Epigenetics and methylation in the rheumatic diseases

Flore Zufferey, MD, Frances M.K. Williams, BSc, MBBS, PHD, FRCP(E), Tim D. Spector, MBBS, MSc, FRCP, MD*

A R T I C L E   I N F O

Keywords:
- Epigenetics
- Methylation
- EWAS
- Autoimmune diseases
- Rheumatic diseases

A B S T R A C T

Objectives: Rheumatic diseases encompass a wide range of conditions of poorly characterized etiopathology, many having both genetic and environmental susceptibility factors. Epigenetic studies are providing new insights into disease pathogenesis. Recent rheumatology literature related to DNA methylation studies—both epigenome-wide and candidate gene—are discussed, as well as methodological issues.

Method: A PubMed search for articles published until April 2013 was conducted using the following keywords: (“methylation” OR “epigenetics”) AND (“rheumatoid arthritis” OR “lupus” OR “autoimmune disease” OR “osteoarthritis” OR “ostesoporosis” OR “osteoarthritis” OR “musculoskeletal disorder”) and EWAS. The reference lists of identified articles were searched for further articles.

Results: Several genome-wide methylation studies have been reported recently, mostly in autoimmune rheumatic diseases. Overall, these studies have identified methylation signatures in disease, clustering of subgroups as well as new and known epigenetic associations. Methodological issues, small sample sizes and reduced coverage of methylation assays render many results preliminary.

Conclusions: There have been a number of epigenetic advances in rheumatic diseases recently. The new technologies and emerging field of epigenome-wide association will provide novel perspectives in disease etiology, diagnosis, classification, and therapy.

© 2013 Elsevier Inc. All rights reserved.

Introduction

The field of epigenetics in medicine has grown exponentially over the last few years, benefiting from the advanced application of high-throughput technologies. Epigenetics has however been harder to define with any consensus [1,2], not least because of the wide variety of regulatory mechanisms including histone modifications, chromatin folding, DNA methylation (DNAm), and miRNA expression (micro-RNAs). Among these, DNA methylation is currently the most well studied because of the availability of suitable technology. DNA methylation forms the subject of this review—for other epigenetic mechanisms in the field, see reviews [3–6].

Rheumatic disease encompasses a wide range of different conditions, from autoimmune, inflammatory diseases to degenerative and mechanical conditions. Disease classification relies mainly on the evolving clinical, radiologic, and biological composite criteria because the etiopathology of few such diseases has been elucidated fully. Although twin and family studies have established genetic risk factors for most, if not all the conditions, genetic variants identified so far explain a small proportion of the total phenotypic variance. Epigenetics has great promise, therefore, in explaining and redefining diseases in the field as well as accounting for the role of some of the environmental factors.

In this review, we aim to provide clinicians with an overview of the current methodological concepts, challenges, and issues in methylation analysis. We present results of methylation studies in the context of rheumatic diseases and discuss future perspectives.

Methylation architecture and distribution

The genome sequence is composed of four nucleotides—A (adenosine), T (thymine), C (cytosine), and G (guanine). DNA methylation (DNAm) refers to the addition of a methyl group to cytosine at the carbon 5 position (5-methylcytosine, 5mC). Other types of methylation marks exist such as hydroxymethylcytosine (5hmC), which has been identified as a modification of developmental importance at least in early life [7]. The global distribution of cytosine modifications across the genome is referred to as the DNA methylome.

In humans, methylation occurs mainly when cytosine is followed by a guanine on the same strand of DNA linked by a phosphate group (called CpGs or CpG dinucleotides), although several other types of methylation, e.g., at CpHpG (H = A, T, or C),
exist as well [8]. Overall distribution of methylation in the genome is bimodal; the majority of CpGs across the genome are methylated (70–80%), whereas a high density of CpGs enriched in the promoter region of the genes (CpG islands) are generally unmethylated [9]. It is therefore challenging to summarize these changes over the genome.

Methylation properties and functions in human biology and disease

Methylation plays a key role in physiologic conditions, and alterations in methylation regulation have been identified in pathological processes [10,11]. Current knowledge indicates that the function of methylation and its relationship to gene expression varies according to the position in the genome as well as in regions with poor or rich CpG content [12,13]. High levels of 5mC in CpG-rich promoter regions are usually strongly associated with transcriptional repression [14], whereas gene body methylation may not be and shows a more context-dependant relationship [15]. DMAs does not act alone on gene regulation but is closely interconnected with other genetic and epigenetic mechanisms (e.g., transcription factors and histone-modifying enzymes), acting together on chromatin structure to determine the state of gene accessibility. Methylation is thought to reinforce the maintenance of closed chromatin through subsequent cell divisions and mediate long-lasting changes in gene expression [13]. The aforementioned regulatory mechanisms rely on two central properties of DNAm: stability (mitotic inheritance) and plasticity. Plasticity refers to the constant adaptation of the methylome to specific cell regulatory processes in response to internal and external stimuli.

Physiological variations in the methylome

Gametogenesis and pre-implantation zygotic phases are both associated with extensive methylation resetting and reprogramming [16]. Later, in post-zygotic somatic development, most blocks of genomic methylation patterns are static across tissues and throughout life, changing only in local context as specific cellular processes are activated or repressed [13]. These include, for example, mechanisms sustaining cellular fate and differentiation, responsible for specific methylome signatures across different cell types (cell-specific methylome) [17]. Aging—which plays a role in many rheumatic diseases—seems also to be associated with global reproducible methylation changes with a tendency towards hypermethylation [18].

Epigenetic epidemiology and human diseases

Parental contributions of imprinted genes are currently the clearest examples of how human disease may result from epigenetic deregulation [19]. Examples of monogenic epigenetic diseases are congenital syndromes such as Beckwith–Wiedemann or Prader–Willi/Angelman syndromes. These are recognizable syndromes associated with congenital abnormalities and/or developmental delay. These syndromes are due to the consequence of the loss of normal imprinted gene regulation in specific subdomains due to different underlying genetic mechanisms (microdeletion and uniparental disomy) [19].

In common diseases, it has been suggested more recently that each individual has many “epigenomes” and that these may play a central role in disease pathogenesis [19,20]. Although there is some evidence of epigenetic change influenced by the environment [21,22], such as smoking in humans [23], it remains largely unknown as to precisely how the environment triggers alterations in the epigenome leading to disease susceptibility [24]. There is also accumulating evidence for genetic influence on DNAm [25–30], which may mediate disease susceptibility.

Disease modeling using concordance rate between twin pairs or family members helps tease apart the respective contributions of genetics and environment in disease variance. Heritability is an estimate of how much of the variance of a trait may be attributed to genetic factors. Despite known heritable influence in the majority of rheumatic diseases (Table 1), concordance rates between identical (MZ) twin pairs are frequently surprisingly low (e.g., 12–15% in rheumatoid arthritis) [31]. This suggests that a variety of genetic and non-genetic factors as well as their complex interactions contribute to disease susceptibility. Among these, epigenetics has recently emerged as a potential explanatory element as signals vary within MZ pairs [32]. This hypothesis is for example sustained for many rheumatic diseases by the observation of gender bias, geographic differential distribution, remission and relapse as well as age-related prevalence but also decline of clinical symptoms with age, all of which are considered good epidemiological, clinical, and etiopathogenetic indicators of epigenetic properties [20].

Epigenome-wide association studies (EWAS)

Epigenome-wide association studies (EWAS), which is related to the better known genome-wide association study (GWAS), aims to identify epigenetic marks associated with disease by performing a hypothesis-free testing across the whole genome [33]. Before discussing the application of EWAS in rheumatic diseases, it is important to consider the current methodological aspects to allow interpretation of the results obtained so far [34]. (Fig.) depicts a flowchart on how EWAS are carried out and the following paragraphs discuss the different points in more details.

Research question

Results of epigenetic studies provide insight into disease mechanisms or potential biomarkers. Indeed, although challenging to detect [35], in contrast to genetic polymorphisms, plastic DNAm properties may help identify much-needed biomarkers for disease prognosis, monitoring progression, or predicting response to treatment with potential direct translation into the clinics. Comparison of patients with positive specific autoimmune antibodies but free of clinical symptoms versus clinically affected or in responders versus non-responders for a specific medication may help identify such biomarkers.

Table 1: Disease concordance rate within MZ twin pairs and heritability estimates—the proportion of phenotypic variance attributed to additive genetic factors—in several rheumatic diseases based on twin studies

<table>
<thead>
<tr>
<th>Disease/trait</th>
<th>MZ concordance rate (%)</th>
<th>Heritability estimates (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>12–15</td>
<td>65</td>
<td>[31]</td>
</tr>
<tr>
<td>SLE</td>
<td>24–56</td>
<td>69</td>
<td>[76,77]</td>
</tr>
<tr>
<td>OA</td>
<td>31–50</td>
<td>50</td>
<td>[78–80]</td>
</tr>
<tr>
<td>Osteoporotic fractures</td>
<td>13–19</td>
<td>27</td>
<td>[81]</td>
</tr>
<tr>
<td>Hyperuricemia</td>
<td>53</td>
<td>49</td>
<td>[74]</td>
</tr>
<tr>
<td>Gout</td>
<td>42</td>
<td>0*</td>
<td>[74]</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>4–32</td>
<td>66</td>
<td>[82]</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>10</td>
<td>—</td>
<td>[83]</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
<td>40–75</td>
<td>94</td>
<td>[84,85]</td>
</tr>
</tbody>
</table>

RA = rheumatoid arthritis; SLE = systemic lupus erythematosus; OA = osteoarthritis.

* 95% confidence interval = 0–61.8%.
Study design

The two most frequently used population-based study designs in EWAS are case-control and MZ discordant twin pairs [33]. The former comprises two groups of unrelated individuals—generally a healthy control group and a case group affected by a disease of interest drawn from the same population. Ideally cases and controls should be matched for age, sex, ethnicity, and other potential confounders [23]. The discordant MZ twin design uses a different approach. Methylation patterns between MZ twin pairs discordant for a disease are compared within each pair. In this situation, there is internal control of genetic and age/cohort variation. In addition, many other factors such as upbringing and social factors are closely matched [36].

Population selection: power and sample size

Power calculation in EWAS is a matter of on-going research and debate. It remains somewhat speculative as the real effect size of methylation differences is not completely known [36]. However, most findings so far have shown that effect size differences are small, at about 5%. Based on the very scarce current literature, around 15–20 MZ discordant twin pairs and several hundreds of cases and controls would be necessary to reach sufficient power [33,37] under specific experimental and study conditions. The discovery sample would also then need replication in a larger independent replication sample. However epigenome-wide calculations are not possible with any degree of accuracy as many known and unknown variables influence power calculation in EWAS that vary across studies depending on factors such as the study design, the coverage of the methylation assay and its variability, confounders, and the trait under consideration [36].

Sample collection

Choice of biological sample

The choice of material sample for DNA extraction will depend on the trait of interest and should be of developmental or functional relevance. More than one cell or tissue of interest may be relevant, for example, in RA, where both fibroblast-like synovio-cytes (FLS) and T cells have been shown to be involved in the pathogenesis [38]. As methylation is determined across different cells from a sample tissue or organ and represents an average measurement across the sampled DNA, cellular heterogeneity may confound the results. This is particularly pertinent for venous whole blood, which although highly accessible, shows great cellular heterogeneity [39]. Moreover, heterogeneity of cell composition may be responsible for systematic bias between cases and controls. Neutropenia which may be drug induced or autoimmune-related as in Felty’s syndrome may be such an example. Saliva samples for DNA have proved problematic as they contain heterogeneous cell composition, like blood, but which can...
vary within the same day, adding considerably to the noise in the analysis.

**Temporality–causality**

Methylation changes found associated with a condition may take place prior to the phenotype but not necessarily be causally linked. In addition, methylation change may occur as a consequence of disease and not necessarily be the triggering step in the disease susceptibility: conclusions on causality cannot be drawn from retrospective data. Ideally, longitudinal cohorts or utilization of Guthrie cards (stored blood for metabolic screening in newborns) will allow the sequence of change to be determined and point to underlying causality. Although methylation may be the link between early life changes and adult disease, no methylation study for any disease has yet proven causality in humans.

**Confounders**

Any lifestyle effect such as smoking or alcohol or pathophysiology of the disease and/or any treatment prior to sample collection could itself generate epigenetic change, which may be of interest or a confounder. There are also recognized technical sources of variation that should be taken into account in both sample collection and assay randomization (especially batch effects) [33]. The use of different array platforms and/or at a different date may account for the observed variability in DNA methylation between cases and controls and give a spurious association, which is called a batch effect.

**Methylation assessment**

Methylation may be assessed at specific candidate gene loci or genome-wide. A wide range of methods have been developed to explore methylation changes across the whole genome in a hypothesis-free manner (reviewed in Refs. [41–44]). We will pinpoint some basic concepts to understand this rapidly evolving topic.

There are two types of platforms in genome-wide methylation analysis: (1) array-based (bisulfite microarrays) and (2) sequencing-based methods (bisulfite sequencing and enrichment-based methods) [42].

Bisulfite sequencing is considered the gold standard, it offers resolution to a single-base and is generally accurate and reproducible, but remains expensive. Enrichment-based methods include for example the use of DNA-methylation-specific antibodies as in methylated DNA immunoprecipitation coupled with high-throughput sequencing (MeDIP-sequencing) to enrich for a fraction of highly methylated (or unmethylated) DNA fragments. It is then followed by quantification of the enrichment of specific fragments by next-generation sequencing.

Bisulfite microarrays are 10-fold cheaper currently and measure DNA methylation at a preselected fraction of Cytosines throughout the genome, each of which is represented by dedicated probes on the microarray. Resolution is high (at a single CpG), but coverage is limited and may not reflect the most interesting areas because methylation is only assessed at the CpG sites on the probes. Illumina Infinium assays are the most widely used method for methylation. The latest version covers around 450,000 CpGs sites, which corresponds to only about 1.5% of CpGs in the human genome; it includes only few CpG-dense regions. More recently, it has been shown that the array only contains less than 9% of the important variable CpG regions [45]. Currently most of the studies published in the rheumatic disease literature have used bisulfite microarray methods.

**Statistical analyses**

Assessment of methylation sites is generally treated as a binary variable: (i) the detection of differently methylated regions (DMRs) after statistical comparisons between groups (e.g., cases versus controls). DMRs may be defined according to statistical significance after comparisons and relative and absolute differences or ratios in DNA methylation between groups and controls. After normalization, absolute DNA methylation levels calculated from bisulfite-sequencing data are generally expressed as beta values (β). β Values conceptually stem for the observation that the distribution of DNA methylation levels across the genome resembles a β distribution with a finite scale of 0–1 (where 0 = absence of methylation and 1 = complete methylation). As for GWAS, multiple testing must be taken into account in EWAS. The principle is adjusted statistically for the number of tests performed according to the coverage of the methylation assay to lower the number of false-positive signals due to multiple comparisons and obtain an adjusted p-value. There is however no clear consensus for epigenome-wide significance level as there is for GWAS yet [33], and approaches remain conservative.

A DMR may be situated in the promoter region or in a gene body, but also in inter-genic regions depending on the coverage and type of methylation assay. From the list of DMRs, mapping the corresponding gene and determining its level of methylation (one gene may have many DMRs, possibly in opposite directions) is a further challenge. Usually, bioinformatic tools such as gene pathway analyses help identify biologically meaningful trends in the lists of DMRs.

**Validation and replication**

Reliability of the data and their significance should be confirmed by validation and replication of the results in another independent sample.

**Integrative analysis with other omics and experimental data**

Comparison of DNA methylation results with gene expression analysis or other epigenetic marks (e.g., histone modifications, miRNAs, and open chromatin state) would provide a useful overall picture of the epigenetic regulatory landscape. Moreover, as methylation status has a genetic component, the addition of genotyping or DNA sequencing data to case–control studies allows the exploration of the relationship between sequence variation and methylation regulation—called Methylation QTLs (MeQTLs).

**Methylation changes in genes and across the genome in rheumatic diseases**

Methylation changes in specific genes have been identified in many rheumatic diseases through candidate gene methylation studies and, more recently, EWAS. In the present review, all published EWAS in rheumatic diseases until April 2013 have been included, along with the most significant candidate gene studies for key diseases. Recent literature on EWAS and rheumatic diseases has mainly focused on autoimmune diseases and includes studies on rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and dermatomyositis (DM). In addition, there are recent studies of osteoporosis (OP) and osteoarthritis (OA) (Table 2).

**General overview**

Major methodological issues render the results published so far preliminary, and interpretation, like in early candidate gene days, should be cautious. Many studies are underpowered, lack replication data, and have low methylation coverage. Indeed, all EWAS...
<table>
<thead>
<tr>
<th>Disease</th>
<th>Study design</th>
<th>Phenotype</th>
<th>Sample size</th>
<th>Replication</th>
<th>Cell/tissue</th>
<th>Methylation assay</th>
<th>Methylation profile</th>
<th>Statistical criteria for DMRs definition</th>
<th>Gene expression platform&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Main results</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>MZ pairs RA</td>
<td>RA</td>
<td>5</td>
<td>–</td>
<td>WBCs</td>
<td>Illumina GoldenGate</td>
<td>= 0</td>
<td>0 p &lt; 0.05 (adjusted for multiple testing) and min mean methylation change of 10%</td>
<td>–</td>
<td>No significant changes</td>
<td>[50]</td>
</tr>
<tr>
<td>CC</td>
<td>RA versus OA</td>
<td>OA</td>
<td>6–5</td>
<td>–</td>
<td>FLS</td>
<td>Illumina 450 K</td>
<td>= 1859</td>
<td>1206 p &lt; 0.05 (adjusted for multiple testing) and average difference in methylation level &gt; 0.1</td>
<td>qRT-PCR</td>
<td>Hypomethylation in genes relevant to RA, such as CHE1L1, CAPN1, STAT3, MAP3K5, MEFV, and WISP3 and hypermethylation in TGFBR2 and FOXO1.</td>
<td>[57]</td>
</tr>
<tr>
<td>CC</td>
<td>RA versus OA</td>
<td>OA</td>
<td>6–6</td>
<td>–</td>
<td>FLS</td>
<td>Illumina 450 K</td>
<td>= 2571</td>
<td>1240 p &lt; 0.05 and absolute value of the median difference between b-values &gt; 0.1</td>
<td>qRT-PCR</td>
<td>Altered DNAm in genes such as IL6R, CAPN1, DPP4, and HRO.</td>
<td>[58]</td>
</tr>
<tr>
<td>CC</td>
<td>ACPA + RA versus controls</td>
<td>354–337 12–12</td>
<td>WBCs</td>
<td>Illumina 450 K</td>
<td>NP 51,476</td>
<td>18,227</td>
<td>–</td>
<td>0.05 adjusted for multiple testing</td>
<td></td>
<td>Two clusters within the MHC region that may, through differential methylation, mediate the risk of RA after correction for cellular heterogeneity. Partial replication in monocyte cell fractions in an independent cohort of 12 cases and 12 controls.</td>
<td>[59]</td>
</tr>
<tr>
<td>SLE</td>
<td>MZ pairs SLE</td>
<td>SLE</td>
<td>5</td>
<td>–</td>
<td>WBCs</td>
<td>Illumina GoldenGate</td>
<td>= 49</td>
<td>49 p &lt; 0.05 (adjusted for multiple testing) and min mean methylation change of 10%</td>
<td>qRT-PCR</td>
<td>Hypomethylation of immune-related genes such as IFNGR2, MMP14, LCN2, CSF3R, PECAM1, CD9, and AIMP2.</td>
<td>[50]</td>
</tr>
<tr>
<td>CC</td>
<td>SLE versus controls</td>
<td>12–12</td>
<td>CD4&lt;sup&gt;+&lt;/sup&gt; T</td>
<td>Illumina 27 K</td>
<td>NP 341</td>
<td>336</td>
<td>p &lt; 0.05 and at least 1.2-fold difference in methylation level</td>
<td>–</td>
<td>Hypomethylation of immune-related genes such as CD9, MMP-9, PDCFR, and BST2. Hypermethylation in RUNX3 involved in T cell maturation and genes involved in folate biosynthesis.</td>
<td>[51]</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>SLE versus controls</td>
<td>12–12</td>
<td>WBCs</td>
<td>Illumina 27 K</td>
<td>NP 2165</td>
<td>2165</td>
<td>&lt; 0.05</td>
<td>–</td>
<td>Hypomethylation of the IL10 and IL12 genes in SLE patients.</td>
<td>[86]</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>SLE versus controls</td>
<td>18</td>
<td>Naive CD4&lt;sup&gt;+&lt;/sup&gt; T</td>
<td>Illumina 450 K</td>
<td>NP 86</td>
<td>47</td>
<td>&lt; 0.001 (adjusted for multiple testing) and at least 1.2-fold difference in methylation level</td>
<td>HumanHT-12 v4 Expression BeadChip array (Illumina)</td>
<td>Hypomethylation in interferon-regulated genes in naive CD4&lt;sup&gt;+&lt;/sup&gt; T cells in SLE, including IFIT1, IFIT3, MX1, STAT1, IFI44L, USP18, TRIM22 and BST2</td>
<td>[52]</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>MZ pairs DM</td>
<td>DM</td>
<td>5</td>
<td>–</td>
<td>WBCs</td>
<td>Illumina GoldenGate</td>
<td>= 0</td>
<td>0 p &lt; 0.05 (adjusted for multiple testing) and min mean methylation change of 10%</td>
<td>–</td>
<td>No significant changes</td>
<td>[50]</td>
</tr>
<tr>
<td>OA</td>
<td>CC</td>
<td>OA versus controls</td>
<td>25–20</td>
<td>Articular cartilage</td>
<td>Illumina 27 K</td>
<td>NP 91</td>
<td>91</td>
<td>&lt; 0.05 (adjusted for multiple testing)</td>
<td>–</td>
<td>DMRs mainly related to the inflammatory/defense response (demethylated in OA). RUNX1 and MSX1 were respectively the 2 most hypo and hypermethylated genes in OA. Majority of hypomethylated genes in OP compared to OA. No classic gene bone candidate but enrichment for GWAS bone trait-associated genes and the homeobox group of genes.</td>
<td>[66]</td>
</tr>
<tr>
<td>OP</td>
<td>CC</td>
<td>Hip fracture versus hip OA</td>
<td>27–26</td>
<td>Trabecular bone</td>
<td>Illumina 27 K</td>
<td>NP 241</td>
<td>228</td>
<td>&lt; 0.05 (adjusted for multiple testing)</td>
<td>–</td>
<td></td>
<td>[68]</td>
</tr>
</tbody>
</table>

MZ pairs stands for MZ discordant twin pairs. Abbreviations: CC = case-control; DMRs = differently methylated regions; DMGs = differently methylated genes; SLE = systemic lupus erythematosus; RA = rheumatoid arthritis; DM = dermatomyositis; OA = osteoarthritis; OP = osteoporosis; WBCs = white blood cells; FLS = fibroblast-like cells; NP = not provided; qRT-PCR = quantitative real-time PCR. Symbols: Similar global methylation level (\(\approx\)), global hypomethylation (\(\sim\)).

<sup>a</sup> In an independent population.

<sup>b</sup> Only gene expression analyses performed on the same population as for DMRs discovery were considered.
in rheumatic diseases published so far, have been generated using methylation arrays of variable coverage. In addition, methods and statistical analyses—notably criteria to define a probe as a DMR—differ greatly between studies rendering meaningful cross-comparison a challenge. Correction for multiple testing or potential confounders has not been taken into account in several studies. In addition, in some instances, the control group was not always composed of healthy individuals, but those with different diseases (e.g., RA versus OA or OA versus OP), making generalizability difficult. Only a few studies have integrated other experiments like gene expression or other omics experiments, which will be necessary to interpret the results in a biologically meaningful manner.

Overall, these studies provide preliminary evidence for specific methylation changes in rheumatic disease and suggest methylation change in novel candidate genes and clustering of disease subgroups according to methylation profile. Average differences in methylation levels (β values) between the groups compared were generally modest (<15%) across all studies, suggesting relatively small effect sizes. When examined, expression data showed an overall inverse—but variable and inconsistent—relationship between DNAm and particular gene expression levels, highlighting again the complexity of regulation of gene expression.

**Systemic lupus erythematosus**

First evidence of altered DNAm status in SLE came from a study that found that T cells from patients with active lupus have a decreased global DNAm level and that 5-azacytidine (5-azaC), an inhibitor of DNA methylation, could induce auto-reactivity in cloned CD4 + T cells and autoimmune syndrome [46,47]. Moreover, classic methylation-sensitive autoimmunity-related genes have been identified in CD4 + T cells in SLE, such as **ITGAL** (CD11A), **PRF1** (perforin), **TNFSF7** (CD70), and **CD40LG** (reviewed in Ref. [48]).

EWASs published so far suggest a tendency towards hypomethylation in SLE compared to healthy controls and support the central role of type I interferon pathway in SLE pathogenesis [49] as interferon-related genes have been identified as differentially methylated across studies. However, major limitations include small sample size, low coverage of the epigenome, and scarce clinical data with heterogeneous case population.

The first EWAS study using the MZ discordant twins design was published in 2010 and included 15 pairs of MZ autoimmune discordant twins (five with RA, five with SLE, and five with DM) [50]. Using a low-coverage methylation array representing 807 gene promoters, the authors identified consistent methylation changes in SLE twins relative to unaffected co-twins in certain genes influencing immune function and a global decrease in DNAm levels. One year later, the methylation profile of CD4 + T cells of 12 unrelated lupus patients and 12 normal controls partially matched for ethnicity was compared [51]. The authors identified 341 differently methylated (DM) CG between cases and controls—the majority were hypomethylated in SLE, notably in known connective tissue-related genes (e.g., **MMP-9** and **PDGFRa**). Of note, PDGFRa autoantibodies are present in less than half of lupus patients and contribute to the development of autoimmune hemolytic anemia.

Using a more powerful assay and two independent cohorts (one for discovery (n = 18) from the same institution as [51] and one for replication (n = 18)), Coit et al. [52] identified and replicated 86 DMRs between naïve T cells from patients and controls (36 pairs) in 47 genes (mainly hypomethylated). A majority of them were regulated by type I interferon. However, gene expression analysis assessed from microarray in a subset of study participants did not reveal different expression levels of these DMGs in naïve CD4, raising the question of their biological significance.

Lin et al. identified global hypomethylation in SLE as compared to controls in a small population of 12 active SLE cases and 12 controls. No correction for multiple testing was apparently applied, and limited clinical and demographic data were provided. Among the 2165 genes with significant methylation changes, the authors focused on **IL1R2** and **IL10** (the latter was also hypomethylated in Ref. [50]) as biologically relevant candidates for further confirmation in 66 SLE and 102 healthy controls. Both genes were hypomethylated in SLE relative to healthy controls but as well in RA patients (n = 12). This finding corroborates recent evidence for overlapping methylome signatures between different autoimmune diseases similar to that found for single nucleotide polymorphisms in GWAS [53].

**Rheumatoid arthritis**

The investigation of DNAm in RA began with relatively small candidate gene studies showing altered methylation status of single gene promoters (generally at a single CpG site), including **DR3**, **IL6**, **IL10**, **IL1R2**, and **CCL12** identified in peripheral blood mononuclear cells and FLSs (reviewed in Ref. [54]). In addition, evidence for epigenetic involvement in RA with respect to gender bias is supported by one study, which shows significantly increased skewed X-chromosome inactivation pattern in female RA patients (n = 106) compared to controls (n = 257) [55]. Also, demethylation of the promoter of the CD40L gene on X chromosome was observed in CD4 + T cells from female (n = 11) but not male (n = 9) RA patients [56].

Several EWAS in RA have been published recently [50,57–59]. They suggest the presence of specific methylation changes in RA, and interesting regions such as the **IL6R** gene and the MCH region which have been identified as differently methylated. However, cross-comparisons are difficult because of the use of different tissue (whole blood versus FLSs) and different “control” groups. Moreover, due to the known clinical heterogeneity of RA, only one study included only ACPA-positive patients, the others included heterogeneous and treated individuals, which may have biased the results.

In the twin discordant study from Javierre et al. [50], there were no significant changes in methylation profile within the five discordant RA twin pairs. More recently, two case–control studies compared methylation profiles derived from a small number of cultured FLSs in patients with RA or OA (n = 6 RA versus n = 5 OA in the first study and n = 5 RA versus n = 6 OA in the second one) [57,58]. The authors provided limited clinical and demographic data, and statistical analyses were not adjusted for age, medication, or batch effects. Having said that, both research groups identified an RA “methylome signature” with partially overlapping genes between both sets, including genes like **MMP20**, **RASSGR2**, **ADAMTS2**, **EFG**, and **TIMP2**. Hypomethylated loci were identified by the first group in key genes relevant to RA. In a study by Nakano et al. [57], gene expression determined by PCR for 7 genes picked up from the DMRs list on the same samples of patients (6 OA and 13 RA FLS lines) showed overall greater expression in RA in comparison with OA, but the relationship for individual genes was variable. de la Rica and co-authors reanalyzed available RA FLSs and OA FLSs array expression data from the Gene Expression Omnibus (GEO) database and compared the results with their DNAm data. Two hundred eight annotated CpGs displayed an inverse correlation between expression and methylation levels. Then, they investigated the expression status of 10 genes displaying a change in DNAm in the set of samples used in this study.
Results showed inconsistent relationship between methylation level and gene expression.

Finally, Liu et al. [59] analyzed a much larger cohort of RA patients (n = 354 and 337 controls) from a population-based cohort. As a major improvement on other studies, cases comprised only ACPA-positive RA patients, free of treatment and the results were replicated in a second independent sample. Of the 51,476 identified DMRs, the authors identified genotype (single nucleotide polymorphisms)-dependant DMRs to help identify methylation that was inferred as a likely result of disease. Of note, the potential causal relationship was not based on observational data (or Mendelian Randomization methods) but was determined statistically using a series of cross-sectional conditional correlation analyses based on the causal inference literature. The authors identified two clusters within the major histocompatibility complex region, whose differential methylation potentially mediates genetic risk for rheumatoid arthritis. As the analyses were performed on whole blood-derived DNA samples cells, a correction was applied for cellular heterogeneity by estimating and adjusting for cell-type proportions, as well as for age, gender, and smoking status. The fact that DMRs were found in the HLA region suggests that the results are likely to be real but the study was powered to find major effects only. The authors partially replicated their findings on additional 12 RA patients and 12 controls on the monocye fraction of peripheral blood cells.

Osteoarthritis (OA)

The role of DNAm in OA has in general focused on methylation status and gene expression at the promoters of candidate genes known to be important for cartilage maintenance (reviewed in Ref. [60]). These in vitro studies have included a number of metalloproteinases at a single CpG (e.g., MMP13 and ADAMTS4 in end-stage OA chondrocytes in one study on 16 patients with OA and 10 controls) [61]. The expression of MMP-13 was further shown to be modulated by epigenetic change in the promoter regions of the leptin gene in a candidate gene study on osteoarthritic articular cartilage from 15 participants [62]. It was also postulated that long-term inflammatory cytokines might induce overexpression of IL1B through the methylation of its promoter at a specific CpG based on an experiment on human cultured chondrocytes. A single nucleotide polymorphism in GDF5 was postulated to mediate GDF5 differential expression in OA through DNA methylation [63] and SOD2 gene—related to the oxidative stress process—may also be regulated by DNA methylation of several CpG sites within its promoter in OA chondrocytes [64]. No changes were seen in total DNA methylation content in OA chondrocytes relative to control by chromatography [65].

Only one unreplicated small EWAS on knee OA (25 cases versus 20 controls, articular cartilage samples) has been published. The authors found global hypomethylation in OA and identified 91 differentially methylated (DM) probes between cases and controls [66]. Correction for age but not for cellular heterogeneity was applied. The authors identified a small sub-group of cases (7 out of 23) with a particular methylation pattern (1357/23,837 probes differentiated this group from the others using a statistics tool for unsupervised class discovery). This sub-group was enriched for genes associated with inflammatory response (hypomethylated) and extracellular matrix constituents (hypermethylated). Although potentially interesting, this finding is based on a small sample and clinical details were not discussed, making these results still preliminary.

Osteoporosis

The contribution of epigenetics and methylation to bone mass and homeostasis is poorly understood, limited mainly to genes of importance in osteoblast and osteoclast differentiation from small candidate gene studies (reviewed in Ref. [67]). Genome-wide methylation profiles from pieces of trabecular bone of the femoral head from women (27 cases, mean age 80 ± 3 years) undergoing hip surgery for fracture (arguably considered as a proxy for OP) were recently compared to those of patients with hip OA (26 patients, mean age 75 ± 6 years) [68]. Of the 225 DMGs (the majority was less methylated in OP than in OA), there was enrichment for GWAS bone trait-associated genes and in genes participating in skeletal development, particularly in the homeobox superfamily of transcription factors. Of note, the leptin gene showed significant methylation difference between OP and OA in this study, although there were no other DMGs in classic bone gene candidates. This unique study without replication and integrative experiments is too preliminary to draw any firm conclusions.

Pain sensitivity

Chronic pain, even if not a disease per se, is a major symptom of most rheumatic diseases and a key therapeutic aim. The concept of central sensitivity syndrome may underlie conditions like fibromyalgia and overlapping conditions such as chronic fatigue syndrome [69]. There is increasing evidence for epigenetic mechanisms in both the development and maintenance of pain states [70]. Heat pain suprathreshold response was determined among 2500 twins from the TwinsUK cohort, as these measures have been shown to reflect propensity to develop chronic pain. Using genome-wide MeDIP-sequencing on whole blood from 25 discordant MZ twin pairs (discovery cohort) for objective measures of heat pain sensitivity and 50 unrelated samples, Bell et al. [71] identified nine DMRs in eight unique regions associated with high or low pain sensitivity. MZ twin and unrelated sample EWAS were performed separately followed by a meta-analysis of combined results. The strongest association signal was in the promoter of the ion channel gene TRPA1, a known pain gene whose product is expressed in peripheral nociceptors [72]. Of interest, most of the top DMRs were not covered on the Illumina HumanMethylation450 BeadChip (used for validation), stressing the importance of increased methylation coverage in EWAS. Variability in WBC sub-type (lymphocytes, neutrophils, basophils, and eosinophils) did not show a major effect on the most significant DMRs in this study. Longitudinal data were obtained for 33 individuals from the MZ twin sample (2–3 years apart) and patterns at nine DMRs showed relatively stable methylation over time.

This study illustrates the efficiency of adequate sample size and in-depth methylation coverage for gene discovery in EWAS. It suggests epigenetic influence on sensitivity to pain in normal individuals, leading to potential future interesting development in pathological conditions of chronic pain states such as chronic back pain or fibromyalgia.

Future perspectives

Other unexplored rheumatic diseases may represent good candidates for epigenetic investigation. For instance, gout has a high and increasing prevalence (≥ 3.9%) [73] and cohorts with extended genomic data already exist. Interestingly, in contrast to high genetic contribution to plasma urate level, gout has a relatively low estimated heritability with an important component of non-shared environment in phenotypic variance, suggesting the environmental influence is strong and possibly epigenetically mediated [74].

Further, the application of alternative study designs might be suitable for some diseases. For instance, comparison of affected
versus unaffected cells within a tissue from the same individual may provide insight into acquired epigenetic modification in a similar approach to cancer. This may be applicable, for example, to Paget’s disease or from different joints in localized forms of inflammatory arthritis. In addition, EWAS, if sufficiently powered with large sample sizes or twin designs with epigenome-wide coverage, have great potential for the identification of markers for disease onset, progression, or response to treatment for diseases such as RA or SLE. Collecting samples at diagnosis before medication starts and then longitudinally comparing the methylation profile between groups of responders versus non-responders may show methylation differences induced by treatment response or previous differences precluding drug resistance (e.g., in RA patients treated with specific DMARDs or biological drugs) [75].

**Conclusion**

Epigenetics is a major part of the future of rheumatology. The increasing number of published epigenome-wide methylation studies brings new challenges of interpretation to the medical and scientific community. Current data from genome-wide methylation analyses suggest specific methylation changes in several rheumatic diseases compared to healthy controls or control groups. These studies have allowed the identification of differentially methylated genes previously implicated in the disease pathology but have also highlighted new gene associations with individual diseases. Improved methodology and generation of methylation data using more powerful techniques with coverage of the methylome are likely to follow and will shed new light on disease pathogenesis. Larger sample sizes, replication, and biologic validation is going to be necessary to integrate these data and allow their potential translation into the clinical arena. Longitudinal data will be advantageous in teasing apart cause and effect. We can however expect advances in epigenetics to aid disease diagnostics, classification, and therapy in the future.

**References**


[66] Correa-Ferreira et al. / Seminars in Arthritis and Rheumatism 2013) 43 (2013) 9 – 16