Biochimica et Biophysica Acta 1829 (2013) 1161–1174

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer

Marta Kulis, Ana C. Queirós, Renée Beekman, José I. Martín-Subero *

Departamento de Anatomía Patológica, Farmacología y Microbiología, Universidad de Barcelona, Unidad de Hematopatología, Hospital Clinic, Institut d'Investigacions Biomédiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

ARTICLE INFO

Article history:
Received 23 May 2013
Received in revised form 2 August 2013
Accepted 5 August 2013
Available online 9 August 2013

Keywords:
DNA methylation
Gene body
Enhancer
Cell differentiation
Cancer
Transcription regulation

ABSTRACT

Ever since the discovery of DNA methylation at cytosine residues, the role of this so called fifth base has been extensively studied and debated. Until recently, the majority of DNA methylation studies focused on the analysis of CpG islands associated to promoter regions. However, with the upcoming possibilities to study DNA methylation in a genome-wide context, this epigenetic mark can now be studied in an unbiased manner. As a result, recent studies have shown that not only promoters but also intragenic and intergenic regions are widely modulated during physiological processes and disease. In particular, it is becoming increasingly clear that DNA methylation in the gene body is not just a passive witness of gene transcription but it seems to be actively involved in multiple gene regulation processes. In this review we discuss the potential role of intragenic DNA methylation in alternative promoter usage, regulation of short and long non-coding RNAs, alternative RNA processing, as well as enhancer activity. Furthermore, we summarize how the intragenic DNA methylome is modified both during normal cell differentiation and neoplastic transformation.

1. Introduction

Back in the 1970s, Riggs and Holliday independently suggested that methylation of cytosines in the context of CpG dinucleotides may represent an epigenetic mark associated with gene silencing [1,2]. Years later, this hypothesis was experimentally demonstrated by several groups [3–7]. Interestingly, already in the 1980s it was recognized that CpG density is not evenly distributed within the genome, but rather shows a bimodal distribution. Regions with an elevated CpG content are called CpG islands (CGIs), and overlap with transcriptional start sites (TSSs) of approximately 60–70% of all human genes [8,9]. In contrast, regions with low CpG density are frequently located outside TSS [10,11]. Furthermore, it was noted in these studies that both DNA methylation and CpG density show a bimodal distribution. In general, CGIs are unmethylated whereas most regions with low CpG density are heavily methylated in normal tissues, something which is now considered general knowledge. Experimental evidence demonstrated that gene expression can be regulated by DNA methylation levels of CGIs in the proximity of TSSs [12,13]. Based on these data, the majority of cancer-related DNA methylation studies focused on the role of CGI hypermethylation as a mechanism of tumor suppressor gene silencing [14–18]. Additionally, already in the 1980s, pioneer studies in cancer cells reported that neoplastic transformation was also associated with global and gene specific loss of DNA methylation [19–22]. Collectively, the studies published so far in cancer epigenomics point to a massive disruption of the DNA methylome in tumor cells as compared to their normal cell counterparts.

DNA methylation can be analyzed by methylation sensitive restriction enzyme digestion, bisulfite conversion based techniques or affinity enrichment of methylated DNA [23–25]. A comprehensive overview of the advantages and disadvantages of these approaches and their applicability is given by Laird [26]. Until recently, these techniques were PCR or microarray-based to analyze single or multiple regions of interest, respectively [27–30]. As a consequence, the great majority of studies designed to investigate DNA methylation dynamics were covering only a small fraction of the DNA methylome, namely CGIs and promoter regions. Nowadays, the use of high-density microarray analyses, e.g., comprehensive high-throughput arrays for relative methylation (CHARM) and Infinium [31,32], and next-generation sequencing based DNA methylation analysis, e.g., methylated DNA immunoprecipitation followed by deep-sequencing (MeDIP-seq), CXXC affinity purification followed by deep-sequencing (CAP-seq), reduced representation bisulfite sequencing (RRBS) and whole genome bisulfite sequencing (WGBS) [33–36], allows us to obtain an unbiased representation of DNA methylation maps throughout the genome. The use of these techniques has started to reveal that DNA methylation has an even broader function than previously anticipated and that its functions may vary in a context-dependent manner [37].

One of the major findings derived from whole-genome studies is that DNA methylation levels in the gene body, i.e. intragenic DNA methylation, widely change during cell differentiation and carcinogenesis. Although the precise role of DNA methylation within the gene body is still

* Corresponding author at: Centre Esther Koplowitz, C/Rosselló 153, 2nd floor, 08036 Barcelona, Spain. Tel.: +34 2275400.
E-mail address: imartins@clinic.ub.es (J.I. Martín-Subero).

1874-9399/$ – see front matter
http://dx.doi.org/10.1016/j.bbagrm.2013.08.001

far from being well understood, recent publications indicate that it may be involved in the regulation of multiple processes, such as transcript elongation, expression of intragenic coding and non-coding transcripts, alternative splicing and enhancer activation. The goal of this review is to provide a comprehensive overview of the current knowledge on intragenic DNA methylation and its association with gene regulation in the context of cell differentiation and neoplastic transformation.

2. DNA methylation and gene expression regulation: General aspects

The molecular mechanisms by which DNA methylation mediates gene silencing have been mainly studied in the context of genes containing CGIs in their promoter region. Many reports have shown that inducing hypomethylation of CGIs, either by genetic deletions of DNA methyltransferases DNMT1, DNMT3A and DNMT3B or by pharmacologic interventions through demethylating agents, results in gene reactivation [7,13,38–41]. How does CGI methylation induce gene silencing? One of the most accepted models is that gene silencing is mediated by proteins containing a methyl-CpG binding domain (MBD). MBD binding to methylated CGIs is followed by recruitment of histone deacetylases, chromatin compaction and gene silencing [42,43]. Another possible model involves DNA methylation dependent transcription factor (TF) binding to their recognition sites; if the DNA is methylated, the TFs cannot bind to the promoter region and therefore, the gene is inactive [44–46].

Of important note is that, although the presence of DNA methylation at CGIs seems to ensure a repressive chromatin environment, its absence does not necessarily associate with gene activation. In fact, although the great majority of CGIs remain unmethylated during development and cell differentiation, only approximately 50% of the genes are expressed in one or more particular cell types [47]. The remaining genes with unmethylated CGIs are either silenced by other epigenetic mechanisms, such as the polycomb repressive complex, which mediates methylation of H3K27 and leads to a closed chromatin state [48], or are not activated due to the lack of TFs. Interestingly, although DNA methylation has frequently been considered a mechanism inducing de novo gene silencing, multiple studies indicate that DNA methylation in the context of CGIs seems to play a role in achieving stable gene inactivation [49–52], whereas gene regulation through histone modifications is more dynamic. Hence, in many cases CGI methylation may be more a consequence rather than a cause of gene repression.

3. Association between intragenic DNA methylation and gene expression

The negative correlation between gene expression and CGI methylation at TSSs is well established. However, this association cannot be extrapolated to other genomic contexts such as CpGs located in the gene body. In general, DNA methylation is thought to block transcription initiation but not elongation. In fact, intragenic nucleosomes with trimethylation of H3K36 (H3K36me3), which is associated with transcript elongation, seem to recruit DNMTs, thus facilitating the methylation of intragenic DNA [53]. Therefore, even if the gene body is highly methylated, which is a frequent finding in normal undifferentiated cells, the gene may be transcribed. This can lead to an apparent contradiction, as DNA methylation in the promoter can be negatively associated with gene expression whereas the methylation status in the gene body of the same gene can show a positive correlation (Fig. 1).

A positive correlation between intragenic DNA methylation and gene expression has been recently observed in multiple genome-wide epigenomic studies, both in the context of cell development and differentiation as well as in cancer cells [34,54–58]. Our whole-genome DNA methylation and gene expression study in chronic lymphocytic leukemia (CLL) has revealed that the methylation status of CpGs in the gene body of around 900 genes shows a significant correlation (positive or negative) with gene expression in the absence of DNA methylation changes in promoter regions [56]. Moreover, a positive association is clear from a study on the X chromosome. Using an elegant allele-specific DNA methylation and expression analysis, it was observed that gene bodies of transcribed genes from the active X chromosome are extensively methylated whereas those at promoter regions are unmethylated [59]. This finding sheds a new light on the common model that allele-specific methylation is restricted to CGIs on the inactive X chromosome.

Recent studies have revealed a mechanistic link between replication timing, gene expression and DNA methylation patterns [60,61], further underlining the positive correlation between intragenic DNA methylation and transcription. It is known that regions replicating during the first half of S-phase have a more open, transcription facilitating chromatin conformation than late replicating regions. Furthermore, early-replicating regions have a higher DNA methylation level and more efficiently maintain their methylation levels during cell division, when compared to late replicating regions [60,62]. Hence, the accumulation of cell divisions results in a reduction of DNA methylation levels in late replicating regions. This phenomenon is enhanced in cancer cells, whose high proliferation leads to hypomethylation of late replicating regions such as those bound to the nuclear lamina [63]. These findings may be explained by the differential expression of methyltransferases [64] and/or a changing availability of methyl group donors during the cell cycle, which may affect the efficient DNA methylation maintenance in late S-phase. This, however, is not the only mechanism to efficiently maintain gene-specific methylation of active genes because certain active genes that reside in late S-phase regions efficiently maintain their methylation status as well [60]. A recent meta-analysis of genome-wide epigenomic and gene expression data from cell lines intends to illustrate the relationship between intragenic DNA methylation and transcription levels [65]. The authors describe that such a relationship is non-linear but has a bell-shape distribution, i.e. the lowest levels of intragenic DNA methylation correspond to both the lowest and highest expressed genes, while the highest methylation levels are associated with genes expressed at intermediate levels. This finding contradicts the data derived from the replication timing studies, hence, additional investigations are required to study the above described phenomena in further detail.

An additional study that combined high-throughput RNA sequencing (RNA-seq) data with DNA methylation profiles obtained by RRBS observed a context-dependent correlation of CpG gene-body methylation, related to whether the CpGs were located in or outside intragenic CGIs. For CpGs outside intragenic CGIs, the methylation status correlated positively with gene expression, while for CpGs located within CGIs, the methylation status can either be positively or negatively correlated with gene expression levels. Only for approximately 15% of the intragenic CGIs presenting a negative correlation, an association with specific gene regulatory processes, e.g., alternative promoter usage or intragenic enhancer activity, was found. No significant enrichment for transcription factor binding sites was determined in the remaining CGIs. Thus, the explanation of the negative correlation of DNA methylation and gene expression in the majority of intragenic CGIs remains unknown and needs further investigation [66].

In summary, both positive and negative correlations between transcription and intragenic DNA methylation have been described. However, of important note is that an association between gene expression and DNA methylation does not necessarily indicate that alterations in DNA methylation cause gene expression level changes. In the following sections of this review we will summarize the different scenarios in which intragenic DNA methylation may play a causative role in transcriptional regulation.

4. Regulation of alternative intragenic promoters

An initial study from 1999 linked the DNA methylation status of an alternative promoter in the *TGF-β3* gene to expression of an alternative