International Journal of Epidemiology, 2015, 1181–1190
doi: 10.1093/ije/dyv072
Advance Access Publication Date: 19 May 2015
Data Resource Profile

Data Resource Profile

Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)

Caroline L Relton,[1,2,3*] Tom Gaunt,[1,2] Wendy McArdle,[2] Karen Ho,[2] Aparna Duggirala,[2] Hashem Shihab,[1,2] Geoff Woodward,[1,2] Oliver Lyttleton,[2] David M Evans,[1,2,4] Wolf Reik,[5,6] Yu-Lee Paul,[5] Gabriella Ficz,[7] Susan E Ozanne,[8] Anil Wipat,[9] Keith Flanagan,[9] Allyson Lister,[9] Bastiaan T Heijmans,[10] Susan M Ring[1,2] and George Davey Smith[1,2]

[1]MRC Integrative Epidemiology Unit, and [2]School of Social and Community Medicine, University of Bristol, Bristol, UK, [3]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, [4]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia, [5]Babraham Institute, Cambridge, UK, [6]Wellcome Trust Sanger Institute, Cambridge, UK, [7]Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK, [8]University of Cambridge Institute of Metabolic Sciences and MRC Metabolic Diseases Unit, Cambridge, UK, [9]School of Computer Science, Newcastle University, Newcastle upon Tyne, UK and [10]Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

*Corresponding author. MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. E-mail: caroline.relton@bristol.ac.uk

Accepted 8 April 2015

Key words: ALSPAC, epigenetics, DNA methylation

Data resource basics

ARIES is a population-based resource of DNA methylation data accessible by the research community, established with the aim of improving our understanding of the determinants and consequences of epigenetic variation, in particular the role of DNA methylation in health and development. To this end, ARIES has profiled genome-wide DNA methylation on samples selected from approximately 1000 mother-child pairs at three time points in children and two time points in their mothers. ARIES has also profiled a wide range of fetal and adult tissues for reference purposes, with which to compare cord blood and peripheral blood DNA methylation patterns. A web portal has been constructed to allow openly accessible browsing of aggregate ARIES DNA methylation data in the context of other publicly available data on many aspects of gene function, using a custom-built data integration system (ARIES-Explorer).

Population-based DNA methylation profiling

ARIES is a sub-study of the Avon Longitudinal Study of Parents and Children (ALSPAC).[1,2] A sub-set of 1018 mother-offspring pairs were included in ARIES and were selected based on availability of DNA samples at two time-for the mother (at an antenatal clinic and at a follow-up clinic when their offspring were mean age 15.5 years) and three time points for the offspring [at birth, childhood, (mean age 7.5 years) and adolescence at mean age 15.5 years]. Figure 1 provides a schematic of the ALSPAC participants and time points profiled as part of ARIES.

ARIES children provided a DNA sample extracted from cord blood drawn from the umbilical cord upon delivery.

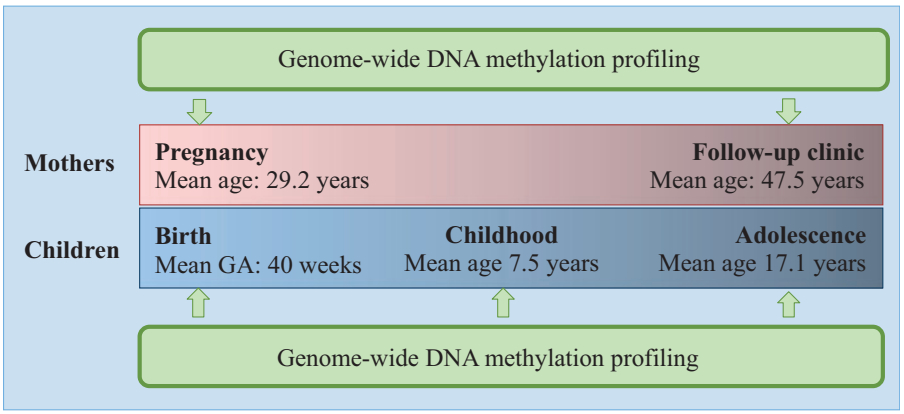

Figure 1. Schematic of the ALSPAC participants included in ARIES. A total of 1018 mother-child pairs have been profiled (v.1 data release 2014) at each of the five time points indicated, using the Illumina Infinium HumanMethylation450K BeadChip array. Ten of the same mother-child pairs have been profiled using whole-genome bisulphite sequencing at each of the five time points shown. GA, gestational age.

A second DNA sample was extracted from peripheral blood drawn at the 7-year clinic visit, and a third at a clinic held either at age 15 or at age 17 years. Maternal samples were provided during the antenatal period [mean 25.7 (SD 9.5) weeks of gestation] and approximately 15–17 years later when mothers accompanied their child to a clinic visit or when they attended a clinic themselves.

Tissue-specific DNA methylation profiling

DNA methylation patterns differ between tissues. To generate a reference resource to facilitate comparison of peripheral blood DNA methylation with that of other tissues, we undertook profiling of DNA methylation from a range of other sources, where possible with paired peripheral blood DNA. Tissues were provided from three sources (as described below) for genome-wide DNA methylation profiling in ARIES using the Illumina Infinium HumanMethylation450 BeadChip (450 K) array. The tissues represented are summarized in Table 1.

i. Adult autopsy tissues. Twelve tissue types were harvested from 16 autopsy cases (10 male, 6 female) and provided for DNA methylation profiling and genotyping (through collaboration with Leiden University Medical Center). Tissue included peripheral blood (collected from the thoracic cavity) to allow cross-reference between blood and other tissue profiles from the same individuals. The samples were collected within 12 h of the death of the subjects with a mean age of 62.8 years. Further details can be found elsewhere.[3] All samples are anonymized and procedures were performed according to the ethical guidelines in the Code for Proper Secondary Use of Human Tissue in The Netherlands (Dutch Federation of Medical Scientific Societies).
ii. Post-mortem brain tissues. Post-mortem brain tissue was provided by the Newcastle Brain Tissue Resource. A total of 20 samples underwent DNA methylation profiling from each of 10 male and 10 female donors. Multiple regions were sampled from post-mortem brain tissue; frontal cortex, temporal cortex, hippocampus, substantia nigra, dorsal raphe nucleus, putamen, hypothalamus, amygdala, cerebellum and peripheral blood DNA was also provided from the same individuals.
iii. Fetal tissues. Fetal tissues were provided by the MRC Human Developmental Biology Resource at Newcastle University from terminations of pregnancy. Twenty different tissues (listed in Table 1) were harvested from two male and two female fetuses of gestational age 12–14 weeks. No associated information is available for the fetal tissues (e.g. maternal age or reason for termination).

Whole-genome bisulphite sequencing

Samples from 10 mother-child pairs selected from the ALSPAC cohort, that underwent Illumina 450 K array analysis at each of the available time points, were also whole genome-bisulphite sequenced (BS-Seq) at the same time points (Figure 1). Samples were selected based on the completeness of the Illumina 450 K data and availability of additional DNA at the same time points.

Gene expression profiling

All tissues undergoing DNA methylation profiling, with the exception of peripheral blood (as no RNA was available in this instance), were considered for gene expression analysis. Samples were limited by the availability of good quality RNA. Those tissue types where successful gene expression data were generated are summarized in Table 1.

Table 1. Tissues profiled as part of ARIES in addition to cord blood and peripheral blood DNA samples from ALSPAC

Tissue	Source					
	Human developmental biology resource		Adult autopsy		Newcastle brain tissue resource	
	$N = 4$		$N = 16$		$N = 20$	
	DNAm	RNA	DNAm	RNA	DNAm	RNA
Blood			x		x	
Skin	x	x	x			
Subcutaneous fat			x			
Omentum			x			
Stomach	x	x				
Small intestine	x	x				
Large intestine	x	x				
Lung	x	x				
Skeletal muscle	x	x	x			
Liver	x	x	x			
Kidney	x	x	x			
Kidney fat			x			
Adrenal gland	x	x				
Heart	x	x	x			
Spleen	x	x	x			
Pancreas	x	x	x			
Eye	x	x				
Gonad (ovary/testis)	x	x				
Bone (rib)	x	x				
Thymus	x	x				
Umbilical cord	x	x				
Chorionic villi	x					
Yolk sac	x					
Brain (frontal cortex)	x	x	x		x	
Brain (temporal cortex)	x	x			x	
Brain (cerebellum)	x	x			x	
Brain (hippocampus)					x	
Brain (substantia nigra)					x	
Brain (dorsal raphe nucleus)					x	
Brain (putamen)					x	
Brain (hypothalamus)					x	
Brain (amygdala)					x	

DNAm, DNA methylation data available; RNA, gene expression data available.

Approximately 1000 ALSPAC lymphoblastoid cell lines derived from blood samples drawn from participants at age 9 years had previously undergone gene expression profiling. There is an overlap of 319 ALSPAC participants for whom both DNA methylation and gene expression data are available.

Data browsing

ARIES-Explorer [www.ariesepigenomics.org.uk] provides a web interface to browse methylation variation between groups of individuals and across time. Supporting this web interface, a highly powerful platform for graph-based data mining and data integration, termed *Entanglement*, was constructed to integrate ARIES DNA methylation data and other information to allow 'on-the-fly' querying of relationships with publicly available epigenomics data sets. This system is highly scalable, allowing large data sources to be integrated through use of disposable virtual machines and no-SQL databases [http://intbio.ncl.ac.uk/?projects=entanglement].

Linkage to existing data

The extensive body of genetic and life-course data available on the ALSPAC study are detailed in the cohort profiles[1,2] and can be browsed via the study website [http://www.bristol.ac.uk/alspac/] which contains details of all the data that are available through a fully searchable data dictionary [http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/].The potential exists to link ARIES DNA methylation data to any of this information, which covers a plethora of measures spanning social, behavioural, psychiatric, environmental, genetic, metabolomic and many other domains.

Ethical approval

Ethical approvals are in place for all sources of biological samples and data in ARIES and include appropriate permissions from the Human Development Biology Resource, Newcastle Brain Tissue Resource and Leiden University Medical Center. Ethical approval for the ALSPAC study was obtained from the ALSPAC Ethics and Law Committee and the local research ethics committees.

Data resource area and population coverage

ALSPAC is a long-term health research project that started in the early 1990 s. It recruited more than 14 000 pregnant women in the former county of Avon, who were due to give birth between April 1991 and December 1992. The study currently holds longitudinal data and biosamples for these women, their offspring and their partners. Table 2 compares the characteristics of mothers in ARIES with those in the rest of ALSPAC. Mothers included in ARIES were slightly older, more likely to have a non-manual occupation and less likely to have smoked throughout pregnancy. The ARIES sub-sample is considered to be reasonably representative of the main study population.

Frequency of database updates

The first public release of ARIES data became available in June 2014. Subsequent annual updates will incorporate all new data generated and processed by that date and will be

Table 2. Comparison of a selection of maternal characteristics in ALSPAC mothers included and not included in ARIES

Characteristic	ALSPAC mothers *included in ARIES* Mean (SD), median (IQR) or %*	ALSPAC mothers not *included in ARIES* Mean (SD), median (IQR) or %*
Offspring sex	($n = 1018$)	($n = 13041$)
Male	48.8%	51.9%
Female	51.2%	48.1%
Gestational age at delivery (weeks)	($n = 1018$)	($n = 13614$)
	40 (39 – 41)	39.3 (2.1)
Parity	($n = 989$)	($n = 12002$)
Nulliparous	46.4%	44.6%
Multiparous	53.6%	55.5%
Age at delivery (years)	($n = 986$)	($n = 10838$)
	29.2 (4.4)	28.2 (4.8)
Occupation	($n = 901$)	($n = 9188$)
Manual occupation	14.0%	20.5%
Non-manual occupation	86.0%	79.5%
Smoking status	($n = 1006$)	($n = 12166$)
Never before or during pregnancy	86.7%	73.3%
Before pregnancy or during 1st trimester only	3.6%	7.3%
Throughout pregnancy	9.7%	19.4%
Reported pre-pregnancy BMI	($n = 944$)	($n = 10633$)
	22.8 (3.7)	22.9 (3.8)

* For mothers or offspring as appropriate.

annotated with version numbers, with all previous versions stored and available on request. Notification of version updates will be posted on the study website.

Additional data generation is under way, with further work planned. This includes Illumina 450 K array analysis of more ALSPAC participants including fathers ($n = 312$ ARIES fathers who have been profiled to date), samples from ARIES mothers attending further follow-up clinics, additional time points for ARIES children with further DNA samples available during infancy, children of ALSPAC participants, various case series and specific exposure sub-groups.

The ARIES-Explorer web portal has the facility to import the browser's own Illumina 450 K array data as a track for visualization and comparison alongside ARIES data in the genome browser. The functionality of ARIES-Explorer is also under development and additional features will become available in due course. Plans are under way to add data from several ethnic groups (South Asian, South African, Brazilian).

Measures

Illumina Infinium HumanMethylation450K BeadChip

DNA methylation was quantified using the Illumina Infinium® HumanMethylation450K BeadChip assay (Illumina 450 K array) (Illumina Inc., CA). Cord and peripheral blood samples (whole blood or buffy coat) were collected according to standard procedures. Following extraction, DNA was bisulfite-converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA) then genome-wide methylation status of over 485 000 CpG sites was measured using the Illumina 450 K array according to the standard protocol. The arrays were scanned using an Illumina iScan and initial quality review was assessed using GenomeStudio (version 2011.1).

Samples from all participant ages in ARIES were distributed across slides using a semi-random approach (sampling criteria were in place to ensure that all time points were represented on each array) to minimize the possibility of confounding by batch effects. In addition, during the data generation process a wide range of batch variables were recorded in a purpose-built laboratory information management system (LIMS). The LIMS also reported quality control (QC) metrics from the standard control probes on the Illumina 450 K array for each sample. Samples failing QC (average probe P-value ≥ 0.01) were excluded from further analysis and scheduled for repeat assay. As an additional QC step, genotype probes were compared with SNP-chip data from the same individual to identify and remove any sample mismatches. For individuals with no genome-wide SNP data, samples

were flagged if there was a sex-mismatch based on X-chromosome methylation.

Tissue samples from autopsy specimens, post-mortem brain and fetal samples were processed in exactly the same way as blood samples described above.

The Illumina 450 K BeadChip assay detects the proportion of molecules methylated at each CpG site on the array. For each sample, the estimated methylation level at each CpG site was expressed as a beta value (β), which is the ratio of the methylated probe intensity and the overall intensity and ranges from 0 (no cytosine methylation) to 1 (complete cytosine methylation). Methylation data were pre-processed using R (version 3.0.1), with background correction and subset quantile normalization performed within each time point using the pipeline described by Touleimat and Tost.[4] Data are available both pre- and post-normalization.

It is widely recognized that white blood cell heterogeneity can confound DNA methylation measurement.[5] The overwhelming majority of ALSPAC biosamples did not have whole blood cell counts taken prior to DNA extraction and therefore the only viable option is to apply a *post hoc* correction for cellular heterogeneity to the DNA methylation data generated. In order to ensure that the results of any downstream analysis are not influenced by variation in cell type fraction between samples, we estimated the fraction of CD8 T cells, CD4 T cells, NK cells, B cells, monocytes and granulocytes in the samples using the *estimateCellCounts* function in the *minfi* Bioconductor package implemented in R.[6] Both cell-type corrected and uncorrected data are available to facilitate use of either data set in a main analysis or the corrected data set as a sensitivity analysis.

Whole-genome bisulphite sequencing

For BS-seq library preparation, DNA samples were fragmented by sonication (Covaris) and adaptor ligated (using Illumina-supplied methylated adaptors and NEBnext library preparation kit). DNA was then bisulphite-treated using the Sigma Imprint kit, according to the manufacturer's instructions (one-step protocol). Final library amplification (16 cycles) was done using Pfu Turbo Cx (Agilent), after which the libraries were gel-purified using QIAGEN Minelute kit. Libraries were sequenced on either an Illumina GAIIx or an Illumina HiSeq using the default RTA analysis software.[7]

Gene expression array analysis

Pre-existing LCL gene expression data were generated using the Illumina HT-12 V3 BeadChip (Illumina Inc., San Diego, CA) which included 48 804 probes.[8] Additional gene expression data on selected tissue samples were generated by ServiceXS (Leiden, The Netherlands) using the same array.

Genotyping

Participants of ALSPAC, including those in ARIES, were previously genotyped. Children were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina Inc., San Diego, CA) by the Wellcome Trust Sanger Institute (WTSI, Cambridge, UK) and the Laboratory Corporation of America (LCA, Burlington, NC, USA) with support from 23andMe. Mothers were genotyped using the Illumina Human660W-quad genome-wide SNP genotyping platform (Illumina Inc., San Diego, CA) at the Centre National de Génotypage (CNG, Paris, France). Imputation has been performed to increase SNP density for all mothers and children. The final imputed data set contains >8 million SNPs with imputation quality >0.8 and minor allele frequency (MAF) > 1%. The ARIES sample also overlaps with participants included in a large-scale whole-genome sequencing initiative to identify low frequency and rare variants, the UK10K project [http://www.uk10k.org/].

Genotyping of autopsy cases was undertaken using the Illumina HumanCoreExome-24 BeadChip SNP array (Illumina Inc., San Diego, CA), at the University of Bristol, following the manufacturer's protocol.

Data resource use

Longitudinal studies provide a powerful opportunity to investigate the role of epigenetic variation in both normal development and the evolution of adverse traits and diseases.[9] To date the use of ARIES data has been motivated by six main areas of activity, illustrative examples of which are provided below. There are currently in excess of 70 projects approved to access and use DNA methylation data in ALSPAC, so it is anticipated that wider usage of ARIES will prompt many further applications of the data in due course.

i. Epigenome-wide association studies (EWAS) have been undertaken to uncover associations between DNA methylation and a range of exposures or outcomes. These have included analysis of the influence of: maternal prenatal smoking on offspring DNA methylation[10] (Figure 2A) and the mediation of lower infant birthweight in smokers;[11] maternal pre-pregnancy BMI on offspring DNA methylation and subsequent childhood adiposity;[12] and maternal vitamin B_{12} status on offspring DNA methylation and subsequent childhood IQ. In an effort to provide a rigorous

evidence base for future epigenetic studies, a Pregnancy and Childhood Epigenetics Consortium has been established (PACE), to facilitate meta-analyses across several exposures and outcomes pertinent to this sector of the population. EWAS of many other exposures and phenotypes have been conducted (e.g. alcohol intake, trace elements, heavy metals, stress, autism spectrum disorders, children born following assisted reproductive technologies, etc.).

ii. Replication studies. Frequent requests are made by other investigators to access ARIES data relating to one or a few loci that have been identified as being associated with a specific trait or exposure in an independent cohort.[13,14] Such replication is an essential step in a study design, and ARIES has developed reciprocal arrangements for replication with several other cohort studies. Applications to access and use ARIES data for replication can be made at any time with no requirement to involve ALSPAC investigators.

iii. Genome-wide association studies have been undertaken to identify common genetic variation associated with DNA methylation variation. A large-scale GWAS has been completed, indicating a stable genetic influence of genotype on DNA methylation across the life course. Findings are currently being prepared for publication and summary statistics for all observed SNP-CpG associations (numbering in excess of 24 000 independent associations at each time point), in the form of a methylation quantitative trait locus (mQTL) catalogue will be freely available online via the study website [www.ariesepigenomics.org.uk]. A Genetics of DNA Methylation Consortium (GoDMC) [www.godmc.org.uk] has recently been established to facilitate meta-analysis of GWAS studies.

iv. Development and application of statistical approaches for analysing DNA methylation data, including longitudinal modelling of DNA methylation over time (Figure 2B) and the application of Mendelian randomization[15] to strengthen causal inference in epigenetic epidemiology (Figure 2C), have been a major focus of ARIES investigators. The vulnerability of epigenetic studies to confounding and reverse causation have been noted previously by us,[16] as has the importance of strengthening causal inference using a variety of approaches.[17-19] Prominent among such strategies is the application of Mendelian randomization,[18] which is beginning to be applied in studies utilizing ARIES data (Figure 2C). The availability of dyads and trios allows family-based approaches to be implemented: for example, the assessment of intra-uterine (and inter-generational) effects on DNA methylation[11-13] as well as assessment of correlation within families. A further area of methodological development using ARIES data is the modelling of change in DNA methylation patterns over time. Studies to date have demonstrated evidence of attenuation of differences in DNA methylation identified at birth, with some methylation differences persisting for much longer than others[10,20] (Figure 2B).

v. Interrogation of functional effects of trait-associated loci identified in genetic association studies has been undertaken using ARIES data. Following the identification of a GWAS association, it is straightforward to 'look-up' the SNP of interest in the ARIES mQTL catalogue to assess whether the SNP of interest is likely to be exerting its influence via perturbation of DNA methylation; i.e. is the GWAS SNP also a mQTL? Such an approach was recently adopted using ARIES mQTL to assess the functional significance of loci associated with thyroid function.[21]

vi. Analysis of tissue-specific DNA methylation patterns in fetal tissues, brain regions, autopsy tissues and peripheral blood are in progress, including a comparative analysis of tissue-specific differentially methylated sites and regions.

Strengths and weaknesses

ARIES is a large, longitudinal population-based epigenetics resource which includes dyads and trios and which is nested within the richly phenotyped ALSPAC, and so provides an unparalleled opportunity to link DNA methylation data to a vast array of exposure and outcome variables as well as model changes in these parameters over time. It is imperative that DNA methylation data are interpreted in light of the underlying architecture of genetic variation,[22] and this is possible given the comprehensive genome-wide SNP data also available in ARIES. The majority of ARIES children are also part of the UK10K sequencing initiative which has generated whole-genome sequence data. The study has provided the first online mQTL catalogue, which will prove invaluable in functional interrogation of loci of interest as well as for the identification of genetic proxies for DNA methylation in Mendelian randomization studies.

ALSPAC and ARIES place high priority on facilitating data access. To this end, summary ARIES data are freely available online and a web portal has been specifically developed to enhance data browsing and integration with other data sources. Individual-level data are also available through standard ALSPAC data access mechanisms which are open to all bona fide researchers.

Limitations of ARIES include its reliance on DNA extracted from peripheral blood. This places constraints

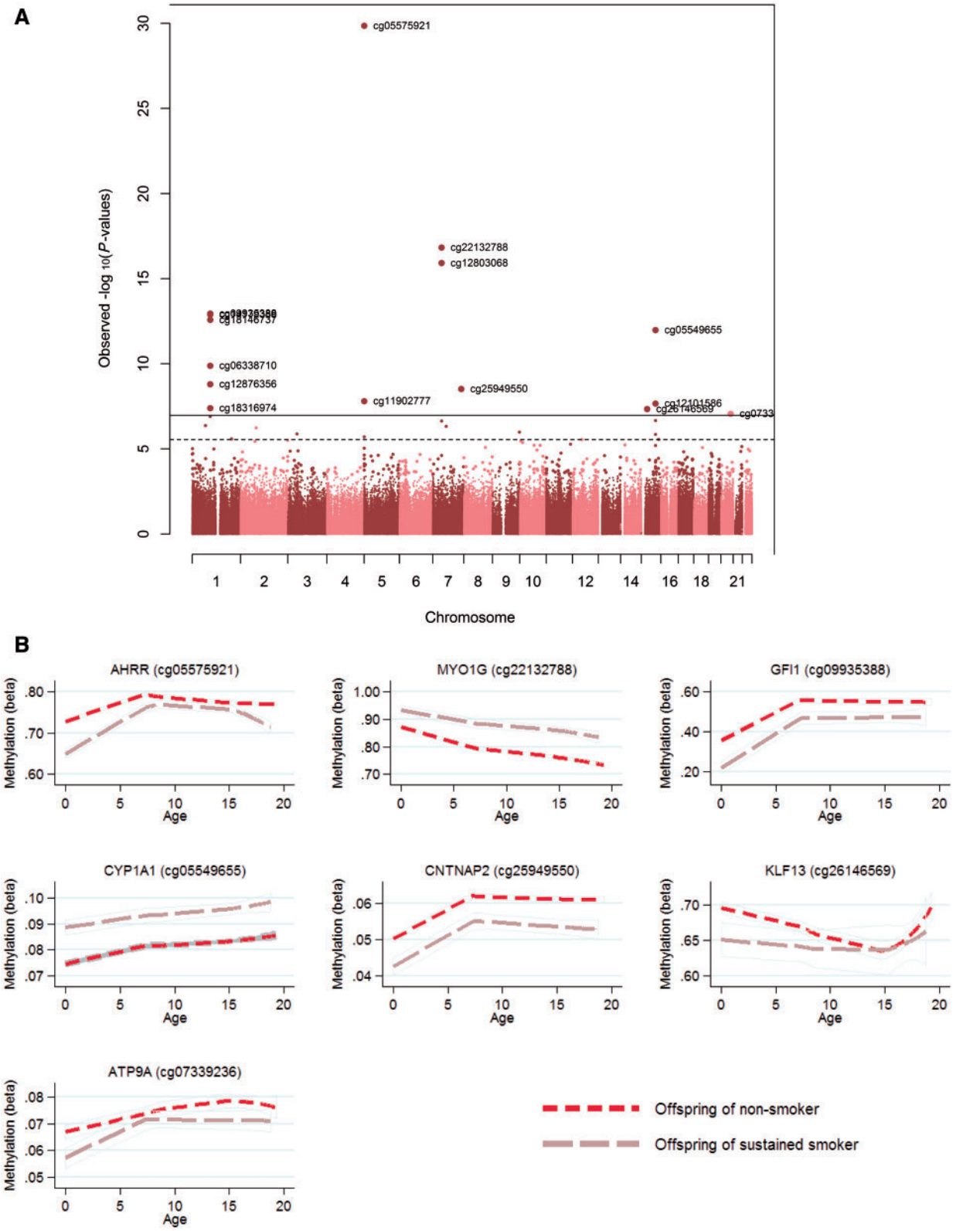

Figure 2. Use of ARIES data. (A) An epigenome-wide association study illustrating that maternal prenatal smoking is associated with DNA methylation differences in neonatal DNA of offspring.[18] (B) Longitudinal modelling of differentially methylated loci over time to assess the persistence of DNA methylation changes arising from prenatal exposure to cigarette smoking.[18] (C) The application of Mendelian randomization to epigenetic data. An example is the analysis of the relationship of maternal vitamin B_{12}, neonatal DNA methylation and IQ at age 8 years. SNP1 acts as a proxy for maternal vitamin B_{12} levels and SNP2 acts as a proxy for site-specific DNA methylation.

(continued)

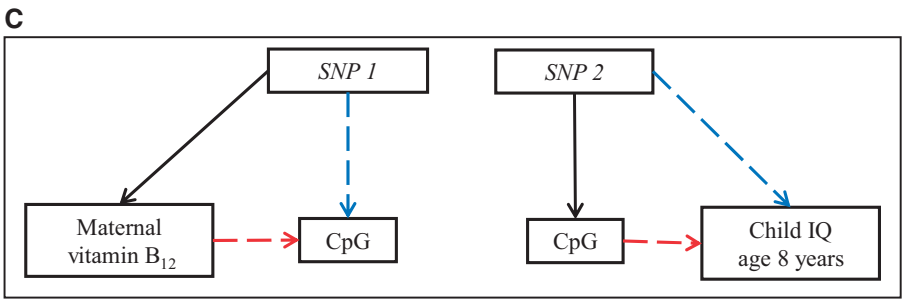

Figure 2. Continued.

upon the inferences that can be made regarding tissue-specific DNA methylation patterns. The creation of a reference data series of numerous tissue types mitigates this limitation to some extent but is an imperfect solution. Other limitations include reliance on the Illumina Infinium HumanMethylation450K BeadChip which has limited coverage (1.7% of all CpG sites in the genome) and recognized technical biases.[23] Complementary methods can be (and have been) used for validation and verification of differential DNA methylation, such as Pyrosequencing or targeted BS-Seq. Applications can be made to access and use biological samples from the ALSPAC bio-repository to undertake additional epigenetic (or other) analyses.

Data resource access

Summary data can be browsed via the ARIES-Explorer web portal [www.ariesepigenomics.org.uk]. Additional linked data are available by following the ALSPAC project approval process. Prospective data users are encouraged to browse the catalogue of existing approved projects listed on the ALSPAC website using [epigenetics] [methylation] keyword search, to apprise themselves of ongoing work and potential overlapping activities in their chosen field. ALSPAC maintains a policy of non-exclusivity, and therefore duplication of effort can occur and is the responsibility of the investigator to avert. At the time of writing, approximately 10% of ALSPAC data requests include DNA methylation.

The ALSPAC access policy and data dictionary can be located via the following links: [http://www.bristol.ac.uk/alspac/researchers/data-access/, http://www.bristol.ac.uk/alspac/researchers/data-access/data-dictionary/]

Illumina 450 K array data are provided as flat (comma-separated value) text files for each time point comprising raw 'beta' (proportion methylated), normalized 'beta', unmethylated probe intensities, methylated probe intensities and detection *P*-value (enabling signal/noise filtering). BS-Seq data are available as 'BedGraph' flat text files comprising proportion methylated, methylated read count and unmethylated read count for each reported locus in the genome.

Illumina 450 K array data can be imported into R and analysed with a wide range of packages built for analysis of 450 K data, including wateRmelon[24] and Minfi[25] for processing/QC and CpGassoc[26] for epigenome-wide association analysis (among many other packages). We have established a user forum to address questions and issues arising from the analysis of epigenetic data, including Illumina 450 K array data, at [https://groups.google.com/forum/#!forum/epigenomicsforum].

All studies applying to use ARIES data linked to other ALSPAC data are requested to give due acknowledgement to both studies.

ARIES in a nutshell

- ARIES was set up to support population-based epigenetic studies. It is the only study of its kind to have large scale serial DNA methylation data on the same individuals and on mother-child pairs.
- Initiated in 2012, ARIES includes 1,000 mother-child pairs from the Avon Longitudinal Study of Parents and Children (ALSPAC) recruited during pregnancy, 1991-92.
- Participants have undergone genome-wide DNA methylation analysis at birth (child and mother), childhood (child only) and adolescence (child and mother). In addition a range of reference tissues from foetal samples, post-mortem brain tissue and autopsy samples have been profiled.
- Data can be linked to a wide range of environmental, social, lifestyle, genetic, phenotypic and other data collected as part of ALSPAC.
- Aggregate data are openly available to browse via the study web portal (www.ariesepigenomics.org.uk) and individual level data are available upon request via ALSPAC (http://www.bristol.ac.uk/alspac/).

Funding

ARIES was funded by the BBSRC (BBI025751/1 and BB/I025263/1). Core programme support for ALSPAC is provided by the Medical Research Council (MRC) and the Wellcome Trust (Grant ref: 102215/2/13/1) and the University of Bristol. Supplementary funding to generate DNA methylation data which is (or will be) included in ARIES has been obtained from the MRC, ESRC, NIH and other sources. ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/2 and MC_UU_12013/8). Collection of the autopsy series was supported by the European Union's Seventh Framework Program IDEAL (FP8/2007–2011) under grant agreement No.259679.

Acknowledgements

We are extremely grateful to all the families who took part in ALSPAC, 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. We thank those who have made a substantive contribution to ARIES, including Laura Barrett, Patience Ezea, Gibran Hemani, Nabila Kazmi, Rebecca Richmond, Gemma Sharp, Roderick Slieker and Matthew Suderman, the Newcastle Brain Tissue Resource, the Human Developmental Biology Resource and the ARIES Scientific Advisory Board (Adrian Bird, Stephan Beck, Jane Rogers, Chris Larminie and John Greally).

Conflict of interest: None declared.

References

1. Fraser A, Macdonald-Wallis C, Tilling K *et al*. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013;**42**:97–110.
2. Boyd A, Golding J, Macleod J *et al*. Cohort Profile: The 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;**42**:111–27.
3. Slieker RC, Bos SD, Goeman JJ *et al*. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* 2013;**6:**26.
4. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 2012;**4**:325–41.
5. Reinius LE, Acevedo N, Joerink M *et al*. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 2012;**7**: e41361.
6. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;**15**:R31.
7. Ficz G, Hore TA, Santos F *et al*. FGF signalling in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* 2013;**13:** 351–59.
8. Grundberg E, Small KS, Hedman ÅK *et al*. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012;**44:**1084–89.
9. Ng JWY, Barrett LM, Wong A, Kuh D, Davey Smith G, Relton CL. The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities. *Genome Biol* 2012; **13:**246.
10. Richmond RC, Simpkin AL, Woodward G *et al*. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse. *Hum Mol Genet* 2015;**24:**2201–17.
11. Kupers L, Xu X, Vaez A *et al*. DNA methylation mediates the effect of maternal smoking during pregnancy on birth weight of the offspring. *Int J Epidemiol* 2015;**44**:1224–37.
12. Sharp G, Lawlor DA, Richmond RC *et al*. 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**:1288–304.
13. Lee KW, Richmond R, Hu P *et al*. Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environ Health Perspect* 2015;**123:** 193–99.
14. Chambers JC, Loh M, Lehne B *et al*. Epigenome-wide association identifies DNA methylation markers in peripheral blood that predict incident Type-2 diabetes amongst Indian Asians and Europeans. *Lancet Diabetes Endocrinol* 2015;**3**: 526–34
15. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *Int J Epidemiol* 2012;**41:**5–9.
16. Kirkbride J, Susser E, Kundakovic M, Kresovich J, Davey Smith G, Relton CL. Prenatal nutrition, epigenetics and schizophrenia risk: Can we test causal effects? *Epigenomics* 2012;**4:** 303–15.
17. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32:**1–22.
18. Richmond RC, Al-Amin A, Davey Smith G, Relton CL. Approaches for drawing causal inferences from epidemiological birth cohorts: A review. *Early Hum Dev* 2014;**90:** 769–80.
19. Relton CL, Davey Smith G. Two step epigenetic Mendelian randomization: a strategy for establishing a causal role for epigenetic factors in pathways to disease *Int J Epidemiol* 2012;**41:** 161–76.
20. Simpkin A, Suderman M, Gaunt T *et al*. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum Mol Genet* 2015;**24:**2201–17.
21. Taylor PN, Porcu E, Chew S *et al*. Whole genome sequence based analysis of thyroid function. *Nature Commun* 2015;**6:** 5681
22. Naeem H, Wong NC, Chatterton Z *et al*. Reducing the risk of false discovery enabling identification of biologically significant genome-wide status using the HumanMethylation450 array. *BMC Genomics* 2014;**15:**51.
23. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 2013;**14:**585–94.

24. Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013;**14:**293.
25. Aryee MJ, Jaffe AE, Corrada-Bravo H *et al*. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30:**1363–69.
26. Barfield RT, Kilaru V, Smith AK, Coneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics* 2012;**28:**1280–81.