This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Genetic and environmental influences on DNA methylation

Castillo-Fernandez, Juan Edgar

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 15. Jan. 2025

Genetic and environmental influences on DNA methylation

Juan Edgar Castillo-Fernandez

Submitted for the degree of Doctor of Philosophy

Department of Twin Research & Genetic Epidemiology

King´s College London

Supervisors:

Dr. Jordana Bell

Prof. Tim Spector

Acknowledgements

This wok was funded and promoted by the European Union's FP7 project EpiTrain. I also received support from the Mexican Secretariat of Public Education and the Mexiquense Council of Science and Technology.

I wish to thank my supervisors Dr. Jordana Bell and Prof. Tim Spector who provided me guidance and support during this process.

A special thanks to the staff that made possible and accessible to the scientific community the resources of TwinsUK and PETS, and to all the volunteers that dedicated their time to provide personal information and biological samples.

Finally, I would like to thank my collaborators, in particular, Dr. Jeff Craig, Dr. Jane Loke, and Prof. Richard Saffery, who hosted me at Murdoch Childrens Research Institute in Australia, for their contributions to the work presented in Chapter 4 and to the CHARGE consortium's Epigenetics Group, especially analysts, for their contributions to the work presented in Chapter 5.

Abstract

Epigenetic mechanisms respond to both genetic and environmental factors, but these underlying effects are not yet fully characterized or completely understood. In this thesis I used twins as a tool for evaluating the effect of genetic and environmental impacts on DNA methylation, a well-known epigenetic mechanism. DNA methylation was studied on a genome-wide scale in human blood samples with a combination of bioinformatics, computational, and statistical approaches. First, genetic influences on DNA methylation profiles were assessed by estimation of the heritability of DNA methylation and identification of methylation-quantitative trait loci in identical and non-identical twins profiled with the commonly used Infinium HumanMethylation450 BeadChip and the new enhancer-enriched Infinium MethylationEPIC Beadchip. Strong genetic effects (heritability > 0.4) were detected for 10% of sites and common genetic variants were identified to affect methylation levels at 22% of sites, from the 771,169 interrogated CpG sites across the genome. Second, I explored influences on the early-life methylome by comparing genome-wide DNA methylation profiles in naturally-conceived twins and twins conceived by *in vitro* fertilisation. Analysis of epigenetic profiles obtained by methylated DNA immunoprecipitation coupled with deep sequencing detected small changes at a gene previously associated with infertility, *TNP1*. Finally, I explored the effect of intrinsic factors on adult DNA methylation profiles, performing epigenome-wide studies of menopause and related phenotypes such as the use of hormone replacement therapy, which have metabolic consequences in middle-aged women. Epigenetic changes at seven CpG sites were associated with hormone replacement therapy. In summary, the results presented in this thesis give insights into genetic and specific environmental influences on the human epigenome.

Statement of work undertaken

I performed all statistical analyses presented in this thesis. The preprocessing of microarrays used in Chapter 3 and 5 was performed by Dr. Pei-Chien Tsai.

The study presented in Chapter 4 was done in collaboration with the Murdoch Childrens Research Institute and BGI-Shenzhen. The work has been published under the title “DNA methylation changes at infertility genes in newborn twins conceived by in vitro fertilisation”. Names and specific contributions are detailed in the annexed manuscript (Annex X). Briefly, samples were collected by the Murdoch Childrens Research Institute, MeDIP-seq data were produced by BGI-Shenzhen, EpiTYPER data were produced in the Murdoch Childrens Research Institute.

Lastly, I led the meta-analyses presented in Chapter 5. The analysis plan was designed with contributions from Dr. Ken Ong, Dr. Cathy Elks, Dr. Michael Mendelson, Dr. Jordana Bell and me. The meta-analysis was conducted using summary statistics kindly provided by the participating CHARGE cohorts. The contributing lead analysts were Carola Marzi (KORA), Michael Mendelson (FHS), Alexia Cardona (EPIC-Norfolk), Cindy Boer (Rotterdam Study), and myself (TwinsUK). I performed the meta-analysis across cohorts.

Publications arising from this thesis

Original research papers

Castillo-Fernandez JE, Loke YJ, Bass-Stringer S, Gao F, Xia Y, Wu H, Lu H, Liu Y, Wang J, Spector TD, Saffery R, Craig JM, Bell JT. DNA methylation changes at infertility genes in newborn twins conceived by in vitro fertilisation. Genome Med. 2017;9:28.

Morris JA, Tsai P-C, Joehanes R, Zheng J, Trajanoska K, Soerensen M, Forgetta V, **Castillo-Fernandez JE**, Frost M, Spector TD, Christensen K, Christiansen L, Rivadeneira F, Tobias JH, Evans DM, Kiel DP, Hsu Y-H, Richards JB, Bell JT. Epigenome-wide association of DNA methylation in whole blood with bone mineral density. J Bone Miner Res. 2017. Advance online publication. doi:10.1002/jbmr.3148.

Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P-C, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE, Aslibekyan S, Brody JA, Kühnel B, Herder C, Almli LM, Zhi D, Wang Y, Huan T, Yao C, Mendelson MM, Joehanes R, Liang L, Love S-A, Guan W, Shah S, McRae AF, Kretschmer A, Prokisch H, Strauch K, Peters A, Visscher PM, Wray NR, Guo X, Wiggins KL, Smith AK, Binder EB, Ressler KJ, Irvin MR, Absher DM, Hernandez D, Ferrucci L, Bandinelli S, Lohman K, Ding J, Trevisi L, Gustafsson S, Sandling JH, Stolk L, Uitterlinden AG, Yet I, **Castillo-Fernandez JE**, Spector TD, Schwartz JD, Vokonas P, Lind L, Li Y, Fornage M, Arnett DK, Wareham NJ, Sotoodehnia N, Ong KK, van Meurs JBJ, Conneely KN, Baccarelli AA, Deary IJ, Bell JT, North KE, Liu Y, Waldenberger M, London SJ, Ingelsson E, Levy D. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2016; February:1–12.

Reviews

Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease. Genome Med. 2014;6:60.

Yet I, Tsai P-C, **Castillo-Fernandez JE**, Carnero-Montoro E, Bell JT. Genetic and environmental impacts on DNA methylation levels in twins. Epigenomics. 2016;8:105–17.

Abbreviations

5fC, 5-formylcytosine

5hmcC, 5-hydroxymethylcytosine

5mC, 5-methylcytosine

ALSPAC, Avon Longitudinal Study of Parents and Children

CAHRGE, Cohorts for Heart and Aging Research in Genomic Epidemiology

CBMCs, Cord Blood Mononuclear Cells

DMP, Differentially Methylated Position

DMR, Differentially Methylated Region

DZ, Dizygotic

EGA, European Genome-phenome Archive

EPIC-Norfolk, European Prospective Investigation of Cancer-Norfolk

EWAS, Epigenome-Wide Association Scans

FDR, False Discovery Rate

FHS, Framingham Heart Study

GIFT, Gamete Intra-Fallopian Transfer

H^2, broad sense heritability

h^2, narrow sense heritability

HRT, Hormone Replacement Therapy

ICC, Intra-class Correlation Coefficient

ICR, Imprinting Control Region

ICSI, Intracytoplasmic Sperm Injection

IVF, In Vitro Fertilization

KORA, Cooperative Health Research in the Augsburg Region

LCL, Lymphoblastoid Cell Line

MAF, Minor Allele Frequency

MeDIP-seq, Methylated DNA Immunoprecipitation coupled with sequencing

meQTLs, methylation-Quantitative Trait Loci

MZ, Monozygotic

PBMC, Peripheral Blood Mononuclear Cell

PETS, Peri/postnatal Epigenetic Twin Study

RRBS, Reduced Representation Bisulfite Sequencing

RS, Rotterdam Study

TwinsUK, UK Adult Twin Registry

WBCs, Whole Blood Cells

WGBS, whole genome bisulfite sequencing

Contents

List of Tables

List of Figures

1. Introduction

Epigenetics emerged during the first half of the twentieth-century as the study of biological mechanisms involved in embryonic development and cell differentiation [1]. More recently, it has been defined as the study of mitotically inherited patterns of gene expression that are not explained by modifications in the DNA sequence [2]. DNA methylation, histone modifications and regulatory RNAs are considered epigenetic mechanisms that have been involved in processes, such as, cell differentiation, imprinting, and X-inactivation [3]. One characteristic of epigenetic patterns is that they are able to change in response to environmental stimuli [3]. Such change will then be inherited through cell division to allow daughter cells to conserve the identity of their parent cell. The dynamics of DNA methylation have been studied specially during early embryogenesis, when the genome undergoes global DNA demethylation required to reach a totipotent state followed by the establishment of new DNA methylation marks necessary for cell differentiation and the commitment of somatic cells to a specific role [4]. Disruption of some methylation or demethylation pathways has shown to disrupt development [4]. Apart from its key role in developmental biology, epigenetics has recently become relevant to epidemiology. It offers the promise to unravel biological mechanisms underlying disease and has potential as a biomarker of disease or disease progression. In 1983, Feinberg and Vogelstein [5] reported epigenetic alterations at the human growth hormone and γ-globin genes in colon cancer patients. Since then, epigenetic alterations have been reported in many other types of cancer [6], autoimmune diseases [7], diabetes [8], Alzheimer's disease [9], Parkinson's disease [10], asthma [11], and multiple other human complex traits [12–14].

One of the best characterized epigenetic mechanisms in humans is DNA methylation of cytosines at CpG dinucleotides. It was proposed as a mechanism of mammalian gene regulation in 1975 [15, 16]. DNA methylation may play different roles, but typically, CpG methylation at the 5' regions of genes reduces gene expression. This down-regulation has been shown to occur due to either the inability of specific transcription factors to bind DNA in the presence of methylated CpGs or through the recruitment of methyl-CpG- binding proteins with transcription repression activity [17–19]. Conversely, in gene body regions, patterns of high methylation have been found in transcriptionally active genes [20]. The next most studied epigenetic mechanisms are the modification of histone residues, which are very diverse and may include acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation, and many others. Combinatorial modifications at selected histone residues trigger specific gene expression activity [21]. Less studied epigenetic regulators include histone variants [22, 23], ATP dependent chromatin remodelling complexes [24], and non-coding RNAs [25]. It is known that different epigenetic regulators might work in concert to regulate gene expression. A well-known example is X-inactivation, in which a non-coding RNA within *Xist* acts in *cis* to silence the X through the recruitment of the chromatin modifier complex PRC2*,* which in turn lays down H3K27me3 [26]. Despite the importance of other epigenetic regulators, in this thesis I focus mainly on the study of DNA methylation given the feasibility of its profiling in a large number of samples.

1.1 DNA methylation

The transfer of a methyl group from an *S*-adenyl methionine methyl donor to the fifth carbon of a cytosine by a DNA methyltransferase enzyme to produce 5-methylcytosine (5mC) is commonly known as DNA methylation [27]. This

covalent modification usually occurs at cytosines that are followed by a guanine (CpG dinucleotide). In the haploid human genome there are ~28 million CpG sites [28], which can be found in a methylated or unmethylated form. The identity of a differentiated cell, which is determined by its gene expression profile, will depend on its unique DNA methylation pattern.

In mammals, *de novo* DNA methylation is laid down by the DNA methyltransferases Dnmt3a and Dnmt3b. The mitotic inheritance of this mark requires another methyltransferase, Dnmt1, which copies methylation patterns from the parental strand to the daughter strand during DNA replication [29]. This is possible thanks to the symmetric property of the CpG dinucleotide, which is also read on the opposite strand –from 5' to 3'- as CpG.

DNA methylation patterns are dynamic, which means that DNA methylation can be reversed. The removal of methylation is achieved by passive or active mechanisms. Passive demethylation refers to the lack of maintenance of 5mC and the subsequent dilution of this modification after a number of cell divisions. In contrast, active mechanisms require the enzymatic removal of 5mC [30]. Although no demethylase capable of converting 5mC to cytosine has been identified, processes that can result in this loss are oxidation and deamination of 5mC by members of the ten-eleven translocation family and the activation-induced cytidine deaminase, respectively, followed by base replacement via DNA repair [31].

The specific DNA methylation pattern of a cell is defined through development. Global and complex waves of DNA de-methylation and re-methylation are observed in the pre-implantation embryo and primordial germ cells. Right after fertilization, the zygote undergoes global DNA demethylation, a characteristic state of pluripotent cells. DNA demethylation occurs independently

in the paternal and maternal pronuclei, happening at a faster rate in the paternal genome [32]. Imprinting marks escape this step of reprogramming given that are parent of origin-specific and essential for development. As the embryo develops, DNA methylation is regained until cells define their relatively stable epigenetic identity as differentiated cells. Primordial germ cells are specified from the epiblast in the post-implantation embryo and have to reset their DNA methylation profiles, including imprinting marks, in order to gain germline potency and imprinting marks specific for male or female gametes [33]. On the contrary, differentiated somatic cells inherit tissue-specific DNA methylation patterns that remain relatively stable in almost every cell type [34]. As seen in longitudinal studies, some changes do occur and are accumulated throughout life [35, 36]. Some of these changes are associated with age, which predominately show a gain in methylation [37]. Whether differential DNA methylation is originated during embryonic development or acquired after birth, DNA methylation variation within populations is common and is of scientific interest to understand the drivers of such variation.

1.2 Factors that influence DNA methylation

DNA methylation patterns are crucial for the correct growth and development of humans, however, there is room for variability thanks to changes that occur throughout life due to genetic, environmental, or stochastic factors.

1.2.1 Genetic influences on DNA methylation

Familial clustering of DNA methylation changes over time suggested that this modification is under genetic control [38]. Twin studies also suggested a genetic influence on DNA methylation since identical or monozygotic (MZ) twins, who share most of their genetic variants, had more similar DNA methylation profiles

than non-identical or dizygotic (DZ) twins, who share in average 50% of their genetic variants [39]. Kaminsky *et al*. [11] found significant differences when estimating locus-specific (~6,000 loci) DNA methylation intra-class correlation coefficient (ICC) differences ($ICC_{MZ} - ICC_{DZ}$) of 40 age- and sex-matched MZ and DZ twin pairs in buccal epithelial cells (mean $ICC_{MZ} - ICC_{DZ} = 0.15 \pm 0.0039$, $P = 1.2 \times 10^{-294}$) and white blood cells (mean $ICC_{MZ} - ICC_{DZ} = 0.0073 \pm 0.0034$, $P = 0.044$). Other twin studies have also observed evidence of genetic influence on DNA methylation for a proportion of CpG sites in the genome.

Early studies suggested that variants of genes involved in the establishment and maintenance of DNA methylation had an impact on global DNA methylation levels [40, 41]. Nevertheless, most of the evidence of genetic influences on DNA methylation comes from the study of quantitative trait loci for DNA methylation (meQTLs). These meQTLs have been reported in brain [42, 43], lymphoblastoid cell lines (LCLs) [44], adipose tissue [45], and lung [46]. The peak abundance of meQTLs is in close proximity to the target CpG site and the proportion of the variance explained can be high, up to 63% in LCLs [44] and up to 79.8% in lung [46].

1.2.2 Environmental influences on DNA methylation

Environmental influences may also contribute to changes in DNA methylation profiles. Tobacco smoking is a commonly studied environmental factor in human epigenetics as a prenatal and adult life exposure. It was first associated with differential DNA methylation in an adult population (50 to 60 years of age) at the *F2RL3* gene, which encodes coagulation factor II receptor-like 3 [47]. Since then, multiple studies, looking at different tissues, have identified and replicated smoking-associated changes in DNA methylation at several genes, including the well replicated *AHRR* gene [48–50]. Interestingly, changes of methylation at

AHRR were also observed and replicated when looking at effects of maternal smoking on DNA methylation profiles of newborns [51, 52].

The number of studies looking for environment-associated changes in human DNA methylation at a genome-wide level is steadily growing. Apart from smoking, other environmental factors considered include influences of alcohol [53, 54], tea and coffee [55], maternal nutrition [56], dietary fat [57, 58], folate [59, 60], fluorinated chemicals [61], season of birth [62, 63] and season of DNA collection [64], air pollution [65], and others.

Epigenetic studies can particularly benefit from the use of twins given that genetic and environmental factors influence DNA methylation profiles. The classical twin study design can be used to study the extent to which genetic and environmental factors influence DNA methylation profiles. The study design of discordant MZ twins can be used to study the association of epigenetics and disease or environmental exposures controlling for many potential confounders such as genetic factors, age, gender, maternal effects, cohort effects, and most in utero and environmental influences.

1.3 Twins and epigenetics

Most complex phenotypes arise as a result of the interplay between genetics and environment. In epidemiology, it is of interest to determine what proportion of the phenotypic variance each of these factors can explain. Classical twin studies make use of MZ and DZ twins to decipher these influences. Since MZ twins are assumed to share virtually 100% of their genetic variants, and DZ twins share on average only 50% of their variants, the difference in phenotype concordance levels between these two groups can be indicative of the genetic influence on the phenotype. Greater phenotype concordance in MZ twins would point to a higher contribution of genetics to the disease.

Heritability in the broad sense has been used to define the fraction of the total phenotypic variance (σ_P^2) in a population that can be attributed to genetic variance (σ_G^2) $\left[H^2 = \frac{\sigma_G^2}{\sigma_P^2}\right]$, but genetic variance can further be partitioned into the variance attributed to additive (σ_A^2), dominant (σ_D^2), and epistatic (σ_I^2) effects. This partition gives origin to the definition of heritability (h^2) in the narrow sense, which only considers the fraction of the total phenotypic variance attributed to additive genetic effects $\left[h^2 = \frac{\sigma_A^2}{\sigma_P^2}\right]$ [66]. The rest of the variation that is not attributed to genetic effects is attributed to environmental influences, which can be divided into shared and unique environmental influences and random error. The classical twin estimate of heritability, known as h^2, is defined as twice the difference between MZ and DZ intra-pair correlation coefficients (ICC) $[h^2 = 2(ICC_{MZ} - ICC_{DZ})]$ [67]. The twin model also assumes that MZ and DZ twins equally share the environment and a fraction of the correlation is due to the shared environmental influences. The fraction corresponding to shared environmental effects is estimated as the difference between the total correlation and the heritability estimate. Finally, the missing variation is attributed to unique environmental effects plus error.

Heritabilities can be calculated for particular epigenetic variants at specific epigenetic loci by treating them like phenotypes. Genome-wide studies using the classical twin design have estimated that the mean heritability of DNA methylation in adult tissues is between 18-20% [3–5] and between 5-12% in neonatal tissues [6]. These results suggest that environmental and stochastic factors may have a greater effect on methylation sites than genetic factors.

1.4 Epigenome-wide association scans (EWAS)

Epigenome-wide association scans (EWAS) aim to determine epigenetic factors across the genome that are associated with complex traits [68]. In the context of DNA methylation this consists of testing the association between DNA methylation and a trait of interest at a large number of methylation sites across the genome.

Common considerations for EWAS are the choice of DNA methylation assay, tissue selection, identification of confounding variables, and study design [68–70]. Options for genome-wide DNA methylation interrogation are array-based methods, enrichment methods, sequencing methods, or a combination of them. The Infinium methylation assays, which include the promoter-enriched HumanMethylation27 BeadChip, the gene centric HumanMethylation450 BeadChip, and the enhancer-enriched MethylationEPIC BeadChip are examples of array-based methods; they offer CpG-site resolution, but are limited to a defined number of sites targeted throughout the genome. Enrichment-based methods can be based on antibody affinity binding followed by sequencing, such as, methylated DNA immunoprecipitation (MeDIP-seq). MeDIP-seq does not offer CpG-site resolution, but allows a larger coverage of the genome since it is not limited to a number of predefined probes. Whole genome bisulfite sequencing (WGBS) is the gold standard for DNA methylation characterization, although its cost is still high for epidemiological studies. The technique is based on the conversion of unmethylated cytosines to uracil followed by sequencing. This method offers single-base resolution, genome-wide coverage, and is not biased towards the capture of specific regions as enrichment-based methods. One of the limitations of bisulfite treatment is its inability to discriminate between 5mC and its oxidized form, 5-hydroxymethylcytosine (5hmC). Protocols of

oxidative bisulfite sequencing have now overcome this problem by oxidizing selectively 5hmC to 5-formylcytosine (5fC), which is then converted to uracil as an unmethylated cytosine [71]. A more cost-effective version of WGBS is reduced representation bisulfite sequencing (RRBS). This method uses methylation-sensitive restriction enzymes to enrich for regions of high CpG density and reduce the amount of DNA to be bisulfite converted and sequenced [72].

The next consideration is the selection of the appropriate tissue since epigenetic patterns are tissue-specific. Tissue availability and the biological question may define this selection. Another consideration is cell heterogeneity of the tissue since methylation differences may represent differences in cell composition rather than methylation changes present across multiple cell types. In a similar way, confounding may be introduced if the population of study is not homogenous with respect to factors that impact on DNA methylation, such as biological (age, gender, and others) or technical effects. Lastly, another important consideration is the study design. A retrospective population-based study allows for large samples, however has the drawbacks of selection bias as the sample might not properly represent the population. An attractive design in epigenetics uses discordant MZ twins, which allows controlling for many confounding variables such as age, gender, and most genetic and environmental effects [73]. Another attractive design is the longitudinal study. This allows inferring causation, for example if samples are collected before and after disease onset, but the limited availability of longitudinal samples is a challenge.

1.5 Project overview

The aim of this project was to understand the factors that drive DNA methylation variation in human populations, specifically genetic and environmental effects.

The first stage of the project looked at genetic influences. I first assessed on a genome-wide level the heritability of DNA methylation in blood samples of identical and non-identical twins profiled with the Illumina's Infinium HumanMethylation450 BeadChip and the enhancer-enriched Infinium MethylationEPIC Beadchip. I then investigated the regions of the genome influencing DNA methylation in a study of methylation quantitative trait loci.

The second component of the thesis aimed to investigate the effect of environment. Given that embryonic stages of development are crucial for the establishment of epigenetic profiles, I explored the effect of perturbations to the cellular environment during conception on the epigenetic profile of newborns. I compared genome-wide DNA methylation profiles of naturally-conceived twins and twins conceived by *in vitro* fertilization to determine if epigenetic changes were introduced as a result of the fertility treatment. Finally, I explored the effect of intrinsic environment, specifically hormonal changes, on adult DNA methylation profiles. To this end I selected menopause and the use of hormone replacement therapy as a proxy measure of hormonal changes in adult females, and both of these factors have been shown to have profound metabolic implications in middle-aged women.

1.5.1 Research hypothesis and objectives

The overall research hypothesis of my thesis is that DNA methylation responds to intrinsic and extrinsic factors from conception to adulthood.

To address this, I tested three specific research hypotheses and aims.

Chapter 3

Research hypothesis: Greater genetic effects will be found in the EPIC array compared to its predecessor the 450K array given that it interrogates more CpG sites at intergenic regions, which are more likely to be under control of genetic variants according to previous studies.

Aims:

- To characterize DNA methylation heritability at different functional regions of the genome using the classical twin study
- To identify the genetic variants contributing to DNA methylation variability

Chapter 4

Research hypothesis: *In vitro* fertilisation overlaps with an important period of global epigenetic reprogramming affecting not only imprinting, therefore, I hypothesize that differences in DNA methylation between newborns naturally-conceived and those conceived by in *vitro* fertiilsation will be observed throughout the genome.

Aims:

- To conduct EWAS of *in vitro* fertilisation.
- To explore the impact of *in vitro* fertilisation on imprinting marks

Chapter 5

Research hypothesis: DNA methylation changes associated to the menopausal transition will be observed in blood given the myriad of metabolic changes that occur in menopausal women.

Aim:

- To conduct EWAS on age at menopause, menopausal status, and use of hormone replacement therapy

2. Materials

The work presented in this thesis is based on three main studies that explore different population samples as described in **Table 2-1**. All studies presented in this thesis were conducted with appropriate ethics approval. All participants provided written informed consent and in case of minors, consent was provided by the parents.

Table 2-1. Samples

Study	Cohort(s)	Description
Chapter 3: Genetic influences on DNA methylation	TwinsUK	DNA source: Whole blood Platform: 450K (n=728) Platform: EPIC (n=226)
Chapter 4: The effect of in vitro fertilisation on DNA methylation profiles at birth	PETS	DNA source: Whole blood (n=98); Cord blood mononuclear cells (n=82) Platform: MeDIP-seq
Chapter 5: DNA methylation changes associated with menopause related traits	CHARGE consortium (TwinsUK, EPIC-Norfolk, ALSPAC, KORA, RS, FHS)*	DNA source: Whole blood Platform: 450K (n≈2500, see chapter 5 for details)

*Participating cohorts: The UK Adult Twin Registry (TwinsUK), The European Prospective Investigation of Cancer (EPIC)-Norfolk, The Avon Longitudinal Study of Parents and Children (ALSPAC), Cooperative Health Research in the Augsburg Region (KORA), Rotterdam Study (RS), and The Framingham Heart Study (FHS)

2.1 The UK Adult Twin Registry (TwinsUK)

Genotyping, DNA methylation, and gene expression data used in Chapters 3 and 5 were collected by the UK Adult Twin Registry (TwinsUK). TwinsUK is a nation-wide registry of adult twins hosted by the Department of Twin Research, King's College London, UK. TwinsUK is the largest twin registry in the UK with over 12,000 volunteers (>80% female, 50:50 zygosity ratio) recruited trough media campaigns without selecting for particular diseases. The registry started in 1992 recruiting middle-aged female twins and from 1995 the invitation was extended to men and women over 18 years old.

2.1.1 TwinsUK phenotype data

The main areas of research in TwinsUK have been the study of common complex diseases and the process of healthy aging. Twins receive detailed disease and life-style questionnaires and are asked to attend clinical assessments where biological samples are taken and multiple phenotypes are measured. The phenotypes collected cover areas of endocrinology, allergy, toxicology, cardiology, gastroenterology, neurology, oncology, radiology, rheumatology, ophthalmology, urology, and obstetrics, among others. Biological samples include, but are not limited to, serum, plasma, fat, skin, urine, and stool. TwinsUK collects hundreds of clinical, biochemical, behavioral, and socio-economic variables, but also 'omics' data including transcriptomics, epigenomics, proteomics, metabolomics, microbiomics, and metagenomics.

The cohort also benefits from the collection of longitudinal data. Between 1992 and 2004 a total of 5725 twins attended a comprehensive baseline visit. Active twins were invited for a follow-up visit between April 2004 and May 2007 and a second follow-up visit that started in August 2007. More information can be found on the registry's website (www.twinsuk.ac.uk) and published cohort profiles [74–77]. Further information with respect to specific TwinsUK phenotypes explored in this thesis is provided in Chapter 3 and 5.

2.1.2 TwinsUK genotype data

More than 5,000 participants of TwinsUK have been genotyped with one of two chips, Illumina HumanHap300 BeadChip and Illumina HumanHap610 QuadChip. Furthermore, additional genetic variants have been imputed in this cohort using the HapMap and 1000 Genomes project reference panels [78, 79]. Further details on the genotype data used in this thesis are provided in Chapter 3.

2.1.3 TwinsUK DNA methylation data

Genome-wide DNA methylation has been interrogated in nearly 1,000 twins using the Infinium HumanMethylation450 BeadChip (450K array) and the Infinium MethylationEPIC BeadChip (EPIC array), which target more than 450,000 and 850,000 CpG sites, respectively. Detailed information on the datasets and quality control analysis of these data is provided in Chapter 3.

2.2 The Peri/postnatal Epigenetic Twin Study (PETS)

The study presented in Chapter 4 used data collected by the Peri/postnatal Epigenetic Twin Study (PETS). The aim of PETS is to study the plasticity of epigenetics during the intrauterine period and early childhood. PETS cohort is comprised of 250 twin pairs and their mothers recruited midway through the second trimester of pregnancy between January 2007 and September 2009 at three hospitals in Melbourne, Australia. Detailed data on periconceptional and parental factors were collected during pregnancy and in the immediate neonatal period. At recruitment parental age and pre-pregnancy weights and heights were recorded along with information on education, socio-economic status, medical history, mode of conception, nutrition, alcohol intake, and smoking. At 24 weeks gestation pregnancy history detailing maternal weight, illnesses, nutritional supplements, medication, alcohol intake and smoking was collected. A routine maternal blood sampling was done at 28 weeks gestation and food frequency and perceived stress questionnaires were collected. At 36 weeks gestation, if applicable, maternal weight was recorded once again. In the immediate neonatal period, pregnancy history was collected again and infant anthropometric measures were taken. During delivery, or during the immediate neonatal period, multiple biological samples were collected, including umbilical cord, cord blood,

buccal epithelium, and placental tissue. Chorionicity and zygosity were determined from first trimester ultrasound scans and placental examination. Given that MZ twins may not always share placentae, zygosity in same-sex dichorionic twins was determined by a 12-marker microsatellite test.

Of the 250 twin pairs, over 40% are MZ, almost 65% were born by caesarean section, and near a fifth of them were conceived by *in vitro* fertilization. The median gestational age and birth weight were 37 weeks (27-40) and 2.5 kg (0.8-3.9), respectively. Follow-ups at 18 months and 5 years of age were planned to take biological samples, clinical measures, and detailed history of the twins' health and nutrition. More detailed information has been described in cohort profiles [80, 81].

2.2.1 Sample and phenotypes

The study presented in Chapter 4 was based on 41 MZ and 66 DZ twins from the PETS cohort. In total, 47 twins were conceived after *in vitro* fertilization (IVF) as determined from questionnaire data. Mean birth weight was 2.6 kg in both groups, IVF and non-IVF. Mean gestational age was of 37 weeks. Mean maternal age was 35 years in the IVF group and 32 years in the non-IVF group. DNA from whole blood cells was available for 98 of these twins, while DNA from cord blood mononuclear cells was available for 82 twins.

2.2.2 PETS DNA methylation profiling

DNA from whole blood cells and cord blood mononuclear cells was subjected to methylated DNA immunoprecipitation coupled with deep sequencing (MeDIP-seq) for DNA methylation profiling. MeDIP-seq data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001002248.

2.3 Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium

The CHARGE consortium emerged to facilitate collaboration between cohorts in the framework of genomic epidemiology. Subjects of the participating cohorts are not selected for any particular disease, but are not necessarily free of diseases. Cohorts from predominately Caucasian populations with data on menopause and hormone replacement therapy contributed to the study presented in Chapter 5.

1. **The UK Adult Twin Registry (TwinsUK)** is described in section 2.1.
2. **The European Prospective Investigation of Cancer (EPIC)-Norfolk** includes men and women from 35 general practices in Norfolk, UK recruited between 1993 and 1997. The baseline health examination was attended by 25,639 participants aged between 40 and 79 years. A second and third health examinations were carried out in 1997-2000 and 2006-2011, respectively. EPIC-Norfolk has longitudinally characterized the population in terms of lifestyle and physiological, cognitive, metabolic and genetic profiles [82].
3. **The Avon Longitudinal Study of Parents and Children (ALSPAC)** includes mothers and children from more than 14,000 pregnancies with expected delivery between 1 April 1991 and 31 December 1992 in the area of Bristol, UK, and who have been followed-up extensively for more than two decades. The study was conceived with the aim of investigating modifiable influences on child health. The ALSPAC phenotype resource includes data from clinical assessments and questionnaires completed by children, mothers, and teachers. The ALSPAC biobank has collected biological samples including blood, urine, hair, toenails, teeth, saliva, and

placenta since early pregnancy. A subset of the cohort has also “omic” profiles, including genotypes, gene expression, and DNA methylation [83].

4. **Cooperative Health Research in the Augsburg Region (KORA)** was established in 1996 in the region of Augsburg, Germany with the aim of collecting data and biosamples for future research in health sciences. Four baseline cross-sectional health surveys were performed at five-year intervals on individuals sampled at random from the local population. The total pool consists of 18,000 participants that have been followed-up over 4-20 years. Apart from the baseline questionnaire, participants receive follow-up postal questionnaires and attend medical examinations. KORA also contains a Biobank of biological samples, KORA-gen, for genetic research [84].
5. **Rotterdam Study (RS)** was initiated in 1990 to study late onset disease in the increasing proportion of elderly people across different populations. The overall aim was to study cardiovascular, ophthalmic, psychiatric, respiratory, dermatological, oncological, endocrine, and hepatic diseases, and others. After three cycles of recruitment in 1990, 2000, and 2006, the cohort consists of nearly 15,000 participants aged 45 and over from the district of Ommoord in the city of Rotterdam, The Netherlands. Extensive examinations, including collection of biological samples, were carried out at baseline and follow-up examinations were repeated every 3-4 years [85].
6. **The Framingham Heart Study (FHS)** was initiated in 1948 in the town of Framingham, Massachusetts, USA. A total of 5,209 participants aged 30-54 joined the Original Cohort. In 1971, children of the Original Cohort and

their spouses were enrolled in the Offspring Cohort. In 2002, grandchildren of the Original Cohort became the Third Generation Cohort. Currently the overall FHS cohort comprises participants from three generations in multigenerational families. All cohorts are examined periodically every 2-4 years. The study started with a focus on cardiovascular traits and has now expanded to include a wider range of human diseases [86].

2.3.1 Phenotypes

Cohorts were asked to perform analyses on women with natural age at menopause (age at last menstrual period in women who are reportedly postmenopausal or have not had a menstrual period in the last 12 months) between 40 and 60 years inclusive and premenopausal women. The phenotypes of interest were menopausal status, age at menarche, duration of reproductive years, and use of hormone replacement therapy and further details are provided in Chapter 5.

2.3.2 DNA methylation profiling

DNA methylation in each of the participating CHARGE cohorts was interrogated in whole blood samples using the Illumina 450K array. Data was background corrected and normalized according to cohort-specific quality control standards. Further description is provided in Chapter 5.

3. Genetic influences on DNA methylation

3.1 Introduction

The extent to which DNA methylation variation is under genetic control can be explored with the estimation of its heritability. Low to high DNA methylation heritability has been shown at different loci of the genome. It was reported that CpG-sites located at the *IGF2/H19* locus exhibit high heritabilities between 20% and 97% [87]. In contrast, low DNA methylation heritability at the major histocompatibility complex was estimated, between 2 and 16% [88]. To date multiple genome-wide studies using the classical twin design have estimated DNA methylation heritability, reporting that the mean heritability of DNA methylation across tested CpG-sites ranges between 18-20% in adult tissues [37, 45, 89] and between 5-12% in neonatal tissues [90]. Non-twin based efforts to estimate heritability have interrogated the proportion of the variance in DNA methylation at a locus that is explained by common genetic variants. The mean genome-wide heritability explained by common variants across tested CpG-sites was estimated at 3% in brain [91] and 7% in blood [89]. Heritability studies have been helpful in determining the extent to which variation in DNA methylation is influenced by genetic variation, however are not informative of the specific mechanisms underlying genetic influences.

Quantitative trait loci (QTLs) studies have provided an insight into these mechanisms by identifying genetic variants associated with DNA methylation levels. DNA methylation QTLs (meQTLs) have been explored on a genome-wide scale using the Infinium HumanMethlation450 BeadChip, identifying local (*cis*) and distal (*trans*) genetic variants associated with methylation levels in multiple samples across a number of cells, tissues, and ages [37, 42, 43, 45, 46, 92, 93].

In adipose tissue from 648 individuals, 98,085 (28.5%) of the 344,303 CpG sites tested had a significant association at 1% FDR with a genetic variant within a 100 kb radius [45]. In lung tissue from 210 individuals, *cis* meQTLs in a radius of 100 kb were detected for 40,650 (12%) of the 338,456 tested CpG sites controlling for FDR at 5% and *trans* meQTLs acting more than 500 kb away from the target site or in a different chromosome were identified for 615 CpG sites (0.1%) [46]. In 64 Yoruba LCLs, 13,915 (4.2%) of the 329,469 tested CpG sites were associated with at least one genetic variant in *cis* [92]. In purified monocytes, neutrophils, and naïve CD4+ T cells from nearly 200 individuals, an average of 9.89% of the tested CpG sites showed an association with a genetic variant within 1 Mb window [94]. In whole blood from 3,841 individuals, *cis* meQTLs were identified for 34.4% of all tested CpG sites [95].

In this chapter I present a two-fold approach, heritability and meQTL analyses, to study the genetic influences on DNA methylation. DNA methylation was profiled in whole blood samples from healthy adult twins using both the Infinium HumanMethylation450 BeadChip (450K), as well as the new Infinium MethylationEPIC BeadChip (EPIC), a microarray with greater coverage of the methylome than its predecessor [96]. The use of EPIC allows for the investigation of genetic effects on enhancers, which were not highly represented in previous versions of the widely-used Illumina arrays.

3.2 Materials and methods

3.2.1 Subjects

Two datasets were explored in this chapter, 450K and EPIC. The 450K study was based on blood samples from 728 healthy female twins from TwinsUK. The sample consisted of 330 MZ and 34 DZ female twin pairs with a mean age of 58 years. The EPIC study was based on blood samples from 226 healthy female

twins from TwinsUK. The sample consisted of 85 pairs of MZ, 19 pairs of DZ twins, and 18 unrelated twins with a mean age of 60 years (39-80 years). In total, 196 individuals were in both datasets.

3.2.2 DNA methylation data

Whole blood DNA methylation data were generated with the Infinium HumanMethylation450 BeadChip and the Infinium MethylationEPIC BeadChip. In both cases the same quality control checks were applied. Briefly, intensity signals were corrected for probe type bias using the BMIQ normalization method [97]. BMIQ normalizes type II probes into type I keeping the relative ranking of beta values of the type II probes by assigning probes to one of three methylation states and then transforming the probabilities of type II probes of belonging to a state into quantiles using the distribution of type I probes. This procedure reduces technical variation and the bias caused by the lower dynamic range of type II probes. Other normalization methods to remove technical variation from Illumina DNA methylation arrays have been proposed. For example, the functional normalization method combines quantile normalization and the use of control probes as surrogates for batch effects, therefore removing technical variation explained by these covariates. Functional normalization is performed in type I and type II probes separately to address the issue of their different distributions. Using data from technical triplicates, functional normalization showed a modest reduction of technical variability in comparison to BMIQ, however effect sizes in association studies showed to be very similar [98].

In this chapter, methylation levels were reported as beta values calculated as the ratio of the methylated probe intensity signal over the sum of the methylated and unmethylated probe intensity signals. CpG sites with a detection p-value > 0.01 were discarded. Only CpG sites located in autosomes were considered for

the analysis. Probes in the array that mapped to multiple locations of the *in silico* bisulfite converted human genome allowing two mismatches were discarded. Probes were also excluded if they targeted polymorphic CpG sites with minor allele frequency (MAF) > 5% in the UK10K haplotype reference panel [99]. A total of 420,230 and 771,169 CpG sites remained for analysis in the 450K and EPIC arrays, respectively. In both datasets white blood cells proportions were estimated from DNA methylation data for plasmablasts, $CD8^{+}CD28^{-}CD45RA^{-}$ T cells, naïve $CD8^{+}$ T cells, $CD4^{+}$ T, natural killers, monocytes, and granulocytes [100].

3.2.3 Genotype data

TwinsUK participants were genotyped with one of two chips, Illumina HumanHap300 BeadChip or Illumina HumanHap610 QuadChip. Detailed quality control and imputation of the genotype data in the larger TwinsUK cohort have been previously described [101, 102]. In this thesis, genotype data from TwinsUK included directly genotyped and imputed genetic variants, where imputed variants were based on the 1000 Genomes project reference panel, as previously described [101]. A total of 6,266,036 variants with a MAF > 5% and IMPUTE info score > 0.7 were used for QTL analysis. Genotypes were duplicated for the co-twin when only a single twin from a MZ pair was genotyped. In this thesis, QTL analysis was performed in the EPIC dataset of 226 individuals. QTL analysis in the 450K dataset was previously performed by other members of the epigenetics research group at the Department of Twin Research, King's College London.

3.2.4 Annotation datasets

CpG islands and their corresponding shores and shelves were obtained from the manifest provided by Illumina (https://support.illumina.com/array/kits.html). CpG sites not located in any of these regions were considered as mapping to “open sea”. Histone peaks and predicted chromatin states (promoters and enhancers) of PBMCs were obtained from the Roadmap Epigenomics project [103]. Transcription factor peaks of GM12878 were obtained from the ENCODE project [104]. PBMCs and GM12878 were selected for being the closest tissues to whole blood available in those datasets.

3.2.5 Estimation of DNA methylation heritability

DNA methylation heritability at each CpG site was estimated in both 450K and EPIC datasets, following the same procedure. Methylation beta values at each CpG site were transformed to standard normal form N(0,1) prior to analysis. Methylation beta values were adjusted for age, cell proportions and technical covariates (plate and position). The observed methylation variance was partitioned into additive genetic (A), common environmental (C), and unique environmental (E) factors using the ACE model [105]. The model was fitted using the OpenMX statistical package [106] in R [107]. In terms of A, C, and E, heritability (h^2) was defined as $h^2 = \frac{A}{A+C+E}$.

3.2.6 Estimation of meQTLs

MeQTLs were estimated in the EPIC dataset of 226 individuals. Methylation beta values at each CpG site were transformed to standard normal form N(0,1) prior to analysis. Methylation beta values were adjusted for age, cell proportions, family structure, and technical covariates (plate and position). The associations

between genetic variants and corrected methylation values were then computed using MatrixEQTL [108] in R [107] as follows:

$$y = \alpha + \beta\ G$$

where y is methylation, α is the y-intercept, β is the slope of the regression line, and G is the additive genotype.

Local (*cis*) effects were investigated if SNPs were within 1Mb from the target CpG site. Distal (*trans*) effects were investigated if the SNP was located more than 1Mb away from the target CpG site, or on a different chromosome. To correct for multiple testing, results were reported using a permutation-based false discovery rate (FDR) correction. FDR was determined by dividing the average number of significant hits in 10 permutations (keeping twin structure) over the number of significant hits in the observed data at a given p-value threshold. Results are presented at a FDR 5% threshold.

3.2.7 Blood cell-type specific meQTLs

Methylation beta values at each CpG site were transformed to standard normal form N(0,1) prior to analysis. Methylation beta values were adjusted for age, family structure, and technical covariates (plate and position). To detect cell-type specific associations, a linear regression with an interaction term between genotype and cell-type proportion was added as follows:

$$y = \alpha + \beta G + \gamma E + \delta GxE$$

where y is methylation, α is the y-intercept, β, γ, δ are slopes, G is the additive genotype, E is the cell proportion, and GxE is an interaction term. The model test for significance of δ. This model has been used previously for the detection of cell-type specific effects in gene expression [109]. To correct for multiple testing, a permutation-based FDR was estimated as described in section 3.2.6 above.

3.3 Results

3.3.1 DNA methylation profiles in twins

We first characterized the patterns of methylation within pairs of twins and pairs of unrelated individuals. Intra-class correlation coefficients (ICCs) were estimated within pairs of MZ, DZ and unrelated twin pairs using the EPIC array, which has a larger coverage of the genome. Higher mean ICC was observed in MZ twins than in DZ twins ($p=1.96x10^{-6}$) and the latter showed higher mean ICC than unrelated individuals (p=0.038) (**Figure 3-1**). At the CpG site level, MZ twin pairs also showed greater ICCs than pairs of DZ or unrelated twins (**Figure 3-2**).

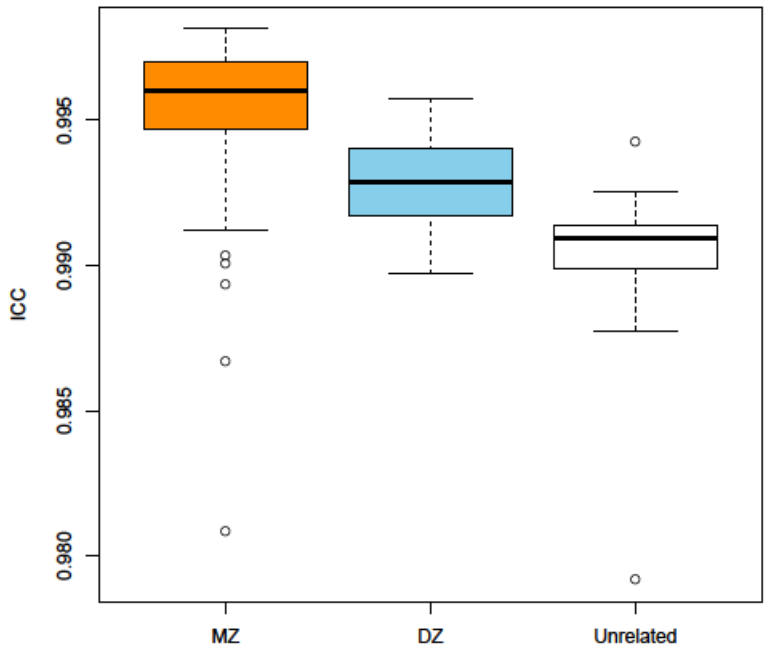


Figure 3-1. DNA methylation profile similarities in twins. The intra-class correlation coefficient (ICC) was estimated within MZ (n=85), DZ (n=19), and unrelated (n=9) twin pairs.

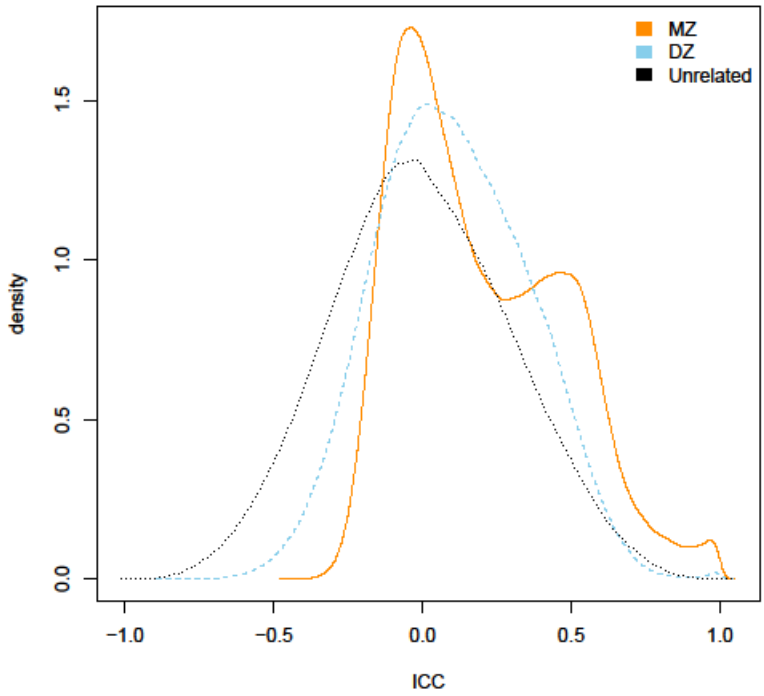


Figure 3-2. Correlation of CpG site methylation in twins in the EPIC dataset. Density plot of CpG site ICCs in MZ (n=85), DZ (n=19), and unrelated (n=9) twin pairs.

3.3.2 Heritability of DNA methylation

DNA methylation heritability was estimated in the two datasets of healthy twins, 450K (728 twins) and EPIC (226 twins). In the 450K dataset, the genome-wide mean heritability (h^2) across 420,230 autosomal CpG sites was estimated at 12%. Altogether, 43,993 CpG sites (10%) had evidence for strong heritability (h^2 > 40%).

The classification of CpG sites by functional regions of the genome showed that sites within enhancers exhibit greater mean heritability than sites within promoters (**Figure 3-3**, **Table 3-1**). The classification by CpG content showed that sites within CpG islands have lower mean heritability than sites within shores, shelves, and the open sea (**Figure 3-4**, **Table 3-1**). Comparing our results with ChIP-seq experiments, we observed that heritability at CpG sites that overlap TFBS peaks is lower than the mean overall heritability and CpG sites

that overlap with two repressive histone marks, H3K27me3 and H3K9me3, have higher mean heritability than the overall (**Table 3-1**).

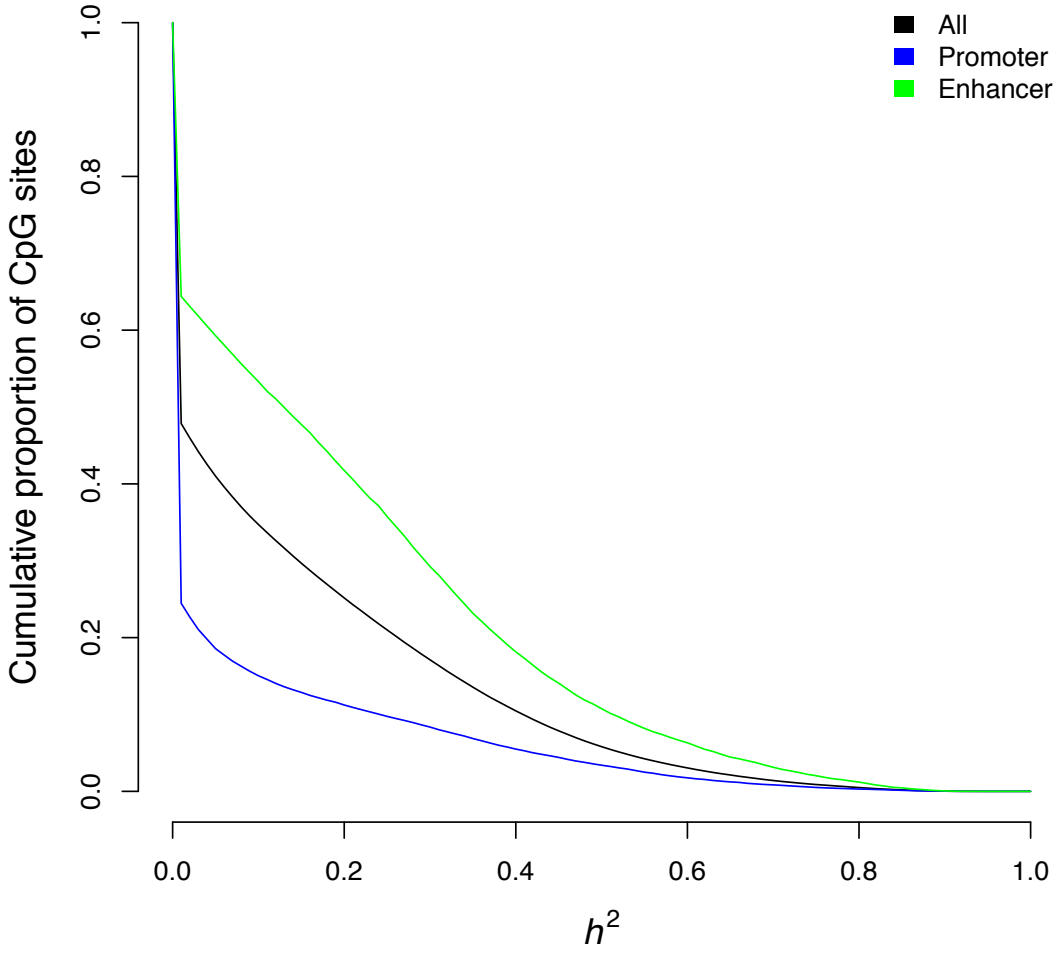


Figure 3-3. Genetic influence on DNA methylation in 450K dataset by functional genomic region. Plot shows the proportion of CpG sites within promoters or enhancers that showed heritability greater or equal to h^2.

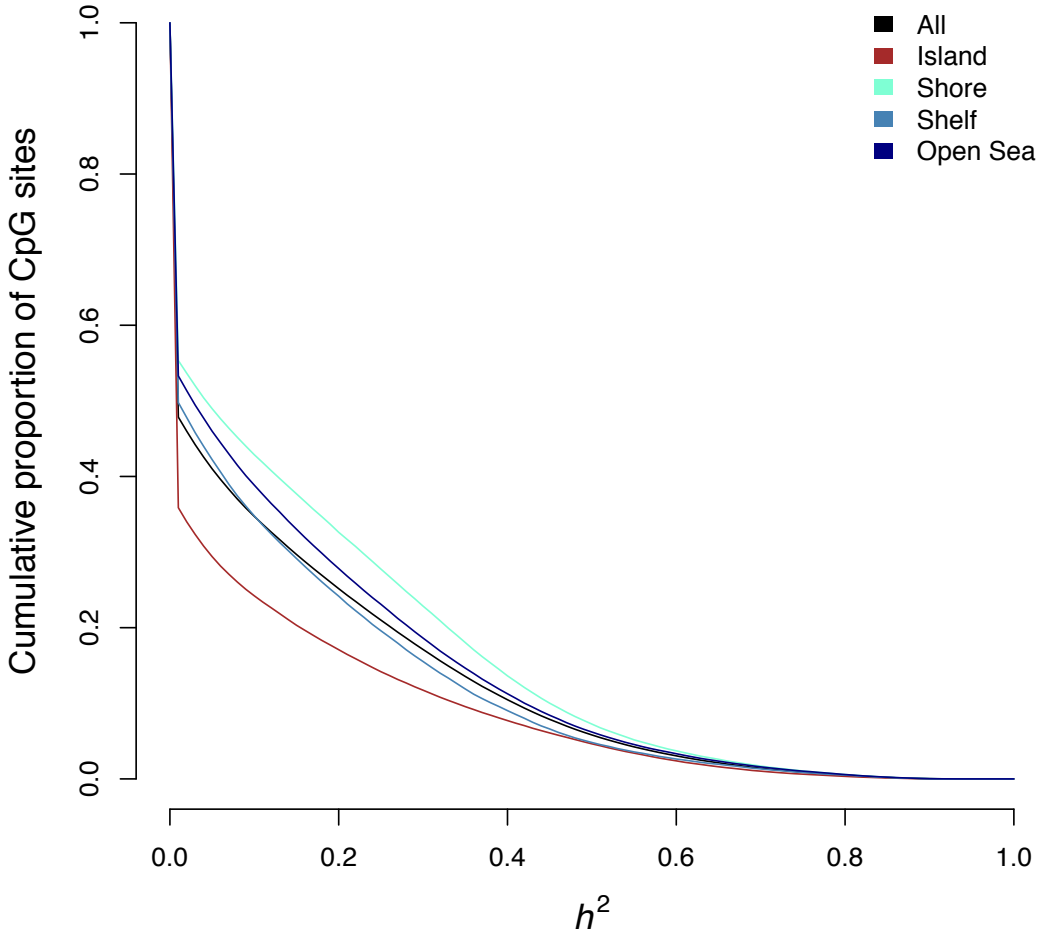


Figure 3-4. Genetic influence on DNA methylation in 450K dataset by CpG content. Plot shows the proportion of CpG sites within CpG islands, shores, shelves, or open sea that showed heritability greater or equal to h^2.

In the EPIC dataset, results showed the same patterns seen with the 450K dataset. The genome-wide mean heritability (h^2) across 771,169 autosomal CpG sites was estimated at 12%. Altogether, 79,408 CpG sites (10%) had evidence for strong heritability (h^2 > 40%). As observed for the 450K results, the classification of CpG sites by functional regions of the genome showed that sites within enhancers exhibit greater mean heritability than sites within promoters (**Figure 3-5**, **Table 3-1**). Classification by CpG content showed that sites within

CpG islands have lower mean heritability than sites within shores, shelves, and the open sea (**Figure 3-6**, **Table 3-1**). Heritability at CpG sites that overlap TFBS peaks is lower than the mean overall heritability and CpG sites that overlap with two repressive histone marks, H3K27me3 and H3K9me3, have higher mean heritability than the overall (**Table 3-1**).

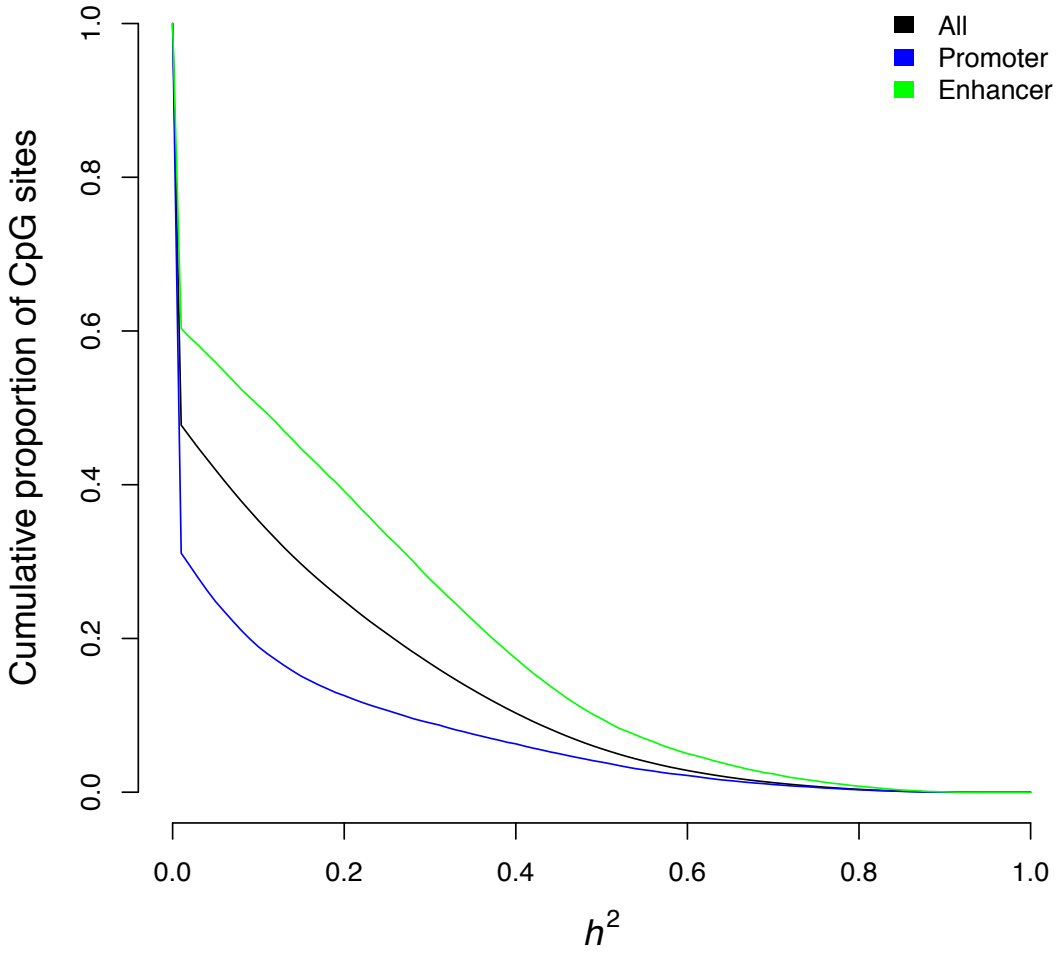


Figure 3-5. Genetic influence on DNA methylation in EPIC dataset by functional genomic region. Plot shows the proportion of CpG sites within promoters or enhancers that showed heritability greater or equal to h^2.

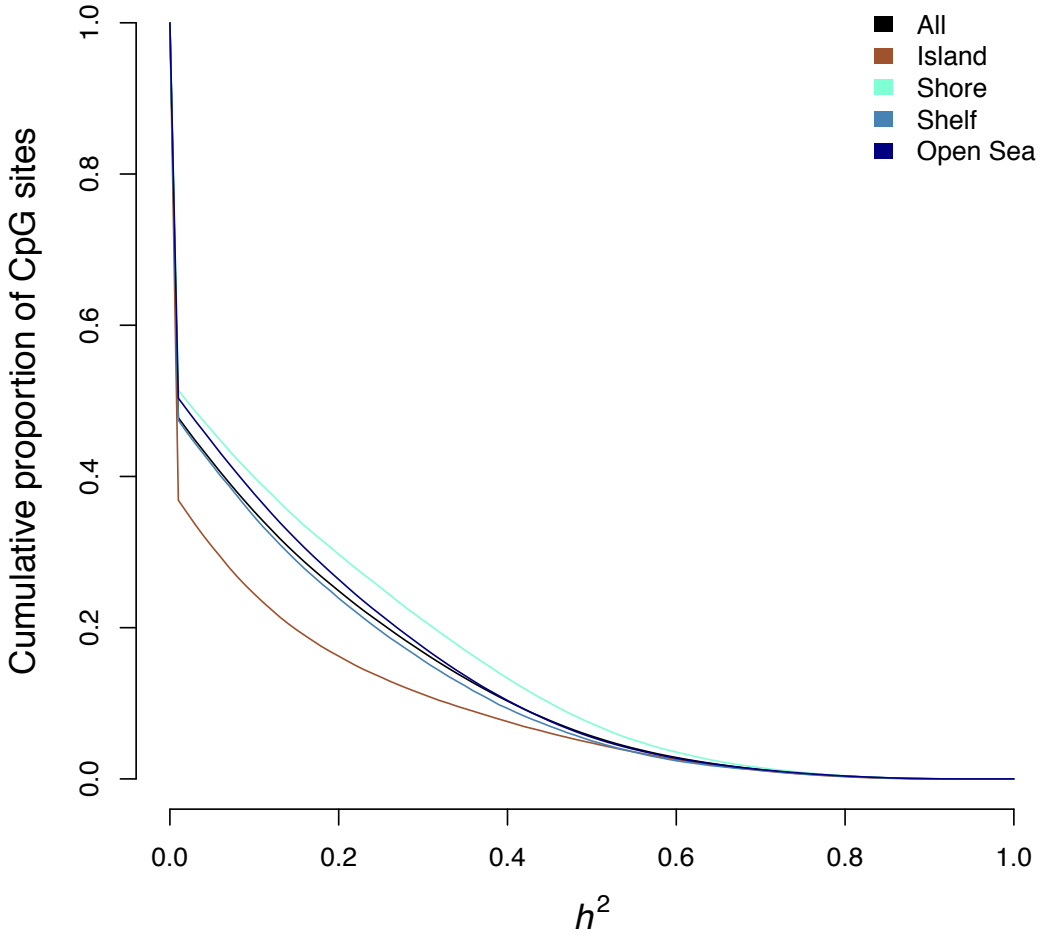


Figure 3-6. Genetic influence on DNA methylation in EPIC dataset by CpG content. Plot shows the proportion of CpG sites within CpG islands, shores, shelves, or open sea that showed heritability greater or equal to h^2.

Table 3-1. Mean heritability of DNA methylation

	EPIC		450K	
Category	**#CpG sites**	**Mean h^2**	**#CpG sites**	**Mean h^2**
All	771,169	0.12	420,230	0.12
Promoter (PBMCs)	39,328	0.07	32,366	0.06
Enhancer (PBMCs)	30,057	0.18	14,842	0.20
Island	147,597	0.09	135,010	0.09
Shore	139,787	0.14	98,751	0.15
Shelf	53,402	0.12	38,925	0.12
Open Sea	430,383	0.13	147,544	0.13
TFBS (GM12878)	157,963	0.09	111,675	0.07
H3K27ac (PBMCs)	598,723	0.12	380,275	0.12
H3K27me3 (PBMCs)	408,331	0.14	262,832	0.15
H3K36me3 (PBMCs)	370,544	0.12	236,421	0.12
H3K4me1 (PBMCs)	601,044	0.12	379,277	0.13
H3K4me3 (PBMCs)	486,043	0.12	331,910	0.12
H3K9ac (PBMCs)	557,277	0.12	359,922	0.12
H3K9me3 (PBMCs)	141,923	0.14	90,784	0.15
New in EPIC	369,415	0.13		
Overlap with 450K	401,754	0.12		

The results between the 450K and EPIC datasets were very consistent and showed the same trend in every category annotated. The correlation of the heritability estimates between the two was of 0.88.

3.3.3 Identification of methylation quantitative trait loci (meQTLs)

Genetic effects were further explored by the identification of meQTLs in the EPIC dataset of 226 twins. In total, 171,156 CpG sites (22%) were found to be affected by *cis* meQTLs and 2,021 CpG sites (0.2%) were found to be affected by *trans* meQTLs at 5% FDR ($P=3.6x10^{-4}$ and $P=1.1x10^{-8}$, respectively). As expected, the median heritability of DNA methylation at CpG sites affected by meQTLs, both

cis and *trans,* was greater than that of the overall (**Figure 3-7**). Altogether, we identified common genetic variants to explain the genetic influence on 38,755 (49%) of the CpG sites with strong DNA methylation heritability (h^2 > 0.4).

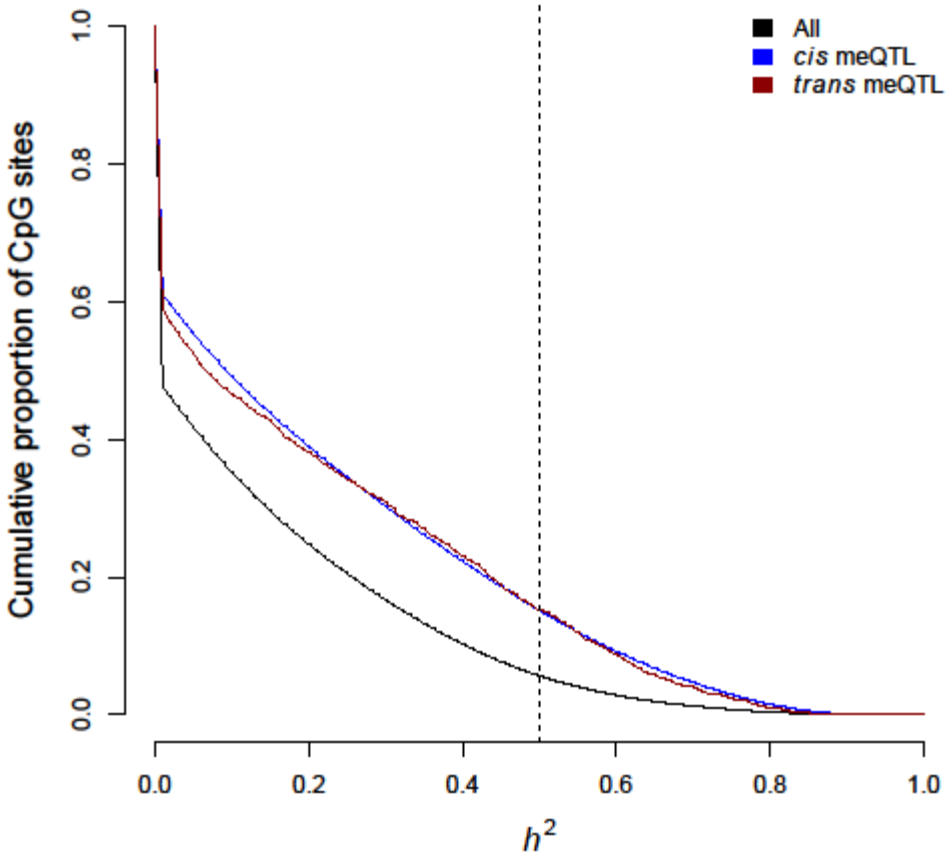


Figure 3-7. Heritability of DNA methylation at CpG sites affected by meQTLs. Proportion of CpG sites with heritability of DNA methylation > h^2 in all analysed CpG sites, CpG sites affected by *cis* meQTLs, and CpG sites affected by *trans* meQTLs.

3.3.4 Annotation of meQTL CpG sites

We then explored where the CpG sites affected by *cis* and *trans* meQTLs were located in the genome. Overall, we observed opposite trends in CpG sites affected by *cis* meQTLs and *trans* meQTLs. CpG sites affected by *cis* meQTLs were depleted in TFBSs, promoters, CpG islands, regions of positive selection, and conserved elements (**Table 3-2**, Fisher's exact test). In contrast, CpG sites with *cis* meQTLs were enriched in shores, shelves, open sea, enhancers, gene bodies and intergenic regions. Concordantly, CpG sites affected by *cis* meQTLs were also depleted in H3K4me3 peaks, which are associated with promoters, and enriched in H3K4me1 peaks, which are associated with enhancer regions [110] . Conversely, CpG sites affected by *trans* meQTLs were enriched in TFBSs, promoters, H3K4me3 peaks, and CpG islands. CpGs with *trans* meQTLs were depleted in shores, shelves, open sea, enhancers, and H3K4me1 peaks. Interestingly, both CpG sites affected by *cis* meQTLs and CpG sites affected by *trans* meQTLs were depleted in exon boundaries. To account for the fact that the methylation variance affects the probability of harboring meQTLs, enrichment analyses were also performed by matching to sets of CpGs with similar variance. Conclusions remained unchanged (Annex I).

3.3.5 Properties of meQTL SNPs

The peak abundance of SNPs that are *cis* meQTLs was found in close proximity to the target CpG site (**Figure 3-8**) and the effect sizes tended to be small with a maximum peak at R^2 = 0.06 (**Figure 3-9**). SNPs were found to be distributed across all the MAF range (**Figure 3-10**). It is important to highlight that in this section, only the most significant association per CpG sites was considered, which is not necessarily the causal genetic variant.

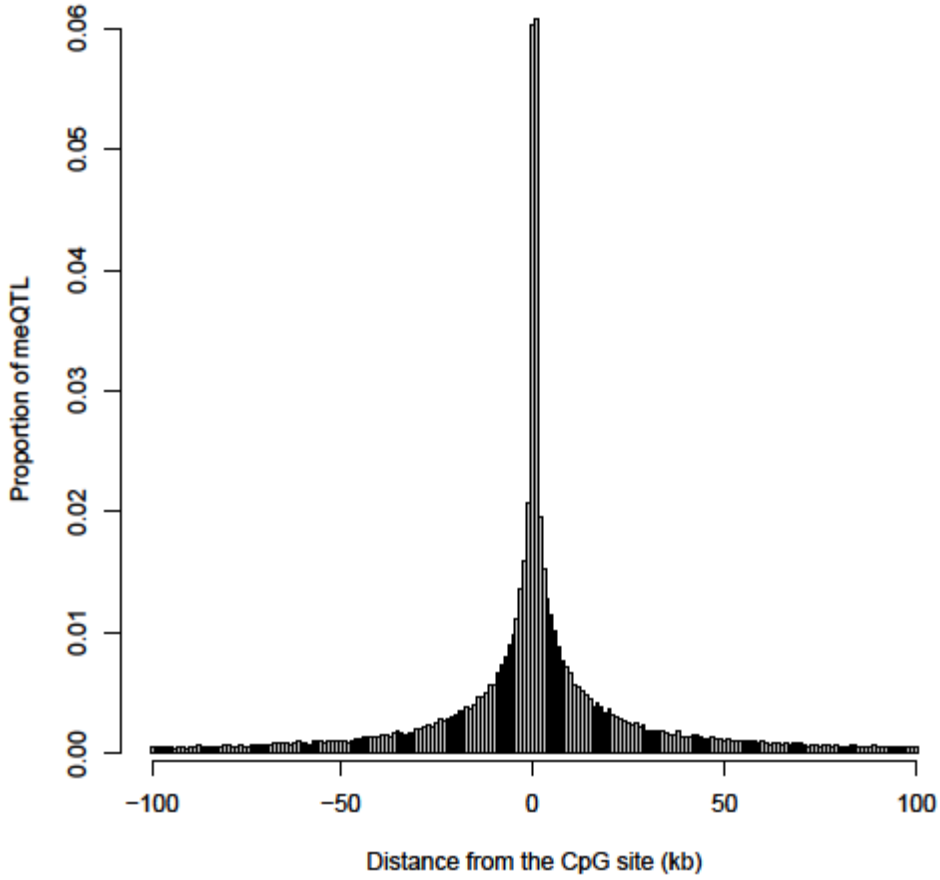


Figure 3-8. Distance of *cis* meQTL SNP to target CpG site. Histogram showing the proportion of CpG sites with a meQTL acting at a given distance. Only the most significantly associated SNP per CpG site is considered in this plot.

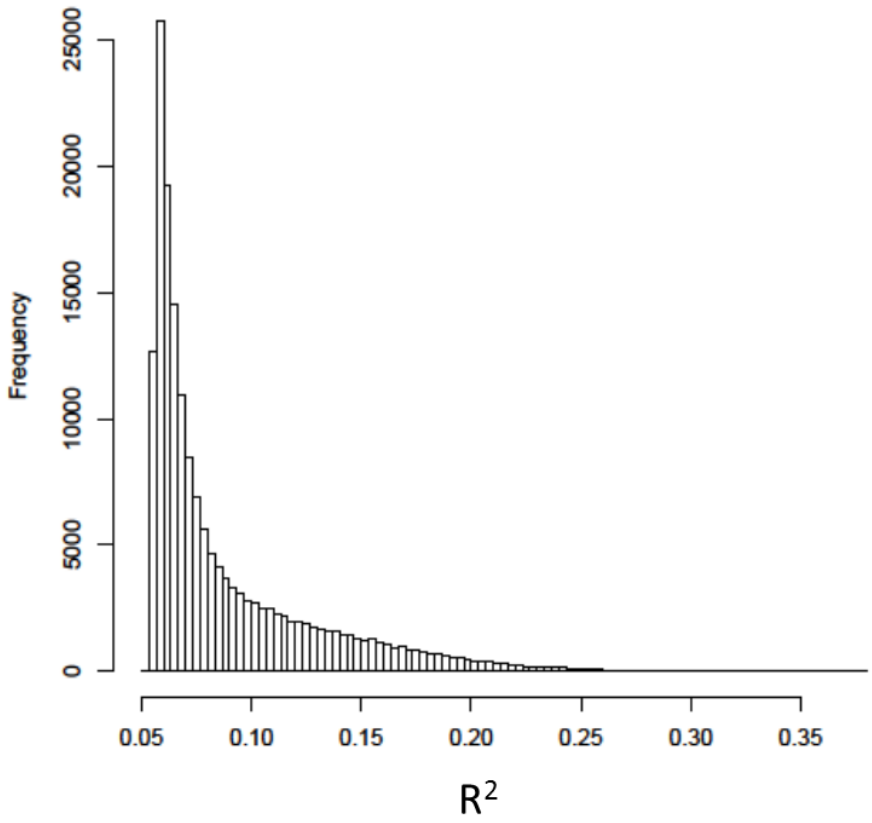


Figure 3-9. Proportion of the variance explained by *cis* meQTLs. Histogram of R-squared showing only the most significant association per CpG site.

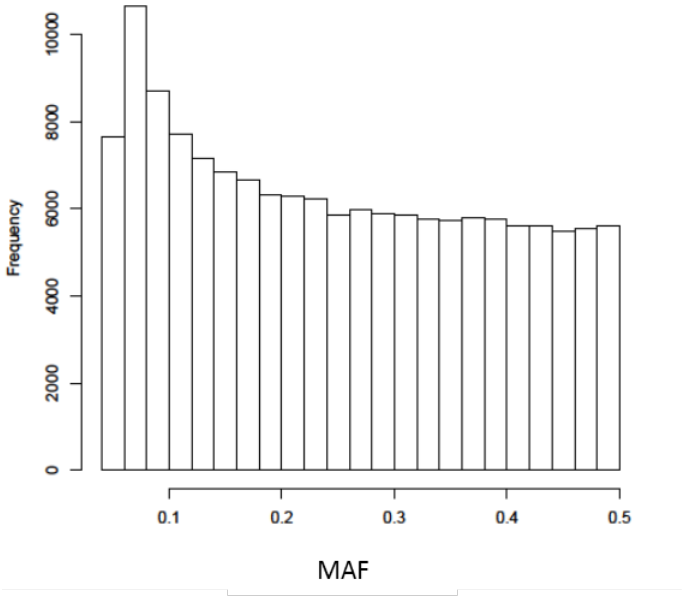


Figure 3-10. MAF distribution of *cis* meQTLs. Histogram of MAFs showing only the most significant association per CpG site.

3.3.6 Blood meQTLs are likely to be shared across blood cell-subtypes

Since the analyses performed so far were in whole blood samples, we further sought to identify blood cell-type specific meQTLs. We performed analyses testing if DNA methylation levels at each CpG site were significantly associated with the interaction between genotype and proportion of each $CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killers, monocytes, and granulocytes. We did not identify any genome-wide significant meQTL for any of the tested interactions at 5% FDR.

Table 3-2. Enrichment and depletion of CpG sites affected by *cis* meQTLs

	All		*cis* meQTLs				*trans* meQTL			
	# CpG sites	Proportion	# CpG sites	Proportion	Odds ratio	P-value	# CpG sites	Proportion	Odds ratio	P-value
TFBS (GM12878)	157,963	0.205	30,374	0.177	▼ 0.799	1.10E-227	630	0.312	▲ 1.761	1.14E-29
H3K27ac (PBMC)	598,723	0.776	133,927	0.782	▲ 1.047	5.87E-12	1,574	0.779	▲ 1.014	8.10E-01
H3K27me3 (PBMC)	408,331	0.529	94,607	0.553	▲ 1.128	4.97E-106	972	0.481	▼ 0.823	1.34E-05
H3K36me3 (PBMC)	370,544	0.480	82,961	0.485	▲ 1.022	7.67E-05	939	0.465	▼ 0.938	1.54E-01
H3K4me1 (PBMC)	601,044	0.779	136,953	0.800	▲ 1.173	4.30E-124	1,506	0.745	▼ 0.827	2.57E-04
H3K4me3 (PBMC)	486,043	0.630	106,098	0.620	▼ 0.945	7.66E-24	1,393	0.689	▲ 1.302	2.96E-08
H3K9ac (PBMC)	557,277	0.723	124,200	0.726	▲ 1.020	1.59E-03	1,534	0.759	▲ 1.210	2.30E-04
H3K9me3 (PBMC)	141,923	0.184	30,848	0.180	▼ 0.968	4.08E-06	386	0.191	▲ 1.047	4.21E-01
Island	147,597	0.191	25,928	0.151	▼ 0.702	0.00E+00	673	0.333	▲ 2.114	2.60E-51
Shore	139,787	0.181	33,104	0.193	▲ 1.109	6.29E-49	325	0.161	▼ 0.865	1.64E-02
Shelf	53,402	0.069	12,190	0.071	▲ 1.040	2.81E-04	114	0.056	▼ 0.803	2.24E-02
Open sea	430,383	0.558	99,934	0.584	▲ 1.145	2.12E-131	909	0.450	▼ 0.647	1.52E-22
Promoter (PBMCs)	39,328	0.051	6,758	0.039	▼ 0.716	1.20E-140	208	0.103	▲ 2.141	4.88E-21
Enhancer (PBMCs)	30,057	0.039	8,523	0.050	▲ 1.408	2.28E-143	52	0.026	▼ 0.651	1.51E-03
TSS1500	114,681	0.149	24,823	0.145	▼ 0.963	1.17E-06	309	0.153	▲ 1.033	5.95E-01
TSS200	73,929	0.096	13,284	0.078	▼ 0.748	3.36E-194	355	0.176	▲ 2.015	1.79E-28
5' UTR	101,401	0.131	20,622	0.120	▼ 0.881	1.58E-53	311	0.154	▲ 1.202	3.34E-03
1st Exon	42,877	0.056	6,963	0.041	▼ 0.666	4.57E-219	204	0.101	▲ 1.911	7.96E-16
Exon boundary	6,594	0.009	1,235	0.007	▼ 0.806	4.37E-12	8	0.004	▼ 0.460	2.08E-02
Gene body	325,070	0.422	72,997	0.426	▲ 1.026	2.43E-06	716	0.354	▼ 0.752	6.99E-10
3' UTR	21,658	0.028	4,618	0.027	▼ 0.949	1.72E-03	47	0.023	▼ 0.824	2.00E-01
Intergenic	214,281	0.278	51,269	0.300	▲ 1.146	7.55E-113	514	0.254	▼ 0.886	1.82E-02
Positive selection	7,418	0.010	1,346	0.008	▼ 0.775	7.03E-18	20	0.010	▲ 1.029	8.20E-01
Conserved elements	148,701	0.193	27,082	0.158	▼ 0.739	0.00E+00	403	0.199	▲ 1.043	4.46E-01

3.4 Discussion

We know from previous studies that DNA methylation at specific loci in the genome can be influenced by genetic effects. The aims of this study were to characterize DNA methylation heritability at different functional regions of the genome using the classical twin study and to identify the genetic variants contributing to DNA methylation variability both by replicating published results on the 450K array and by using the new EPIC array to identify novel effects.

The patterns of heritability observed in this study were in line with a published study conducted in whole blood [89], where they observed that highly heritable sites ($h^2 \geq 0.5$) showed an enrichment in CpG island shores and depletion at CpG islands. One of the key novel results from this chapter was the assessment of DNA methylation heritability on a genome-wide scale using the new EPIC array, which targets enhancers. Strong heritability ($h^2 > 0.4$) was observed for at least 10% of the CpG sites in the array, and for almost half of those sites we were able to identify common genetic variants influencing their methylation levels.

The strongest genetic effects on DNA methylation were observed at CpG sites away from transcription start sites and outside of the CpG islands. These results suggest that DNA methylation variation at promoters is mostly due to non-genetic effects, which may include environmental or stochastic factors. It is important to note that these low heritability estimates observed mainly at promoters may arise from the lack of actual variation. These CpG sites exhibit very low methylation values and the observed variation might be only the result of measurement error. Others have addressed this issue by looking only at the top 10% variable CpG sites interrogated by the array, which limits the analysis to a reduced subset of CpG sites [45].

Further annotation of CpG sites using predicted chromatin states of PMCs showed that enhancers are the regions of the genome where genetics effects are stronger. Conversely, TFBSs are regions where weak genetic effects are observed.

The second main novel result in this chapter was the identification of common genetic variants that influence DNA methylation levels at CpG sites profiled on the EPIC array. By looking at local and distal effects of genetic variants through QTL analysis we were able to observe that the affected CpG sites are located in very different genomic regions, which might indicate different mechanisms of gene regulation exerted by genetic variation. It would be interesting as well to investigate if *cis* and *trans* meQTLs fall within particular regions of the genome. From work done by Dr. Elena Carnero-Montoro (manuscript in preparation) characterizing meQTLs identified with the 450K dataset, we know that *cis* meQTLs are enriched in TFBSs, while *trans* meQTLs are depleted. It has been planned to extend the approach used on the 450K meQTLs to characterize novel EPIC meQTL SNPs identified in this chapter.

In this study we looked at whole blood and used estimated cell proportions to investigate if this could mediate an association with genotype. We did not identify blood cell-type specific meQTLs, which may suggest a lack of power [109], but also suggests that cell-specific genetic effects may not be common as shown in a recent study looking at almost 200 samples of purified monocytes, neutrophils, and naïve $CD4^+$ T cells that reported less than 0.1% of CpG sites having a cell-type specific QTL [94]. In this same study, cell-type specific histone and expression QTLs were found to be more abundant. In addition, one study following a similar approach using whole blood and a proxy for cell counts also showed evidence of cell-type specific expression QTLs [109]. These results

suggest that gene regulation mechanisms of cell-type specific QTLs may not be mediated by DNA methylation to a great extent.

The main strength of this study was the use of the new EPIC array that allows the exploration of almost twice the number of CpG sites compared to its predecessor the 450K array. However, the number of samples profiled with this platform at the moment is a limitation. Currently, more samples are being profiled in TwinsUK with the idea of combining them with these 226 samples and extending the analyses to include the functional annotation of SNPs that are meQTLs.

4. The effect of *in vitro* fertilisation on DNA methylation profiles at birth

4.1 Introduction

As the frequency of *in vitro* fertilisation (IVF) treatment increases worldwide, much research effort has focused on exploring both short and long-term health outcomes associated with conception via IVF, with contradictory results. A number of studies have observed associations with adverse perinatal and obstetric outcomes including low birth weight, preterm birth, perinatal mortality, congenital malformations, placental complications, and increased frequency of imprinting disorders such as Angelman syndrome and Beckwith-Wiedemann syndrome [111–114]. On the other hand, parallel efforts have reported that these associations are not attributed to IVF treatment itself, but rather to multiple pregnancy or parental subfertility, both common factors in IVF births [115, 116]. Further research is required to identify potential factors associated with conception via IVF, including not only health outcomes, but also biological consequences such as epigenetic modifications.

Given that birth weight and imprinting disorders are controlled at least in part by epigenetic factors [90, 117], IVF may have an influence on epigenetic profiles, potentially resulting in changes that persist well after birth and over the life course. Epigenetic mechanisms are considered possible mediators of the developmental origins of health and disease [118], therefore an assessment of the influence of IVF on DNA methylation profiles may give some insights into mechanisms underlying potential related health outcomes. Establishment of DNA methylation profiles in the germ line and embryo takes place early in development [119]. Theoretically, this epigenetic reprogramming could therefore be influenced by IVF-related interventions that occur very early, prior to

blastocyst implantation. Indeed, induction of ovulation, embryo culturing, and cryopreservation, among others, have all been linked to specific alterations in DNA methylation in mice, although results are somewhat inconsistent [120–122].

Most studies in humans comparing naturally- and IVF-conceived newborns have interrogated DNA methylation alterations targeting almost exclusively imprinted Differentially Methylated Regions (DMRs). These studies have reported hypomethylation at the *KvDMR1* DMR in peripheral blood and cord blood of IVF children [123, 124]. Reduced methylation was also observed at the *H19* DMR in placentas and buccal epithelium of IVF newborns [80, 125]. At the *MEST* DMR, hypermethylation was observed in maternal peripheral blood and umbilical cord blood of IVF newborns [126]. However, a different study reported hypomethylation at the *MEST* DMR in placentas of IVF cases [125].

High-throughput approaches using bead array technology have also interrogated DNA methylation in IVF in a genome-wide manner. Katari *et al.,* 2009, reported differential methylation at 78 genes in cord blood and 40 in placenta with at least two differentially methylated CpG sites ($P \leq 0.08$) when looking across the promoters of 736 genes (GoldenGate Array, Illumina) in 10 cases and 13 controls [127]. A more extensive study using the promoter-enriched Illumina Infinium HumanMethylation27 bead array in cord blood samples from 10 IVF cases and 8 controls, reported a total of 24 genes with at least two differentially methylated CpG sites ($P<0.05$) [128]. More recently, a study used the genome-wide Illumina Infinium HumanMethylation450 bead array in samples from 38 IVF-conceived newborns followed by fresh embryo transfer, 38 IVF-conceived followed by cryopreserved embryo transfer, 18 born to subfertile parents after conception by intrauterine insemination, and 43 controls born to fertile parents [129]. This platform interrogates CpG sites across the whole

genome although with a limited coverage since it targets gene-centric annotations [130]. The authors identified differential methylation at multiple sites, including metastable epialleles.

Here, we interrogated evidence for differential methylation between IVF and non-IVF newborn twins in a more comprehensive manner by conducting epigenome-wide association scans (EWAS) [68] using methylated DNA immunoprecipitation followed by deep sequencing (MeDIP-seq) [131] genome-wide in samples from cord blood, and its mononuclear fraction, collected at birth from IVF and non-IVF twins. The use of twins in this study allowed the partition of the observed variance in DNA methylation into genetic and environmental factors. The approach also avoids potential spurious associations due to an imbalanced number of multiple and single pregnancies between conception method groups.

4.2 Methods

4.2.1 Subjects and sample collection

The study included 47 IVF and 60 non-IVF newborn twins (from 54 twin pairs) from the Peri/postnatal Epigenetic Twins Study (PETS), Melbourne, Australia. Recruitment and full study procedure has been described previously [81, 132]. Cord blood was collected at birth and used to process mononuclear cells by Ficoll gradient centrifugation as described previously [133]. Whole blood cells (WBCs) from cord blood were available for a total of 98 twins (40 IVF and 58 non-IVF) and cord blood mononuclear cells (CBMCs) for a total of 82 twins (35 IVF and 47 non-IVF). Maternal age and method of conception were determined via questionnaire at recruitment (18-20 weeks gestation). Twins of mothers who said yes to IVF or ICSI treatment were classified as IVF regardless of the use of ovulation induction medication or other fertility treatments. Maternal smoking

status was collected via questionnaire on recruitment and at 24 and 36 weeks of pregnancy. Birth weight was collected during the immediate neonatal period. Zygosity and chorionicity were determined by physical examination of the inter-placental membranes at birth, and by genetic test when required, as described previously [81, 132]. For a subset of the whole blood samples (n=54, 22 IVF and 32 non-IVF), blood cell subtype counts were obtained through automatic differential counting. Pregnancy complications were recorded and are shown in **Table 4-1**.

Table 4-1. Pregnancy complications

	Occurrence (N)	
Pregnancy complications	**non-IVF**	**IVF**
Gestational diabetes	0	1
Vacuum-assisted vaginal delivery	0	1
Intrauterine growth restriction	1	2
Shortened cervix	1	1
Elevated blood pressure	0	2
Twin-to-twin transfusion syndrome	1	0
Hyperemesis	1	0

4.2.2 DNA methylation profiling

MeDIP-seq was performed at BGI-Shenzhen, Shenzhen, China. Extracted DNA was fragmented using a Covaris sonication system and sequencing libraries were prepared from 5μg fragmented genomic DNA. End repair, <A> base addition and adaptor ligation steps were performed using Illumina's Single-End DNA Sample Prep kit. Adaptor-ligated DNA was immunoprecipitated by anti-5mC using a commercial antibody (Diagenode) and MeDIP products were validated by quantitative PCR. MeDIP DNA was purified with ZYMO DNA Clean & Concentrator-5 columns and amplified using adaptor-mediated PCR. DNA fragments between 200 and 500 bp in size were gel-excised, and the amplification quality and quantity were evaluated by Agilent BioAnalyzer analysis. The libraries were subjected to highly parallel 50-bp single-end

sequencing on the Illumina GAII platform. All sequencing data passed initial quality checks for base composition (no exclusions) using FASTQC v0.10.0. For each individual, ~30 million reads were generated and mapped onto hg19 using BWA. After removing duplicates, we filtered data using quality score Q10. We quantified methylation levels using MEDIPS [134] producing the mean relative methylation score (RPM) in 500-bp bins (overlap of 250 bp) across the genome. Altogether, there were 11,524,145 windows and these were used for the analyses.

4.2.3 Epigenome-wide IVF-DMR analyses

Data were analyzed by subsets looking at each tissue type independently and looking at all samples together in order to identify tissue-independent and tissue-shared IVF-associated differentially methylated regions (IVF-DMRs) as shown in **Figure 4-1**. Methylation bins with RPM values equal to zero in more than 50% of the individuals were excluded resulting in 9,592,803 and 9,285,089 bins to be analyzed in WBCs and CBMCs, respectively. Transformed to standard normal form (N(0,1)) methylation scores in each genomic bin were regressed using a linear mixed-effects model to account for twin structure (lme4 package [135] in R [107]). Tissue type, birth weight, sex, maternal smoking, DNA purity, DNA concentration, and the loadings of the first five principal components were used as covariates and included as fixed effects in the model. Family and zygosity were included as random effects. The principal components were included to account for unknown sources of variation, such as cell heterogeneity. Correction for multiple testing was performed by a Benjamini-Hochberg FDR calculation.

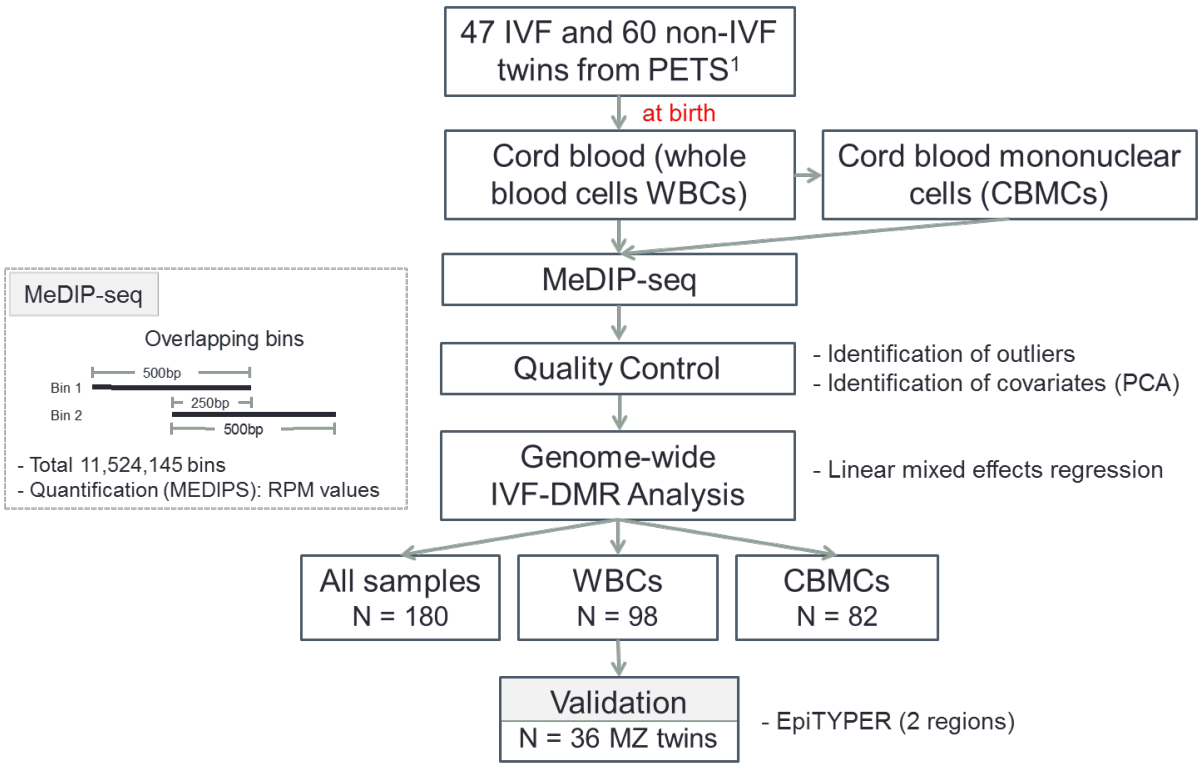


Figure 4-1. Analysis flow chart. DNA from whole blood cells (WBCs) and cord blood mononuclear cells (CBMCs) of IVF and non-IVF twins was subjected to MeDIP-seq. Relative methylation scores (RPM) were calculated in 500 bp bins. Principal component analysis (PCA) was used to identify covariates and global patterns of methylation. IVF-DMR analyses were conducted in all samples available and in WBCs and CBMCs separately. Validation of two IVF-associated differentially methylated regions (IVF-DMRs) was performed with EpiTYPER in monozygotic (MZ) twins.

4.2.4 Variance decomposition of WBC IVF-DMRs

The contribution of additive genetic (A), common environmental (C), and unique environmental (E) factors to DNA methylation was estimated using the ACE model based on the classical twin design [105]. The model was fitted using the OpenMX statistical package [106]. RPM values without adjustment for covariates were used to estimate the ACE proportions.

4.2.5 Statistical analysis

Pairwise correlations and PCA were performed using RPM values across all bins with values > 0 in at least 50% of the samples. Hierarchical clustering was performed using Euclidean distance as measure of dissimilarity and average linkage clustering.

4.2.6 Validation with EpiTYPER

Validation of two top signals in genes previously linked to infertility, *TNP1* and *C9orf3*, was pursued with EpiTYPER. Five hundred nanograms of genomic DNA were bisulphite converted using the MethylEasy Exceed Rapid Bisulphite Modification Kit (Human Genetic Signatures, North Ryde, NSW, Australia). Primers to target the regions in *TNP1* and *C9orf3* were designed using the EpiDesigner tool (Sequenom Inc., Herston, QLD, Australia). The *H19 CTCF6* region was assayed in a previous study [133]. Primers, genomic coordinates and PCR conditions are shown in **Table 4-2**. Methylation levels were determined by EpiTYPER on the MassARRAY System (Sequenom Inc., Herston, QLD, Australia). Differential methylation analysis considered the average of 2-3 technical replicates and was performed using data on single CpG sites.

Table 4-2. PCR conditions

	Assay 1 (*TNP1*)	Assay 2 (*C9orf3*)
Location	chr2:217727235-217727594	chr9:97503908-97504381
Product size	360	474
Number of analysable CpG sites	2	2
Left primer	aggaagagagTTGTGTGAAATTATGTTTTATGTTTGT	aggaagagagATTTTATTTTTAGTGGTATGGTTTT
Right primer	cagtaatacgactcactatagggagaaggctCCCTACCTTAAAATAACCCCACTTA	cagtaatacgactcactatagggagaaggctACCTTCTAAATAAAACTCCCTATATAATC
PCR conditions	95°C for 10 min	95°C for 10 min
	5 cycles of 95°C for 20 s, 59-60°C for 30 s, and 72°C for 2 min	5 cycles of 95°C for 20 s, 56-58°C for 30 s, and 72°C for 2 min
	40 cycles of 95°C for 20 s, 59-60°C for 30 s, and 72°C for 2 min	40 cycles of 95°C for 20 s, 56-58°C for 30 s, and 72°C for 2 min
	72°C for 10 min	72°C for 10 min

4.3 Results

4.3.1 Genome-wide methylation profiles in twins

We profiled DNA methylation levels from a total of 107 newborn twins (47 conceived via IVF and 60 conceived *in vivo*) in whole blood and in cord blood mononuclear cells (WBCs and CBMCs, respectively). Details of any fertility treatment used and demographic characteristics that represent potential confounders of DNA methylation levels at birth such as, sex, birth weight, maternal age, and maternal smoking status are shown in **Table 4-3**.

Table 4-3. Sample description

Group	Total number of twins (number of complete sets)	Zygosity and chorionicity[a]	Sex[b]	Birth weight (kg) mean(s.d.)	Maternal age (years) mean(s.d.)	Maternal smoking (% smokers)	Ovarian stimulation	ICSI	GIFT	Frozen embryo
WBCs										
IVF	40 (20)	10 MZ MC 30 DZ DC	18(F) 22(M)	2.57(4.77)	36(4)	20%	6(No) 34(Yes)	22(No) 18(Yes)	-	28(No) 12(Yes)
Non-IVF	58 (29)	14 MZ MC 12 MZ DC 32 DZ DC	34(F) 24(M)	2.58(3.99)	32(5)	28%	56(No) 2(Yes)	-	56(N) 2(Y)	-
CBMCs										
IVF	35 (16)	9 MZ MC 1 MZ DC 25 DZ DC	16(F) 19(M)	2.50(4.42)	35(5)	23%	1(No) 34(Yes)	14(No) 21(Yes)	-	24(No) 11(Yes)
Non-IVF	47 (22)	12 MZ MC 10 MZ DC 25 DZ DC	30(F) 17(M)	2.60(3.48)	32(4)	28%	45(No) 2(Yes)	-	45(N) 2(Y)	-

[a]MZ, monozygotic; DZ, dizygotic; MC, monochorionic; DC, dichorionic
[b]F, female; M, male
[c]N, no; Y, yes
ICSI, intracytoplasmic sperm injection; GIFT, gamete intra-fallopian transfer

Hierarchical clustering of pairwise correlations of genome-wide methylation profiles showed that global patterns of methylation were not influenced by tissue type, method of conception, chorionicity, or zygosity (**Figure 4-2**).

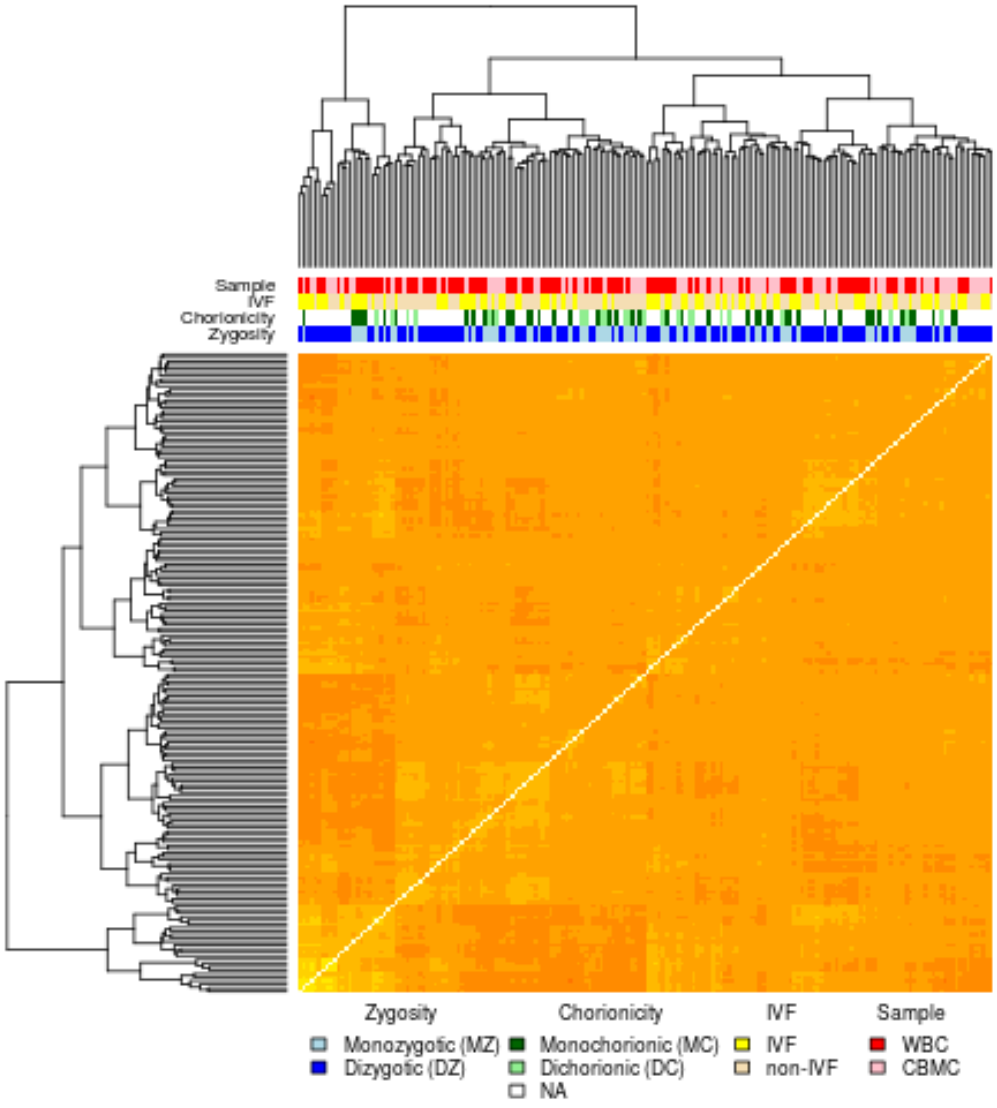


Figure 4-2. Heatmap of pairwise correlations of genome-wide methylation profiles. Hierarchical clustering was performed on pairwise correlations of methylation profiles. Top bars were coloured to indicate sample type, method of conception, chorionicity, and zygosity.

We then explored genome-wide patterns of DNA methylation variability in the dataset. Principal component analysis was used to identify factors that were significantly associated with genome-wide variability in DNA methylation profiles.

The first five principal components in the dataset, which explained ~13% of the total variance in DNA methylation, were at least nominally associated (P<0.05) with sample type (WBCs vs CMBCs), birth weight, maternal smoking, and conception method (**Figure 4-3a**).

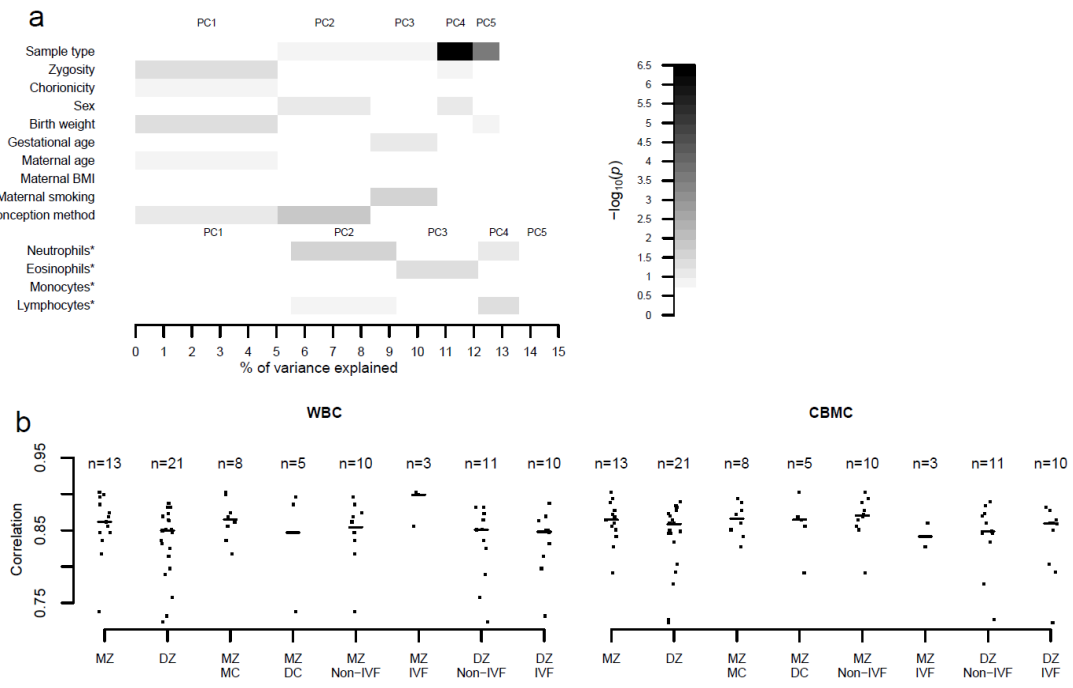


Figure 4-3. Global methylation patterns. (a) Biological factors associated with principal components of variation of methylation profiles. Variables marked with * were only available in a subset of the sample (n=54). (b) Within-pair methylation correlation in WBCs and CBMCs.

We next estimated the within twin-pair correlation of methylation profiles in twin pairs available in both datasets. In concordance with previous studies [90], we observed higher median correlation within monozygotic (MZ) twin pairs compared to dizygotic (DZ) twin pairs (**Figure 4-3b**). Previous studies have shown that twin chorionicity can have an effect on within-pair DNA methylation differences, but not with consistent direction of effect across tissues [39, 90, 133]. In our study, we did not observe significant chorionicity-related methylation

differences (Figure 1b), but the number of MZ twins within chorionicity categories was relatively low (n=8 monochorionic and n=5 dichorionic pairs). Interestingly, the method of conception showed methylation profile differences within MZ twin pairs. MZ IVF twins had higher median correlation compared to MZ non-IVF twins in WBC, but the opposite trend was observed in CBMCs, and in both cases the MZ IVF sample was small (n=3).

4.3.2 IVF-DMRs in CBMCs and WBCs

In order to identify tissue-independent and tissue-specific IVF-associated DMRs, we compared DNA methylation profiles in WBCs and CBMCs in relation to method of conception adjusting for birth weight, sex, maternal smoking, and the first 5 principal components, which capture partly cell heterogeneity (**Figure 4-3b**). Epigenome-wide analyses of DNA methylation in relation to method of conception did not identify genome-wide significant signals in the CBMCs subset or in the combined CBMC and WBC datasets, after correction for multiple testing (data not shown). In WBCs alone, one significant DMR was observed at a false discovery rate (FDR) of 5% (**Figure 4-4**). This was located ~3kb upstream of *TNP1* (chr2:217,726,751-217,727,250), which encodes a transition nuclear protein that replaces histones and is subsequently replaced by protamines during spermiogenesis. A deletion in the promoter region of this gene, which reduces its expression, has been reported in infertile men [136]. Methylation upstream of *TNP1* might have an impact on its expression. In mice, methylation changes during spermatogenesis have been observed at *TNP1,* which suggests a role of methylation in the regulation of this gene [137].

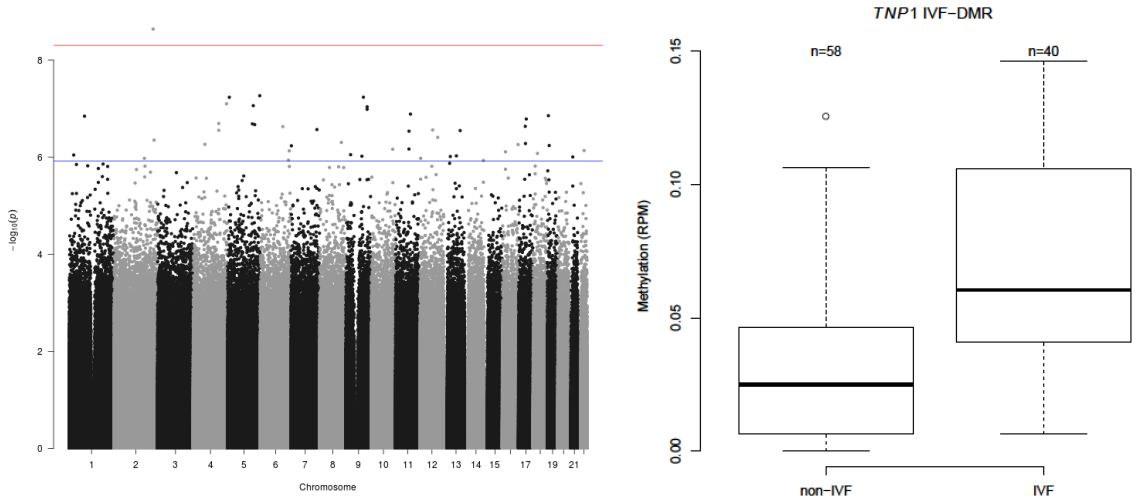


Figure 4-4. TNP1 IVF-DMR. Left, manhattan plot of the WBC IVF-DMR analysis showing 1 genome-wide significant hit at 5% FDR (red line) and 46 suggestive hits at 25% FDR (blue line). Right, boxplot of the methylation values (RPM) at the top IVF-DRM identified ~3kb upstream of TNP1.

To explore the biological characteristics of the top-ranked results in the IVF epigenome-wide analyses we selected a more liberal threshold of FDR 25%, at which 46 IVF-DMRs were identified (**Table 4-4**). Interestingly, the third-ranked DMR genome-wide was located in the first intron of *C9orf3* (chr9:97,504,001-97,504,500), which has been associated with polycystic ovary syndrome in women [138] and development of erectile dysfunction after radiotherapy for prostate cancer in men [139]. Another signal within this list was located in intron 1 of *STOX2* (chr4:184,814,001-184,814,500), whose reduced expression has been implicated in pre-eclampsia [140]. Since adverse perinatal outcomes may be associated with maternal age, we further adjusted for this covariate and observed that the 46 FDR 25% WBC IVF-DMRs remained significant (**Table 4-4**).

Table 4-4. FDR 25% WBC IVF-DMRs

Chromosome	Start	End	p	FDR adjusted p	p (adjusted for maternal age)	Gene name*	Gene start*	Gene end*
chr2	217726751	217727250	2.30E-09	0.0221	6.40E-10	*AC007557.1*	217735495	217736362
						TNP1	217724181	217724787
chr5	178761751	178762250	5.43E-08	0.1244	1.76E-05	*ADAMTS2*	178537852	178772431
chr9	97504001	97504500	5.83E-08	0.1244	5.14E-07	*C9orf3*	97488983	97849441
chr5	9275751	9276250	5.86E-08	0.1244	1.67E-06	*SEMA5A*	9035138	9546187
chr4	184814001	184814500	7.95E-08	0.1244	4.96E-07	*STOX2*	184774584	184944679
chr5	142488501	142489000	8.73E-08	0.1244	5.76E-07	*ARHGAP26*	142149949	142608576
chr9	118148751	118149250	9.20E-08	0.1244	4.61E-06	*DEC1*	117904097	118164923
chr9	118149001	118149500	1.04E-07	0.1244	3.47E-06	*DEC1*	117904097	118164923
chr11	82654251	82654750	1.30E-07	0.1250	5.26E-08	*C11orf82*	82611017	82669319
						PRCP	82534544	82681626
						RAB30	82684175	82782965
chr19	6165251	6165750	1.40E-07	0.1250	1.00E-06	*RFX2*	5993175	6199583
						ACSBG2	6135258	6193112
						MLLT1	6212966	6279959
chr1	85522251	85522750	1.43E-07	0.1250	2.56E-07	*WDR63*	85464830	85598821
						MCOLN3	85483765	85514182
chr17	42569001	42569500	1.64E-07	0.1274	1.90E-05	*GPATCH8*	42472652	42580798
chr4	141606501	141607000	2.03E-07	0.1274	1.59E-05	*TBC1D9*	141541919	141677274
chr5	137736001	137736500	2.06E-07	0.1274	1.13E-06	*REEP2*	137774706	137782658
						KDM3B	137688285	137772717
chr5	150614501	150615000	2.14E-07	0.1274	2.23E-07	*SLC36A3*	150656323	150683327
						GM2A	150591711	150650001
						CCDC69	150560613	150603706

chr17	36918251	36918750	2.32E-07	0.1274	2.02E-07	*MLLT6*	36861795	36886056
						CISD3	36886488	36891297
						CWC25	36956687	36981734
						PIP4K2B	36921942	36956379
						PCGF2	36890150	36906070
						CTB-58E17.5	36905613	36906969
						PSMB3	36908989	36920484
						AC006449.1	36884086	36884451
chr6	126138251	126138750	2.36E-07	0.1274	2.12E-06	*NCOA7*	126102307	126252266
chr7	144431251	144431750	2.70E-07	0.1274	5.29E-09	*TPK1*	144149034	144533488
chr12	70937251	70937750	2.76E-07	0.1274	1.83E-06	*PTPRB*	70910630	71031220
chr4	141606251	141606750	2.80E-07	0.1274	1.50E-05	*TBC1D9*	141541919	141677274
chr13	90019001	90019500	2.84E-07	0.1274	5.94E-07	-	-	-
chr11	74179001	74179500	2.92E-07	0.1274	1.61E-09	*LIPT2*	74202757	74204778
						POLD3	74204896	74380162
						KCNE3	74165886	74178774
chr12	99153001	99153500	3.93E-07	0.1637	2.44E-07	*ANKS1B*	99120235	100378432
						APAF1	99038919	99129204
chr2	223336751	223337250	4.47E-07	0.1788	2.00E-06	*SGPP2*	223289236	223425667
chr8	120972001	120972500	4.98E-07	0.1869	3.26E-08	*DEPTOR*	120885957	121063152
chr17	38047001	38047500	5.25E-07	0.1869	7.83E-06	*GSDMB*	38060848	38076107
						ZPBP2	38024417	38034149
						IKZF3	37921198	38020441
						ORMDL3	38077294	38083854
chr4	64626751	64627250	5.45E-07	0.1869	2.67E-06	-	-	-
chr16	87256751	87257250	5.51E-07	0.1869	2.12E-08	*C16orf95*	87117168	87351022
chr19	10656751	10657250	5.77E-07	0.1869	1.42E-06	*CDKN2D*	10677138	10679735
						ATG4D	10654571	10664094

						KEAP1	10596796	10614417
						AP1M2	10683347	10697991
						KRI1	10663761	10676713
						S1PR5	10623623	10628607
chr7	2487251	2487750	5.85E-07	0.1869	2.03E-05	*CHST12*	2443223	2474242
chr11	74178751	74179250	6.82E-07	0.2059	1.31E-07	*KCNE3*	74165886	74178774
						LIPT2	74202757	74204778
						POLD3	74204896	74380162
chr10	119176501	119177000	6.87E-07	0.2059	1.11E-06	*PDZD8*	119040000	119134978
chr22	34755251	34755750	7.32E-07	0.2094	1.76E-05	-	-	-
chr6	161664751	161665250	7.42E-07	0.2094	2.93E-06	*AGPAT4*	161551011	161695093
chr16	17161751	17162250	7.80E-07	0.2137	2.41E-06	*XYLT1*	17195626	17564738
chr18	23695001	23695500	8.39E-07	0.2235	1.97E-06	*PSMA8*	23713816	23773319
						SS18	23596578	23671181
chr9	26364751	26365250	8.90E-07	0.2267	1.02E-06	-	-	-
chr1	25227001	25227500	9.05E-07	0.2267	1.61E-05	*RUNX3*	25226002	25291612
chr13	68877251	68877750	9.43E-07	0.2267	2.19E-05	-	-	-
chr9	89126501	89127000	9.59E-07	0.2267	7.00E-06	-	-	-
chr13	35317501	35318000	9.72E-07	0.2267	2.51E-05	-	-	-
chr21	19575001	19575500	9.92E-07	0.2267	2.31E-06	*CHODL*	19273580	19639690
chr2	169470001	169470500	1.06E-06	0.2307	4.35E-07	*CERS6*	169312372	169631644
chr12	4310251	4310750	1.06E-06	0.2307	8.13E-05	-	-	-
chr6	157136501	157137000	1.15E-06	0.2448	1.61E-06	*ARID1B*	157099063	157531913
chr14	104067251	104067750	1.17E-06	0.2448	8.21E-06	*APOPT1*	104029299	104073860
						BAG5	104022881	104029168
						KLC1	104028233	104167888
						RP11-73M18.2	104029299	104152261

*From GENCODE v19

The non-IVF group included a small number (n=4) of newborns conceived with other types of fertility treatments not equivalent to IVF, such as gamete intra-fallopian transfer (GIFT) and ovarian stimulation. We re-analysed the 46 FDR 25% WBC IVF-DMRs excluding GIFT (n=2) and non-IVF ovarian stimulation (n=2) controls, and observed that conclusions remained unchanged (Annex II).

Hierarchical clustering using DNA methylation levels at these 46 FDR 25% DMRs alone classified 38 out of 40 IVF-twins by method of conception (

Figure 4-5). We also explored these signals with respect to functional annotations. A total of 10 FDR 25% WBC IVF-DMRs overlapped CpG sites previously shown to be dynamic during development [34], 20 overlapped DNase I hypersensitivity sites (wgEncodeRegDnaseClusteredV3) [141], one overlapped a CpG island (cpgIslandEx) [142], and none overlapped with candidate metastable epialleles [143] (

Figure 4-5).

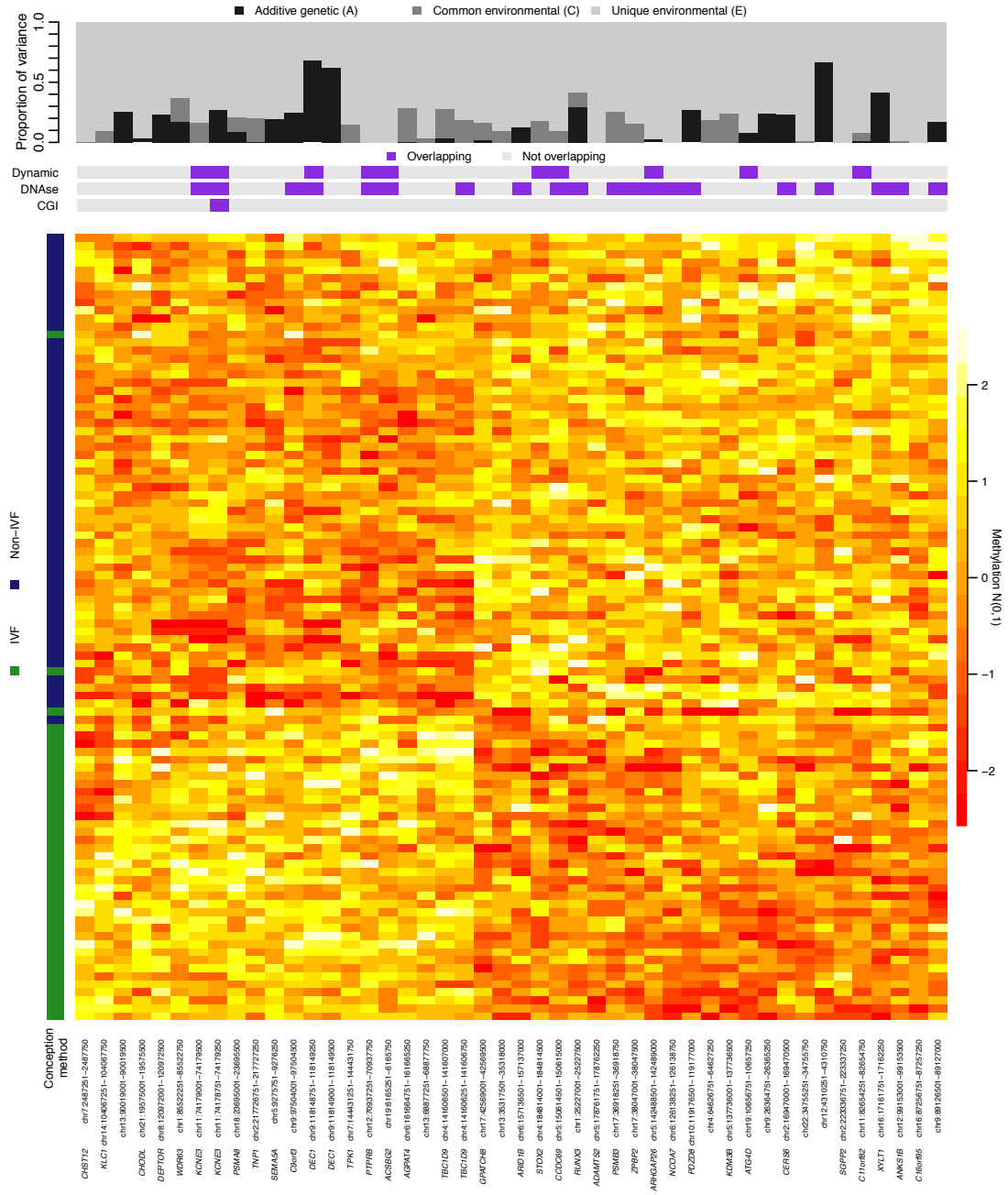


Figure 4-5. WBC IVF-DMRs. Heatmap rows correspond to the 98 WBC samples while columns correspond to the 46 FDR 25% WBC IVF-DMRs. The vertical colour bar indicates method of conception (IVF, green; non-IVF, blue). Top panel shows the fraction of variance explained by additive genetic (A), shared environmental (C), and unique environmental factors (E). Horizontal colour bars indicate overlap (violet) or absence (gray) of dynamic CpG sites, DNase I hypersensitivity sites, or CpG islands with the DMR.

Cell type-specific DNA methylation can impact the profiles observed in a population of cells, such as in a whole blood sample, and we therefore accounted

for blood cell type heterogeneity using a twofold approach. First, we performed principal component analysis on the methylation levels of the entire set of WBC samples, and our main EWAS analyses above are corrected for the first five principal components (PCs), which likely capture variation attributed to technical and biological factors, potentially including cell heterogeneity. To assess whether the first five PCs capture cell heterogeneity, blood cell subtype counts were obtained through automatic differential counting for a subset of the WBC samples (n=54 twins, 22 IVF and 32 non-IVF) and these were compared against the distributions of the first 5 PCs. Proportion of neutrophils, eosinophils and lymphocytes were associated ($P<0.05$) with the loadings of the second, third and fourth principal components, respectively (**Figure 4-3a**). Therefore, since the EWAS model used in this study took into account the loadings of the first five principal components, these analyses already take into account the influence of cell heterogeneity to a certain extent.

Second, we re-analysed the 46 FDR 25% WBC IVF-DMRs in the subset of 54 WBC samples with available cell counts, adjusting for the proportion of neutrophils, eosinophils, monocytes, and lymphocytes. We also performed analyses adjusting for the loadings of the first five principal components within this dataset alone. Most results were concordant when comparing across all models (Annex III) and only five out the 46 FDR 25% WBC IVF-DMRs were not significant ($P>0.05$) after adjusting for cell proportions in the subset of 54 individuals (chr8:120,972,001-120,972,500, chr7:2,487,251-2,487,750, chr18:23,695,001-23,695,500, chr12:4,310,251-4,310,750, and chr14:104,067,251-104,067,750).

4.3.3 Variance decomposition of WBC IVF-DMRs

Given that epigenetic changes were potentially affecting infertility genes, we wanted to investigate if the findings may capture a genetic signature affecting DNA methylation that could be transmitted to offspring. We applied twin variance decomposition analyses to partition the total epigenetic variance into additive genetic (A), and common (C) and unique (E) environmental components (ACE) [105]. The ACE model was used to determine the contribution of genetics, shared intrauterine environment due to shared maternal influences, and non-shared (twin-specific) or stochastic factors to epigenetic variation. The mean contribution of additive genetic effects (narrow-sense heritability) to DNA methylation across the genome in different tissues from newborns has been previously estimated to be between 0.05 and 0.12 [90]. Here we estimated the average genome-wide narrow-sense heritability for DNA methylation in WBCs at 0.06. At the 46 FDR 25% WBC IVF-DMRs, the major contributors to DNA methylation variation were non-shared or stochastic events (

Figure 4-5). However, several FDR 25% IVF-DMRs had evidence for heritability (A > 0.4), suggestive of genetic effects underlying specific IVF-associated DNA methylation changes. These included an intronic region in *DEC1* (chr9:118,148,751-118,149,500), a region 33kb away from *XYLT1* (chr16:17,161,751-17,162,250), and an intergenic region in chromosome 12 (chr12:4,310,251-4,310,750). When looking at the two DMRs associated to infertility genes, DNA methylation variation showed no evidence for genetic effects (A=0) near *TNP1*, while heritability at the DMR in *C9orf3* was estimated at 0.25.

4.3.4 Effects of IVF on imprinting

Previous studies have explored DNA methylation patterns in IVF births specifically at imprinting control regions (ICRs). We therefore assessed whether there was an enrichment of differential methylation effects at 34 known ICRs [144] in our genome-wide results, but no enrichment was observed (P>0.05). However, when we explored individual signals at candidate IVF-DMRs we were able to replicate one previously reported ICR IVF-associated DMR. Concordantly with previous IVF methylation studies in placental tissue [125] and buccal epithelium [80], we observed hypomethylation at the 6th CTCF binding site within the *H19*/*IGF2* (*H19 CTCF6*) DMR (**Figure 4-6**). This association was observed in CBMCs (P=0.01), but not in WBCs (data not shown).

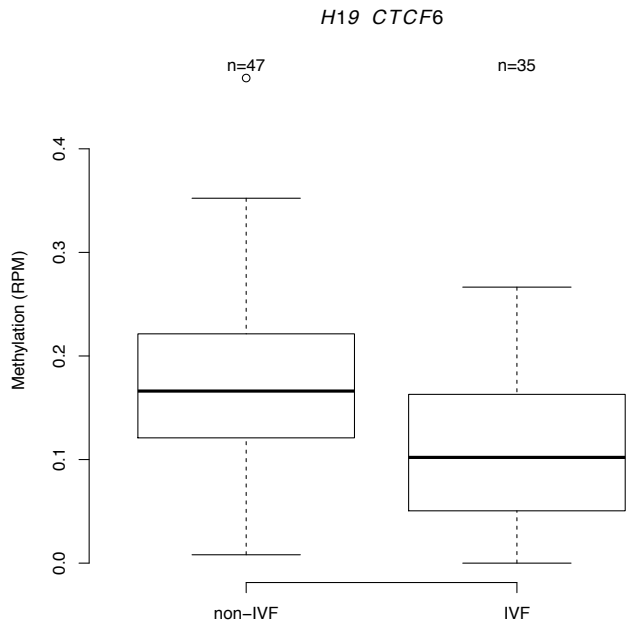


Figure 4-6. H19 CTCF6 IVF-DMR replication. Boxplot of the methylation values (RPM) at the sixth CTCF binding site within the H19 DMR (H19 CTCF6).

4.3.5 Effects of intracytoplasmic sperm injection (ICSI)

Intracytoplasmic sperm injection (ICSI) is a technique in IVF used to treat couples with male-factor infertility [145]. In contrast to conventional IVF where fertilisation

occurs by placing spermatozoa near an egg, ICSI consists of the direct injection of a selected single sperm cell into the egg. This manipulation may introduce additional risk factors [146]. To assess the effect of ICSI on the 46 FDR 25% WBC IVF-DMRs we adjusted for the use of this technology and also compared the ICSI and the conventional IVF groups separately against the non-IVF group. After adjustment for ICSI, p-values increased at several FDR 25% IVF-DMRs (**Table 4-5**), suggesting that ICSI or paternal infertility might have a role in these methylation changes. One FDR 25% IVF-DMR signal (chr1:85,522,251-85,522,750) appeared stronger after adjustment, suggesting either a female infertility effect or that ICSI prevents or corrects a methylation change that occurs in conventional IVF. This DMR was located upstream of *WDR63,* a gene mainly expressed in testis, fallopian tube, and adrenal gland [147].

Table 4-5. Effect of intracytoplasmic sperm injection on the FDR 25% WBC IVF-DMRs

			IVF (n=40) vs non-IVF (n=58)			IVF (n=34) vs non-IVF (n=58) adjusted for ICSI			conventional IVF (n=16) vs non-IVF (n=58)			ICSI (n=18) vs non-IVF (n=58)		
Chromosome	Start	End	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
chr2	217726751	217727250	1.18	0.19	2.30E-09	1.45	0.26	7.71E-08	-1.42	0.26	7.44E-08	-1.16	0.27	1.29E-05
chr5	178761751	178762250	-1.08	0.2	5.43E-08	-1.09	0.29	8.23E-05	0.87	0.30	1.89E-03	1.11	0.27	2.52E-05
chr9	97504001	97504500	1.07	0.2	5.83E-08	0.77	0.27	2.77E-03	-0.88	0.31	2.67E-03	-1.30	0.25	1.95E-07
chr5	9275751	9276250	1.09	0.19	5.86E-08	1.17	0.27	1.04E-05	-1.13	0.29	6.50E-05	-1.13	0.27	2.75E-05
chr4	184814001	184814500	-0.96	0.18	7.95E-08	-0.97	0.24	3.94E-05	0.94	0.26	1.37E-04	1.15	0.23	8.35E-07
chr5	142488501	142489000	-1.11	0.2	8.73E-08	-1.1	0.28	7.60E-05	1.25	0.31	5.16E-05	1.08	0.28	7.10E-05
chr9	118148751	118149250	1.09	0.2	9.20E-08	1.35	0.28	1.28E-06	-1.30	0.29	8.90E-06	-1.02	0.28	1.68E-04
chr9	118149001	118149500	1.08	0.2	1.04E-07	1.31	0.28	2.22E-06	-1.23	0.29	2.00E-05	-1.05	0.29	1.69E-04
chr11	82654251	82654750	-1.05	0.19	1.30E-07	-0.97	0.29	4.68E-04	0.95	0.31	1.53E-03	0.88	0.27	5.94E-04
chr19	6165251	6165750	0.94	0.18	1.40E-07	0.89	0.25	1.90E-04	-0.88	0.28	9.35E-04	-1.02	0.23	7.36E-06
chr1	85522251	85522750	0.99	0.18	1.43E-07	1.54	0.25	4.48E-09	-1.53	0.25	7.41E-09	-0.80	0.28	2.32E-03
chr17	42569001	42569500	-1.07	0.19	1.64E-07	-1.25	0.27	2.21E-06	1.30	0.28	3.81E-06	1.21	0.26	2.44E-06
chr4	141606501	141607000	1.09	0.19	2.03E-07	1.26	0.26	3.56E-06	-1.33	0.30	1.05E-05	-1.13	0.26	1.83E-05
chr5	137736001	137736500	-1.03	0.19	2.06E-07	-1.11	0.28	4.12E-05	1.13	0.30	1.39E-04	1.10	0.27	3.78E-05
chr5	150614501	150615000	-1.08	0.21	2.14E-07	-1.2	0.29	2.80E-05	1.24	0.31	3.18E-05	1.07	0.29	1.21E-04
chr17	36918251	36918750	-1.1	0.21	2.32E-07	-0.96	0.29	6.01E-04	1.05	0.31	4.88E-04	1.03	0.28	1.83E-04
chr6	126138251	126138750	-0.99	0.19	2.36E-07	-0.75	0.27	3.29E-03	0.79	0.30	4.92E-03	1.22	0.25	8.94E-07
chr7	144431251	144431750	1.02	0.19	2.70E-07	0.84	0.28	1.30E-03	-0.85	0.28	1.48E-03	-1.17	0.27	1.77E-05
chr12	70937251	70937750	0.88	0.17	2.76E-07	0.72	0.23	1.12E-03	-0.87	0.26	5.26E-04	-0.94	0.20	3.10E-06
chr4	141606251	141606750	1.01	0.19	2.80E-07	1.16	0.26	8.05E-06	-1.31	0.29	6.93E-06	-1.04	0.25	2.14E-05
chr13	90019001	90019500	1.06	0.2	2.84E-07	1.25	0.29	8.88E-06	-1.36	0.30	5.09E-06	-0.84	0.29	2.20E-03
chr11	74179001	74179500	1.07	0.2	2.92E-07	1.3	0.29	7.25E-06	-1.37	0.29	3.15E-06	-1.06	0.29	1.76E-04

chr12	99153001	99153500	-1.01	0.2	3.93E-07	-1.02	0.29	2.02E-04	1.02	0.31	4.59E-04	1.24	0.27	3.77E-06
chr2	223336751	223337250	-1.01	0.2	4.47E-07	-1.41	0.28	4.00E-07	1.50	0.29	1.69E-07	0.77	0.29	4.65E-03
chr8	120972001	120972500	0.97	0.19	4.98E-07	0.83	0.27	1.22E-03	-0.87	0.29	1.43E-03	-0.96	0.26	1.81E-04
chr17	38047001	38047500	-1.04	0.2	5.25E-07	-1.3	0.28	4.51E-06	1.31	0.27	1.21E-06	0.74	0.28	5.39E-03
chr4	64626751	64627250	-0.99	0.19	5.45E-07	-0.87	0.28	9.92E-04	0.86	0.29	1.61E-03	0.98	0.26	1.36E-04
chr16	87256751	87257250	-0.83	0.16	5.51E-07	-0.97	0.23	1.75E-05	0.94	0.25	9.13E-05	0.82	0.24	3.37E-04
chr19	10656751	10657250	-1.03	0.2	5.77E-07	-1.09	0.28	6.47E-05	1.17	0.29	3.80E-05	1.13	0.27	2.18E-05
chr7	2487251	2487750	0.8	0.16	5.85E-07	0.73	0.23	1.06E-03	-0.81	0.24	5.61E-04	-0.94	0.23	3.49E-05
chr11	74178751	74179250	0.99	0.2	6.82E-07	1.19	0.28	2.15E-05	-1.15	0.28	2.45E-05	-0.82	0.28	2.37E-03
chr10	119176501	119177000	-1.04	0.2	6.87E-07	-1.03	0.29	3.41E-04	1.16	0.31	1.69E-04	1.14	0.28	4.83E-05
chr22	34755251	34755750	-0.98	0.2	7.32E-07	-0.8	0.28	2.35E-03	0.97	0.31	1.21E-03	1.15	0.25	3.95E-06
chr6	161664751	161665250	0.98	0.19	7.42E-07	0.9	0.26	3.89E-04	-0.83	0.29	2.48E-03	-0.81	0.25	7.71E-04
chr16	17161751	17162250	-1.01	0.2	7.80E-07	-0.99	0.28	2.19E-04	1.17	0.30	4.48E-05	1.08	0.28	1.16E-04
chr18	23695001	23695500	1.04	0.21	8.39E-07	0.95	0.29	5.83E-04	-1.04	0.32	6.05E-04	-1.11	0.28	3.96E-05
chr9	26364751	26365250	-0.95	0.19	8.90E-07	-0.96	0.28	2.78E-04	1.13	0.30	9.90E-05	0.96	0.28	2.77E-04
chr1	25227001	25227500	-0.83	0.17	9.05E-07	-0.87	0.25	3.81E-04	0.78	0.25	1.39E-03	0.99	0.25	5.78E-05
chr13	68877251	68877750	0.97	0.2	9.43E-07	1.01	0.28	1.44E-04	-1.24	0.28	1.05E-05	-0.84	0.28	1.39E-03
chr9	89126501	89127000	-0.92	0.19	9.59E-07	-1.14	0.26	9.49E-06	1.10	0.28	6.62E-05	0.76	0.26	2.12E-03
chr13	35317501	35318000	-0.96	0.2	9.72E-07	-0.54	0.29	4.44E-02	0.54	0.31	6.52E-02	1.04	0.27	7.61E-05
chr21	19575001	19575500	0.96	0.19	9.92E-07	0.88	0.27	8.63E-04	-0.96	0.30	8.25E-04	-1.13	0.26	1.24E-05
chr2	169470001	169470500	-1.02	0.21	1.06E-06	-1.31	0.28	2.08E-06	1.33	0.29	6.99E-06	1.04	0.27	5.91E-05
chr12	4310251	4310750	-0.99	0.2	1.06E-06	-0.91	0.29	8.79E-04	1.02	0.30	4.54E-04	0.87	0.30	1.99E-03
chr6	157136501	157137000	-0.96	0.19	1.15E-06	-0.78	0.27	2.19E-03	0.91	0.30	1.37E-03	1.30	0.25	1.62E-07
chr14	104067251	104067750	0.71	0.14	1.17E-06	0.81	0.21	7.35E-05	-0.75	0.22	3.89E-04	-0.58	0.24	9.44E-03

4.3.6 Validation of IVF-DMRs

We pursued validation of the differential methylation signals at the top associated DMR (located ~3kb upstream of *TNP1*) and at the third-ranked DMR (located in *C9orf3*), both in or near genes previously linked to infertility. Altogether, four CpG sites were targeted for validation using Sequenom's EpiTYPER technology.

For the DMR in *C9orf3,* we were able to target two CpG sites within the most-associated 500bp bin in this locus (**Figure 4-7**). We assayed methylation levels in 36 MZ twins included in the discovery EWAS, and observed significantly higher methylation in the IVF group, concordant with the MeDIP-seq analysis, at both tested CpG-sites in the *C9orf3 locus* (P=0.02 and P=0.03 respectively), therefore validating this signal using a different methylation profiling approach (**Figure 4-8**).

For the *TNP1* DMR we were unable to target CpGs within the most associated 500bp bin*,* and we therefore selected two of the closest CpG sites contained within the second most associated DMR in that locus (**Figure 4-7**). Within the sample of 36 MZ twins we also observed higher methylation in the IVF group, consistent with the MeDIP-seq signal, with effects close to nominal significance (P=0.08) (**Figure 4-8**). However, correlation between the MeDIP-seq signal at the most-associated DMR in *TNP1* and the EpiTYPER methylation values was, as expected, relatively low as we were unable to target CpG sites within this most-associated DMR (correlation of 0.18 and 0 at the two tested CpG-sites). We profiled additional samples from DZ twin pairs, but did not obtain validation of the signal.

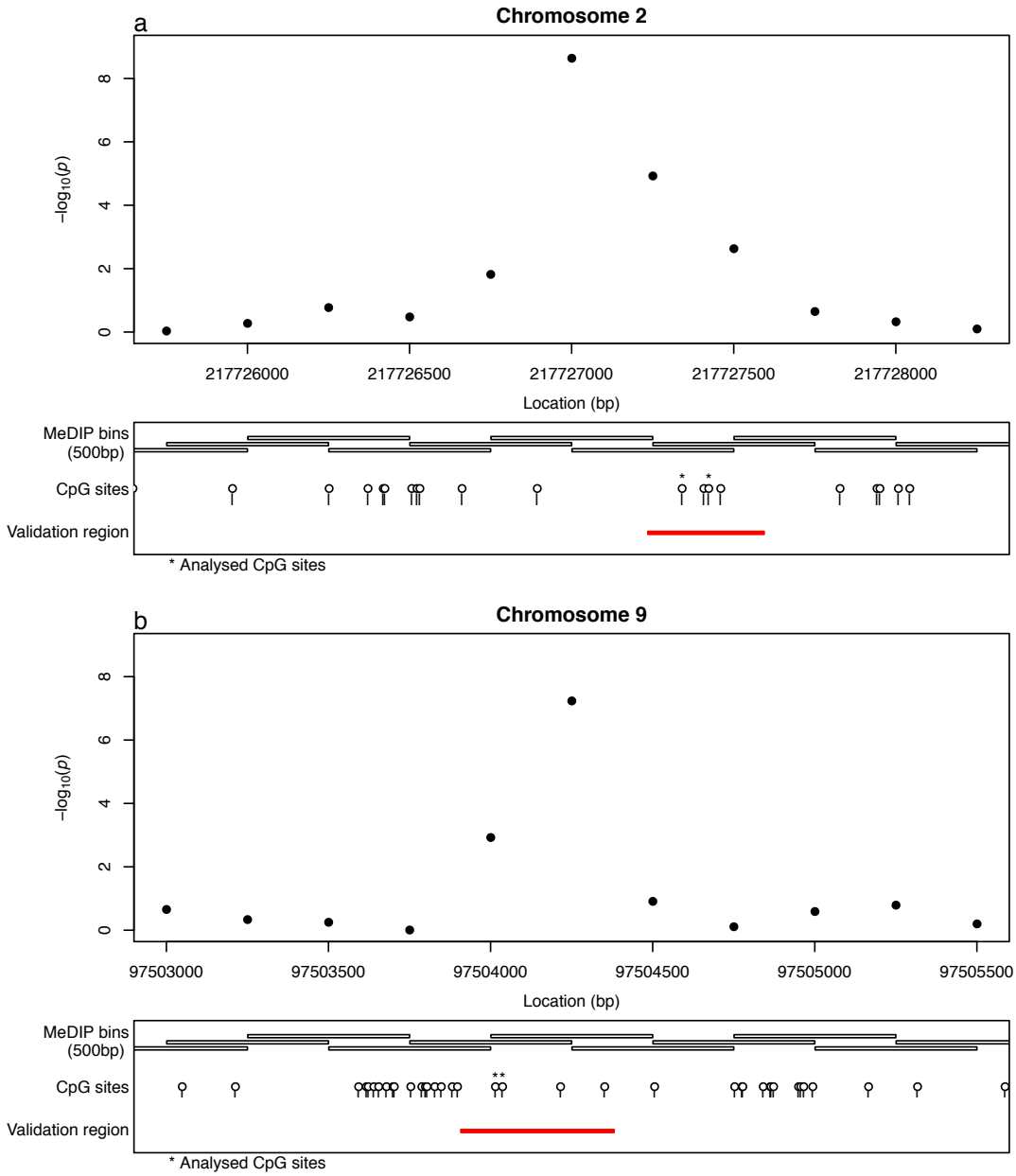


Figure 4-7. CpG sites targeted for validation. Top panel shows EWAS signals neighbouring the most associated DMR located near TNP1 (a) and C9orf3 (b) genes. Bottom panel shows the corresponding 500bp bins tested during EWAS, the CpG sites contained in those regions, and the targeted region for PCR.

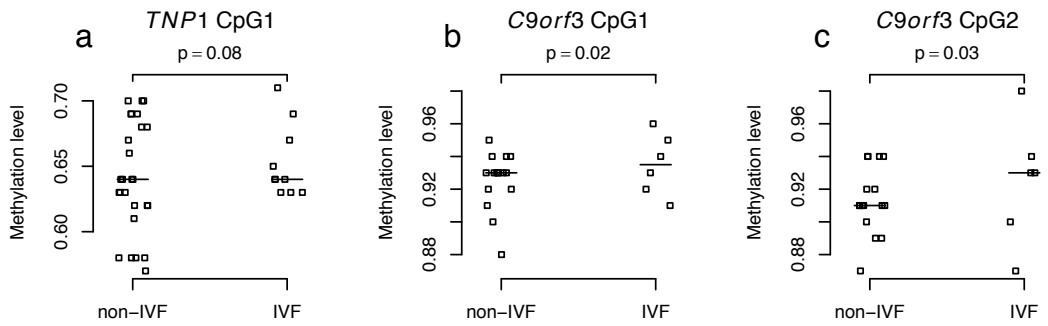


Figure 4-8. IVF vs non-IVF DNA methylation differences using Sequenom's EpiTYPER technology. Horizontal lines show the group median.

We also considered the effect of ICSI compared to conventional IVF in MZ twins in the validation dataset. We observed significantly higher methylation in the ICSI group at the first CpG of the targeted region near *TNP1* and at the first CpG site of *C9orf3* (**Figure 4-9**).

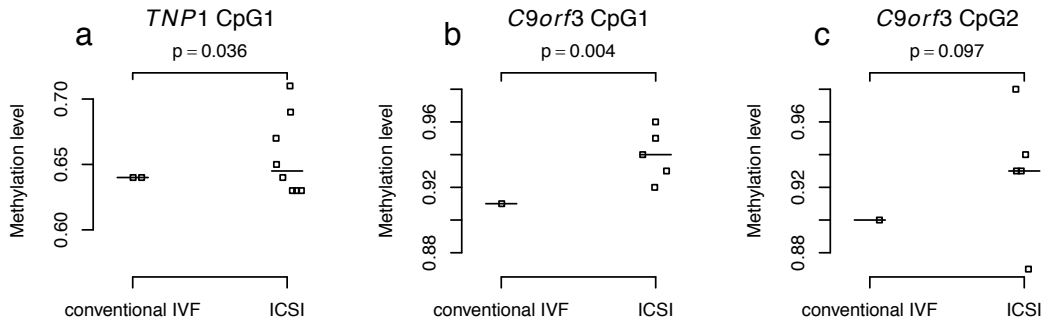


Figure 4-9. ICSI vs conventional IVF DNA methylation differences using Sequenom's EpiTYPER technology. Horizontal lines show the group median.

Lastly, we also compared methylation in relation to conception method at the *H19 CTCF6* DMR in a reduced subset of CBMCs samples (n=42 twins) using EpiTYPER. When comparing IVF to non-IVF twins (**Figure 4-10**) we observed a difference with the same direction of effect as in the MeDIP-seq analysis, although not significant (P=0.19). Interestingly, when comparing naturally-

conceived twins to twins that were conceived with any type of medical help (**Figure 4-10**), i.e. not exclusively IVF, the difference reached nominal significance (P=0.04), suggesting that differential methylation at this region is associated to parental subfertility rather than IVF conception.

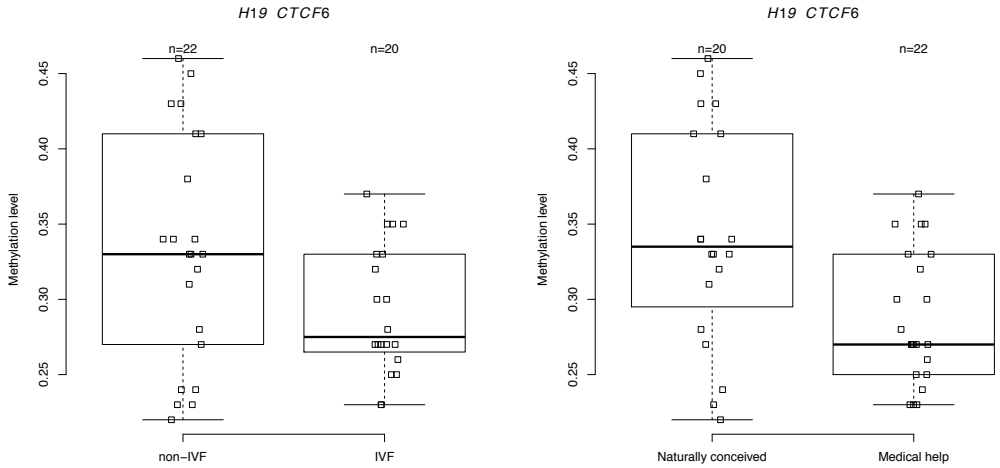


Figure 4-10. H19 CTCF6 IVF-DMR replication using Sequenom's EpiTYPER technology. Boxplot of the methylation values at the sixth CTCF binding site within the H19 DMR (H19 CTCF6) comparing IVF vs non-IVF newborns (left) and the group using any type of medical help vs naturally conceived newborns (right).

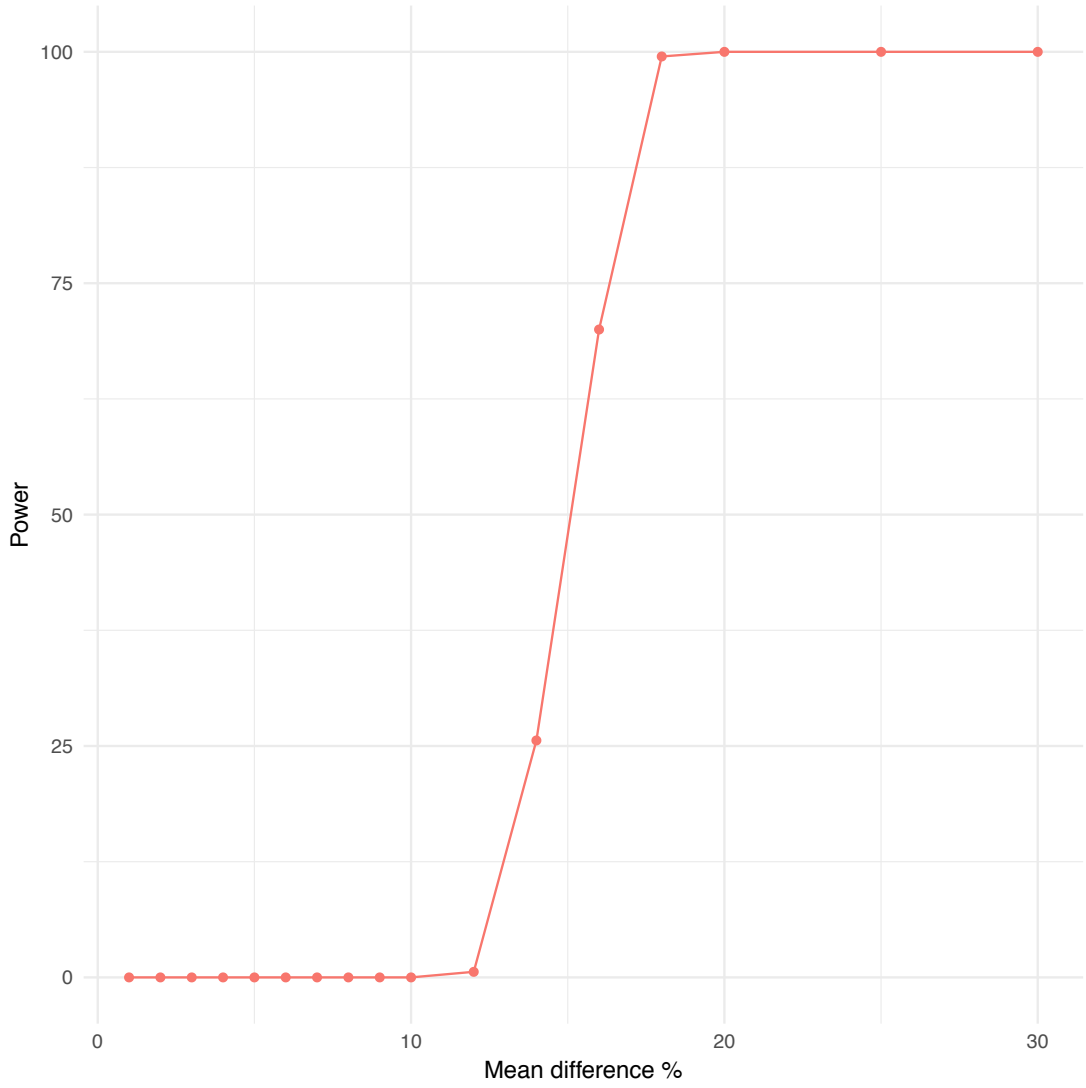


Figure 4-11. Power estimation. Based on simulations by Tsai & Bell (2015) for a sample size of 50 pairs of cases and controls at a genome-wide significance threshold of 1 x 10^{-6}.

4.4 Discussion

Since IVF procedures are carried out during an important period of epigenetic reprogramming in early development, we hypothesized that IVF may induce epigenetic differences that persist to birth. We were able to identify significant and suggestive differentially methylated regions related to IVF conception (IVF-DMRs) in WBCs, although our results suggest that at least some of these changes may be linked to parental subfertility, which is confounded with IVF treatment. For this reason, even though we observed some small differences

in DNA methylation, we were unable to prove that they have a relationship with IVF itself. The observation that IVF-DMRs were identified close to genes implicated in fertility and reproduction suggests that a genetic signature influencing DNA methylation could be transmitted from parent to offspring. To assess this further, we estimated the heritability of the IVF-DMRs. We observed that the IVF-DMR located in *C9orf3*, a gene associated with polycystic ovary syndrome, was estimated to have a heritability at 25% and eight other FDR 25% WBC IVF-DMRs showed heritability greater than this (

Figure 4-5).

The results presented in this study could be explained by different scenarios i) IVF is not associated with changes in DNA methylation and any differences observed are only due to parental subfertility ii) IVF is in fact the cause of the methylation changes, the changes at fertility genes were found by chance and have no relationship with the fertility status of the parents iii) both parental subfertility and IVF are associated with changes in DNA methylation. The first scenario can be tested by comparing the methylation profiles of subjects born to subfertile couples after IVF with subjects conceived by IVF from gametes of fertile donors. To test the second scenario, it would be necessary to compare profiles of non-IVF subjects born to fertile parents with subjects conceived by IVF from gametes of fertile donors. The third scenario would be proved if differences are observed in both comparisons.

Epigenetic states of metastable epialleles in mammals are mitotically inherited after establishment in early development, therefore shared across tissues, and can cause expression variability within isogenic individuals [148]. A study in humans looking for systematic inter-individual variation in DNA methylation across tissues from two different lineages identified 109 candidate

metastable epialleles [143]. Nutritional conditions during conception have been shown to be important to the establishment of epigenetic states at some of these metastable epialleles [56]. If an influence of IVF on the epigenetic marks of these alleles exists, it could potentially cause lifetime lasting effects. A previous study, which included newborns from single and multiple pregnancies, identified DNA methylation differences in IVF-conception at candidate metastable epialleles, although at different epialleles to those affected by maternal nutritional factors [129]. In our study, none of the 109 candidate metastable epialleles overlapped with the 46 FDR 25% WBC IVF-DMRs. This discrepancy could be attributed to differences between single and multiple pregnancies or to low power to detect such changes.

Our results also showed that IVF-DMRs, including hypomethylation of the regulatory region of *H19*, were generally not shared between WBCs and CBMCs. This observation suggests that the epigenetic differences reported here likely did not appear during early development or that these effects are not fixed and can revert in a cell type-specific manner. CBMCs, in contrast to WBCs, lack the granulocyte fraction, which is the predominant group of cells in the blood. Thus, the IVF-DMRs may be granulocyte-specific or at least in part influenced by this group of cells.

To date, there has been mixed evidence on the effect of IVF at imprinted genes and their regulatory regions. Some studies have reported DNA methylation changes or increased variability at these imprinted regions [80, 124, 125], while others have reported no associated changes [126, 149]. We observed that there is not an overall destabilization of methylation patterns in ICRs, but specific DMRs, such as the *H19* DMR, can show a weak but nominally significant association with the method of conception. Previous studies have reported

similar observations, that is, changes in methylation at some imprinted regions, but not in the majority [126, 129]. It is unknown if these changes occur due to IVF since imprinting defects have been previously described in sperm of infertile men, including hypomethylation of the *H19 CTCF6* DMR [150]. Loke *et al.*, 2015, reported that hypomethylation at this locus in buccal epithelium of newborns in the IVF group was driven by the subgroup conceived by ICSI [80]. However, it is difficult to dissect whether the observed effect on DNA methylation of ICSI-conceived newborns is due to the technique itself or to male infertility. Whitelaw *et al*., 2014, found higher levels of *SNRPN* methylation in buccal cells of ICSI-conceived newborns and these were associated with longer duration of infertility in the parents [151]. In our data, we observed that the difference at the *H19 CTCF6* DMR was greater when considering any type of medical help during conception supporting the idea that parental subfertility is the driver of methylation changes at this region. Information about the indication for ART, the use of donor eggs or sperm, and the fertility status of parents in the control group would be required to further assess the effect of parental subfertility.

Adverse perinatal outcomes and increased frequency of imprinting disorders has also been observed in offspring of couples with a history of subfertility that were able to conceive naturally [152–154]. However, studies that controlled for parental subfertility by comparing siblings in which one was conceived naturally and the other by IVF also observed an effect [155]. It is therefore likely that both parental subfertility and IVF may induce epigenetic changes, as observed in another genome-wide study that found DNA methylation differences between IVF-conceived newborns and a group conceived through intrauterine insemination (infertile controls), but also between the latter and naturally-conceived newborns (fertile controls) [129]. In addition, a

study looking at 37 candidate CpG sites identified seven that were differentially methylated when comparing an IVF-conceived group born to parents without male infertility that used donor oocytes to naturally-conceived newborns [156].

Finally, two IVF-DMRs associated to infertility (*TNP1* and *C9orf3*) were targeted for validation. Differential methylation was validated at the *C9orf3* gene. However, validation of the *TNP1* region was hampered by our inability to target CpG-sites within the most-associated DMR in this locus. We attempted validation at *TNP1* by targeting CpG sites in the neighbouring 500bp bin, and observed consistent direction of association close to nominal significance.

In this study the non-IVF group included a set of twins conceived after GIFT and another set conceived after ovarian stimulation not followed by IVF. GIFT and ovarian stimulation are fertility treatments not equivalent to IVF since fertilisation still occurs in the fallopian tubes. We showed that our results were not affected by the inclusion of these data, potentially because they were represented in small numbers, only four out 58 samples.

There are several limitations in this study. Firstly, it is known that cell composition may represent a confounding variable in EWAS [157]. Our results use principal component analysis anticipating that these will capture cell heterogeneity, and follow-up of our findings in a subset of twins with available cell counts showed that the majority of findings remained significant after adjustment for cell heterogeneity. Second, although MeDIP-seq has the strength of genome-wide coverage, it lacks base-pair resolution, instead generating methylation scores across genomic regions. However, it has been reported that methylation of neighbouring CpG sites is correlated over distances up to 1,000 bp [158], suggesting that the approach may be able to capture a good proportion of the methylation variance in a genomic region. Third, although this study

includes a sample size larger than most previous studies exploring IVF, contemporary EWAS study designs generally require larger number of cases and controls to achieve sufficient power to detect small to moderate effect sizes [68, 159]. As shown in **Figure 4-11**, our sample size of approximately 50 cases and 50 controls allows us to detect changes in mean methylation of 15% with a power of 50%. For changes in mean methylation below 10% our study was totally underpowered. Lastly, our approach cannot conclusively determine the cause of the observed IVF-associated methylation changes. Future studies of IVF-associated regions in animal models, where genetic differences and infertility diseases can be discarded, could help identify if these changes were caused by IVF itself.

In conclusion, we observed evidence for differences in DNA methylation between IVF and non-IVF twins on a genome-wide scale. A strength of this study design is that it allowed us to also estimate the contribution of genetic and environmental factors towards DNA methylation levels at the IVF-associated loci. The inclusion of only twin pregnancies also avoided biases present in studies that consider single and multiple pregnancies together. Multiple pregnancies are more common after IVF. Therefore, the differences observed when studying singleton and twin births together may be confounded with the higher risks of adverse perinatal outcomes in multiple pregnancy births, rather than IVF itself. Nevertheless, we were unable to dissect whether methylation changes were likely caused by IVF, or were due to the underlying parental subfertility, or other factors. These scenarios require further study exploring the stability of these DMRs over time, their relationship with gene expression, and their potential role in health and disease.

5. DNA methylation changes associated with menopause related traits

5.1 Introduction

Menopause represents the end of the natural reproductive period in women. The timing of this trait, which ranges between 40 and 58 years of age with an average of 51 [160], has been associated with different health outcomes. Early menopause has been associated with lower bone mineral density [161], while a late onset has been associated with increased risk for breast cancer [162]. It is hypothesized that the biological mechanisms behind the associations of osteoporosis and breast cancer with age at menopause are related to estrogen levels [163]. Osteoporosis has been found associated with estrogen deficiencies [164], while breast cancer with higher estrogen exposure [165].

Genetic factors and lifestyle influence the timing of menopause. A twin study estimated a heritability of 63% for age at menopause [166] and genome-wide association studies have identified genetic variants that explain a fraction of this heritability. Stolk *et al*. (2009) identified three loci associated with age at natural menopause using 2,979 European women [167] and He et al. (2009) identified four loci (two of them also identified by Stolk *et al.*) using 9,112 women of European ancestry [168]. Two of these loci, *BRSK1/TMEM150B and MCM8*, were then replicated in Hispanic women (n=3,468) [169]. A larger effort using 38,968 women of European ancestry identified four out of the five previously reported loci and another 13 loci associated with age at menopause (**Table 5-1**) [170]. In replication cohorts, these 17 regions explained from 2.5 to 4.1% of the observed variance in age at natural menopause [170]. The loci associated with age at menopause in women of European ancestry appear to be also relevant for women of other ethnic backgrounds. A study conducted in 6,510 women of

African ancestry replicated six of the loci identified previously in women of European ancestry [171], and did not identify novel associations. In a Chinese sample of 3,533 women, associations of eight SNPs from studies in women of European ancestry were also replicated [172].

Table 5-1. Genome-wide significant SNPs associated with age at menopause in the largest meta-analysis with women of European ancestry (n=38,968)

SNP	Minor/Major allele	MAF	Effect per minor allele (years)	Gene
rs4246511	T/C	0.271	0.289	*RHBDL2*[c,d,]
rs1635501	C/T	0.478	−0.188	*EXO1*
rs2303369	T/C	0.388	−0.174	*FNDC4*
rs10183486	T/C	0.366	−0.219	*TLK1*
rs4693089	G/A	0.486	0.209	*HELQ*
rs890835	A/C	0.112	0.266	*RNF44*
rs365132	T/G	0.49	0.275	*UIMC1*[a,c,d]
rs2153157	A/G	0.492	0.184	*SYCP2L*[a]
rs1046089	A/G	0.353	−0.226	*BAT2*
rs2517388	G/T	0.174	0.274	*ASH2L*[d]
rs2277339	G/T	0.102	−0.394	*PRIM1*[c]
rs3736830	G/C	0.157	−0.243	*KPNA3*
rs4886238	A/G	0.334	0.172	*TDRD3*
rs2307449	G/T	0.405	−0.167	*POLG*[d]
rs11668344	G/A	0.363	−0.416	*TMEM150B*[a,b,c,d]
rs12461110	A/G	0.356	−0.174	*NLRP11*[d]
rs16991615	A/G	0.069	0.971	*MCM8*[a,b,c]

[a]Loci identified by previous smaller studies in women of European ancestry
[b]Loci replicated in Hispanic women
[c]Loci replicated in African-American women
[d]Loci replicated in Chinese women

Despite the shared genetic architecture of age at menopause across different ethnic groups, it has been suggested that ethnicity is an influencing factor in the timing of natural menopause. The Multiethnic Cohort Study in the United States found that, compared to Caucasian, menopause occurs earlier

among Latinas and Native Hawai'ians and later among Japanese Americans [160]. No significant differences were observed between Caucasians and African Americans. Other factors that have also been found to be associated with age at menopause are smoking, age at menarche, parity, and body mass index (BMI). Ex-smokers at age 45 and current smokers at age 45 showed an earlier menopause than non-smokers. Greater age at menarche and greater number of births were seen in women with later age at menarche, and lower BMI was associated with earlier age at menopause [160].

Furthermore, it has been suggested that there may be a secular trend of later age at menopause. A study including women from the population of Gothenburg, Sweden born in 1908, 1914, 1918, 1922, and 1930, suggested and upward trend in the mean age at menopause of 0.1 years per birth year [173]. Another study also observed an increase in the age at menopause looking at 5-year birth cohorts between 1915 and 1939. The mean age at menopause in each cohort was 49.1, 49.5, 50.1, 50.1, and 50.5, respectively [174].

Since DNA methylation profiles can be influenced by genetic variation, external environment, and intrinsic factors, we hypothesized that DNA methylation changes may also be associated with age at menopause or the menopausal transition. Several lines of evidence suggest that DNA methylation changes may occur with menopause. First, DNA methylation is under genetic control, and as such it may be one process mediating the effects of genetic variants identified through GWAS to be associated with age at menopause. Second, DNA methylation has been associated with a number of environmental factors, including smoking [49, 50] and BMI [175], which are both also known to associate with age at menopause. Third, menopause changes the intrinsic organismal environment, specifically circulating hormones [176], which may

impact the level of DNA methylation in blood cell subtypes. Whether DNA methylation changes contribute towards menopause, or whether they are markers of menopause, their identification is of interest both in terms of improving our understanding of the biological mechanisms involved in or with respect to prediction.

Here we set out to identify DNA methylation variation associated with menopause and related traits. We used whole blood DNA methylation profiles characterized with the Infinium HumanMethylation450 Beadchip to conduct epigenome-wide association scans (EWAS) of self-reported age at menopause, early vs late menopause onset, menopausal status, reproductive period (as an indicator of lifetime exposure to endogenous estrogens), and current use of hormone replacement therapy (as an exogenous source of estrogens). We then meta-analyzed EWAS results from five European cohorts.

5.2 Materials and methods

5.2.1 Subjects and phenotype collection

The study was conducted with data from six different cohort members of CHARGE consortium. These included the UK Adult Twin Registry (TwinsUK), the European Prospective Investigation of Cancer (EPIC)-Norfolk, the Avon Longitudinal Study of Parents and Children (ALSPAC), the Cooperative Health Research in the Augsburg Region (KORA), the Rotterdam Study (RS), and the Framingham Heart Study (FHS). All but the FHS cohort were European-based. Further cohort descriptions are provided in Chapter 2.

Five phenotypes related to menopause were explored in this study. These were defined as indicated below:

1. Continuous age at natural menopause (**ANM**) – self-reported age at last menstrual period in women who are reportedly postmenopausal or have not had a menstrual period in the last 12 months.
2. **Duration** of reproductive years – the continuous length of time between menarche, defined as the self-reported age at first menstrual period, and menopause, as defined in 1 above.
3. **Early vs late menopause** – subject were classified into two groups, where those who underwent early menopause were defined as individuals with age at natural menopause between 40-44.99 years, and those in the late menopause category were defined as subjects with age at menopause between 55-60 years.
4. **Menopausal status** at time of DNA sampling– subjects were classified into two groups, where the first group were premenopausal women who have not yet undergone menopause. The second group included peri/postmenopausal women, where perimenopausal was defined as < 2 years since last menstrual period, and postmenopausal being > 2 years since last menstrual period.
5. Hormone Replacement Therapy (**HRT) use** – was explored in postmenopausal women only and these were classified in two groups. The first group included postmenopausal women on self-reported HRT at time of DNA sampling, and the second group included postmenopausal HRT non-users.

In all cohorts we also excluded subjects according to the following criteria:

1. Individuals with age at menopause < 40 years or > 60 years
2. Individuals with medically (treatment or surgically) induced menopause

3. Individuals using HRT before menopause
4. Individuals with age at menarche < 8 years or > 18 years

Phenotype data were obtained from self-reported questionnaires in each cohort, and quality control checks at the phenotype level data were performed within cohorts by each lead cohort analyst in the CHARGE consortium according to phenotype definitions provided by us.

Since these variables had not been explored before in the TwinsUK sample with DNA methylation data, I went through multiple longitudinal electronic records collected between 1992 and 2013 to determine age at menarche, age at menopause, menopausal status at DNA sampling, and HRT use onset. When possible, age at menopause was verified against longitudinal records of menstrual periods and menopausal status. It was verified that no menstrual period was recorded after the self-reported age at menopause and that the participant consistently reported a premenopausal status or postmenopausal status before and after, respectively, the self-reported age at menopause.

The final number of participants per cohort and per phenotype used in downstream analyses is shown in **Table 5-2**.

Table 5-2. Total number of participants per cohort

Cohorts	Age at menopause	Duration of reproductive period	Early vs late menopause	Menopausal status	HRT use
Discovery					
TwinsUK	269	225	-	319	171
EPIC-Norfolk	306	420	75	546	348
ALSPAC	65	65	-	514	577
KORA	517	517	340	761	517

RS	209	201	59	393	-
Total	**1,366**	**1,428**	**474**	**2,533**	**1613**
Replication					
FHS	963	899	160	-	938

5.2.2 DNA methylation data

DNA was extracted from whole blood samples. Genome-wide methylation data was obtained with the Infinium HumanMethylation450 BeadChip. Intensity signals were background corrected and normalized before the estimation of DNA methylation levels. DNA methylation levels were expressed as beta-values (ratio of the methylated probe intensity signal over the sum of the methylated and unmethylated probe intensity signals). Probes with detection p-value > 0.01 in more than 5% of the samples and samples with > 5% of probes with detection p-values > 0.01 were excluded from the analysis. Additional probe filtering included exclusion of probes that map to multiple regions of the *in silico* bisulfite converted human genome allowing up to two mismatches at any position and exclusion of probes that target a polymorphic CpG site, including the extension base in Type I probes, with a minor allele frequency > 5% in the UK10K haplotype reference panel [99]. DNA methylation data were also used to estimate blood cell proportions ($CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killers, monocytes, and granulocytes) [177] within each cohort.

5.2.3 Statistical analyses in TwinsUK data

The observed variance in age at menopause was partitioned into additive genetic (A), common environmental (C), and unique environmental (E) factors using the ACE model [105]. The model was fitted using the OpenMX statistical package [106] in R [107].

5.2.4 meQTLs analysis

Previously reported GWAS hits for age at menopause (**Table 5-1**) were searched against the meQTLs identified in Chapter 3 of this thesis.

5.2.5 EWAS

EWAS were performed within each cohort and results were meta-analyzed across cohorts. EWAS within cohort were performed by the lead-analyst representing the cohort in the CHARGE epigenetics working group consortium. The lead analysts were Carola Marzi (KORA), Michael Mendelson (FHS), Alexia Cardona (EPIC-Norfolk), Cindy Boer (Rotterdam Study), and myself (TwinsUK).

I provided the EWAS pipeline for this analysis with inputs from Dr. Ken Ong (University of Cambridge), Dr. Cathy Elks (University of Cambridge), Dr. Michael Mendelson (Framingham Heart Study), and Dr. Jordana Bell (King's College London). Briefly, the EWAS pipeline first outlines data quality control steps and probe exclusion criteria as outlined in section 5.2.2 above. Following quality control, DNA methylation beta-values at each CpG-site were transformed to standard normal form N(0,1) prior to analysis. Linear mixed models were used to test the association between DNA methylation and each phenotype. Methylation was used as the response variable and phenotype, technical covariates, biological covariates, and family structure, if appropriate, were used as predictors. Technical covariates (plate and position in the plate) were included as random effects. Family structure was also included as random effects in studies with family cohort designs, in this case TwinsUK and FHS. Biological covariates such as age, smoking status (current, never, or former) at the time of blood collection, and estimated whole blood cell proportions ($CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killers, monocytes, and granulocytes) were included in the model as fixed effects.

5.2.6 Meta-analysis

Following individual-cohort EWAS analysis, the results were uploaded to a web-based file-sharing system and I performed the meta-analysis. Individual-cohort results were checked in terms of correlation between reported p-values and expected p-values, skewness and kurtosis of the estimates, distribution of p-values, and genomic inflation factor. Manhattan plots per cohort are shown in Annex IV-VIII. An inverse variance-weighted random-effects meta-analysis implemented with GWAMA [178] was performed with statistics from the five European cohorts.

5.3 Results

5.3.1 Age at menopause is a heritable trait

Using a sample of 582 twins from TwinsUK (291 MZ and 29 DZ twin pairs) we partitioned the variance in age at menopause into additive genetic, common environmental, and unique environmental factors using the ACE model in OpenMX [106]. We estimated that the additive genetic component accounted for 53% of the observed variance, while the rest was attributed to unique environmental factors.

5.3.2 Age at menopause is associated with smoking status

It is known that tobacco smoking and BMI are associated with DNA methylation at multiple loci in the genome [49, 175, 179]. Associations of smoking and BMI with age at menopause have also been reported [160]. Given that smoking or BMI may confound the association between DNA methylation and age at menopause, we investigated in TwinsUK data if that association was observed. We did observe a significant association (P=0.015) between smoking and age at

menopause in a sample of 582 twins (**Figure 5-1**). Length of reproductive period and use of HRT were not significantly associated with smoking (p>0.05), and no significant association was observed between age at menopause, length of reproductive period, or use of HRT with BMI (p>0.05).

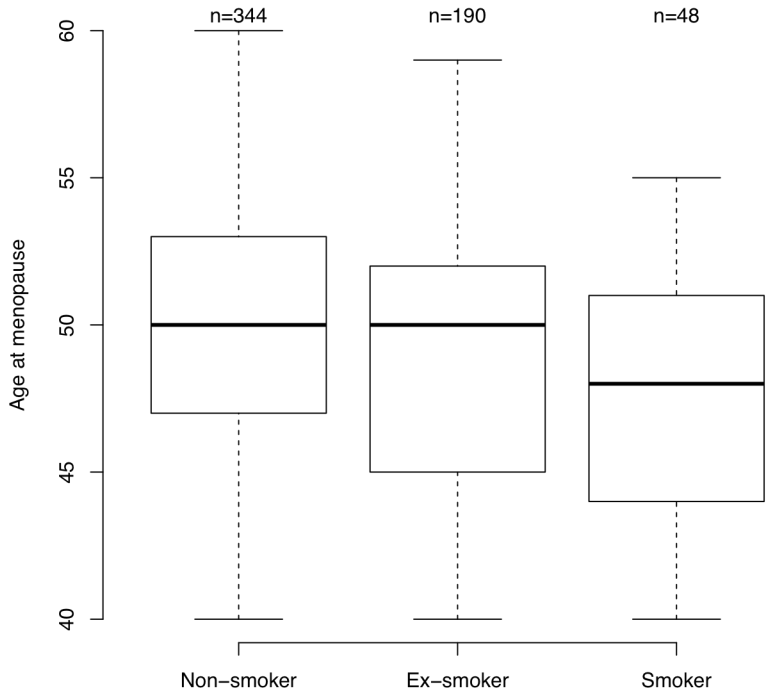


Figure 5-1. Age at menopause is associated with smoking status. Boxplot showing age at menopause in non-smokers, ex-smokers, and smokers (n=582).

5.3.3 GWAS hits for age at menopause are meQTLs

The lead SNP at each of the 17 loci identified through GWAS to be associated with age at menopause (**Table 5-1**) were searched against the list of *cis* meQTLs and *trans* meQTLs generated in Chapter 3 and in a database of meQTLs generated with a greater sample size [95]. In total, I found evidence of meQTL effects, either in cis or trans, for 13 out of the 17 SNPs (**Table 5-3**).

Table 5-3. GWAS hits for age at menopause and influenced CpG sites

SNP	CpG sites	CpG site Genes
rs1635501	cg04706276; cg16121177; cg13345151; cg03450102; cg08574246	*EXO1*

rs3736830	cg20399252; cg20826914; cg04663916; cg13070676	*EBPL*
rs4886238	cg16113134; cg01612771	
rs2307449	cg01877814; cg24365117; cg27640114; cg18831671; cg02557652	*POLG; FANCI; ABHD2*
rs11668344	cg16555964; cg14496450; cg26104073	*TMEM150B; HRNBP3;*
rs10183486	cg07673935; cg08741063	*TLK1*
rs2303369	cg14021192; cg12648201; cg21248554; cg12000995; cg15118510; cg04845466; cg05102552; cg26034919; cg24768116; cg05696406; cg17158414; cg05484376; cg13211152; cg11618577; cg20102877; cg26350635; cg01969012; cg10337841; cg07881013; cg04015759; cg20744217; cg22608170; cg03929741; cg18886436	*PPM1G; FTHL3; KRTCAP3; IFT172; NRBP1; SNX17; FNDC4; MRPL33; CORO1B; MAPK4; EFNB2; ZNF385A*
rs4693089	cg13524797	*AGPAT9*
rs365132	cg17809377; cg06783121; cg06882562; cg08690709; cg14100313; cg13798109; cg17143007	*HK3; UIMC1*
rs2153157	cg23049448; cg09105523; cg08480294	*SYCP2L*
rs1046089	cg17006042; cg20794828; cg03144490; cg24239961; cg25769566; cg04835051; cg00586094; cg17494781; cg13740929; cg07845406; cg06460587; cg22318514; cg18264486; cg04736217; cg19563932; cg24208375; cg07455790; cg14110444; cg17731470; cg18576957; cg02304584; cg13892322; cg05182583; cg05857999; cg11540476; cg09993780; cg26821115; cg08624648; cg17552909; cg18808760; cg01620082; cg06606381; cg17862947; cg01255021; cg17841099; cg00624589; cg22322277; cg02753903; cg06099246; cg23331010; cg07575812; cg18651192; cg04788957; cg15007120; cg19589727; cg00943909	*BAT3; LY6G6C; CSNK2B; LY6G5B; HSPA1B; LSM2; HSPA1A; HSPA1L; LY6G5C; C6orf48; SNORD52; BAT2; C6orf25; CLIC1; AIF1; FBRSL1; SLC39A1; CREB3L4; RYR1; STG3GAL6; SH3PXD2A; FLJ25363; GNAS;*
rs2277339	cg11592377; cg26545245	*PRIM1*
rs16991615	cg23629183; cg23636682	*MINA; FBXO22*

5.3.4 Differentially methylated positions associated with current HRT use in five European cohorts

Meta-analysis of the five European cohorts in the discovery set (TwinsUK, EPIC-Norfolk, ALSPAC, KORA, and Rotterdam Study) did not show any genome-wide significant differentially methylated position (DMP) for age at menopause,

duration of reproductive period, early vs late menopause, or menopausal status (data not shown).

The meta-analysis for use of HRT identified seven genome-wide significant DMPs at a false discovery rate (FDR) of 5% (**Figure 5-2**, **Figure 5-3**, **Table 5-4**). Funnel plots for the seven DMPs are shown in Annex IX.

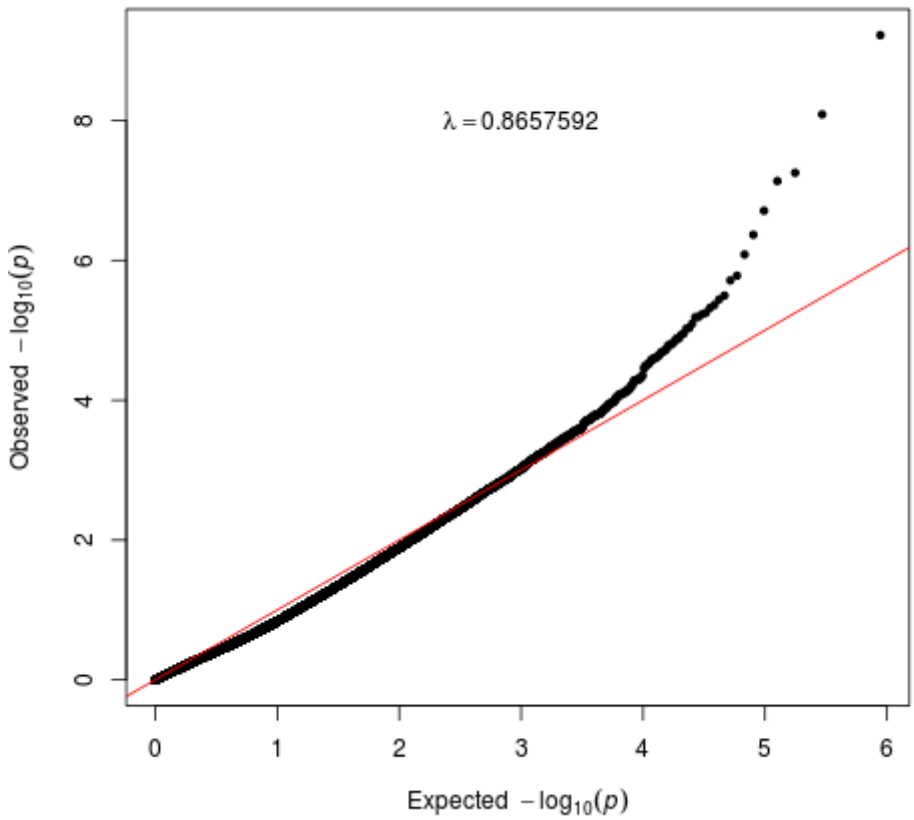


Figure 5-2. HRT use meta-analysis EWAS Q-Q plot. Methylation levels at 441,568 CpG sites were tested for an association with current use of HRT. Results are based on analysis of 1,613 samples from four cohorts.

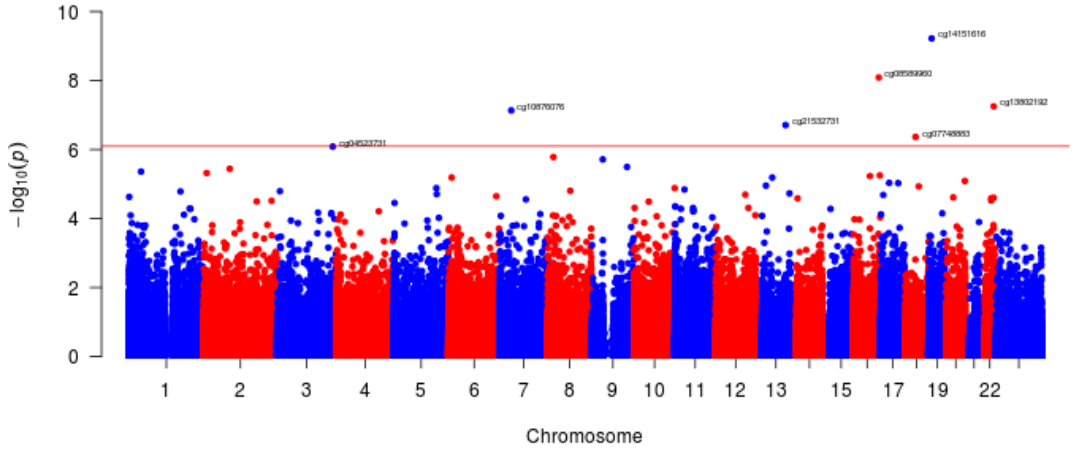


Figure 5-3. HRT use EWAS Manhattan plot. At a false discovery rate (FDR) of 5%, seven CpG sites were identified as HRT-associated DRMs.

Table 5-4. HRT-associated DMPs

CpG site	beta	se	p	i2	n	Effects	Chr	Position	Gene	Location	CpG island
cg14151616	-0.007	0.001	6.00E-10	0	1,096	---?	19	10,426,492	*FDX1L*	Body	Island
cg08589960	-0.041	0.007	8.16E-09	0	1,096	--+?	16	83,986,755	*OSGIN1*	TSS200;5'UTR;Body	Open sea
cg13802192	-0.006	0.001	5.58E-08	0	1,612	--+-	22	46,692,554	*GTSE1*	TSS200	Island
cg10876076	-0.012	0.002	7.32E-08	0	1,613	----	7	37,960,873	*EPDR1*	Body	Island
cg21532731	-0.014	0.003	1.95E-07	0	1,096	---?	13	99,305,879	-	NA	Open sea
cg07748883	0.028	0.006	4.30E-07	0	1,613	+--+	18	34,914,606	*BRUNOL4*	Body	N_Shelf
cg04523731	0.014	0.003	8.23E-07	0	1,613	++++	3	185,212,897	*TMEM41A*	Body	N_Shelf

We then looked for evidence of genetic influence on the methylation levels of the identified HRT-associated DMPs using the estimates presented in Chapter 3 of this thesis. Only cg04523731 showed evidence of very low heritability and of being affected by genetic variants (**Table 5-5**).

Table 5-5. Genetic influences on HRT-associated DMPs

CpG site	Heritability	meQTL
cg14151616	0.00	-
cg08589960	0.00	-
cg13802192	0.00	-
cg10876076	0.00	-
cg21532731	0.00	-
cg07748883	0.00	-
cg04523731	0.04	rs7619196

We attempted replication of the seven HRT-associated DMPs with data from the Framingham Heart Study (FHS), an American cohort of mostly European ancestry. We did not observe significant differences at any of the HRT-associated DMPs in the replication cohort **Table 5-6**.

Table 5-6. Replication of HRT-associated DMPs

CpG site	beta	se	t-statistic	p	n
cg14151616	0.000	0.001	-0.202	0.840	938
cg08589960	0.000	0.001	0.238	0.812	938
cg13802192	0.001	0.001	0.748	0.455	938
cg10876076	0.002	0.001	1.803	0.072	938
cg21532731	0.000	0.001	0.193	0.847	938
cg07748883	0.002	0.004	0.571	0.569	938
cg04523731	0.002	0.003	0.788	0.431	938

5.4 Discussion

Given that age at menopause is a highly heritable trait and at least 17 loci have been associated with age at menopause through GWAS [170], we hypothesized that the biological mechanisms of some of these genetic variants influencing the timing of menopause may be mediated by DNA methylation. Out of the 17 GWAS loci associated with age at menopause, 13 were meQTLs with *cis* or *trans* effects. The target CpG sites of these meQTLs were located in genes such as *EBPL*, gene potentially involved in steroid biosynthesis [159], *ABHD2* and *AGPAT9*, genes also associated with obesity-related traits [180], and *PPM1G*, gene also associated with hypertriglyceridemia [181]. This observation suggest that lipid metabolism is an important component of age at menopause. Whether DNA methylation plays a functional role in age at menopause through lipid metabolism is still unknown and should be explored further.

The evidence from multiple European cohorts did not support our hypothesis. The EWAS approach did not identify any genome-wide significant DMP associated with age at menopause in blood using a sample size of 1,366 suggesting that either timing of menopause is not mediated by DNA methylation, or that the effects are not observed in blood, or that our study did not have enough power to detect the differences in DNA methylation. Similarly, we were not able to identify differential methylation associated with extreme ages at natural menopause (early vs late), duration of the reproductive period, or menopausal status. Having in mind that blood may not be the most relevant tissue to look at changes associated with menopause, we could test the hypothesis that other tissues more relevant to the metabolic changes observed during and after menopause, such as adipose or bone tissue, show DNA methylation changes even though these are not observed in blood. I will follow

up this study on adipose tissue considering that sample availability will not be as high as it was with blood. In this new scenario sample size will limit our power to detect methylation changes, however we expect these changes to be large enough to be identified.

Unlike the rest of the studied phenotypes, HRT-use showed association with DNA methylation levels in blood. However, these changes were observed in the European discovery set, but not in the American replication cohort. It is important to highlight that there are different types of HRT and the lack of replication could indicate a difference in the type of HRT prescribed in Europe and the United States. We also did not observe strong genetic influences on the methylation levels of HRT-associated DMPs, suggesting that these differences might be induced by the use of HRT and are not confounded by genetics.

HRT is used to treat the deficiency of estrogen, which is common in postmenopausal women. Estrogens are known to play a role in reproductive function, but are not limited to it; they also have neuroprotective, antithrombotic, and lipid lowering effects [182]. Studies looking at the effects of HRT are inconclusive, but have suggested a protective role against osteoarthritis [183], cognitive decline [184], and muscle weakness [185]. However, other studies have associated the use of HRT with the late onset of diseases such as asthma [186]. One study has looked for differences in DNA methylation and gene expression in blood of MZ twins discordant for HRT use and identified changes in both methylation and expression at five genes, all of which have been previously implicated with bone mineral content or body adiposity [187]. None of the HRT-associated DMPs identified in this chapter overlapped with those identified in discordant MZ twins, however the metabolic changes due to estrogens are vastly and better characterization of HRT prescriptions are needed

to better assess their effects. DNA methylation levels in whole blood cells might be susceptible to HRT use given that circulating levels of hormones are modified. A closer look at the specific hormones producing the changes in DNA methylation might be more informative of the biological mechanisms involved.

6. Conclusions

Since the post-genomic era, genetic epidemiology has been successful identifying genetic variants that explain a proportion of the variation in human traits and diseases. However, for the large majority of human traits a large proportion of the phenotypic variance has not yet been explained. A significant component of this unexplained phenotypic variance is likely to be attributed to environmental factors that have not yet been accounted for. Common complex diseases, as most human traits, are influenced by both genetic and environmental factors. Epigenetics offers a new framework to look at complex traits and diseases in order to understand the biological consequences of genetic variation, environmental exposures, and the interaction between the two. This thesis aimed to characterize genetic and environmental influences on the epigenome by performing twin studies on DNA methylation profiles. Three main results were obtained.

In the first part of this thesis as I was able to characterize the proportion of the variance in DNA methylation attributed to genetic and environmental factors at almost 800,000 CpG sites distributed across the genome. Furthermore, I was able to identify functional regions of the genome that are more susceptible to genetic influences, such as enhancers. I also identified common genetic variants that impact DNA methylation levels. We observed that at least 10% of the interrogated CpG sites have a strong genetic component on DNA methylation (h^2 > 40%), and in almost half of these heritable CpG sites we were able to identify meQTLs.

In the second part of this thesis, I presented a study looking at the effect of assisted reproductive technology on DNA methylation. Infertility and subfertility are common conditions in human populations and a global public health issue as

recognized by the World Health Organization [188]. The 12-month prevalence of infertility in developed countries has been estimated at 9% and it has also been estimated that more than half of the affected population seeks medical help [189]. Advances in the manipulation of the fertilization process have allowed the development of *in vitro* fertilization (IVF) and related techniques, such as, intracytoplasmic sperm injection (ICSI). The first baby conceived with the use of IVF was born in 1978 and since then IVF has been satisfactorily adopted as a fertility treatment. Data from the European IVF-monitoring Consortium from 1997 to 2010 has shown an uninterrupted increase in the number of IVF cycles per year [190]. From these data, it is also estimated that just in Europe the number of IVF births between 2003 and 2010 is near to 600,000. In 2010, it meant that IVF accounted for 0.6% to 5.9% of births in different countries of Europe [190]. My work explored the hypothesis that IVF introduces DNA methylation changes that persist after birth given that the treatment overlaps with an important period of epigenetic reprogramming. Our study was the first large-scale epigenome-wide analysis of IVF of its kind to date, and with a sample size larger than most of other studies looking at IVF. We showed that differences in DNA methylation of blood cells are very small and are likely not exclusively due to the fertility treatment.

The last part of the thesis looked at epigenetic changes associated with menopause and related phenotypes in women. We observed that the majority (13 out of 17) of loci previously associated through GWAS with age at menopause had an effect on the methylation levels of multiple CpG sites. However, the only differentially methylated positions (DMPs) that were identified in blood through EWAS, were associated with the use of hormone replacement therapy (HRT). HRT is used to treat estrogen deficiencies in postmenopausal

women and modifies the levels of circulating hormones, which might result in DNA methylation changes in blood cells.

6.1 Strengths and limitations

Of the main strengths of this thesis was the use of twins as a unique study design in human population analyses of DNA methylation as a complex trait. Twins were a valuable tool exploited in every results chapter presented in this thesis. In Chapter 3 the use of twins allowed the estimation of heritability of DNA methylation. In Chapter 4, the use of twins allowed to control for the effect of multiple pregnancy, which is a common confounder in studies seeking to compare natural conceptions with those achieved after assisted reproductive technology. In addition, twins also allowed to assess the contribution of genetics and environment to methylation at the suggestive IVF-DMRs identified. Lastly, in Chapter 5, twins were helpful to corroborate that age at menopause is a heritable trait.

With regard to DNA methylation profiling techniques, we relied on two main technologies with different strengths and limitations. One was microarray based technology (450K and EPIC arrays), which has the advantage of single base resolution, but a limited coverage (approximately 450,000 and 850,000 CpG sites). The 450K array also has the advantage of being vastly used to profile DNA methylation in a great number of cohorts, which allowed us to conduct the largest study to date looking at menopause and related phenotypes such as HRT. In Chapter 3, the greater coverage of the EPIC array was a strength that allowed the investigation of genetic impacts on enhancer regions that have not been explored before in any other study. The other DNA methylation profiling technique was MeDIP-seq, which has genome-wide coverage, but lacks single

base resolution. The use of this technique in Chapter 4 allowed us to look at regions not covered by microarray-based platforms.

There were also several limitations of the work presented in this thesis. First, despite the strengths of the twin model, it has been suggested that the twin-based epigenetic heritability estimation is affected by the fact that MZ twins have more similar epigenetic starting points than DZ twins [191]. However, we were able to validate some of the twin-based heritability estimated genetic effects with a methodology that did not depend on the assumptions of the classical twin model, that is, by the identification of meQTLs.

The meQTL study did rely on another assumption, that MZ co-twins have identical DNA sequences. Although MZ twins are often referred to as being genetically identical, post-zygotic mutation events can occur. Somatic point mutations and copy number variations have been found in normal concordant as well as in disease-discordant MZ twin pairs [192]. However, these events are extremely rare and should not hinder the identification of meQTLs. Somatic point mutations during early development occur with a frequency of 1.2×10^{-7} per base pair per twin pair [193].

Another limitation was sample size, which usually results in limited statistical power to detect heritability or differential methylation. Visscher (2004) revisited the statistical power of the classical twin design to detect heritability (h^2) under the ACE model [194]. To detect high heritability ($h^2 > 0.4$) with power of 95% when $C = 0.5$ and using an equal proportion of MZ and DZ twin pairs, a total of 74 twin pairs are needed. However, the contribution of common environmental factors is usually much lower than 50%. Under a more realistic scenario, to detect $h^2 > 0.4$ with power of 95% when $C = 0.1$, a total of 650 twin pairs are needed. Power calculations have also been performed for the detection of differential

methylation in the context of case-control EWAS [195]. It has been estimated that a total of 20 case-control pairs at single-locus significance and 112 case-control pairs at genome-wide significance are needed to reach power of 80% to detect a 0.1 mean methylation difference. The methylation differences at the two validated CpG sites located in the *C9orf3* IVF-DMR using a subset of MZ twins (n=22) were of only 0.09 and 0.012, which gives a power of 20% at single-locus significance.

We did not identify any genome-wide significant DMP associated with age at menopause possibly because our study did not have enough power to detect the differences in DNA methylation levels at the CpG-sites targeted by the Illumina 450K array. GWAS signals for age at menopause have been identified using samples of at least 3,000 and up to almost 39,000 individuals, therefore these are relatively modest genetic effects and maximizing power is of key concern when conducting an EWAS. Similarly, negative results were found for duration of reproductive period, early vs late menopause, and menopausal status. Another factor that might hinder the identification of DMPs is data collection. For most cohorts, data was self-reported and susceptible to recall bias.

It is also important to state that the studies presented in this thesis were looking at DNA in blood samples and given the cell-type specificity of epigenetic marks, these results are not necessarily true for other tissues. The advantages of looking at peripheral blood are the ease of its collection and the potential clinical use of the identified DMRs or DMPs as biomarkers. Also because of the cell-type specificity of epigenetic marks and because blood is a mixture of cell types, adjustment for cell heterogeneity was performed in every study. We did not use actual cell counts, but instead estimated their proportions from epigenetic data which has proven to be a good proxy [177].

6.2 Future perspectives

Twins are a unique resource in the field of epigenetics and essential tools to elucidate the molecular etiology of diseases. Most findings so far come from studying unrelated individuals, however a number of studies have also explored epigenetic profiles in twins during normal development, ageing, and in the context of disease. Some examples include the identification of methylation changes at *DOK7* in breast cancer, *TRPA1* in pain susceptibility, *ST6GALNAC1* in bipolar disorder, *ZBTB20* in major depressive disorder, and several genes identified in autism spectrum disorder. One of the difficulties that twin studies face is the collection of enough individuals with the specific trait to be studied. Collaborations with other twin cohorts could help overcome this problem and allow the use of twin models that were not exploited in this thesis such as the discordant MZ twins design [73].

The IVF study in particular could benefit from follow-up studies such as integration of gene expression data and assessment of the longitudinal stability of methylation at the suggestive IVF-DMR. The PETS cohort has collected samples of peripheral blood from the twins at 18 months and 5 years of age, which would allow this exploration in the future. In addition, given that the first stage in most IVF cycles is ovulation induction with hormones and given that HRT-associated DMPs were identified, it would be interesting to explore whether there are persistent effects on DNA methylation patterns of the mothers due to ovulation induction treatments.

For the study looking at menopause, an interesting follow-up would be to look at relevant tissues such as fat or bone, as it is known that menopause is characterized by drastic metabolic changes associated with lipids and bone mineral density [196].

It would be worth looking also at more environmental exposures to identify changes in DNA methylation such as the ones identified for HRT use. We cannot assume causality in this particular study, but assuming there is, DNA methylation could be used to identify past environmental exposures, which has a potential use in personalized medicine.

Lastly, although we investigated the genetic influences with a two-fold approach, validated two IVF-DMRs with a different technology, and identified HRT-associated DMPs through meta-analysis, these studies still require replication in independent cohorts.

7. References

1. Waddington CH. The epigenotype. Endeavour. 1942;1:18–20.
2. Holliday R. Epigenetics: an overview. Dev Genet. 1994;15:453–7. doi:10.1002/dvg.1020150602.
3. Holliday R. Epigenetics: A Historical Overview. Epigenetics. 2006;1:76–80. doi:10.4161/epi.1.2.2762.
4. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. 2014;28:812–28. doi:10.1101/gad.234294.113.
5. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301:89–92. doi:10.1038/301089a0.
6. Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004;4:143–53. doi:10.1038/nrc1279.
7. Ballestar E. Epigenetics lessons from twins: prospects for autoimmune disease. Clin Rev Allergy Immunol. 2010;39:30–41. doi:10.1007/s12016-009-8168-4.
8. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet. 2011;7:e1002300. doi:10.1371/journal.pgen.1002300.
9. Sanchez-Mut J V., Aso E, Panayotis N, Lott I, Dierssen M, Rabano A, et al. DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. Brain. 2013;136:3018–27. doi:10.1093/brain/awt237.
10. Masliah E, Dumaop W, Galasko D, Desplats P. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. Epigenetics. 2013;8:1030–8. doi:10.4161/epi.25865.
11. Salam MT, Zhang Y, Begum K. Epigenetics and childhood asthma: current evidence and future research directions. Epigenomics. 2012;4:415–29. doi:10.2217/epi.12.32.
12. Kaminsky Z, Petronis A, Wang S-C, Levine B, Ghaffar O, Floden D, et al. Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. Twin Res Hum Genet. 2008;11:1–11. doi:10.1375/twin.11.1.1.
13. Heyn H, Ferreira HJ, Bassas L, Bonache S, Sayols S, Sandoval J, et al. Epigenetic disruption of the PIWI pathway in human spermatogenic disorders. PLoS One. 2012;7:e47892. doi:10.1371/journal.pone.0047892.
14. Zhang R, Miao Q, Wang C, Zhao R, Li W, Haile CN, et al. Genome-wide DNA methylation analysis in alcohol dependence. Addict Biol. 2013;18:392–403. doi:10.1111/adb.12037.
15. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science (80-). 1975;187:226–32. http://www.ncbi.nlm.nih.gov/pubmed/1111098.
16. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Genome Res. 1975;14:9–25. doi:10.1159/000130315.
17. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. 1989;3:612–9. http://www.ncbi.nlm.nih.gov/pubmed/2545524. Accessed 25 Feb 2014.
18. Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell. 1991;64:1123–34.

http://www.ncbi.nlm.nih.gov/pubmed/2004419. Accessed 25 Feb 2014.
19. Nan X, Campoy FJ, Bird A. MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. Cell. 1997;88:471–81. doi:10.1016/S0092-8674(00)81887-5.
20. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol. 2009;27:361–8. doi:10.1038/nbt.1533.
21. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403:41–5. doi:10.1038/47412.
22. Sullivan KF, Hechenberger M, Masri K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol. 1994;127:581–92. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2120219&tool=pmcentrez&rendertype=abstract. Accessed 25 Feb 2014.
23. Barrero MJ, Sese B, Martí M, Izpisua Belmonte JC. Macro histone variants are critical for the differentiation of human pluripotent cells. J Biol Chem. 2013;288:16110–6. doi:10.1074/jbc.M113.466144.
24. Hirschhorn JN, Brown SA, Clark CD, Winston F. Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. 1992;6:2288–98. http://www.ncbi.nlm.nih.gov/pubmed/1459453. Accessed 25 Feb 2014.
25. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell. 2008;31:785–99. doi:10.1016/j.molcel.2008.09.003.
26. Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322:750–6. doi:10.1126/science.1163045.
27. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38:23–38. doi:10.1038/npp.2012.112.
28. Lövkvist C, Dodd IB, Sneppen K, Haerter JO. DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Res. 2016;44:5123–32. doi:10.1093/nar/gkw124.
29. Vertino PM, Sekowski JA, Coll JM, Applegren N, Han S, Hickey RJ, et al. DNMT1 is a component of a multiprotein DNA replication complex. Cell Cycle. 2002;1:416–23. doi:10.4161/cc.1.6.270.
30. Auclair G, Weber M. Mechanisms of DNA methylation and demethylation in mammals. Biochimie. 2012;94:2202–11. doi:10.1016/j.biochi.2012.05.016.
31. Bhutani N, Burns DM, Blau HM. DNA demethylation dynamics. Cell. 2011;146:866–72. doi:10.1016/j.cell.2011.08.042.
32. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. The DNA methylation landscape of human early embryos. Nature. 2014;511:606–10. doi:10.1038/nature13544.
33. Tang WC, Dietmann S, Irie N, Leitch H, Floros V, Bradshaw C, et al. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell. 2015;161:1453–67. doi:10.1016/j.cell.2015.04.053.
34. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500:477–81. doi:10.1038/nature12433.
35. Martino D, Loke YJ, Gordon L, Ollikainen M, Cruickshank MN, Saffery R, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. Genome biology. 2013;14:R42. doi:10.1186/gb-2013-14-5-r42.

36. Florath I, Butterbach K, Müller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. Hum Mol Genet. 2014;23:1186–201. doi:10.1093/hmg/ddt531.
37. Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 2012;8:e1002629. doi:10.1371/journal.pgen.1002629.
38. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, et al. Intra-individual change over time in DNA methylation with familial clustering. JAMA. 2008;299:2877–83. doi:10.1001/jama.299.24.2877.
39. Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AHC, et al. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009;41:240–5. doi:10.1038/ng.286.
40. Friso S, Girelli D, Trabetti E, Olivieri O, Guarini P, Pignatti PF, et al. The MTHFR 1298A>C polymorphism and genomic DNA methylation in human lymphocytes. Cancer Epidemiol Biomarkers Prev. 2005;14:938–43. doi:10.1158/1055-9965.EPI-04-0601.
41. El-Maarri O, Kareta MS, Mikeska T, Becker T, Diaz-Lacava A, Junen J, et al. A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. Hum Mol Genet. 2009;18:1755–68. doi:10.1093/hmg/ddp088.
42. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai S-L, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 2010;6:e1000952. doi:10.1371/journal.pgen.1000952.
43. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. Am J Hum Genet. 2010;86:411–9. doi:10.1016/j.ajhg.2010.02.005.
44. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12:R10. doi:10.1186/gb-2011-12-1-r10.
45. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet. 2013;93:876–90. doi:10.1016/j.ajhg.2013.10.004.
46. Shi J, Marconett CN, Duan J, Hyland PL, Li P, Wang Z, et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nat Commun. 2014;5:3365. doi:10.1038/ncomms4365.
47. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet. 2011;88:450–7. doi:10.1016/j.ajhg.2011.03.003.
48. Monick MM, Beach SRH, Plume J, Sears R, Gerrard M, Brody GH, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet. 2012;159B:141–51. doi:10.1002/ajmg.b.32021.
49. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. PLoS One. 2013;8:e63812. doi:10.1371/journal.pone.0063812.
50. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated

with smoking. Hum Mol Genet. 2013;22:843–51. doi:10.1093/hmg/dds488.
51. Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect. 2012;120:1425–31. doi:10.1289/ehp.1205412.
52. Markunas C a, Xu Z, Harlid S, Wade P a, Lie RT, Taylor J a, et al. Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy. Environ Health Perspect. 2014; November 2013. doi:10.1289/ehp.1307892.
53. Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SRH. The impact of recent alcohol use on genome wide DNA methylation signatures. Front Genet. 2012;3:54. doi:10.3389/fgene.2012.00054.
54. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P-C, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2016; February:1–12. doi:10.1038/mp.2016.192.
55. Ek WE, Tobi EW, Ahsan M, Lampa E, Ponzi E, Kyrtopoulos SA, et al. Tea and coffee consumption in relation to DNA methylation in four European cohorts. Hum Mol Genet. 2017. doi:10.1093/hmg/ddx194.
56. Dominguez-Salas P, Moore SE, Baker MS, Bergen AW, Cox SE, Dyer RA, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. Nat Commun. 2014;5:3746. doi:10.1038/ncomms4746.
57. Aslibekyan S, Wiener HW, Havel PJ, Stanhope KL, O'Brien DM, Hopkins SE, et al. DNA methylation patterns are associated with n-3 fatty acid intake in Yup'ik people. J Nutr. 2014;144:425–30. doi:10.3945/jn.113.187203.
58. Voisin S, Almén MS, Moschonis G, Chrousos GP, Manios Y, Schiöth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. Eur J Hum Genet. 2014. doi:10.1038/ejhg.2014.139.
59. Binder AM, Michels KB. The causal effect of red blood cell folate on genome-wide methylation in cord blood: a Mendelian randomization approach. BMC Bioinformatics. 2013;14:353. doi:10.1186/1471-2105-14-353.
60. Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J. 2014. doi:10.1096/fj.13-249029.
61. Kingsley SL, Kelsey KT, Butler R, Chen A, Eliot MN, Romano ME, et al. Maternal serum PFOA concentration and DNA methylation in cord blood: A pilot study. Environ Res. 2017;158:174–8. doi:10.1016/j.envres.2017.06.013.
62. Dugu?? PA, Geurts YM, Milne RL, Lockett GA, Zhang H, Karmaus W, et al. Is there an association between season of birth and blood DNA methylation in adulthood? Allergy: European Journal of Allergy and Clinical Immunology. 2016;71:1501–4. doi:10.1111/all.12949.
63. Lockett GA, Soto-Ram?rez N, Ray MA, Everson TM, Xu C-J, Patil VK, et al. Association of season of birth with DNA methylation and allergic disease. Allergy. 2016;71:1314–24. doi:10.1111/all.12882.
64. Lim ASP, Klein H-U, Yu L, Chibnik LB, Ali S, Xu J, et al. Diurnal and seasonal molecular rhythms in human neocortex and their relation to Alzheimer's disease. Nat Commun. 2017;8:14931. doi:10.1038/ncomms14931.
65. Gruzieva O, Xu C-J, Breton C V., Annesi-Maesano I, Ant? JM, Auffray C, et al. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. Environ Health Perspect. 2016;125:104–10.

doi:10.1289/EHP36.
66. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008;9:255–66. doi:10.1038/nrg2322.
67. Falconer DS. Introduction to Quantitative Genetics. Glasgow: Robert MacLehose and Company Limited; 1960. doi:10.1002/bimj.19620040211.
68. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41. doi:10.1038/nrg3000.
69. Tsai P-C, Spector TD, Bell JT. Using epigenome-wide association scans of DNA methylation in age-related complex human traits. Epigenomics. 2012;4:511–26. doi:10.2217/epi.12.45.
70. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods. 2013;10:949–55. doi:10.1038/nmeth.2632.
71. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012;336:934–7. doi:10.1126/science.1220671.
72. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33:5868–77. doi:10.1093/nar/gki901.
73. Bell JT, Spector TD. A twin approach to unraveling epigenetics. Trends Genet. 2011;27:116–25. doi:10.1016/j.tig.2010.12.005.
74. Spector TD, Williams FMK. The UK Adult Twin Registry (TwinsUK). Twin Res Hum Genet. 2006;9:899–906. doi:10.1375/183242706779462462.
75. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). Twin Res Hum Genet. 2013;16:144–9. doi:10.1017/thg.2012.89.
76. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. Int J Epidemiol. 2013;42:76–85. doi:10.1093/ije/dyr207.
77. Spector TD, MacGregor AJ. The St. Thomas' UK Adult Twin Registry. Twin Res. 2002;5:440–3. doi:10.1375/136905202320906246.
78. 1000 Genomes Project Consortium T 1000 GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65. doi:10.1038/nature11632.
79. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437:1299–320. doi:10.1038/nature04226.
80. Loke YJ, Galati JC, Saffery R, Craig JM. Association of in vitro fertilization with global and IGF2/H19 methylation variation in newborn twins. J Dev Orig Health Dis. 2015;6:115–24. doi:10.1017/S2040174415000161.
81. Saffery R, Morley R, Carlin JB, Joo J-HE, Ollikainen M, Novakovic B, et al. Cohort profile: The peri/post-natal epigenetic twins study. Int J Epidemiol. 2012;41:55–61. doi:10.1093/ije/dyr140.
82. Hayat SA, Luben R, Keevil VL, Moore S, Dalzell N, Bhaniani A, et al. Cohort Profile: A prospective cohort study of objective physical and cognitive capability and visual health in an ageing population of men and women in Norfolk (EPIC-Norfolk 3). Int J Epidemiol. 2014;43:1063–72. doi:10.1093/ije/dyt086.
83. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: The "Children of the 90s"-The index offspring of the avon longitudinal study of parents and children. Int J Epidemiol. 2013;42:111–27.

doi:10.1093/ije/dys064.
84. Holle R, Happich M, L?wel H, Wichmann H, MONICA/KORA Study Group. KORA - A Research Platform for Population Based Health Research. Das Gesundheitswes. 2005;67 S 01:19–25. doi:10.1055/s-2005-858235.
85. Hofman A, Murad SD, Van Duijn CM, Franco OH, Goedegebure A, Arfan Ikram M, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol. 2013;28:889–926. doi:10.1007/s10654-013-9866-z.
86. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol. 2015;44:1800–13. doi:10.1093/ije/dyv337.
87. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum Mol Genet. 2007;16:547–54. doi:10.1093/hmg/ddm010.
88. Gervin K, Hammerø M, Akselsen HE, Moe R, Nygård H, Brandt I, et al. Extensive variation and low heritability of DNA methylation identified in a twin study. Genome Res. 2011;21:1813–21. doi:10.1101/gr.119685.110.
89. van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan C V, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun. 2016;7:11115. doi:10.1038/ncomms11115.
90. Gordon L, Joo JE, Powell JE, Ollikainen M, Novakovic B, Li X, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome Res. 2012;22:1395–406. doi:10.1101/gr.136598.111.
91. Quon G, Lippert C, Heckerman D, Listgarten J. Patterns of methylation heritability in a genome-wide analysis of four brain regions. Nucleic Acids Res. 2013;41:2095–104. doi:10.1093/nar/gks1449.
92. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, et al. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. PLoS Genet. 2014;10:e1004663. doi:10.1371/journal.pgen.1004663.
93. Drong AW, Nicholson G, Hedman AK, Meduri E, Grundberg E, Small KS, et al. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. PLoS One. 2013;8:e55923. doi:10.1371/journal.pone.0055923.
94. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell. 2016;167:1398–1414.e24. doi:10.1016/j.cell.2016.10.026.
95. Bonder MJ, Luijk R, Zhernakova D V, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017;49:131–8. doi:10.1038/ng.3721.
96. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8:389–99. doi:10.2217/epi.15.114.
97. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29:189–96. doi:10.1093/bioinformatics/bts680.
98. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503. doi:10.1186/s13059-014-0503-2.

99. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526:82–90. doi:10.1038/nature14962.
100. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:R115. doi:10.1186/gb-2013-14-10-r115.
101. Beaumont M, Goodrich JK, Jackson MA, Yet I, Davenport ER, Vieira-Silva S, et al. Heritable components of the human fecal microbiome are associated with visceral fat. Genome Biol. 2016;17:189. doi:10.1186/s13059-016-1052-7.
102. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46:543–50. doi:10.1038/ng.2982.
103. Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30. doi:10.1038/nature14248.
104. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74. doi:10.1038/nature11247.
105. Rijsdijk F V, Sham PC. Analytic approaches to twin data using structural equation models. Brief Bioinform. 2002;3:119–33. http://www.ncbi.nlm.nih.gov/pubmed/12139432.
106. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika. 2011;76:306–17. doi:10.1007/s11336-010-9200-6.
107. R Core Team. R: A Language and Environment for Statistical Computing. 2016. http://www.r-project.org.
108. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353–8. doi:10.1093/bioinformatics/bts163.
109. Westra H-J, Arends D, Esko T, Peters MJ, Schurmann C, Schramm K, et al. Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. 2015;11:e1005223. doi:10.1371/journal.pgen.1005223.
110. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39:311–8. doi:10.1038/ng1966.
111. Pandey S, Shetty A, Hamilton M, Bhattacharya S, Maheshwari A. Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. Hum Reprod Update. 2012;18:485–503. doi:10.1093/humupd/dms018.
112. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjaerven R, Vatten LJ. Increased risk of placenta previa in pregnancies following IVF/ICSI; a comparison of ART and non-ART pregnancies in the same mother. Hum Reprod. 2006;21:2353–8. doi:10.1093/humrep/del153.
113. Cox GF, Bürger J, Lip V, Mau UA, Sperling K, Wu B-L, et al. Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet. 2002;71:162–4. doi:10.1086/341096.
114. Maher ER, Brueton LA, Bowdin SC, Luharia A, Cooper W, Cole TR, et al. Beckwith-Wiedemann syndrome and assisted reproduction technology (ART). J Med Genet. 2003;40:62–4. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1735252&tool=pmcentrez&rendertype=abstract. Accessed 22 Nov 2014.
115. Apantaku O, Chandrasekaran I, Bentick B. Obstetric outcome of singleton pregnancies achieved with in vitro fertilisation and intracytoplasmic sperm

injection: experience from a district general hospital. J Obstet Gynaecol. 2008;28:398–402. doi:10.1080/01443610802091792.
116. De Neubourg D, Gerris J, Mangelschots K, Van Royen E, Vercruyssen M, Steylemans A, et al. The obstetrical and neonatal outcome of babies born after single-embryo transfer in IVF/ICSI compares favourably to spontaneously conceived babies. Hum Reprod. 2006;21:1041–6. doi:10.1093/humrep/dei424.
117. Weksberg R. Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. Hum Mol Genet. 2002;11:1317–25. doi:10.1093/hmg/11.11.1317.
118. Saffery R. Epigenetic change as the major mediator of fetal programming in humans: Are we there yet? Ann Nutr Metab. 2014;64:203–7. doi:10.1159/000365020.
119. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, et al. DNA methylation dynamics of the human preimplantation embryo. Nature. 2014;511:611–5. doi:10.1038/nature13581.
120. Fauque P. Ovulation induction and epigenetic anomalies. Fertil Steril. 2013;99:616–23. doi:10.1016/j.fertnstert.2012.12.047.
121. Doherty AS, Mann MR, Tremblay KD, Bartolomei MS, Schultz RM. Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod. 2000;62:1526–35. http://www.ncbi.nlm.nih.gov/pubmed/10819752. Accessed 22 Nov 2014.
122. Wang Z, Xu L, He F. Embryo vitrification affects the methylation of the H19/Igf2 differentially methylated domain and the expression of H19 and Igf2. Fertil Steril. 2010;93:2729–33. doi:10.1016/j.fertnstert.2010.03.025.
123. Gomes MV, Huber J, Ferriani RA, Amaral Neto AM, Ramos ES. Abnormal methylation at the KvDMR1 imprinting control region in clinically normal children conceived by assisted reproductive technologies. Mol Hum Reprod. 2009;15:471–7. doi:10.1093/molehr/gap038.
124. Li L, Wang L, Le F, Liu X, Yu P, Sheng J, et al. Evaluation of DNA methylation status at differentially methylated regions in IVF-conceived newborn twins. Fertil Steril. 2011;95:1975–9. doi:10.1016/j.fertnstert.2011.01.173.
125. Nelissen ECM, Dumoulin JCM, Daunay A, Evers JLH, Tost J, van Montfoort APA. Placentas from pregnancies conceived by IVF/ICSI have a reduced DNA methylation level at the H19 and MEST differentially methylated regions. Hum Reprod. 2013;28:1117–26. doi:10.1093/humrep/des459.
126. Tierling S, Souren NY, Gries J, Loporto C, Groth M, Lutsik P, et al. Assisted reproductive technologies do not enhance the variability of DNA methylation imprints in human. J Med Genet. 2010;47:371–6. doi:10.1136/jmg.2009.073189.
127. Katari S, Turan N, Bibikova M, Erinle O, Chalian R, Foster M, et al. DNA methylation and gene expression differences in children conceived in vitro or in vivo. Hum Mol Genet. 2009;18:3769–78. doi:10.1093/hmg/ddp319.
128. Melamed N, Choufani S, Wilkins-Haug LE, Koren G, Weksberg R. Comparison of Genome-Wide and Gene-Specific DNA Methylation between ART and Naturally Conceived Pregnancies. Epigenetics. 2015;10:474–83. doi:10.4161/15592294.2014.988041.
129. Estill MS, Bolnick JM, Waterland RA, Bolnick AD, Diamond MP, Krawetz SA. Assisted reproductive technology alters deoxyribonucleic acid methylation profiles in bloodspots of newborn infants. Fertil Steril. 2016;106:629–639.e10. doi:10.1016/j.fertnstert.2016.05.006.
130. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98:288–95.

doi:10.1016/j.ygeno.2011.07.007.
131. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37:853–62. doi:10.1038/ng1598.
132. Loke YJ, Novakovic B, Ollikainen M, Wallace EM, Umstad MP, Permezel M, et al. The Peri/postnatal Epigenetic Twins Study (PETS). Twin Res Hum Genet. 2013;16:13–20. doi:10.1017/thg.2012.114.
133. Ollikainen M, Smith KR, Joo EJ-H, Ng HK, Andronikos R, Novakovic B, et al. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet. 2010;19:4176–88. doi:10.1093/hmg/ddq336.
134. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res. 2010;20:1441–50. doi:10.1101/gr.110114.110.
135. Bates D, Maechler M, Bolker B, Walker S. {lme4}: Linear mixed-effects models using Eigen and S4. 2014. http://cran.r-project.org/package=lme4.
136. Miyagawa Y, Nishimura H, Tsujimura A, Matsuoka Y, Matsumiya K, Okuyama A, et al. Single-nucleotide polymorphisms and mutation analyses of the TNP1 and TNP2 genes of fertile and infertile human male populations. J Androl. 2005;26:779–86. doi:10.2164/jandrol.05069.
137. Trasler JM, Hake LE, Johnson P a, Alcivar a a, Millette CF, Hecht NB. DNA methylation and demethylation events during meiotic prophase in the mouse testis. Mol Cell Biol. 1990;10:1828–34. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=362295&tool=pmcentrez&rendertype=abstract.
138. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet. 2012;44:1020–5. doi:10.1038/ng.2384.
139. Kerns SL, Ostrer H, Stock R, Li W, Moore J, Pearlman A, et al. Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys. 2010;78:1292–300. doi:10.1016/j.ijrobp.2010.07.036.
140. Fenstad MH, Johnson MP, Løset M, Mundal SB, Roten LT, Eide IP, et al. STOX2 but not STOX1 is differentially expressed in decidua from pre-eclamptic women: data from the Second Nord-Trondelag Health Study. Mol Hum Reprod. 2010;16:960–8. doi:10.1093/molehr/gaq064.
141. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489:75–82. doi:10.1038/nature11232.
142. Gardiner-Garden M, Frommer M. CpG Islands in vertebrate genomes. J Mol Biol. 1987;196:261–82. doi:10.1016/0022-2836(87)90689-9.
143. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol. 2015;16:118. doi:10.1186/s13059-015-0660-y.
144. Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics. 3:89–96.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2492384&tool=pmcentrez&rendertype=abstract. Accessed 7 May 2015.
145. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet (London, England). 1992;340:17–8. http://www.ncbi.nlm.nih.gov/pubmed/1351601. Accessed 24 Jan 2016.
146. Palermo GD, Neri Q V, Rosenwaks Z. Safety of intracytoplasmic sperm injection. Methods Mol Biol. 2014;1154:549–62. doi:10.1007/978-1-4939-0659-8_26.
147. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80-). 2015;347:1260419–1260419. doi:10.1126/science.1260419.
148. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet. 2002;18:348–51. http://www.ncbi.nlm.nih.gov/pubmed/12127774. Accessed 23 Jun 2016.
149. Oliver VF, Miles HL, Cutfield WS, Hofman PL, Ludgate JL, Morison IM. Defects in imprinting and genome-wide DNA methylation are not common in the in vitro fertilization population. Fertil Steril. 2012;97:147–53.e7. doi:10.1016/j.fertnstert.2011.10.027.
150. Boissonnas CC, Abdalaoui H El, Haelewyn V, Fauque P, Dupont JM, Gut I, et al. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. Eur J Hum Genet. 2010;18:73–80. doi:10.1038/ejhg.2009.117.
151. Whitelaw N, Bhattacharya S, Hoad G, Horgan GW, Hamilton M, Haggarty P. Epigenetic status in the offspring of spontaneous and assisted conception. Hum Reprod. 2014;29:1452–8. doi:10.1093/humrep/deu094.
152. Ludwig M, Katalinic A, Gross S, Sutcliffe A, Varon R, Horsthemke B. Increased prevalence of imprinting defects in patients with Angelman syndrome born to subfertile couples. J Med Genet. 2005;42:289–91. doi:10.1136/jmg.2004.026930.
153. Doornbos ME, Maas SM, McDonnell J, Vermeiden JPW, Hennekam RCM. Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. Hum Reprod. 2007;22:2476–80. doi:10.1093/humrep/dem172.
154. Jaques AM, Amor DJ, Baker HWG, Healy DL, Ukoumunne OC, Breheny S, et al. Adverse obstetric and perinatal outcomes in subfertile women conceiving without assisted reproductive technologies. Fertil Steril. 2010;94:2674–9. doi:10.1016/j.fertnstert.2010.02.043.
155. Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Söderström-Anttila V, et al. Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis. Hum Reprod Update. 19:87–104. doi:10.1093/humupd/dms044.
156. Song S, Ghosh J, Mainigi M, Turan N, Weinerman R, Truongcao M, et al. DNA methylation differences between in vitro- and in vivo-conceived children are associated with ART procedures rather than infertility. Clin Epigenetics. 2015;7:41. doi:10.1186/s13148-015-0071-7.
157. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15:R31. doi:10.1186/gb-2014-15-2-r31.
158. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38:1378–85. doi:10.1038/ng1909.
159. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015.

doi:10.1093/ije/dyv041.
160. Henderson KD, Bernstein L, Henderson B, Kolonel L, Pike MC. Predictors of the Timing of Natural Menopause in the Multiethnic Cohort Study. Am J Epidemiol. 2008;167:1287–94. doi:10.1093/aje/kwn046.
161. Kritz-Silverstein D, Barrett-Connor E. Early menopause, number of reproductive years, and bone mineral density in postmenopausal women. Am J Public Health. 1993;83:983–8. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1694784&tool=pmcentrez&rendertype=abstract. Accessed 20 May 2015.
162. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13:1141–51. doi:10.1016/S1470-2045(12)70425-4.
163. Burger HG, Hale GE, Robertson DM, Dennerstein L. A review of hormonal changes during the menopausal transition: focus on findings from the Melbourne Women's Midlife Health Project. Hum Reprod Update. 2007;13:559–65. doi:10.1093/humupd/dmm020.
164. Sapir-Koren R, Livshits G. Postmenopausal osteoporosis in rheumatoid arthritis: The estrogen deficiency-immune mechanisms link. Bone. 2017;103:102–15. doi:10.1016/j.bone.2017.06.020.
165. Dall GV, Britt KL. Estrogen Effects on the Mammary Gland in Early and Late Life and Breast Cancer Risk. Front Oncol. 2017;7:110. doi:10.3389/fonc.2017.00110.
166. Snieder H, MacGregor AJ, Spector TD. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. J Clin Endocrinol Metab. 1998;83:1875–80. doi:10.1210/jcem.83.6.4890.
167. Stolk L, Zhai G, van Meurs JBJ, Verbiest MMPJ, Visser JA, Estrada K, et al. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. Nat Genet. 2009;41:645–7. doi:10.1038/ng.387.
168. He C, Kraft P, Chen C, Buring JE, Paré G, Hankinson SE, et al. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat Genet. 2009;41:724–8. doi:10.1038/ng.385.
169. Chen CTL, Fernández-Rhodes L, Brzyski RG, Carlson CS, Chen Z, Heiss G, et al. Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study. Hum Mol Genet. 2012;21:1419–32. doi:10.1093/hmg/ddr570.
170. Stolk L, Perry JRB, Chasman DI, He C, Mangino M, Sulem P, et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nat Genet. 2012;44:260–8. doi:10.1038/ng.1051.
171. Chen CTL, Liu C-T, Chen GK, Andrews JS, Arnold AM, Dreyfus J, et al. Meta-analysis of loci associated with age at natural menopause in African-American women. Hum Mol Genet. 2014;23:3327–42. doi:10.1093/hmg/ddu041.
172. Shen C, Delahanty RJ, Gao Y-T, Lu W, Xiang Y-B, Zheng Y, et al. Evaluating GWAS-identified SNPs for age at natural menopause among chinese women. PLoS One. 2013;8:e58766. doi:10.1371/journal.pone.0058766.
173. Rödström K, Bengtsson C, Milsom I, Lissner L, Sundh V, Bjoürkelund C. Evidence for a secular trend in menopausal age: a population study of women in Gothenburg. Menopause. 2003;10:538–43. doi:10.1097/01.GME.0000094395.59028.0F.
174. Nichols HB, Trentham-Dietz A, Hampton JM, Titus-Ernstoff L, Egan KM, Willett WC, et al. From menarche to menopause: trends among US Women born

from 1912 to 1969. Am J Epidemiol. 2006;164:1003–11. doi:10.1093/aje/kwj282.
175. 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. 2016;541:81–6. doi:10.1038/nature20784.
176. Simpson ER. Sources of estrogen and their importance. J Steroid Biochem Mol Biol. 2003;86:225–30. http://www.ncbi.nlm.nih.gov/pubmed/14623515. Accessed 4 Jul 2017.
177. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86. doi:10.1186/1471-2105-13-86.
178. M?gi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010;11:288. doi:10.1186/1471-2105-11-288.
179. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology. 2013;24:712–6. doi:10.1097/EDE.0b013e31829d5cb3.
180. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One. 2012;7:e51954. doi:10.1371/journal.pone.0051954.
181. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, et al. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet. 2013;50:298–308. doi:10.1136/jmedgenet-2012-101461.
182. Billeci AMR, Paciaroni M, Caso V, Agnelli G. Hormone replacement therapy and stroke. Curr Vasc Pharmacol. 2008;6:112–23. http://www.ncbi.nlm.nih.gov/pubmed/18393913. Accessed 5 Jul 2017.
183. Spector TD, Nandra D, Hart DJ, Doyle D V. Is hormone replacement therapy protective for hand and knee osteoarthritis in women?: The Chingford Study. Ann Rheum Dis. 1997;56:432–4. http://www.ncbi.nlm.nih.gov/pubmed/9486006. Accessed 5 Jul 2017.
184. Skoog I, Gustafson D. HRT and dementia. J Epidemiol Biostat. 1999;4:227–51; discussion 252. http://www.ncbi.nlm.nih.gov/pubmed/10695961. Accessed 5 Jul 2017.
185. Ronkainen PHA, Kovanen V, Alen M, Pollanen E, Palonen E-M, Ankarberg-Lindgren C, et al. Postmenopausal hormone replacement therapy modifies skeletal muscle composition and function: a study with monozygotic twin pairs. J Appl Physiol. 2009;107:25–33. doi:10.1152/japplphysiol.91518.2008.
186. Romieu I, Fabre A, Fournier A, Kauffmann F, Varraso R, Mesrine S, et al. Postmenopausal hormone therapy and asthma onset in the E3N cohort. Thorax. 2010;65:292–7. doi:10.1136/thx.2009.116079.
187. Bahl A, Pöllänen E, Ismail K, Sipilä S, Mikkola TM, Berglund E, et al. Hormone Replacement Therapy Associated White Blood Cell DNA Methylation and Gene Expression are Associated With Within-Pair Differences of Body Adiposity and Bone Mass. Twin Res Hum Genet. 2015;18:647–61. doi:10.1017/thg.2015.82.
188. World Health Organization. WHO | Infertility is a global public health issue. WHO. 2014. http://www.who.int/reproductivehealth/topics/infertility/perspective/en/. Accessed 9 Jun 2017.
189. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Hum Reprod. 2007;22:1506–12. doi:10.1093/humrep/dem046.
190. Kupka MS, Ferraretti AP, de Mouzon J, Erb K, D'Hooghe T, Castilla JA, et al.

Assisted reproductive technology in Europe, 2010: results generated from European registers by ESHRE†. Hum Reprod. 2014;29:2099–113. doi:10.1093/humrep/deu175.
191. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010;465:721–7. doi:10.1038/nature09230.
192. Bruder CEG, Piotrowski A, Gijsbers AACJ, Andersson R, Erickson S, Diaz de Ståhl T, et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet. 2008;82:763–71. doi:10.1016/j.ajhg.2007.12.011.
193. Li R, Montpetit A, Rousseau M, Wu SYM, Greenwood CMT, Spector TD, et al. Somatic point mutations occurring early in development: a monozygotic twin study. J Med Genet. 2013;0:1–7. doi:10.1136/jmedgenet-2013-101712.
194. Visscher PM. Power of the classical twin design revisited. Twin Res. 2004;7:505–12. doi:10.1375/1369052042335250.
195. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015;44:1429. doi:10.1093/ije/dyv041.
196. Murano T, Izumi S, Kika G, Haque SF, Okuwaki S, Mori A, et al. Impact of menopause on lipid and bone metabolism and effect of hormone replacement therapy. Tokai J Exp Clin Med. 2003;28:109–19. http://www.ncbi.nlm.nih.gov/pubmed/15055403. Accessed 1 Jul 2017.

Annex

I. Enrichment analysis accounting for variance

	cis meQTLs				*trans* meQTLs			
	# CpG sites	Fold Enrichment	Min	Max	# CpG sites	Fold Enrichment	Min	Max
TFBS (GM12878)	30,374 ▼	0.895	33,677	34,313	630 ▲	1.518	373	465
H3K27ac (PBMCs)	133,927 ▲	1.009	132,347	133,052	1,574 ▲	1.002	1,526	1,617
H3K27me3 (PBMCs)	94,607 ▲	1.027	91,736	92,511	972 ▼	0.913	1,003	1,127
H3K36me3 (PBMCs)	82,961 ▲	1.022	80,747	81,612	939 ▼	0.963	914	1,024
H3K4me1 (PBMCs)	136,953 ▲	1.023	133,658	134,268	1,506 ▼	0.959	1,524	1,610
H3K4me3 (PBMCs)	106,098 ▼	0.988	107,072	107,829	1,393 ▲	1.095	1,211	1,327
H3K9ac (PBMCs)	124,200 ▲	1.008	122,869	123,682	1,534 ▲	1.052	1,403	1,500
H3K9me3 (PBMCs)	30,848 ▼	0.963	31,678	32,334	386 ▲	1.038	333	434
Island	25,928 ▼	0.840	30,568	31,163	673 ▲	1.725	333	421
Shore	33,104 ▲	1.048	31,280	31,933	325 ▼	0.893	315	396
Shelf	12,190 ▲	1.040	11,520	11,979	114 ▼	0.813	111	179
Open sea	99,934 ▲	1.031	96,573	97,349	909 ▼	0.807	1,068	1,209
Promoter (PBMCs)	6,758 ▼	0.821	8,072	8,382	208 ▲	2.013	81	143
Enhancer (PBMCs)	8,523 ▲	1.180	7,057	7,387	52 ▼	0.669	51	98
TSS1500	24,823 ▼	0.983	24,877	25,625	309 ▲	1.029	259	344
TSS200	13,284 ▼	0.856	15,292	15,775	355 ▲	1.828	152	229
5' UTR	20,622 ▼	0.937	21,688	22,334	311 ▲	1.171	223	314
1st Exon	6,963 ▼	0.772	8,816	9,235	204 ▲	1.768	94	139
Exon boundary	1,235 ▼	0.888	1,318	1,470	8 ▼	0.463	7	31
Gene body	72,997 ▲	1.018	71,358	72,193	716 ▼	0.839	797	911
3' UTR	4,618 ▼	0.983	4,579	4,854	47 ▼	0.842	37	75
Intergenic	51,269 ▲	1.041	48,898	49,594	514 ▼	0.919	498	604
Positive selection	1,346 ▼	0.836	1,516	1,688	20 ▼	0.970	11	35
Conserved elements	27,082 ▼	0.834	32,177	32,761	403 ▲	1.033	336	435

II. Re-analysis of FDR 25% WBC IVF-DMRs excluding other fertility treatments from non-IVF group (n=94, 54 non-IVF and 40 IVF)

Chromosome	Start	End	Estimate	SE	p
chr2	217726751	217727250	1.17	0.19	9.66E-09
chr5	178761751	178762250	-1.09	0.20	5.69E-08
chr9	97504001	97504500	1.05	0.20	1.22E-07
chr5	9275751	9276250	1.09	0.19	1.17E-07
chr4	184814001	184814500	-0.94	0.18	2.00E-07
chr5	142488501	142489000	-1.09	0.20	1.05E-07
chr9	118148751	118149250	1.05	0.20	3.42E-07
chr9	118149001	118149500	1.04	0.20	1.52E-07
chr11	82654251	82654750	-1.07	0.20	1.57E-07
chr19	6165251	6165750	0.93	0.18	2.38E-07
chr1	85522251	85522750	0.97	0.18	3.36E-07
chr17	42569001	42569500	-1.10	0.19	7.18E-08
chr4	141606501	141607000	1.07	0.20	5.32E-07
chr5	137736001	137736500	-1.02	0.20	4.39E-07
chr5	150614501	150615000	-1.05	0.20	3.18E-07
chr17	36918251	36918750	-1.11	0.21	4.66E-07
chr6	126138251	126138750	-1.01	0.19	1.00E-07

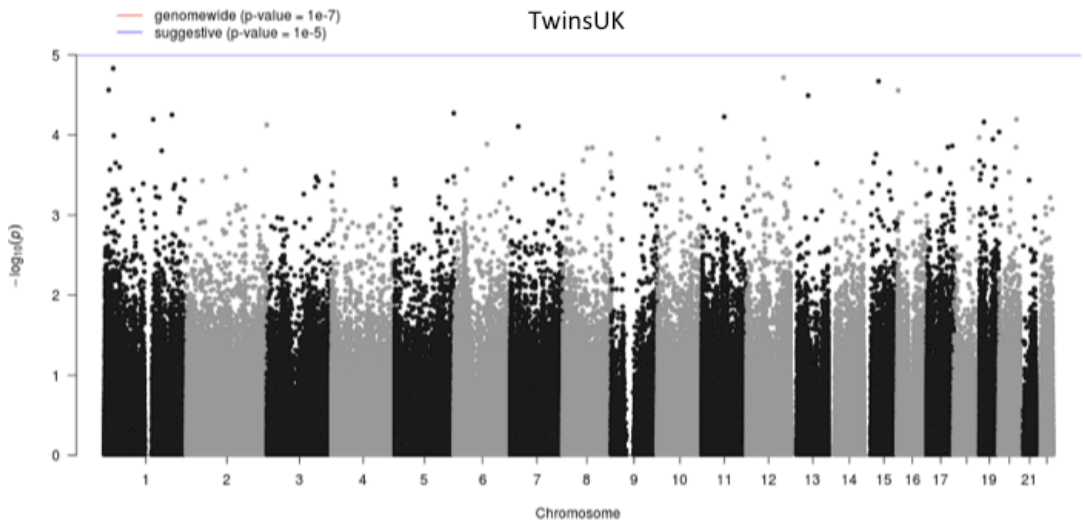
chr7	144431251	144431750	1.00	0.19	6.16E-07
chr12	70937251	70937750	0.90	0.18	5.51E-07
chr4	141606251	141606750	1.02	0.19	3.96E-07
chr13	90019001	90019500	1.08	0.20	1.77E-07
chr11	74179001	74179500	1.08	0.20	2.81E-07
chr12	99153001	99153500	-0.99	0.20	8.01E-07
chr2	223336751	223337250	-1.01	0.20	4.73E-07
chr8	120972001	120972500	1.01	0.19	1.51E-07
chr17	38047001	38047500	-1.06	0.20	5.53E-07
chr4	64626751	64627250	-0.99	0.19	5.24E-07
chr16	87256751	87257250	-0.83	0.17	9.12E-07
chr19	10656751	10657250	-1.02	0.20	4.44E-07
chr7	2487251	2487750	0.78	0.16	1.36E-06
chr11	74178751	74179250	1.01	0.20	5.79E-07
chr10	119176501	119177000	-1.05	0.21	8.45E-07
chr22	34755251	34755750	-0.98	0.20	1.25E-06
chr6	161664751	161665250	0.97	0.19	1.04E-06
chr16	17161751	17162250	-1.02	0.20	6.74E-07
chr18	23695001	23695500	1.02	0.21	1.07E-06
chr9	26364751	26365250	-0.96	0.20	8.40E-07
chr1	25227001	25227500	-0.82	0.17	1.82E-06
chr13	68877251	68877750	0.98	0.20	1.61E-06
chr9	89126501	89127000	-0.93	0.19	6.93E-07
chr13	35317501	35318000	-0.97	0.19	6.46E-07
chr21	19575001	19575500	1.00	0.19	3.04E-07
chr2	169470001	169470500	-1.03	0.20	6.63E-07
chr12	4310251	4310750	-0.96	0.20	2.47E-06
chr6	157136501	157137000	-0.97	0.19	1.27E-06
chr14	104067251	104067750	0.70	0.14	1.78E-06

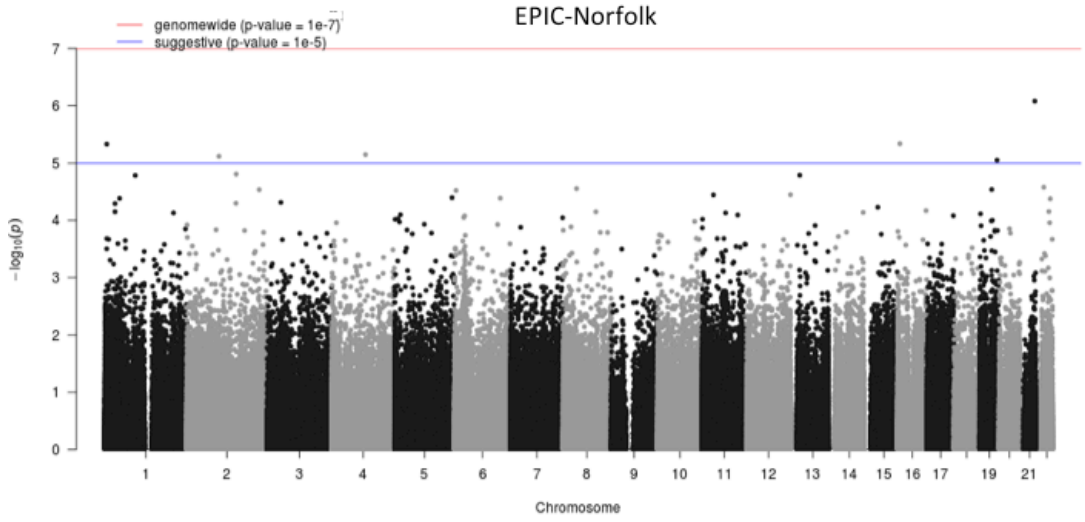
III. Re-analysis of FDR 25% WBC IVF-DMRs in subset with cell counts

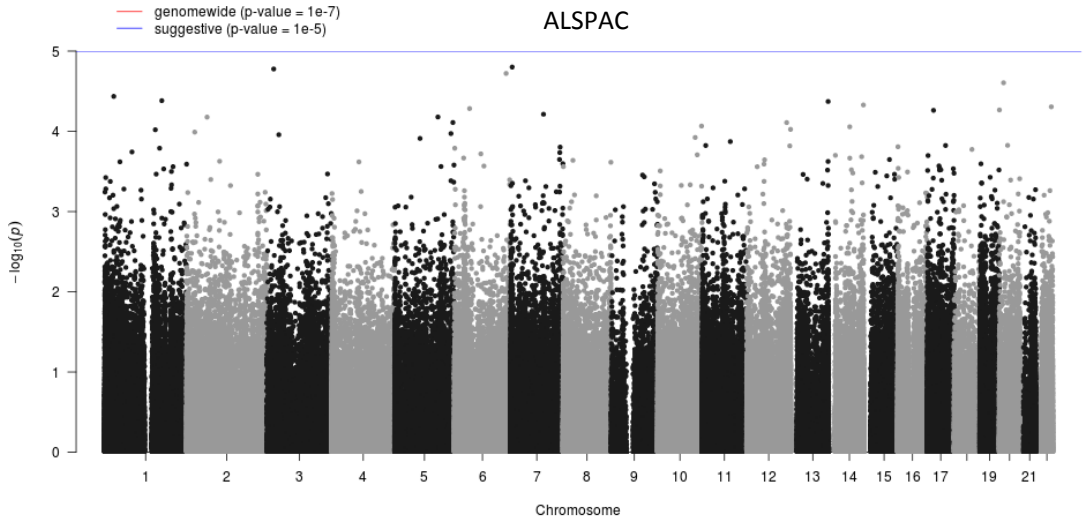
			Adjusted for 5 PCs (n=98, 40 IVF and 58 non-IVF)			Adjusted for 5 PCs (n=54, 22 IVF and 32 non-IVF)			Adjusted for cell proportions (n=54, 22 IVF and 32 non-IVF)		
Chromosome	Start	End	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
chr2	217726751	217727250	1.18	0.19	2.30E-09	0.98	0.28	2.78E-04	0.91	0.27	4.43E-04
chr5	178761751	178762250	-1.08	0.2	5.43E-08	-1.19	0.26	3.37E-06	-1.1	0.26	1.44E-05
chr9	97504001	97504500	1.07	0.2	5.83E-08	1.14	0.28	1.85E-05	0.91	0.27	4.56E-04
chr5	9275751	9276250	1.09	0.19	5.86E-08	1.17	0.25	5.74E-06	1.11	0.25	6.28E-06
chr4	184814001	184814500	-0.96	0.18	7.95E-08	-1.06	0.22	1.00E+00	-1.06	0.26	6.25E-05
chr5	142488501	142489000	-1.11	0.2	8.73E-08	-1.04	0.29	2.14E-04	-0.93	0.3	1.02E-03
chr9	118148751	118149250	1.09	0.2	9.20E-08	1	0.29	3.18E-04	0.91	0.3	1.18E-03
chr9	118149001	118149500	1.08	0.2	1.04E-07	1.02	0.3	2.41E-04	0.94	0.3	8.19E-04
chr11	82654251	82654750	-1.05	0.19	1.30E-07	-1.1	0.26	1.20E-05	-1.08	0.26	2.99E-05
chr19	6165251	6165750	0.94	0.18	1.40E-07	1.05	0.24	7.97E-06	0.82	0.28	2.11E-03
chr1	85522251	85522750	0.99	0.18	1.43E-07	0.85	0.28	1.07E-03	0.97	0.26	1.24E-04
chr17	42569001	42569500	-1.07	0.19	1.64E-07	-0.97	0.28	2.15E-04	-0.72	0.29	7.61E-03
chr4	141606501	141607000	1.09	0.19	2.03E-07	0.81	0.26	1.22E-03	0.54	0.28	3.38E-02
chr5	137736001	137736500	-1.03	0.19	2.06E-07	-1.06	0.27	7.31E-05	-0.79	0.29	2.85E-03
chr5	150614501	150615000	-1.08	0.21	2.14E-07	-1.14	0.28	2.43E-05	-1.3	0.26	1.86E-06
chr17	36918251	36918750	-1.1	0.21	2.32E-07	-1.12	0.31	2.47E-04	-0.75	0.27	3.45E-03
chr6	126138251	126138750	-0.99	0.19	2.36E-07	-1.23	0.25	7.72E-07	-1.19	0.24	1.05E-06
chr7	144431251	144431750	1.02	0.19	2.70E-07	1.12	0.27	4.08E-05	1.21	0.27	7.42E-06
chr12	70937251	70937750	0.88	0.17	2.76E-07	0.87	0.25	4.24E-04	0.78	0.28	2.30E-03
chr4	141606251	141606750	1.01	0.19	2.80E-07	0.76	0.28	3.19E-03	0.56	0.28	2.97E-02
chr13	90019001	90019500	1.06	0.2	2.84E-07	1	0.28	1.44E-04	1.02	0.26	6.44E-05
chr11	74179001	74179500	1.07	0.2	2.92E-07	0.9	0.28	4.89E-04	0.86	0.28	1.27E-03

chr12	99153001	99153500	-1.01	0.2	3.93E-07	-0.97	0.24	4.62E-05	-0.96	0.27	2.30E-04
chr2	223336751	223337250	-1.01	0.2	4.47E-07	-0.99	0.27	8.56E-05	-0.78	0.28	3.05E-03
chr8	120972001	120972500	0.97	0.19	4.98E-07	0.53	0.23	1.21E-02	0.4	0.28	9.42E-02
chr17	38047001	38047500	-1.04	0.2	5.25E-07	-1.34	0.26	2.51E-06	-1.34	0.25	3.52E-06
chr4	64626751	64627250	-0.99	0.19	5.45E-07	-0.95	0.25	1.07E-04	-0.74	0.32	1.19E-02
chr16	87256751	87257250	-0.83	0.16	5.51E-07	-1.06	0.19	7.18E-08	-1.06	0.28	7.47E-05
chr19	10656751	10657250	-1.03	0.2	5.77E-07	-1.06	0.27	4.01E-05	-0.86	0.26	4.40E-04
chr7	2487251	2487750	0.8	0.16	5.85E-07	0.82	0.19	1.74E-05	0.29	0.29	2.64E-01
chr11	74178751	74179250	0.99	0.2	6.82E-07	0.89	0.28	8.08E-04	0.88	0.28	9.17E-04
chr10	119176501	119177000	-1.04	0.2	6.87E-07	-1.18	0.27	1.49E-05	-1.03	0.27	1.09E-04
chr22	34755251	34755750	-0.98	0.2	7.32E-07	-0.85	0.29	1.70E-03	-0.81	0.26	1.09E-03
chr6	161664751	161665250	0.98	0.19	7.42E-07	0.94	0.3	1.05E-03	0.62	0.31	2.70E-02
chr16	17161751	17162250	-1.01	0.2	7.80E-07	-0.94	0.25	1.13E-04	-0.87	0.28	1.52E-03
chr18	23695001	23695500	1.04	0.21	8.39E-07	0.79	0.33	1.18E-02	0.4	0.29	1.20E-01
chr9	26364751	26365250	-0.95	0.19	8.90E-07	-1.02	0.26	4.45E-05	-0.78	0.26	1.28E-03
chr1	25227001	25227500	-0.83	0.17	9.05E-07	-0.71	0.22	4.89E-04	-1.04	0.28	2.73E-04
chr13	68877251	68877750	0.97	0.2	9.43E-07	1.07	0.25	1.21E-05	0.9	0.29	1.08E-03
chr9	89126501	89127000	-0.92	0.19	9.59E-07	-0.57	0.25	1.36E-02	-0.64	0.3	2.33E-02
chr13	35317501	35318000	-0.96	0.2	9.72E-07	-0.77	0.29	3.49E-03	-0.57	0.29	3.03E-02
chr21	19575001	19575500	0.96	0.19	9.92E-07	1.13	0.22	3.90E-07	0.76	0.24	7.41E-04
chr2	169470001	169470500	-1.02	0.21	1.06E-06	-1.12	0.29	1.22E-04	-0.88	0.28	1.96E-03
chr12	4310251	4310750	-0.99	0.2	1.06E-06	-0.99	0.29	2.65E-04	-0.8	0.29	1.00E+00
chr6	157136501	157137000	-0.96	0.19	1.15E-06	-0.98	0.24	1.88E-05	-1.01	0.26	4.46E-05
chr14	104067251	104067750	0.71	0.14	1.17E-06	0.73	0.2	1.42E-04	0.19	0.3	4.84E-01

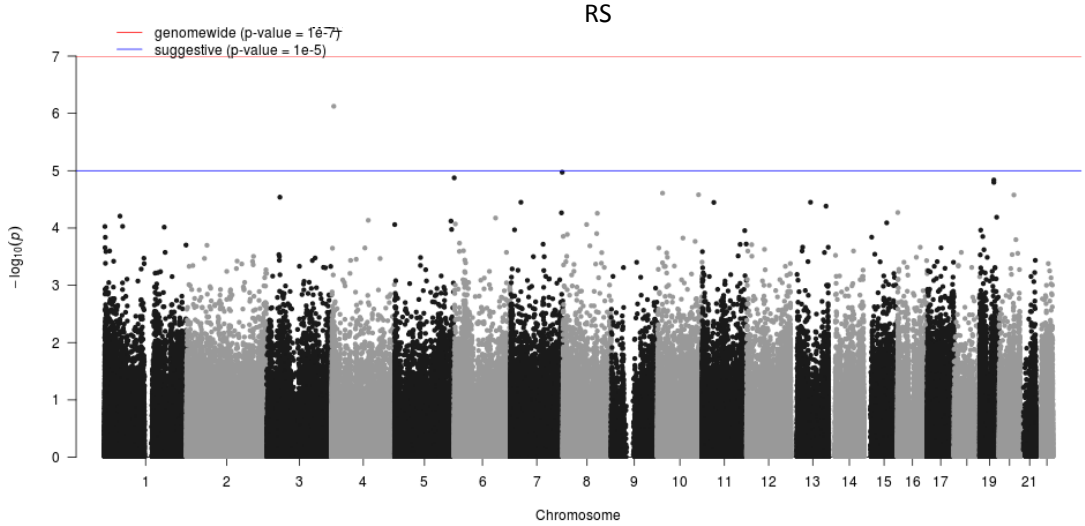
IV. Age at menopause EWAS Manhattan plots



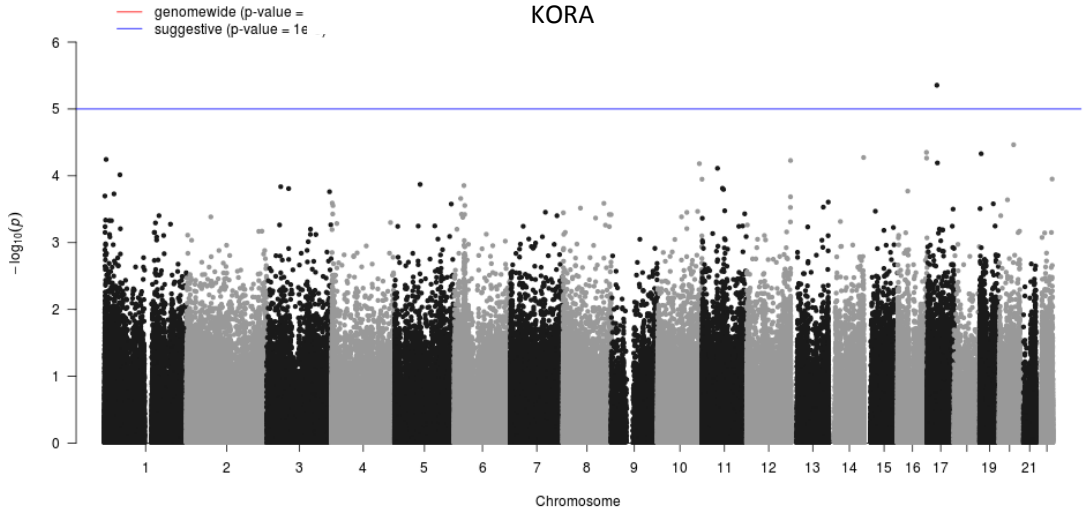




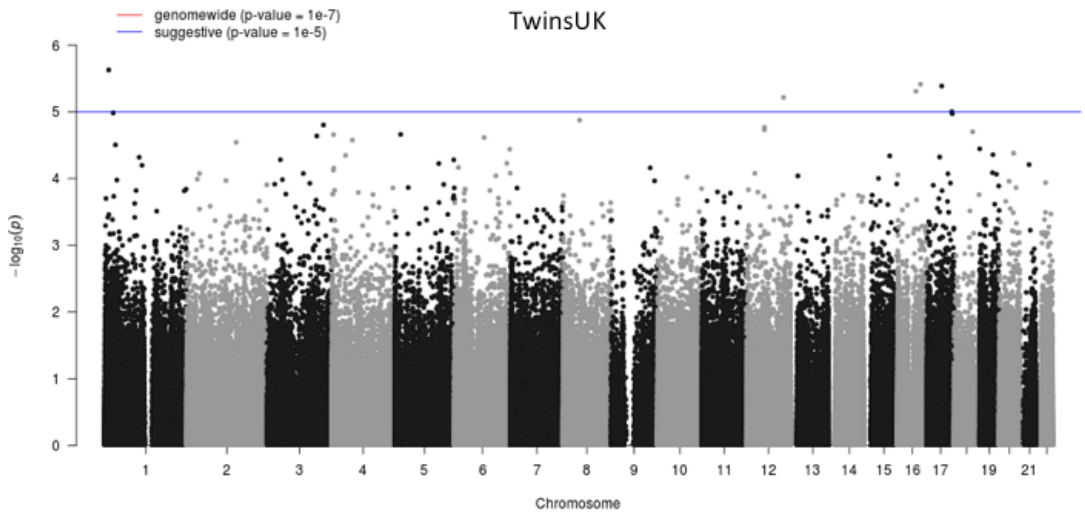
RS

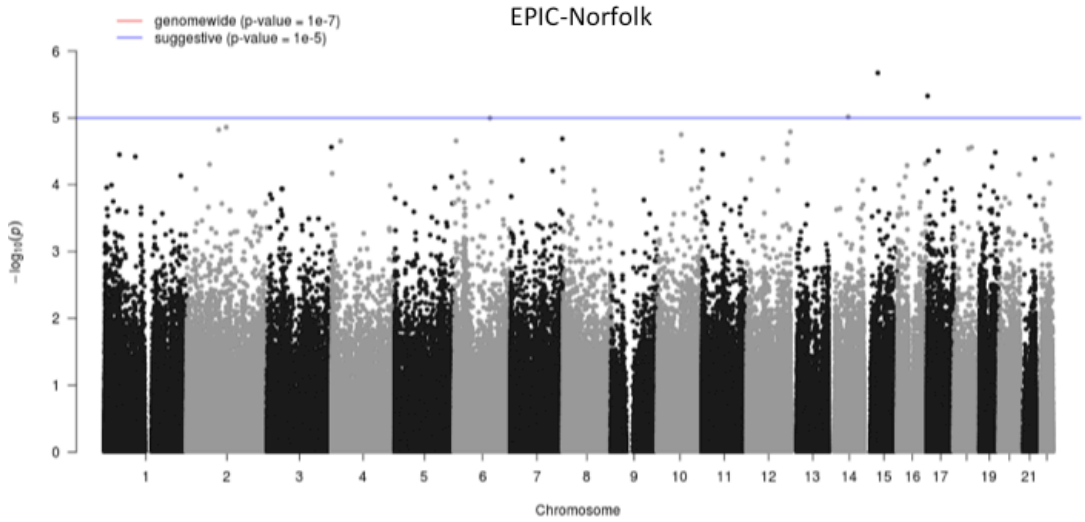

genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
$-\log_{10}(p)$
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

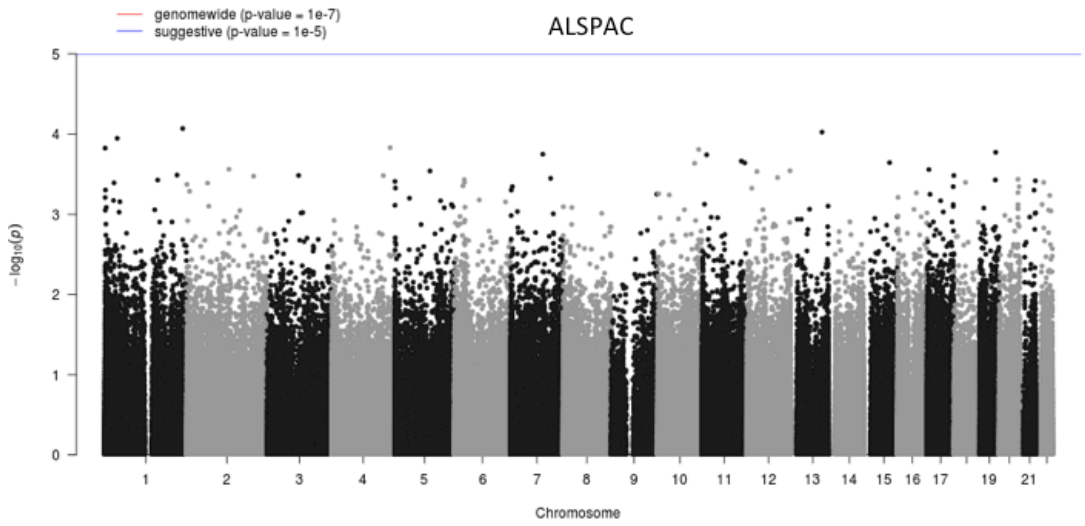
KORA

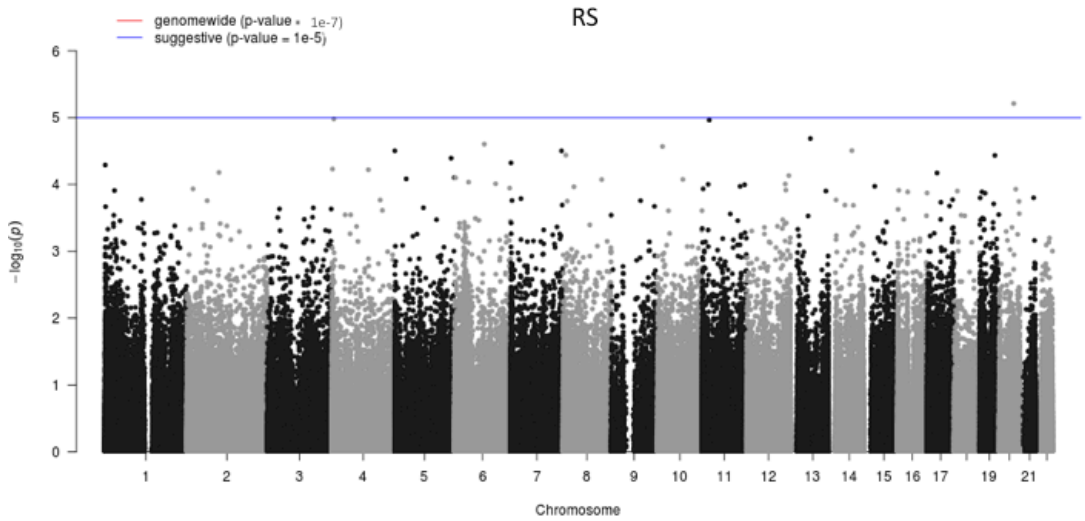

genomewide (p-value =
suggestive (p-value = 1e
$-\log_{10}(p)$
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

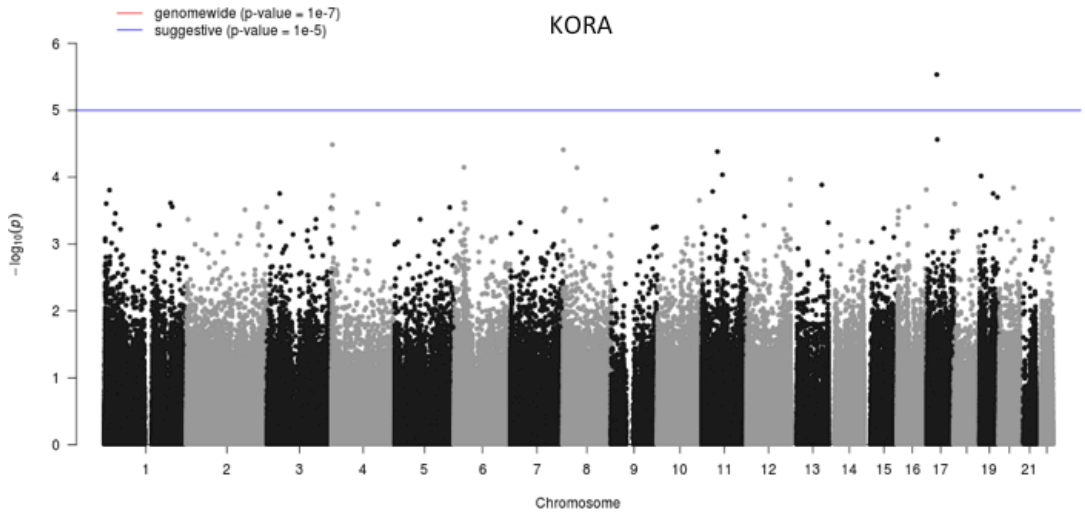
V. Length of reproductive period EWAS Manhattan plots



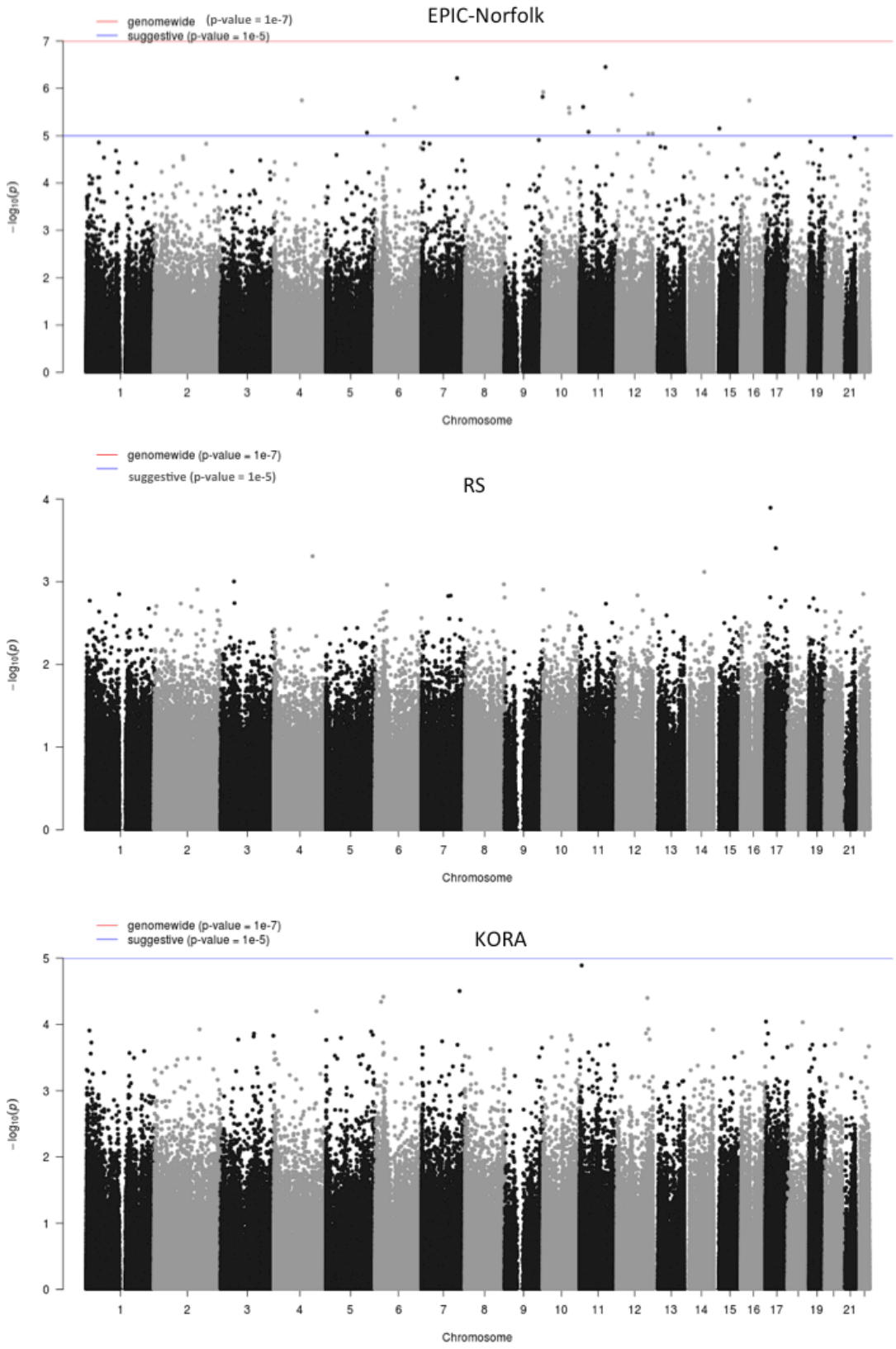




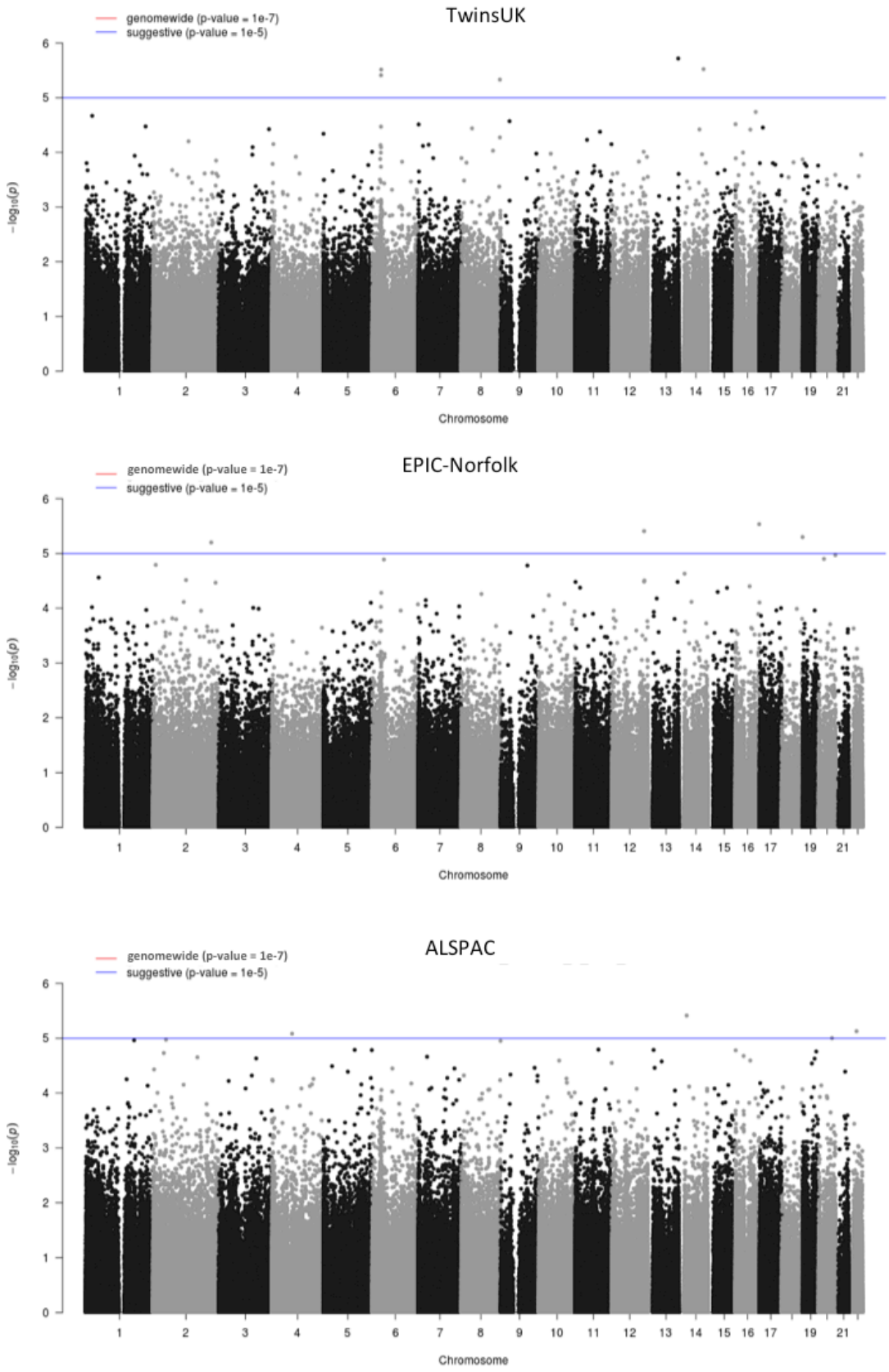

RS
genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
$-\log_{10}(p)$
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

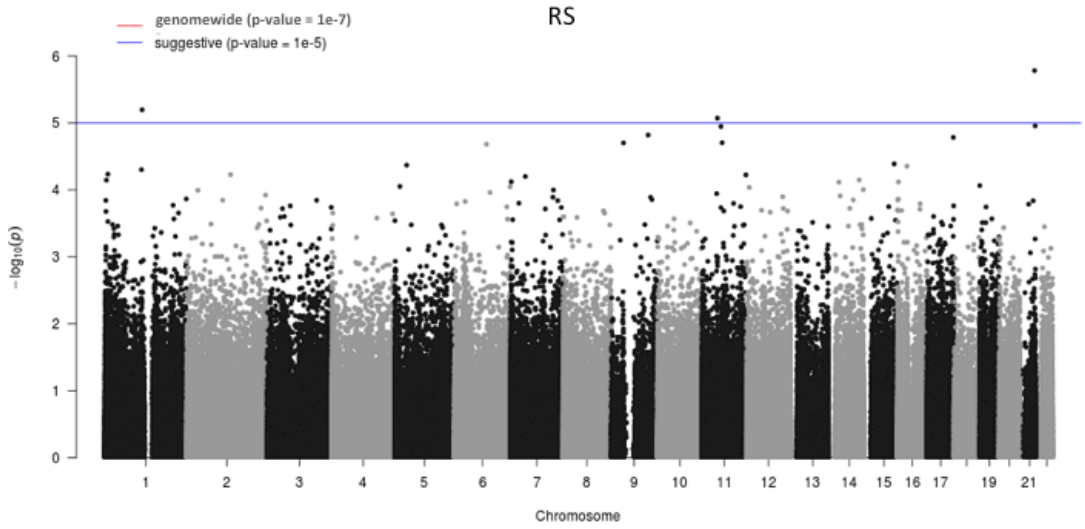

KORA
genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
$-\log_{10}(p)$
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

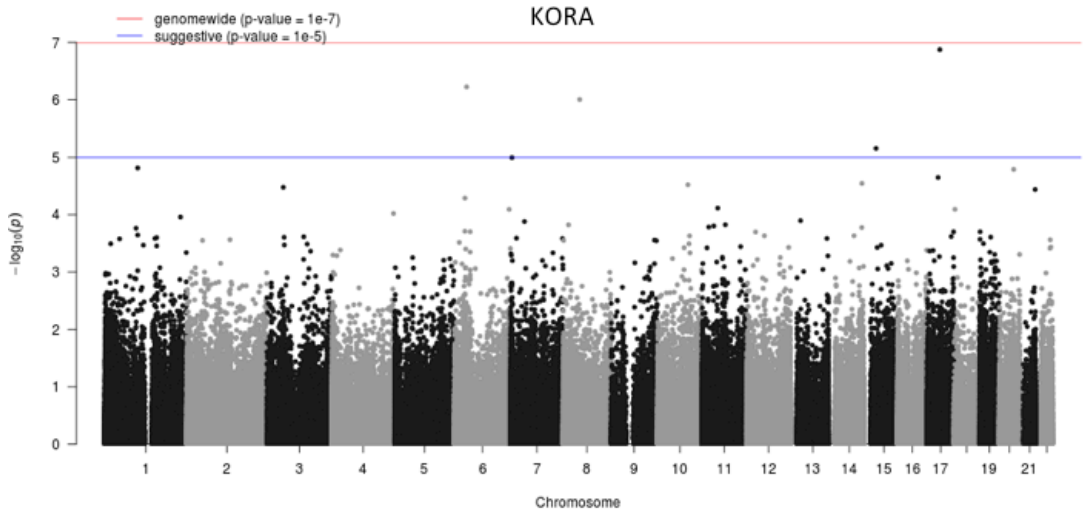
VI. Early vs late menopause EWAS Manhattan plots



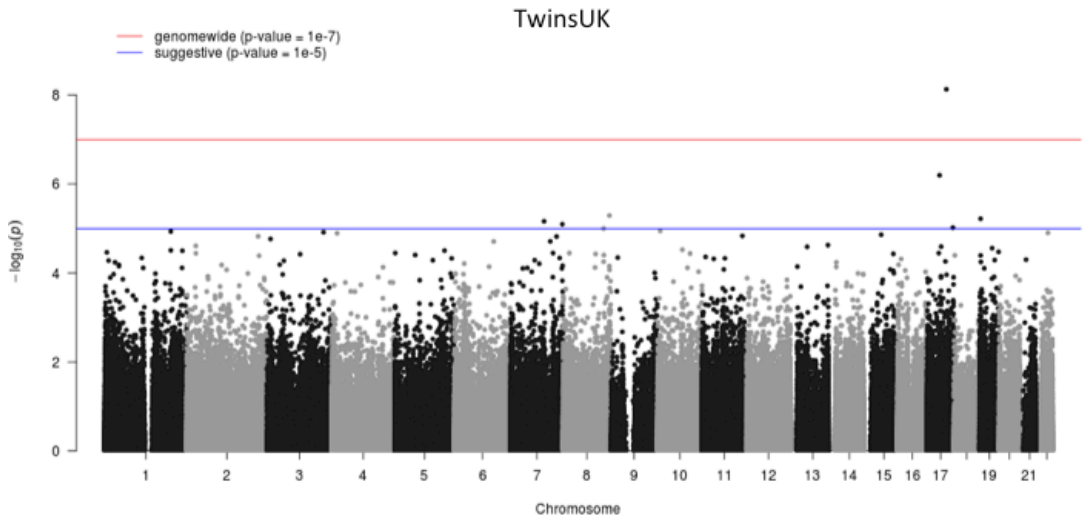
VII. Menopausal status EWAS Manhattan plots

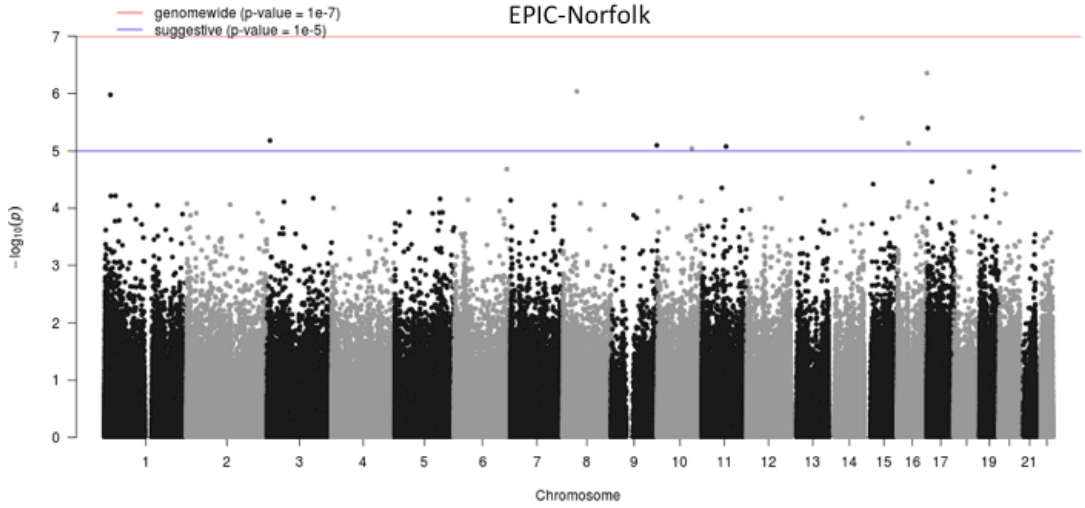


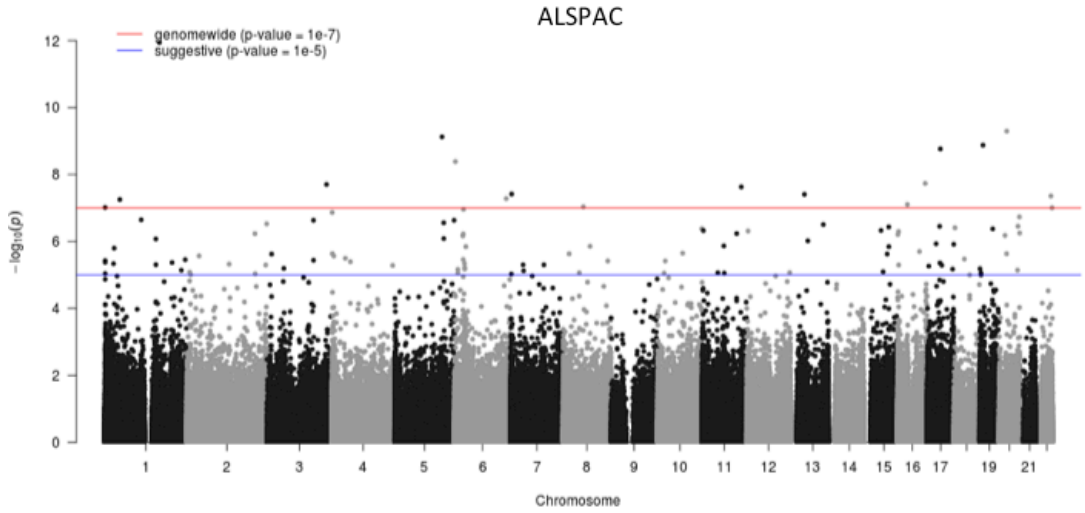

RS
genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
$-\log_{10}(p)$
Chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21

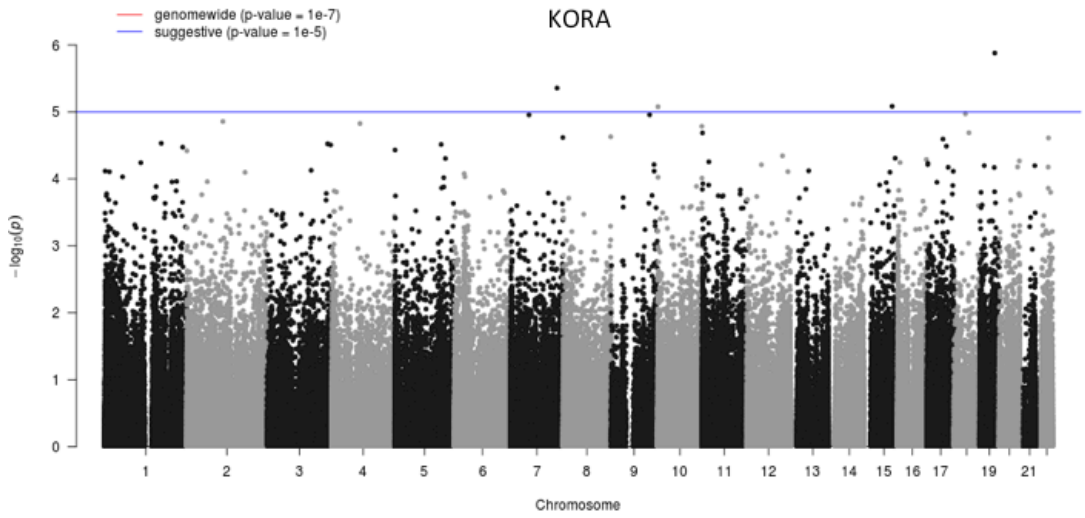

KORA
genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
$-\log_{10}(p)$
Chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 21

VIII. HRT use EWAS Manhattan plots

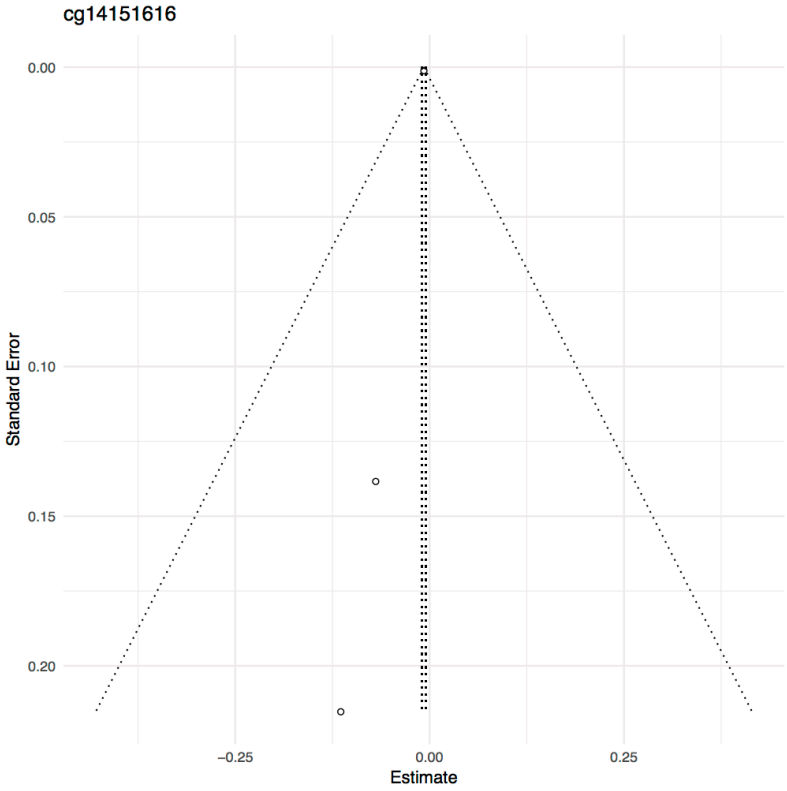


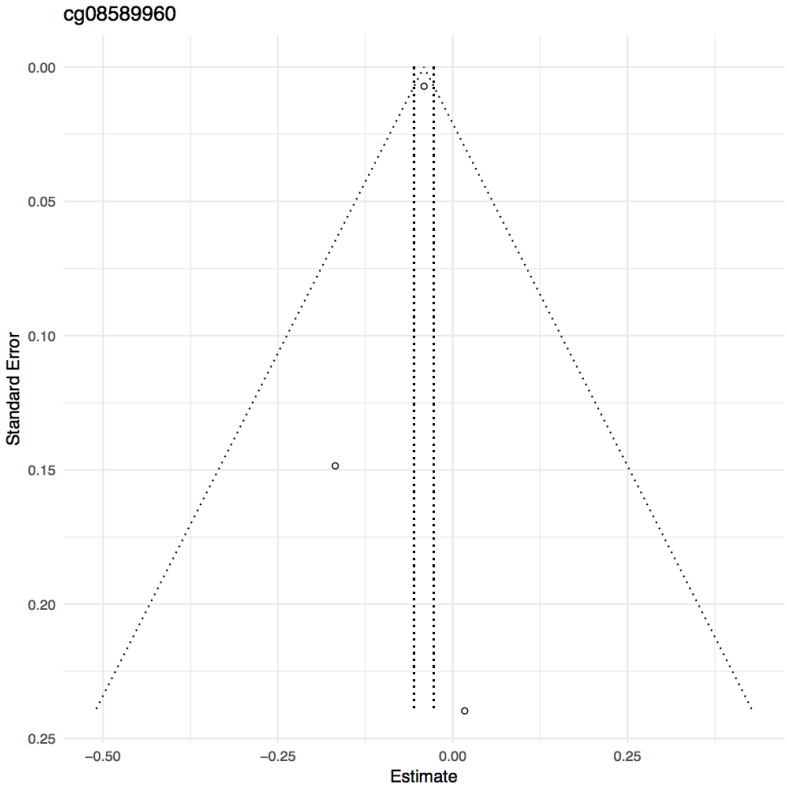


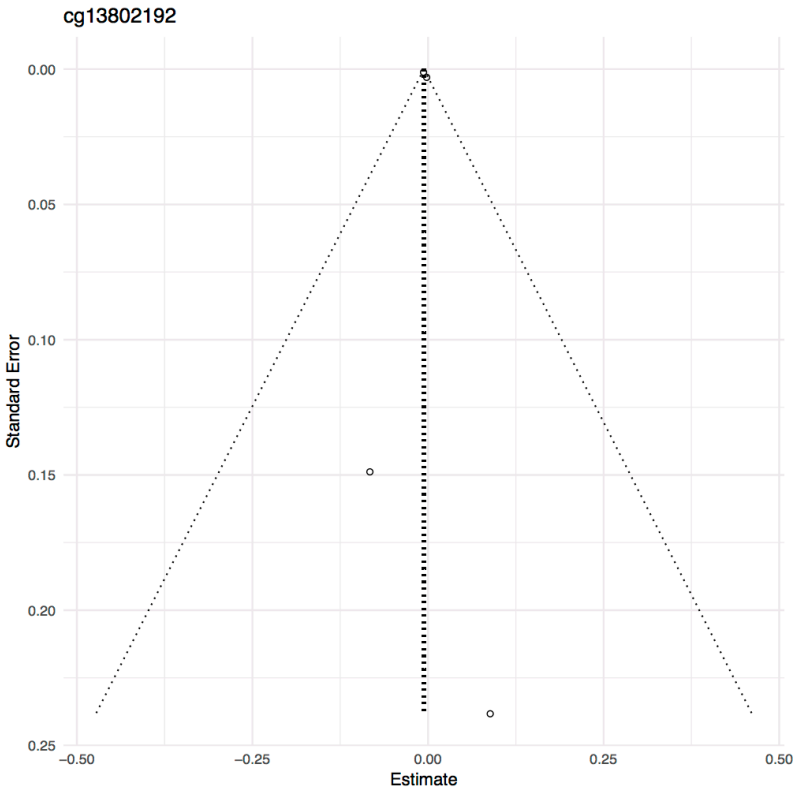


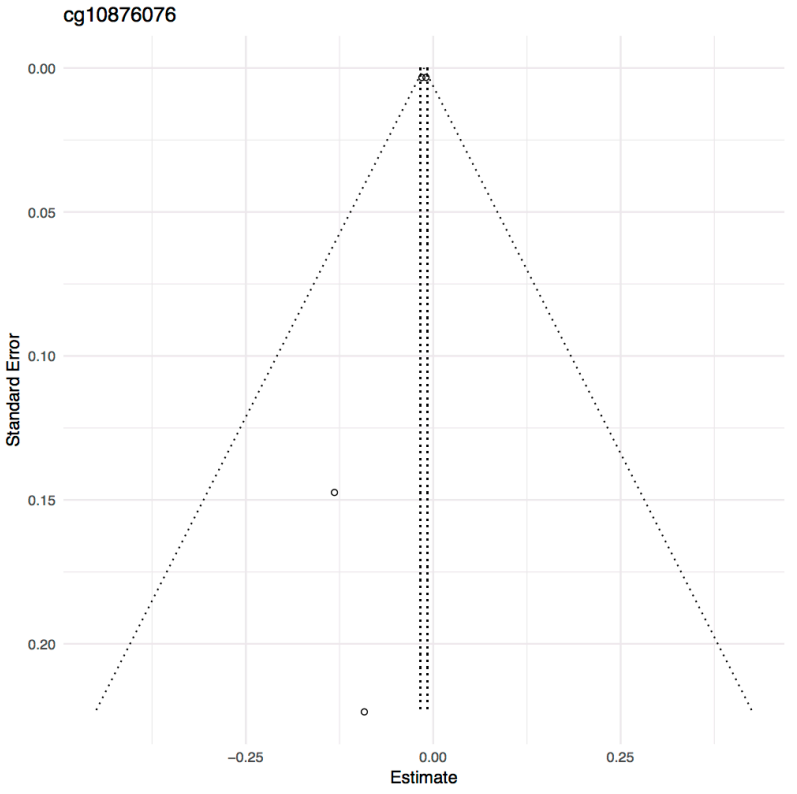

genomewide (p-value = 1e-7)
suggestive (p-value = 1e-5)
KORA
−log10(p)
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

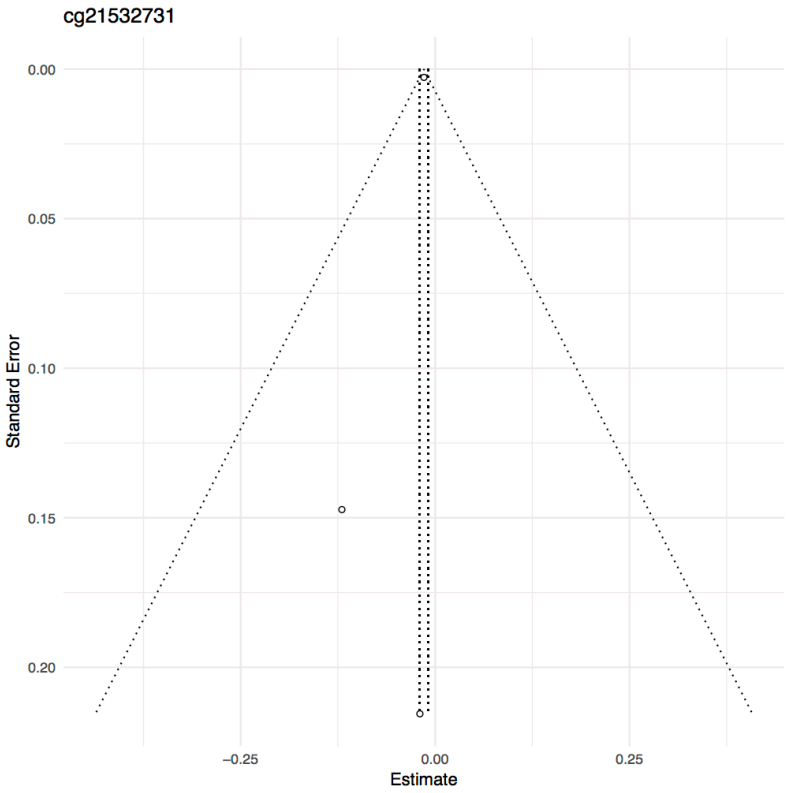
IX. HRT-associated DMPs funnel plots

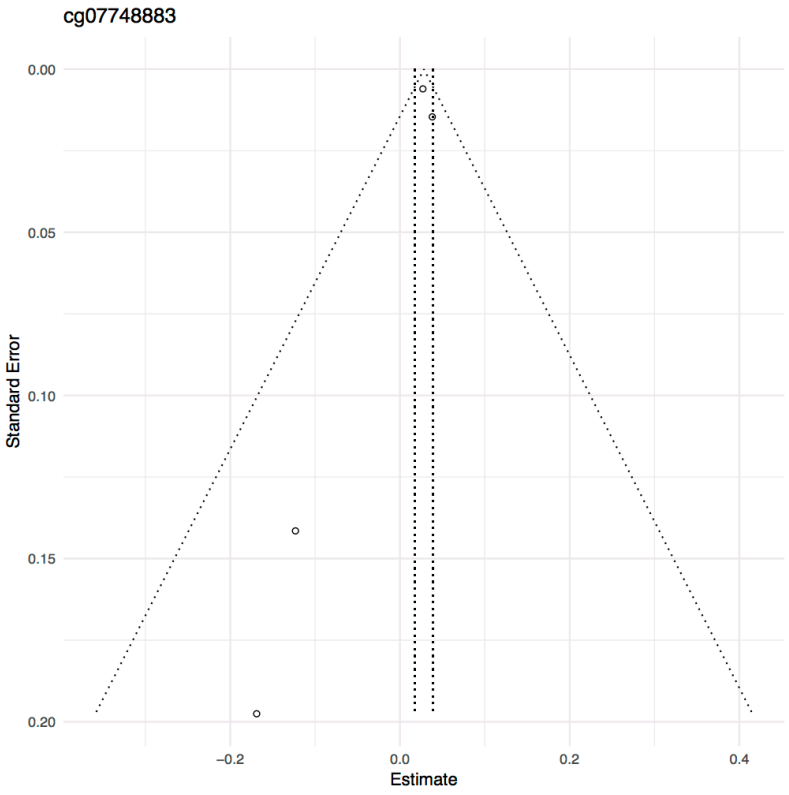


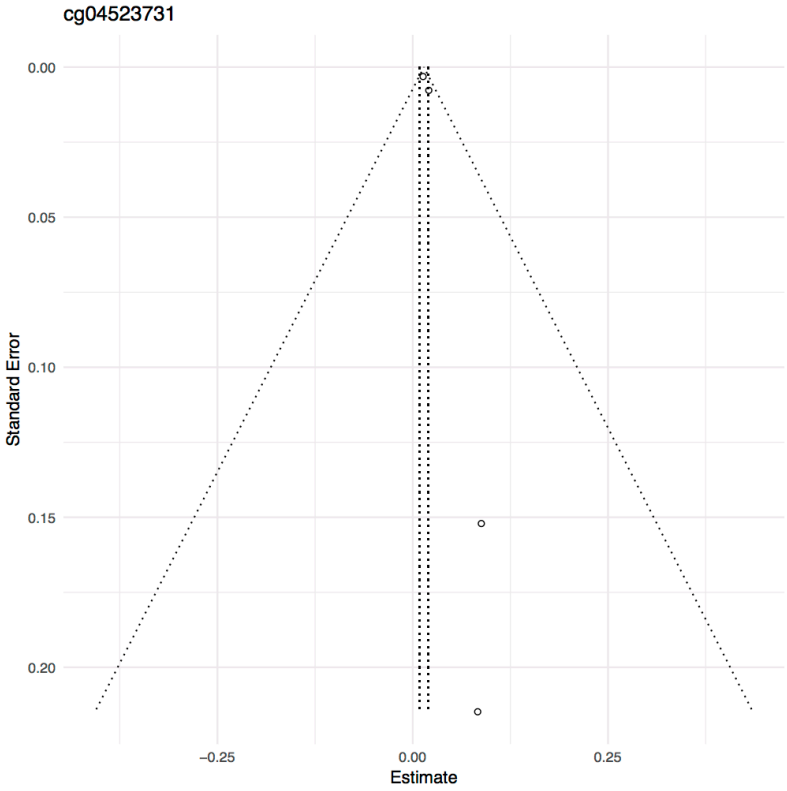



cg13802192
Standard Error
0.00
0.05
0.10
0.15
0.20
0.25
−0.50
−0.25
0.00
0.25
0.50
Estimate


cg10876076
Standard Error
0.00
0.05
0.10
0.15
0.20
−0.25
0.00
0.25
Estimate


cg21532731
0.00
0.05
0.10
0.15
0.20
Standard Error
−0.25
0.00
0.25
Estimate


cg07748883
0.00
0.05
0.10
0.15
0.20
Standard Error
−0.2
0.0
0.2
0.4
Estimate



X. DNA methylation changes at infertility genes in newborn twins conceived by in vitro fertilization (attached manuscript)

Castillo-Fernandez et al. Genome Medicine (2017) 9:28
DOI 10.1186/s13073-017-0413-5

Genome Medicine

RESEARCH Open Access

DNA methylation changes at infertility genes in newborn twins conceived by *in vitro* fertilisation

Juan E. Castillo-Fernandez[1], Yuk Jing Loke[2], Sebastian Bass-Stringer[2], Fei Gao[3], Yudong Xia[3], Honglong Wu[3], Hanlin Lu[3], Yuan Liu[3], Jun Wang[3,4,5,6], Tim D. Spector[1*], Richard Saffery[7,8*], Jeffrey M. Craig[2,7*†] and Jordana T. Bell[1*†]

Abstract

Background: The association of *in vitro* fertilisation (IVF) and DNA methylation has been studied predominantly at regulatory regions of imprinted genes and at just thousands of the ~28 million CpG sites in the human genome.

Methods: We investigated the links between IVF and DNA methylation patterns in whole cord blood cells ($n = 98$) and cord blood mononuclear cells ($n = 82$) from newborn twins using genome-wide methylated DNA immunoprecipitation coupled with deep sequencing.

Results: At a false discovery rate (FDR) of 5%, we identified one significant whole blood DNA methylation change linked to conception via IVF, which was located ~3 kb upstream of *TNP1*, a gene previously linked to male infertility. The 46 most strongly associated signals (FDR of 25%) included a second region in a gene also previously linked to infertility, *C9orf3*, suggesting that our findings may in part capture the effect of parental subfertility. Using twin modelling, we observed that individual-specific environmental factors appear to be the main overall contributors of methylation variability at the FDR 25% IVF-associated differentially methylated regions, although evidence for methylation heritability was also obtained at several of these regions. We replicated previous findings of differential methylation associated with IVF at the *H19/IGF2* region in cord blood mononuclear cells, and we validated the signal at *C9orf3* in monozygotic twins. We also explored the impact of intracytoplasmic sperm injection on the FDR 25% signals for potential effects specific to male or female infertility factors.

Conclusions: To our knowledge, this is the most comprehensive study of DNA methylation profiles at birth and IVF conception to date, and our results show evidence for epigenetic modifications that may in part reflect parental subfertility.

Keywords: *In vitro* fertilisation (IVF), Epigenomics, DNA methylation, MeDIP-seq

* Correspondence: tim.spector@kcl.ac.uk; richard.saffery@mcri.edu.au; jeff.craig@mcri.edu.au; jordana.bell@kcl.ac.uk
†Equal contributors
[1]Department of Twin Research and Genetic Epidemiology, King's College London, SE1 7EH London, UK
[7]Department of Paediatrics, University of Melbourne, Parkville, VIC, Australia
[2]Early Life Epigenetics, Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, VIC, Australia
Full list of author information is available at the end of the article

BioMed Central

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

As the frequency of *in vitro* fertilisation (IVF) treatment increases worldwide, much research effort has focused on exploring both short- and long-term health outcomes associated with conception via IVF, with contradictory results. A number of studies have observed associations with adverse perinatal and obstetric outcomes, including low birth weight, preterm birth, perinatal mortality, congenital malformations, placental complications, and increased frequency of imprinting disorders such as Angelman syndrome and Beckwith-Wiedemann syndrome [1–4]. On the other hand, parallel efforts have reported that these associations are not attributed to IVF treatment itself, but rather to multiple pregnancy or parental subfertility, both common factors in IVF births [5, 6]. Further research is required to identify potential factors associated with conception via IVF, including not only health outcomes but also biological consequences such as epigenetic modifications.

Given that birth weight and imprinting disorders are controlled at least in part by epigenetic factors [7, 8], IVF may have an influence on epigenetic profiles, potentially resulting in changes that persist well after birth and over the life course. Epigenetic mechanisms are considered possible mediators of the developmental origins of health and disease [9]; therefore, an assessment of the influence of IVF on DNA methylation profiles may give some insights into mechanisms underlying potential related health outcomes. Establishment of DNA methylation profiles in the germ line and embryo takes place early in development [10]. Theoretically, this epigenetic reprogramming could therefore be influenced by IVF-related interventions that occur very early, prior to blastocyst implantation. Indeed, induction of ovulation, embryo culturing, and cryopreservation, among others, have all been linked to specific alterations in DNA methylation in mice, although results are somewhat inconsistent [11–13].

Most studies in humans comparing naturally and IVF-conceived newborns have interrogated DNA methylation alterations targeting almost exclusively imprinted differentially methylated regions (DMRs). These studies have reported increased epigenetic variability at the *KvDMR1*, *PEG1*, and *H19* DMRs in umbilical cord blood [14], hypomethylation of the *H19* and *MEST* DMRs in placenta [15], and hypomethylation of the *H19* DMR in buccal epithelium [16] in individuals conceived by IVF. High-throughput approaches using bead array technology have also interrogated DNA methylation in IVF in a genome-wide manner. Katari et al. [17] reported differential methylation at 78 genes in cord blood and 40 in placenta with at least two differentially methylated CpG sites ($P \leq 0.08$) when looking across the promoters of 736 genes (GoldenGate Array, Illumina) in ten cases and 13 controls. A more extensive study using the promoter-enriched Illumina Infinium HumanMethylation27 bead array in cord blood samples from ten IVF cases and eight controls reported a total of 24 genes with at least two differentially methylated CpG sites ($P < 0.05$) [18]. More recently, a study used the genome-wide Illumina Infinium HumanMethylation450 bead array in samples from 38 IVF-conceived newborns followed by fresh embryo transfer, 38 IVF-conceived followed by cryopreserved embryo transfer, 18 born to subfertile parents after conception by intrauterine insemination, and 43 controls born to fertile parents [19]. This platform interrogates CpG sites across the whole genome, although with a limited coverage since it targets gene-centric annotations [20]. The authors identified differential methylation at multiple sites, including metastable epialleles.

Here, we interrogated evidence for differential methylation between IVF and non-IVF newborn twins in a more comprehensive manner by conducting epigenome-wide association scans (EWAS) [21] using methylated DNA immunoprecipitation followed by deep sequencing (MeDIP-seq) [22] genome-wide in samples from cord blood, and its mononuclear fraction, collected at birth from IVF and non-IVF twins. The use of twins in this study allowed the partition of the observed variance in DNA methylation into genetic and environmental factors. The approach also avoids potential spurious associations due to an imbalanced number of multiple and single pregnancies between conception method groups.

Methods

Subjects and sample collection

The study included 47 IVF and 60 non-IVF newborn twins (from 54 twin pairs) from the Peri/postnatal Epigenetic Twins Study (PETS), Melbourne, Australia. Recruitment and full study procedure have been described previously [23, 24]. Cord blood was collected at birth and used to process mononuclear cells by Ficoll gradient centrifugation as described previously [25]. Whole blood cells (WBCs) from cord blood were available for a total of 98 twins (40 IVF and 58 non-IVF) and cord blood mononuclear cells (CBMCs) for a total of 82 twins (35 IVF and 47 non-IVF). Maternal age and method of conception were determined via questionnaire at recruitment (18–20 weeks gestation). Twins of mothers who said yes to IVF or intracytoplasmic sperm injection (ICSI) treatment were classified as IVF regardless of the use of ovulation induction medication or other fertility treatments. Maternal smoking status was collected via questionnaire on recruitment and at 24 and 36 weeks of pregnancy. Birth weight was collected during the immediate neonatal period. Zygosity and chorionicity were determined by physical examination of the inter-placental membranes at birth, and by genetic

test when required, as described previously [23, 24]. Pregnancy complications were recorded and are shown in Additional file 1: Table S1.

DNA methylation profiling

MeDIP-seq was performed at BGI-Shenzhen, Shenzhen, China. Extracted DNA was fragmented using a Covaris sonication system and sequencing libraries were prepared from 5 μg fragmented genomic DNA. End repair, <A > base addition and adaptor ligation steps were performed using Illumina's Single-End DNA Sample Prep kit. Adaptor-ligated DNA was immunoprecipitated by anti-5mC using a commercial antibody (Diagenode) and MeDIP products were validated by quantitative PCR. MeDIP DNA was purified with ZYMO DNA Clean & Concentrator-5 columns and amplified using adaptor-mediated PCR. DNA fragments between 200 and 500 bp in size were gel-excised, and the amplification quality and quantity were evaluated by Agilent BioAnalyzer analysis. The libraries were subjected to highly parallel 50-bp single-end sequencing on the Illumina GAII platform. All sequencing data passed initial quality checks for base composition (no exclusions) using FASTQC v0.10.0. For each individual, ~30 million reads were generated and mapped onto hg19 using BWA. After removing duplicates, we filtered data using quality score Q10. We quantified methylation levels using MEDIPS [26], producing the mean relative methylation score (RPM) in 500-bp bins (overlap of 250 bp) across the genome. Altogether, there were 11,524,145 windows and these were used for the analyses. Bins with RPM values of zero in more than 50% of the samples were excluded, resulting in 9,592,803 (WBC) and 9,285,089 (CBMC) bins used in downstream analyses.

Epigenome-wide IVF-DMR analyses

Normalised (N(0,1)) methylation scores in each genomic bin were regressed using a linear mixed-effects model to account for twin structure (lme4 package [27] in R [28]). Tissue type, birth weight, sex, maternal smoking, 260/280 ratio, DNA concentration, and the loadings of the first five principal components were used as covariates and included as fixed effects in the model. Family and zygosity were included as random effects in the linear mixed model. The principal components were included to account for unknown sources of variation, such as cell heterogeneity. Correction for multiple testing was performed by a Benjamini-Hochberg false discovery rate (FDR) calculation.

Variance decomposition of WBC IVF-DMRs

The contribution of additive genetic (A), common environmental (C), and unique environmental (E) factors to DNA methylation was estimated using the ACE model based on the classic twin design [29]. The model was fitted using the OpenMX statistical package [30]. RPM values without adjustment for covariates were used to estimate the ACE proportions.

Statistical analysis

Pairwise correlations and principal components analysis were performed using RPM values across all bins with values > 0 in at least 50% of the samples. Hierarchical clustering was performed using Euclidean distance as a measure of dissimilarity and average linkage clustering.

Validation analysis

Genomic DNA (500 ng) was bisulphite converted using the MethylEasy Exceed Rapid Bisulphite Modification Kit (Human Genetic Signatures, North Ryde, NSW, Australia). Primers to target the regions in *TNP1* and *C9orf3* were designed using the EpiDesigner tool (Sequenom Inc., Herston, QLD, Australia). The *H19 CTCF6* region was the same used in a previous study [25]. Primers, genomic coordinates, and PCR conditions are shown in Additional file 1: Table S2. Methylation levels were determined by EpiTYPER on the MassARRAY System (Sequenom Inc., Herston, QLD, Australia). Statistical analysis considered the average of two to three technical replicates and were performed using data on single CpG sites.

Results

Genome-wide methylation profiles in twins

We profiled DNA methylation levels from a total of 107 newborn twins (47 conceived via IVF and 60 conceived in vivo) in WBCs and CBMCs. Details of any fertility treatment used and demographic characteristics that represent potential confounders of DNA methylation levels at birth, such as sex, birth weight, maternal age, and maternal smoking status, are shown in Table 1. We first explored the genome-wide patterns of DNA methylation variability in the dataset. Principal component analysis was used to identify factors that were significantly associated with genome-wide variability in DNA methylation profiles. The first five principal components in the dataset, which explained ~13% of the total variance in DNA methylation, were at least nominally associated ($P < 0.05$) with sample type (WBCs versus CMBCs), birth weight, maternal smoking, and conception method (Fig. 1a).

We next estimated the within twin-pair correlation patterns in methylation profiles of twin pairs available in both datasets using Pearson's correlation. In concordance with previous studies [7], we observed higher median correlation within monozygotic (MZ) twin pairs compared to dizygotic (DZ) twin pairs (Fig. 1b). Previous studies have shown that twin chorionicity can have an effect on within-pair DNA

Table 1 Breakdown of samples used for the identification of IVF-DMRs and potential covariates

Group	Total number of twins (number of complete sets)	Zygosity and chorionicity[a]	Sex[b]	Birth weight (kg): mean(sd)	Maternal age (years): mean (sd)	Maternal smoking (percentage smokers)	Ovarian stimulation	ICSI	GIFT	Frozen embryo
WBCs										
IVF	40 (20)	10 MZ MC 30 DZ DC	18 (F) 22 (M)	2.57 (4.77)	36 (4)	20%	6 (No) 34 (Yes)	22 (No) 18 (Yes)	-	28 (No) 12 (Yes)
Non-IVF	58 (29)	14 MZ MC 12 MZ DC 32 DZ DC	34 (F) 24 (M)	2.58 (3.99)	32 (5)	28%	56 (No) 2 (Yes)	-	56 (No) 2 (Yes)	-
CBMCs										
IVF	35 (16)	9 MZ MC 1 MZ DC 25 DZ DC	16 (F) 19 (M)	2.50 (4.42)	35 (5)	23%	1 (No) 34 (Yes)	14 (No) 21 (Yes)	-	24 (No) 11 (Yes)
Non-IVF	47 (22)	12 MZ MC 10 MZ DC 25 DZ DC	30 (F) 17 (M)	2.60 (3.48)	32 (4)	28%	45 (No) 2 (Yes)	-	45 (No) 2 (Yes)	-

[a]*MZ* monozygotic, *DZ* dizygotic, *MC* monochorionic, *DC* dichorionic
[b]*F* female, *M* male
GIFT gamete intra-fallopian transfer, *ICSI* intracytoplasmic sperm injection, *sd* standard deviation

methylation differences, but not with consistent direction of effect across tissues [7, 25, 31]. In our study, we did not observe significant chorionicity-related methylation differences (Fig. 1b), but the number of MZ twins within chorionicity categories was relatively low ($n = 8$ monochorionic and $n = 5$ dichorionic pairs). Interestingly, the method of conception showed methylation profile differences within MZ twin pairs. MZ IVF twins had higher median correlation compared to MZ non-IVF twins in WBCs, but the opposite trend was observed in CBMCs, and in both cases the MZ IVF sample was very small ($n = 3$).

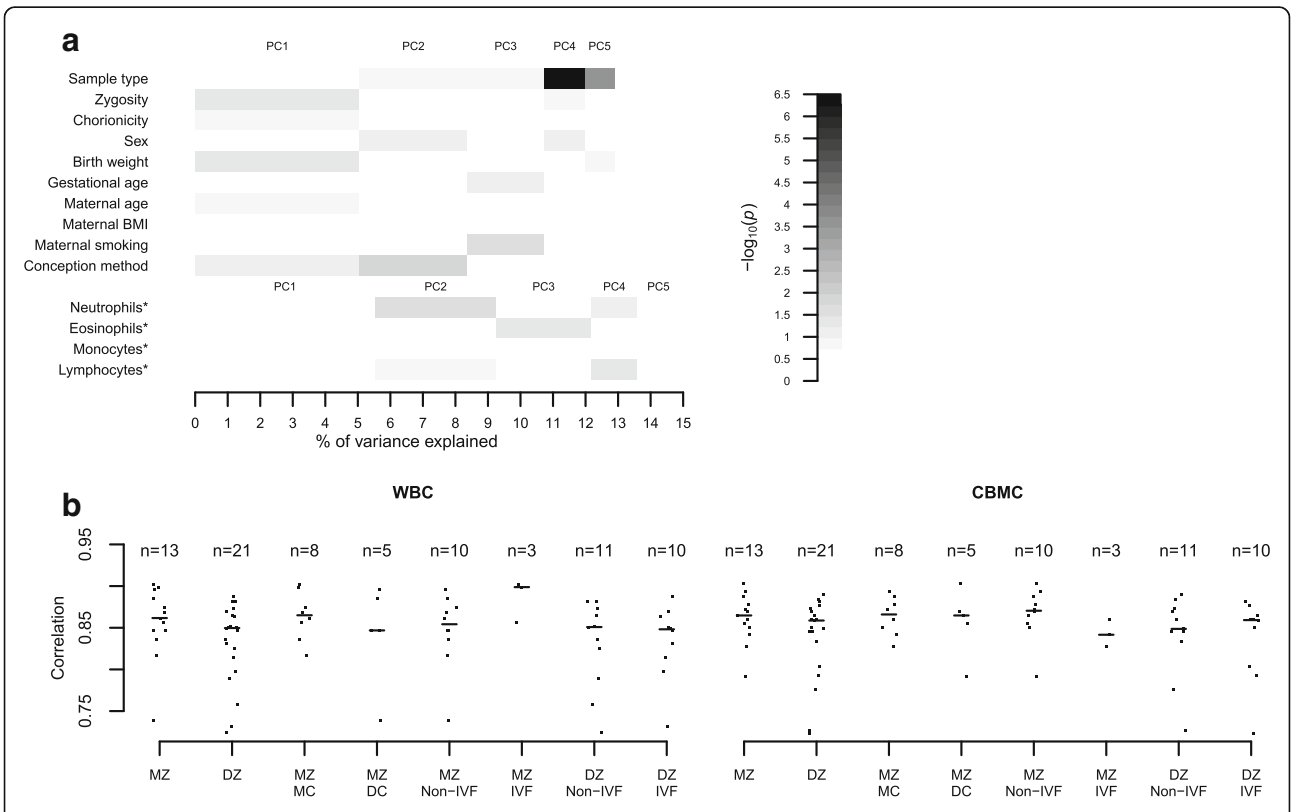


Fig. 1 Global methylation patterns. **a** Biological factors associated with principal components of variation of methylation profiles. Variables marked with an *asterisk* were only available in a subset of the sample ($n = 54$). **b** Within-pair methylation correlation in WBCs and CBMCs. *BMI* body mass index, *DC* dichorionic, *DZ* dizygotic, *MC* monochorionic, *MZ* monozygotic, *PC* principal component

IVF-DMRs in CBMCs and WBCs

In order to identify tissue-independent and tissue-specific IVF-associated DMRs, we compared DNA methylation profiles in WBCs and CBMCs in relation to method of conception adjusting for birth weight, sex, maternal smoking, and the first five principal components, which partly capture cell heterogeneity. Epigenome-wide analyses of DNA methylation in relation to method of conception did not identify genome-wide significant signals in the CBMCs subset or in the combined CBMC and WBC datasets, after correction for multiple testing. In WBCs alone, one significant DMR was observed at a FDR of 5% (Fig. 2). This was located ~3 kb upstream of *TNP1* (chr2:217,726,751–217,727,250), which encodes a transition nuclear protein that replaces histones and is subsequently replaced by protamines during spermiogenesis. A deletion in the promoter region of this gene, which reduces its expression, has been reported in infertile men [32]. Methylation upstream of *TNP1* might have an impact on its expression. In mice, methylation changes during spermatogenesis have been observed at *TNP1*, which suggests a role of methylation in the regulation of this gene [33]. To explore the biological characteristics of the top-ranked results in the IVF epigenome-wide analyses we selected a more liberal threshold of FDR 25%, at which 46 IVF-DMRs were identified (Table 2). Interestingly, the third-ranked DMR genome-wide (Additional file 1: Figure S1) was located in the first intron of *C9orf3* (chr9:97,504,001–97,504,500), which has been associated with polycystic ovary syndrome in women [34] and development of erectile dysfunction after radiotherapy for prostate cancer in men [35]. Another signal within this list was located in intron 1 of *STOX2* (chr4:184,814,001–184,814,500), whose reduced expression has been implicated in pre-eclampsia [36]. Since adverse perinatal outcomes may be associated with maternal age, we further adjusted for this covariate and observed that the 46 FDR 25% WBC IVF-DMRs remained significant (Table 2).

The non-IVF group included a small number ($n = 4$) of newborns conceived with other types of fertility treatments not equivalent to IVF, such as gamete intrafallopian transfer (GIFT) and ovarian stimulation. We re-analysed the 46 FDR 25% WBC IVF-DMRs excluding GIFT ($n = 2$) and non-IVF ovarian stimulation ($n = 2$) controls and observed that conclusions remained unchanged (Additional file 1: Table S3).

Hierarchical clustering using DNA methylation levels at these 46 FDR 25% DMRs alone grouped twins by method of conception, assigning 38 out of 40 IVF twins and 57 out of 58 non-IVF twins to the correct group (Fig. 3). We also explored these signals with respect to functional annotations. A total of ten FDR 25% WBC IVF-DMRs overlapped CpG sites previously shown to be dynamic during development [37], 20 overlapped DNase I hypersensitivity sites (wgEncodeRegDnaseClusteredV3) [38], one overlapped a CpG island (cpgIslandEx) [39],

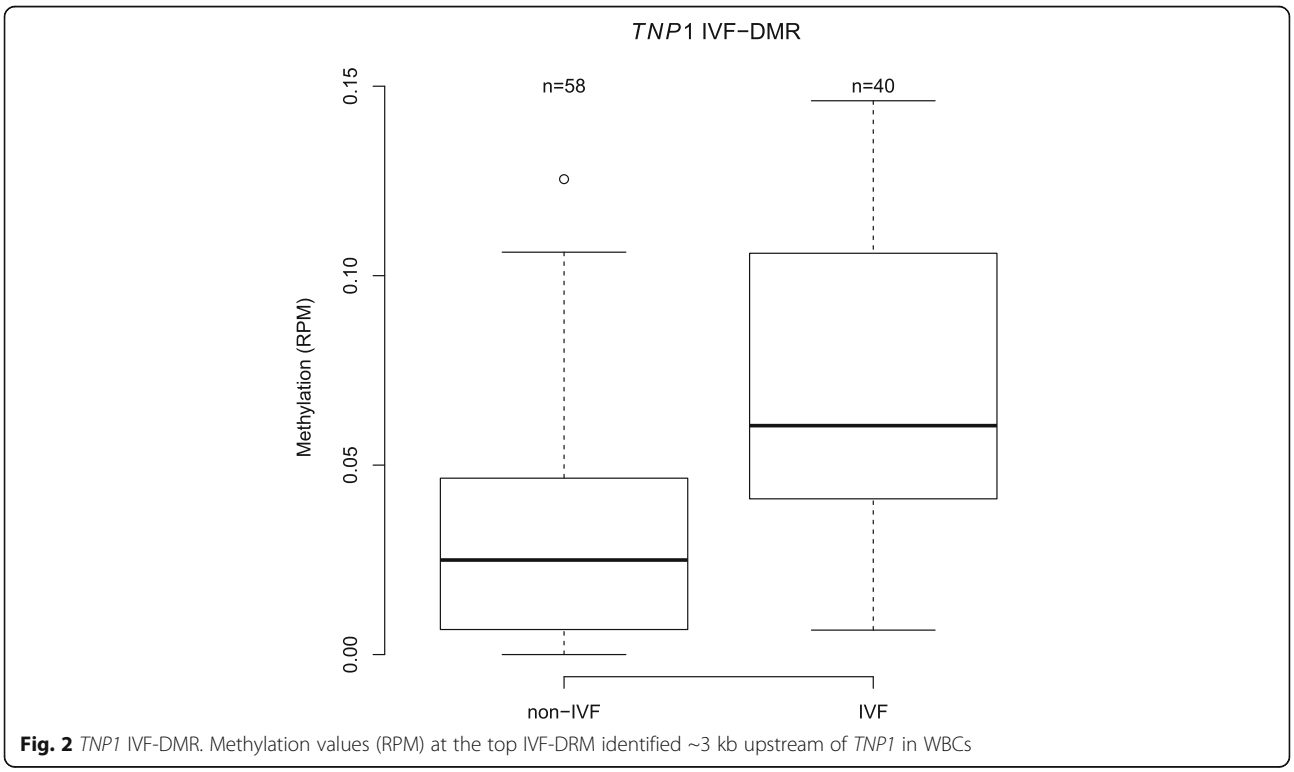


Fig. 2 *TNP1* IVF-DMR. Methylation values (RPM) at the top IVF-DRM identified ~3 kb upstream of *TNP1* in WBCs

Table 2 FDR 25% WBC IVF-DMRs

Chromosome	Start	End	*P*	FDR adjusted *P*	*P* (adjusted for maternal age)	Gene name[a]	Gene start[a]	Gene end[a]
Chr2	217726751	217727250	2.30E-09	0.0221	6.40E-10	*AC007557.1*	217735495	217736362
						TNP1	217724181	217724787
Chr5	178761751	178762250	5.43E-08	0.1244	1.76E-05	*ADAMTS2*	178537852	178772431
Chr9	97504001	97504500	5.83E-08	0.1244	5.14E-07	*C9orf3*	97488983	97849441
Chr5	9275751	9276250	5.86E-08	0.1244	1.67E-06	*SEMA5A*	9035138	9546187
Chr4	184814001	184814500	7.95E-08	0.1244	4.96E-07	*STOX2*	184774584	184944679
Chr5	142488501	142489000	8.73E-08	0.1244	5.76E-07	*ARHGAP26*	142149949	142608576
Chr9	118148751	118149250	9.20E-08	0.1244	4.61E-06	*DEC1*	117904097	118164923
Chr9	118149001	118149500	1.04E-07	0.1244	3.47E-06	*DEC1*	117904097	118164923
Chr11	82654251	82654750	1.30E-07	0.1250	5.26E-08	*C11orf82*	82611017	82669319
						PRCP	82534544	82681626
						RAB30	82684175	82782965
Chr19	6165251	6165750	1.40E-07	0.1250	1.00E-06	*RFX2*	5993175	6199583
						ACSBG2	6135258	6193112
						MLLT1	6212966	6279959
Chr1	85522251	85522750	1.43E-07	0.1250	2.56E-07	*WDR63*	85464830	85598821
						MCOLN3	85483765	85514182
Chr17	42569001	42569500	1.64E-07	0.1274	1.90E-05	*GPATCH8*	42472652	42580798
Chr4	141606501	141607000	2.03E-07	0.1274	1.59E-05	*TBC1D9*	141541919	141677274
Chr5	137736001	137736500	2.06E-07	0.1274	1.13E-06	*REEP2*	137774706	137782658
						KDM3B	137688285	137772717
Chr5	150614501	150615000	2.14E-07	0.1274	2.23E-07	*SLC36A3*	150656323	150683327
						GM2A	150591711	150650001
						CCDC69	150560613	150603706
Chr17	36918251	36918750	2.32E-07	0.1274	2.02E-07	*MLLT6*	36861795	36886056
						CISD3	36886488	36891297
						CWC25	36956687	36981734
						PIP4K2B	36921942	36956379
						PCGF2	36890150	36906070
						CTB-58E17.5	36905613	36906969
						PSMB3	36908989	36920484
						AC006449.1	36884086	36884451
Chr6	126138251	126138750	2.36E-07	0.1274	2.12E-06	*NCOA7*	126102307	126252266
Chr7	144431251	144431750	2.70E-07	0.1274	5.29E-09	*TPK1*	144149034	144533488
Chr12	70937251	70937750	2.76E-07	0.1274	1.83E-06	*PTPRB*	70910630	71031220
Chr4	141606251	141606750	2.80E-07	0.1274	1.50E-05	*TBC1D9*	141541919	141677274
Chr13	90019001	90019500	2.84E-07	0.1274	5.94E-07	-	-	-
Chr11	74179001	74179500	2.92E-07	0.1274	1.61E-09	*LIPT2*	74202757	74204778
						POLD3	74204896	74380162
						KCNE3	74165886	74178774
Chr12	99153001	99153500	3.93E-07	0.1637	2.44E-07	*ANKS1B*	99120235	100378432
						APAF1	99038919	99129204
Chr2	223336751	223337250	4.47E-07	0.1788	2.00E-06	*SGPP2*	223289236	223425667
Chr8	120972001	120972500	4.98E-07	0.1869	3.26E-08	*DEPTOR*	120885957	121063152

Table 2 FDR 25% WBC IVF-DMRs *(Continued)*

Chr17	38047001	38047500	5.25E-07	0.1869	7.83E-06	*GSDMB*	38060848	38076107
						ZPBP2	38024417	38034149
						IKZF3	37921198	38020441
						ORMDL3	38077294	38083854
Chr4	64626751	64627250	5.45E-07	0.1869	2.67E-06	-	-	-
Chr16	87256751	87257250	5.51E-07	0.1869	2.12E-08	*C16orf95*	87117168	87351022
Chr19	10656751	10657250	5.77E-07	0.1869	1.42E-06	*CDKN2D*	10677138	10679735
						ATG4D	10654571	10664094
						KEAP1	10596796	10614417
						AP1M2	10683347	10697991
						KRI1	10663761	10676713
						S1PR5	10623623	10628607
Chr7	2487251	2487750	5.85E-07	0.1869	2.03E-05	*CHST12*	2443223	2474242
Chr11	74178751	74179250	6.82E-07	0.2059	1.31E-07	*KCNE3*	74165886	74178774
						LIPT2	74202757	74204778
						POLD3	74204896	74380162
Chr10	119176501	119177000	6.87E-07	0.2059	1.11E-06	*PDZD8*	119040000	119134978
Chr22	34755251	34755750	7.32E-07	0.2094	1.76E-05	-	-	-
Chr6	161664751	161665250	7.42E-07	0.2094	2.93E-06	*AGPAT4*	161551011	161695093
Chr16	17161751	17162250	7.80E-07	0.2137	2.41E-06	*XYLT1*	17195626	17564738
Chr18	23695001	23695500	8.39E-07	0.2235	1.97E-06	*PSMA8*	23713816	23773319
						SS18	23596578	23671181
Chr9	26364751	26365250	8.90E-07	0.2267	1.02E-06	-	-	-
Chr1	25227001	25227500	9.05E-07	0.2267	1.61E-05	*RUNX3*	25226002	25291612
Chr13	68877251	68877750	9.43E-07	0.2267	2.19E-05	-	-	-
Chr9	89126501	89127000	9.59E-07	0.2267	7.00E-06	-	-	-
Chr13	35317501	35318000	9.72E-07	0.2267	2.51E-05	-	-	-
Chr21	19575001	19575500	9.92E-07	0.2267	2.31E-06	*CHODL*	19273580	19639690
Chr2	169470001	169470500	1.06E-06	0.2307	4.35E-07	*CERS6*	169312372	169631644
Chr12	4310251	4310750	1.06E-06	0.2307	8.13E-05	-	-	-
Chr6	157136501	157137000	1.15E-06	0.2448	1.61E-06	*ARID1B*	157099063	157531913
Chr14	104067251	104067750	1.17E-06	0.2448	8.21E-06	*APOPT1*	104029299	104073860
						BAG5	104022881	104029168
						KLC1	104028233	104167888
						RP11-73 M18.2	104029299	104152261

[a]From GENCODE v19

and none overlapped with candidate metastable epialleles [40] (Fig. 3).

Cell type-specific DNA methylation can impact the profiles observed in a population of cells, such as in a whole blood sample, and we therefore accounted for blood cell type heterogeneity using a twofold approach. First, we performed principal component analysis on the methylation levels of the entire set of WBC samples, and our main EWAS analyses above are corrected for the first five principal components, which likely capture variation attributed to technical and biological factors, potentially including cell heterogeneity. To assess whether the first five principal components capture cell heterogeneity, blood cell subtype counts were obtained through automatic differential counting for a subset of the WBC samples ($n = 54$ twins, 22 IVF, and 32 non-IVF) and these were compared against the distributions of the first five principal components. The proportion of neutrophils, eosinophils, and lymphocytes were associated ($P < 0.05$) with the loadings of the second, third, and fourth principal

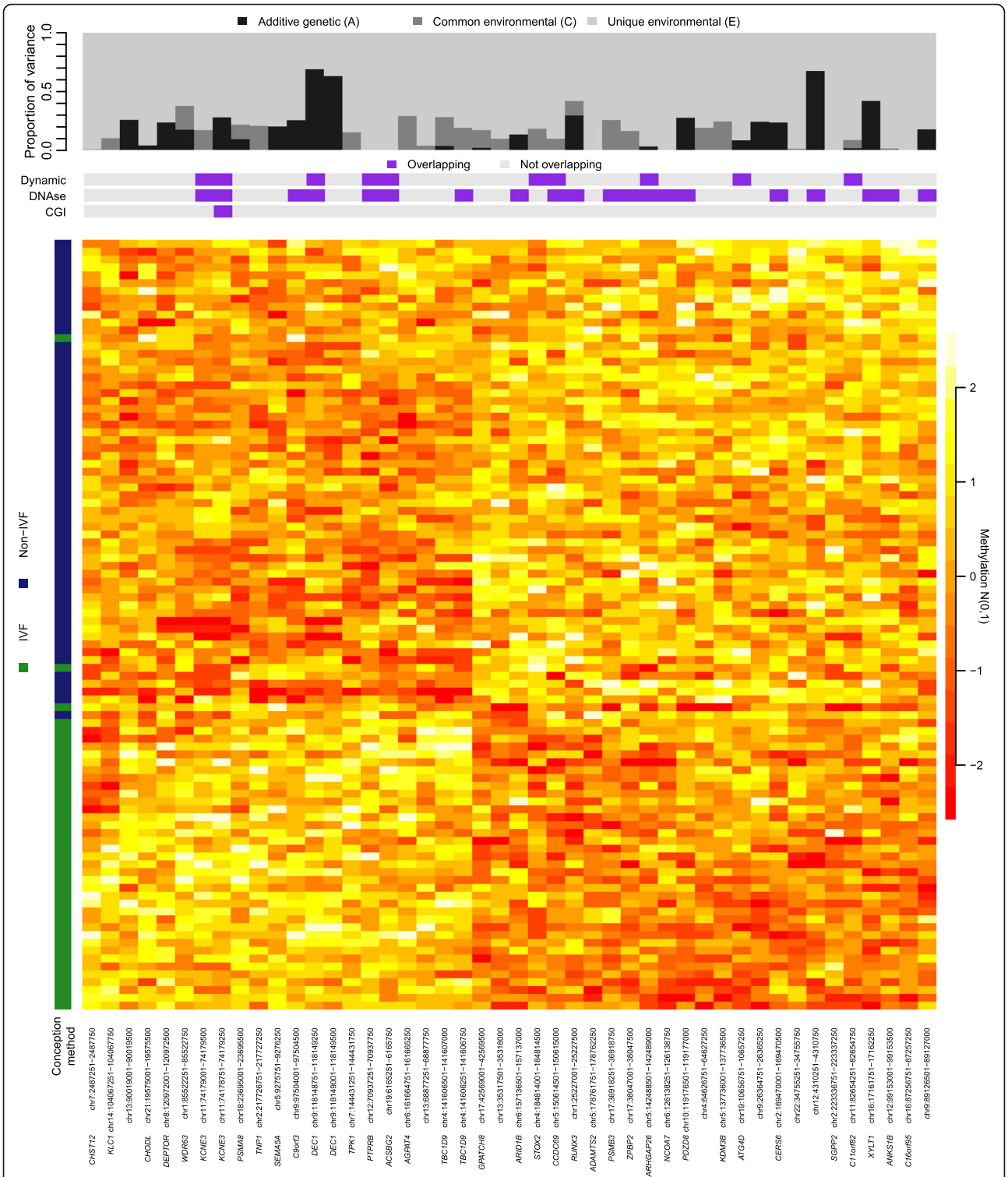


Fig. 3 WBC IVF-DMRs. Heatmap rows correspond to the 98 WBC samples while columns correspond to the 46 FDR 25% WBC IVF-DMRs. The *vertical colour bar* indicates method of conception (IVF, *green*; non-IVF, *blue*). *Top panel* shows the fraction of variance explained by additive genetic (A), shared environmental (C), and unique environmental (E) factors. *Horizontal colour bars* indicate overlap (*violet*) or absence (*gray*) of dynamic CpG sites, DNase I hypersensitivity sites, or CpG islands with the DMR

components, respectively (Fig. 1a). Therefore, since the EWAS model used in this study took into account the loadings of the first five principal components, these analyses already take into account the influence of cell heterogeneity to a certain extent.

Second, we re-analysed the 46 FDR 25% WBC IVF-DMRs in the subset of 54 WBC samples with available cell counts, adjusting for the proportion of neutrophils, eosinophils, monocytes, and lymphocytes. We also performed analyses adjusting for the loadings of the first five principal components within this dataset alone. Most results were concordant when comparing across all models (Additional file 1: Table S4) and only five out the 46 FDR 25% WBC IVF-DMRs were not significant ($P > 0.05$) after adjusting for cell proportions (chr8:120,972,001–120,972,500, chr7:2,487,251–2,487,750, chr18:23,695,001–23,695,500, chr12:4,310,251–4,310,750, and chr14:104,067,251–104,067,750).

Variance decomposition of WBC IVF-DMRs

Given that epigenetic changes were potentially affecting infertility genes, we wanted to investigate if the findings may capture a genetic signature affecting DNA methylation that could be transmitted to offspring. We applied twin variance decomposition analyses to partition the total epigenetic variance into additive genetic (A) and common (C) and unique (E) environmental components (ACE) [29]. The ACE model was used to determine the contribution of genetics, shared intrauterine environment due to shared maternal influences, and non-shared (twin-specific) or stochastic factors to epigenetic variation. The mean contribution of additive genetic effects (narrow-sense heritability) to DNA methylation across the genome in different tissues from newborns has been previously estimated to be between 0.05 and 0.12 [7]. Here we estimated the average genome-wide narrow-sense heritability for DNA methylation in WBCs at 0.06. At the 46 FDR 25% WBC IVF-DMRs, the major contributors to DNA methylation variation were non-shared or stochastic events (Fig. 3). However, several FDR 25% IVF-DMRs had evidence for heritability ($A > 0.4$), suggestive of genetic effects underlying specific IVF-associated DNA methylation changes. These included an intronic region in *DEC1* (chr9:118,148,751–118,149,500), a region 33 kb away from *XYLT1* (chr16:17,161,751–17,162,250), and an intergenic region in chromosome 12 (chr12:4,310,251–4,310,750). When looking at the two DMRs associated with infertility genes, DNA methylation variation showed no evidence for genetic effects ($A = 0$) near *TNP1*, while heritability at the DMR in *C9orf3* was estimated at 0.25.

Effects of IVF on imprinting

Previous studies have explored DNA methylation patterns in IVF births specifically at imprinting control regions (ICRs). We therefore assessed whether there was an enrichment of differential methylation effects at 34 known ICRs [41] in our genome-wide results, but no enrichment was observed ($P > 0.05$). However, when we explored individual signals at candidate IVF-DMRs we were able to replicate one previously reported ICR IVF-associated DMR. Concordantly with previous IVF methylation studies in placental tissue [15] and buccal epithelium [16], we observed hypomethylation in IVF twins at the sixth CTCF binding site within the *H19*/*IGF2* (*H19 CTCF6*) DMR (Additional file 1: Figure S2). This association was observed in CBMCs ($P = 0.01$), but not in WBCs.

Effects of ICSI

ICSI is a technique in IVF used to treat couples with male-factor infertility [42]. In contrast to conventional IVF where fertilisation occurs by placing spermatozoa near an egg, ICSI consists of the direct injection of a selected single sperm cell into the egg. This manipulation may introduce additional risk factors [43]. To assess the effect of ICSI on the 46 FDR 25% WBC IVF-DMRs we adjusted for the use of this technology and also compared the ICSI and the conventional IVF groups separately against the non-IVF group. After adjustment for ICSI, the association weakened at several FDR 25% IVF-DMRs (Table 3), suggesting that ICSI or paternal infertility might have a role in these methylation changes. One FDR 25% IVF-DMR signal (chr1:85,522,251–85,522,750) appeared stronger after adjustment, suggesting either a female infertility effect or that ICSI prevents or corrects a methylation change that occurs in conventional IVF. This DMR was located upstream of *WDR63*, a gene mainly expressed in testis, fallopian tube, and adrenal gland [44].

Validation of IVF-DMRs

We pursued validation of the differential methylation signals at the top associated DMR (located ~3 kb upstream of *TNP1*) and at the third-ranked DMR (located in *C9orf3*), both in or near genes previously linked to infertility. Altogether, four CpG sites were targeted for validation using Sequenom's EpiTYPER technology.

For the DMR in *C9orf3*, we were able to target two CpG sites within the most-associated 500-bp bin in this locus (Additional file 1: Figure S3). We assayed methylation levels in 36 MZ twins included in the discovery EWAS and observed significantly higher methylation in the IVF group, concordant with the MeDIP-seq analysis, at both tested CpG sites in the *C9orf3 locus* ($P = 0.02$ and 0.03, respectively), therefore validating this signal using a different methylation profiling approach (Additional file 1: Figure S4).

For the *TNP1* DMR we were unable to target CpGs within the most associated 500-bp bin, and we therefore

Table 3 Effect of ICSI on the FDR 25% WBC IVF-DMRs

			IVF ($n = 40$) versus non-IVF ($n = 58$)			IVF ($n = 34$) versus non-IVF ($n = 58$) adjusted for ICSI			Conventional IVF ($n = 16$) versus non-IVF ($n = 58$)			ICSI ($n = 18$) versus non-IVF ($n = 58$)		
Chromosome	Start	End	Estimate	SE	*P*	Estimate	SE	*P*	Estimate	SE	*P*	Estimate	SE	*P*
Chr2	217726751	217727250	1.18	0.19	2.30E-09	1.45	0.26	7.71E-08	−1.42	0.26	7.44E-08	−1.16	0.27	1.29E-05
Chr5	178761751	178762250	−1.08	0.2	5.43E-08	−1.09	0.29	8.23E-05	0.87	0.30	1.89E-03	1.11	0.27	2.52E-05
Chr9	97504001	97504500	1.07	0.2	5.83E-08	0.77	0.27	2.77E-03	−0.88	0.31	2.67E-03	−1.30	0.25	1.95E-07
Chr5	9275751	9276250	1.09	0.19	5.86E-08	1.17	0.27	1.04E-05	−1.13	0.29	6.50E-05	−1.13	0.27	2.75E-05
Chr4	184814001	184814500	−0.96	0.18	7.95E-08	−0.97	0.24	3.94E-05	0.94	0.26	1.37E-04	1.15	0.23	8.35E-07
Chr5	142488501	142489000	−1.11	0.2	8.73E-08	−1.1	0.28	7.60E-05	1.25	0.31	5.16E-05	1.08	0.28	7.10E-05
Chr9	118148751	118149250	1.09	0.2	9.20E-08	1.35	0.28	1.28E-06	−1.30	0.29	8.90E-06	−1.02	0.28	1.68E-04
Chr9	118149001	118149500	1.08	0.2	1.04E-07	1.31	0.28	2.22E-06	−1.23	0.29	2.00E-05	−1.05	0.29	1.69E-04
Chr11	82654251	82654750	−1.05	0.19	1.30E-07	−0.97	0.29	4.68E-04	0.95	0.31	1.53E-03	0.88	0.27	5.94E-04
Chr19	6165251	6165750	0.94	0.18	1.40E-07	0.89	0.25	1.90E-04	−0.88	0.28	9.35E-04	−1.02	0.23	7.36E-06
Chr1	85522251	85522750	0.99	0.18	1.43E-07	1.54	0.25	4.48E-09	−1.53	0.25	7.41E-09	−0.80	0.28	2.32E-03
Chr17	42569001	42569500	−1.07	0.19	1.64E-07	−1.25	0.27	2.21E-06	1.30	0.28	3.81E-06	1.21	0.26	2.44E-06
Chr4	141606501	141607000	1.09	0.19	2.03E-07	1.26	0.26	3.56E-06	−1.33	0.30	1.05E-05	−1.13	0.26	1.83E-05
Chr5	137736001	137736500	−1.03	0.19	2.06E-07	−1.11	0.28	4.12E-05	1.13	0.30	1.39E-04	1.10	0.27	3.78E-05
Chr5	150614501	150615000	−1.08	0.21	2.14E-07	−1.2	0.29	2.80E-05	1.24	0.31	3.18E-05	1.07	0.29	1.21E-04
Chr17	36918251	36918750	−1.1	0.21	2.32E-07	−0.96	0.29	6.01E-04	1.05	0.31	4.88E-04	1.03	0.28	1.83E-04
Chr6	126138251	126138750	−0.99	0.19	2.36E-07	−0.75	0.27	3.29E-03	0.79	0.30	4.92E-03	1.22	0.25	8.94E-07
Chr7	144431251	144431750	1.02	0.19	2.70E-07	0.84	0.28	1.30E-03	−0.85	0.28	1.48E-03	−1.17	0.27	1.77E-05
Chr12	70937251	70937750	0.88	0.17	2.76E-07	0.72	0.23	1.12E-03	−0.87	0.26	5.26E-04	−0.94	0.20	3.10E-06
Chr4	141606251	141606750	1.01	0.19	2.80E-07	1.16	0.26	8.05E-06	−1.31	0.29	6.93E-06	−1.04	0.25	2.14E-05
Chr13	90019001	90019500	1.06	0.2	2.84E-07	1.25	0.29	8.88E-06	−1.36	0.30	5.09E-06	−0.84	0.29	2.20E-03
Chr11	74179001	74179500	1.07	0.2	2.92E-07	1.3	0.29	7.25E-06	−1.37	0.29	3.15E-06	−1.06	0.29	1.76E-04
Chr12	99153001	99153500	−1.01	0.2	3.93E-07	−1.02	0.29	2.02E-04	1.02	0.31	4.59E-04	1.24	0.27	3.77E-06
Chr2	223336751	223337250	−1.01	0.2	4.47E-07	−1.41	0.28	4.00E-07	1.50	0.29	1.69E-07	0.77	0.29	4.65E-03
Chr8	120972001	120972500	0.97	0.19	4.98E-07	0.83	0.27	1.22E-03	−0.87	0.29	1.43E-03	−0.96	0.26	1.81E-04
Chr17	38047001	38047500	−1.04	0.2	5.25E-07	−1.3	0.28	4.51E-06	1.31	0.27	1.21E-06	0.74	0.28	5.39E-03
Chr4	64626751	64627250	−0.99	0.19	5.45E-07	−0.87	0.28	9.92E-04	0.86	0.29	1.61E-03	0.98	0.26	1.36E-04
Chr16	87256751	87257250	−0.83	0.16	5.51E-07	−0.97	0.23	1.75E-05	0.94	0.25	9.13E-05	0.82	0.24	3.37E-04
Chr19	10656751	10657250	−1.03	0.2	5.77E-07	−1.09	0.28	6.47E-05	1.17	0.29	3.80E-05	1.13	0.27	2.18E-05
Chr7	2487251	2487750	0.8	0.16	5.85E-07	0.73	0.23	1.06E-03	−0.81	0.24	5.61E-04	−0.94	0.23	3.49E-05
Chr11	74178751	74179250	0.99	0.2	6.82E-07	1.19	0.28	2.15E-05	−1.15	0.28	2.45E-05	−0.82	0.28	2.37E-03
Chr10	119176501	119177000	−1.04	0.2	6.87E-07	−1.03	0.29	3.41E-04	1.16	0.31	1.69E-04	1.14	0.28	4.83E-05
Chr22	34755251	34755750	−0.98	0.2	7.32E-07	−0.8	0.28	2.35E-03	0.97	0.31	1.21E-03	1.15	0.25	3.95E-06
Chr6	161664751	161665250	0.98	0.19	7.42E-07	0.9	0.26	3.89E-04	−0.83	0.29	2.48E-03	−0.81	0.25	7.71E-04
Chr16	17161751	17162250	−1.01	0.2	7.80E-07	−0.99	0.28	2.19E-04	1.17	0.30	4.48E-05	1.08	0.28	1.16E-04
Chr18	23695001	23695500	1.04	0.21	8.39E-07	0.95	0.29	5.83E-04	−1.04	0.32	6.05E-04	−1.11	0.28	3.96E-05
Chr9	26364751	26365250	−0.95	0.19	8.90E-07	−0.96	0.28	2.78E-04	1.13	0.30	9.90E-05	0.96	0.28	2.77E-04
Chr1	25227001	25227500	−0.83	0.17	9.05E-07	−0.87	0.25	3.81E-04	0.78	0.25	1.39E-03	0.99	0.25	5.78E-05
Chr13	68877251	68877750	0.97	0.2	9.43E-07	1.01	0.28	1.44E-04	−1.24	0.28	1.05E-05	−0.84	0.28	1.39E-03
Chr9	89126501	89127000	−0.92	0.19	9.59E-07	−1.14	0.26	9.49E-06	1.10	0.28	6.62E-05	0.76	0.26	2.12E-03
Chr13	35317501	35318000	−0.96	0.2	9.72E-07	−0.54	0.29	4.44E-02	0.54	0.31	6.52E-02	1.04	0.27	7.61E-05
Chr21	19575001	19575500	0.96	0.19	9.92E-07	0.88	0.27	8.63E-04	−0.96	0.30	8.25E-04	−1.13	0.26	1.24E-05

Table 3 Effect of ICSI on the FDR 25% WBC IVF-DMRs *(Continued)*

Chr2	169470001	169470500	−1.02	0.21	1.06E-06	−1.31	0.28	2.08E-06	1.33	0.29	6.99E-06	1.04	0.27	5.91E-05
Chr12	4310251	4310750	−0.99	0.2	1.06E-06	−0.91	0.29	8.79E-04	1.02	0.30	4.54E-04	0.87	0.30	1.99E-03
Chr6	157136501	157137000	−0.96	0.19	1.15E-06	−0.78	0.27	2.19E-03	0.91	0.30	1.37E-03	1.30	0.25	1.62E-07
Chr14	104067251	104067750	0.71	0.14	1.17E-06	0.81	0.21	7.35E-05	−0.75	0.22	3.89E-04	−0.58	0.24	9.44E-03

selected two of the closest CpG sites contained within the second most associated DMR in that locus (Additional file 1: Figure S3). Within the sample of 36 MZ twins we also observed higher methylation in the IVF group, consistent with the MeDIP-seq signal, with effects close to nominal significance ($P = 0.08$; Additional file 1: Figure S4). However, correlation between the MeDIP-seq signal at the most-associated DMR in *TNP1* and the EpiTYPER methylation values was, as expected, relatively low as we were unable to target CpG sites within this most-associated DMR (correlation of 0.18 and 0 at the two tested CpG sites). We profiled additional samples from DZ twin pairs but did not obtain validation of the signal.

We also considered the effect of ICSI compared to conventional IVF in MZ twins in the validation dataset. We observed significantly higher methylation in the ICSI group at the first CpG of the targeted region near *TNP1* and at the first CpG site of *C9orf3* (Additional file 1: Figure S5).

Lastly, we also compared methylation in relation to conception method at the *H19 CTCF6* DMR in a reduced subset of CBMCs samples (n = 42 twins) using EpiTYPER. When comparing IVF to non-IVF twins (Additional file 1: Figure S6) we observed a difference with the same direction of effect as in the MeDIP-seq analysis, although not significant ($P = 0.19$). Interestingly, when comparing naturally conceived twins to twins that were conceived with any type of medical help (Additional file 1: Figure S6), i.e. not exclusively IVF, the difference reached nominal significance ($P = 0.04$), suggesting that differential methylation at this region is associated with parental subfertility rather than IVF conception.

Discussion

Since IVF procedures are carried out during an important period of epigenetic reprogramming in early development, we hypothesised that IVF may induce epigenetic differences that persist to birth. We were able to identify significant and suggestive DMRs related to IVF conception (IVF-DMRs) in WBCs, although our results suggest that at least some of these changes may be linked to parental subfertility, which is confounded with IVF treatment. The observation that IVF-DMRs were identified close to genes implicated in fertility and reproduction suggests that a genetic signature influencing DNA methylation could be transmitted from parent to offspring. To assess this further, we estimated the heritability of the IVF-DMRs. We observed that the IVF-DMR located in *C9orf3*, a gene associated with polycystic ovary syndrome, was estimated to have a heritability at 25% and eight other FDR 25% WBC IVF-DMRs showed heritability greater than this (Fig. 3).

Epigenetic states of metastable epialleles in mammals are mitotically inherited after establishment in early development, therefore shared across tissues, and can cause expression variability within isogenic individuals [45]. A study in humans looking for systematic inter-individual variation in DNA methylation across tissues from two different lineages identified 109 candidate metastable epialleles [40]. Nutritional conditions during conception have been shown to be important to the establishment of epigenetic states at some of these metastable epialleles [46]. If an influence of IVF on the epigenetic marks of these alleles exists, it could potentially cause long lasting effects. A previous study, which included newborns from single and multiple pregnancies, identified DNA methylation differences in IVF conception at candidate metastable epialleles, although at different epialleles to those affected by maternal nutritional factors [19]. In our study, none of the 109 candidate metastable epialleles overlapped with the 46 FDR 25% WBC IVF-DMRs. This discrepancy could be attributed to differences between single and multiple pregnancies or to low power to detect such changes.

Our results also showed that IVF-DMRs, including hypomethylation of the regulatory region of *H19*, were generally not shared between WBCs and CBMCs. This observation suggests that the epigenetic differences reported here likely did not appear during early development or that these effects are not fixed and can revert in a cell type-specific manner. CBMCs, in contrast to WBCs, lack the granulocyte fraction, which is the predominant group of cells in the blood. Thus, the IVF-DMRs may be granulocyte-specific or at least in part influenced by this group of cells.

To date, there has been mixed evidence on the effect of IVF at imprinted genes and their regulatory regions. Some studies have reported DNA methylation changes or increased variability at these imprinted regions [14–16], while others have reported no associated changes [47, 48]. We observed that there is not an overall destabilisation of

methylation patterns in ICRs, but specific DMRs, such as the *H19* DMR, can show a weak but nominally significant association with the method of conception. Previous studies have reported similar observations, that is, changes in methylation at some imprinted regions, but not in the majority [19, 48]. It is unknown if these changes occur due to IVF since imprinting defects have been previously described in sperm of infertile men, including hypomethylation of the *H19 CTCF6* DMR [49]. Loke et al. [16] reported that hypomethylation at this locus in buccal epithelium of newborns in the IVF group was driven by the subgroup conceived by ICSI. However, it is difficult to dissect whether the observed effect on DNA methylation of ICSI-conceived newborns is due to the technique itself or to male infertility. Whitelaw et al. [50] found higher levels of *SNRPN* methylation in buccal cells of ICSI-conceived newborns and these were associated with longer duration of infertility in the parents. In our data, we observed that the difference at the *H19 CTCF6* DMR was greater when considering any type of medical help during conception, supporting the idea that parental subfertility is the driver of methylation changes at this region. Information about the indication for assisted reproductive technology, the use of donor eggs or sperm, and the fertility status of parents in the control group would be required to further assess the effect of parental subfertility.

Adverse perinatal outcomes and increased frequency of imprinting disorders have also been observed in offspring of couples with a history of subfertility that were able to conceive naturally [51–53]. However, studies that controlled for parental subfertility by comparing siblings in which one was conceived naturally and the other by IVF also observed an effect [54]. It is likely, therefore, that both parental subfertility and IVF may induce epigenetic changes, as observed in another genome-wide study that found DNA methylation differences between IVF-conceived newborns and a group conceived through intrauterine insemination (infertile controls), but also between the latter and naturally conceived newborns (fertile controls) [19]. In addition, a study looking at 37 candidate CpG sites identified seven that were differentially methylated when comparing an IVF-conceived group born to parents without male infertility that used donor oocytes to naturally conceived newborns [55].

Finally, two IVF-DMRs associated with infertility (*TNP1* and *C9orf3*) were targeted for validation. Differential methylation was validated at the *C9orf3* gene. However, validation of the *TNP1* region was hampered by our inability to target CpG sites within the most-associated DMR in this locus. We attempted validation at *TNP1* by targeting CpG sites in the neighbouring 500-bp bin and observed consistent direction of association close to nominal significance.

In this study the non-IVF group included a set of twins conceived after GIFT and another set conceived after ovarian stimulation not followed by IVF. GIFT and ovarian stimulation are fertility treatments not equivalent to IVF since fertilisation still occurs in the fallopian tubes. We showed that our results were not affected by the inclusion of these data, potentially because they were represented in small numbers, only four out of 58 samples.

Limitations

There are several limitations in this study. First, it is known that cell composition may represent a confounding variable in EWAS [56]. Our results use principal component analysis anticipating that these will capture cell heterogeneity, and follow-up of our findings in a subset of twins with available cell counts showed that the majority of findings remained significant after adjustment for cell heterogeneity. Second, although MeDIP-seq has the strength of genome-wide coverage, it lacks base-pair resolution, instead generating methylation scores across genomic regions. However, it has been reported that methylation of neighbouring CpG sites is correlated over distances up to 1000 bp [57], suggesting that the approach may be able to capture a good proportion of the methylation variance in a genomic region. Third, although this study includes a sample size larger than most previous studies exploring IVF, contemporary EWAS study designs generally require larger numbers of cases and controls to achieve sufficient power to detect small to moderate effect sizes [21, 58]. Lastly, our approach cannot conclusively determine the cause of the observed IVF-associated methylation changes. Future studies of IVF-associated regions in animal models, where genetic differences and infertility diseases can be discarded, could help identify if these changes were caused by IVF itself.

Conclusions

We observed evidence for differences in DNA methylation between IVF and non-IVF twins on a genome-wide scale. A strength of this study design is that it allowed us to also estimate the contribution of genetic and environmental factors towards DNA methylation levels at the IVF-associated loci. The inclusion of only twin pregnancies also avoided biases present in studies that consider single and multiple pregnancies together. Multiple pregnancies are more common after IVF. Therefore, the differences observed when studying singleton and twin births together may be confounded with the higher risks of adverse perinatal outcomes in multiple pregnancy births, rather than IVF itself. Nevertheless, we were unable to dissect whether methylation changes were likely caused by IVF, or were due to the underlying parental subfertility, or other factors. These scenarios require further study exploring the stability of these DMRs over time, their relation with gene expression, and their potential role in health and disease.

Additional file

Additional file 1: Supplementary tables and figures. **Table S1** Pregnancy complications and other outcomes. **Table S2** PCR assays. **Table S3** Re-analysis of FDR 25% WBC IVF-DMRs excluding other fertility treatments from non-IVF group. **Table S4** Re-analysis of FDR 25% WBC IVF-DMRs in subset with cell counts. **Figure S1** *C9orf3* FDR 25% IVF-DMR. **Figure S2** *H19 CTCF6* IVF-DMR replication. **Figure S3** CpG sites targeted for validation. **Figure S4** IVF versus non-IVF DNA methylation differences using Sequenom's EpiTYPER technology. **Figure S5** ICSI versus conventional IVF DNA methylation differences using Sequenom's EpiTYPER technology. **Figure S6** *H19 CTCF6* IVF-DMR replication using Sequenom's EpiTYPER technology. (DOCX 104 kb)

Abbreviations

BMI: Body mass index; CBMC: Cord blood mononuclear cell; DMR: Differentially methylated region; DZ: Dizygotic; EWAS: Epigenome-wide association scans; FDR: False discovery rate; GIFT: Gamete intra-fallopian transfer; ICR: Imprinting control region; ICSI: Intracytoplasmic sperm injection; IVF: *In vitro* fertilisation; MeDIP-seq: Methylated DNA immunoprecipitation coupled with sequencing; MZ: Monozygotic; PETS: Peri/postnatal Epigenetic Twins Study; SD: Standard deviation; WBC: Whole blood cell

Acknowledgements

We wish to thank Ruth Morley, John Carlin, Mark Umstad, Euan Wallace, Mark Permezel, Sarah Healy, Tina Vaiano, Nicole Brooks, Jennifer Foord, Sheila Holland, Anne Krastev, Siva Illancheran, Joanne Mockler, Anna Czajko, and Geraldine McIlroy for contributions to PETS, and all mothers and twins that participated in this study.

Funding

PETS was supported by grants from the Australian National Health and Medical Research Council (grant numbers 437015 and 607358 to JC and RS), the Financial Markets Foundation for Children (grant number 032-2007) and by the Victorian Government's Operational Infrastructure Support Program. RS is supported by a NHMRC Senior Research Fellowship. JCF was funded by EU-FP7 project EpiTrain (316758). TDS is an NIHR Senior Investigator and a European Research Council Senior Research Investigator (ERC 250157). The study also received support from the TwinsUK resource, funded by the Wellcome Trust, with support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The study was also supported by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Economic and Social Research Council (ESRC) (grant number ES/N000404/1 to JTB).

Availability of data and material

MeDIP-seq data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001002248.

Authors' contributions

YJL, JC, and RS provided the PETS samples. FG, YX, HW, HL, YL, and JW were responsible for the sequencing and quality control of the MeDIP-seq. JCF and JTB conducted the EWAS analyses. SBS, YJL, and JCF performed the validation analysis. JCF and JTB prepared the manuscript with contributions from TDS, JC, RS, and YJL. JTB, TDS, JC, and RS supervised this work.
All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Ethics approval and consent to participate

This study complies with the Declaration of Helsinki. Written informed consent was obtained from all participants under approvals from the Royal Women's Hospital Human Research Ethics Committee (project number 06/21), Mercy Hospital for Women Human Research Ethics Committee (project number R06/30), and Monash Health Human Research Ethics Committee (project number 06117C), Melbourne, Australia.

Author details

[1]Department of Twin Research and Genetic Epidemiology, King's College London, SE1 7EH London, UK. [2]Early Life Epigenetics, Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, VIC, Australia. [3]BGI-Shenzhen, Shenzhen, China. [4]King Abdulaziz University, Jeddah 22254, Saudi Arabia. [5]Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. [6]The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen DK-2200, Denmark. [7]Department of Paediatrics, University of Melbourne, Parkville, VIC, Australia. [8]Cancer, Disease and Developmental Epigenetics, Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, VIC, Australia.

Received: 13 July 2016 Accepted: 4 February 2017
Published online: 24 March 2017

References

1. Pandey S, Shetty A, Hamilton M, Bhattacharya S, Maheshwari A. Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. Hum Reprod Update. 2012;18:485–503.
2. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjaerven R, Vatten LJ. Increased risk of placenta previa in pregnancies following IVF/ICSI; a comparison of ART and non-ART pregnancies in the same mother. Hum Reprod. 2006;21:2353–8.
3. Cox GF, Bürger J, Lip V, Mau UA, Sperling K, Wu B-L, Horsthemke B. Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet. 2002;71:162–4.
4. Maher ER, Brueton LA, Bowdin SC, Luharia A, Cooper W, Cole TR, Macdonald F, Sampson JR, Barratt CL, Reik W, Hawkins MM. Beckwith-Wiedemann syndrome and assisted reproduction technology (ART). J Med Genet. 2003;40:62–4.
5. Apantaku O, Chandrasekaran I, Bentick B. Obstetric outcome of singleton pregnancies achieved with *in vitro* fertilisation and intracytoplasmic sperm injection: experience from a district general hospital. J Obstet Gynaecol. 2008;28:398–402.
6. De Neubourg D, Gerris J, Mangelschots K, Van Royen E, Vercruyssen M, Steylemans A, Elseviers M. The obstetrical and neonatal outcome of babies born after single-embryo transfer in IVF/ICSI compares favourably to spontaneously conceived babies. Hum Reprod. 2006;21:1041–6.
7. Gordon L, Joo JE, Powell JE, Ollikainen M, Novakovic B, Li X, Andronikos R, Cruickshank MN, Conneely KN, Smith AK, Alisch RS, Morley R, Visscher PM, Craig JM, Saffery R. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome Res. 2012;22:1395–406.
8. Weksberg R. Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. Hum Mol Genet. 2002;11:1317–25.
9. Saffery R. Epigenetic change as the major mediator of fetal programming in humans: are we there yet? Ann Nutr Metab. 2014;64:203–7.
10. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A. DNA methylation dynamics of the human preimplantation embryo. Nature. 2014;511:611–5.
11. Fauque P. Ovulation induction and epigenetic anomalies. Fertil Steril. 2013;99:616–23.
12. Doherty AS, Mann MR, Tremblay KD, Bartolomei MS, Schultz RM. Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod. 2000;62:1526–35.
13. Wang Z, Xu L, He F. Embryo vitrification affects the methylation of the H19/Igf2 differentially methylated domain and the expression of H19 and Igf2. Fertil Steril. 2010;93:2729–33.
14. Li L, Wang L, Le F, Liu X, Yu P, Sheng J, Huang H, Jin F. Evaluation of DNA methylation status at differentially methylated regions in IVF-conceived newborn twins. Fertil Steril. 2011;95:1975–9.
15. Nelissen ECM, Dumoulin JCM, Daunay A, Evers JLH, Tost J, van Montfoort APA. Placentas from pregnancies conceived by IVF/ICSI have a reduced DNA methylation level at the H19 and MEST differentially methylated regions. Hum Reprod. 2013;28:1117–26.
16. Loke YJ, Galati JC, Saffery R, Craig JM. Association of *in vitro* fertilization with global and IGF2/H19 methylation variation in newborn twins. J Dev Orig Health Dis. 2015;6:115–24.

17. Katari S, Turan N, Bibikova M, Erinle O, Chalian R, Foster M, Gaughan JP, Coutifaris C, Sapienza C. DNA methylation and gene expression differences in children conceived *in vitro* or in vivo. Hum Mol Genet. 2009;18:3769–78.
18. Melamed N, Choufani S, Wilkins-Haug LE, Koren G, Weksberg R. Comparison of genome-wide and gene-specific DNA methylation between ART and naturally conceived pregnancies. Epigenetics. 2015;10:474–83.
19. Estill MS, Bolnick JM, Waterland RA, Bolnick AD, Diamond MP, Krawetz SA. Assisted reproductive technology alters deoxyribonucleic acid methylation profiles in bloodspots of newborn infants. Fertil Steril. 2016;106:629–39.e10.
20. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan J-B, Shen R. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98:288–95.
21. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41.
22. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37:853–62.
23. Saffery R, Morley R, Carlin JB, Joo J-HE, Ollikainen M, Novakovic B, Andronikos R, Li X, Loke YJ, Carson N, Wallace EM, Umstad MP, Permezel M, Galati JC, Craig JM. Cohort profile: The peri/post-natal epigenetic twins study. Int J Epidemiol. 2012;41:55–61.
24. Loke YJ, Novakovic B, Ollikainen M, Wallace EM, Umstad MP, Permezel M, Morley R, Ponsonby A-L, Gordon L, Galati JC, Saffery R, Craig JM. The Peri/postnatal Epigenetic Twins Study (PETS). Twin Res Hum Genet. 2013;16:13–20.
25. Ollikainen M, Smith KR, Joo EJ-H, Ng HK, Andronikos R, Novakovic B, Abdul Aziz NK, Carlin JB, Morley R, Saffery R, Craig JM. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet. 2010;19:4176–88.
26. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, Herwig R, Adjaye J. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res. 2010;20:1441–50.
27. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67:1–48.
28. R Core Team: R: a Language and Environment for Statistical Computing. 2014. http://www.R-project.org/.
29. Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. Brief Bioinform. 2002;3:119–33.
30. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J. OpenMx: an open source extended structural equation modeling framework. Psychometrika. 2011;76:306–17.
31. Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AHC, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG, Petronis A. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009;41:240–5.
32. Miyagawa Y, Nishimura H, Tsujimura A, Matsuoka Y, Matsumiya K, Okuyama A, Nishimune Y, Tanaka H. Single-nucleotide polymorphisms and mutation analyses of the TNP1 and TNP2 genes of fertile and infertile human male populations. J Androl. 2005;26:779–86.
33. Trasler JM, Hake LE, Johnson PA, Alcivar AA, Millette CF, Hecht NB. DNA methylation and demethylation events during meiotic prophase in the mouse testis. Mol Cell Biol. 1990;10:1828–34.
34. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, Zhang B, Liang X, Li T, Chen J, Shen J, Zhao J, You L, Gao X, Zhu D, Zhao X, Yan Y, Qin Y, Li W, Yan J, Wang Q, Zhao J, Geng L, Ma J, Zhao Y, He G, Zhang A, Zou S, Yang A, Liu J, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet. 2012;44:1020–5.
35. Kerns SL, Ostrer H, Stock R, Li W, Moore J, Pearlman A, Campbell C, Shao Y, Stone N, Kusnetz L, Rosenstein BS. Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys. 2010;78:1292–300.
36. Fenstad MH, Johnson MP, Løset M, Mundal SB, Roten LT, Eide IP, Bjørge L, Sande RK, Johansson AK, Dyer TD, Forsmo S, Blangero J, Moses EK, Austgulen R. STOX2 but not STOX1 is differentially expressed in decidua from pre-eclamptic women: data from the Second Nord-Trondelag Health Study. Mol Hum Reprod. 2010;16:960–8.
37. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500:477–81.
38. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee B-K, Lee K, London D, Lotakis D, Neph S, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489:75–82.
39. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol. 1987;196:261–82.
40. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, Coarfa C, Hernandez-Vargas H, Castelino JM, Routledge MN, Gong YY, Herceg Z, Lee YS, Lee K, Moore SE, Fulford AJ, Prentice AM, Waterland RA. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol. 2015;16:118.
41. Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics. 2008;3:89–96.
42. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet. 1992;340:17–8.
43. Palermo GD, Neri QV, Rosenwaks Z. Safety of intracytoplasmic sperm injection. Methods Mol Biol. 2014;1154:549–62.
44. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.
45. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet. 2002;18:348–51.
46. Dominguez-Salas P, Moore SE, Baker MS, Bergen AW, Cox SE, Dyer RA, Fulford AJ, Guan Y, Laritsky E, Silver MJ, Swan GE, Zeisel SH, Innis SM, Waterland RA, Prentice AM, Hennig BJ. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. Nat Commun. 2014;5:3746.
47. Oliver VF, Miles HL, Cutfield WS, Hofman PL, Ludgate JL, Morison IM. Defects in imprinting and genome-wide DNA methylation are not common in the *in vitro* fertilization population. Fertil Steril. 2012;97:147–53. e7.
48. Tierling S, Souren NY, Gries J, Loporto C, Groth M, Lutsik P, Neitzel H, Utz-Billing I, Gillessen-Kaesbach G, Kentenich H, Griesinger G, Sperling K, Schwinger E, Walter J. Assisted reproductive technologies do not enhance the variability of DNA methylation imprints in human. J Med Genet. 2010;47:371–6.
49. Boissonnas CC, El Abdalaoui H, Haelewyn V, Fauque P, Dupont JM, Gut I, Vaiman D, Jouannet P, Tost J, Jammes H. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. Eur J Hum Genet. 2010;18:73–80.
50. Whitelaw N, Bhattacharya S, Hoad G, Horgan GW, Hamilton M, Haggarty P. Epigenetic status in the offspring of spontaneous and assisted conception. Hum Reprod. 2014;29:1452–8.
51. Ludwig M, Katalinic A, Gross S, Sutcliffe A, Varon R, Horsthemke B. Increased prevalence of imprinting defects in patients with Angelman syndrome born to subfertile couples. J Med Genet. 2005;42:289–91.
52. Doornbos ME, Maas SM, McDonnell J, Vermeiden JPW, Hennekam RCM. Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. Hum Reprod. 2007;22:2476–80.
53. Jaques AM, Amor DJ, Baker HWG, Healy DL, Ukoumunne OC, Breheny S, Garrett C, Halliday JL. Adverse obstetric and perinatal outcomes in subfertile women conceiving without assisted reproductive technologies. Fertil Steril. 2010;94:2674–9.
54. Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Söderström-Anttila V, Nygren KG, Hazekamp J, Bergh C. Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis. Hum Reprod Update. 2013;19:87–104.
55. Song S, Ghosh J, Mainigi M, Turan N, Weinerman R, Truongcao M, Coutifaris C, Sapienza C. DNA methylation differences between *in vitro*- and in vivo-conceived children are associated with ART procedures rather than infertility. Clin Epigenetics. 2015;7:41.

56. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15:R31.
57. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38:1378–85.
58. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015;44:1429.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit


BioMed Central