



Timms, J. A., Relton, C. L., Sharp, G. C., Rankin, J., Stratthdee, G., & McKay, J. A. (2019). Exploring a potential mechanistic role of DNA methylation in the relationship between in utero and post-natal environmental exposures and risk of childhood acute lymphoblastic leukaemia. *International Journal of Cancer*, 145(11), 2933-2943. <https://doi.org/10.1002/ijc.32203>

Peer reviewed version

Link to published version (if available):  
[10.1002/ijc.32203](https://doi.org/10.1002/ijc.32203)

[Link to publication record on the Bristol Research Portal](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at <https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.32203> . Please refer to any applicable terms of use of the publisher.

## University of Bristol – Bristol Research Portal

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

**Exploring a potential mechanistic role of DNA methylation in the relationship between *in utero* and post-natal environmental exposures and risk of childhood acute lymphoblastic leukaemia**

Jessica A Timms<sup>1</sup>, Caroline L Relton<sup>2</sup>, Gemma C Sharp<sup>2</sup>, Judith Rankin<sup>1</sup>, Gordon Strathdee<sup>3</sup>, Jill A M<sup>c</sup>Kay<sup>1,4</sup>

<sup>1</sup>*Institute of Health & Society, Newcastle University, U.K.*, <sup>2</sup>*MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol*, <sup>3</sup>*Northern Institute for Cancer Research, Newcastle University, U.K.*, <sup>4</sup>*Department of Applied Sciences, Northumbria University.*

**Corresponding author:** Jessica Ann Timms, Research Oncology, King's College London, 3rd Floor, Bermondsey Wing, Guy's Hospital, London, SE1 9RT; email: Jessica.timms@kcl.ac.uk

**Keywords:** radiation; alcohol; smoking; caffeine; ALSPAC

**Word count:** 4653

**Financial support:** Funding was provided via a studentship from the Institute of Health and Society, Newcastle University, UK, and from the North of England Children's Cancer Research charity (NECCR). CL Relton is supported by the UK Medical Research Council Integrative Epidemiology Unit (MC\_UU\_12013\_2) and Cancer Research UK (C18281/A19169). The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Jessica A Timms, Caroline L Relton, Gemma C Sharp, Judith Rankin, Gordon Strathdee, and Jill A M<sup>c</sup>Kay will serve as guarantors for the contents of this paper. The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1; and BB/I025263/1). No writing assistance was utilized in the production of this manuscript.

**Brief description – “Novelty and Impact”**

Childhood acute lymphoblastic leukaemia (ALL) is a multi “hit” disease, whereby additional “hits” are required to transform pre-malignant cells carrying ALL-associated genetic translocations. We investigated the potential that environmental exposures associated with increased ALL risk influenced changes in DNA methylation patterns in offspring, and their concordance with methylation changes identified in ALL disease. We found statistically significant directionally concordant changes in methylation for maternal radiation exposure, alcohol, sugary caffeinated drinks; and childhood day care attendance.

## Abstract

The aetiology of childhood acute lymphoblastic leukaemia (ALL) is unclear. Genetic abnormalities have been identified in a number of ALL cases, although these alone are not sufficient for leukaemic transformation. Various *in utero* and post-natal environmental exposures have been suggested to alter risk of childhood ALL. DNA methylation patterns can be influenced by environmental exposures, and are reported to be altered in ALL, suggesting a potential mediating mechanism between environment and ALL disease risk. To investigate this, we used a 'meet in the middle' approach, investigating the overlap between exposure-associated and disease-associated methylation change. Genome-wide DNA methylation changes in response to possible ALL-risk exposures (i.e. breast feeding, infection history, day care attendance, maternal smoking, alcohol, caffeine, folic acid, iron, and radiation exposure) were investigated in a sub-population of the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort using an epigenome-wide association study (EWAS) approach (n=861-927), and compared to a list of ALL disease-associated methylation changes compiled from published data. Hypergeometric probability tests suggested that the number of directionally concordant gene methylation changes observed in ALL disease and in response to the following exposures; maternal radiation exposure (p=0.001), alcohol intake (p=0.006); sugary caffeinated drink intake during pregnancy (p=0.045); and infant day care attendance (p=0.003), were not due to chance. Data presented suggests that DNA methylation may be one mediating mechanism in the multiple hit pathway needed for ALL disease manifestation.

## Introduction

Acute lymphoblastic leukaemia (ALL) is the most common form of childhood cancer [1]. Although survival rates have improved, with 5-year survival rate increasing from 60-90% [2], overall incidence of childhood ALL has been slowly increasing [3]. Genetic aberrations such as chromosomal translocations and hyperdiploidy have been suggested to play a role in ALL disease development, yet these events alone are not sufficient for disease manifestation. For example, a translocation resulting in the *TEL-AML1* fusion gene associated with B-cell lineage ALL [4] has been found to be present in roughly 1% of newborns at birth [5], which is approximately 100-fold higher incidence rate than that of childhood B-cell lineage ALL. The peak incidence of childhood ALL, between 2-5 years of age, suggests that *in utero* and early life events, including environmental factors, may contribute to disease risk. Blood chimerism in twins, occurring via the transfer of hematopoietic stem cells between the foetuses, results in concordant twin cases of leukaemia [6]; these have been identified as having the same fusion-gene sequence or *IGH/TCR* sequences, supporting the notion of ALL originating *in utero*. Concordance in identical twins is estimated to be 10%, yet 70% of these

share the same placenta, suggesting secondary events are required, and have a major impact on the likelihood of leukaemia development [7]. Furthermore, a recent study observed spatial clustering of ALL cases, using addresses of mothers during pregnancy, suggesting that environmental factors may play a role in disease development [8]. Various *in utero* and early childhood exposures have been investigated in the aetiology of childhood ALL [9, 10]. These include breast feeding [11], infection history [12], childcare/day care attendance [11, 12], maternal smoking [13, 14], alcohol [15, 16], caffeine [15, 17], folic acid [18, 19], iron [20], radiation [8], household paints [21], chemicals [22], pesticides [23, 24], and herbicides [25]. The weight of supporting evidence for the role of each exposure in the aetiology of childhood ALL varies. The evidence for exposure to pesticides and paints demonstrate consistently positive associations with increased risk of childhood leukaemia (including data from a recent meta-analysis for paint) [21]. Evidence for a role of maternal caffeine and folate intake is weaker, but emerging, with recent meta-analysis suggesting high caffeine intake may be a risk factor [17], with folate supplementation suggested to be protective [26]. Despite some evidence from earlier studies suggesting maternal smoking and alcohol intake may influence risk, more recent meta-analysis found no association between these exposures and childhood leukaemia risk [27]. Studies investigating iron intake during pregnancy and leukaemia risk have been small scale and therefore are limited in their ability to provide robust evidence [28-30]. Likewise, the number of studies investigating maternal and infant radiation exposure is limited, and thus despite associations being found between maternal and infant radiation exposure and childhood cancer risk [31, 32], further investigations are warranted. A number of studies, including a recent meta-analysis [11], suggest that breast feeding and exposure to infection in early life may be protective against leukaemia development.

Whilst the strength of the associations for some of these environmental factors and ALL risk is weak and evidence limited, this may be due to lack of available environmental data, limited number of cases for this rare disease (i.e. studies underpowered), and inaccuracies associated with collecting environmental exposure data retrospectively. The investigation of intermediate biomarkers linking an exposure and the disease outcome may therefore help to strengthen the evidence for the association of a given exposure and disease outcome, but also may offer novel insights into the underlying causal pathway to disease.

Epigenetics play a key role in human health and disease [33-35]. Epigenetic modifications such as DNA methylation, the most extensively studied epigenetic modification [36], and alterations to chromatic structure are able to modulate gene expression [37]. Methylation of cytosine residues located within CpG dinucleotides across the genome play a crucial role in cellular identity by inhibiting the binding ability of transcriptomic machinery to DNA, thus influencing gene expression

[38, 39]. Each individual's epigenome is subject to change induced via genetic, environmental, and stochastic factors throughout the life course [40]. However the epigenome is particularly susceptible to dysregulation via environmental influences during gestation, neonatal development, puberty, and old age [41]. The establishment of DNA methylation patterns occurs during embryonic development, therefore this may be an acutely critical and vulnerable period during which environmental influences are of importance for resultant DNA methylation profiles, and possible health consequences in later life [42].

Aberrant DNA methylation in ALL is well established [39, 43-45], particularly at promoter associated CpG islands, which results in hundreds of genes involved in critical molecular pathways being silenced [46]. ALL is a heterogeneous disease and is subtyped based on its cellular phenotype and cytogenetic aberrations. A core set of CpG sites have been identified as aberrantly methylated throughout all subtypes, as well as subtype specific DNA methylation patterns [39, 47].

As DNA methylation can be altered by environmental exposures, and is known to be altered in ALL itself, it is plausible that DNA methylation may act as one potential mediating mechanism via which *in utero* and early life exposures are able to increase the risk of developing ALL. We therefore hypothesised that if DNA methylation were a mediating mechanism between environmental risk factors and ALL, there would be common methylation changes associated with both a given exposure and ALL. To investigate this hypothesis we employed a 'meet in the middle' approach [48]. Firstly, an epigenome-wide association study (EWAS) was carried out to identify associations between individual CpG site methylation and ALL risk exposures [49]. Exposure-associated CpG sites were mapped to genes and compared with previously identified genes from the literature which have reported altered DNA methylation in ALL compared to healthy individuals. Hypergeometric testing was used to assess the probability that the resultant overlap between exposure-associated and disease-associated methylation was not due to chance. The resultant data suggests that DNA methylation may play a mediating role in the causal pathway linking ALL risk exposures with disease risk.

## **Materials and Methods**

### **Study population**

The ALSPAC cohort is a prospective birth cohort comprising 14541 pregnancies recruited from the Bristol area in England with expected delivery dates between 1st April 1991 and 31st December 1992 [50, 51]. A large amount of data has been collected including genetic, epigenetic, biological, social and other environmental exposures in relation to health and developmental outcomes. Blood

samples were collected from all consenting mothers (during pregnancy and follow up clinic) and their children (birth; childhood, age seven years; and adolescence, age 17 years) at several clinics. The study website contains details of all the data that is available through a fully searchable data dictionary (<http://www.bristol.ac.uk/alspac/researchers/data-access/data-dictionary/>). The study was observational and aimed to intervene as little as possible in the normal course of pregnancy and childhood [52]. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

The Accessible Resource for Integrated Epigenomic Studies (ARIES) is a sub-population of the ALSPAC cohort [53]. ARIES comprises of 1018 mother child pairs who were selected based on the availability of matched-samples at multiple time points required for the study. Genome-wide DNA methylation analysis was carried out on blood samples (cord/peripheral) from these individuals at birth and age seven.

### **Identification of environmental factors associated with an increased risk of ALL**

Literature searches were previously carried out to identify exposures associated with ALL risk and their potential to influence changes in DNA methylation using the PubMed and ScienceDirect databases (1<sup>st</sup> January 1987- 31<sup>st</sup> March 2018). This was carried out in three stages, firstly, reviews discussing ALL aetiology were screened for environmental exposures potentially associated with increased ALL risk. From this initial screen a list of key words were compiled (Box 1). Secondly, a more rigorous literature search was carried out using the key words to identify evidence for an association between environmental exposures and ALL risk. Thirdly, the identified potential ALL risk exposures were used to further investigate the literature to find any relationships between the identified exposures and changes in DNA methylation patterns (Box 1). References from these papers, including the reviews were also screened. No language restrictions were imposed. To be as inclusive as possible in our investigations, we examined all potential risk factors for which there has been at least one study providing evidence (even if limited) of an association between an environmental factor and ALL risk.

Box 1. Literature searches carried out, and key words used to identify ALL risk exposures and potential exposure associated changes in DNA methylation.

Literature search	Key words
1. Identification of ALL risk exposures	Leukemia, acute lymphoblastic, leukaemia
2. Identification of evidence supporting an association between ALL and an exposure	Leukemia, acute lymphoblastic, leukaemia, radiation, smoking, alcohol, folate, folic acid,

	iron, coffee, caffeine, herbicides, pesticides, household chemicals, chemicals, household paints, paints, child care, day care, breast feeding, birth weight, infection history, infection, virus, and bacterial
3. Identification of relationships between ALL risk exposures and changes in DNA methylation	DNA methylation, and exposures identified in literature search 2.

### Identification of genes with altered methylation in ALL disease

Methylation changes in ALL were identified from previously published literature [39, 54], the significant CpG sites identified were mapped to genes (see [39, 54] for details). Studies which utilised the Illumina Infinium® HumanMethylation450k BeadChip assay (450K BeadChip) to measure DNA methylation, were used as they directly correlated with the measurements carried out in the ALSPAC cohort. This resulted in a total of 2173 genes which had reported significant differential methylation in ALL disease compared to healthy controls (Table S1).

Additionally, for comparison, a list of gene loci with altered methylation in ALL disease from all large published studies [43-45, 55-58], i.e. with no exclusions based on methodologies used to measure methylation was also compiled and assessed in relation to variation in methylation observed in relation to environmental factors in the ALSPAC cohort.

### Environmental exposure data available in the ALSPAC cohort

The environmental data used in this study was selected from nine questionnaires carried out as part of the ALSPAC study, and these ranged from early gestation, throughout pregnancy, and throughout the life of the offspring (still ongoing) (Figure 1). The impact of *in utero* environmental exposures including maternal alcohol intake, smoking, radiation exposure, folic acid supplementation, coffee consumption, sugary caffeinated drink consumption and iron supplementation during pregnancy on DNA methylation status in offspring cord blood were assessed (Table S2 for cohort characteristics and details of measure of exposure variables). No data were available for pesticide exposure, so this exposure was not assessed in this analysis. The influence of early life post-natal exposures i.e. day care attendance at 8, 38, and 54 months, breast feeding, and offspring reported illness by the mothers on DNA methylation in the blood at age seven were also assessed (Table S3). The data used in this study was selected to best represent each exposure under investigation (where data were available), and in line with previously published data (see Table S4 for summary of evidence and

indication of risk level, and Table S5 for justification of variable selection). An additional search was carried out to identify larger methylation studies (450K BeadChip) which identified exposure associated methylation changes (FDR corrected and cell-type adjusted).

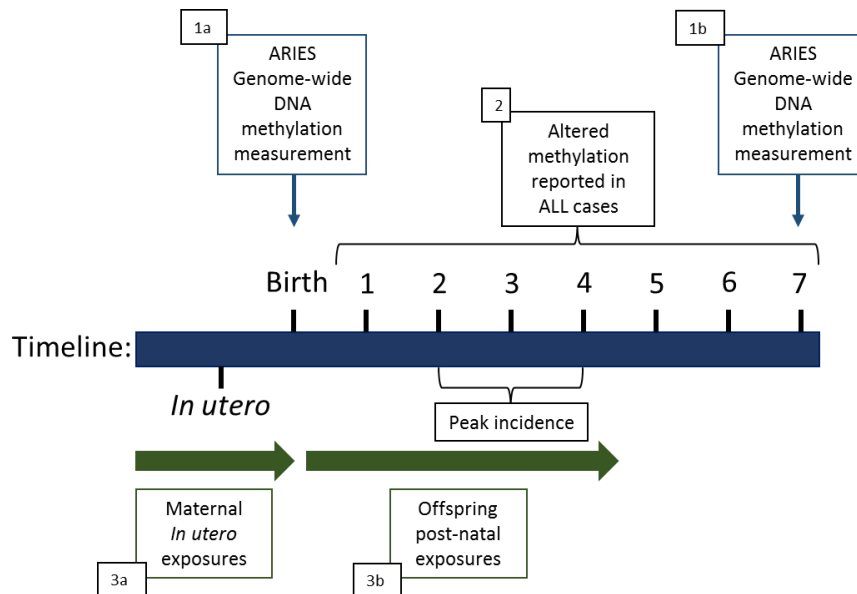


Figure 1. Study design: 1a) 450K BeadChip was used to measure DNA methylation in cord blood at birth ( $n = 861$ ); and 1b) in blood age 7 ( $n = 927$ ); 2) altered DNA methylation in ALL disease identified in previous studies [39, 54]; 3a) questionnaires carried out for the ALSPAC study were used to identify *in utero* exposures (8-28 weeks gestation) previously associated with ALL risk (Table S2); and 3b) post-natal exposures previously associated with ALL risk (8-58 months of age) (Table S3).

### The Illumina Infinium® HumanMethylation450k BeadChip assay, data pre-processing and normalisation

Genome-wide DNA methylation data was previously generated for the ARIES sub-population using the 450K BeadChip. The arrays were scanned using an Illumina iScan and reviewed using GenomeStudio (version 2011.1). To minimise the possibility of confounding by batch effects, samples were semi-randomly arranged on the chip, making sure samples from each time point were represented on each chip. Data were pre-processed using R (version 3.0.1) at the MRC Integrative Epidemiology Unit, University of Bristol [53]. The Watermelon package [59] was used according to the subset quantile normalisation approach described by Touleimat & Tost [60] to reduce non-biological differences between probes. Data were normalised using functional normalization [61], resulting in a total of 861 cord blood samples and 927 age seven samples for analyses.

In the cord blood, cell-type proportion estimation was used to adjust for cell type effects. Publicly available expression signature data were used, this data can be found in the NCBI Gene Expression



Omnibus repository, GSE68456 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68456>) [62]. For the age seven analyses the estimateCellCounts function in the minfi R package [63] was used, this is based on the model developed by Houseman *et al* [64]. The estimated cell-type proportion was added to the regression model to adjust for the potential effect of this variable.

### **Multiple linear regression analyses of DNA methylation at individual CpG sites and environmental factors**

Multiple linear regression analysis was used to investigate associations between DNA methylation at individual CpG sites and environmental exposures of interest. For each exposure, the resulting p-values were corrected for multiple testing by controlling the false discovery rate (FDR) at 5% ( $p \leq 0.05$ ), first described by Benjamini and Hochberg [65]. DNA methylation was modelled as a continuous variable (outcome), in a multivariate regression model accounting for potential confounders (sex, parity, gestation, cell type, and batch). These were selected as relevant covariates, as well as being in line with previous multiple linear regression models [66].

FDb.InfiniumMethylation.hg19 R package (FDb.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes) was used to annotate the methylation data by mapping CpG sites to genes for further analyses. If CpG sites were mapped to multiple genes, both genes were included in the analyses, and unmappable CpGs were excluded.

### **Integration of lists of gene loci with altered methylation due to an ALL risk exposure and a list of gene loci with altered methylation in ALL**

Molbiotools (<http://www.molbiotools.com/listcompare.html>) was used to integrate the list of gene loci with altered methylation in ALL, with the lists of gene loci with significantly altered methylation (FDR-adjusted  $p \leq 0.05$ ) for each ALL risk exposure analysed using the gene symbol for comparison. Molbiotools was also used to identify gene loci which overlapped between multiple exposures, as well as gene loci which were altered in ALL and multiple exposures. Hypergeometric tests were carried out using the GeneProf calculator (<http://www.geneprof.org>) to assess the probability that the observed overlapping changes in methylation found in relation to a given exposure and in ALL were likely to be significant ( $p \leq 0.05$ ) and not due to a chance observation. An  $n$  of 21231 (i.e. the number of genes represented on the array) was used as the 'population size', with 2173 (i.e. number of genes with altered methylation in ALL) as the 'number of successes in the population'.  $N$  used for 'sample sizes' and 'number of successes in sample' were the number of genes found to have altered methylation in response to a given exposure and the number of genes found to overlap between methylation change associated with a given exposure and in ALL respectively. Where larger data sets with exposure-associated methylation changes (FDR corrected and cell type adjusted) and disease-

associated methylation changes were available the same integration process discussed above was also carried out.

### **Gene Ontology enrichment and pathway analysis**

The Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 Beta (<https://david.ncicrf.gov/>) [67] was used to carry out Gene Ontology (GO) and Kyoto Encyclopaedia of genes and genomes (KEGG) pathway analysis to investigate the biological pathways and processes that may be affected through DNA methylation changes associated with individual *in utero* and post-natal exposures (FDR-adjusted  $p \leq 0.05$ ). GO and pathway analysis were carried out for each individual exposure using the resultant list of significant genes found to have altered methylation in relation to the given exposure. Furthermore, for those exposures where genes were also found to have altered methylation in ALL [39, 54], this analysis was also carried out using the list of overlapping genes.

### **Results**

#### **Epigenome-wide association study investigating the influence of ALL risk exposures on DNA methylation**

For the majority of *in utero* exposures examined, methylation was found to be significantly altered across a range of gene loci in cord blood at birth (FDR-adjusted  $p \leq 0.05$ ) (Table 1), with most changes observed to be associated with maternal radiation exposure (n CpGs=288), alcohol intake (n CpGs=192), and consumption of sugary caffeinated drinks (n CpGs=66) during pregnancy. Only maternal iron intake whilst pregnant had no significantly associated methylation changes. The split between direction of methylation change in response to radiation and smoking (throughout pregnancy, at 3 months, and pre-pregnancy) exposures were roughly equal (i.e. 51% and 55%, 53%, 43% hypo- and 49% and 45%, 47%, 57% hypermethylated CpGs respectively). Whilst alcohol intake, sugary caffeinated drinks, and coffee consumption had larger proportions of hypermethylated compared to hypomethylated CpGs (82%, 73% and 60% respectively) (Table 1 and Table S6). Only folic acid supplementation resulted in a higher percentage (89%) of hypomethylated CpGs (Table 1 and Table S6). Gene ontology and pathway analysis carried out for each of the gene lists generated for individual *in utero* exposures only identified one potentially significantly affected pathway ('Amoebiasis') in response to the methylation changes which occurred in response to radiation exposure (Table 2). No other biological processes or pathways were suggested to be affected by methylation changes associated with other *in utero* exposures. Two additional studies were identified with larger data sets for maternal smoking (n CpGs=6073 and genes=3176) [68] and maternal plasma folate (n CpGs=443 and genes=229) [69] associated methylation changes. Both

maternal smoking and maternal plasma folate studies had a significant overlap of gene loci with exposure associated- and ALL-associated methylation changes ( $4.95 \times 10^{-11}$ ,  $4.32 \times 10^{-13}$ , respectively) (Table 3).

At age seven, post-natal exposures used as proxy measures for infection i.e. day care attendance (8 months of age) and maternally reported cold (6 months of age) were found to be significantly associated with altered methylation of 11 and 60 gene loci respectively (Table 1). Post-natal exposure to day care attendance at 38 months and 54 months, and breast feeding had no significantly associated methylation changes. Whilst the direction of methylation change in response to day care attendance was approximately equal i.e. 45% hypo- vs 55% hypermethylated CpGs; the majority of CpGs with altered methylation in response to having cold symptoms were hypomethylated (73%) (Table 1 and Table S6). GO and pathway analysis revealed that methylation changes associated with recorded cold exposure may significantly affect the MAPK signalling pathway (Table 4), but changes associated with day care attendance and breast feeding were not found to be associated with any biological processes or pathways.

#### **DNA methylation changes in ALL and integration with potential ALL risk exposure associated changes in methylation**

Hypergeometric probability tests were used to ascertain if the degree of overlap between methylation changes observed in ALL and those associated with a potential risk exposure were statistically significant or likely to be due to chance. For over half of the exposures investigated (Table 1) there was a significant overlap between genes with both exposure-associated and disease-associated methylation change.

The concordance of the change in direction of methylation influenced by exposures and in ALL was not uniform across exposures. Whilst concordance was high (i.e. over 70%) for radiation exposure, alcohol intake, sugary caffeinated drinks consumption and day care attendance, for all other exposures concordance rates were intermediate (smoking and folate supplementation) or even low (reported cold symptoms and coffee consumption) (Table 1). We therefore carried out further hypergeometric tests to ascertain if the overlap between concordant methylation changes observed in ALL and those associated with a potential risk exposure were a statistically significant finding or likely to be due to chance (Table S7). These analyses suggested there was a statistically significant overlap between genes with methylation change in ALL and those found to have the same direction of methylation change in response to *in utero* exposure to radiation exposure, alcohol intake, sugary caffeinated drinks consumption, and to day care attendance at 8 months (Table 1). An unexpected disparity between findings for maternal coffee and sugary caffeinated drinks consumption have

been reported in this study. This may allude to something other than caffeine driving the larger number of concordant changes identified for sugary caffeinated drinks consumption. The final column in Table 1 shows the overlap between genes with exposure-associated changes in methylation and changes in methylation in ALL disease from all published studies with no restrictions on the methods used to measure methylation. There is a lack of overlap (except for radiation exposure associated gene loci) compared to that observed in studies only using the 450K BeadChip array. The disparities between studies are likely to be partially or largely driven by the differences in technologies used to measure DNA methylation [43-45, 55-58], making the comparisons between data sets difficult, therefore this additional analysis should be interpreted with care.

Comparison of the data sets found that 7 gene loci (*KCNQ1DN*, *MAD1L1*, *ROBO3*, *SEPT9*, *SNX8*, *TBX2* and *ZC3H3*) had altered methylation in ALL and in response to more than one potential ALL risk exposure, although the direction of methylation change was only found to be conserved between ALL and exposures for one gene (*TBX2*). Pathway analysis did not identify any significantly affected pathways when using the lists of genes which had both disease and exposure-associated altered methylation, most likely due to an inadequate number of genes overlapping between each exposure and ALL disease for this type of analysis or limitation of relying on gene mapping approaches.

## **Discussion**

Due to the rarity of childhood ALL it is very difficult to investigate causal pathways in disease aetiology. The early onset of disease, with peak incidence between 2-5 years of age, *in utero* and early life periods are likely to be crucial for disease development. Therefore, to improve understanding of this complex disease pathway, better knowledge of the differences between individuals who possess/experience a first 'hit' (e.g. ALL-associated genetic translocation) and develop disease, compared to those that do, yet do not develop ALL will be key. However, as such genetic abnormalities are rare in the general population (1% for the *TEL-AML1* fusion gene [5]), and fresh whole blood samples are required to measure such abnormalities, to the best of our knowledge, there is currently no single cohort possessing the numbers, data and samples required to investigate ALL disease development in this manner. We therefore applied a 'meet in the middle' approach [48] to investigate one plausible mechanism in the disease pathway to link possible risk factors to ALL cases i.e. DNA methylation changes.

We hypothesised that DNA methylation may act as a potential mediating mechanism between environmental exposures and childhood ALL risk. The majority of ALL risk exposures investigated in this study were shown to influence DNA methylation at multiple gene loci, with some degree of

overlap found with disease-associated methylation changes. In demonstrating that methylation changes associated with these potential risk exposures occur in genes which also have altered methylation in overt disease, we provide evidence for a mediating role of DNA methylation in the causal disease pathway. Hypergeometric tests suggest that for over half of the exposures investigated (i.e. *in utero* radiation exposure, maternal alcohol intake, maternal sugary caffeinated drinks intake, maternal smoking and infant day care attendance at 8 month), these concomitant gene methylation changes were not likely to be due to chance. Moreover, 4/5 of those exposures were found to have significant concordant changes in direction of methylation associated with exposure and ALL disease (with the exception of maternal smoking). In addition, we also compared our ALL-disease associated methylation signature with two previously published studies that investigated variation in methylation in association with maternal smoking and plasma folate [68, 69] (Table 3). This analysis provided additional evidence demonstrating that maternal smoking and plasma folate were associated with a large number of methylation changes, and that a significant number of these changes were also found in ALL disease. We would therefore suggest that for radiation exposure, maternal alcohol intake, maternal sugary caffeinated drinks intake, smoking, and folic acid intake, it is plausible that methylation changes which occur as a result of these exposures may contribute to the aetiology of ALL.

Whilst, in this context significant associations were also observed for day care attendance, the causal pathway towards disease may be less clear. It has been suggested that infections during early life (<1yr) 'prime' the immune system, enabling it to elicit an appropriate immune response later in childhood and adult life [70]. Without this 'priming', it is thought that exposure to infection later in life is likely to lead to increased proliferation and error which may lead to the development of leukaemia in individuals harbouring genetic aberrations. Greaves [71] suggested a two-hit model for role of infections, specifically in B cell precursor ALL. The proposed mechanism is multifactorial, involving inherited genetics, modulators of risk (i.e. diet), and exposure to infections in early-life. He further proposes that developed societies restriction of early-life microbial exposure has led to an increase in cases, and may therefore, in-part, be preventable. As such, day care attendance has been utilised as a proxy measure for infection, with inverse associations reported between attendance and ALL risk [12]. If day care attendance is associated with lower risk of ALL, we would hypothesise that DNA methylation changes associated with day care attendance would be protective, and therefore show the opposite direction of methylation change to that observed in disease. However, here we observed concordant changes across 5/6 genes (i.e. day care related methylation changes were associated with ALL risk not protection). Whilst the reason for this perhaps unexpected outcome is unclear, there may be several rational explanations for this finding. In this study

recording of day care attendance was taken at ~8 months. Therefore individuals who attended day care thereafter would have been identified as non-attenders. There are therefore likely to have been a reasonable proportion of infants who attended day care between the ages of 8 month and 1 year i.e. after data collection and prior to usual state nursery attendance schedule, which may have influenced DNA methylation patterns of those individual, and may have skewed the outcome. There is also the possibility that other/additional exposures modify the effect of day care exposure, for example breast feeding has been shown to modify the effect of day care attendance on risk of type 1 diabetes, where day care attendance was only associated with a decreased risk with increasing breast feeding duration, and increased risk with non-breastfed children [72]. Furthermore, the use of a proxy measure of infection, may mean that the observed differences in DNA methylation may be due to other confounding factors, rather than infection itself i.e. children who attend day care may belong to different socio-demographic groups than those who do not, and also have a higher degree of social interaction [73]. A selective increase in ALL incidence in children 1-4 years of age has been shown during socioeconomic transition [74]. The authors further suggested that difference in socioeconomic status (SES) may lead to an increase in a genetically distinct subtype of ALL susceptibility, suggesting hyperdiploid ALL as a candidate. Additional studies have proposed that DNA methylation may act as a modifiable mechanism by which factors such as SES become physically embodied [75], resulting in altered disease risk between SES groups. A spurious result must also be considered, given that for day care there were only a small number of individuals exposed i.e. only 46/871 were reported to attended day care at 8 month of age. This may also be true for maternal radiation exposure were only 29/849 were reported to be exposed. Whilst we concede that it is possible that these observations may be due to chance, we would argue that the perinatal window represents a period of a high degree of plasticity for the developing epigenome compared to later time points during childhood and therefore it is plausible that some of the observations reported here may be real due to the timing of the exposure. In particular, the magnitude and significance of change identified for at least two of the six genes whose methylation were influenced by day care attendance (and are altered in ALL) (TMEM132C, -7% ( $p=0.003$ ); and GLB1L, -2% ( $p=0.03$ )), implies a true association with this exposure, and potential candidates for future studies. Indeed, when using reported cold symptoms in the first 6 months of life as a measure of infection, which perhaps may be a less biased, although imperfect measure, we observed a larger number of methylation changes (i.e. 75 CpGs as opposed to just 11 for day care attendance). Despite a lack of statistically significant overlap between ALL-associated genes, there was a lower degree of concordant changes (33%), whereby 6 genes displayed hypomethylation in response to cold

symptoms and hypermethylation in ALL, supporting the notion that early immune stimulation has a protective effect against ALL.

There are several plausible explanations as to why we did not observe a significant overlap between disease associated methylation change and exposure associated methylation changes for all exposures assessed. The simplest explanation, that DNA methylation may not be a key mechanism linking those particular exposures (i.e. folate supplementation, coffee consumption, and reported cold symptoms) and ALL risk. Whilst this is possible, we suggest alternative explanations for these findings. In this study we applied the 'meet in the middle' approach, due to rarity and difficulty obtaining appropriate data and samples within one cohort. The integration and overlap of the data sets analysed therefore originated from a combination of prospective and retrospective study designs. Moreover, in using data from different studies, the exposed and disease populations are likely to be different, which may have influenced study outcomes. The sample size may not have been sufficient to detect small exposure-associated methylation changes, thus a larger cohort may be needed to detect these small effects, for instance when investigating these associations in larger cohorts [54, 55] (Table 3) significant associations were found.

Here, ALL-associated exposures were measured via questionnaires completed by mothers of the offspring from whom DNA methylation data were available. This method of exposure measurement is not the most accurate due to reporting bias [76], and therefore it is plausible that we were unable to detect some methylation changes associated with some of these exposures due to the inaccuracy of the exposure data. All of the questions had categorised answers and mothers had to select the most appropriate, and no extra/unique data were collected regarding these exposures. In particular, the measure used to assess folate status was based on use of supplements. The utility of this measure can therefore easily be questioned, as dietary intakes of folate are likely to vary widely between individuals. Therefore, there is likely to be a high degree of variation between individuals both taking and not taking supplements, which would not have been captured using this method of assessment, and thus influence our outcome measure. In addition, the use of cold reporting in the first 6 months, whilst an objective measure, does not take into account other infections that may have been encountered during early life and consequently influence methylation patterns. Therefore, future studies using biomarkers or medical records to measure exposures may help improve accuracy of detecting methylation changes in response to some environmental exposures.

Also, whilst the use of hypergeometric tests gives a statistically based rationale to suggest that a higher degree of overlap between ALL and exposure associated genes is not due to a chance finding and therefore is likely to be a significant finding, it also could be argued that a biological rationale

may be equally likely, whereby key methylation changes which result in a functional biological outcome are more important than a plethora of changes. For example, whilst the overlap for concordant genes for both smoking exposure and ALL was not statistically significant, it is plausible that hypermethylation of the *CYP1A1* gene caused by smoking during or pre-pregnancy (and observed in ALL) has a biologically functional outcome (detoxification of compounds from tobacco smoke), which may be in the causal pathway towards disease. The protein encoded by the *SEPT9* gene is involved in cytokinesis, and is crucial for the separation of daughter cells via mediation of midbody abscission [77], it has previously been implicated in cancers, including ALL, as well as potentially contributing to drug resistance [78]. Hypomethylation, as found in ALL and associated with maternal alcohol consumption in this study, could lead to over expression and enhanced cell survival. The *TBX2* gene encodes a transcriptional repressor involved in the determination/maintenance of cell fate [79], found negatively dysregulated in B-cell acute lymphoblastic leukaemia [80], as well as in this study shown via hypomethylation in ALL, associated with maternal alcohol intake, and sugary caffeinated drink consumption. Using this rationale, it is possible that any or all of the concordant changes in methylation we have observed between risk factors and ALL, may be implicated in disease progression if functionally relevant in the causal pathway towards ALL manifestation.

Due to the difficulties in being able to obtain data and biological samples from childhood ALL cases prior to diagnosis there has been limited research investigating the potential mechanistic implications of risk exposures on disease development and progression. Here we provide evidence to suggest that a range of exposures which have been investigated in the context of ALL risk are associated with epigenetic variation. Further evaluation of this variation is required using more reliable exposure measurements, through replication in additional cohorts, and implementing causal analysis methods. International collaborations will also be key in generating the numbers required to capture sufficient ALL cases for a meaningful analysis. Furthermore, the analyses carried out here provides supporting evidence for the ability of environmental exposures to drive disease-associated changes in DNA methylation. If these changes occur pre-disease (possibly via multiple exposures), and in combination with other potential ALL-associated risk factors or 'hits', they could explain one contributing mechanistic pathway driving ALL disease incidence. Further work aiming to assess methylation patterns in ALL cases prior to disease diagnosis is warranted to provide additional evidence for a role of DNA methylation in the causal pathway towards disease, and with a view to develop predictive biomarkers of disease for high risk groups, and to strengthen the causal association between exposures and ALL risk to inform public health policy.



## Acknowledgements and author contributions

**Concept and design:** Jessica A Timms, Caroline L Relton, Gemma C Sharp, Judith Rankin, Gordon Strathdee, Jill A McKay

**Development of methodology:** Jessica A Timms, Jill A McKay, Gemma C Sharp

**Acquisition of data:** Jessica A Timms, Caroline L Relton, Gemma C Sharp, Jill A McKay

**Analysis and interpretation of data:** Jessica A Timms, Gordon Strathdee, Jill A McKay

**Writing, review, and/or revision of the manuscript:** Jessica A Timms, Caroline L Relton, Judith Rankin, Gordon Strathdee, Jill A McKay

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

## Conflict of interest

The authors have no other relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with subject matter or materials discussed in the manuscript apart from those disclosed.

## References

1. Siegel, D.A., S.J. Henley, J. Li, L.A. Pollack, E.A. Van Dyne, and A. White, *Rates and Trends of Pediatric Acute Lymphoblastic Leukemia - United States, 2001-2014*. MMWR Morb Mortal Wkly Rep, 2017. **66**(36): p. 950-954.
2. Board, P.D.Q.P.T.E., *Childhood Acute Lymphoblastic Leukemia Treatment (PDQ(R)): Health Professional Version*, in *PDQ Cancer Information Summaries*. 2002-2017, National Cancer Institute (US): Bethesda (MD).
3. Linet, M.S., L.M. Brown, S.M. Mbulaiteye, D. Check, E. Ostroumova, A. Landgren, and S.S. Devesa, *International long-term trends and recent patterns in the incidence of leukemias and lymphomas among children and adolescents ages 0-19 years*. Int J Cancer, 2016. **138**(8): p. 1862-74.
4. Alm, S.J., C. Engvall, J. Asp, L. Palmqvist, J. Abrahamsson, and L. Fogelstrand, *Minimal residual disease monitoring in childhood B lymphoblastic leukemia with t(12;21)(p13;q22); ETV6-RUNX1: concordant results using quantitation of fusion transcript and flow cytometry*. Int J Lab Hematol, 2017. **39**(2): p. 121-128.
5. Pui, C.-H., L.L. Robison, and A.T. Look, *Acute lymphoblastic leukaemia*. The Lancet, 2008. **371**(9617): p. 1030-1043.
6. Greaves, M.F. and J. Wiemels, *Origins of chromosome translocations in childhood leukaemia*. Nat Rev Cancer, 2003. **3**(9): p. 639-49.

7. Greaves, M.F., A.T. Maia, J.L. Wiemels, and A.M. Ford, *Leukemia in twins: lessons in natural history*. Blood, 2003. **102**(7): p. 2321-33.
8. Carceles-Alvarez, A., J.A. Ortega-Garcia, F.A. Lopez-Hernandez, M. Orozco-Llamas, B. Espinosa-Lopez, E. Tobarra-Sanchez, and L. Alvarez, *Spatial clustering of childhood leukaemia with the integration of the Paediatric Environmental History*. Environ Res, 2017. **156**: p. 605-612.
9. Wiemels, J., *Perspectives on the causes of childhood leukemia*. Chemico-Biological Interactions, 2012. **196**(3): p. 59-67.
10. Whitehead, T.P., C. Metayer, J.L. Wiemels, A.W. Singer, and M.D. Miller, *Childhood Leukemia and Primary Prevention*. Curr Probl Pediatr Adolesc Health Care, 2016. **46**(10): p. 317-352.
11. Rudant, J., T. Lightfoot, K.Y. Urayama, E. Petridou, J.D. Dockerty, C. Magnani, E. Milne, L.G. Spector, L.J. Ashton, N. Dessypris, A.Y. Kang, M. Miller, R. Rondelli, J. Simpson, E. Stiakaki, L. Orsi, E. Roman, C. Metayer, C. Infante-Rivard, and J. Clavel, *Childhood acute lymphoblastic leukemia and indicators of early immune stimulation: a Childhood Leukemia International Consortium study*. Am J Epidemiol, 2015. **181**(8): p. 549-62.
12. Ajrouche, R., J. Rudant, L. Orsi, A. Petit, A. Baruchel, A. Lambilliotte, M. Gambart, G. Michel, Y. Bertrand, S. Ducassou, V. Gandemer, C. Paillard, L. Saumet, N. Blin, D. Hemon, and J. Clavel, *Childhood acute lymphoblastic leukaemia and indicators of early immune stimulation: the Estelle study (SFCE)*. Br J Cancer, 2015. **112**(6): p. 1017-26.
13. de Smith, A.J., M. Kaur, S. Gonseth, A. Endicott, S. Selvin, L. Zhang, R. Roy, X. Shao, H.M. Hansen, A.Y. Kang, K.M. Walsh, G.V. Dahl, R. McKean-Cowdin, C. Metayer, and J.L. Wiemels, *Correlates of Prenatal and Early-Life Tobacco Smoke Exposure and Frequency of Common Gene Deletions in Childhood Acute Lymphoblastic Leukemia*. Cancer Research, 2017. **77**(7): p. 1674.
14. Rumrich, I.K., M. Viluksela, K. Vahakangas, M. Gissler, H.M. Surcel, and O. Hanninen, *Maternal Smoking and the Risk of Cancer in Early Life - A Meta-Analysis*. PLoS One, 2016. **11**(11): p. e0165040.
15. Menegaux, F., C. Steffen, S. Bellec, A. Baruchel, B. Lescoeur, G. Leverger, B. Nelken, N. Philippe, D. Sommelet, D. Hemon, and J. Clavel, *Maternal coffee and alcohol consumption during pregnancy, parental smoking and risk of childhood acute leukaemia*. Cancer Detect Prev, 2005. **29**(6): p. 487-93.
16. Karalexi, M.A., N. Dessypris, T.P. Thomopoulos, E. Ntouvelis, M. Kantzanou, A.A. Diamantaras, M. Moschovi, M. Baka, E. Hatzipantelis, M. Kourti, S. Polychronopoulou, E. Stiakaki, A.M. Mora, V. Wunsch-Filho, C. Infante-Rivard, D. Loutradis, and E.T. Petridou, *Parental alcohol consumption and risk of leukemia in the offspring: a systematic review and meta-analysis*. Eur J Cancer Prev, 2017. **26**(5): p. 433-441.
17. Thomopoulos, T.P., E. Ntouvelis, A.A. Diamantaras, M. Tzanoudaki, M. Baka, E. Hatzipantelis, M. Kourti, S. Polychronopoulou, V. Sidi, E. Stiakaki, M. Moschovi, M. Kantzanou, and E.T. Petridou, *Maternal and childhood consumption of coffee, tea and cola beverages in association with childhood leukemia: a meta-analysis*. Cancer Epidemiol, 2015. **39**(6): p. 1047-59.
18. Kałużna, E.M., E. Strauss, B. Świątek-Kościelna, O. Zając-Spychała, E. Gowin, J.S. Nowak, J. Rembowska, and D. Januszkiewicz-Lewandowska, *The methylenetetrahydrofolate reductase 677T-1298C haplotype is a risk factor for acute lymphoblastic leukemia in children*. Medicine, 2017. **96**(51): p. e9290.
19. Dessypris, N., M.A. Karalexi, E. Ntouvelis, A.A. Diamantaras, V. Papadakis, M. Baka, E. Hatzipantelis, M. Kourti, M. Moschovi, S. Polychronopoulou, V. Sidi, E. Stiakaki, and E.T. Petridou, *Association of maternal and index child's diet with subsequent leukemia risk: A systematic review and meta analysis*. Cancer Epidemiol, 2017. **47**: p. 64-75.

20. El-Rashedi, F.H., M.A. El-Hawy, S.M. El-Hefnawy, and M.M. Mohammed, *HFE gene mutation and iron overload in Egyptian pediatric acute lymphoblastic leukemia survivors: a single-center study*. Hematology, 2017. **22**(7): p. 398-404.
21. Bailey, H.D., C. Metayer, E. Milne, E.T. Petridou, C. Infante-Rivard, L.G. Spector, J. Clavel, J.D. Dockerty, L. Zhang, B.K. Armstrong, J. Rudant, L. Fritschi, A. Amigou, E. Hatzipantelis, A.Y. Kang, E. Stiakaki, and J. Schuz, *Home paint exposures and risk of childhood acute lymphoblastic leukemia: findings from the Childhood Leukemia International Consortium*. Cancer Causes Control, 2015. **26**(9): p. 1257-70.
22. Heck, J.E., A.S. Park, J. Qiu, M. Cockburn, and B. Ritz, *Risk of leukemia in relation to exposure to ambient air toxics in pregnancy and early childhood*. Int J Hyg Environ Health, 2014. **217**(6): p. 662-8.
23. Gunier, R.B., A. Kang, S.K. Hammond, K. Reinier, C.S. Lea, J.S. Chang, M. Does, G. Scelo, J. Kirsch, V. Crouse, R. Cooper, P. Quinlan, and C. Metayer, *A task-based assessment of parental occupational exposure to pesticides and childhood acute lymphoblastic leukemia*. Environmental Research, 2017. **156**: p. 57-62.
24. Bailey, H.D., C. Infante-Rivard, C. Metayer, J. Clavel, T. Lightfoot, P. Kaatsch, E. Roman, C. Magnani, L.G. Spector, E. Th Petridou, E. Milne, J.D. Dockerty, L. Miligi, B.K. Armstrong, J. Rudant, L. Fritschi, J. Simpson, L. Zhang, R. Rondelli, M. Baka, L. Orsi, M. Moschovi, A.Y. Kang, and J. Schuz, *Home pesticide exposures and risk of childhood leukemia: Findings from the childhood leukemia international consortium*. Int J Cancer, 2015.
25. Metayer, C., J.S. Colt, P.A. Buffler, H.D. Reed, S. Selvin, V. Crouse, and M.H. Ward, *Exposure to herbicides in house dust and risk of childhood acute lymphoblastic leukemia*. J Expo Sci Environ Epidemiol, 2013. **23**(4): p. 363-70.
26. Metayer, C., E. Milne, J.D. Dockerty, J. Clavel, M.S. Pombo-de-Oliveira, C. Wesseling, L.G. Spector, J. Schuz, E. Petridou, S. Ezzat, B.K. Armstrong, J. Rudant, S. Koifman, P. Kaatsch, M. Moschovi, W.M. Rashed, S. Selvin, K. McCauley, R.J. Hung, A.Y. Kang, and C. Infante-Rivard, *Maternal supplementation with folic acid and other vitamins and risk of leukemia in offspring: a Childhood Leukemia International Consortium study*. Epidemiology, 2014. **25**(6): p. 811-22.
27. Orsi, L., J. Rudant, R. Ajrouche, G. Leverger, A. Baruchel, B. Nelken, M. Pasquet, G. Michel, Y. Bertrand, S. Ducassou, V. Gandemer, P. Lutz, L. Saumet, P. Moreau, D. Hemon, and J. Clavel, *Parental smoking, maternal alcohol, coffee and tea consumption during pregnancy, and childhood acute leukemia: the ESTELLE study*. Cancer Causes Control, 2015. **26**(7): p. 1003-17.
28. Dorak, M.T., A.K. Burnett, M. Worwood, A.M. Sproul, and B.E.S. Gibson, *The C282Y Mutation of HFE Is Another Male-Specific Risk Factor for Childhood Acute Lymphoblastic Leukemia*. Blood, 1999. **94**(11): p. 3957-3958.
29. Thompson, J.R., P.F. Gerald, M.L.N. Willoughby, and B.K. Armstrong, *Maternal folate supplementation in pregnancy and protection against acute lymphoblastic leukaemia in childhood: a case-control study*. The Lancet, 2001. **358**(9297): p. 1935-1940.
30. Kennedy, A.E., K.Y. Kamdar, P.J. Lupo, M.F. Okcu, M.E. Scheurer, M.K. Baum, and M.T. Dorak, *Examination of HFE associations with childhood leukemia risk and extension to other iron regulatory genes*. Leuk Res, 2014. **38**(9): p. 1055-60.
31. Brent, R.L., *Carcinogenic risks of prenatal ionizing radiation*. Semin Fetal Neonatal Med, 2014. **19**(3): p. 203-13.
32. Pearce, M.S., J.A. Salotti, M.P. Little, K. McHugh, C. Lee, K.P. Kim, N.L. Howe, C.M. Ronckers, P. Rajaraman, A.W. Sir Craft, L. Parker, and A. Berrington de Gonzalez, *Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study*. Lancet, 2012. **380**(9840): p. 499-505.
33. McKay, J.A. and J.C. Mathers, *Diet induced epigenetic changes and their implications for health*. Acta Physiol (Oxf), 2011. **202**(2): p. 103-18.

34. Vickers, M.H., *Early life nutrition, epigenetics and programming of later life disease*. *Nutrients*, 2014. **6**(6): p. 2165-78.
35. Bakulski, K.M. and M.D. Fallin, *Epigenetic epidemiology: promises for public health research*. *Environ Mol Mutagen*, 2014. **55**(3): p. 171-83.
36. Hattori, N. and T. Ushijima, *Compendium of aberrant DNA methylation and histone modifications in cancer*. *Biochem Biophys Res Commun*, 2014. **455**(1-2): p. 3-9.
37. Mill, J. and B.T. Heijmans, *From promises to practical strategies in epigenetic epidemiology*. *Nat Rev Genet*, 2013. **14**(8): p. 585-94.
38. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond*. *Nat Rev Genet*, 2012. **13**(7): p. 484-492.
39. Nordlund, J., C. Backlin, P. Wahlberg, S. Busche, E. Berglund, M.-L. Eloranta, T. Flaegstad, E. Forestier, B.-M. Frost, A. Harila-Saari, M. Heyman, O. Jonsson, R. Larsson, J. Palle, L. Ronnblom, K. Schmiegelow, D. Sinnett, S. Soderhall, T. Pastinen, M. Gustafsson, G. Lonnerholm, and A.-C. Syvanen, *Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia*. *Genome Biology*, 2013. **14**(9): p. r105.
40. Xu, C.-J., M.J. Bonder, C. Söderhäll, M. Bustamante, N. Baiz, U. Gehring, S.A. Jankipersadsing, P. van der Vlies, C.C. van Diemen, B. van Rijkom, J. Just, I. Kull, J. Kere, J.M. Antó, J. Bousquet, A. Zhernakova, C. Wijmenga, I. Annesi-Maesano, J. Sunyer, E. Melén, Y. Li, D.S. Postma, and G.H. Koppelman, *The emerging landscape of dynamic DNA methylation in early childhood*. *BMC Genomics*, 2017. **18**(1): p. 25.
41. Dolinoy, D.C., R. Das, J.R. Weidman, and R.L. Jirtle, *Metastable epialleles, imprinting, and the fetal origins of adult diseases*. *Pediatr Res*, 2007. **61**(5 Pt 2): p. 30r-37r.
42. Mathers, J.C. and J.A. McKay, *Epigenetics - potential contribution to fetal programming*. *Adv Exp Med Biol*, 2009. **646**: p. 119-23.
43. Wong, N.C., D. Ashley, Z. Chatterton, M. Parkinson-Bates, H.K. Ng, M.S. Halemba, A. Kowalczyk, J. Bedo, Q. Wang, K. Bell, E. Algar, J.M. Craig, and R. Saffery, *A distinct DNA methylation signature defines pediatric pre-B cell acute lymphoblastic leukemia*. *Epigenetics*, 2012. **7**(6): p. 535-41.
44. Dunwell, T., L. Hesson, T.A. Rauch, L. Wang, R.E. Clark, A. Dallol, D. Gentle, D. Catchpoole, E.R. Maher, G.P. Pfeifer, and F. Latif, *A genome-wide screen identifies frequently methylated genes in haematological and epithelial cancers*. *Mol Cancer*, 2010. **9**: p. 44.
45. Davidsson, J., H. Lilljebjorn, A. Andersson, S. Veerla, J. Heldrup, M. Behrendtz, T. Fioretos, and B. Johansson, *The DNA methylome of pediatric acute lymphoblastic leukemia*. *Hum Mol Genet*, 2009. **18**(21): p. 4054-65.
46. Garcia-Manero, G., H. Yang, S.-Q. Kuang, S. O'Brien, D. Thomas, and H. Kantarjian, *Epigenetics of acute lymphocytic leukemia*. *Seminars in hematology*, 2009. **46**(1): p. 10.1053/j.seminhematol.2008.09.008.
47. Nordlund, J., C.L. Backlin, V. Zachariadis, L. Cavelier, J. Dahlberg, I. Ofverholm, G. Barbany, A. Nordgren, E. Overnas, J. Abrahamsson, T. Flaegstad, M.M. Heyman, O.G. Jonsson, J. Kanerva, R. Larsson, J. Palle, K. Schmiegelow, M.G. Gustafsson, G. Lonnerholm, E. Forestier, and A.C. Syvanen, *DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia*. *Clin Epigenetics*, 2015. **7**(1): p. 11.
48. Vineis, P. and F. Perera, *Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old*. *Cancer Epidemiol Biomarkers Prev*, 2007. **16**(10): p. 1954-65.
49. Timms, J.A., C.L. Relton, J. Rankin, G. Strathdee, and J.A. McKay, *DNA methylation as a potential mediator of environmental risks in the development of childhood acute lymphoblastic leukemia*. *Epigenomics*, 2016. **8**(4): p. 519-36.
50. Boyd, A., J. Golding, J. Macleod, D.A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring, and G. Davey Smith, *Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children*. *Int J Epidemiol*, 2013. **42**(1): p. 111-27.

51. Fraser, A., C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey Smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S.M. Nelson, and D.A. Lawlor, *Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort*. *Int J Epidemiol*, 2013. **42**(1): p. 97-110.
52. Golding, J., M. Pembrey, and R. Jones, *ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology*. *Paediatr Perinat Epidemiol*, 2001. **15**(1): p. 74-87.
53. Relton, C.L., T. Gaunt, W. McArdle, K. Ho, A. Duggirala, H. Shihab, G. Woodward, O. Lyttleton, D.M. Evans, W. Reik, Y.L. Paul, G. Ficz, S.E. Ozanne, A. Wipat, K. Flanagan, A. Lister, B.T. Heijmans, S.M. Ring, and G. Davey Smith, *Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)*. *Int J Epidemiol*, 2015.
54. Chatterton, Z., L. Morenos, F. Mechinaud, D.M. Ashley, J.M. Craig, A. Sexton-Oates, M.S. Halemba, M. Parkinson-Bates, J. Ng, D. Morrison, W.L. Carroll, R. Saffery, and N.C. Wong, *Epigenetic deregulation in pediatric acute lymphoblastic leukemia*. *Epigenetics*, 2014. **9**(3): p. 459-67.
55. Figueroa, M.E., S.C. Chen, A.K. Andersson, L.A. Phillips, Y. Li, J. Sotzen, M. Kundu, J.R. Downing, A. Melnick, and C.G. Mullighan, *Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia*. *J Clin Invest*, 2013. **123**(7): p. 3099-111.
56. Sandoval, J., H. Heyn, J. Mendez-Gonzalez, A. Gomez, S. Moran, M. Baiget, M. Melo, I. Badell, J.F. Nomdedeu, and M. Esteller, *Genome-wide DNA methylation profiling predicts relapse in childhood B-cell acute lymphoblastic leukaemia*. *Br J Haematol*, 2013. **160**(3): p. 406-9.
57. Almamun, M., B.T. Levinson, A.C. van Swaay, N.T. Johnson, S.D. McKay, G.L. Arthur, J.W. Davis, and K.H. Taylor, *Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia*. *Epigenetics*, 2015. **10**(9): p. 882-90.
58. Schafer, E., R. Irizarry, S. Negi, E. McIntyre, D. Small, M.E. Figueroa, A. Melnick, and P. Brown, *Promoter hypermethylation in MLL-r infant acute lymphoblastic leukemia: biology and therapeutic targeting*. *Blood*, 2010. **115**(23): p. 4798-809.
59. Pidsley, R., Y.W. CC, M. Volta, K. Lunnon, J. Mill, and L.C. Schalkwyk, *A data-driven approach to preprocessing Illumina 450K methylation array data*. *BMC Genomics*, 2013. **14**: p. 293.
60. Touleimat, N. and J. Tost, *Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation*. *Epigenomics*, 2012. **4**(3): p. 325-41.
61. Fortin, J.P., A. Labbe, M. Lemire, B.W. Zanke, T.J. Hudson, E.J. Fertig, C.M. Greenwood, and K.D. Hansen, *Functional normalization of 450k methylation array data improves replication in large cancer studies*. *Genome Biol*, 2014. **15**(12): p. 503.
62. de Goede, O.M., H.R. Razzaghian, E.M. Price, M.J. Jones, M.S. Kobor, W.P. Robinson, and P.M. Lavoie, *Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells*. *Clin Epigenetics*, 2015. **7**: p. 95.
63. Aryee, M.J., A.E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A.P. Feinberg, K.D. Hansen, and R.A. Irizarry, *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. *Bioinformatics*, 2014. **30**(10): p. 1363-9.
64. Houseman, E.A., W.P. Accomando, D.C. Koestler, B.C. Christensen, C.J. Marsit, H.H. Nelson, J.K. Wiencke, and K.T. Kelsey, *DNA methylation arrays as surrogate measures of cell mixture distribution*. *BMC Bioinformatics*, 2012. **13**: p. 86.
65. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995. **57**(1): p. 289-300.
66. Sharp, G.C., D.A. Lawlor, R.C. Richmond, A. Fraser, A. Simpkin, M. Suderman, H.A. Shihab, O. Lyttleton, W. McArdle, S.M. Ring, T.R. Gaunt, G. Davey Smith, and C.L. Relton, *Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring*

- adiposity: findings from the Avon Longitudinal Study of Parents and Children*. *Int J Epidemiol*, 2015. **44**(4): p. 1288-304.
67. Sherman, B.T., W. Huang da, Q. Tan, Y. Guo, S. Bour, D. Liu, R. Stephens, M.W. Baseler, H.C. Lane, and R.A. Lempicki, *DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis*. *BMC Bioinformatics*, 2007. **8**: p. 426.
  68. Joubert, Bonnie R., Janine F. Felix, P. Yousefi, Kelly M. Bakulski, Allan C. Just, C. Breton, S.E. Reese, Christina A. Markunas, Rebecca C. Richmond, C.-J. Xu, Leanne K. Küpers, Sam S. Oh, C. Hoyo, O. Gruzieva, C. Söderhäll, Lucas A. Salas, N. Baiz, H. Zhang, J. Lepeule, C. Ruiz, S. Ligthart, T. Wang, Jack A. Taylor, L. Duijts, Gemma C. Sharp, Soesma A. Jankipersadsing, Roy M. Nilsen, A. Vaez, M.D. Fallin, D. Hu, Augusto A. Litonjua, Bernard F. Fuemmeler, K. Huen, J. Kere, I. Kull, Monica C. Munthe-Kaas, U. Gehring, M. Bustamante, Marie J. Saurel-Coubizolles, Bilal M. Quraishi, J. Ren, J. Tost, Juan R. Gonzalez, Marjolein J. Peters, Siri E. Håberg, Z. Xu, Joyce B. van Meurs, Tom R. Gaunt, M. Kerkhof, E. Corpeleijn, Andrew P. Feinberg, C. Eng, Andrea A. Baccarelli, Sara E. Benjamin Neelon, A. Bradman, Simon K. Merid, A. Bergström, Z. Herceg, H. Hernandez-Vargas, B. Brunekreef, M. Pinart, B. Heude, S. Ewart, J. Yao, N. Lemonnier, Oscar H. Franco, Michael C. Wu, A. Hofman, W. McArdle, P. Van der Vlies, F. Falahi, Matthew W. Gillman, Lisa F. Barcellos, A. Kumar, M. Wickman, S. Guerra, M.-A. Charles, J. Holloway, C. Auffray, Henning W. Tiemeier, George D. Smith, D. Postma, M.-F. Hivert, B. Eskenazi, M. Vrijheid, H. Arshad, Josep M. Antó, A. Dehghan, W. Karmaus, I. Annesi-Maesano, J. Sunyer, A. Ghantous, G. Pershagen, N. Holland, Susan K. Murphy, Dawn L. DeMeo, Esteban G. Burchard, C. Ladd-Acosta, H. Snieder, W. Nystad, Gerard H. Koppelman, Caroline L. Relton, Vincent W.V. Jaddoe, A. Wilcox, E. Melén and Stephanie J. London, *DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis*. *The American Journal of Human Genetics*, 2016. **98**(4): p. 680-696.
  69. Joubert, B.R., H.T. den Dekker, J.F. Felix, J. Bohlin, S. Ligthart, E. Beckett, H. Tiemeier, J.B. van Meurs, A.G. Uitterlinden, A. Hofman, S.E. Haberg, S.E. Reese, M.J. Peters, B.K. Andreassen, E.A. Steegers, R.M. Nilsen, S.E. Vollset, O. Middtun, P.M. Ueland, O.H. Franco, A. Dehghan, J.C. de Jongste, M.C. Wu, T. Wang, S.D. Peddada, V.W. Jaddoe, W. Nystad, L. Duijts, and S.J. London, *Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns*. *Nat Commun*, 2016. **7**: p. 10577.
  70. Greaves, M., *Infection, immune responses and the aetiology of childhood leukaemia*. *Nat Rev Cancer*, 2006. **6**(3): p. 193-203.
  71. Greaves, M., *A causal mechanism for childhood acute lymphoblastic leukaemia*. *Nat Rev Cancer*, 2018.
  72. Hall, K., B. Frederiksen, M. Rewers, and J.M. Norris, *Daycare Attendance, Breastfeeding, and the Development of Type 1 Diabetes: The Diabetes Autoimmunity Study in the Young*. *BioMed Research International*, 2015. **2015**: p. 203947.
  73. Crosnoe, R., K.M. Purtell, P. Davis-Kean, A. Ansari, and A.D. Benner, *The Selection of Children from Low-Income Families into Preschool*. *Developmental psychology*, 2016. **52**(4): p. 599-612.
  74. Hrušák, O., J. Trka, J. Zuna, A. Poloučková, T. Kalina, and J. Starý, *Acute lymphoblastic leukemia incidence during socioeconomic transition: selective increase in children from 1 to 4 years*. *Leukemia*, 2002. **16**: p. 720.
  75. Needham, B.L., J.A. Smith, W. Zhao, X. Wang, B. Mukherjee, S.L. Kardia, C.A. Shively, T.E. Seeman, Y. Liu, and A.V. Diez Roux, *Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis*. *Epigenetics*, 2015. **10**(10): p. 958-69.

76. Connor Gorber, S., S. Schofield-Hurwitz, J. Hardt, G. Levasseur, and M. Tremblay, *The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status*. *Nicotine Tob Res*, 2009. **11**(1): p. 12-24.
77. Estey, M.P., C. Di Ciano-Oliveira, C.D. Froese, M.T. Bejide, and W.S. Trimble, *Distinct roles of septins in cytokinesis: SEPT9 mediates midbody abscission*. *J Cell Biol*, 2010. **191**(4): p. 741-9.
78. Costa, O., P. Schneider, L. Coquet, P. Chan, D. Penther, E. Legrand, T. Jouenne, M. Vasse, and J.-P. Vannier, *Proteomic profile of pre - B2 lymphoblasts from children with acute lymphoblastic leukemia (ALL) in relation with the translocation (12; 21)*. *Clinical Proteomics*, 2014. **11**(1): p. 31-31.
79. Bilican, B. and C.R. Goding, *Cell cycle regulation of the T-box transcription factor tbx2*. *Experimental Cell Research*, 2006. **312**(12): p. 2358-2366.
80. Lee, S.-T., M.O. Muench, M.E. Fomin, J. Xiao, M. Zhou, A. de Smith, J.I. Martín-Subero, S. Heath, E.A. Houseman, R. Roy, M. Wrensch, J. Wiencke, C. Metayer, and J.L. Wiemels, *Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures*. *Nucleic Acids Research*, 2015. **43**(5): p. 2590-2602.