

RESEARCH

Open Access



# The X-factor in ART: does the use of assisted reproductive technologies influence DNA methylation on the X chromosome?

Julia Romanowska<sup>1,2\*</sup>, Haakon E. Nustad<sup>1,3</sup>, Christian M. Page<sup>1,4</sup>, William R. P. Denault<sup>1,5</sup>, Yunsung Lee<sup>1</sup>, Maria C. Magnus<sup>1</sup>, Kristine L. Haftorn<sup>1</sup>, Miriam Gjerdevik<sup>1,6</sup>, Boris Novakovic<sup>7,8</sup>, Richard Saffery<sup>7,8</sup>, Håkon K. Gjessing<sup>1,2</sup>, Robert Lyle<sup>1,9</sup>, Per Magnus<sup>1</sup>, Siri E. Håberg<sup>1</sup> and Astanand Jugessur<sup>1,2</sup>

## Abstract

**Background** Assisted reproductive technologies (ART) may perturb DNA methylation (DNAm) in early embryonic development. Although a handful of epigenome-wide association studies of ART have been published, none have investigated CpGs on the X chromosome. To bridge this knowledge gap, we leveraged one of the largest collections of mother–father–newborn trios of ART and non-ART (natural) conceptions to date to investigate sex-specific DNAm differences on the X chromosome. The discovery cohort consisted of 982 ART and 963 non-ART trios from the Norwegian Mother, Father, and Child Cohort Study (MoBa). To verify our results from the MoBa cohort, we used an external cohort of 149 ART and 58 non-ART neonates from the Australian ‘Clinical review of the Health of adults conceived following Assisted Reproductive Technologies’ (CHART) study. The Illumina EPIC array was used to measure DNAm in both datasets. In the MoBa cohort, we performed a set of X-chromosome-wide association studies (‘XWASs’ hereafter) to search for sex-specific DNAm differences between ART and non-ART newborns. We tested several models to investigate the influence of various confounders, including parental DNAm. We also searched for differentially methylated regions (DMRs) and regions of co-methylation flanking the most significant CpGs. Additionally, we ran an analogous model to our main model on the external CHART dataset.

**Results** In the MoBa cohort, we found more differentially methylated CpGs and DMRs in girls than boys. Most of the associations persisted after controlling for parental DNAm and other confounders. Many of the significant CpGs and DMRs were in gene-promoter regions, and several of the genes linked to these CpGs are expressed in tissues relevant for both ART and sex (testis, placenta, and fallopian tube). We found no support for parental DNAm-dependent features as an explanation for the observed associations in the newborns. The most significant CpG in the boys-only analysis was in *UBE2DNL*, which is expressed in testes but with unknown function. The most significant CpGs in the girls-only analysis were in *EIF2S3* and *AMOT*. These three loci also displayed differential DNAm in the CHART cohort.

**Conclusions** Genes that co-localized with the significant CpGs and DMRs associated with ART are implicated in several key biological processes (e.g., neurodevelopment) and disorders (e.g., intellectual disability and autism). These connections are particularly compelling in light of previous findings indicating that neurodevelopmental outcomes differ in ART-conceived children compared to those naturally conceived.

\*Correspondence:

Julia Romanowska

Julia.Romanowska@uib.no

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. 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 in a credit line to the data.

**Keywords** Epigenetics, Assisted reproductive technology (ART), The Norwegian Mother, Father and Child Cohort Study (MoBa), DNA methylation, Epigenome-wide association study, X chromosome, X-chromosome-wide association study (XWAS), Trio study design, Illumina EPIC array, X chromosome inactivation (XCI)

## Background

The use of assisted reproductive technologies (ART) has been on the rise in most parts of the world since the first baby was born to *in vitro* fertilization (IVF) in 1978 [1, 2]. The trend of declining fecundity and greater reliance on ART to conceive is expected to persist in the future, as egg-freezing gains more acceptance in contemporary societies and more couples choose to postpone child-bearing [3–5]. As the clinical and laboratory procedures for ART coincide with the developmental window in which the early embryo undergoes extensive epigenetic remodeling [6–8], it is critical to determine whether the ART procedures themselves or some underlying mechanisms related to parental characteristics (e.g., parental infertility) are responsible for the observed epigenetic differences between ART and non-ART newborns. A number of epigenome-wide association studies (EWASs) of ART have been published in recent years [9–17] and have already contributed substantially to our current understanding of epigenetic changes associated with ART. However, none of these studies have investigated the effect of epigenetic markers on the X chromosome.

Until recently, most genome-wide association studies (GWASs) were also performed almost exclusively on autosomes, leaving out single-nucleotide polymorphisms (SNPs) on the X chromosome, even though this chromosome constitutes ~5% of the human genome and houses ~1000 genes, several of which have been associated with complex traits [18, 19]. The main reason for this exclusion is that the initial methods for GWAS were primarily designed for autosomal markers, as analyzing different X chromosome contents in males and females comes with its own set of analytic challenges [20]. To fill this knowledge gap, we and others have developed a suite of biostatistical tools for analyzing X-linked SNPs both individually and as haplotypes [21–31]. Currently, there is a similar trend of systematic exclusion of CpGs on the sex chromosomes in the vast majority of EWASs, which may result in overlooking important associations.

There are several reasons why X chromosome markers are less tractable to analyze than autosomal markers. First, one needs to account for X chromosome inactivation (XCI) in which one of the X chromosomes in female somatic cells is randomly selected and transcriptionally inactivated in early embryonic development [32, 33]. This crucial mechanism ensures a balanced dosage of X-linked genes in males

and females [34–36]. However, XCI is not complete in humans, with approximately 12% of the genes reported to escape XCI and a further 15% differing in their XCI status across individuals, tissues, and cells [33, 37–39]. Second, the analysis of X-linked markers is complicated by genes in the pseudoautosomal regions (PARs), which are expressed in a similar fashion to autosomal genes as a consequence of escaping XCI [40, 41]. Third, the gradual loss of the X chromosome with age [42] may further complicate the analysis of X-linked markers when comparing cohorts that differ significantly with age.

Despite these challenges, taking X chromosome markers into account in a GWAS or EWAS is important based on the following observations: (a) genes on the X chromosome are known to play essential roles in transcriptional regulation of autosomal genes [43, 44], (b) several traits show a consistently higher prevalence in one sex, and (c) there are distinct physical differences between the sexes (sexual dimorphism) [45]. All of these features might stem from sex-specific differences which may be especially relevant for differences occurring *prior* to gonadal differentiation, i.e., differences that are solely attributable to sex chromosome content rather than those induced by gonadal and hormonal changes [34, 35, 46]. Although a wide variety of traits are known to exhibit sex-specific DNA methylation (DNAm) signatures on the autosomes [47–57], less is known about the presence of such signatures on the X chromosome, possibly due to the overall lack of focus on X-linked markers and the dearth of X-chromosome-wide association studies (XWASs) conducted to date. The few XWASs published thus far include an investigation of CpGs influenced by cigarette smoking, an exploration of differential chronological aging in males versus females, and a study of DNAm changes associated with aging on the X and Y chromosomes [58–61].

Given these important knowledge gaps, our main objective was to examine sex-specific differences in DNAm profiles on the X chromosome by contrasting ART and naturally conceived newborns. We used one of the largest case–control collection of mother–father–newborn trios of ART and non-ART conceptions to date [62], stemming from the Norwegian Mother, Father, and Child Cohort Study (MoBa) [63]. The analyses were stratified by sex and adjusted for potential confounding factors (mother’s age, smoking

status, BMI, and primiparity, as well as parental DNAm at each CpG). As an external check of our results from the MoBa cohort, we analyzed data from the Australian ‘Clinical review of the Health of 22–33 years old conceived with and without ART’ (CHART) cohort [64, 65], which was the only available external ART cohort to which we could compare our MoBa results.

### Results

In the discovery cohort (MoBa), we analyzed DNAm data from 982 ART and 963 non-ART mother–father–newborn trios. These data were generated on the Illumina EPIC platform using DNA extracted from peripheral blood in adults and cord blood in newborns (for details, see Methods and ref. [62]). Our main aim was to identify differences in DNAm in ART versus non-ART newborns, both at a single-CpG level across the entire X chromosome (XWAS) and at a regional level where we searched for differentially methylated regions (DMRs). All the analyses were performed separately for boys and girls.

In the main model, we adjusted for known confounders (mother’s age, smoking status, BMI, and primiparity). In addition, we tested three adjusted models, where we included: (i) parental DNAm at each CpG, (ii) birthweight and gestational age of the newborn, and (iii) all the covariates from (i) and (ii). All the analyses were stratified by

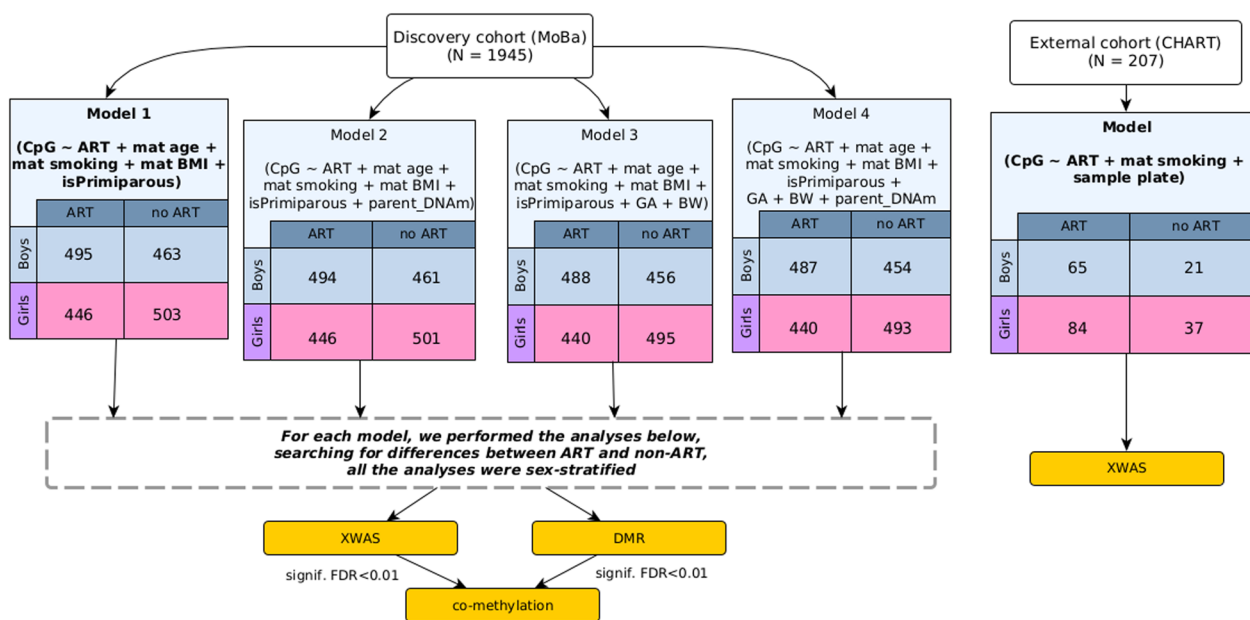
sex. We also explored co-methylation patterns between significant CpGs identified by the above analyses as well as other CpGs in the immediate flanking regions.

In the external cohort (CHART), we analyzed DNAm data, also generated on the EPIC array, from 149 ART-conceived and 58 non-ART newborns. These analyses are outlined in Fig. 1 and detailed in the Methods section.

### Differences between ART and non-ART newborns and their parents in the MoBa cohort

The ART parents were older than the non-ART parents, and the ART newborns weighed less than the non-ART newborns (Table 1). Fewer of the ART mothers smoked during pregnancy than the non-ART mothers, but, intriguingly, a higher proportion of the ART mothers were *past* smokers.

Figure 2 highlights the general trends in the XWAS results for boys and girls separately, before and after controlling for inflation using the R package BACON (see Methods for details). Effect sizes for the CpGs were dominated by global hypomethylation (overall lower DNAm level in the ART newborns). Figure 2 also illustrates the efficacy of BACON in reducing inflation in the *p* values. A similar figure showing the results of all the models tested can be found in Additional file 1: Figure S3.



**Fig. 1** Overview of the analytic pipeline. We refer to **Model 1** as the ‘main model’ throughout this paper and focus primarily on significant findings from this model. The other models are referred to as ‘adjusted models’ and treated as sensitivity analyses. Abbreviations used in the figure: CpG = cytosine-phosphate-guanine; ART = assisted reproductive technologies; mat = maternal; BMI = body mass index; BW = birthweight; GA = gestational age; XWAS = X-chromosome-wide association study; DMR = differentially methylated region; FDR = false discovery rate; MoBa = The Mother, Father, and Child Cohort Study; and CHART = The ‘Clinical review of the Health of adults conceived following Assisted Reproductive Technologies’ study

**Table 1** Characteristics of the discovery cohort (MoBa)

Characteristic	Non-ART (N = 983) <sup>1</sup>	ART (N = 962) <sup>1</sup>	p value <sup>2</sup>
Male (%)	470 (48%)	505 (52%)	0.039
Maternal age (years)	30 (27–33)	33 (31–36)	<0.001
Paternal age (years)	32 (29–36)	35 (32–38)	<0.001
Gestational age (weeks)	40 (39–41)	40 (39–41)	0.090
(No data)	4	0	
Birth weight (g)	3650 (3330–3970)	3540 (3190–3850)	<0.001
(No data)	0	1	
First child	461 (47%)	673 (70%)	<0.001
Maternal BMI (kg/m <sup>2</sup> )	23 (21–26)	23 (21–26)	0.4
(No data)	14	17	
Maternal smoking			<0.001
Never	490 (50%)	494 (52%)	
Past smoker	253 (26%)	358 (37%)	
1st trimester	132 (13%)	62 (6.5%)	
1st trimester and after	104 (11%)	44 (4.6%)	
(No data)	4	4	

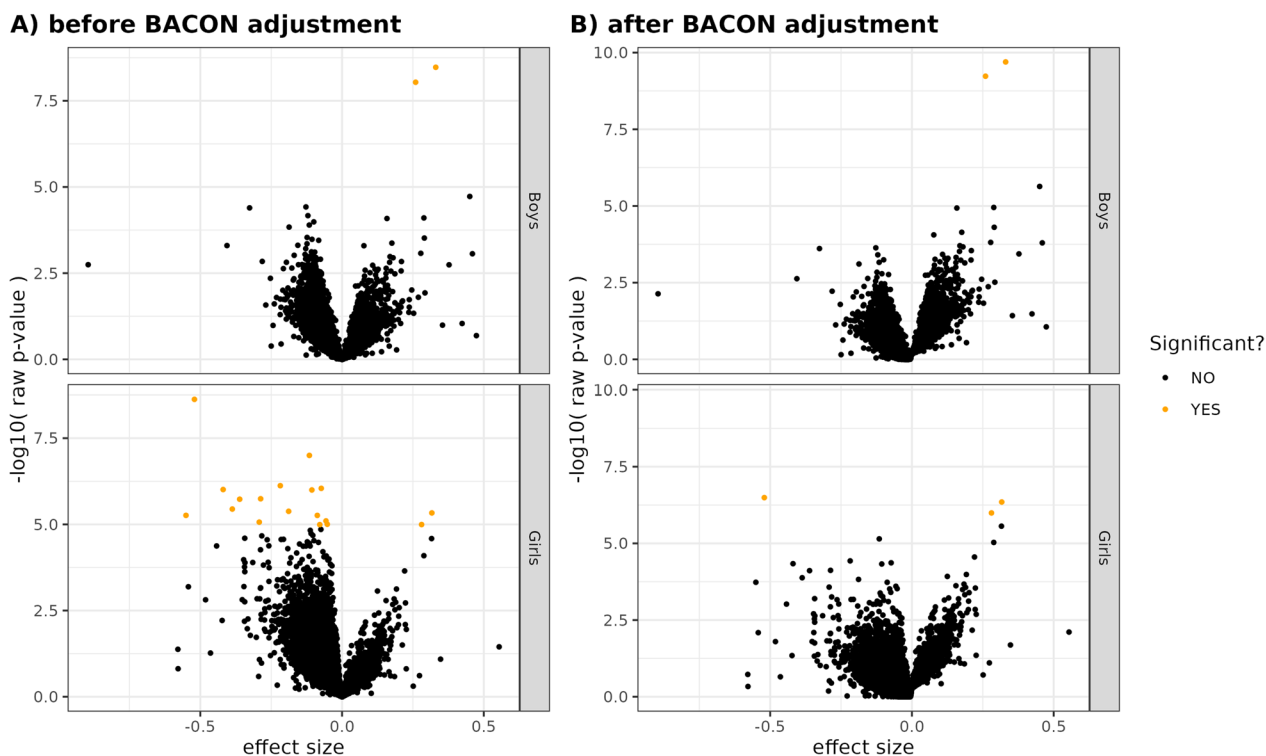
<sup>1</sup> n (%); median (IQR)

<sup>2</sup> Pearson’s chi-squared test; Wilcoxon rank-sum test

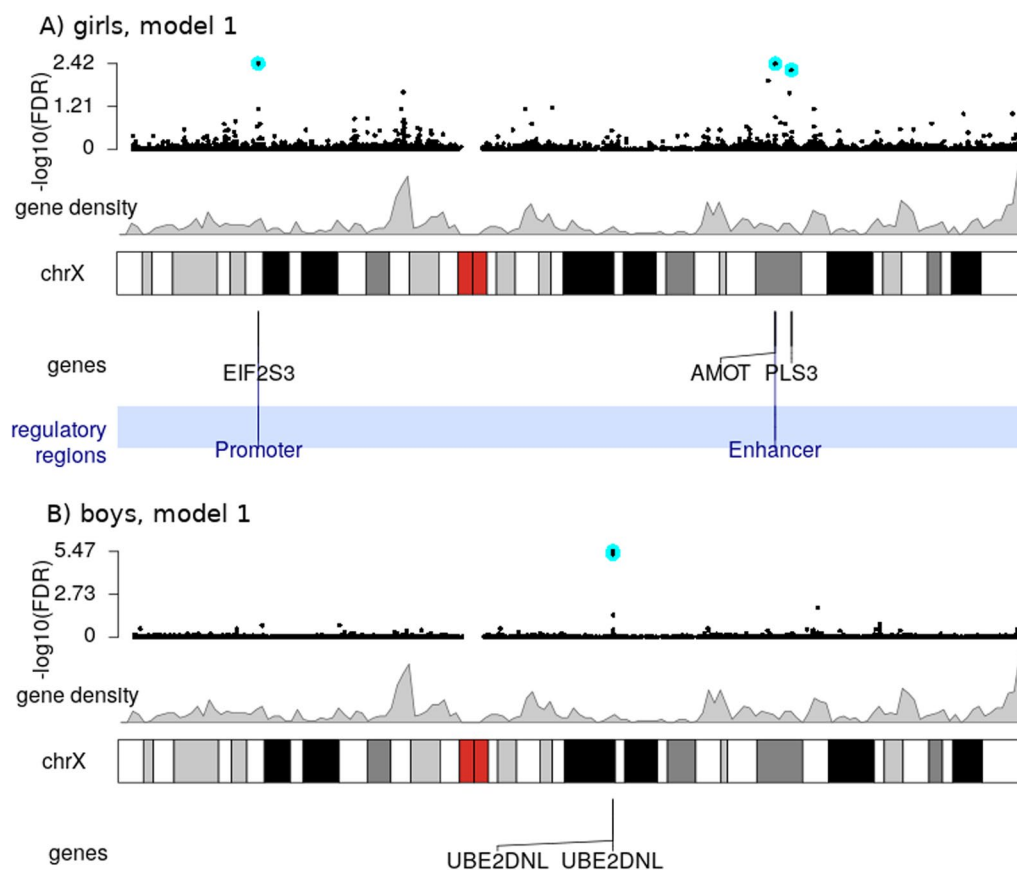
**Several CpGs were significantly differentially methylated between ART and non-ART in the MoBa cohort**

We identified three significantly differentially methylated CpGs in the girls-only analysis (cg25034591, cg13866977, and cg26175661) and two CpGs in the boys-only analysis (cg00920314, cg04516011), all at a false discovery rate-adjusted (FDR-adjusted) p value < 0.01. Strikingly, there was no overlap in the location of the significant findings between the girls-only and boys-only analyses (see Fig. 3 for the location of findings from the main model and Additional file 2 for the significant results for all four statistical models). Additionally, tables with all the results are available in GitHub, at <https://github.com/folkehelseinstituttet/X-factor-ART>.

Adjusting for parental DNAm in the MoBa sample enabled ruling out parental characteristics as the reason for the observed DNAm differences. When the results of the main model (Model 1 of Fig. 1) were contrasted with those of the adjusted models (Models 2–4 of Fig. 1), there was no significant change in the findings in the boys-only analyses (Additional file 2). By contrast, only cg25034591 and cg13866977 remained significant across all models in the girls-only analyses. These results suggest that the differential methylation at these sites is more likely the



**Fig. 2** Effect sizes versus  $-\log_{10} p$  values for each of the X-linked CpGs included in the analysis of the MoBa cohort. Significant findings at FDR < 0.01 are highlighted in orange. The data are presented before (A) and after (B) p value adjustment with the BACON algorithm. This figure shows the results of the analysis when applying the main model, while all the models are included in Additional file 1: Figure S3.



**Fig. 3** Results of the XWAS of girls (**A**) and boys (**B**) based on the main model (**Model 1**;  $\text{CpG} \sim \text{ART} + \text{maternal age} + \text{maternal smoking} + \text{maternal BMI} + \text{primiparity}$ ). The top plot in each panel is a Manhattan plot of all the tested CpGs. The genomic locations of the most significant findings ( $\text{FDR} < 0.01$ ) are highlighted by cyan-colored circles. Immediately beneath is a line plot of gene density, the chromosomal bands, and any genes and/or regulatory regions that overlap with the significant findings

result of the ART procedures themselves rather than parental DNAm.

We performed bootstrap analyses to evaluate the consistency with which the significant CpGs were retained (see Methods for details). The two significant CpGs in the boys-only analyses (cg00920314 and cg04516011) showed a high degree of consistency. They were significant 54% and 47% of the time, respectively, which is substantially higher than the next CpG on the ranked list (cg00243584 at 9%). In the girls-only analyses, cg25034591 was significant in 51% of the bootstrap samples, but the other two CpGs were not as consistent (cg13866977 at 25% and cg26175661 at 19%, occupying positions six and 14 on the ranked list, respectively). The full list of CpGs found to be significant at least once, and the proportion of times a given CpG was found to be significant, are provided in Additional file 4 for the boys-only analyses and in Additional file 5 for the girls-only analyses. These tables are also provided in the GitHub repository.

The two significant CpGs detected in the boys-only analysis are adjacent and located within the gene ‘Ubiquitin conjugating enzyme E2 D N-terminal like’ (*UBE2DNL*) (Fig. 3). In contrast, the significant CpGs in the girls-only analysis are located in different chromosomal regions, i.e., within ‘Eukaryotic translation initiation factor 2 subunit gamma’ (*EIF2S3*), ‘Angiomotin’ (*AMOT*), and ‘Plastin 3’ (*PLS3*) (Fig. 3). Two of the CpGs are located within promoter regions and one within an enhancer. See Table 2 for a summary of the genes.

#### Patterns of co-methylation around the significant CpGs in the MoBa cohort

Analyzing clusters of DNAm can be more informative than scrutinizing one CpG at a time, as it may, for example, help in identifying co-methylation patterns and regions that may be important from a population-epigenetic perspective [66]. Accordingly, we examined regions of 50 kb around each significant CpG detected in our

**Table 2** Summary of the genes and loci identified in the current XWAS

Sex	Gene name (location; ensembl ID)	Full gene name (MIM entry)	Gene function <sup>1</sup>	Refs
Girls	<i>EIF2S3</i> (Xp22.11; ENSG00000130741)	Eukaryotic translation initiation factor 2 subunit gamma (MIM:300161)	This gene encodes the core subunit of eukaryotic translation initiation factor 2 (eIF2). This gamma subunit is the largest component of a heterotrimeric GTP-binding protein which is essential for protein synthesis. Hemizygous mutations in <i>EIF2S3</i> cause an X-linked syndrome called 'mental retardation, epileptic seizures, hypogonadism and -genitalism, microcephaly, and obesity' (MEHMO). <i>EIF2S3</i> has also been reported to escape XCI.	[33, 76, 77, 143, 144]
Girls	<i>AMOT</i> (Xq23; ENSG00000126016)	Angiomotin (MIM:300410)	This gene belongs to the motin family of angiostatin-binding proteins containing conserved coiled-coil domains and C-terminal PDZ binding motifs. <i>AMOT</i> is predominantly expressed in endothelial cells of capillaries and in larger vessels of the placenta where it may mediate the inhibitory effect of angiostatin on tube formation and the migration of endothelial cells toward growth factors during the formation of new blood vessels.	[145–147]
Girls	<i>PLS3</i> (Xq23; ENSG00000102024)	Plastin 3 (MIM:300131)	Plastins comprise a family of actin-binding proteins that are conserved throughout eukaryote evolution. They are expressed in most tissues of higher eukaryotes. Two ubiquitous plastin isoforms (L and T) have been identified in humans.	
Boys	<i>UBE2DNL</i> (Xq21.1; ENSG00000229547)	Ubiquitin conjugating enzyme E2 N-terminal like (pseudogene)	<i>UBE2DNL</i> is labeled 'pseudogene' in various gene databases, but it is reported to be expressed in testis.	Not applicable

<sup>1</sup> Information on gene function was collated from various sources, including NCBI's Entrez Gene (<https://www.ncbi.nlm.nih.gov/gene>), Gene Cards (<https://www.genecards.org/>), the Online Inheritance in Man (OMIM) (<https://omim.org/>), and the cited references in the last column

XWAS (Figures 4–6 show important examples). This led to the identification of clusters of positively correlated CpGs often mapping to a promoter region (see Fig. 4). We also observed clusters of CpGs within gene body regions, such as cg13866977 which was positively correlated with 16 other CpGs across the *AMOT* region (Fig. 5). Overall, the direction of correlation patterns in the promoter region and gene body was opposite to each other, which is as expected and consistent with gene expression patterns that are typical for these regions [67, 68]. To illustrate, a cluster of three CpGs in the promoter region near *EIF2S3*, with positive correlation between the CpGs, was also positively correlated with cg25034591 (Fig. 4). In contrast, a different cluster of four CpGs within the *EIF2S3* gene body, which were positively correlated with one another, was *negatively* correlated with cg25034591. The co-methylation analysis of the significant findings in the boys-only XWAS (Fig. 6) indicated that both significant CpGs are located within a cluster of correlated CpGs and are also part of a DMR located at chr X:84,189,179–84,189,658 (GRCh37) that harbors four CpGs (see the section below and Fig. 7).

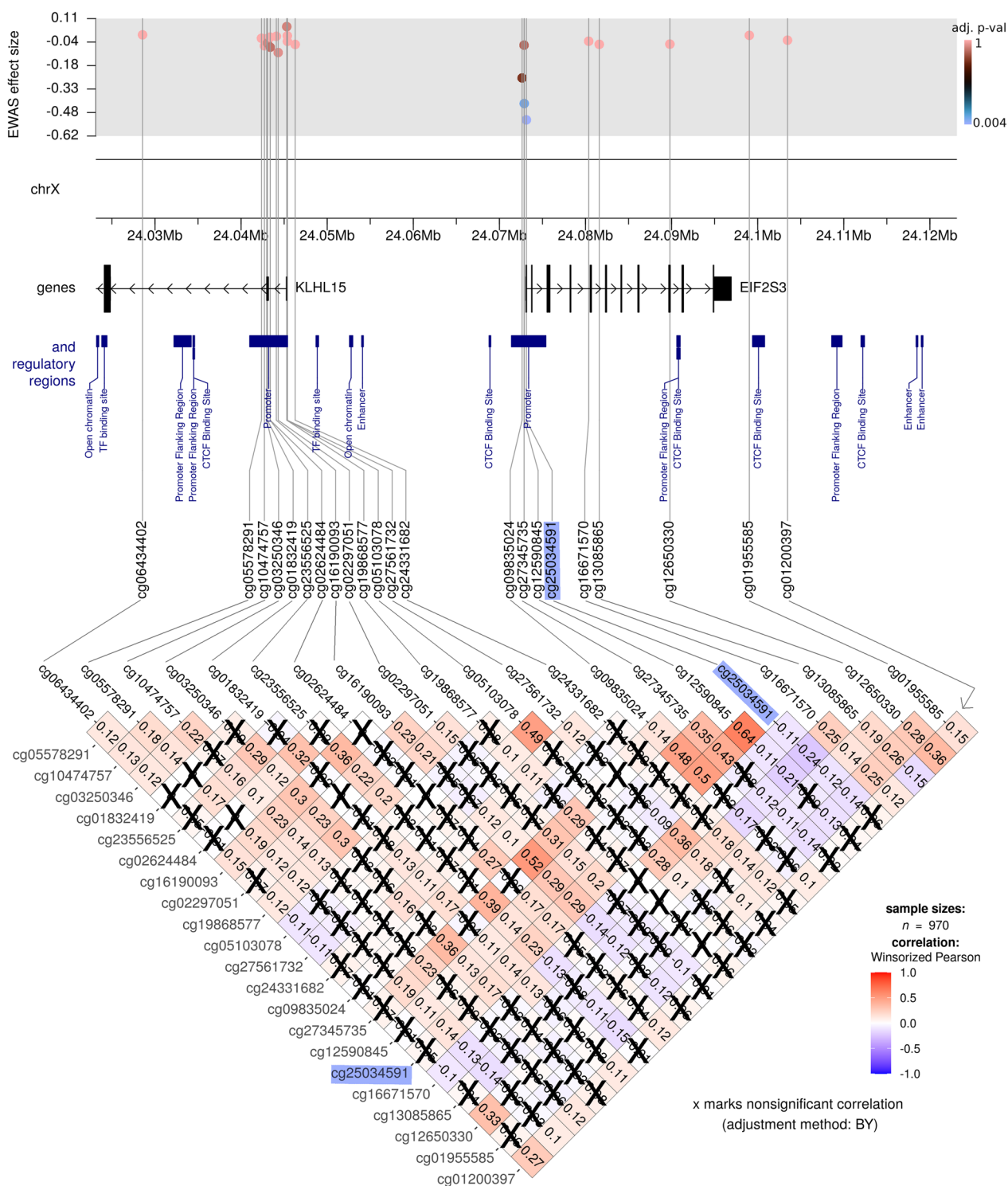
#### DMR analysis in ART and non-ART newborns in the MoBa cohort

We identified 12 significant DMRs in the girls-only analysis and three in the boys-only analysis (main model, Fig. 7

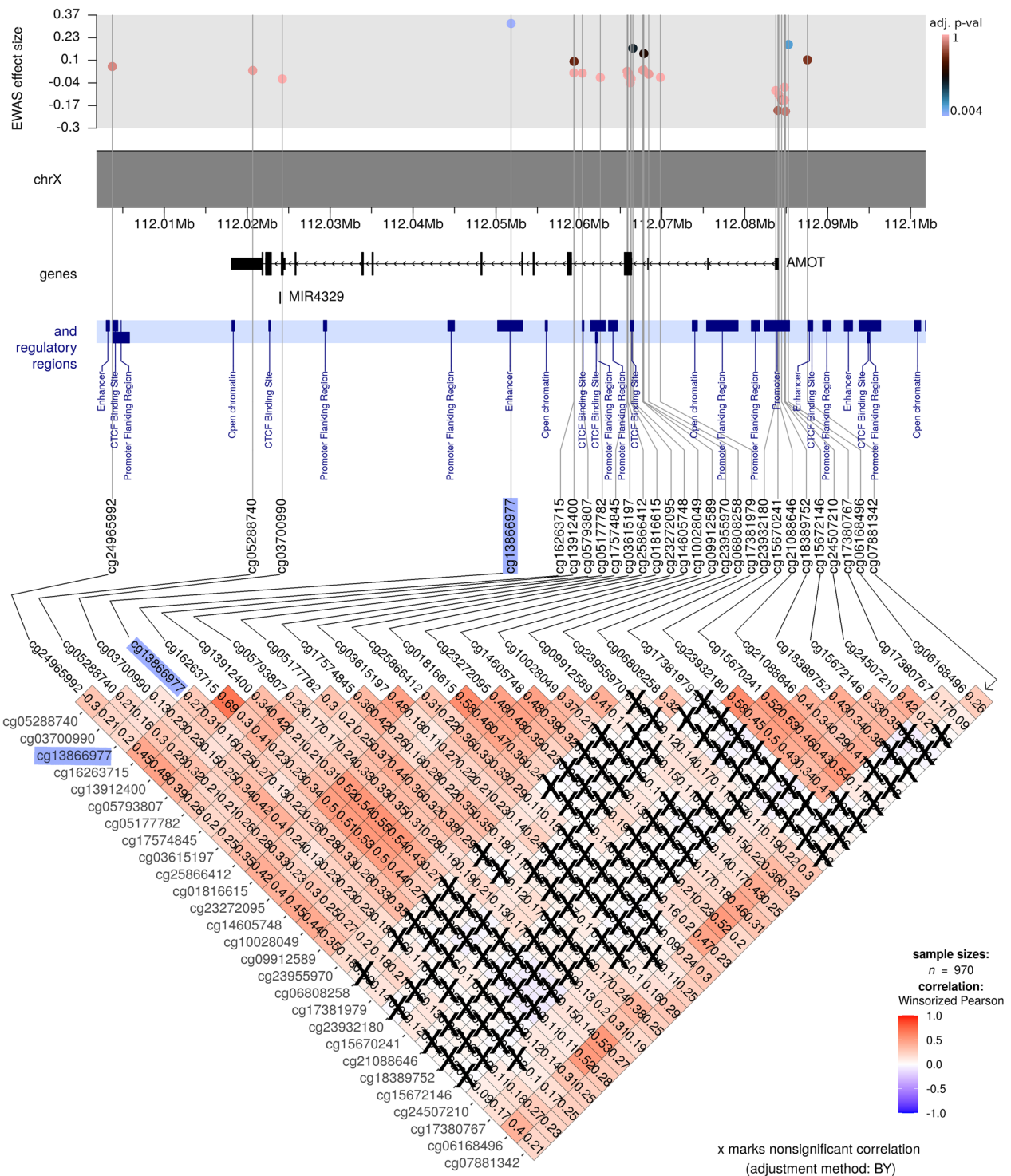
and Additional file 3). We considered a DMR as being statistically significant if it contained three or more CpGs and had an FDR-adjusted *p* value < 0.01. The number of significant DMRs varied only slightly between the main and the adjusted model (see Additional file 1: Figures S4–S6 and the table with all significant DMR findings in Additional file 3); notably, we found eight DMRs in the girls-only analysis and two in the boys-only analysis that were shared across all the models tested. The majority of these DMRs were located in promoter regions. See Additional file 3 for more details as well as the GitHub repository for all the results. In one instance, a DMR included a CpG that was significantly associated with the ART phenotype in boys (Fig. 6). Overall, however, there was only one instance where the DMRs in boys and girls were near each other (Fig. 8, panel A).

#### Analyses in an independent cohort—CHART

To check our results from the MoBa cohort, we ran an analogous model to our main model on an independent dataset from the Australian CHART study (<https://lifecourse.melbournechildrens.com/cohorts/art/>, see Fig. 1 and Methods). Due to the substantially smaller sample size of the CHART cohort (149 ART and 58 non-ART newborns), none of the associations were statistically significant (FDR < 0.01) in boys and



**Fig. 4** Co-methylation patterns in the region flanking the most significant XWAS finding (cg25034591; highlighted in blue) in the girls-only analysis. The top part of the figure shows the effect sizes and FDR-adjusted  $p$  values for the CpGs within 50 kb of cg25034591. The X chromosome coordinates are provided directly underneath. The middle part of the plot shows the location of genes and regulatory regions based on GRCh37 annotations. The bottom part of the figure shows the DNAm correlation matrix, where the color gradient indicates the strength and direction of correlation of DNAm level for each pair of CpGs. Note that the correlation coefficients are provided inside each matrix element (each diamond), and nonsignificant correlations are crossed out (significance was estimated based on adjusted  $p$ -values, using the Benjamini-Yakuteli method)

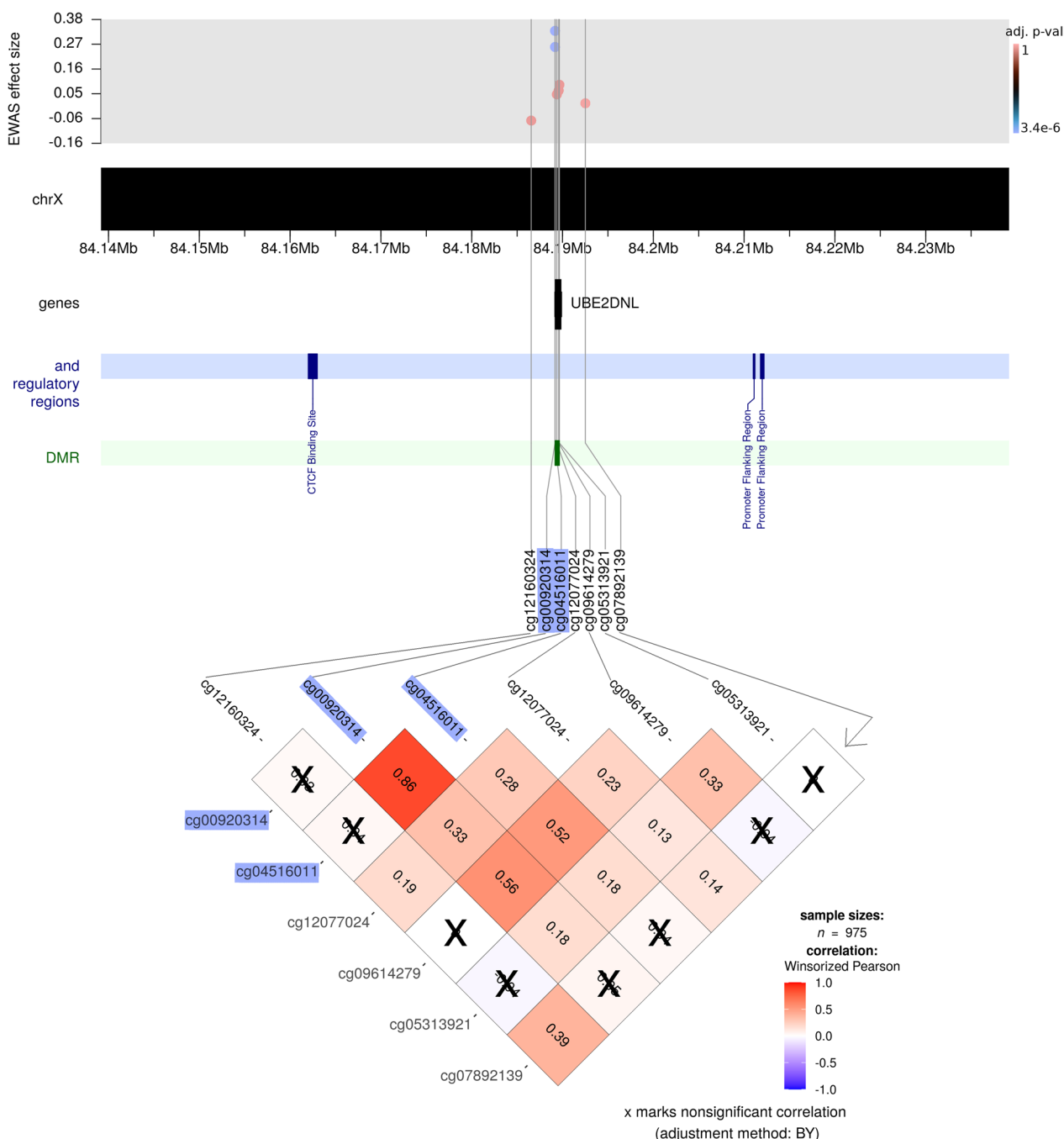


**Fig. 5** Co-methylation patterns among CpGs within 50 kb of cg13866977 (highlighted in blue). The rest of the figure legend is similar to that of Fig. 4 and will therefore not be repeated here or in the remaining co-methylation figures below

there were only three significant associations in girls (cg26018312, cg25174364, and cg15133558; Fig. 9). While these CpGs are located in different areas than

the significant results from the MoBa dataset, they were not linked to any gene or regulatory region, and we were thus unable to analyze these further. However,

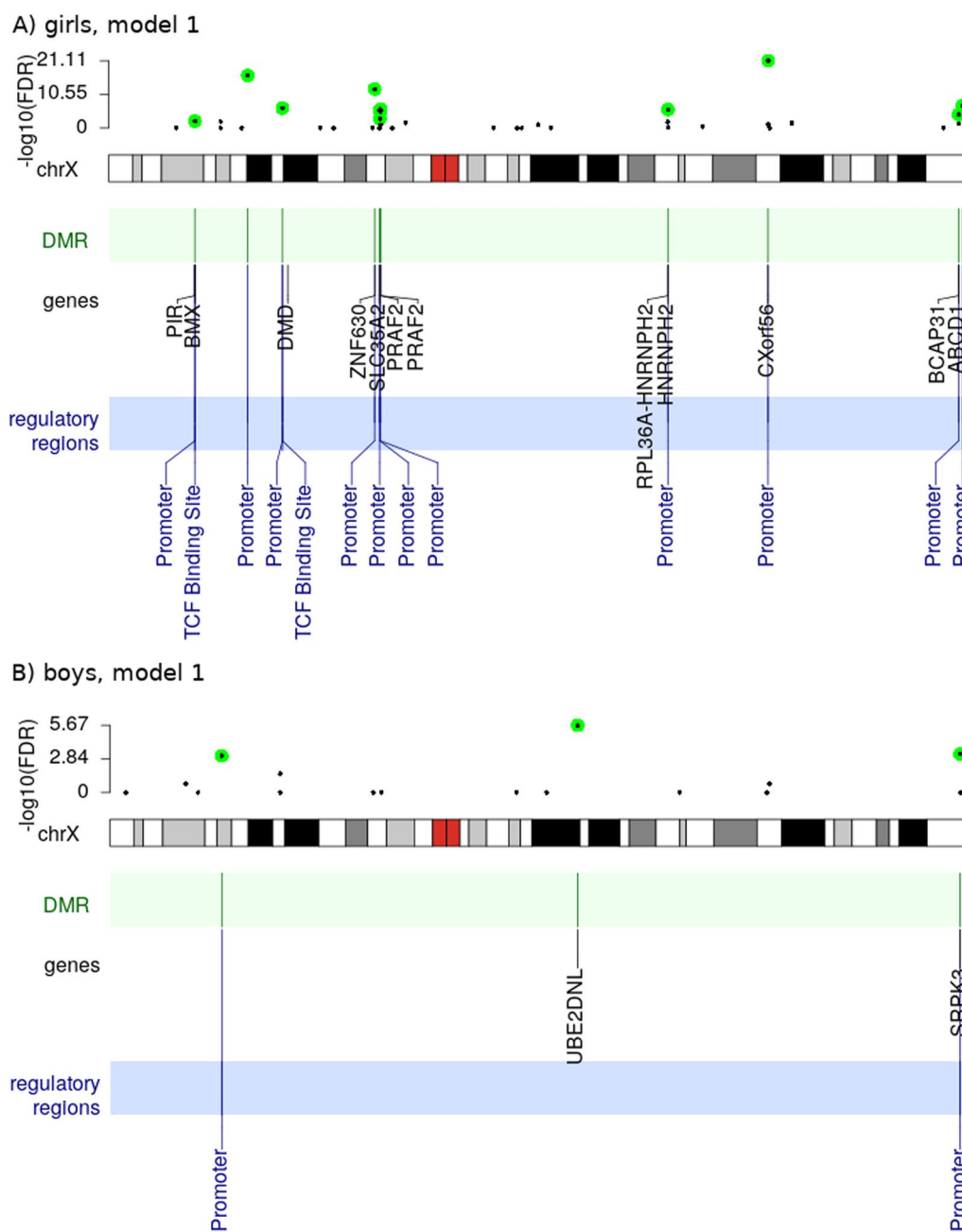




**Fig. 6** Co-methylation patterns in the area flanking the most significant XWAS finding in the boys-only analysis (cg00920314; highlighted in blue). As with the co-methylation figures for the girls-only analyses, the top part of the plot shows the effect sizes and *p* values for the CpGs within 50 kb of cg00920314, which also includes the next most significant CpG, cg04516011 (also highlighted in blue). In addition, the dark-green bar in the plot indicates a DMR in this region

the results did point to changes in DNAm associated with ART status at CpGs in *EIF2S3* and *AMOT* for girls (Additional file 1: Figure S7). Here, the CpGs differed from those in the MoBa cohort because of differences in the quality of DNAm readings in the two

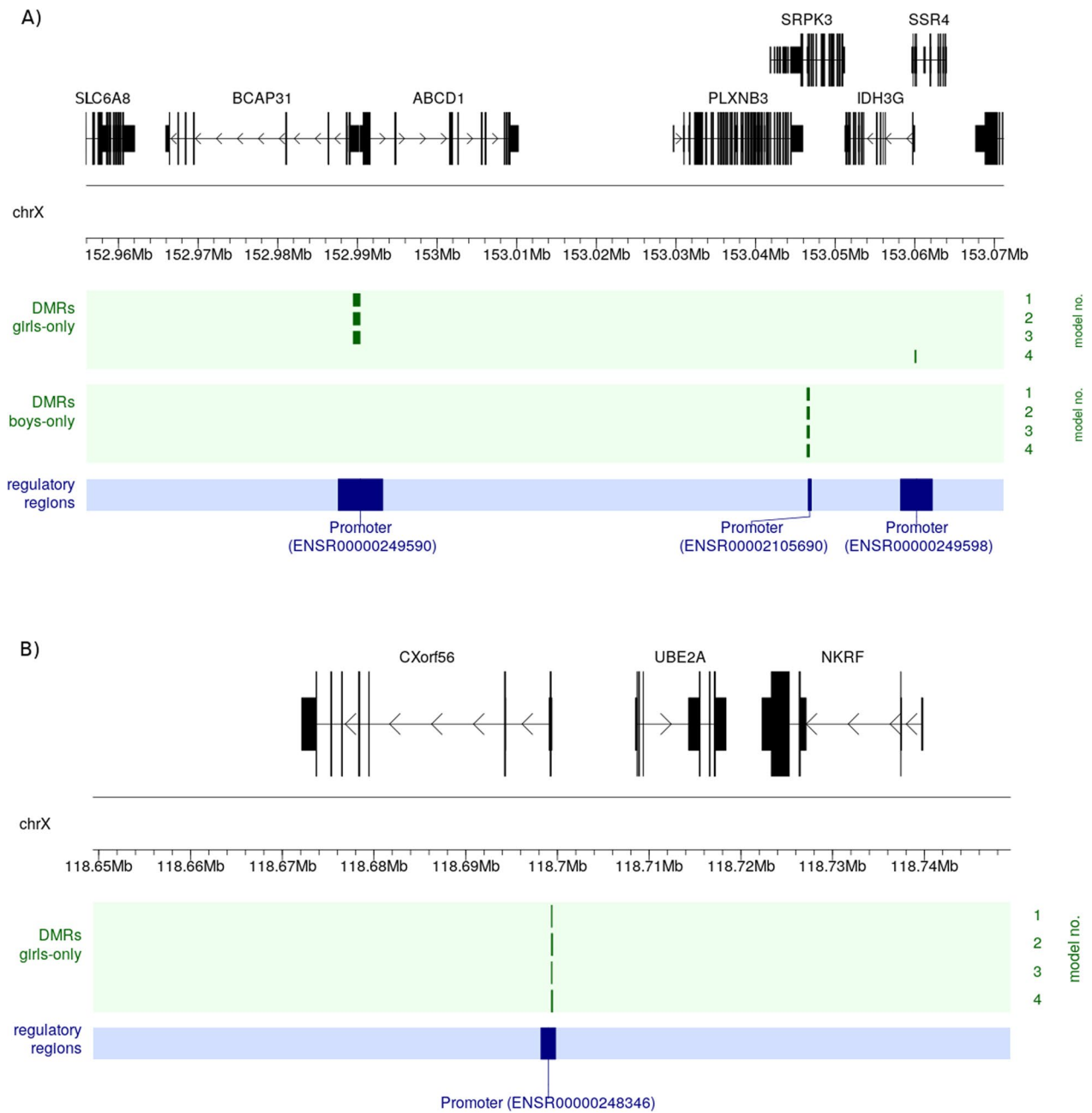
cohorts. We show the CpGs within the *AMOT* gene that were also available in the MoBa dataset in Additional file 1: Figure S8. Moreover, there was a 401 bp-long DMR in *UBE2DNL* in the boys-only analyses, which contained four hypomethylated probes



**Fig. 7** Location of DMRs on the X chromosome. The girls-only analysis is shown in panel A and the boys-only analysis in panel B. Note that we only show the results of the main model (**Model 1**). The top part of each panel shows  $p$  values for all the DMRs that contain at least three CpGs. The FDR-adjusted  $p$  values  $< 0.01$  are marked in green. The bottom part of each panel marks genes and regulatory regions harbored by the significant DMRs

(data not shown). Finally, we observed differences in DNAm levels at cg04516011 and cg00920314 in boys (Additional file 1: Figure S9), which were also identified in the larger MoBa sample (see Additional file 1:

Figure S10 for a comparison). All the results, including  $p$  values and effect sizes, are available online (GitHub repository at <https://github.com/folkhelseinstituttet/X-factor-ART>).

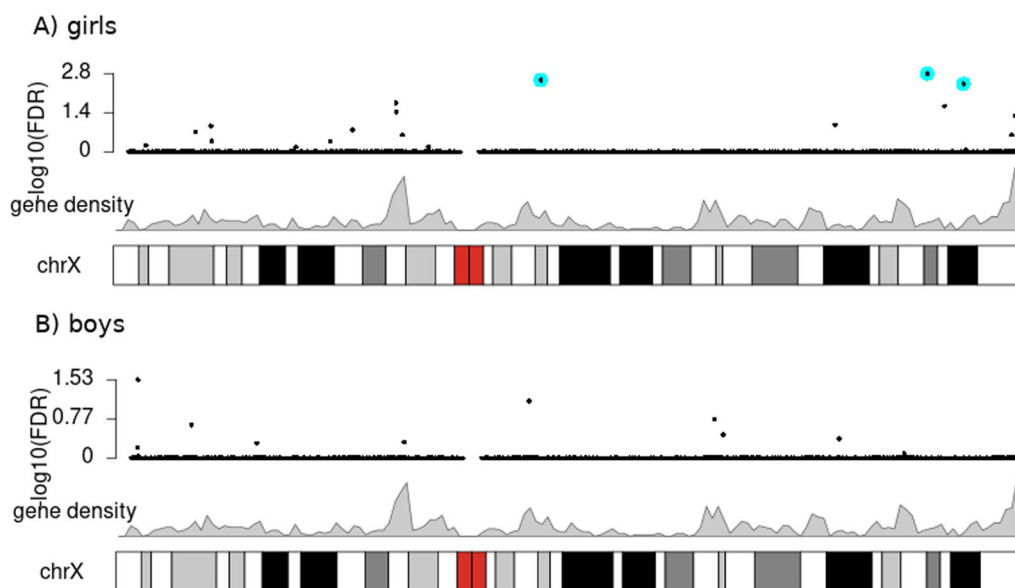


**Fig. 8** Zooming in on genomic features around selected significant DMRs. **A** These significant DMRs detected in the girls-only and boys-only analyses where the only ones localized in the vicinity of each other; the DMRs' positions are marked by green vertical bars in four lines, where each line corresponds to a specific statistical model (**Models 1–4**; right-hand side). **B** The most significant DMR detected in the girls-only analyses

### Discussion

We investigated sex-specific differences in DNAm levels on the X chromosome between newborns conceived through the use of ART and those conceived naturally. Equipped with the largest collection of ART trios to date, we searched for DNAm differences at single-CpG sites as well as in regions. Specifically, we ran

four separate models to check for the effect of various potential confounders, including parental DNAm. We also performed an external check of the results from the MoBa cohort using the main model on a smaller, independent cohort of ART newborns from Australia (CHART).



**Fig. 9** Results of the XWAS of girls (A) and boys (B) from CHART cohort, based on an analogous model to the main model ( $CpG \sim ART + maternal\ smoking$ ). The top plot in each panel is a Manhattan plot of all the tested CpGs. The genomic locations of the most significant findings ( $FDR < 0.01$ ) are highlighted as cyan-colored circles. Immediately beneath is a line plot of gene density and chromosomal bands

#### Characteristics of the ART and non-ART participants in the MoBa cohort.

The ART parents in our study were older than the non-ART parents, and the ART newborns weighed less than the non-ART newborns. Both of these observations are consistent with previous findings [69–73]. Furthermore, fewer of the ART mothers smoked during pregnancy than the non-ART mothers, but, intriguingly, a higher proportion of the ART mothers were past smokers. The observation that more ART mothers were past smokers is noteworthy in light of previous findings linking smoking to impaired fertility in both men and women [74]. Furthermore, a meta-analysis of 21 studies [75] revealed significant associations between smoking at the time of ART treatment and lower success rate for a number of clinical outcomes of ART. Specifically, smoking was associated with lower odds of live birth per cycle, lower odds of clinical pregnancy per cycle, higher odds of spontaneous miscarriage, and higher odds of ectopic pregnancy [75].

#### Significant sex-specific DNAm differences in ART and non-ART newborns

The results of our XWASs of the MoBa data showed significant sex-specific DNAm differences in ART and non-ART newborns. These differences remained significant even after adjusting for several confounders known to be associated with cord-blood DNAm. The results also revealed more differentially methylated CpGs and DMRs in girls than boys, with a slightly lower overall X-chromosome-wide methylation. The differentially methylated

CpGs in our study were mostly located in promoters controlling genes involved in several key developmental processes (e.g., neurodevelopment) and disorders (e.g., intellectual disability and autism).

#### Differential DNAm at cg25034591 suggests upregulation of several genes involved in transcription and translation processes

In the MoBa cohort, the most significant CpG associated with ART in the girls-only analyses, cg25034591, is located in *EIF2S3* and a promoter region (ensembl ID: ENSR00000245352). This promoter region regulates ten genes (<https://www.genecards.org/Search/Keyword?queryString=ENSR00000245352>, see Additional file 1: Table S1), five of which encode a highly interconnected group of proteins with important functions in the regulation of transcription and translation (<https://version-11-5.string-db.org/cgi/network?networkId=b4AYV01yPalh>). The DNAm patterns around cg25034591 form two distinct clusters, one containing a set of positively correlated downstream CpGs and the other a set of negatively correlated upstream CpGs (Fig. 4). This pattern indicates that cg25034591 does not act alone, but operates in concert with other neighboring CpGs. This result was also supported by the independent data from the CHART cohort, where two other CpGs in *EIF2S3* displayed marked differences in DNAm in ART versus non-ART girls (Additional file 1: Figure S7). These CpGs were not present in the MoBa sample analyses because they had been excluded after quality control.

Mutations in *EIF2S3* cause MEHMO, a rare X-linked syndrome characterized by intellectual disability, epilepsy, hypogonadism, hypogenitalism, microcephaly, and obesity [76–78]. Interestingly, both *EIF2S3* and cg25034591 have been reported to escape XCI [33, 59]. We also find evidence for this in our data; notably, the  $\beta$ -values for DNAm at cg25034591 were within the range 0.00009–0.018 in ART-conceived girls and within 0.00016–0.032 in those naturally conceived. It is thus plausible that ART interferes with the escape of XCI at this CpG, leading to an upregulation of genes controlled by the promoter ENSR00000245352. Interestingly, one study reported that impaired imprinted X-chromosome inactivation was responsible for the skewed sex ratio observed after in vitro fertilization [79]. It would therefore be worthwhile to examine XCI in more detail in future analyses.

#### Interpreting the relevance of the findings in the context of ART

The second most significant CpG in girls, cg13866977, lies within a regulatory region and an intron of *AMOT*. This CpG was originally annotated to a region defined as an ‘enhancer’ in the GRCh37 (hg19) version of the genome, but was subsequently changed to ‘promoter flanking region’ in the newer GRCh38 (hg38) genome build (ensembl regulatory ID: ENSR00000912938). It is not unusual for the definition and location of an annotation to change from one genome version to another, especially when the distinction between a promoter and an enhancer becomes blurred as a result of sharing several properties and functions [80]. A perhaps more suitable annotation for cg13866977 in this case would have been ‘transcription regulatory element.’ Furthermore, GeneHancer [81] lists this regulatory region as a putative enhancer (<https://www.genecards.org/Search/Keyword?queryString=ENSR00000912938>) for four genes, one of which is *AMOT* (Additional File 1: Table S1). *AMOT* is a member of the motin family of angiostatin-binding proteins. This gene is especially relevant for ART since it is expressed in placental vessels and the endothelial cells of capillaries, with reported links to premature births [82]. Nevertheless, interpreting the relevance of this finding in the context of ART is not straightforward.

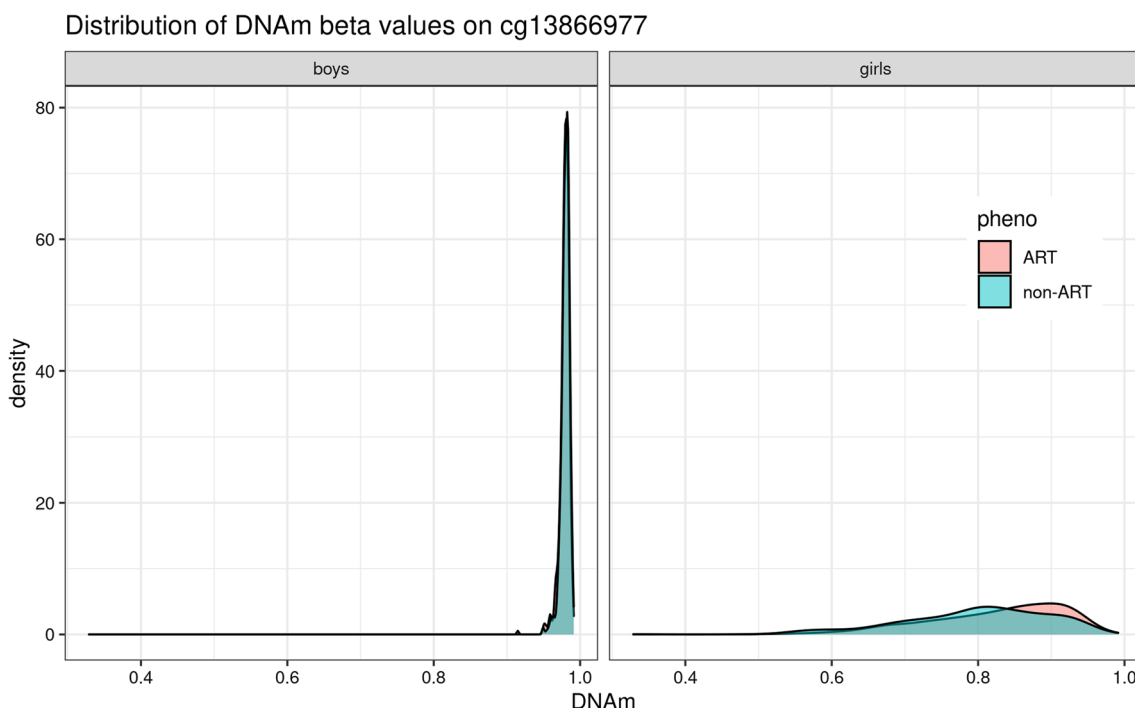
The above-mentioned promoter, ENSR00000912938, is particularly active in six different types of tissues, including the placenta. However, there is no evidence of its activity in cord blood. These observations are based on the ensembl visualization of experimental data showing various histone marker states and DNase1 activity for this promoter ([http://www.ensembl.org/Homo\\_sapiens/Regulation/Summary?db=core;fdb=funcgen;r=X:112806973-112809972;rf=ENSR00000912938](http://www.ensembl.org/Homo_sapiens/Regulation/Summary?db=core;fdb=funcgen;r=X:112806973-112809972;rf=ENSR00000912938)). Furthermore,

according to the Genotype-Tissue Expression (GTEx) database [83], neither *AMOT* nor ‘LHFPL tetraspan subfamily member 1’ (*LHFPL1*), another protein-coding gene controlled by this regulatory region, is transcribed in blood, which is paradoxical given that the DNAm data in both the MoBa and the CHART cohort were generated from newborn’s cord blood.

Our results also showed that the DNAm level at cg13866977 was close to 1.0 in boys, implying that the cytosine at this site is fully methylated. In girls, it was mostly above 0.7 (Fig. 10). Since DNAm signals mainly reflect the level of transcription, we investigated whether transcription factors (TFs) predicted to bind to cg13866977 preferentially bind to the unmethylated or methylated sequence. The output of the search in JASPAR and MeDReaders indicated that none of the seven TFs bind to the methylated sequence (Additional file 6, also available online in the GitHub repository). This suggests that high methylation at this CpG might signal the inactivation of this regulatory region. Moreover, the methylation state was higher among girls conceived by ART than those conceived naturally (effect size = 0.32). The effect size did not change appreciably when we adjusted for parental DNAm at this site (effect size = 0.33). Again, the independent dataset from the CHART study showed a similar trend of association (Additional file 1: Figure S7), except for two other CpGs, cg05177782 and cg09912589, that were positively correlated with cg13866977 (see Fig. 5). Although these results suggest that the regulatory region within *AMOT* is less active after the ART procedure in girls, the specific function of this activation remains to be elucidated.

#### Differential DNAm in boys point to a pseudogene with unknown function

The most significant CpGs in the boys-only analysis, cg00920314 and cg04516011, were both located within *UBE2DNL*. The NCBI gene database (<https://www.ncbi.nlm.nih.gov/gene>) classifies *UBE2DNL* as a pseudogene, but the ExpressionAtlas database [84] (<https://www.ebi.ac.uk/gxa/>) reports that it is expressed in testes. According to UniProt, *UBE2DNL* is inactive because it lacks a cytosine in the active site (<https://www.uniprot.org/uniprot/otkb/Q8IWF7>). There is mounting evidence pointing to pseudogenes as playing important roles in gene regulation instead of simply being evolutionary relics of inactive genes [85, 86]. This notion has garnered additional support through the application of high-throughput sequencing technologies enabling genome-wide characterizations of pseudogenes [87–89]. Similar to our findings, several of the transcribed pseudogenes identified in a previous study by Zheng et al. [87] were also found to be either transcribed exclusively in testes or were



**Fig. 10** The distribution of methylation  $\beta$  values at cg13866977 differs between girls and boys. This CpG is located within a regulatory region and the *AMOT* gene. There is a significant difference in methylation level between girls conceived by ART and girls conceived naturally

particularly active in those tissues. This pattern of testis-specific pseudogene transcription has also been reported by others [90, 91]. The association with *UBE2DNL* in our data also appears to be credible for three reasons: 1) the CpGs remained significant even after adjusting for covariates (see Additional file 2), 2) there is a highly significant DMR in this region, and 3) the data from the independent CHART cohort showed a DMR in this pseudogene (data not shown) and differential DNAm at these two CpGs in ART vs. non-ART boys (Additional file 1: Figure S8).

#### DMRs co-located with genes involved in key developmental processes

The most significant DMR (chrX:118,699,347–118,699,412 in GRCh37) in the girls-only analyses of the MoBa cohort is located within the promoter of three genes (see Fig. 8B and Additional file 1: Table S2 for details). This promoter is active in cord blood, and the genes linked to this promoter are important for immune response, mitochondrial processes, and chromosome segregation. (See the references in Additional file 1: Table S2.) Specifically, one of the genes controlled by this promoter is ‘STING1 ER exit protein 1’ (*STEEP1*, previously called *CXorf56*, Fig. 8B). Mutations in *STEEP1* cause X-linked intellectual disability and other neurological disorders [92, 93]. The DMR encompassing *STEEP1* was also found to have a lower level of DNAm in ART

compared to non-ART newborns, suggesting that the expression of these genes might be up-regulated by ART. The link with *STEEP1* needs to be verified in other similar cohorts when they become available.

Another significant DMR (chrX:152,989,492–152,990,345 in GRCh37) in the girls-only analyses was co-located with a promoter (ensembl ID: ENSR00000249590, see Fig. 8A), which, according to a search in GeneHancer DB, is either a putative promoter or enhancer for nine genes (see Additional file 1: Table S2 for more details). Six of these genes encode proteins that form part of a network, according to the results of text mining and co-expression arrays (STRING DB, <https://version-11-5.string-db.org/cgi/network?networkId=bWgRg6ih0nDV>). Deletions or duplications in many of these genes have been reported to cause different impairments and diseases [94–100], with autism featuring prominently among these clinical manifestations. Right next to this DMR (chrX:152,989,492–152,990,345 in GRCh37), the boys-only analysis revealed another significant DMR (chrX:153,046,451–153,046,767 in GRCh37) co-located with a promoter (ensembl ID: ENSR00002105690, see Fig. 8A) that is active in cord blood. This promoter regulates four genes that are important in the developmental processes of various tissues, including neurons (Additional file 1: Table S2). These indirect connections with autism and neurodevelopment are particularly noteworthy, given previous reports indicating

that neurodevelopmental outcomes differ in children conceived by ART [101, 102], but not always [103, 104].

Lastly, we also checked for any common features among all the genes that co-localized with all the significant DMRs. The STRING protein–protein interaction database (<https://version-11-5.string-db.org/cgi/network?networkId=b1fEsljWdfBy>) indicates that five of the 13 genes found in all the significant DMRs in the girls-only analysis are involved in the X-linked monogenic disease (DOID:0050735, <https://diseases.jensenlab.org/>).

### Strengths and weaknesses

A major strength of our study is the large size of the MoBa sample, enabling a more powerful exploration of questions related to ART and infertility. Additionally, the trio design enabled adjusting for parental DNAm in the regression models, which is essential to correct for other DNAm-dependent parental characteristics as a possible reason for the observed associations in the newborns. Another strength is the mandatory reporting of any use of ART to the Norwegian Medical Birth Registry, including the specifics of the ART procedure used to achieve pregnancy. This ensures virtually complete case ascertainment and a detailed assessment of different ART procedures. Combined with the comprehensive data from questionnaires on relevant covariates, the depth of information on these trios is unparalleled. Furthermore, the DNAm data were generated on the more comprehensive Illumina EPIC array, which is a significant technical leap over its predecessors (Illumina's 27K and 450K Beadchips) in terms of its genomic coverage of regulatory elements, reliability, and reproducibility [105]. One shortcoming of our study is the lack of a well-powered replication cohort with which to compare and validate our findings. EWASs of ART have been far and few between. To our knowledge, the only available dataset was CHART—a small cohort from Australia. Due to differences in the established quality control and analysis pipelines for MoBa and CHART, we were not able to apply the exact same model to both samples. Nevertheless, the model applied to CHART was a close approximation to the main model applied to MoBa. Reassuringly, the results of the main XWAS model in the CHART cohort showed the same trends as observed in the MoBa cohort, despite CHART being significantly smaller and stemming from a different population than MoBa.

### Conclusions

To summarize, our results showed that, for newborns conceived with the help of ART, there were more differentially methylated CpGs and DMRs in girls than boys, with a slightly lower genome-wide methylation in girls

and the opposite pattern in boys. Adjustment for several confounders known to be associated with cord-blood DNAm did not affect the associations, nor did adjustment for parental DNAm, which makes it less likely that parental characteristics were responsible for the observed associations in the newborns. Moreover, our downstream bioinformatic analyses revealed that several of the identified genes were expressed in tissues that are relevant for ART and sex. Furthermore, a number of the genes were associated with neurodevelopment and intellectual disability, which is consistent with previous reports of significant differences in neurodevelopment between newborns conceived by ART and those conceived naturally. More generally, our study fills an important knowledge gap in that it provides an easily adaptable analytic pipeline to investigate the contributions of X-linked CpGs to subfertility and other traits. Its application to the reanalysis of previously published EWASs, such as those available in the EWAS Open Platform [106] and the GEO repository, may facilitate the discovery of additional genes and loci that might have been missed by focusing solely on autosomal CpGs.

### Methods

#### Discovery cohort—MoBa

MoBa is a large population-based pregnancy cohort study in which pregnant women were recruited across Norway from 1999 through 2008 [63]. Fathers were invited from 2001 onward, which explains the lower number of fathers (75,000) compared to mothers and newborns. The participation rate was 41% among the MoBa mothers. Overall, MoBa includes 114,000 children, 95,000 mothers, and 75,000 fathers. Blood samples were initially drawn from the parents at approximately 18 weeks of gestation, and later from the mother and the umbilical cord after delivery [107]. The current analyses were done on a subset of the MoBa data which was included in the 'Study of Assisted Reproductive Technology' (START) project [62]. The current MoBa dataset included 963 trios in which the newborn was conceived using ART and 982 randomly sampled trios in which the newborn was conceived naturally (i.e., by coitus). DNAm in both of these groups of trios was measured using the Illumina Infinium Methylation EPIC BeadChip (Illumina, San Diego, USA) with ~850,000 CpG sites. The inclusion criteria consisted of all of the following: 1) The child was born in the period 2001–2009, 2) the child was a singleton newborn with a record in the Norwegian Medical Birth Registry, 3) the mother filled out and returned the first MoBa questionnaire at around week 17 of gestation, and 4) blood samples were available for the whole trio (child, mother, and father).

In Norway, fertility clinics are mandated to report any ART conception to the national birth registry. We defined ART as ‘any ART’ (excluding intrauterine insemination) and coded it as a binary variable (ART vs. non-ART). As information on the ART procedure was missing for 79 of the trios, these were excluded from the analysis.

#### DNAm measurements in the discovery cohort

DNA samples from the ART and non-ART trios in the MoBa cohort were shipped to Life & Brain GmbH in Bonn, Germany, for further sample processing and measurement of DNAm on the Illumina Infinium Methylation EPIC BeadChip platform (Illumina, San Diego, USA). Extensive details regarding the quality control (QC) pipeline used for data cleaning have been provided in our previous work [62]. Briefly, we established a QC pipeline based on the `RnBeads` package [108] using the statistical programming language R [109]. Cross-hybridizing probes and probes in which the last three bases overlapped with a SNP were removed from the analyses. Additionally, probes with a detection  $p$  value above 0.01 were removed. The `greedy` algorithm was then applied to remove probes and samples showing outlying DNAm values. This procedure minimizes the false positive rate and maximizes the sensitivity when the retained measurements are considered as prediction for the reliable ones. The remaining DNAm data were corrected for background noise using the `enmix.oob` function [110].

We extracted DNAm data on the X chromosome only and applied BMIQ [111] to normalize the Type I and Type II probes [112]. We then checked for multimodality of DNAm per CpG for girls and boys separately using the `gaphunter` function in the `minfi` R package [113, 114]. Crucially, the QC functions applied to the data did not combine any information across samples, which is essential to keep the analyses separate for males and females due to their distinct modalities. The total number of probes on the X chromosome remaining for the current analyses was 16,841, out of the initial 19,090 X chromosome probes present on the EPIC array.

Figure 1 provides an overview of the analytic pipeline and study population, and the sections below provide additional details.

#### Statistical analyses in the discovery cohort

**Regression models.** In preparation for the XWAS of the MoBa cohort, we used the `logit2()` function from R package `minfi` [114] to transform  $\beta$ -values for DNAm into M-values, since M-values are more amenable to statistical tests [115]. Four regression models were fit for boys and girls separately to estimate the difference in methylation levels between the ART and non-ART newborns. This stratification by sex is necessary because of

the distinctly different overall DNAm profiles for girls and boys on the X chromosome. In previously published studies, a number of variables were reported to be associated with DNAm in cord blood and with the use of ART, including mother’s age, smoking status, BMI, and whether she was primiparous. These variables were included as potential confounders in the model, i.e.,  $CpG \sim ART + \text{maternal age} + \text{maternal smoking} + \text{maternal BMI} + \text{primiparity}$  (referred to as the ‘main model’; see also Fig. 1). Although all samples were randomly placed on the bisulfite conversion plates before measuring DNAm, the regression model also included plate ID as a random effect to adjust for batch effects. As DNAm levels associated with parental infertility may confound the XWAS results in the newborns, we ran additional models where we adjusted for maternal methylation in the boys-only analysis and for both maternal and paternal methylation in the girls-only analysis (referred to as the ‘adjusted model’). These adjustments were included in the models as fixed effects, separately for each CpG; for more details, please refer to our previous publication [62]. Moreover, we extended each of the two aforementioned models by including further adjustments for gestational age and birthweight (see Fig. 1). Linear mixed models were implemented using the `rint.reg` function in the R package `Rfast` [116, 117].

**Controlling for inflation of the test statistics.** Additional file 1: Figure S1 depicts the density curves of the DNAm values ( $\beta$ -values) in the newborns according to sex and type of probe on the Illumina EPIC array. The methylation patterns are distinctively different in males and females, as has also been reported by other studies (e.g., [59]). Notably, the middle portion of the distribution for females typically exhibits a bump, as a consequence of XCI, whereas males exhibit higher densities at the opposite ends of the distribution. As females have two copies of the X chromosome, and one copy is silenced through XCI, the distribution of the average DNAm is flatter in females than males.

As pointed out by several reports [118–120], large-scale hypothesis testing of high-dimensional data (e.g., those stemming from a GWAS, EWAS, or XWAS) may be prone to heavily inflated type I error when using the theoretical null distribution to assess the significance of the  $p$  values. We, therefore, used the R package BACON [120] to re-scale the raw z-statistics from the XWAS. BACON is a Bayesian method that controls the false positive rate and accounts for potentially poorly calibrated test statistics while preserving statistical power. We chose BACON over competing methods because it is flexible and can handle a larger proportion of true associations [120, 121]. As can be seen in Additional file 1: Figure S2, BACON reduced inflation substantially in girls



but had a negligible effect in boys. After this correction, we applied a false discovery rate (FDR) < 0.01 to select CpGs that were significantly associated with ART in our sample.

#### Consistency of significant findings in the discovery cohort

We applied a bootstrapping scheme to the XWAS results to evaluate the consistency with which a significant CpG was identified as being significant. We created 1000 bootstrap samples with replacement separately for girls and boys, ensuring an equal proportion of ART and non-ART cases as in the original MoBa dataset. We then reran the analysis using the same main model for each of the bootstrap samples and determined the proportion of times each CpG was found to be significant (using the same significance threshold as previously).

#### Co-methylated CpGs and DMR detection in the discovery cohort

We retrieved the annotation tracks from Ensembl BioMart [122] (<http://www.ensembl.org>) using the R package `biomaRt` [123, 124] and generated a regional plot of the association results. This regional plot was subsequently combined with a co-methylation (correlation) plot of neighboring CpG sites flanking the significant CpGs. The correlation of DNAm values was calculated and plotted using `ggstatsplot` [125]. The rationale for this analysis is that if the biological functions of two CpGs are correlated, their DNAm levels are expected to change in the same way between ART and non-ART samples.

We chose the `dmrff` R package [126] to identify differentially methylated regions (DMRs). This choice was based on the results of a recent paper demonstrating the superior performance of `dmrff` to four frequently used methods for DMR detection: `DMRcate`, `comb-p`, `seqIm`, and `GlobalP` [127]. Finally, the R package `karyoploteR` (part of Bioconductor [128]) was used to visualize genomic features superimposed on a linear representation of the X chromosome [129].

#### Downstream bioinformatic analyses in the discovery cohort

The most significant CpGs and DMRs (both at FDR < 0.01) from the above analyses were subjected to a series of downstream bioinformatic analyses to unravel the biological processes that might be influenced by DNAm at these CpGs. Briefly, we searched the MeDReaders database [130] (<http://medreader.org/>) and adapted data from Yin et al. [131] to retrieve information about transcription factors (TFs) that preferentially bind to methylated DNA.

(The table is available in the GitHub repository.) We also used JASPAR2022 [132] tracks in the ensembl genome browser ([http://grch37.ensembl.org/Homo\\_sapiens/](http://grch37.ensembl.org/Homo_sapiens/)) to check which TFs bind to the significant CpGs detected in our XWASs. GeneHancer [81] was used to find possible targets of promoter and enhancer regions that are co-located with our results. Further, to obtain information on mRNA transcription and protein expression, we searched ExpressionAtlas [133] (<https://www.ebi.ac.uk/gxa/home>) and HumanProteinAtlas [134] (<https://www.proteinatlas.org/>). Finally, information about gene and protein functions and interactions was gathered via GeneCards [135], UniProt [136], and STRING db [137].

#### External cohort and analysis

To perform an external check of the findings from the MoBa cohort, we analyzed the DNAm data from the Australian CHART cohort. CHART consists of 547 adults conceived with the use of IVF and 549 naturally conceived controls [64, 65, 138]. In a subsample of 149 ART-conceived and 58 non-ART neonates (see Fig. 1), DNAm was measured in DNA isolated from neonatal blood spots (Guthrie spots) using the Illumina EPIC array. Data were preprocessed using the `MissMethyl` R package [139], and low-quality and cross-reactive probes were removed from further analysis [15]. Cell composition was estimated using the Bakulski cord-blood cell reference method [140]. Maternal smoking during pregnancy was predicted using a DNA Methylation score [141]. Linear regression modeling was performed using the `limma` R package [142], with the model:  $\text{CpG} \sim \text{ART} + \text{maternal smoking} + \text{sample\_plate}$ . The analyses were run separately for boys and girls. The preprocessing and analysis pipelines for the CHART and MoBa datasets differ because these were established through separate research processes, and the data were collected differently, were stored in different countries, and could not be accessed by the same researchers.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-023-00484-6>.

**Additional file 1.** Supplementary tables and figures.

**Additional file 2.** All significant single CpGs.

**Additional file 3.** All significant DMRs.

**Additional file 4.** Bootstrapping results for boys-only analyses.

**Additional file 5.** Bootstrapping results for girls-only analyses.

**Additional file 6.** Transcription factor classification by JASPAR and MeDReaders.

### Acknowledgements

MoBa is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families who take part in this ongoing cohort study.

### Author contributions

H.E.N. and C.M.P. preprocessed raw DNAm data and performed quality control. J.R. and H.E.N. ran analyses and interpreted the results, as well as wrote the manuscript and prepared the code and data for sharing in the GitHub repository. A.J. formulated the research question, drafted the manuscript, and was a major contributor to the manuscript overall. J.B., W.R.P.D., Y.L., M.C.M., K.L.H., M.G., H.K.G., R.L., P.M., S.E.H. helped shape the final version of the manuscript. J.B., A.J., J.R., and H.E.N. were heavily involved in discussing the results. M.G., C.M.P., W.R.P.D., and H.K.G. checked the validity of the statistical methods used. B.N. and R.S. performed analyses on the external dataset, CHART, and created figures, as well as contributed to the relevant parts of the manuscript. P.M. and S.E.H. secured funding necessary for obtaining raw DNA methylation data. All authors read and approved the final manuscript.

### Funding

Open access funding was provided by the University of Bergen. The START project was funded by the Research Council of Norway (RCN), through its Centers of Excellence funding scheme (project number 262700), and by intramural funding at the Norwegian Institute of Public Health (NIPH). BN is supported by an NHMRC Investigator Grant (APP1173314). The funding bodies did not play any role in the design of the study, data collection, analysis, interpretation of the results, and the writing of this manuscript.

### Availability of data and materials

The MoBa and START data are available from the NIPH, but restrictions apply regarding the availability of these data, which were originally used under specific approvals for the current study and are therefore not publicly available. Data are, however, available from the authors upon reasonable request and after approval by relevant authorities and the NIPH. The CHART data have been deposited in the Gene Expression Omnibus repository, under accession number GSE131433. The scripts for all the above analyses as well as all the results of calculations are available in the GitHub repository at <https://github.com/folkhelseinstituttet/X-factor-ART>.

### Declarations

#### Ethics approval and consent to participate

The establishment of MoBa and the initial data collection were based on a license from the Norwegian Data Protection Agency and additional approvals from the Regional Committees for Medical and Health Research Ethics (REK) in Norway. The MoBa cohort is now based on regulations in the Norwegian Health Registry Act. The current study was approved by the Southeast branch of REK in Norway (Reference Number 2017/1362). The Australian CHART study was approved by the Human Research Ethics Committee (RCH HREC Project 33163) of the Royal Children's Hospital, Melbourne, Australia.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. <sup>2</sup>Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway. <sup>3</sup>Deeplnsight, 0154 Oslo, Norway. <sup>4</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. <sup>5</sup>Department of Human Genetics, University of Chicago, Chicago, IL, USA. <sup>6</sup>Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway. <sup>7</sup>Murdoch Children's Research Institute, Melbourne, Australia. <sup>8</sup>Department of Paediatrics, University of Melbourne, Melbourne, Australia. <sup>9</sup>Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway.

Received: 9 January 2023 Accepted: 10 April 2023

Published online: 21 April 2023

### References

- De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, et al. ART in Europe, 2014: results generated from European registries by ESHRE: the European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod*. 2018;33(9):1586–601.
- Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. *Lancet*. 1978;2(8085):366.
- Balasz J, Gratacós E. Delayed childbearing: effects on fertility and the outcome of pregnancy. *Curr Opin Obstetrics Gynecol*. 2012;24(3):187–93.
- Leridon H, Slama R. The impact of a decline in fecundity and of pregnancy postponement on final number of children and demand for assisted reproduction technology. *Hum Reprod*. 2008;23(6):1312–9. <https://doi.org/10.1093/humrep/den106>.
- Schmidt L, Sobotka T, Bentzen JG, Nyboe Andersen A. ESHRE Reproduction and Society Task Force. Demographic and medical consequences of the postponement of parenthood. *Hum Reprod Update*. 2012;18(1):29–43.
- 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(7511):606–10.
- Mani S, Ghosh J, Coutifaris C, Sapienza C, Mainigi M. Epigenetic changes and assisted reproductive technologies. *Epigenetics*. 2020;15(1–2):12–25.
- von Meyenn F, Reik W. Forget the parents: epigenetic reprogramming in human germ cells. *Cell*. 2015;161(6):1248–51.
- Castillo-Fernandez JE, Loke YJ, Bass-Stringer S, Gao F, Xia Y, Wu H, et al. DNA methylation changes at infertility genes in newborn twins conceived by in vitro fertilisation. *Genome Med*. 2017;9(1):28. <https://doi.org/10.1186/s13073-017-0413-5>.
- Choufani S, Turinsky AL, Melamed N, Greenblatt E, Brudno M, Bérard A, et al. Impact of assisted reproduction, infertility, sex and paternal factors on the placental DNA methylome. *Hum Mol Genet*. 2019;28(3):372–85.
- El Hajj N, Haertle L, Dittrich M, Denk S, Lehnen H, Hahn T, et al. DNA methylation signatures in cord blood of ICSI children. *Hum Reprod*. 2017;32(8):1761–9.
- 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(3):629–39.
- Gentilini D, Somigliana E, Pagliardini L, Rabelotti E, Garagnani P, Bernardinelli L, et al. Multifactorial analysis of the stochastic epigenetic variability in cord blood confirmed an impact of common behavioral and environmental factors but not of in vitro conception. *Clin Epigenetics*. 2018;10(1):77. <https://doi.org/10.1186/s13148-018-0510-3>.
- 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(6):474–83. <https://doi.org/10.4161/15592294.2014.988041>.
- Novakovic B, Lewis S, Halliday J, Kennedy J, Burgner DP, Czajko A, et al. Assisted reproductive technologies are associated with limited epigenetic variation at birth that largely resolves by adulthood. *Nat Commun*. 2019;10(1):3922.
- 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.
- Yeung EH, Mendola P, Sundaram R, Zeng X, Guan W, Tsai MY, et al. Conception by fertility treatment and offspring deoxyribonucleic acid methylation. *Fertil Steril*. 2021;116(2):493–504.
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McClay K, Muzny D, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434(7031):325–37.
- Sidorenko J, Kassam I, Kemper KE, Zeng J, Lloyd-Jones LR, Montgomery GW, et al. The effect of X-linked dosage compensation on complex trait variation. *Nat Commun*. 2019;10(1):3009.
- Balaton BP, Dixon-McDougall T, Peeters SB, Brown CJ. The exceptional nature of the X chromosome. *Hum Mol Genet*. 2018;27(R2):R242–9.
- Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, et al. Accounting for eXentricities: analysis of the X Chromosome in GWAS Reveals

- X-linked genes implicated in autoimmune diseases. *PLoS ONE*. 2014;9(12): e113684. <https://doi.org/10.1371/journal.pone.0113684>.
22. Chung RH, Morris RW, Zhang L, Li YJ, Martin ER. X-APL: an improved family-based test of association in the presence of linkage for the X chromosome. *Am J Hum Genet*. 2007;80(1):59–68.
  23. Ding J, Lin S, Liu Y. Monte Carlo pedigree disequilibrium test for markers on the X chromosome. *Am J Hum Genet*. 2006;79(3):567–73.
  24. Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z, et al. XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J Hered*. 2015;106(5):666–71.
  25. Ho GYF, Bailey-Wilson JE. The transmission/disequilibrium test for linkage on the X chromosome. *Am J Hum Genet*. 2000;66(3):1158–60.
  26. Horvath S, Laird NM, Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction for X-chromosomal markers. *Am J Hum Genet*. 2000;66(3):1161–7.
  27. Jugessur A, Skare Ø, Lie RT, Wilcox AJ, Christensen K, Christiansen L, et al. X-linked genes and risk of orofacial clefts: Evidence from two population-based studies in Scandinavia. *PLoS ONE*. 2012;7(6):1–12.
  28. Skare Ø, Gjessing HK, Gjerdevik M, Haaland ØA, Romanowska J, Lie RT, et al. A new approach to chromosome-wide analysis of X-linked markers identifies new associations in Asian and European case-parent triads of orofacial clefts. *PLoS ONE*. 2017;12(9): e0183772. <https://doi.org/10.1371/journal.pone.0183772>.
  29. Skare Ø, Lie RT, Haaland ØA, Gjerdevik M, Romanowska J, Gjessing HK, et al. Analysis of parent-of-origin effects on the X chromosome in Asian and European orofacial cleft triads identifies associations with DMD, FGF13, EGFL6, and additional loci at Xp22.2. *Front Genet*. 2018;9(February). <https://doi.org/10.3389/fgene.2018.00025/full>.
  30. Zhang L, Martin ER, Chung RH, Li YJ, Morris RW. X-LRT: a likelihood approach to estimate genetic risks and test association with X-linked markers using a case-parents design. *Genet Epidemiol*. 2008;32(4):370–80. <https://doi.org/10.1002/gepi.20311>.
  31. Zhang L, Martin ER, Morris RW, Li YJ. Association test for X-linked QTL in family-based designs. *Am J Hum Genet*. 2009;84(4):431–44.
  32. Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet*. 2011;12(8):565–75. <https://doi.org/10.1038/nrg3032>.
  33. Navarro-Cobos MJ, Balaton BP, Brown CJ. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. *Am J Med Genet C Semin Med Genet*. 2020;184(2):226–38.
  34. Bermejo-Alvarez P, Ramos-Ibeas P, Gutierrez-Adan A. Solving the “X” in embryos and stem cells. *Stem Cells Dev*. 2012;21(8):1215–24. <https://doi.org/10.1089/scd.2011.0685>.
  35. The Migeon BR. Role of X inactivation and cellular mosaicism in Women’s Health and sex-specific diseases. *JAMA*. 2006;295(12):1428. <https://doi.org/10.1001/jama.295.12.1428>.
  36. Posytnick BJ, Brown CJ. Escape from X-chromosome inactivation: an evolutionary perspective. *Front Cell Dev Biol*; 2019. p. 7.
  37. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet*. 2015;24(6):1528–39. <https://doi.org/10.1093/hmg/ddu564>.
  38. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, et al. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017;550(7675):244–8.
  39. Wainer Katsir K, Linnal M. Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics*. 2019;20(1):201.
  40. Helena Mangs A, Morris B. The human pseudoautosomal region (PAR): origin, function and future. *Curr Genom*. 2007;8(2):129–36.
  41. Raudsepp T, Chowdhary BP. The Eutherian pseudoautosomal region. *Cytogenet Genome Res*. 2015;147(2–3):81–94.
  42. Russell LM, Strike P, Browne CE, Jacobs PA. X chromosome loss and ageing. *Cytogenet Genome Res*. 2007;116(3):181–5.
  43. Bermejo-Alvarez P, Rizo D, Lonergan P, Gutierrez-Adan A. Transcriptional sexual dimorphism during preimplantation embryo development and its consequences for developmental competence and adult health and disease. *Reproduction*. 2011;141(5):563–70.
  44. Kobayashi S, Isotani A, Mise N, Yamamoto M, Fujihara Y, Kaseda K, et al. Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, *Rhox5/Pem*, at preimplantation stages. *Curr Biol*. 2006;16(2):166–72.
  45. Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun*. 2017;8(1):15475.
  46. Williams TM, Carroll SB. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat Rev Genet*. 2009;10(11):797–804.
  47. El-Maarri O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, et al. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet*. 2007;122(5):505–14. <https://doi.org/10.1007/s00439-007-0430-3>.
  48. McCarthy MM, Auger AP, Bale TL, De Vries GJ, Dunn GA, Forger NG, et al. The epigenetics of sex differences in the brain. *J Neurosci*. 2009;29(41):12815–23. <https://doi.org/10.1523/JNEUROSCI.3331-09.2009>.
  49. Nino CL, Perez GF, Isaza N, Gutierrez MJ, Gomez JL, Nino G. Characterization of sex-based DNA methylation signatures in the airways during early life. *Sci Rep*. 2018;8(1):5526.
  50. Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin*. 2013;6(1):4. <https://doi.org/10.1186/1756-8935-6-4>.
  51. Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW. Sex differential in methylation patterns of selected genes in Singapore Chinese. *Hum Genet*. 2005;117(4):402–3.
  52. Singmann P, Shem-Tov D, Wahl S, Gallert H, Fiorito G, Shin SY, et al. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin*. 2015;8(1):43.
  53. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O’Donovan MC, et al. Methylomic trajectories across human fetal brain development. *Genome Res*. 2015;25(3):338–52. <https://doi.org/10.1101/gr.180273.114>.
  54. Suderman M, Simpkin A, Sharp G, Gaunt T, Lyttleton O, McArdle W, et al. Sex-associated autosomal DNA methylation differences are widespread and stable throughout childhood. *bioRxiv*. 2017. Available from: <https://www.biorxiv.org/content/early/2017/03/19/118265>.
  55. Wiencke JK, Zheng S, Lafuente A, Lafuente MJ, Grudzen C, Wrensch MR, et al. Aberrant methylation of p16INK4a in anatomic and gender-specific subtypes of sporadic colorectal cancer. *Cancer Epidemiol Biomark Prevent*. 1999;8(6):501–6.
  56. Xu H, Wang F, Liu Y, Yu Y, Gelernter J, Zhang H. Sex-biased methylome and transcriptome in human prefrontal cortex. *Hum Mol Genet*. 2014;23(5):1260–70. <https://doi.org/10.1093/hmg/ddt516>.
  57. Yousefi P, Huen K, Davé V, Barcellos L, Eskenazi B, Holland N. Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics*. 2015;16(1):911. <https://doi.org/10.1186/s12864-015-2034-y>.
  58. Klebaner D, Huang Y, Hui Q, Taylor JY, Goldberg J, Vaccarino V, et al. X chromosome-wide analysis identifies DNA methylation sites influenced by cigarette smoking. *Clin Epigenetics*. 2016;8(1). <https://doi.org/10.1186/s13148-016-0189-2>.
  59. Li S, Lund JB, Christensen K, Baumbach J, Mengel-From J, Kruse T, et al. Exploratory analysis of age and sex dependent DNA methylation patterns on the X-chromosome in whole blood samples. *Genome Med*. 2020;12(1):39. <https://doi.org/10.1186/s13073-020-00736-3>.
  60. McCartney DL, Zhang F, Hillary RF, Zhang Q, Stevenson AJ, Walker RM, et al. An epigenome-wide association study of sex-specific chronological ageing. *Genome Med*. 2020;12(1):1. <https://doi.org/10.1186/s13073-019-0693-z>.
  61. Kananen L, Marttila S. Ageing-associated changes in DNA methylation in X and Y chromosomes. *Epigenetics Chromatin*. 2021;14(1):33. <https://doi.org/10.1186/s13072-021-00407-6>.
  62. Häberg SE, Page CM, Lee Y, Nustad HE, Magnus MC, Haftorn KL, et al. DNA methylation in newborns conceived by assisted reproductive technology. *Nat Commun*. 2022;13(1):1896.
  63. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, et al. Cohort profile update: the norwegian mother and child cohort study (MoBa). *Int J Epidemiol*. 2016;45(2):382–8. <https://doi.org/10.1093/ije/dyw029>.
  64. Lewis S, Kennedy J, Burgner D, McLachlan R, Ranganathan S, Hammarberg K, et al. Clinical review of 24–35 year olds conceived with

- and without in vitro fertilization: study protocol. *Reprod Health*. 2017;9(14):117.
65. Halliday J, Lewis S, Kennedy J, Burgner DP, Juonala M, Hammarberg K, et al. Health of adults aged 22 to 35 years conceived by assisted reproductive technology. *Fertil Steril*. 2019;112(1):130–9.
  66. Liu Y, Li X, Aryee MJ, Ekström TJ, Padyukov L, Klareskog L, et al. GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am J Hum Genet*. 2014;94(4):485–95.
  67. Brenet F, Moh M, Funk P, Feisterstein E, Viale AJ, Socci ND, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE*. 2011;6(1).
  68. Yang X, Han H, DeCarvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*. 2014;26(4):577–90. <https://doi.org/10.1016/j.ccr.2014.07.028>.
  69. Goisis A, Håberg SE, Hanevik HI, Magnus MC, Kravdal Ø. The demographics of assisted reproductive technology births in a Nordic country. *Hum Reprod*. 2020;35(6):1441–50.
  70. Magnus MC, Wilcox AJ, Fadum EA, Gjessing HK, Opdahl S, Juliusson PB, et al. Growth in children conceived by ART. *Hum Reprod*. 2021;36(4):1074–82.
  71. Hann M, Roberts SA, D'Souza SW, Clayton P, Macklon N, Brison DR. The growth of assisted reproductive treatment-conceived children from birth to 5 years: a national cohort study. *BMC Med*. 2018;16(1):224. <https://doi.org/10.1186/s12916-018-1203-7>.
  72. De Geyter C, De Geyter M, Steimann S, Zhang H, Holzgreve W. Comparative birth weights of singletons born after assisted reproduction and natural conception in previously infertile women. *Hum Reprod*. 2006;21(3):705–12.
  73. Tong VT, Kissin DM, Bernson D, Copeland G, Boulet SL, Zhang Y, et al. With maternal smoking among women, use without, of assisted reproductive technologies. *J Womens Health*. 2016;25(10):1066–72. <https://doi.org/10.1089/jwh.2015.5662>.
  74. Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion (US). *The Health Consequences of Smoking — 50 Years of Progress: A Report of the Surgeon General*. Centers for Disease Control and Prevention (US), Atlanta, GA; 2014.
  75. Waylen AL, Metwally M, Jones GL, Wilkinson AJ, Ledger WL. Effects of cigarette smoking upon clinical outcomes of assisted reproduction: a meta-analysis. *Hum Reprod Update*. 2008;15(1):31–44. <https://doi.org/10.1093/humupd/dmn046>.
  76. Moortgat S, Désir J, Benoit V, Boulanger S, Pendeville H, Nassogne MC, et al. Two novel EIF253 mutations associated with syndromic intellectual disability with severe microcephaly, growth retardation, and epilepsy. *Am J Med Genet A*. 2016;170(11):2927–33. <https://doi.org/10.1002/ajmg.a.37792>.
  77. Moortgat S, Manfroid I, Pendeville H, Freeman S, Bourdouxhe J, Benoit V, et al. Broadening the phenotypic spectrum and physiological insights related to EIF253 variants. *Hum Mutat*. 2021;42(7):827–34. <https://doi.org/10.1002/humu.24215>.
  78. Skopkova M, Hennig F, Shin BS, Turner CE, Stanikova D, Brennerova K, et al. EIF253 mutations associated with severe X-linked intellectual disability syndrome MEHMO. *Hum Mutat*. 2017;38(4):409–25.
  79. Tan K, An L, Miao K, Ren L, Hou Z, Tao L, et al. Impaired imprinted X chromosome inactivation is responsible for the skewed sex ratio following in vitro fertilization. *Proc Natl Acad Sci USA*. 2016;113:3197–202.
  80. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet*. 2020;21(2):71–87. <https://doi.org/10.1038/s41576-019-0173-8>.
  81. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017(1):1665–80. <https://doi.org/10.1093/database/bax028>.
  82. Casavant SG, Cong X, Moore J, Starkweather A. Associations between preterm infant stress, epigenetic alteration, telomere length and neurodevelopmental outcomes: A systematic review. *Early Human Dev*. 2019;131:63–74.
  83. GTEx Consortium, Laboratory DACCLWG, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13.
  84. Moreno P, Fexova S, George N, Manning JR, Miao Z, Mohammed S, et al. Expression Atlas update: gene and protein expression in multiple species. *Nucleic Acids Res*. 2022;50(D1):D129–40.
  85. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Francisco Carter DR. Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA*. 2011;17(5):792–8. <https://doi.org/10.1261/rna.265831.1>.
  86. Balakirev ES, Ayala FJ. Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet*. 2003;37(1):123–51. <https://doi.org/10.1146/annurev.genet.37.040103.103949>.
  87. Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, et al. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res*. 2007;17(6):839–51. <https://doi.org/10.1101/gr.5586307>.
  88. Svensson Ö, Arvestad L, Genome-Wide LJ. Survey for biologically functional pseudogenes. *PLoS Comput Biol*. 2006;2(5):e46. <https://doi.org/10.1371/journal.pcbi.0020046>.
  89. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu X, et al. The GENCODE pseudogene resource. *Genome Biol*. 2012;13(9):R51. <https://doi.org/10.1186/gb-2012-13-9-r51>.
  90. Reymond A, Marigo V, Yaylaoglu MB, Leoni A, Ucla C, Scamuffa N, et al. Human chromosome 21 gene expression atlas in the mouse. *Nature*. 2002;420(6915):582–6.
  91. Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA. The mouse gene encoding the testis-specific isoform of poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. *J Mol Evol*. 1998;47(3):275–81. <https://doi.org/10.1007/PL00006385>.
  92. Rocha ME, Silveira TRD, Sasaki E, Sás DM, Lourenço CM, Kandaswamy KK, et al. Novel clinical and genetic insight into CXorf56-associated intellectual disability. *Eur J Hum Genet*. 2020;28(3):367–72.
  93. Wolańska E, Pollak A, Rydzanicz M, Pesz K, Klaniewska M, Rozensztrauch A, et al. The role of the reanalysis of genetic test results in the diagnosis of dysmorphic syndrome caused by inherited Xq24 deletion including the UBE2A and CXorf56 genes. *Genes*. 2021;12(3):350.
  94. San Antonio-Arce V, Fenollar-Cortés M, Oancea Ionescu R, DeSantos-Moreno T, Gallego-Merlo J, Illana Cámara FJ, et al. MECP2 Duplications in symptomatic females. *child neurology open*. 2016;3:2329048X1663067. <https://doi.org/10.1177/2329048X16630673>.
  95. Firouzabadi SG, Kariminejad R, Vameghi R, Darvish H, Ghaedi H, Banihashemi S, et al. Copy number variants in patients with autism and additional clinical features: report of VIPR2 duplication and a novel microduplication syndrome. *Mol Neurobiol*. 2017;54(9):7019–27. <https://doi.org/10.1007/s12035-016-0202-y>.
  96. Rosenberg C, Freitas ÉL, Uehara DT, Auricchio MTBM, Costa SS, Oiticica J, et al. Genomic copy number alterations in non-syndromic hearing loss. *Clin Genet*. 2016;89(4):473–7.
  97. Piton A, Gauthier J, Hamdan FF, Lafrenière RG, Yang Y, Henrion E, et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry*. 2011;16(8):867–80.
  98. Fernández RM, Núñez-Torres R, González-Meneses A, Antiñolo G, Borrego S. Novel association of severe neonatal encephalopathy and Hirschsprung disease in a male with a duplication at the Xq28 region. *BMC Med Genet*. 2010;11(1):137. <https://doi.org/10.1186/1471-2350-11-137>.
  99. Iwasa M, Yamagata T, Mizuguchi M, Itoh M, Matsumoto A, Hironaka M, et al. Contiguous ABCD1 DXS1357E deletion syndrome: report of an autopsy case. *Neuropathology*. 2013;33(3):292–8. <https://doi.org/10.1111/j.1440-1789.2012.01348.x>.
  100. Cukier HN, Lee JM, Ma D, Young JI, Mayo V, Butler BL, et al. The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. *Autism Res*. 2012;5(6):385–97. <https://doi.org/10.1002/aur.1251>.
  101. Bergh C, Wennerholm UB. Long-term health of children conceived after assisted reproductive technology. *Upsala J Med Sci*. 2020;125(2):152–7.
  102. Aref-Eshghi E, Kerkhof J, Pedro VP, Barat-Houari M, Ruiz-Pallares N, Andrau JC, et al. Evaluation of DNA methylation epigenotypes for diagnosis and phenotype correlations in 42 mendelian neurodevelopmental disorders. *Am J Hum Genet*. 2020;106(3):356–70.
  103. Farhi A, Gabis LV, Frank S, Glasser S, Hirsh-Yechezkel G, Brinton L, et al. Cognitive achievements in school-age children born following assisted

- reproductive technology treatments: a prospective study. *Early Human Dev.* 2021;155: 105327.
104. Roychoudhury S, Lodha A, Synnes A, Abou Mehrem A, Canning R, Banihani R, et al. Neurodevelopmental outcomes of preterm infants conceived by assisted reproductive technology. *Am J Obstet Gynecol.* 2021;225(3):276.e1-276.
  105. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 2016;17(1):208. <https://doi.org/10.1186/s13059-016-1066-1>.
  106. Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res.* 2022;50(D1):D1004-9.
  107. Paltiel L, Anita H, Skjerdet T, Harbak K, Bækken S, Nina Kristin S, et al. The biobank of the Norwegian Mother and Child Cohort Study - present status. *Norsk Epidemiologi.* 2014;24(1-2).
  108. Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, et al. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol.* 2019;20(1):55. <https://doi.org/10.1186/s13059-019-1664-9>.
  109. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2022. Available from: <https://www.R-project.org>.
  110. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.* 2016;44(3): e20. <https://doi.org/10.1093/nar/gkv907>.
  111. 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(2):189-96. <https://doi.org/10.1093/bioinformatics/bts680>.
  112. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research.* 2016;gkw967. <https://doi.org/10.1093/nar/gkw967>.
  113. Andrews SV, Ladd-Acosta C, Feinberg AP, Hansen KD, Fallin MD. "Gap hunting" to characterize clustered probe signals in Illumina methylation array data. *Epigenetics Chromatin.* 2016;9(1):1-21.
  114. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics.* 12 2016;33:btw691. <https://doi.org/10.1093/bioinformatics/btw691>.
  115. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* 2010;11(1):587. <https://doi.org/10.1186/1471-2105-11-587>.
  116. Tsagris M, Papadakis M. Taking R to its limits: 70+ tips. *PeerJ Preprints.* 2018;6.
  117. Tsagris M, Papadakis M. Forward regression in R: from the extreme slow to the extreme fast. *J Data Sci.* 2021;2(16):771-80.
  118. Efron B. Large-scale simultaneous hypothesis testing. *J Am Stat Assoc.* 2004;99(465):96-104. <https://doi.org/10.1198/016214504000000089>.
  119. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet.* 2019;51(4):592-9.
  120. van Iterson M, van Zwet EV, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017;18(1):19. <https://doi.org/10.1186/s13059-016-1131-9>.
  121. Brägelmann J, Lorenzo BJ. A comparative analysis of cell-type adjustment methods for epigenome-wide association studies based on simulated and real data sets. *Brief Bioinform.* 2019;20(6):2055-65.
  122. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amodè MR, et al. Ensembl 2021. *Nucleic Acids Res.* 2021;49(D1):D884-91.
  123. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics.* 2005;21(16):3439-40. <https://doi.org/10.1093/bioinformatics/bti525>.
  124. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009;4(8):1184-91.
  125. Patil I. Visualizations with statistical details: the 'ggstatsplot' approach. *J Open Source Softw.* 2021;6(61):3167. <https://doi.org/10.21105/joss.03167>.
  126. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K, Dmrrf: identifying differentially methylated regions efficiently with power and control. *bioRxiv.* 2018. p. 1-26.
  127. Lent S, Cardenas A, Rifas-Shiman SL, Perron P, Bouchard L, Liu CT, et al. Detecting differentially methylated regions with multiple distinct associations. *Epigenomics.* 2021;13(6):451-64.
  128. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.* 2015;12(2):115-21.
  129. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics.* 2017;33(19):3088-90.
  130. Wang G, Luo X, Wang J, Wan J, Xia S, Zhu H, et al. MeDReaders: a database for transcription factors that bind to methylated DNA. *Nucleic Acids Res.* 2018;46(D1):D146-51.
  131. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science.* 2017;356(6337):eaaj2239. <https://doi.org/10.1126/science.aaj2239>.
  132. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2022;50(D1):D165-73.
  133. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, et al. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res.* 2019. <https://doi.org/10.1093/nar/gkz947/5609521>.
  134. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science (New York, NY).* 2020;367(6482).
  135. Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. The GeneCards Suite. In: *Practical Guide to Life Science Databases.* Singapore: Springer; 2021. p. 27-56. [https://doi.org/10.1007/978-981-16-5812-9\\_2](https://doi.org/10.1007/978-981-16-5812-9_2).
  136. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480-9.
  137. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-13.
  138. Juonala M, Lewis S, McLachlan R, Hammarberg K, Kennedy J, Saffery R, et al. American Heart Association ideal cardiovascular health score and subclinical atherosclerosis in 22-35-year-old adults conceived with and without assisted reproductive technologies. *Hum Reprod.* 2020;35(1):232-9.
  139. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2016;1(32):286-8.
  140. Bakulski KM, Feinberg JL, Andrews SV, Yang J, Brown S, McKenney SL, et al. DNA methylation of cord blood cell types: applications for mixed cell birth studies. *Epigenetics.* 2016;11:354-62.
  141. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Häberg SE, et al. DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. *Environ Health Perspect.* 2017;125:760-6.
  142. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;4(43):e47-7.
  143. Ehrmann IE, Ellis PS, Mazeyrat S, Duthie S, Brockdorff N, Mattei MG, et al. Characterization of genes encoding translation initiation factor eIF-2 in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. *Hum Mol Genet.* 1998;7(11):1725-37. <https://doi.org/10.1093/hmg/7.11.1725>.
  144. Sánchez JM, Gómez-Redondo I, Browne JA, Planells B, Gutiérrez-Adán A, Lonergan P. MicroRNAs in amniotic fluid and maternal blood plasma associated with sex determination and early gonad differentiation in cattle. *Biol Reprod.* 2021;105(2):345-58.

145. Farrell A, Alahari S, Ermini L, Tagliaferro A, Litvack M, Post M, et al. Faulty oxygen sensing disrupts angiotensin function in trophoblast cell migration and predisposes to preeclampsia. *JCI Insight*. 2019;4(8).
146. Leung CY, Zernicka-Goetz M. Angiotensin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat Commun*. 2013;4(1):2251.
147. Moleirinho S, Guerrant W, Kissil JL. The Angiotensins—from discovery to function. *FEBS Lett*. 2014;588(16):2693–703.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

