LDL-Cholesterol, HDL-Cholesterol and triglycerides levels were measured with a colorimetric method (Vitros 5.1 fs, Ortho clinical diagnostics) (27), HbA1c levels were assessed in full blood with HPLC (Tosoh G8) (26), fasting C-peptide and insulin were measured at the Hormone Laboratory, Oslo University Hospital, with non-competing immunofluorometric assays (DELFI, PerkinElmer Life Sciences, Wallac Oy, Turku, Finland) (32). Serum 25(OH)D was analysed by competitive RIA (DiaSorin) at visit 1 and visit 2 (33), and S-leptin was analyzed by HADCYMAG-61K based on Luminex® xMAP® technology (34), at the Hormone Laboratory, Oslo University Hospital. Serum vitamin B12 and folate were measured with Electrochemiluminescence (ECLIA) assays, Roche, at Medical Biochemistry, Oslo University Hospital. HOMA-IR and HOMA-B were estimated by Oxford University HOMA Calculator 2.2 using fasting glucose and C-peptide.

Objectively measured physical activity

Physical activity (PA) was objectively measured from visit 1 to 3 using SenseWear™ Pro3 Armband (SWA) (BodyMedia Inc, Pittsbur, PA, USA) (35). Data from women with at least one valid day (defined as \geq 19.2h) were considered valid (36). Physical activity was characterized

as Sedentary behaviour (< 1.5 metabolic equivalents (METs)) light intensity (1.5 to <3 METS) or moderate or intense? (≥ 3 METs) (35, 37).

Offspring data

The STORK G study also collected gestational age, abdominal circumference, head circumference, femur length and estimated fetal weight by ultrasound measured on three different time points during pregnancy (38). We have detailed anthropometric measurements at birth such as birthweight, head circumference, abdominal circumference, crown-heel length and neonatal skinfolds measured with a Holtain T/W Skinfold Caliper (Holtain Ltd., Crymych) (39). Measurements of weight and length/height were collected during routine follow-up of the Mother - and Child Health Clinics when the children were 6 weeks old and thereafter at the 3, 6, 12, 15, 24 and 48 months visits. Further register-based follow-up is planned.

Venous serum cord blood samples were collected at birth and stored at -80°C . Several sections from the placenta and umbilical cord have been sampled, and stored as Formalin-fixated-paraffin embedded blocks. Currently, an ongoing pilot study of 80 FFPE placentas (40 South Asian, 40 European) demonstrate that DNA with adequate quality can be extracted from the Formalin-fixated-paraffin embedded blocks to successfully run pyrosequencing (Sletner, unpublished). Furthermore, we have frozen placenta biopsies of about 1/3 of the women's offspring.

DNA extraction

In STORK G, at gestational week 28 ± 2 , DNA from peripheral blood leukocytes was extracted continuously throughout the data collection, at the Hormone Laboratory, Oslo University Hospital, using a salting out procedure (40), and stored at -80°C .

Genetic data

The samples were genotyped using the Illumina CoreExome chip, by the Department of Clinical Sciences, Clinical Research Centre, Lund University, Malmö, Sweden (41). Of the 664 genotyped samples those with low call rate (i.e. < 95%, $n=0$), extreme heterozygosity ($> |\text{mean} \pm (3 \times \text{SD})|$, $n=1$), mismatched gender ($n=24$) or cryptic relatedness (i.e. one individual (chosen at

random) from each related pair, defined as genome-wide Identity by descent (IBD) > 0.185 (n=6) were excluded from analyses. Genetic ethnic origin was defined by ancestry informative principal component analysis based on the variance-standardized relationship matrix generated in PLINK 1.9 software package (42) (<https://www.cog-genomics.org/plink/1.9/>)).

Variants with call rate <95% (10081 SNPs), out of Hardy-Weinberg equilibrium (exact $p < 10^{-6}$, 1971 SNPs) or with low minor allele frequency (MAF) <1% (245221 SNPs) were removed before imputation. Quality control was performed using the PLINK 1.9 software package (42). After quality control, 293914 variants were left for imputation.

Imputation in European and South Asian samples was performed as follows: The GWAS scaffold was mapped to NCBI build 37 of the human genome, and imputation to the 1000G reference panel (Phase3- <http://www.well.ox.ac.uk/~wrayner/tools/>) was performed using IMPUTEv2.3.2 (43).

Epigenome-wide DNA methylation

Europeans and South Asians were the largest ethnic groups in STORK G, and South Asians of special interest due to their higher risk of gestational diabetes (34). In EPIPREG, we quantified DNA methylation in maternal peripheral blood leukocytes in gestational week 28 ± 1.2 in all Europeans (n=312) and South Asians (n=168) participating in STORK G who were genotyped and had fasting glucose data recorded. DNA samples were bisulfite converted using EZ DNA Methylation™ Kit (Zymo Research, Tustin, CA, USA) before added onto Infinium MethylationEPIC Kit (Illumina, San Diego, CA, USA) at the Department of Clinical Sciences, Clinical Research Centre, Lund University, Malmö, Sweden. Raw signal intensities of each probe were extracted using Illumina's GenomeStudio Software. The methylation level at each site was represented as a beta (β) value of the fluorescent intensity ratio ranging from 0 (not methylated) to 1 (completely methylated). Meffil R package (44) (<https://cran.r-project.org/>) was used for quality control, normalization and reporting of beta values. During QC, we removed 8 samples: 1 due to sex mismatch (predicted sex outliers > 5SD), 1 outlier in control probes bisulfite 1 and bisulfite 2 (>5 SD), and 6 outliers from the methylated/unmethylated ratio comparison (>3 SD). Furthermore, 1299 probes with a detection p-value <0.01, and bead count <3 were removed. We used functional normalization, adjusting for effects of different batches, plates, columns and rows. A total of 307 European and 165 South Asian women and 864 560 probes of the array passed the QC.

Pyrosequencing

Random samples of 30 women were selected for technical validation of four CpGs sites by bisulfite pyrosequencing. The four CpG sites were chosen from preliminary top associations with fasting glucose (cg08098128, cg14120215), 2-hour glucose (cg19327414) and BMI (cg17148978). DNA samples were first bisulphite converted per Qiagen Bisulfite conversion protocol (45) using 500 ng of DNA. Short DNA sequences that contained the CpG site of interest were amplified by PCR using PyroMark PCR kit from QUIAGEN following the manufacturer instructions (46). The PCR was performed on the 30 samples in doublets with two positive controls (Unmethylated converted DNA and methylated converted DNA) and a negative control. Pyrosequencing was performed using the PyroMark Q48 Autoprep per the instructions provided in the user manual (47). Pearson correlation was used to compare the pyrosequencing results with the methylation values of the EPIC array.

Follow- up study of the women 10-12 years after delivery

A 10-year follow-up of the women who attended STORK G is currently ongoing and expected to finish in 2021. The main aims are to assess the incidence of prediabetes and T2D and explore changes in risk factors for T2D (and CVD) over the last 10 years. We measure weight, height, physical inactivity, and blood pressure, and collect data on self-reported smoking, and chronic diseases/conditions. Dried blood spots are biobanked and about 60% meet for fasting blood samples. We estimate to reach a sample of 350 women – 50% of those eligible.

FINDINGS TO DATE

In the EPIPREG sample, we excluded women without fasting plasma glucose available in week 28±2, and those without genotype data due to low DNA concentrations or problems with DNA extraction. For Europeans, we were able to quantify DNA methylation in 99 % of the eligible samples (empty wells=2 and full plates=1), 88 % of the total number of Europeans participating in STORK G. For South Asians, methylation status could be determined in 100 % of the eligible

subjects, representing 88% of total South Asians participating in STORK G (Fig. 1). Hence, EPIPREG resulted in DNA methylation data of 312 Europeans and 168 South Asians.

Some characteristics of the mothers and offspring of the EPIPREG sample are shown in table 1. When comparing the clinical characteristics of the women and their offspring with and without DNA methylation data, the average of sedentary hours and percentage of truncal fat were significantly higher in the European subjects included in EPIPREG, whereas gestational week, hours of light physical activity, and 25-hydroxyvitamin D levels were higher in the excluded samples (supplementary table 1). The differences were small and generally followed the same trend as the overall STORK G missing data analysis (26). In South Asians we did not detect any significant differences between the women and their offspring included in EPIPREG vs the excluded individuals (supplementary table 2).

Genetic principal components distinctly separated Europeans and South Asian women (figure 2), which fully corresponded with the self-reported ethnicity. The technical validation of DNA methylation measured with the chip versus bisulphite pyrosequencing showed high agreement ($r=0.98$, $p<0.001$) (Figure 3).

STRENGTHS AND LIMITATIONS

EPIPREG is a population-based sample of European and South Asian pregnant women with epigenome-wide DNA methylation in maternal peripheral blood leukocytes. EPIPREG has detailed phenotype data from both the mother and the offspring, as well as genotype data. DNA methylation measured with the epigenome-wide chip showed high agreement with bisulphite pyrosequencing.

EPIPREG is among the larger studies that have collected epigenome-wide DNA methylation data in maternal blood during pregnancy (5, 48). The inclusion of women with both European and South Asian ethnic background enables interesting studies into the role of DNA methylation in ethnic disparities in health. Furthermore, EPIPREG has both genome-wide genotype and DNA methylation data also allowing for methylation quantitative trait loci (mQTL) analysis.

EPIPREG's population-based design and comprehensive phenotyping allows for gaining representative data about the associations between DNA methylation and a wide range of phenotypic traits, exposures and outcomes. Maternal phenotypes were collected in gestational

weeks 15 and 28, and 3 months after delivery. Furthermore, there is an ongoing follow in some women 10 to 12 years after pregnancy. In the offspring we have anthropometric data recorded in utero, at birth and during the first four years of life, as well as serum cord blood and placental tissue biobanked. Lastly, we have permission for linkage with Norwegian national registries using the personal identification number.

Despite EPIPREG's relatively large sample, our study has limited statistical power for EWAS, but should be well suited for meta-analysis efforts, or to serve as a replication cohort. Another limitation is that DNA methylation is only measured in gestational week 28 \pm 2.

COLLABORATION

EPIPREG may serve as a useful sample for generation of new hypotheses about associations between DNA-methylation and phenotypic traits relating to GDM, for replication of findings from other studies, and for meta-analysis efforts. We are currently welcoming collaborations with cohorts with similar data and researchers interested in collaboration are welcome to contact Christine Sommer, or visit our website: www.epipreg.no.

ACKNOWLEDGMENTS

We would like to thank the women who participated in the STORK Groruddalen study, Maria Sterner, Malin Neptin, and Gabriella Gremesperger at the Genomics Diabetes and Endocrinology CRC, Malmö, for experiments.

EPIPREG is supported by the South Eastern Norway Regional Health Authority, and the Norwegian Diabetes Association. G.H.M. is supported by the Norwegian Research Council (Post doctoral mobility research grant 287198), and have received funding support by Nils Normans minnegave.

REFERENCES

1. Conerly M, Grady WM. Insights into the role of DNA methylation in disease through the use of mouse models. *Dis Model Mech*. 2010;3(5-6):290-7.
2. Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation. *Epigenet Chromatin*. 2016;9.
3. Moore LD, Le T, Fan GP. DNA Methylation and Its Basic Function. *Neuropsychopharmacol*. 2013;38(1):23-38.
4. Demerath EW, Guan WH, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. 2015;24(15):4464-79.
5. Opsahl JO, Moen G-H, Qvigstad E, Böttcher Y, Birkeland KI, Sommer C. Epigenetic signatures associated with maternal body mass index or gestational weight gain: a systematic review. *Journal of Developmental Origins of Health and Disease*. 2020:1-11.
6. Sun DJY, Zhang T, Su SY, Hao G, Chen T, Li QZ, et al. Body Mass Index Drives Changes in DNA Methylation A Longitudinal Study. *Circ Res*. 2019;125(9):824-33.
7. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017;541(7635):81-+.
8. Elliott HR, Sharp GC, Relton CL, Lawlor DA. Epigenetics and gestational diabetes: a review of epigenetic epidemiology studies and their use to explore epigenetic mediation and improve prediction. *Diabetologia*. 2019;62(12):2171-8.
9. Florath I, Butterbach K, Heiss J, Bewerunge-Hudler M, Zhang Y, Schottker B, et al. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia*. 2016;59(1):130-8.
10. Cardona A, Day FR, Perry JRB, Loh M, Chu ARY, Lehne B, et al. Epigenome-Wide Association Study of Incident Type 2 Diabetes in a British Population: EPIC-Norfolk Study. *Diabetes*. 2019;68(12):2315-26.
11. Al Muftah WA, Al-Shafai M, Zaghlool SB, Visconti A, Tsai PC, Kumar P, et al. Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clin Epigenetics*. 2016;8.
12. Xie T, Gorenjak V, Stathopoulou MG, Dade S, Marouli E, Masson C, et al. Epigenome-Wide Association Study (EWAS) of Blood Lipids in Healthy Population from STANISLAS Family Study (SFS). *Int J Mol Sci*. 2019;20(5).
13. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics*. 2017;9.
14. Kazmi N, Elliott HR, Burrows K, Tillin T, Hughes AD, Chaturvedi N, et al. Associations between high blood pressure and DNA methylation. *Plos One*. 2020;15(1).
15. Richard MA, Huan TX, Ligthart S, Gondalia R, Jhun MA, Brody JA, et al. DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum Genet*. 2017;101(6):888-902.
16. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics-Uk*. 2016;8(5):599-618.
17. Dugue PA, Jung CH, Joo JE, Wang XC, Wong EM, Makalic E, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. *Epigenetics-Uk*. 2020;15(4).
18. Bruckmann C, Islam SA, MacIsaac JL, Morin AM, Karle KN, Di Santo A, et al. DNA methylation signatures of chronic alcohol dependence in purified CD3(+) T-cells of patients undergoing alcohol treatment. *Sci Rep-Uk*. 2017;7.

19. Christensen BC, Kelsey KT, Zheng SC, Houseman EA, Marsit CJ, Wrensch MR, et al. Breast Cancer DNA Methylation Profiles Are Associated with Tumor Size and Alcohol and Folate Intake. *Plos Genet.* 2010;6(7).
20. Williams D. Pregnancy: a stress test for life. *Curr Opin Obstet Gyn.* 2003;15(6):465-71.
21. Catalano PM, Tyzbir ED, Roman NM, Amini SB, Sims EAH. Longitudinal Changes in Insulin Release and Insulin Resistance in Nonobese Pregnant-Women. *Am J Obstet Gynecol.* 1991;165(6):1667-72.
22. Buchanan TA, Xiang AH. Gestational diabetes mellitus. *J Clin Invest.* 2005;115(3):485-91.
23. Bilhartz TD, Bilhartz PA, Bilhartz TN, Bilhartz RD. Making Use of a Natural Stress Test: Pregnancy and Cardiovascular Risk. *J Womens Health.* 2011;20(5):695-701.
24. Jenum AK, Sletner L, Voldner N, Vangen S, Morkrid K, Andersen LF, et al. The STORK Groruddalen research programme: A population-based cohort study of gestational diabetes, physical activity, and obesity in pregnancy in a multiethnic population. Rationale, methods, study population, and participation rates. *Scand J Public Healt.* 2010;38:60-70.
25. Sommer C, Sletner L, Jenum AK, Morkrid K, Andersen LF, Birkeland KI, et al. Ethnic differences in maternal dietary patterns are largely explained by socio-economic score and integration score: a population-based study. *Food Nutr Res.* 2013;57.
26. Jenum AK, Morkrid K, Sletner L, Vangen S, Torper JL, Nakstad B, et al. Impact of ethnicity on gestational diabetes identified with the WHO and the modified International Association of Diabetes and Pregnancy Study Groups criteria: a population-based cohort study (vol 166, pg 317, 2012). *Eur J Endocrinol.* 2012;166(3):561-.
27. Sommer C, Sletner L, Morkrid K, Jenum AK, Birkeland KI. Effects of early pregnancy BMI, mid-gestational weight gain, glucose and lipid levels in pregnancy on offspring's birth weight and subcutaneous fat: a population-based cohort study. *Bmc Pregnancy Childb.* 2015;15.
28. Sommer C, Morkrid K, Jenum AK, Sletner L, Mosdøl A, Birkeland KI. Weight gain, total fat gain and regional fat gain during pregnancy and the association with gestational diabetes: a population-based cohort study. *Int J Obes (Lond).* 2014;38(1):76-81.
29. Waage CW, Mdala I, Jenum AK, Michelsen TM, Birkeland KI, Sletner L. Ethnic differences in blood pressure from early pregnancy to postpartum: a Norwegian cohort study. *J Hypertens.* 2016;34(6):1151-9.
30. WHO. Definition and classification of diabetes mellitus and its complications. Report of a WHO consultation. Part 1: Diagnosis and classification of diabetes mellitus. WHO. 1999.
31. Agarwal MM, Boulvain M, Coetzee E, Colagiuri S, Falavigna M, Hod M, et al. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: A World Health Organization Guideline. *Diabetes Res Clin Pr.* 2014;103(3):341-63.
32. Morkrid K, Jenum AK, Sletner L, Vardal MH, Waage CW, Nakstad B, et al. Failure to increase insulin secretory capacity during pregnancy-induced insulin resistance is associated with ethnicity and gestational diabetes. *Eur J Endocrinol.* 2012;167(4):579-88.
33. Eggemoen AR, Jenum AK, Mdala I, Knutsen KV, Lagerlov P, Sletner L. Vitamin D levels during pregnancy and associations with birth weight and body composition of the newborn: a longitudinal multiethnic population-based study. *Brit J Nutr.* 2017;117(7):985-93.
34. Sommer C, Jenum AK, Waage CW, Morkrid K, Sletner L, Birkeland KI. Ethnic differences in BMI, subcutaneous fat, and serum leptin levels during and after pregnancy and risk of gestational diabetes. *Eur J Endocrinol.* 2015;172(6):649-56.
35. Berntsen S, Richardsen KR, Morkrid K, Sletner L, Birkeland KI, Jenum AK. Objectively recorded physical activity in early pregnancy: A multiethnic population-based study. *Scand J Med Sci Spor.* 2014;24(3):594-601.

36. Morkrid K, Jenum AK, Berntsen S, Sletner L, Richardsen KR, Vangen S, et al. Objectively recorded physical activity and the association with gestational diabetes. *Scand J Med Sci Spor.* 2014;24(5):E389-E97.
37. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults: Guidance for Prescribing Exercise. *Med Sci Sport Exer.* 2011;43(7):1334-59.
38. Sletner L, Rasmussen S, Jenum AK, Nakstad B, Jensen OHR, Vangen S. Ethnic differences in fetal size and growth in a multi-ethnic population. *Early Human Development.* 2015;91(9):547-54.
39. Sletner L, Nakstad B, Yajnik CS, Mørkrid K, Vangen S, Vårdal MH, et al. Ethnic differences in neonatal body composition in a multi-ethnic population and the impact of parental factors: a population-based cohort study. *Plos One.* 2013;8(8):e73058.
40. Miller SA, Dykes DD, Polesky HF. A Simple Salting out Procedure for Extracting DNA from Human Nucleated Cells. *Nucleic Acids Res.* 1988;16(3):1215-.
41. Moen GH, LeBlanc M, Sommer C, Prasad RB, Lekvas T, Normann KR, et al. Genetic determinants of glucose levels in pregnancy: genetic risk scores analysis and GWAS in the Norwegian STORK cohort. *Eur J Endocrinol.* 2018;179(6):363-72.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira Manuel A R, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics.* 2007;81(3):559-75.
43. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS genetics.* 2009;5(6):e1000529.
44. Min JL, Hemani G, Smith GD, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics.* 2018;34(23):3983-9.
45. Qiagen. EpiTec (R) Bisulfite handbook. 2014.
46. Qiagen. PyroMark PCR Handbook. 2009.
47. Qiagen. PyroMark (R) Q48 Autoprep User Manual. 2016.
48. Moen GH, Sommer C, Prasad RB, Sletner L, Groop L, Qvigstad E, et al. Epigenetic modifications and gestational diabetes: a systematic review of published literature. *Eur J Endocrinol.* 2017;176(5):R247-R67.

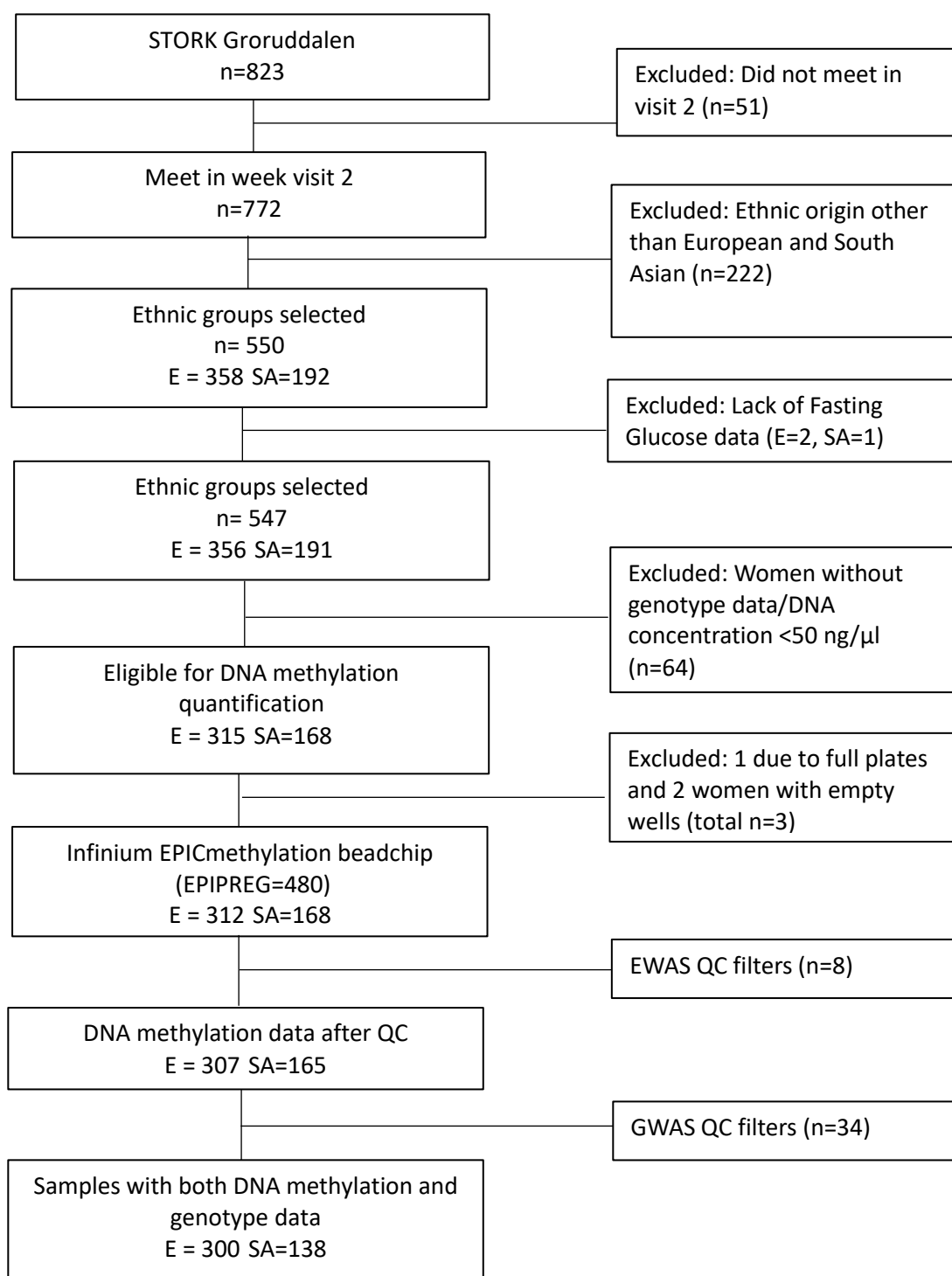


Figure 1. Flow chart of the EPIPREG sample. E= European, SA = South Asian

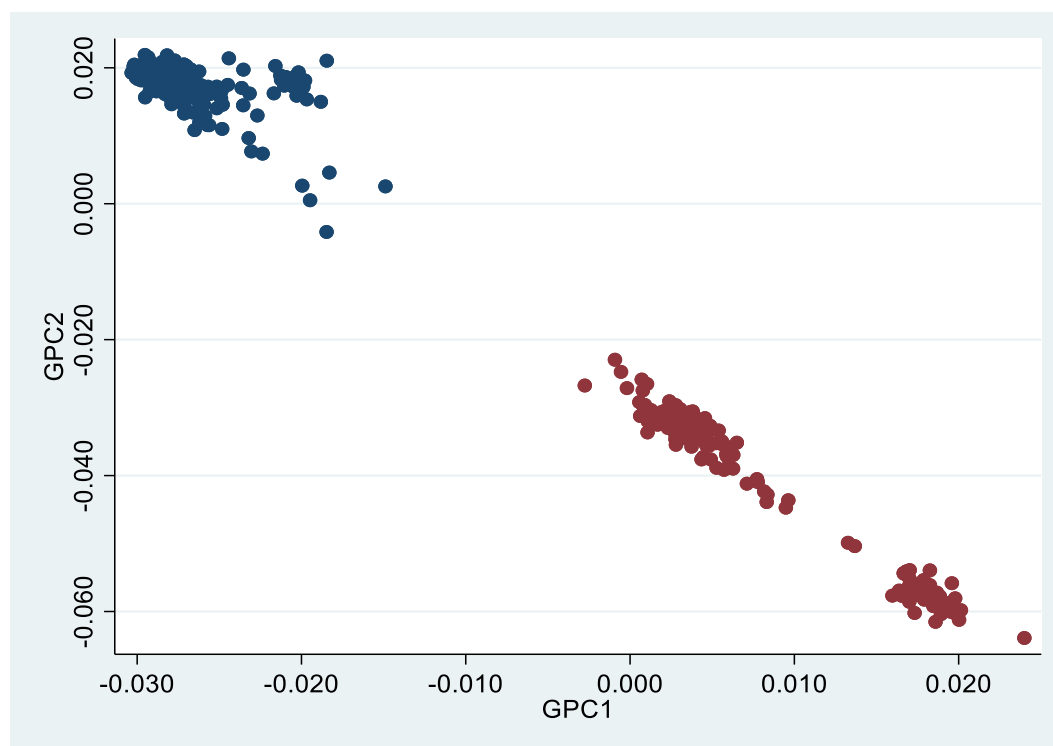


Figure 2: Scatter dot plot of Genetic PC1 (GPC1) and PC2 (GPC2) (n=438). Blue dots are Europeans, red dots are South Asians, based on self-reported ethnicity.

Table 1: Characteristics of the EPIPREG sample. Data from gestational visit 2 for cross-sectional associations with DNA methylation data, otherwise specified. Data are presented in mean (SD) for normally distributed variables and median [IQR] for non-normal variables. Categorical variables are presented by frequency (%).

Variable	N	Europeans, n=312	South Asians, n=168
Age	480	30.1 (4.6)	28.2 (4.6)
Weeks' gestation	480	28.1 (1.2)	28.2 (1.2)
Height (cm)	480	167.4 (5.8)	159.9 (5.7)
Smoking status	476		
Current		20 (6.5)	1 (0.6)
3 months pre-pregnancy		80 (25.9)	2 (1.2)
Former*		87 (28.2)	9 (5.4)
Never		122 (39.5)	155 (92.8)
Pre-pregnancy alcohol intake, n (%)	472	234 (76.2)	5 (3.0)
Pre-pregnancy BMI (kg/m ²)	474	24.6 (4.9)	23.8 (4.1)
Actual BMI (kg/m ²)	478	27.8 (4.7)	26.8 (4.1)
Total fat (%)	465	29.6 (9.7)	26.2 (8.2)
Truncal fat (%)	465	15.9 (5.5)	14.0 (5.4)
Systolic blood pressure (mmHg)	480	107.0 (9.6)	101.1 (8.7)
Diastolic blood pressure (mmHg)	480	68.4 (7.1)	66.1 (6.9)
First degree relative with diabetes	474	43 (14.0)	79 (47.3)
GDM (WHO ₂₀₁₃), n (%)	480	76 (24.4)	70 (41.7)
GDM (WHO ₁₉₉₉), n (%)	478	37 (11.9)	25 (15.1)
Fasting glucose (mmol/L)	480	4.7 (0.6)	5.0 (0.6)
2 hour glucose (mmol/L)	477	6.0 (1.4)	6.4 (1.5)
HbA1c (%)	475	5.1 (0.3)	5.3 (0.3)
Fasting Insulin (pmol/L)	474	48.0 [33.0, 70.2]	71.0 [57.0, 100.5]
Fasting C-peptide (pmol/L)	474	712.0 [560.2, 901.8]	855.5 [688.0, 1067.5]
HOMA-B	474	173.2 [151.3, 199.7]	179.1 [154.9, 207.9]
HOMA-IR	474	1.5 [1.2, 1.9]	1.8 [1.4, 2.3]
Total cholesterol (mmol/L)	480	6.4 (1.1)	6.0 (1.0)
HDL-cholesterol (mmol/L)	480	1.9 (0.4)	1.9 (0.4)
LDL-cholesterol (mmol/L)	473	3.7 (1.0)	3.3 (0.9)
Triglycerides (mmol/L)	480	2.0 (0.7)	2.0 (0.7)
Folate (nmol/l)	471	13.0 [9.9, 18.0]	12.0 [9.4, 16.0]
Vitamin B12 (µmol/L)	472	211.0 [169.2, 253.0]	186.0 [150.0, 232.2]
Vitamin D (µmol/L)	472	69.9 (27.8)	45.6 (22.2)
Leptin (µmol/L)	476	1599.5 [966.9, 2497.7]	2276.9 [1532.8, 3295.8]
Folate (nmol/l)	471	13.0 [9.9, 18.0]	12.0 [9.4, 16.0]
Vitamin B12 (µmol/L)	472	211.0 [169.2, 253.0]	186.0 [150.0, 232.2]
Sedentary time (hours/day)**	382	18 (1.6)	17.8 (1.7)
Light physical activity (hours/day)**	382	4.3 (1.3)	4.6 (1.3)
Moderate-intense physical activity (hours/day)**	382	1.2 (0.8)	1.0 (0.9)

Offspring data			
Gestational age (days)	475	280.9 (11.7)	277.1 (12.6)
Female sex, n (%)	462	147 (49.0)	79 (48.8)
Birth weight (g)	472	3582.2 (529.4)	3211.7 (512.2)
Birth length (cm)	432	50.1 (2.3)	49.3 (2.2)
Neonatal sum of skinfolds (mm)	359	18.8 (4.0)	17.0 (3.6)

Despite 8 individuals did not pass the QC procedure, the data of the 480 individuals are presented for informative purposes

WHO = World health organization. GDM = gestational diabetes mellitus. BMI = body mass index. HDL = high-density lipoproteins. LDL = low-density lipoproteins.

* Ex-smokers and occasional smokers who did not smoke 3 months before pregnancy

* *Women with at least one valid day of registered physical activity (Armband)

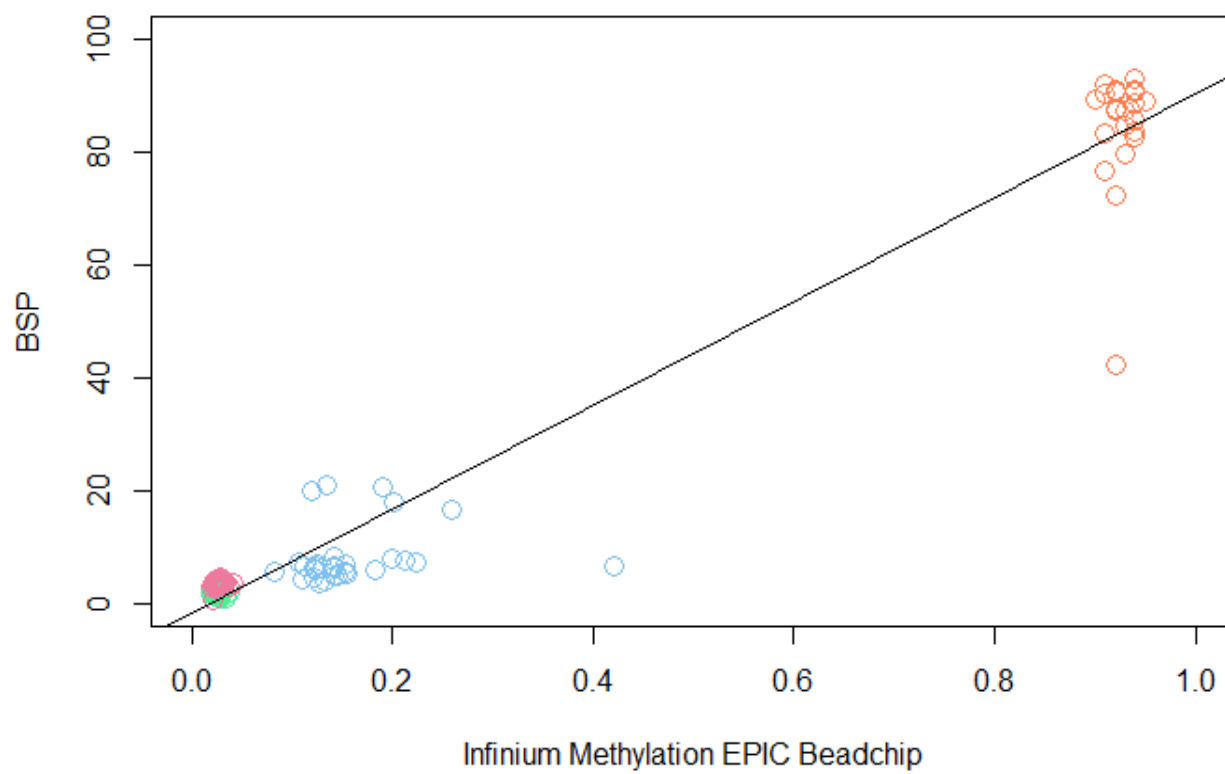


Figure 3. Scatter plot showing the relationship between the DNAm values of the four selected CpG sites quantified with the EPIC chip (x-axis) and with Bisulphite Pyrosequencing (y-axis). Dots colour code: Pink: cg02098128, green: cg17148978, blue: cg14120215, orange: cg19327414.