

REVIEW

Open Access



Best practices in DNA methylation: lessons from inflammatory bowel disease, psoriasis and ankylosing spondylitis

Jessica M. Whyte¹, Jonathan J. Ellis¹, Matthew A. Brown^{1,2*†}  and Tony J. Kenna^{1†}

Abstract

Advances in genomic technology have enabled a greater understanding of the genetics of common immune-mediated diseases such as ankylosing spondylitis (AS), inflammatory bowel disease (IBD) and psoriasis. The substantial overlap in genetically identified pathogenic pathways has been demonstrated between these diseases. However, to date, gene discovery approaches have only mapped a minority of the heritability of these common diseases, and most disease-associated variants have been found to be non-coding, suggesting mechanisms of disease-association through transcriptional regulatory effects.

Epigenetics is a major interface between genetic and environmental modifiers of disease and strongly influence transcription. DNA methylation is a well-characterised epigenetic mechanism, and a highly stable epigenetic marker, that is implicated in disease pathogenesis. DNA methylation is an under-investigated area in immune-mediated diseases, and many studies in the field are affected by experimental design limitations, related to study design, technical limitations of the methylation typing methods employed, and statistical issues. This has resulted in both sparsity of investigations into disease-related changes in DNA methylation, a paucity of robust findings, and difficulties comparing studies in the same disease.

In this review, we cover the basics of DNA methylation establishment and control, and the methods used to examine it. We examine the current state of DNA methylation studies in AS, IBD and psoriasis; the limitations of previous studies; and the best practices for DNA methylation studies. The purpose of this review is to assist with proper experimental design and consistency of approach in future studies to enable a better understanding of the functional role of DNA methylation in immune-mediated disease.

Keywords: DNA methylation, Epigenetics, Ankylosing spondylitis, Psoriasis, Inflammatory bowel disease (IBD), Human

Background

Ankylosing spondylitis (AS), inflammatory bowel disease (IBD) and psoriasis are characterised by systemic or organ-specific failures of the regulatory pathways of the immune system resulting in uncontrolled inflammation. These ‘seronegative’ diseases are closely related clinically and often co-occur in patients and families. Several GWAS in these diseases have identified the same genes that confer

susceptibility to disease, including human leukocyte antigen (*HLA*), which plays a vital role in adaptive immunity and tolerance, IL-23 receptor (*IL23R*), DNA methyltransferase 3A (*DNMT3A*), *DNMT3B*, *DNMT3L* and several genes involved in the JAK-STAT pathway [1–4]. A cross-disease genetic study of five seronegative diseases (AS, ulcerative colitis, Crohn’s disease, psoriasis and primary sclerosing cholangitis) identified pleiotropic genes and shared pathogenic pathways between these diseases [4].

Despite the numerous disease-associated variants identified in AS, IBD and psoriasis, they cumulatively explain only a small proportion (<28%) of the heritability of these diseases [4]. Potential reasons for this ‘missing heritability’ include large numbers of variants of smaller effect yet to be identified, rare variants being missed by

* Correspondence: matt.brown@qut.edu.au

†Matthew A. Brown and Tony J. Kenna are senior authors and joint contributors to the study.

¹Institute of Health and Biomedical Innovation, Queensland University of Technology, Woolloongabba, Queensland, Australia

²Translational Research Institute, Princess Alexandra Hospital, 37 Kent Street, Woolloongabba, Queensland 4102, Australia



available genotyping arrays, copy number variation (CNV), insertion/deletion events, gene-gene interactions and epigenetic factors. Epigenetics refers to functional modifications to DNA other than base sequence coding and includes histone modifications, non-coding RNA interactions with transcriptional and translational machinery and DNA methylation. Epigenetic variation is a dynamic and responsive process that occurs throughout life and in each individual cell and tissue within a single organism.

In this review, we focus on DNA methylation as it has the most robust measurement methods of all forms of epigenetic variation, making it a tractable epigenetic form to study. We discuss the current state of epigenetic research in AS, IBD and psoriasis, and the key aspects of study design that are relevant to dissecting the pathogenic mechanisms involved in these diseases.

What is DNA methylation?

DNA methylation refers to the addition of a methyl group (CH₃) to a cytosine to form 5-methylcytosine (5mC). Predominantly, methylation occurs on cytosine phosphate guanine-paired bases (CpGs). The majority of methylated CpGs occur in CpG islands, dense CpG regions of DNA between 300 and 3000 bp. CpG sites

located 2 kb upstream or downstream of a CpG island are defined as CpG shores, and CpG sites 2 kb beyond these shores are defined as CpG shelves. Regions outside this 4 kb stretch are referred to as the ‘open sea’. DNA methylation can occur in cytosines outside CpG sites, but whether these are recognised as methylated sites is unclear [5]. Generally, CpGs in differentiated cells are uniformly methylated or unmethylated between homologous chromosomes and within cell populations, resulting in a bimodal distribution.

DNA methylation is a highly stable chemical marker that is maintained through mitosis. It is moderately heritable between generations. Maternal exposure to environmental factors can affect embryos in utero and, for female embryos, also affect their gametes, as female reproductive cells are fixed at birth [6]. As cells undergo two cycles of demethylation during reproduction, it is unclear whether these changes can be transmitted beyond these generations [6, 7].

The process of methylation establishment and maintenance is carried out by the DNA methyltransferase (DNMT) protein family (see Fig. 1). DNMT family proteins have a CXXC domain that recognises unmethylated CpGs [8]. The de novo establishment of DNA methylation is carried

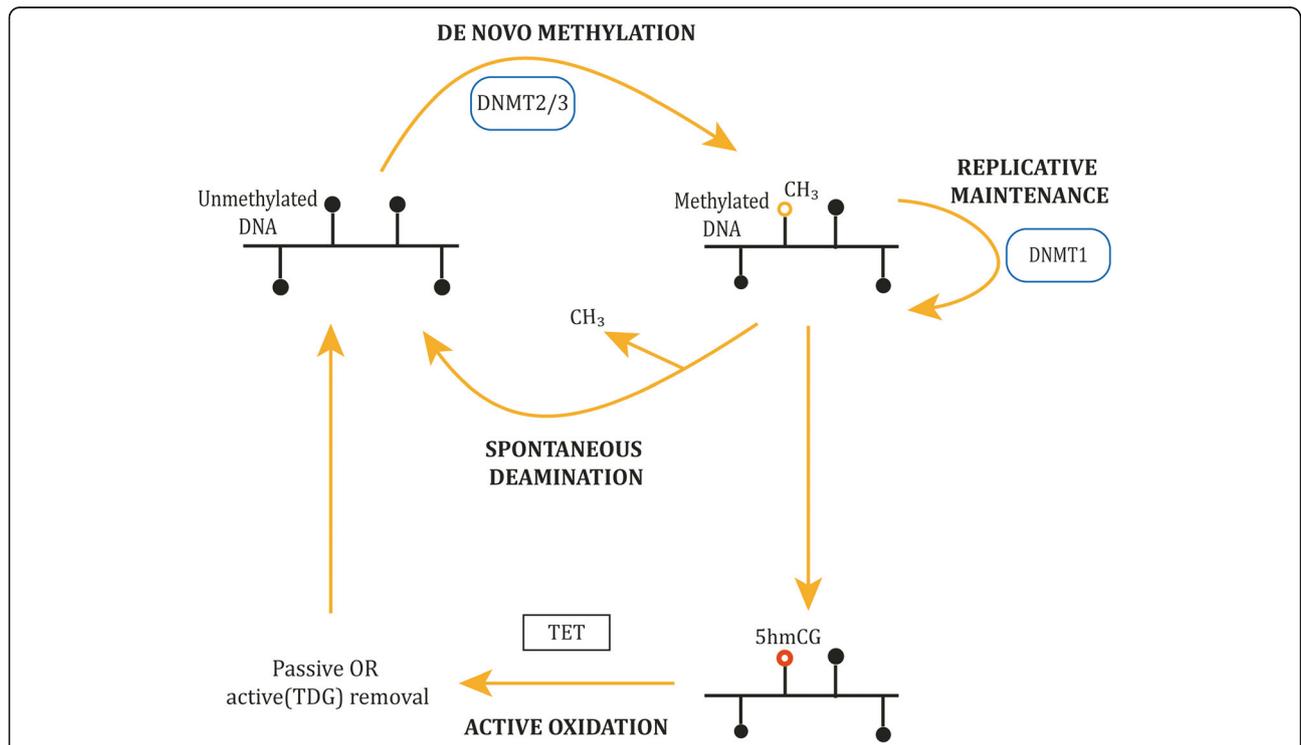


Fig. 1 The process of DNA methylation addition, maintenance and removal. The de novo establishment of DNA methylation is carried out by DNMT2, DNMT3A, DNMT3B and DNMT3L. Once established, DNA methylation requires maintenance to prevent loss of methylation either spontaneously through ‘passive’/spontaneous deamination or actively by the ten-eleven translocation (TET) enzymes. The TET protein family directly remove DNA methylation markers through successive oxidation steps followed by removal of thymine DNA glycosylase (TDG)

out by DNMT2, DNMT3A, DNMT3B and DNMT3-like (DNMT3L). Once established, DNA methylation requires maintenance to prevent loss of methylation either spontaneously through 'passive' deamination or through active deamination by the recently discovered ten-eleven translocation (TET) enzymes. Deamination is the spontaneous loss of an amino group. Methylated cytosines deaminate to thymine, a stable alternative nucleotide, resulting in the gradual depletion of cytosine in the genome. Cytosine to thymine mutations are the most common mutations observed in mammals. The asymmetrically methylated paired bases resulting from deamination are preferentially recognised and methylated by DNMT1, which also maintains DNA methylation during cellular replication.

Initially considered a 'switch' for gene activation or silencing, the exact effect of DNA methylation on gene expression is highly context-dependent, with promoter DNA methylation associated with gene silencing, whereas DNA methylation in the effector region is associated with gene activation. DNA methylation has been investigated in disease for many purposes as follows: as a potential biomarker for disease outcome/severity, to elucidate transcriptional control and to determine the effect of genetic changes on function. Local DNA sequence is the primary determinant of DNA methylation state [9]. Disease-associated SNPs can therefore alter DNA methylation patterns to affect gene expression and cellular function, either in *cis* (at the gene itself), or *trans* (indirectly, often distantly).

Factors influencing DNA methylation

Analysis of DNA methylation in disease greatly benefits from an experimental design that considers the intrinsic and extrinsic factors that affect DNA methylation. We highlight below some of the well-established and identifiable influencers of DNA methylation.

Age

Embryonic development is a crucial period for the establishment of DNA methylation patterns, and prenatal exposure to stressful environmental factors can alter DNA methylation patterns in adulthood (some sex-specific) [10, 11]. Subsequent age-related changes are observed in all cell and tissue types. These changes are indicative of biological ageing to a greater extent than chronological ageing and are related to both cellular intrinsic factors as well as changes in cell composition in whole blood [12, 13].

Sex

Whilst differences in DNA methylation of sex chromosomes and sex-related genes are well characterised, identifying differences in DNA methylation of other autosomal sites between sexes has been inconsistent [14]. Further

research is needed to confirm these findings and to determine if these differences persist to adulthood.

Smoking

Smoking markedly influences DNA methylation. Smoking-related changes in DNA methylation are correlated with smoking severity, whilst some CpG sites revert to never smoking DNA methylation levels within 12 weeks after smoking cessation and others persist after 10 years [15]. These changes are cell type-specific [16]. Second-hand smoke exposure affects DNA methylation similarly to direct exposure [17]. Emerging data indicates e-cigarettes, or vaping, also alter epigenetic and transcriptional regulation [18].

Medication

Many medications used in the treatment of inflammatory diseases influence DNA methylation, including prescribed and 'complementary' medications. For example, glucocorticoids, non-steroidal anti-inflammatory drugs (NSAIDs), sulfasalazine, methotrexate and tumour necrosis factor- α inhibitors all alter DNA methylation [19, 20].

Alcohol and diet

Alcohol is known to alter DNA methylation even at low levels (1 drink/week); however, the persistence and breadth of its effect is poorly understood [21]. Dietary factors can also affect DNA methylation, although this has predominantly been examined in obesity or cancer intervention studies. LINE (long interspersed nuclear element), a surrogate for global DNA methylation levels, has been indicated as a general diet response marker [22]. A single study was able to link levels of LINE methylation to serum glucose levels, and the level of hydroxymethylation to BMI, waist circumference and total cholesterol [23]. BMI can alter DNA methylation levels as a consequence of adiposity [24]. Adiposity is associated with increased inflammation markers and can predispose to inflammatory diseases [25].

Sample cell type

Cell types each have unique DNA methylation profiles, and differences in tissue DNA methylation patterns are strongly influenced by the specific mixtures of cell types within tissues [26]. Further, the effect of environmental factors and disease on DNA methylation patterns is cell type-specific. In silico deconvolution methods are available for estimating cell proportions, and many of these methods have been described in a recent review which also provides insight into when different methods should be applied [27]. However, these methods do not enable the identification of cell-type specific changes in DNA

methylation nor control for the dilution of cell specific signals within complex samples.

Measuring DNA methylation

The gold standard method for characterising DNA methylation is bisulfite conversion, which converts unmethylated (but not methylated) cytosine to uracil enabling the identification of methylated regions through sequencing. An additional oxidative reaction can be used to identify hydroxymethylated regions (5hmC) but is rarely used due to additional cost and input requirements. DNA methylation measurement techniques can be broadly categorised as bisulfite based, DNA methylation-sensitive restriction enzyme based, methylcytosine-specific antibody methods, enrichment methods or direct sequencing (see Table 1). Bisulfite methods are favoured for genome-wide studies; however, as most of the genome is incapable of being methylated, this is often paired with enrichment or selection methods.

BeadChip arrays are commonly used due to the high-throughput capability, relative economy per sample per CpG, reproducibility and the inclusion of regions of known function and disease significance. For example, the most recent version Illumina Human MethylationEPIC array covers over 850,000 CpG sites at a cost of approximately US\$480 per sample [28]. The EPIC array incorporates a greater number of CpGs in the open sea, cytosine nucleotide guanine (CNG) sites and enhancers identified in FANTOM5 (functional annotation of the mammalian genome 5). The probes cover approximately 3% of CpG sites in the genome, but more than 70% of the RefSeq (NCBI reference sequence database) identified transcriptional start sites (TSS) 5'UTR and 3'UTR CpG sites [28]. Novel methods for DNA methylation detection using optical or electrochemical biosensing are being developed, alongside single cell technology. Selection of a DNA methylation measurement method should balance the practicalities of cost alongside the coverage and specificity offered.

Analysing DNA methylation data

Genome-wide methylation studies are complex. Investigators will often seek to use pre-existing cohorts for experiments. This is not recommended as statistical methods cannot compensate for complete confounding variables, for example where all cases are smokers and all controls are non-smokers.

Quality control

Data from DNA methylation measurements differ depending on the type of experiment. Array-based methylation can be expressed as either β values, the ratio of methylated probe intensity to overall probe intensity, or M values, the \log_2 of the ratio of methylated to

unmethylated probes. M values are preferable for statistical analysis, whereas β values are suited to displaying data. In sequencing-based experiments, coverage, or tag count, can be used as a measure of methylation.

Batch variation is particularly an issue with highly sensitive methods, such as BeadChip arrays. Whilst *in silico* methods can be used to overcome batch variation [29], it is always preferable to control for batch variation through properly designed experiments and sample randomisation. Recommendations for sequencing coverage vary between methods, largely influenced by the number of nucleotides assayed. Guidelines for coverage have been developed by the NIH Roadmap Epigenomics project and the ENCODE project [30, 31]. The most common method of DNA methylation measurement, Illumina BeadChip arrays, requires specific QC prior to analysis. Illumina provides Genome Studio for processing DNA methylation data, although several free analysis pipelines are available. For enrichment-based high throughput sequencing assays, measuring the CpG enrichment within the captured DNA is a useful QC step. Cut-offs for individual CpGs (or regions) should be applied, as sufficient average coverage does not guarantee sufficient coverage of individual CpGs.

If publically available datasets or previous versions of arrays are to be used, careful QC should be carried out to curate the datasets to samples controlling for phenotypic differences and for plate variation. Illumina BeadChips may be compared due to the high level of similarity between different iterations, but plate variation is always an issue.

DNA methylation analysis

The major approaches to analysing changes in DNA methylation are as individual CpGs (differentially methylated positions (DMPs)) and differentially methylated regions (DMRs). DMR can be impactful due to the close regulation of proximally located CpGs; however, interpreting DMR can be difficult if regions contain both hypo- and hyper-methylated positions, due to the function of specific regions (e.g. effector vs promoter regions). CpGs can be easily defined as increased or decreased due to the binomial distribution, although the increasing number of CpGs being analysed requires a large sample size to achieve statistical significance after multiple testing corrections, such as the Benjamini-Hochberg false discovery rate (FDR).

Whilst it is always preferable to control for measured covariates through experimental design, various *in silico* methods have been developed to identify DMPs whilst controlling for confounding variables. These include linear regression methods incorporating initial tests for confounding variables such as either principal component analysis (PCA) or multi-dimensional scaling (MDS)

Table 1 Summary table of methods for the detection of DNA methylation

Approach	Method	Relative cost	Throughput	Resolution	Advantages	Disadvantages
Bisulfite-based methods	Methylation-specific PCR (MSP)	+	+	Region	<ul style="list-style-type: none"> • Cheap • Easy • Relatively quick 	<ul style="list-style-type: none"> • Single-gene resolution • No high-throughput capability • Amplification-based
	Human MethylationEPIC BeadChip array	++	+++	Base pair/region	<ul style="list-style-type: none"> • High throughput • Targeted to functionally important regions 	<ul style="list-style-type: none"> • Probe variation • Selected regions (biased)
	Single-cell nucleosome methylation transcription sequencing (scNMTseq)	++++	++++	Base pair	<ul style="list-style-type: none"> • Single-cell resolution • Low-input numbers • Nucleosome, epigenetic and transcription from a single cell 	<ul style="list-style-type: none"> • Expensive • Analysis methods complex • Amplification required
	Reduced representation bisulfite sequencing (RRBS)	++	++	Base pair	<ul style="list-style-type: none"> • Covers CpG dense regions 	<ul style="list-style-type: none"> • Specific base sequence selection due to enzymatic cut sites • Cannot distinguish 5mC from 5hmC
	Pyrosequencing	+++	++++	Base pair	<ul style="list-style-type: none"> • Quantitative • High throughput 	<ul style="list-style-type: none"> • Expensive • Non-targeted
Methyl-CpG isolation-based methods	Whole genome bisulfite sequencing (WGBS)	++++	+++++	Base pair	<ul style="list-style-type: none"> • Genome-wide coverage • Sequence/SNP information • A large amount of data 	<ul style="list-style-type: none"> • Expensive • A large amount of data • Large areas that are incapable of being methylated
	Methylated DNA immunoprecipitation (MeDIP)	++	++	Region	<ul style="list-style-type: none"> • Can incorporate with PCR/microarray/NGS • 5mC in dense, less dense and repeat regions are covered • Antibody-based selection is independent of sequence 	<ul style="list-style-type: none"> • Lower base-pair resolution (~ 150 bp) • Potential antibody non-specific interactions • Antibody-based selection biased towards hypermethylated regions • Unmethylated regions can only be interpreted from the absence of signal
	methyl-CpG-binding domain-isolated genome sequencing (MIGS)	++	++	Region	<ul style="list-style-type: none"> • Genome-wide 5mC coverage 	<ul style="list-style-type: none"> • Lower base-pair resolution (~ 150 bp) • No 5hmC coverage • Bias towards hypermethylated regions
	Combined bisulfite conversion and restriction analysis (COBRA)	++	+	Base pair/region	<ul style="list-style-type: none"> • Small input • Targeted approach 	<ul style="list-style-type: none"> • No high-throughput capability
	HpaII tiny fragment enrichment by ligation-mediated PCR (HELP-tagging assay)	++	+	Base pair	<ul style="list-style-type: none"> • Enrichment of areas of interest • Easy 	<ul style="list-style-type: none"> • HpaII only recognises CCGG when the middle cytosine is unmethylated
	Methylation-specific amplification microarray (MSAM)	++	++	Base pair	<ul style="list-style-type: none"> • Broad coverage • Customizable 	<ul style="list-style-type: none"> • Amplification-based • Biased to regions selected • Non-direct measure
Kinetics-based methods	PacBio single-molecule, real-time (SMRT) sequencing	+++	++++	Base pair	<ul style="list-style-type: none"> • Long reads • Bisulfite conversion free 	<ul style="list-style-type: none"> • Expensive • Covers
	Mass spectrometry	++++	+	Base pair	<ul style="list-style-type: none"> • Direct • Amplification-free method 	<ul style="list-style-type: none"> • Expensive • Input requirement high without additional targeted selection
Optical biosensing	Fluorescence resonance emission transfer (FRET)	++++	++	Base pair	<ul style="list-style-type: none"> • Amplification-free method • Capable of miniaturisation 	<ul style="list-style-type: none"> • Expensive • Not user-friendly data analysis packages
	Surface plasmon resonance (SPR)	++++	+++	Base pair	<ul style="list-style-type: none"> • Amplification-free method 	<ul style="list-style-type: none"> • Expensive • Difficult to analyse

Table 1 Summary table of methods for the detection of DNA methylation (*Continued*)

Approach	Method	Relative cost	Throughput	Resolution	Advantages	Disadvantages
Electrochemical biosensing	Graphene or gold affinity methods	+++	+++	Base pair	<ul style="list-style-type: none"> • Bisulfite conversion free • Low-sample input • Amplification-free method • Bisulfite conversion free method • Low-sample input • Rapid throughput 	<ul style="list-style-type: none"> • Complex to run • Less established • Few commercially available

. The recently developed OSCA method was specifically developed for methylation-association studies, employing a mixed linear model-based method to detect DMPs whilst fitting all other probes as random-effect components to account for the effects of confounders [32]. This approach has been shown to control for observed and unobserved confounders better than standard approaches, retaining statistical power whilst avoiding inflation of association findings, and is an important advance in the field.

DNA methylation is co-regulated alongside other epigenetic data, such as gene expression and chromatin accessibility. Therefore, examining DNA methylation within the context of other 'omics data, such as disease variants to identify methylation quantitative trait loci (meQTL), enables greater functional understanding of DNA methylation changes. These datasets can be used to create a comprehensive map of genetic, transcriptional and epigenetic data, and the relationships between each other. Network analysis can then be applied to search for pathways enriched for disease-related changes.

DNA methylation studies in AS, IBD and psoriasis

The current state of DNA methylation studies in AS, IBD and psoriasis provides information on previously identified shared pathways and avenues for future research. As these diseases have been found through genotypic studies to have pleiotropic genes (the same gene with different effects on function), it is more useful to discuss each disease individually and then address overlapping pathways. The key studies are summarised in Table 2.

Ankylosing spondylitis

The first DNA methylation study in AS was published in 2014, and there has been a scarcity of studies since that time. Only a single multigene study has been published with a small cohort of 5 individuals with grade 4 bilateral sacroiliitis, complete fusion of both sacroiliac joints and 5 age- and sex-matched controls [44]. 1915 DMP were identified, the most significant of which was *HLA-DQB1*, a MHC class II responsible for exogenous peptide display, which had previously been associated with AS radiographic severity and age of onset, but not within Han

Chinese individuals. Unfortunately, this study failed to account for *HLA-B*27* status, the strongest AS-associated genetic loci, which is in strong LD with *HLA-DQB1* [45]. Therefore, these findings require further validation.

The single-gene studies have been a mix of genetically associated and broadly inflammatory-related genes: the suppression of cytokine signalling 1 (*SOCS1**), *DNMT1* [42], *B-cell chronic lymphocytic leukaemia/lymphoma 11B (BCL11B)* [43], *IFN regulatory factor 8 (IRF8)* [46] and *IL12B** [47] (*genetically associated with AS). The study on *SOCS1* acknowledged that the differences observed in *SOCS1* in cell-free DNA was likely due to inflammation-driven cellular apoptosis in AS patients. The study of *IL12B* tested the performance of methylation as a biomarker for AS. The area under the curve of the ROC analysis was statistically significant but not of sufficient magnitude to be clinically useful (AUC = 0.65), and no validation study was done. The methylation differences were restricted to HLA-B27 positive cases, suggesting either that the methylation is affected by HLA-B27 status or that the sample size of HLA-B27-negative cases ($n = 16$) was too small. As the HLA-B27 status of the controls was unknown, it is unclear which explanation applies [47]. None of the studies investigated whether the AS-associated SNPs at these loci operated through effects on methylation.

Whilst many of these studies attempted to connect these changes with HLA-B*27 status, none performed HLA-B*27 typing in healthy controls, who generally have a significantly lower prevalence of HLA-B*27. All focused on promoter regions and identified inverse correlations with mRNA levels for the affected gene. Thus, for the majority of the 116 known AS-associated variants, there is no robust information about methylation effects. AS is known to be associated with *DNMT3A*, *DNMT3B* and *DNMT3L* [4], indicating that it is likely that methylation variation is important in AS pathogenesis. Overall, only a sparse investigation of DNA methylation has been performed in AS to date, and those studies have suffered from failures to collect appropriate information at study onset (such as HLA-B*27 status) or to use appropriate sample types.

Table 2 Seminal papers on DNA methylation in IBD, psoriasis and AS. Study design elements and key findings are outlined

Condition	Reference	Study design	Patients	Controls	Tissue	Measurement method	Key findings
Inflammatory bowel disease	Hasler [33]	MZ CC	20 UC discordant MZ twins 50 inflamed UC; 30 non-inflamed UC	50 HC 25 inflamed disease controls 30 non-inflamed disease controls	Intestinal mucosa tissue	Illumina Methylation 27 BeadChip Custom tiling array; pyrosequencing	Integrated DEG data, MeDIP-seq data and Illumina Methylation 450K data was used to identify 61 IBD-associated loci harbouring DMP in cis of a DEG. All were novel candidate risk loci for IBD including SPINK4, THY1 and CFI.
	McDermott [34]	CC	150 IBD patients; 24 paediatric IBD	40 HC; 22 paediatric HC	PBMC; colonic mucosa tissue	Illumina Methylation 450K BeadChip; pyrosequencing	3196 probes were significantly differentiated between CD patients and healthy controls, and 1418 probes between UC patients and controls. The most significant DMP for both groups was in the 5'UTR of TIFAB.
	Ventham [35]	CC	240 IBD patients	74 symptomatic HC 117 HC	Whole blood, CD4+ T cells, CD8+ T cells, CD14+ monocytes	Illumina Methylation 450K BeadChip	Samples clustered by cell type separately from PBMCs. The top DMP in PBMCs was driven by changes in CD14+ monocytes, not either T cell subset. Conversely, some significant DMP in cell subsets were not detectable in whole PBMC. DMP were significantly associated with known IBD-risk loci.
	Howell [36]	CC	66 paediatric IBD	30 paediatric HC	Bowel mucosal biopsies	Illumina Methylation 450K or EPIC BeadChip; pyrosequencing	Studied DNA methylation, gene expression and gut microbiota from a single cohort at multiple intestinal sites. Site-specific signatures were observed for DNA methylation and gene expression in the ascending colon and sigmoid colon compared to the terminal ileum.
	Somineni [37]	CC	164 paediatric CD	74 paediatric HC	Peripheral blood	Illumina Methylation EPIC BeadChip	Previous findings were replicated including <i>VMP1</i> , <i>RP56K2</i> , <i>ITGB2</i> , and <i>TXK</i> , alongside known therapeutic targets (<i>TNF</i> , <i>JAK3</i> , <i>IL12B</i> , <i>IL23A</i> , and <i>IL1R1</i>). 194 DMP had a strong evidence of genetic effect. The Crohn's methylation signature was a powerful predictor of disease (AUC = 0.91), but had no prognostic utility.
Psoriasis	Gerwin [38]	MZ	27 MZ pairs discordant for psoriasis	27 HC	CD4+ cell and CD8+ cell from PBMCs	Illumina Methylation 27 BeadChip	DNA methylation highly correlated between monozygotic twins in CD4+ T cells and CD8+ T cells. No differences were identified between individual CpG sites or overall methylation per gene. When combined with gene expression data, cell-specific differences were identified, including <i>IL13</i> , <i>ALOX5AP</i> , <i>PTH1L</i> and <i>TNFSF11</i> .
	Zhou (1) [39]	CC	114 psoriasis patients	62 HC	Skin biopsies PP and PN	Illumina Methylation 450K BeadChip; Sequenom Epityper system	129 SNP-CpG pairs achieved statistical significance between psoriatic and healthy control peripheral blood, constituting 28 unique meQTLs and 34 unique CpGs. 11 SNP-CpG pairs passed CIT constituting 3 unique CpG sites within C1orf106, DMBX1 and SIK3.
	Zhou (2) [40]	CC	114 Psoriasis patients (41 PP+PN)	62 HC	skin biopsies; Previous methylation data	Illumina Methylation 450K BeadChip	1514 DMP were identified between PP and PN, only 426 DMP were identified between PP and PN, and none between PN and NN. During replication, 9 sites reached significance (<i>CYP251</i> , <i>ECE1</i> , <i>EIF2C2</i> , <i>MAAM1C1</i> , <i>DLGAP4</i> , <i>S100A9</i> , <i>S100A7A</i> and <i>S100A5</i>), but none replicated in PBMCs.
	Pollock [41]	CC	23 psoriasis patients without PsA; 13 PsA individuals	18 HC	Sperm	Illumina Infinium Human Methylation 450K BeadChip	2467 DMP identified between PsA and healthy controls, compared to 574 DMP between psoriasis and controls. DMR were enriched for the MHC complex. <i>IL22</i> was associated with joint and skin involvement and was the only site to have correlated methylation levels between sperm and blood.
Ankylosing spondylitis	Aslani [42]	CC	40 AS	40 HC	PBMC	MSP	DNMT1 promoter was hypermethylated in cases compared to controls. No significant difference between HLA-B*27+ and HLA-B*27- patients, or disease activity scores.
	Karami [43]	CC	50 AS	50 HC	PBMC	MSP	Cases had decreased <i>BCL11B</i> expression associated with increased

Table 2 Seminal papers on DNA methylation in IBD, psoriasis and AS. Study design elements and key findings are outlined (Continued)

Condition	Reference	Study design	Patients	Controls	Tissue	Measurement method	Key findings
	Hao [44]	CC	10 AS	10 HC	PBMC	Illumina Methylation 450K BeadChip	methylation in the promoter region. 1915 CpGs were identified as differentially methylated between cases and controls. The most significant was HLA-DQB1, previously associated with AS radiographic severity and age of onset. Increased DNA methylation associated with decreased HLA-DQB1 expression.

Inflammatory bowel disease

DNA methylation research in IBD is the most robust of the three diseases, with over 30 papers published in the last decade. There is strong indirect evidence that methylation effects are likely to play a significant role in IBD pathogenesis, with 33% of the heritability of ulcerative colitis, and 30.5% of the heritability of Crohn's disease, being associated with SNPs affecting methylation levels (mQTL) [48]. IBD studies have had a strong focus on DNA methylation as a biomarker for progression to colorectal cancer [49–51]. In this review, we will be focusing on those studies that investigate methylation involvement in the pathogenesis of IBD itself. The differences between the IBD subsets, ulcerative colitis and Crohn's disease have been heavily studied. Consistent with the substantial genetic overlap between these subsets, multigene DNA methylation studies ascertained overlap in the differentially methylated genes in both subsets [34, 35, 52–58].

Multigene DNA methylation studies in IBD have been highly variable in terms of the genes observed. Similar pathways have been identified including those related to tissue/skeletal morphogenesis (e.g. *PRICKLE1*, *SOX11*, *TGFBR3*, *NKX2-3*) [54, 56, 58, 59], immune pathways such as the Wnt/NF- κ B (e.g. *PITX2*, *RARB*, *ROR1*, *FOXA2*) [58, 60], IL-23 pathways (e.g. *STAT3*, *BCL3*, *OSM*, *TLR4*) [53, 61] and inflammation-associated genes (e.g. *ITGB2*, *SAA1*, *IFITM1*, *ITGB2*) [54, 55, 58, 62, 63].

Methylation studies in IBD are challenged by the accessibility of tissues, variability of disease treatment and obtainment of appropriate controls. Medication usage in IBD treatment varies to a greater extent than those used in psoriasis and AS, complicating analysis and potentially explaining the failure for most studies to consider it as a covariate. Uninflamed and healthy control biopsies are often obtained from different gastrointestinal regions to the inflamed samples. This introduces variation based on the location of samples which can confound disease associations with location differences. Howell et al. specifically examined this issue and identified site-specific signatures for DNA methylation and gene expression between the ascending colon and sigmoid colon compared to the terminal ileum [36]. Signatures in each region were also found to be disease subset specific, which agrees with the differential distribution of each subset within the gastrointestinal system. This variability is likely also linked to the different cell types within each region. Ventham et al. examined both whole blood and isolated immune cell subsets (CD4+ T cells, CD8+ T cells and CD14+ monocytes) [35]. Gene expression and DNA methylation were highly clustered by cell type in the principal component analysis (PCA), and all individual cell types clustered separately from whole blood. The signals identified in whole blood were often diluted cell-

specific signals, as seen with *RPS6KA2*, the top DMP in whole blood, which was only significantly differentially methylation in CD14+ monocytes. Conversely, *HDAC4* which was also significantly differentially methylated in CD14+ monocytes was not significantly different in whole blood. Ventham et al. further demonstrated that IBD-associated DMP co-localised with known IBD-associated GWAS loci, significantly more than randomly generated bins with similar probe density [35]. These IBD-associated DMPs were used to identify 326 cis meQTLs, with two SNPs associated with *VMPI* both in linkage disequilibrium with a known IBD-susceptibility allele (rs1292053). This study used causal interference testing (CIT) to investigate if DNA methylation was a mediator between *VMPI* genotype and phenotype or if this was independent or consequential to genotype [35, 64]. The authors cited insufficient sample size as a likely reason for the inability to prove the relationship, an issue echoed by many papers in all three diseases. As this is the largest DNA methylation study in IBD, psoriasis or AS, it indicates the need to reassess the current sample sizes considered for DNA methylation.

Somineni et al. recently reported on methylation differences in peripheral blood mononuclear cells in 164 treatment-naive paediatric IBD cases and 74 non-IBD controls [37]. They found 1189 differentially methylated sites (false discovery rate < 0.05) and replicated Ventham's previous findings at *VMPI*, *SBNO2*, *RPS6KA2*, *ITGB2* and *TXK*. Of these, 194 showed genetic effects on the change in methylation, and using Mendelian randomisation approaches, they found evidence to suggest that three of these, two involving the gene *GPR31* and one involving *RNASET2*, showed evidence of a causal relationship with Crohn's disease, replicating a prior finding at *RNASET2* [65]. Their Crohn's disease methylation signature overlapped strongly with previously reported signatures associated with the chronic low-grade inflammatory disease as assessed by serum CRP measured in subjects with a broad range of diseases including cardiovascular disease and diabetes. Finally, they demonstrated that the methylation signature was a powerful discriminator between cases and controls (AUC = 0.91), but had no utility in predicting prognosis. Howell et al. similarly found that methylation signatures had the high discriminatory capacity in paediatric IBD, with sigmoid colon methylation having AUC = 0.94 [36]. Whether this signature is as predictive in subjects with suppressed disease, or in comparison with other forms of colitis, and the extent of overlap in the methylation signatures these two publications employed, awaits further study.

The link between IBD-associated genotypes and DNA methylation has also been investigated in several single-gene studies. The ulcerative colitis severity region-associated rs1861494 T allele in *IFNG* was associated

with this *IFNG* promoter methylation loss, and more severe disease outcomes in both ulcerative colitis and Crohn's disease [66]. *NKX2-3* (rs1190140), *IL-17A* (IVS1+18) and *STAT4* (rs7574865) all cause loss or gain of a CpG site [61, 67, 68]. In *STAT4* and *IL17A* the T risk alleles result in loss of CpG sites and associated increased expression of the cytokine in both colonic tissue and PBMCs [68]. These studies have provided insight into the mechanisms through which these SNPs can affect function; however, they are limited in their coverage and only encompass loci associated with coding regions.

Psoriasis

Fewer methylation studies have been performed in psoriasis than IBD. Psoriasis has the benefit of a relatively accessible and easily defined sample types, skin and T cells (known to play a primary role in psoriasis pathogenesis) [38, 69–73]. Additionally, most of the studies in psoriasis have focused on multiple genes, likely enabled by the smaller number of genetic loci associated with psoriasis (~60 loci) [3]. This has resulted in a focus on integrated datasets, such as gene expression and genotyping. As expected, the differentially methylated genes identified in the skin are mostly associated with epidermal and keratinocyte differentiation, and cellular regeneration [39, 74–76]. Four skin biopsy-based studies identified members of the *S100A* family, part of the epidermal differentiation complex, as top differentially methylated hits [39, 40, 75–78]. Many studies have failed to identify differences in skin biopsies in affected non-inflamed skin compared to healthy controls [39, 75, 76]. It is therefore worth considering if the skin-related changes in psoriasis are not causative, but rather responsive to psoriasis pathogenesis. Only Zhou et al. examined both skin punch biopsies and PBMCs, and none of the psoriasis-associated DMP replicated between the sample types [39].

Differentially methylated genes in immune cell types have differed from those associated in skin; however, the pathways identified in both samples encompass cellular adhesion and intracellular signalling pathways and inflammatory pathways (such as the IL-23 and STAT/JAK signalling pathways which are consistent with psoriasis-associated genetic variants (e.g. *IL13*, *ALOX5AP*, *PTHLH* and *TNFSF11*) [38, 70, 75, 76]). This was observed in Zhou et al. (2017) where 3 unique CpG sites were identified with 11 associated SNP-CpG pairs within *C1orf106*, *DMBX1* and *SIK3* [39, 40, 64]. *C1orf106* regulates adherence junction stability, although the key gene at this locus remains unclear [79]. *SIK3* is associated with an mTOR signalling cascade involved in skeletogenesis and is an upstream regulator of HDAC. *DMBX1* is involved in cell cycle regulation. Unfortunately, this was the only study to incorporate both genetic and DNA

methylation analysis, and thus, for most psoriasis-associated loci, there is no information as to whether genetic variants operate through methylation to influence disease. The few single-gene studies have identified differential methylation in the promoter regions of *p15*, *p16*, *p21*, *ID4*, *IFNG* and *HLA-C* [69, 72, 77, 78, 80]. All had associated changes in gene expression. PASI score is associated with increased p16 and *HLA-C* promoter methylation and decreased *HLA-DRB1* promoter methylation [77, 78, 81]. These findings suggest that more comprehensive studies of methylation in psoriasis are likely to be productive.

A novel study examining methylation variation in the sperm in psoriasis (PsC), psoriatic arthritis (PsA) and healthy controls aimed to determine if the parent of origin heritability effects in these diseases were due to methylation effects [41]. Studying the sperm rather than peripheral blood or skin also obviated the issue of cellular heterogeneity. At a false discovery rate <0.05, 574 differentially methylated sites were observed comparing PsC patients and controls, 2467 between PsA patients vs. controls and 342 between PsA and PsC patients. The major histocompatibility complex was enriched for these sites, and there were several strong biological candidates for involvement in these diseases amongst the sites. Whilst several CpG were examined in the blood, only *IL22* had a direct correlation between methylation levels in both tissues. Whether the changes observed in the semen correlate with the changes in cell types likely to be involved in psoriasis will require further study.

Limitations of previous studies

Overall, the studies examining DNA methylation in IBD, psoriasis and AS have suffered from similar issues. Cohort selection has often been driven by practical considerations; however, there has also been an insufficient recording of these factors for in silico analysis or adjustment. Smoking is a known risk factor for these diseases and affects disease severity and treatment response, yet only 5 IBD studies even recorded smoking status. An overarching theme is the insufficient sample sizes used, cited by most papers as interfering with identification of disease-associated signals. This is linked to the ongoing issue of sample type used (tissue or circulating cells). A question which is yet to be addressed is the most appropriate sample type in each disease. The frequency with which these issues are discussed as a potential barrier to analysis indicates that these issues are pervasive in the study design for these diseases.

Best practices

This review of studies in these three related diseases highlights some high-quality publications and also that there are a high number of isolated findings which have

either not replicated, or no replication study has been performed of them. It is clear from this review that for the field to make meaningful contributions to health research, some simple, well-defined best practice guidelines are needed.

Study focus

DNA methylation studies generate a large amount of data and are affected by numerous factors. Therefore, it is important to clearly define an achievable question or goal for this information from the outset. This will also enable the collection of additional functional information, such as gene expression or genotype, where possible.

Cohort selection

Reproducibility and statistical power are best when variations in factors that affect DNA methylation exogenous to the hypothesis being tested are controlled for. The factors outlined above should be considered when planning what criteria will be used for cohort selection and analysis. Age and sex should be matched between case and control cohorts; however, other factors including smoking and medication should be controlled where possible or recorded when not. Medication can be controlled through selecting for treatment naïve individuals, using paired samples pre-and post-treatment, matched case studies or controlling for medication use *in silico* (where appropriate).

Although these factors can be adjusted for *in silico* each factor that has to be adjusted for reduces the power of the study to identify disease-relevant signals.

Therefore, proper cohort selection is always preferable to *in silico* methods.

Sample type

Sample types are often selected due to practical considerations such as accessibility, invasiveness and expense. This has led to a broad use of circulating immune cells. However, a key assumption is that the profiled tissue is the most disease relevant. The primary site of the disease remains contended in these diseases and no robust comparison of circulating and tissue resident cells has been performed, leaving the utility of circulating cells as substitutes for primary disease site largely unknown. Whilst these circulating cells may be relevant to disease, they may not encompass tissue-specific methylation changes relevant to disease. It is therefore imperative to examine cell types in isolation, to enable comparisons between tissue resident and circulating cells, and to control for cell-specific DNA methylation patterns.

Sample size

Due to a large number of statistical tests now being carried out in multi-gene studies, stringent levels of genome-wide significance should be sought. Various cut-offs have been suggested in the literature: 1×10^{-6} [82], 1×10^{-7} [83] and 5×10^{-8} [84]. A 10% change in DNA methylation at a single CpG would require 21 paired case controls, but the same change at genome-wide significance (1×10^{-6}) requires over 100 case-control pairs (Fig. 2). Multiple testing adjustments should be included in these calculations. Sample size requirements should

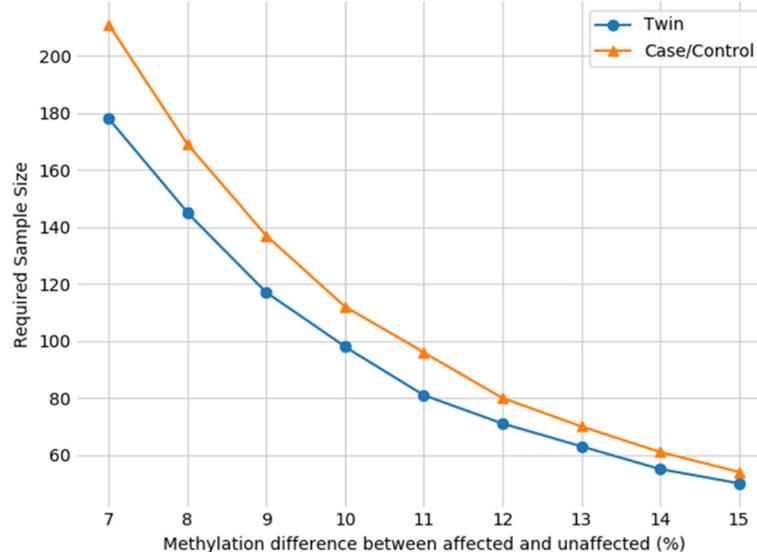


Fig. 2 Sample size requirements for genome-wide significance. Estimated sample sizes, expressed as the number of pairs, either twin pair or case-control, required to reach 80% power in twin and case/control designs using a genome-wide significance threshold of 1×10^{-6} . Data taken from Tsai and Bell 2015

be determined prior to the study outset using either pilot studies or similar cohorts to determine effect size (methylation different between groups).

Method selection

Although over a third of studies in IBD, psoriasis and AS used MSP, it is recommended that this method only be selected where candidate genes have already been identified and the method is being used to investigate the relationship between genotype or gene expression with methylation of that gene. Otherwise, a multigene approach is always preferable. Each method has individual advantages and disadvantages, as outlined in Table 1. It is important to consider the limitations of each method prior to selection. This should then influence the analysis and quality control steps that will be used. If data is to be integrated with another study, care should be taken to select similar cohorts and methods for measurements. Variation due to differences in the factors discussed above should be tested prior to analysis. It may be beneficial to treat these as a validation cohort rather than to integrate such data.

Integration with 'omics data

It is informative to examine DNA methylation in the context of genomic or transcriptomic data as this provides context and functional information. It also provides a method to prioritise DMP. Currently, there has been limited integration of genotype with DNA methylation data. This has made it difficult to identify the relationship between the genetic basis of these diseases and changes at a DNA methylation level. It is also suggested that the integration of microbiota and cytokine data be considered, due to the strong link between gut microbiota changes and all three diseases.

Replication/validation

Technical verification is recommended prior to undertaking biological validation, as it is generally less time consuming and expensive than biological validation. PCR-based techniques are commonly used for technical validation of BeadChip arrays. It is noteworthy that only twin studies or juvenile cohorts carried out biological validation [33, 52, 54]. This is particularly relevant in DNA methylation due to high levels of inter-individual variation and tissue-specific DNA methylation patterns.

Conclusion

Despite evidence in twin studies that epigenetics is a major mechanism through which immune-mediated disease-associated variants affect function, the role of DNA methylation, the most well-characterised epigenetic mechanism, in immune-mediated diseases remains largely unknown. DNA methylation remains an under-investigated biological

mechanism in AS, IBD and psoriasis. Similar pathways have been identified as differentially methylated in each of these diseases, and most have been previously identified through genetic association studies. However, studies to date have been limited by study design issues, including poor cohort selection, improper controls and insufficient statistical power.

Future studies should incorporate the factors outlined in this review and take steps prior to study outset to perform sample size calculations, cohort selection criteria, and the selection of appropriate measurement techniques and analysis pipelines that are based around those techniques. These should be followed by biological and technical validation to ensure that results are held to a robust standard. The use of consistent and appropriate experimental design will enable the identification of disease-relevant changes in DNA methylation and provide functional information to inform rational treatment design in these immune-mediated diseases.

Abbreviations

5hmC: 5-Hydroxymethylcytosine; 5mC: 5-Methylcytosine; AS: Ankylosing spondylitis; BMI: Body mass index; CC: Case-control study; CD: Crohn's disease; CIT: Causal interference testing; CNG: Cytosine nucleotide guanine; COBRA: Combined bisulfite restriction analysis; CpG: Cytosine phosphate guanine-paired bases; DMP: Differentially methylated position; DMR: Differentially methylated region; DNMT: DNA methyltransferase; FANTOM5: Functional annotation of the mammalian genome 5; FDR: False discovery rate; FRET: Fluorescence resonance emission transfer; GWAS: Genome-Wide Association Study; HC: Healthy control(s); HELP-tagging: HpaII tiny fragment enrichment by ligation-mediated PCR; HLA: Human leukocyte antigen; IBD: Inflammatory bowel disease; LINE: Long interspersed nuclear element; MDS: Multi-dimensional scaling; MeDIP-seq: Methylated DNA immunoprecipitation sequencing; meQTL: Methylation quantitative trait loci; MIGS: Methyl-CpG-binding domain-isolated genome sequencing; MSAM: Methylation-specific amplification microarray; MSP: Methylation-specific PCR; MZ: Monozygotic twin study; NN: Normal non-psoriatic samples; NSAIDs: Non-steroidal anti-inflammatory drugs; PBMC: Peripheral blood mononuclear cells; PCA: Principal component analysis; PN: Psoriatic uninvolved samples; PP: Psoriatic involved samples; QC: Quality control; RefSeq: NCBI reference Sequence database; RRBS: Reduced representation bisulfite sequencing; scNMTseq: Single-cell nucleosome, methylation and transcription sequencing; SPR: Surface plasmon resonance; TDG: Thymine DNA glycosylase; TET: Ten-eleven translocation; TSS: Transcriptional start site; UC: Ulcerative colitis; WGBS: Whole genome bisulfite sequencing

Acknowledgements

None applicable

Authors' contributions

JW performed the research for this manuscript and created the tables. JW and JE created the figures. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

Funding

None

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Published online: 03 June 2019

References

- International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet.* 2013;45(7):730–8.
- Liu JZ, Van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47(9):979–86.
- Tsoi LC, Stuart PE, Tian C, et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nature Communications* 2017;8:15382.
- Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet.* 2016;48(5):510–8.
- Tomizawa SI, Kobayashi H, Watanabe T, et al. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development.* 2011;138(5):811–20.
- Van Otterdijk SD, Michels KB. Transgenerational epigenetic inheritance in mammals: how good is the evidence? *FASEB J.* 2016;30(7):2457–65.
- Radford EJ. Exploring the extent and scope of epigenetic inheritance. *Nat Rev Endocrinol.* 2018;14(6):345–55.
- Lin IG, Han L, Taghva A, O'Brien LE, Hsieh CL. Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. *Mol Cell Biol.* 2002;22(3):704–23.
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet.* 2011;43(11):1091–7.
- Murphy SK, Adigun A, Huang Z, et al. Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. *Gene.* 2012;494(1):36–43.
- Tobi EW, Lumey LH, Talens RP, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet.* 2009;18(21):4046–53.
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014;15(2):R31.
- Gervin K, Andreassen BK, Hjørthaug HS, et al. Intra-individual changes in DNA methylation not mediated by cell-type composition are correlated with aging during childhood. *Clin Epigenetics.* 2016;8(1):110.
- McCarthy NS, Melton PE, Cadby G, et al. Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genomics.* 2014;15(1):981.
- Tsaprouni LG, Yang T-P, Bell J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics.* 2014;9(10):1382–96.
- Su D, Wang X, Campbell MR, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLoS One.* 2016; 11(12):e0166486.
- Kim S. Overview of cotinine cutoff values for smoking status classification. *Int J Environ Res Public Health.* 2016;13(12):1236.
- Martin EM, Clapp PW, Rebuli ME, et al. E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke. *Am J Physiol Lung Cell Mol Physiol.* 2016;311(1):L135–L44.
- Wilson LE, Kim S, Xu Z, Harlid S, Sandler DP, Taylor JA. Non-steroidal anti-inflammatory drug use and genomic DNA methylation in blood. *PLoS One.* 2015;10(9):e0138920.
- de Andres MC, Perez-Pampin E, Calaza M, et al. Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. *Arthritis Res Ther.* 2015;17:233.
- Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SRH. The impact of recent alcohol use on genome wide DNA methylation signatures. *Front Genet.* 2012;3:54.
- Zhang FF, Morabia A, Carroll J, et al. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. *J Nutr.* 2011;141(6):1165–71.
- Nicoletti CF, Nonino CB, de Oliveira BAP, et al. DNA methylation and hydroxymethylation levels in relation to two weight loss strategies: energy-restricted diet or bariatric surgery. *Obes Surg.* 2016;26(3):603–11.
- Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature.* 2017;541(7635):81–6.
- Wesseltoft-Rao N, Holven KB, Telle-Hansen VH, et al. Measurements of body fat is associated with markers of inflammation, insulin resistance and lipid levels in both overweight and in lean, healthy subjects. *e-SPEN J.* 2012;7(6): e234–e40.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
- Kaushal A, Zhang H, Karmaus WJJ, et al. Comparison of different cell type correction methods for genome-scale epigenetics studies. *Bmc Bioinformatics.* 2017;18:216.
- Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics.* 2016;8(3):389–99.
- Fortin J-P, Triche T, Hansen K. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array. *bioRxiv.* 2017;33(4):558–60.
- Australo-Anglo-American Spondyloarthritis C, Reveille JD, Sims AM, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet.* 2010;42(2):123–7.
- ENCODE. ENCODE Standards and guidelines for whole genome shotgun bisulfite sequencing (WGBS)2015. <https://www.encodeproject.org/about/experiment-guidelines/> Accessed.
- Zhang F, Chen W, Zhu Z, et al. OSCA: a tool for omic-data-based complex trait analysis. *bioRxiv* 2019:445163.
- Häsler R, Feng Z, Bäckdahl L, et al. A functional methylome map of ulcerative colitis. *Genome Res.* 2012;22(11):2130–7.
- McDermott E, Ryan EJ, Tosetto M, et al. DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. *J Crohn's Colitis.* 2016;10(1):77–86.
- Venthom NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun.* 2016;7:13507.
- Howell KJ, Kraiczky J, Nayak KM, et al. DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. *Gastroenterology.* 2018;154(3):585–98.
- Somineni HK, Venkateswaran S, Kilaru V, et al. Blood-derived DNA methylation signatures of Crohn disease and severity of intestinal inflammation. *Gastroenterology.* 2019;156(8):2254–65 e3.
- Gervin K, Vigeland MD, Matingsdal M, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet.* 2012;8(1):e1002454.
- Zhou F, Wang W, Shen C, et al. Epigenome-wide association analysis identified nine skin DNA methylation loci for psoriasis. *J Invest Dermatol.* 2016;136(4):779–87.
- Zhou F, Shen C, Xu J, et al. Epigenome-wide association data implicates DNA methylation-mediated genetic risk in psoriasis. *Clin Epigenetics.* 2016;8(1):131.
- Pollock RA, Zaman L, Chandran V, Gladman DD. Epigenome-wide analysis of sperm cells identifies IL22 as a possible germ line risk locus for psoriatic arthritis. *PLoS One.* 2019;14(2):e0212043.
- Aslani S, Mahmoudi M, Garshabi M, Jamshidi AR, Karami J, Nicknam MH. Evaluation of DNMT1 gene expression profile and methylation of its promoter region in patients with ankylosing spondylitis. *Clin Rheumatol.* 2016;35(11):2723–31.
- Karami J, Mahmoudi M, Amirzargar A, et al. Promoter hypermethylation of BCL11B gene correlates with downregulation of gene transcription in ankylosing spondylitis patients. *Genes Immun.* 2017;18(3):170–5.
- Hao J, Liu Y, Xu J, et al. Genome-wide DNA methylation profile analysis identifies differentially methylated loci associated with ankylosing spondylitis. *Arthritis Res Ther.* 2017;19(1):177.
- Pimentel-Santos FM, Matos M, Ligeiro D, et al. HLA alleles and HLA-B27 haplotypes associated with susceptibility and severity of ankylosing spondylitis in a Portuguese population. *Tissue Antigens.* 2013;82(6):374–9.
- Chen M, Wu M, Hu X, et al. Ankylosing spondylitis is associated with aberrant DNA methylation of IFN regulatory factor 8 gene promoter region. *Clin Rheumatol.* 2019. <https://doi.org/10.1007/s10067-019-04505-5>.
- Zhang X, Lu J, Pan Z, et al. DNA methylation and transcriptome signature of the IL12B gene in ankylosing spondylitis. *Int Immunopharmacol.* 2019;71:109–14.

48. McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 replicated DNA methylation QTL. *Sci Rep.* 2018;8(1):17605.
49. Papadia C, Louwagie J, Del Rio P, et al. FOXE1 and SYNE1 genes hypermethylation panel as promising biomarker in colitis-associated colorectal neoplasia. *Inflamm Bowel Dis.* 2014;20(2):271–7.
50. Tominaga K, Fujii S, Mukawa K, et al. Prediction of colorectal neoplasia by quantitative methylation analysis of estrogen receptor gene in nonneoplastic epithelium from patients with ulcerative colitis. *Clin Cancer Res.* 2005;11(24):8880–5.
51. Konishi K, Shen L, Wang S, Meltzer SJ, Harpaz N, Issa JJ. Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias. *Gastroenterology.* 2007;132(4):1254–60.
52. Harris RA, Nagy-Szakal D, Pedersen N, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis.* 2012;18(12):2334–41.
53. Lin Z, Hegarty JP, Yu W, et al. Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Dig Dis Sci.* 2012;57(12):3145–53.
54. Harris RA, Nagy-Szakal D, Mir SAV, et al. DNA methylation-associated colonic mucosal immune and defense responses in treatment-naïve pediatric ulcerative colitis. *Epigenetics.* 2014;9(8):1131–7.
55. Karatzas PS, Mantzaris GJ, Safioleas M, Gazouli M. DNA methylation profile of genes involved in inflammation and autoimmunity in inflammatory bowel disease. *Medicine (United States).* 2014;93(28):e309.
56. Kang K, Bae JH, Han K, Kim ES, Kim TO, Yi JM. A genome-wide methylation approach identifies a new hypermethylated gene panel in ulcerative colitis. *Int J Mol Sci.* 2016;17(8):1291.
57. Kraiczky J, Nayak K, Ross A, et al. Assessing DNA methylation in the developing human intestinal epithelium: potential link to inflammatory bowel disease. *Mucosal Immunol.* 2016;9(3):647–58.
58. Sadler T, Bhasin JM, Xu Y, et al. Genome-wide analysis of DNA methylation and gene expression defines molecular characteristics of Crohn's disease-associated fibrosis. *Clin Epigenetics.* 2016;8(1):30.
59. Cooke J, Zhang H, Greger L, et al. Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflamm Bowel Dis.* 2012;18(11):2128–37.
60. Mokarram P, Kavousipour S, Sarabi MM, et al. MGMT-B gene promoter hypermethylation in patients with inflammatory bowel disease - a novel finding. *Asian Pac J Cancer Prev.* 2015;16(5):1945–52.
61. Kim SW, Kim ES, Moon CM, et al. Genetic polymorphisms of IL-23R and IL-17A and novel insights into their associations with inflammatory bowel disease. *Gut.* 2011;60(11):1527–36.
62. Smith PJ, Levine AP, Dunne J, et al. Mucosal transcriptomics implicates under expression of BRINP3 in the pathogenesis of ulcerative colitis. *Inflamm Bowel Dis.* 2014;20(10):1802–12.
63. Li Yim AYW, Duijvis NW, Zhao J, et al. Peripheral blood methylation profiling of female Crohn's disease patients. *Clin Epigenetics.* 2016;8(1):65.
64. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet.* 2009;10:23.
65. Hannon E, Weedon M, Bray N, O'Donovan M, Mill J. Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. *Am J Hum Genet.* 2017;100(6):954–9.
66. Gonsky R, Deem RL, Landers CJ, Haritunians T, Yang S, Targan SR. IFNG rs1861494 polymorphism is associated with IBD disease severity and functional changes in both IFNG methylation and protein secretion. *Inflamm Bowel Dis.* 2014;20(10):1794–801.
67. John G, Hegarty JP, Yu W, et al. NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol Genet Metab.* 2011;104(1–2):174–9.
68. Kim SW, Kim ES, Moon CM, Kim TI, Kim WH, Cheon JH. Abnormal genetic and epigenetic changes in signal transducer and activator of transcription 4 in the pathogenesis of inflammatory bowel diseases. *Dig Dis Sci.* 2012;57(10):2600–7.
69. Brandt D, Serгон M, Abraham S, Mäbert K, Hedrich CM. TCR+ CD3+ CD4– CD8– effector T cells in psoriasis. *Clin Immunol.* 2017;181:51–9.
70. Duan Z, Pan F, Zeng Z, et al. Interleukin-23 receptor genetic polymorphisms and ankylosing spondylitis susceptibility: a meta-analysis. *Rheumatol Int.* 2012;32(5):1209–14.
71. Park GT, Han J, Park SG, Kim S, Kim TY. DNA methylation analysis of CD4+ T cells in patients with psoriasis. *Arch Dermatol Res.* 2014;306(3):259–68.
72. Zhang K, Zhang R, Li X, Yin G, Niu X. Promoter methylation status of p15 and p21 genes in HPP-CFCs of bone marrow of patients with psoriasis. *Eur J Dermatol.* 2009;19(2):141–6.
73. Zhang K, Zhang R, Li X, Yin G, Niu X, Hou R. The mRNA expression and promoter methylation status of the p16 gene in colony-forming cells with high proliferative potential in patients with psoriasis. *Clin Exp Dermatol.* 2007;32(6):702–8.
74. Hou R, Yin G, An P, et al. DNA methylation of dermal MSCs in psoriasis: identification of epigenetically dysregulated genes. *J Dermatol Sci.* 2013;72(2):103–9.
75. Roberson EDO, Liu Y, Ryan C, et al. A subset of methylated CpG sites differentiate psoriatic from normal skin. *J Invest Dermatol.* 2012;132(3 PART 1):583–92.
76. Zhang P, Zhao M, Liang G, et al. Whole-genome DNA methylation in skin lesions from patients with psoriasis vulgaris. *J Autoimmun.* 2013;41:17–24.
77. Chen M, Wang Y, Yao X, et al. Hypermethylation of HLA-C may be an epigenetic marker in psoriasis. *J Dermatol Sci.* 2016;83(1):10–6.
78. Zong W, Ge Y, Han Y, Yang X, Li Q, Chen M. Hypomethylation of HLA-DRB1 and its clinical significance in psoriasis. *Oncotarget.* 2017;8(7):12323–32.
79. Mohanan V, Nakata T, Desch AN, et al. C1orf106 is a colitis risk gene that regulates stability of epithelial adherens junctions. *Science.* 2018;359(6380):1161–6.
80. Ruchusatsawat K, Wongpiyabovorn J, Protjaroen P, et al. Parakeratosis in skin is associated with loss of inhibitor of differentiation 4 via promoter methylation. *Hum Pathol.* 2011;42(12):1878–87.
81. Chen M, Chen ZQ, Cui PG, et al. The methylation pattern of p16INK4a gene promoter in psoriatic epidermis and its clinical significance. *Br J Dermatol.* 2008;158(5):987–93.
82. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol.* 2015;44(4):1429–41.
83. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* 2015;16(1):37.
84. Lu AT, Hannon E, Levine ME, et al. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun.* 2017;8:15353.

Publisher's Note

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