

DNA Methylation: Genomewide Distribution, Regulatory Mechanism and Therapy Target

D. S. Kaplun^{1,2}, D. N. Kaluzhny³, E. B. Prokhortchouk^{1,2}, S. V. Zhenilo^{1,2}

¹Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, 119071 Russia

²Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119071 Russia

³Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

*E-mail: zhenilo@biengi.ac.ru

Received October 07, 2022; in final form, November 29, 2022

DOI: 10.32607/actanaturae.11822

Copyright © 2022 National Research University Higher School of Economics. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT DNA methylation is the most important epigenetic modification involved in the regulation of transcription, imprinting, establishment of X-inactivation, and the formation of a chromatin structure. DNA methylation in the genome is often associated with transcriptional repression and the formation of closed heterochromatin. However, the results of genome-wide studies of the DNA methylation pattern and transcriptional activity of genes have nudged us toward reconsidering this paradigm, since the promoters of many genes remain active despite their methylation. The differences in the DNA methylation distribution in normal and pathological conditions allow us to consider methylation as a diagnostic marker or a therapy target. In this regard, the need to investigate the factors affecting DNA methylation and those involved in its interpretation becomes pressing. Recently, a large number of protein factors have been uncovered, whose ability to bind to DNA depends on their methylation. Many of these proteins act not only as transcriptional activators or repressors, but also affect the level of DNA methylation. These factors are considered potential therapeutic targets for the treatment of diseases resulting from either a change in DNA methylation or a change in the interpretation of its methylation level. In addition to protein factors, a secondary DNA structure can also affect its methylation and can be considered as a therapy target. In this review, the latest research into the DNA methylation landscape in the genome has been summarized to discuss why some DNA regions avoid methylation and what factors can affect its level or interpretation and, therefore, can be considered a therapy target.

KEYWORDS DNA methylation, DNA methyltransferases, G-quadruplexes, TET dioxygenases, methyl-DNA binding proteins.

ABBREVIATIONS TF – transcription factors.

INTRODUCTION

Cytosine is referred to as the fifth DNA base, and cytosine residue methylation is the most common DNA modification in mammalian cells. Cytosine residues in CpG dinucleotides are most often subject to methylation. However, the methylated cytosines outside CpG dinucleotides may account for 25–50% of all mC in stem cells and neurons [1]. In mammals, about 70–80% of cytosines in CpG dinucleotides are methylated [2]. *De novo* DNA methylation is catalyzed by the DNMT3a/3b DNA methyltransferases responsible for methylation in different ge-

nome regions and that are not interchangeable [3, 4]. DNA methylation during replication is maintained by DNMT1 DNA methyltransferase. DNA demethylation occurs both passively, during cell division, and actively, due to enzyme activity. The key factors involved in active demethylation are TET1,2,3 dioxygenases. TET proteins oxidize methylcytosine to hydroxymethylcytosine and, then, formylcytosine and carboxycytosine, which then produce cytosine as a result of excision repair by thymine-DNA glycosylase (TDG/NEIL) (*Fig. 1*) [5]. Methylcytosine derivatives are not only considered as intermediate states

between methylated and non-methylated bases, but also as DNA modifications affecting the binding of transcription factors, as they are involved in gene expression regulation (Methylcytosine derivatives are discussed in survey [6]).

The key changes in DNA methylation during the organism's development are associated with cell differentiation. Differentiated cells are typically characterized by stable DNA methylation patterns, which can still vary due to external stimuli, various pathological processes, and ageing [7–11]. Dynamic DNA methylation changes in differentiated cells are also observed during memory formation and training in neural cells [12, 13]. DNA methylation in differentiated cells turns out to be stable in the remaining cases. Thus, DNA methylation can be considered as a target for therapy and the diagnostics of various pathogenetic conditions based on DNA methylation abnormalities affecting gene transcription.

The key features of DNA methylation distribution in the genome are presented in this survey. The factors affecting DNA methylation onset, maintenance, and demethylation are analyzed based on recently published data. The possibility for therapeutic use of these factors is discussed.

1. THE DNA METHYLATION DISTRIBUTION PATTERN IN MAMMALIAN CELLS

About 90% of all methylated CpG sites in mammalian genomes are located in various repeating sequences, such as satellite repeats and mobile elements [14]. The largest number of CpG-rich repeating elements

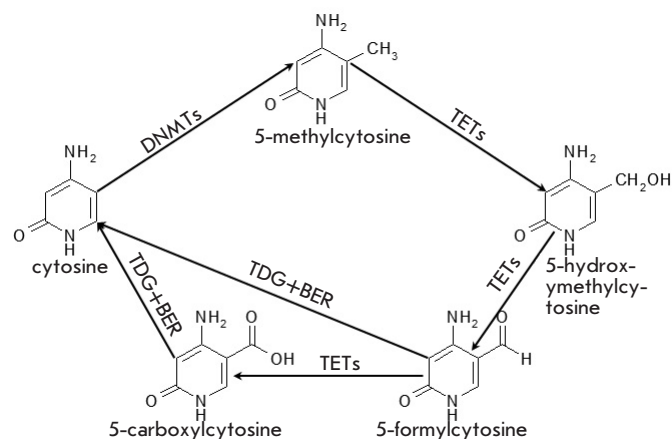


Fig. 1. Cytosine methylation and demethylation scheme

are found in structural chromosomal regions: centromere, pericentromere, and subtelomere (Fig. 2A). Genome-wide nonopore sequencing in humans has made it possible to not only read the sequences of repeating elements, but also to analyze their methylation in the genome: so, a significant degree of methylation has been observed under normal conditions [2, 15]. It is of note that methylation of the duplicated/repeating sequences located in various chromosomal regions may differ significantly [2]; i.e., a specific methylation pattern of repeating sequences is not only determined by the sequence repeating it-

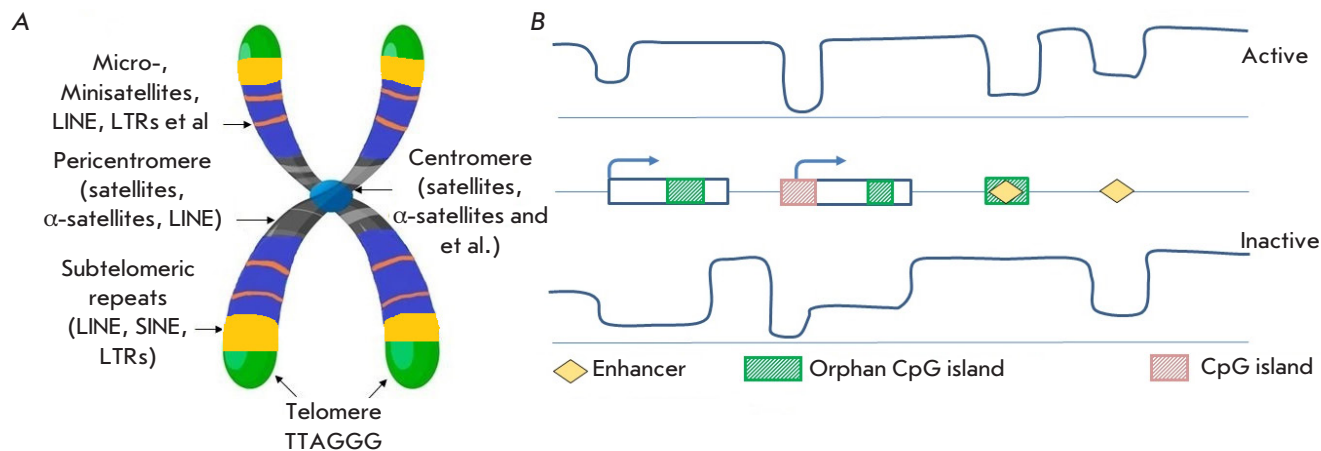


Fig. 2. Landscape of the DNA repetitive sequences. (A) Location of various repeating sequences on a chromosome. (B) DNA methylation profile in the genome, depending on the activity of promoters and enhancers and the presence of CpG islands

self, but by its chromosomal surroundings as well. Hypomethylation of various repeating elements is characteristic of various pathological conditions, including oncogenesis, immunodeficiency, as well as autoimmune, neurological, and mental disorders [7–9, 16, 17]. The necessity of satellite repeat methylation in centromere and pericentromere regions is associated with the correctness of chromosome disjunction in replication [18]. In contrast, methylation of mobile elements, transposons, and retrotransposons is aimed at suppressing their transcription. Demethylation of these repeats results in their active transcription and transposition, which fosters genome instability. It is possible that this is a redundant mechanism, because early transposons and retrotransposons are typically characterized by mutations and deletions in the sequences coding for transposase, which leads to inactive protein formation.

The mammalian genome includes CpG dinucleotides that avoid methylation. These CpG sites are usually included in the so-called CpG islands that are DNA regions where the GC pair content exceeds 50%, while the expected-to-observed CpG content is above 0.6. About 60% of the promoters include CpG islands. Lysine 4 residue trimethylation in the histone H3 molecule (H3K4me3) is an active chromatin modification typical for these regions, regardless of promoter activity [19]. Active chromatin is a DNA region where histone modifications, such as acetylation and H3K4me3, lead to DNA accessibility for transcription activators. The presence of H3K4me3 in the promoter regions of inactive genes facilitates transcription initiation but not mRNA synthesis. Meanwhile, there is a series of inactive gene promoters, including non-methylated CpG islands, where H3K4me3 is not detected. The genes located in clusters with three or more homologous genes coding for olfactory receptors, keratins, apolipoproteins, interleukins, and leukocyte antigens are most commonly attributed to this class [19]. Methylation of CpG islands in the promoter regions correlates with transcription suppression and may occur both under normal (e.g., during organism development) and pathological conditions [20]. For instance, malignant cell transformation and metastasis are typically characterized by hypermethylation of CpG islands in the promoters of oncosuppressor genes; i.e., the proteins involved in cellular adhesion and DNA repair. In most cases, such hypermethylation results in transcription suppression. However, it should be noted that promoter hypermethylation in tumors may occur in the genes considered transcriptionally inactive in the same tissue under normal conditions. In other words, their hypermethylation has no effect on ex-

pression suppression but, rather, reinforces their inactive status [21].

Promoters that include a small quantity of CpG dinucleotides are typical of tissue-specific genes and the genes involved in organism development. Methylation in these promoters does not always correlate with transcription suppression [22]. In a comparative analysis of brain and retinal cells, methylation of 66% of differentially methylated promoters correlated negatively with transcription. Thus, methylation in these promoters corresponds to transcription suppression. At the same time, promoter methylation was observed in 34% of transcriptionally active genes in [22].

The CpG islands that do not overlap with promoter regions are called orphan islands. The number of such islands is about half that of the promoter islands. Orphan islands often include H3K4me3 active chromatin modification and can initiate new transcripts [23]. Many orphan islands are subject to methylation during organism development, which makes them lose active chromatin modifications. Methylation in an orphan island inside a gene prevents the occurrence of transcription initiation sites inside the gene and correlates with active transcription [24]. Methylation inside genes may prevent Polycomb protein binding in the PRC2 repressor complex, which facilitates active transcription as well [25]. About 90% of orphan islands may act as tissue-specific enhancers [26]. The presence of a CpG island amplifies the enhancer's regulatory activity [27]. Active enhancers that include orphan islands are hypomethylated, while the classical enhancers operating in all tissue types show variable methylation [27] (*Fig. 2B*).

Methylation maps created for the whole genomic DNA in various cell types and the information regulatory activity of the elements make it possible to consider DNA methylation as a tool for transcription activity regulation with regard to correction or identification of the various pathogenetic states associated with changes in DNA methylation.

2. DNA METHYLATION HOMEOSTASIS

DNA methylation homeostasis is based on a complex regulatory network that balances methylation and demethylation. The key mechanisms maintaining homeostasis in cellular proliferation and differentiation are as follows: 1) passive genome-wide demethylation and maintenance of the methylation pattern by DNMT1 during replication and 2) targeted *de novo* methylation and active demethylation in specific regions. The factors involved in homeostasis are discussed in the present Chapter.

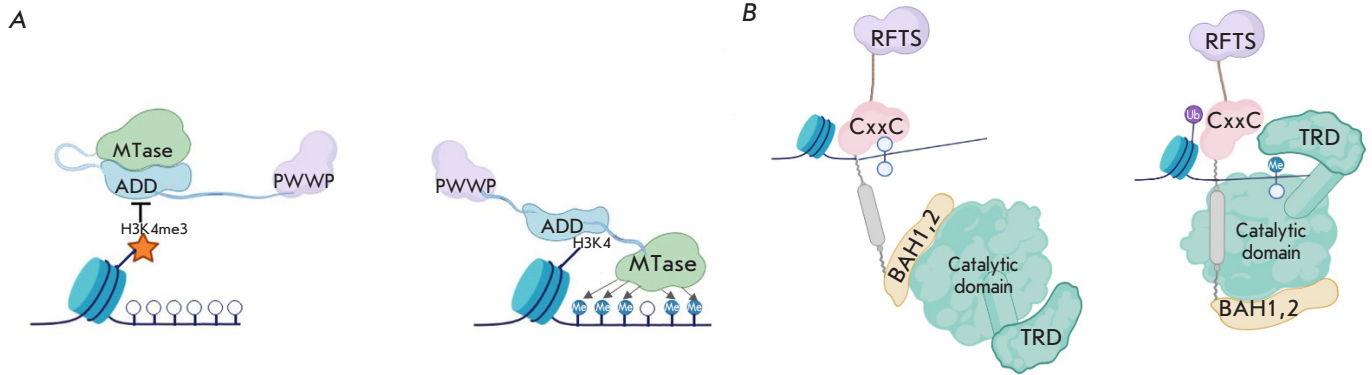


Fig. 3. Binding scheme (A) DNMT3a/3b to unmodified H3K4, the presence of H3K4me3 prevents the ADD domain from binding to DNA, which leads to autoinhibition of the enzyme; (B) DNMT1 to DNA, the interaction with unmethylated DNA leads to inhibition of the catalytic domain [29, 30]

2.1. Maintaining the non-methylated state in DNA regions

About 20% of CpG dinucleotides, most of them CpG islands, avoid methylation. The main factors preventing their methylation include histone modifications, DNA intercatations with certain transcription factors (TF), and the DNA primary and secondary structure.

2.1.1. H3 lysine 4 trimethylation. The presence of trimethylated H3K4 is among the reasons explaining the stability of CpG islands against *de novo* methylation regardless of the transcriptional activity of the region. H3K4me3 prevents the attraction of *de novo* DNA methyltransferases DNMT3a/3b and their regulator, DNMT3L, showing no own catalytic activity to the DNA [28]. DNA methyltransferases DNMT3a/3b include a catalytic domain showing methyltransferase activity (MTase), as well as ADD and PWWP domains involved in chromatin binding. DNA methyltransferases, when in their DNA-unbound form, are inactive due to autoinhibition: the ADD domain interacts with the catalytic domain hindering its activity (Fig. 3). The ADD domain is unable to interact with H3K4me3. At the same time, non-modified H3K4 interacts with the ADD domain of DNMT3a/3b, thereby disrupting ADD binding to the catalytic domain and facilitating the manifestation of methyltransferase activity [28, 29]: Thus, DNA methylation and H3K4 methylation are mutually exclusive phenomena (Fig. 3A).

The case when DNA methyltransferase DNMT1 is maintained is different. DNMT1 is localized in promoter regions, including non-methylated CpG islands,

and is not involved in their methylation. DNMT1 includes the following domains: RFTS (replication foci-targeting sequence), ZF-CxxC, two BAH- (bromo-adjacent homology) domains, and the catalytic domain. The CxxC domain of DNMT1 may bind to sequences including non-methylated CpG dinucleotides. Meanwhile, the BAH1 domain physically intervenes through the interaction between the catalytic domain and the DNA, thereby preventing *de novo* methylation (Fig. 3B) [30].

A particular feature of CpG islands is their ability to bind to TF and enzymes containing the ZF-CxxC domain (CFP1, MLL1/2, KDM2A/2B, TET1/TET3, DNMT1) [31]. Many of these factors are represented by or bind to the histone methyltransferases that modify H3K4, which hinders the attraction of DNA methyltransferases. It should be noted that the lower the gene promoter activity, the higher the need for H3K4me3 to maintain its non-methylated state [32, 33].

2.1.2. TET dioxygenases. TET dioxygenases (ten-eleven translocation) are the enzymes that oxidize methylcytosine for the subsequent excision repair. TET proteins are attracted to the DNA through various mechanisms. TET bind to the CpG islands by their CxxC domain or other transcription factors with a CxxC domain. TET proteins may also be attracted to DNA without the involvement of CpG islands, through messenger proteins, such as Klf4, Nanog, REST, GADD45, CEBPa, etc.; e.g., TET1 and TET2 are attracted to DNA by binding to the TF Nanog, leading to the demethylation of the regulatory gene

regions involved in the maintenance of the pluripotent cellular state [34]. Notably, TET proteins, similar to many CxxC-containing proteins, affect H3K4 trimethylation. TET interact with OGT transferase (O-GLCNAc transferase), which in turn forms a complex with SET1 and MLL histone methyltransferases trimethylating H3K4 [35].

The so-called pioneer factors play a major part in DNA demethylation by TET dioxygenases [36]. They interact with closed, inactive chromatin and change its accessibility for transcription activators. They show their peak activity during organism development, immune system maturation, oncogenesis, and somatic cell reprogramming. Pioneer factors include FOXA1, FOXO, Sox, Pax, GATA, Oct4, PU1, CEBP α , and other TF [37]. The key feature of these factors is their ability to recognize not just a DNA sequence, but a DNA region in a nucleosome context as well [38, 39]. This explains why DNA methylation is not always critical for pioneer factor attraction. In fact, many pioneer factors are methylation-insensitive or have recognition sites that do not contain CpG dinucleotides, which is illustrated by the cases of ASCL1 and FOXA1 [40, 41]. Nevertheless, the pioneer factors Oct4 and Klf4 interact both with sequences not containing CpG and sites containing CpG. In the latter case, Oct4 and Klf4 only bind to the methylated sites [42]. The pioneer factors capable of forming complexes with TET dioxygenases include Klf4, CEBP α , and TF2CP11 [37]. The functional significance of TET2 interaction with Klf4 and CEBP in the process of somatic cell reprogramming has been demonstrated: e.g., the pioneer factors Klf4 and CEBP α attract TET2 dioxygenase to methylated enhancer sequences, which leads to their demethylation and activation [37]. Here, methylation decrease in certain chromatin regions, including in the Klf4 binding sites, is followed by chromatin remodeling. *TET2* knockout cells are not subject to reprogramming [37]. Thus, DNA demethylation by TET enzymes is among the key stages of cellular reprogramming.

Despite the involvement of TET proteins in demethylation in many regions, their removal does not lead to catastrophic changes in the genome-wide DNA methylation level. The main DNA methylation changes in the case of TET knockout have to do with distal regulatory elements and enhancer sequences [43].

2.1.3. DNA secondary structure. Changes in conformation – aka DNA secondary structure – are among the factors contributing to the maintenance of a non-methylated state in CpG islands.

One of these factors is an R-loop, which is an RNA-DNA hybrid and a displaced DNA strand. GADD45A binding to an R-loop in the promoter of the tumor suppressor gene TCF21 attracts TET1, facilitating local demethylation in the region [44]. Thus, the DNA secondary structure may affect DNA demethylation by binding to TET dioxygenases.

G-quadruplexes can also have an effect on the methylation of CpG islands and the CpG dinucleotides not included in the islands. It is an established fact that the GC-rich regulatory regions of eukaryotic genomes are capable of changing local DNA conformation by arranging themselves into alternative structures in the form of G-quadruplexes (G4) [45]. The secondary G-quadruplex (G4) structure is formed by guanine-rich sequences. The G-G base-paired Hoogsteen interaction results in guanine quartet formation, and stacks of such quartets stabilized by potassium cations form the G4 core. The thermodynamic stability of these structures depends on the nucleotide sequence and sometimes exceeds that of the DNA double helix. There are several theoretical and experimental approaches to determining potential G4 regions. Stable G4s formed in the genomic DNA in the presence of potassium ions act as the barrier for DNA polymerase. It often becomes an obstacle for PCR amplification in genome regions including GC-rich sites prone to the formation of G4 structures [46]. The approach based on high-performance sequencing of the errors occurring in the presence of potassium ions is currently considered the best in the experimental prediction of the G4 refolding potential in genomic DNA [47]. A change in the DNA conformation affects its physical and chemical properties and the affinity of various proteins specific to a certain nucleotide sequence. Methylation in the CpG context may change the energy barrier for a transition between the DNA double helix and non-canonical DNA structures, in particular G4 [48]. About 30% of CpG islands include nucleotide sequences capable of forming G4 structures (*Fig. 4A*). Intragenic CpG islands are relatively rich in quadruplex sequences, while the probability of their occurrence in intergenic CpG islands is low. The highest G4 density significantly above the average for all promoters (*Fig. 4C*) is detected in promoter CpG islands (*Fig. 4B*). The maximum G4 density is observed near the transcription start site (TSS). Decreasing G4 occurrence in promoter regions with no CpG islands may be related to the differences in the GC-contents between the promoters overlapping with CpG islands and those removed from them (*Fig. 4D*). The probability of encountering a potential G4-quadruplex

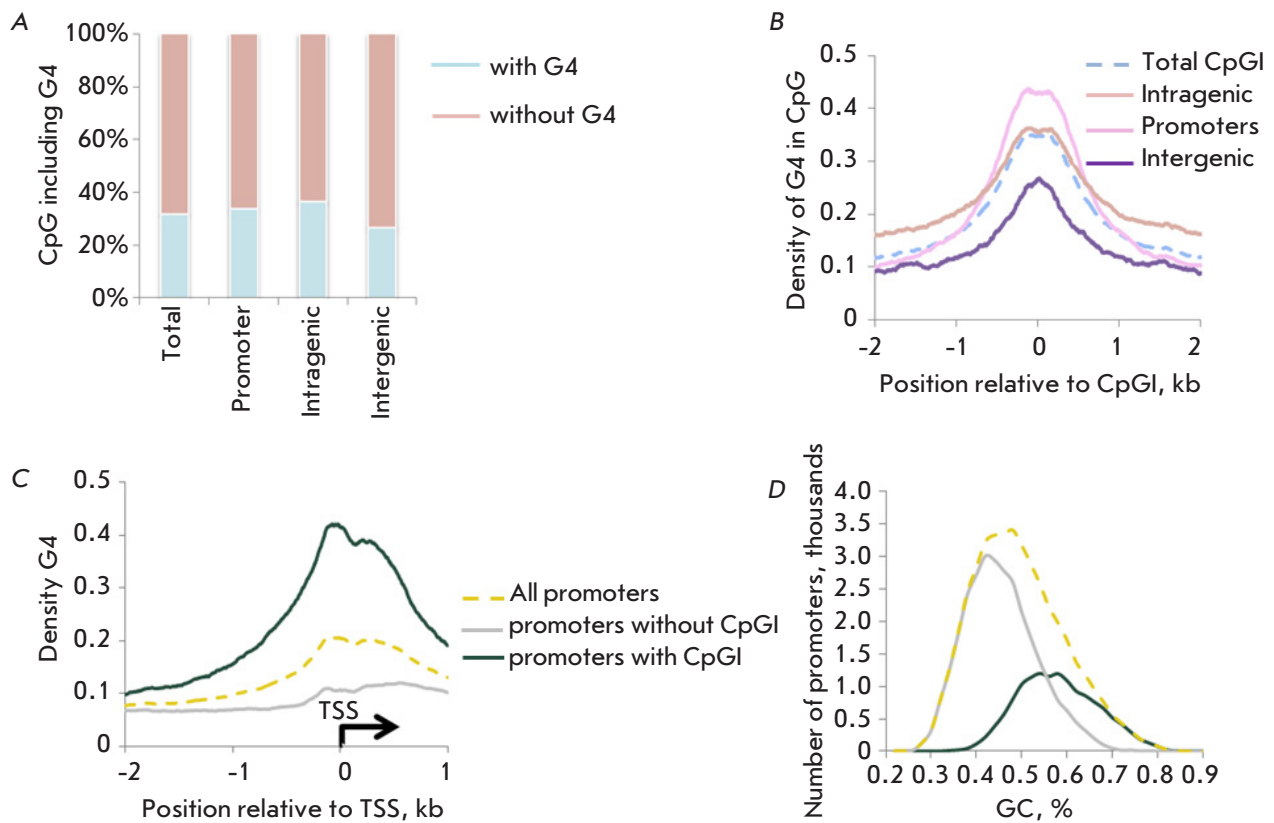


Fig. 4. Distribution of potential G4 sequences in CpG islands. (A) The proportion of CpG islands with G4, (B) the distribution density of G4 near CpG islands depending on the localization in the genome. (C) G4 density and (D) GC composition in promoter regions depending on the presence of CpG islands

depends significantly on the GC-content, even in a randomly generated nucleotide sequence. The probability of encountering a potential G4 in a random sequence with a GC-content of 40% is about one G4 per a million base pairs, while increasing the GC-content in a random sequence to 70% increases the G4 occurrence probability to one per one thousand base pairs [49]. The probability of encountering a G4 sequence in a higher organism genome is above average. These sequences may have an important regulatory role, which is somewhat confirmed by positive evolutionary selection [50]. The presence of G4 in promoters is often associated with transcription suppression [51]. Nevertheless, G4 in stem cells is detected in active promoters and the sites interacting with them; i.e., enhancers, superenhancers, and TF binding sites determining the cell type. In addition to active regulatory ele-

ments, G4s are found in regions with bivalent chromatin modifications; i.e., the ones containing both active and inactive chromatin modifications. A decrease in the detected G4 structures associated with cell differentiation correlates with the occurrence of a closed chromatin [48, 52]. Quadruplex structures may interact with DNA-methyltransferases DNMT1, DNMT3A, and DNMT3B *in vitro* [53, 54]. Indeed, non-methylated sequences in CpG islands containing quadruplexes are rich in DNMT1 binding sites. Meanwhile, interaction between DNMT1 and G4 leads to its DNA methyltransferase inactivation [53]. Thus, G4 formation hinders DNA methylation. This is confirmed by the correlation between the presence of stable quadruplexes in open chromatin and DNA hypomethylation. This correlation is primarily characteristic of sites with a low GC content. Relatively low methylation is also typi-

cal for CpG islands in a closed chromatin containing quadruplexes, compared to regions free of quadruplexes [55].

It is still unclear what DNMT1 activity – specifically the binding to non-methylated CpG sites, when domain positioning hinders the catalytic activity, or interactions with non-canonical DNA structures – is critical in maintaining the non-methylated status of CpG islands. Notably, there are genome regions where DNMT1 binding to DNA manifests *de novo* methyltransferase activity. These regions include LTR-retrotransposons enriched with H3K9me3 and TRIM28. Here, DNMT1 *de novo* activity is regulated by UHRF1 [56]. Thus, the presence of co-factors is also critical for the manifestation of DNMT1 *de novo* activity, in addition to domain positioning.

2.1.4. Competition between transcription factors and DNA methyltransferases. TF binding to DNA can occur with attraction of DNA methyltransferases, thereby protecting DNA from methylation. Sp1 is the classic example of this competition between TF and DNA methyltransferase binding. Sp1 interacts with the non-methylated sequences CCGCCC CpG islands are enriched with and intervenes, with attraction of DNA methyltransferase [57]. Mutation in an Sp1 binding site leads to its increased methylation and reduced transcription [58]. Thus, Sp1 is considered as TF obstructing the methylation of CpG islands. However, unavailability of a recent genome-wide analysis of DNA methylation with Sp1 removed makes it impossible to confirm whether Sp1 is necessary for maintaining the non-methylated status in multiple CpG islands.

CTCF is another factor contributing to the maintenance of a non-methylated DNA state. CTCF is identified as TF binding to non-methylated sequences and capable of acting both as transcription activator and repressor. CTCF also acts as insulator; i.e., it blocks enhancer action on promoters and, therefore, is involved in chromatin structure formation [59]. CTCF binds to non-methylated alleles in imprinted loci, disrupting the enhancer-promoter interaction. CTCF binding to the non-methylated maternal allele in the H19/Igf2 locus is not only critical in terms of enhancer-promoter interaction, it also affects the maintenance of the maternal allele in a non-methylated state. Mutations in the CTCF binding sites in this locus resulted in increased methylation of the maternal allele after ovum fertilization, but methylation in the H19/Igf2 locus in germ cells was not disrupted in [60]. CTCF reduction in the oocytes mediated by RNA interference (RNAi) resulted in increased maternal allele methylation in the locus

of interest in [61, 62]. Thus, CTCF turns out to be critical in maintaining the maternal allele in the H19/Igf2 locus in a non-methylated state. CTCF loss in cancer cells leads to hypermethylation in the protein binding sites as well [63]. According to the genome-wide analysis, CTCF is primarily localized in the non-methylated or poorly methylated regions in the stem cells of mice. Nevertheless, some CTCF binding sites are found to be highly methylated [64]. It turns out that methylation intervenes with the CTCF-DNA interaction only at specific positions of the binding site [65]. Mutation in the methylated CTCF binding sites does not cause changes in the methylation level, despite the fact that the presence of CTCF in methylated sequences correlates with lower methylation compared to the regions lacking CTCF recognition sites [66]. Thus, the interaction between CTCF and methylated sequences has nothing to do with maintenance of the methylation level in these regions. It should be noted that DNA methyltransferase knock-out cells with a reduced DNA methylation level showed no redistribution of CTCF binding sites onto demethylated regions in [67]. Thus, the DNA methylation, on its own, is not an obstacle to CTCF binding. The sites were found in the imprinted H19/Igf2 locus, which CTCF can bind to *in vitro* regardless of their methylation level. It is possible that CTCF is not detected on the methylated allele *in vivo* due to competing binding of methyl-sensitive proteins [68]. Thus, CTCF shows varying DNA binding activity but binding to non-methylated sequences maintains the sequences' low methylation level.

The search for factors protecting DNA from hypermethylation, akin to CTCF or Sp1, could make it possible to study new mechanisms for maintaining the DNA in a non-methylated state and consider them as targets for manipulating DNA methylation and the transcription activity of genes in conditions associated with DNA methylation abnormalities.

2.2. Maintaining DNA regions in methylated state

In this chapter, the processes of DNA methylation onset and maintenance are discussed. They are critical to various repeating sequences, imprinted sites, and regulatory elements. *De novo* DNA methylation involves the DNMT3a and DNMT3b methyltransferases, but, as mentioned above, DNMT1 may manifest *de novo* activity as well. DNMT3a DNA methyltransferase is responsible for methylation onset in the repeating sequences, regulatory elements, and gene bodies acting as Polycomb protein targets. DNMT3b is critical for methylation onset in the regions of satellite repeats and sequences on inactivated X chromosomes [3, 4]. Histone modifications and interac-

tions with transcription factors are important for DNA methyltransferase attraction. Long non-coding RNA and PIWI-interacting non-coding RNA can act as messengers regulating *de novo* methyltransferase binding to DNA as well [69].

2.2.1. Histone modifications. DNMT3 attraction to DNA is achieved using various mechanisms, including histone modifications. As mentioned above, non-modified H3K4 facilitates the binding of DNA methyltransferases through the ADD domain and amplifies their catalytic activation. In addition, DNA methylation is regulated by H3K36me3/me2 histone modifications. DNMT3 methylates the CpG-rich intragenic sequences of actively transcribed genes in the regions characterized by the presence of H3K36me3-modified histones. DNA binding and DNMT3a-mediated methylation in intergenic regions requires H3K36me2. The PWWP domain of DNMT3 is responsible for the interactions with H3K36me2/me3 [70, 71].

DNMT3 binding to heterochromatin and repeating sequences is mediated by H3K9 methylation. DNA methyltransferases are attracted to DNA due to interaction with the histone methyltransferases methylating H3K9 (Suv39h1/2, G9a/GLP, Setdb1) and binding to the HP1 α and HP1 β proteins recognizing methylated H3K9 [72].

2.2.2. Transcription factors attract DNMT to DNA. DNMT3a and DNMT3b are not interchangeable, and their mutations and deletions result in methylation changes in general and specific regions [3, 4]. This has to do with the fact that they are attracted to DNA through interaction with various TFs. As of now, a lot of TFs are being discovered which are capable of interacting with one or both DNA methyltransferases or can be included in a complex with them without interacting directly [73]. Interestingly, these TFs only affect methylation in a limited number of direct targets, which are in many cases restricted to individual target genes. As a result, these TFs may be considered as targets for selective regulation of target gene methylation. Let us discuss some of these factors.

GCNF

GCNF (germ cell nuclear factor) participates in methylation onset and maintenance in various promoter regions by directly interacting with DNMT3a/3b methyltransferases [74]. In addition, GCNF may indirectly attract DNMT3 methyltransferases. GCNF in stem cell differentiation binds to the *Oct4* promoter and interacts with MBD2 and

MBD3, which in turn are included in a single complex with DNMT3. This leads to *Oct4* methylation and its transcription suppression in differentiated cells. Since MBD2/MBD3 cannot bind to *Oct4* during stem cell differentiation with GCNF knockout, the gene remains active [75]. GCNF ability to regulate *Oct4* methylation may be used to analyze a cellular pluripotency status. For instance, GCNF promoter demethylation is observed in somatic cell reprogramming, which enables gene activation during cell differentiation that effectively suppresses *Oct4* transcription. These pluripotent cells are mature, but if their reprogramming is not completed, then GCNF promoter methylation is maintained, gene activation does not occur during cell differentiation, and *Oct4* remains active in differentiated cells, rendering them potentially oncogenic. Thus, GCNF, or more specifically its promoter methylation, can be considered a maturity marker for pluripotent cells.

Kaiso (ZBTB33)

Proteins containing a zinc finger domain often act not only as methyl-DNA-binding proteins, but also as factors contributing to DNA methylation homeostasis [42, 76]. A particular feature of these proteins is their ability to recognize both methylated and non-methylated regions often different in terms of their nucleotide sequences. The zinc finger structure makes it possible to specifically recognize a methylated CG site, most often in a certain context for each TF [77]. The first established proteins to include zinc finger domains interacting with methylated sequences were Kaiso-like proteins: Kaiso (ZBTB33), ZBTB4, and ZBTB38. In addition to zinc fingers, they include the BTB/POZ domain responsible for the protein-protein interaction at their N-end [78–80]. Later, other zinc-finger proteins capable of interacting with the methylated DNA were discovered, including Znf57, CTCF, Klf4, Wt1, and Egr1. The strongest affinity to the methylated DNA is demonstrated by Kaiso and Znf57, binding to methylated sequences over 20 times better than to non-methylated sequences. At the same time, the sensitivity to methylated sequences in the remaining zinc-finger proteins is only 1.5–3 times as high or equal to that for non-methylated sequences [81, 82].

Kaiso binds to methylated sequences and regions including CTGCNA [78, 80]. This protein can act as a transcription repressor, with the BTB/POZ domain at the N-end attracting the NcoR and SMRT corepressor complexes, and as transcription activator [83–85]. Imprinted H19/Igf2 locus is a target for Kaiso that binds to the methylated allele of the locus, and

its removal results in ICR1 methylation decrease in the locus [86, 87]. It is possible that methylation decrease following Kaiso removal is due to competition with CTCF, which in turn can bind to methylated sequences and cause their demethylation. In case of Kaiso knockout, methylation decrease is observed in the *Oct4* promoter in the embryonic fibroblasts of mice and the *TRIM25* promoter in human embryonic renal cells, gene bodies, enhancers, and regions not containing histone modifications [83, 88, 89]. It is shown that *TRIM25* promoter demethylation caused by Kaiso removal is reversible by the expression of exogenous Kaiso, which can be included in a complex with DNMT3a/3b [83, 89]. Notably, Kaiso removal in cancer renal cells in humans causes a slight genome-wide methylation increase. This uniform distribution may be associated with the decrease in TET1 dioxygenase transcription; i.e., Kaiso can shift DNA methylation in both directions. Thus, Kaiso not only maintains the required methylation level, but also participates in methylation onset in various loci by interacting with DNA methyltransferases 3a and 3b [89].

The regulating role of Kaiso in DNA methylation may also be associated with its ability to interact with the ubiquitin-like proteins SUMO1,2,3. The SUMO proteins covalently bind to lysine residues in the target proteins, similarly to ubiquitin. Unlike ubiquitination, SUMOylation usually does not cause protein degradation, while affecting cellular localization, activity, and interaction with other factors. Kaiso SUMOylation affects its transcription properties [83]. The presence of six SIM-SUMO interacting motifs in the Kaiso amino acid sequence and non-covalent interaction between Kaiso and SUMO1 allow us to assume that Kaiso can act as a E3 SUMO ligase. SIM sites are sequences of several hydrophobic amino acid residues surrounded by serine or acidic amino acid residues. The so-called non-canonical E3 SUMO ligases include SIM and non-covalently interact with SUMO [90]. Many proteins are SUMOylated in so-called PML and/or PcG bodies [90, 91]. Kaiso is localized in PcG bodies in the case of exogenous SUMO expression [92]. This allows us to assume that Kaiso not only participates in transcription regulation and DNA methylation maintenance, but may also participate in activity regulation of other factors by affecting their post-translation modifications. For example, SUMOylation of DNA methyltransferases increases their catalytic activity, thereby facilitating an increase in DNA methylation [93]. On the other hand, SUMOylation of the XRC11 excision repair protein is required for effective removal of 5-formyl-

and 5-carboxycytosines in stem cell differentiation and subsequently effective DNA demethylation [94]. That is why studying Kaiso in terms of E3 SUMO ligase and searching for its potential targets makes it possible to uncover new activity regulation mechanisms for various factors, including the proteins contributing to DNA methylation.

Znf57

Unlike Kaiso, Znf57 contains a KRAB (Krueppel-associated box) domain at the N-end. Znf57 binding to methylated sequences using the KRAB domain attracts the TRIM28 (KAP1) corepressor, which forms a complex with H3K9 histone methyltransferase SETDB1 and DNA methyltransferases DNMT1 (maintenance) and DNMT3a/3b (*de novo*) [95]. This repressor complex is formed in the transposon region, imprinted loci, and on inactive enhancers [96, 97]. Znf57 removal causes demethylation in imprinted loci and embryonic death [96]. It should be noted that Znf57 is responsible for methylation maintenance, but not onset.

UHRF1

UHRF1 plays a key role in DNA methylation maintenance in replication. This explains its expression pattern: UHRF1 is only detected in actively dividing cells (for example, spinal cord cells), where DNA methylation onset in the daughter strand is required in replication, and not detected in terminally differentiated cells (neurons, hepatocytes). UHRF1 binds to methylated and semi-methylated DNA using the SRA domain (SET and RING-associated domain). UHRF1 also includes several domains participating in protein-protein interactions: UBL (ubiquitin-like domain), TTD (tandem tudor domain), PHD (plant homeodomain), and RING (a really interesting new gene domain). These domains ensure interaction with maintenance DNA methyltransferase DNMT1, PCNA, histone deacetylase HDAC1, histone methyltransferases G9a, and SUV39H1, PARP1, etc. [98]. UHRF1 binding to semi-methylated DNA in replication ubiquitinates H3K18 and H3K23 and attracts DNMT1 methyltransferase for methylation establishment in the daughter DNA strand. DNMT1 activity is regulated by interaction with H3K18ub and H3K23ub [99]. In pathogenetic tumor conditions, UHRF1 may also affect methylation onset in the promoters of some genes [100]. UHRF1 removal leads to genome instability, G2/M phase arrest, and apoptosis. The absence of double strand break repairs is observed in cells as well [101]. Thus, UHRF1 contributes to DNA methylation establishment and maintenance.

MBD proteins

Methyl-DNA binding proteins with MBD (methyl DNA binding domain) are found among proteins not only recognizing methylated DNA, but also contributing to the binding site methylation. Most MBD proteins are involved in the formation and functioning of the nervous system. There are only four factors in this family (MBD1, MBD2, MBD4, and MeCP2) capable of binding to methylated DNA. These MBD proteins show the strongest affinity to methylated CpG islands [102]. In most cases, these proteins act as interpreters of methylation: i.e. they attract corepressors or compete with transcription activators for DNA binding. However, some recent studies show that these factors can also contribute to DNA methylation establishment and maintenance. It was shown that MeCP2 knockout leads to the occurrence of both hypo- and hypermethylated regions in various types of neurons in mice [103]. The mechanism behind the effect of MBD proteins on the methylation level is yet to be studied. MBD1 regulates methylation in the promoters of the *Htr2c* serotonin receptor gene and *bFGF* growth factor [104, 105]. MBD1 knockout leads to *Htr2c* reactivation, which is considered among the causes of deviations in hippocampal neurogenesis, learning disorders, and occurrence of autism symptoms associated with social behavioral changes, attention deficit, and serotonin activation abnormalities in gene-knockout animals [104]. Reactivation of the *bFGF* growth factor in the case of MBD1 knockout affects the ability to maintain the pluripotent state of stem cells, whose regulation is important for the subsequent differentiation into cells of the nervous system [105].

Thus, there are factors, such as UHRF1, that contribute to genome-wide methylation maintenance and ones (MBD, Kaiso, and GCNF proteins) that regulate methylation in a specific variety of targets. The latter are of special interest, since identification of their binding sites, whose methylation level is affected by the inactivation or mutations of these factors, could make it possible to manipulate the methylation levels in their targets by changing their activity. Interestingly, the desired changes in DNA methylation may also be regulated by activity modulation of DNA methyltransferase using post-translation modifications: Kaiso is a potential E3 SUMO ligase.

3. DNA METHYLATION EDITING

Using advanced DNA editing methods to change methylation levels in certain regions is one of the ways used to alter their transcription activity. This approach makes it possible to alter promoter and

enhancer activity using mutated dCas9 endonuclease incapable of DNA cutting. To ensure DNA hypermethylation, dCas9 is bound to the catalytic domain of DNMT3, whose methyltransferase activity is targeted on the region of interest [106]. The TALEN and zinc finger domain may be used instead of dCas9, but the editing system based around dCas9 remains the most accessible one. The main problems with this editing technique are as follows: 1) the methylation level is not high enough, and 2) DNA demethylation occurs after a certain number of cell divisions. To solve these problems, DNMT3L acting as a cofactor amplifying DNA methylation is added to the catalytic domain of DNMT3. A high methylation level is maintained during a lasting cell division by introducing the chimeric construct dCas9-Ezh2 or dCas9-KRAB into the cells. Ezh2 trimethylates H3K27, and the KRAB domain of Znf57 acts as a base that can be used to assemble a repressor complex that modifies histones and methylates DNA [107]. It is also necessary to identify which factor – Ezh2 or KRAB – would be more effective in suppressing the transcription activity in the region of interest [107].

To ensure DNA demethylation, dCas9 is bound to the catalytic domain of TET proteins [108]. Introduction of the catalytic domain of TET causes not only demethylation, but also 5-hydroxymethylcytosine formation, which contributes to TF attraction [109]. To achieve a more reliable DNA demethylation impact (without cytosine intermediates), dCas9 may be bound to the catalytic domain of ROS1 DNA glycosylase *Arabidopsis* [110].

The key advantage of DNA methylation editing, compared to DNA editing, is that the nucleotide sequence remains intact while only DNA modification changes. These changes are reversible, and almost any sequence in the genome can be edited.

4. DNA METHYLATION AND PATHOGENETIC CONDITIONS

In recent years, the relationship between the regulatory mechanism of DNA methylation and various pathogenetic conditions, especially oncogenesis, rheumatoid arthritis development, and various neurological diseases, has been uncovered [11, 111]. Two categories of clinical significance of the changes in the DNA methylation level can be identified. The first one includes the cases where DNA methylation may act as a marker for a developing pathogenetic condition. The second one includes cases where changes in DNA methylation and the activity of methyl-DNA binding proteins affect the course and progression of the condition.

4.1. DNA methylation as a diagnostic and predictive marker of disease progression

The DNA regions whose methylation changes can be detected in damaged organs or tissues, blood genomic DNA, DNA from various body fluids, and circulating-free DNA are selected as markers of disease progression. Markers making it possible to quite accurately predict oncological diseases at their early stages, evaluate the effect of therapy, detect recurrent cases, and even identify tumor types in some cases, have been selected [112–114].

4.2. DNA methylation as a target for therapy in various pathogenetic conditions

Hypermethylation in the CpG islands located in suppressor gene promoters, leading to their inactivation, is often detected during oncogenesis. Tumor suppressor genes can be activated, albeit inconsistently, through promoter demethylation. For instance, 5-azacitidine reducing DNA methylation is used as an active substance in decitabine used as therapy in acute myeloid leukemia and myelodysplastic syndrome. However, instead of targeting a specific gene, this drug affects the whole genome, causing its instability and damaging the DNA, which may have severe consequences for the patient [115]. Methylation in the promoters of tumor suppressor genes may be reduced by inactivation of the catalytic activity of the maintenance DNA methyltransferase. Inhibitors of DNMT1 DNA methyltransferase RG108 and SG102 are less toxic than 5-azacitidine. They do not change methylation in satellite repeats but affect promoter demethylation, including in some suppressor genes [116, 117]. The key limitation of these inhibitors is the small quantity of targets; i.e., the regulatory elements of suppressor genes. The catalytic activity of DNMT1 may also be suppressed using oligonucleotides that form quadruplex structures [53]. Attempts are made to manipulate DNA methylation using the editing system. The bottleneck of this approach is the delivery of dCas9 or its analogues to target organs and tissues [118]. Hepatocytes, where the editing system can be delivered via injection (for example, tail vein injection in mice), are one of the most accessible targets. Attempts to reduce methylation in the *Fgf21* promoter in the liver of mice have been described. *Fgf21* codes for the factor participating in glucose and cholesterol metabolism. Introduction of dCas9 with the catalytic domain of TET1 resulted in a short-term methylation decrease in the promoter on the sixth day after injection, and as early as the 14th day the methylation level was restored in [119]. Thus, stable DNA methylation editing in a living organism is yet to be achieved.

4.3. Methyl-DNA binding proteins as new targets for therapy

When selecting a therapy target, one should take into consideration how critical is the inactivation of a factor to the organism. Knockout or mutations in the methyl-DNA binding proteins MBD1, MBD2, MeCP2, and Kaiso result primarily in behavioral deviations not disrupting vital processes, which may be reversed upon restoration of protein expression as in the case of MeCP2 [120, 121]. Inactivation of these proteins changes the general methylation level insignificantly and does not lead to genome instability and reactivation of repeating elements. Hence, the MBD proteins Kaiso and their homologue ZBTB4 enjoy an advantage as potential targets. The search for the target genes of these factors associated with pathogenetic conditions seems a promising line of research.

For instance, investigation of the binding sites in methyl-DNA binding proteins made it possible to identify the gamma globin gene as a methyl-dependent target. A gradual transition of hemoglobin types occurs during the human organism's development: the epsilon globin gene is transcribed in the embryonic period; gamma globin – at birth; and beta globin – in adulthood. Patients with the sickle-cell disease and beta thalassemia show an abnormal expression of or mutations in the beta globin gene, leading to severe consequences. Reactivation of the normal form of gamma globin would make it possible to restore a normal hemoglobin level in the blood. The methyl-DNA binding protein MBD2 regulates the attraction of the NuRD corepressor complex to the promoter of the gamma globin gene in blood cells and maintains it in an inactive state in adults [122]. MBD2 removal leads to a 20-fold increase in the expression of the gamma globin gene [123]. Transcription of the gamma globin gene may be activated by disrupting MBD2 binding to the NuRD corepressor complex and its components using inhibitors (*Fig. 5*). Various models have shown that exclusive inactivation of MBD2 does not affect the body function. *MBD2*-knockout mice demonstrate disrupted maternal behavior while nurturing and feeding their offspring [120, 124]. Aside from this, MBD2 removal does not cause any pronounced neurological deviations. Therefore, we can expect MBD2 inhibition to not cause severe side-effects in humans. Thus, the methyl-DNA binding repressor activity of MBD2 may be used for hemoglobin level restoration in patients with sickle-cell disease and beta thalassemia. However, inactivating the methyl-DNA binding protein case cited above is not always necessary. For instance, mutations in or inactivation of

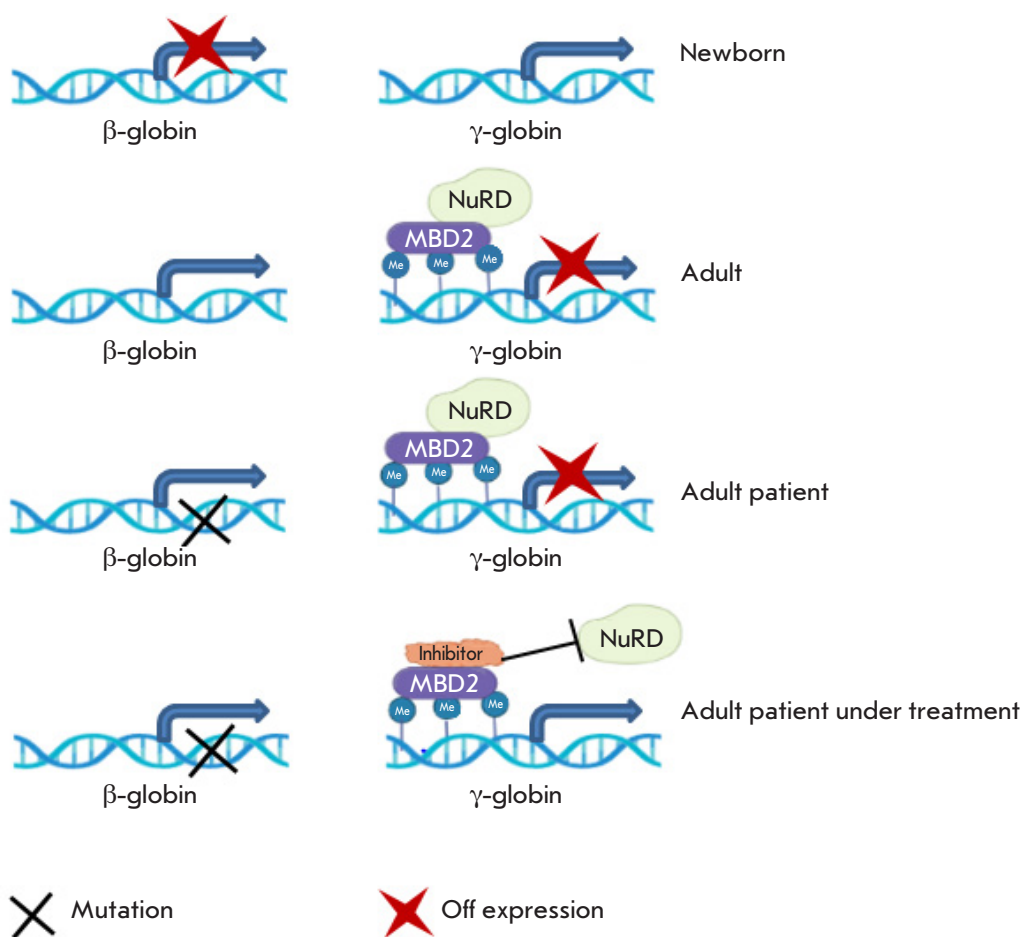


Fig. 5. Model of the functional significance of the interaction between the MBD2 protein and the NuRD repressor complex in the regulation of the gamma globin gene in beta-thalassemia [122, 123]

the methyl-DNA binding protein MeCP2 lead to Rett syndrome development. MeCP2 knockout in mice, similarly to mutations in this gene in humans, causes neurological changes. Notably, changes occurring in nerve cells due to MeCP2 removal or mutation are reversible [125]. The MeCP2 mutations identified in patients with Rett syndrome include, among others, point mutations causing MeCP2 degradation but not affecting the structure of its DNA-binding and repressor domains [126]. When stabilized, this protein can still fulfill its functions [127]. A search for small molecules binding to MeCP2 ubiquitination sites could make it possible to prevent its ubiquitination, with subsequent degradation, and restore the protein's functional activity.

Thus, the search for and characterization of the binding sites in methyl-DNA binding proteins are necessary for the identification of potential targets

whose activity is regulated by DNA methylation and the formation of repressor complexes. Further analysis of the various pathogenetic conditions associated with the target genes of methyl-DNA binding proteins allows us to consider methyl-DNA binding proteins as targets for therapy, while investigation of the mutations in methyl-DNA binding proteins makes it possible to understand when functional changes caused by mutations can be compensated, and when that is impossible.

CONCLUSIONS

DNA methylation is a regulatory element critical to gene expression, genome stabilization, inactivation of repeating sequences, establishment of imprinting, and X-inactivation. Advanced genome-wide sequencing methods allowed us to determine the DNA methylation pattern across the whole genome, in-

cluding various repeating sequences. It opened new opportunities in terms of the identification and characterization of regulatory elements whose activity may be disrupted by various pathogenetic conditions. As of now, a lot of TFs participating in methylation onset and maintenance, demethylation, or interpretation of methylated DNA have been discovered. Methylation can facilitate TF attraction or interfere with it; i.e., the DNA methylation level affects selection of the protein factors interacting with DNA and alternating between attraction of transcription activators and repressors. Discovery of new DNA methylation-dependent factors and investigation of the activating and repressor complexes they are included in allow us to consider these factors as

new therapy targets to be manipulated to achieve a more nuanced effect compared to genome-wide inhibition of DNA methylation. Thus, the study of new methyl-DNA sensitive proteins could make it possible to identify new approaches and therapeutic targets for the management of various pathogenetic conditions associated with DNA methylation onset and regulatory changes. ●

This work was supported by the Russian Foundation for Basic Research (project No. 19-29-04139) (Chapters 3–4) and the Russian Scientific Foundation (project No.19-74-30026) (Chapters 1–2).

Authors have no conflict of interest to declare.

REFERENCES

- Lister R., Mukamel E.A., Nery J.R., Urich M., Puddifoot C.A., Johnson N.D., Lucero J., Huang Y., Dwork A.J., Schultz M.D., et al. // *Science*. 2013. V. 341. № 6146. 1237905.
- Gershman A., Sauria M.E.G., Guitart X., Vollger M.R., Hook P.W., Hoyt S.J., Jaun M., Shumate A., Razaghi R., Koren S., et al. // *Science*. 2022. V. 376. № 6588. eabj5089.
- Yagi M., Kabata M., Tanaka A., Ukai T., Ohta S., Nakabayashi K., Shimizu M., Hata K., Meissner A., Yamamoto T., et al. // *Nat. Commun.* 2020. V. 11. № 1. P. 3199.
- Kato Y., Kaneda M., Hata K., Kumaki K., Hisano M., Kohara Y., Okano M., Li E., Nozaki M., Sasaki H. // *Hum. Mol. Genet.* 2007. V. 16. № 19. P. 2272–2280.
- Dodd T., Yan C., Kossmann B.R., Martin K., Ivanov I. // *Proc. Natl. Acad. Sci. USA*. 2018. V. 115. № 23. P. 5974–5979.
- Shi D.-Q., Ali I., Tang J., Yang W.-C. // *Front. Genet.* 2017. V. 8. P. 100.
- Velasco G., Grillo G., Touleimat N., Ferry L., Ivkovic I., Ribierre F., Deleuze J.-F., Chantalat S., Picard C., Francastel C. // *Hum. Mol. Genet.* 2018. V. 27. № 14. P. 2409–2424.
- Sun Z., Wu Y., Ordog T., Baheti S., Nie J., Duan X., Hojo K., Kocher J.-P., Dyck P.J., Klein C.J. // *Epigenetics*. 2014. V. 9. № 8. P. 1184–1193.
- Wang X., Zhao C., Zhang C., Mei X., Song J., Sun Y., Wu Z., Shi W. // *Cell Commun. Signal.* 2019. V. 17. № 1. P. 1183–1193.
- Salameh Y., Bejaoui Y., Hajj E.N. // *Front. Genet.* 2020. V. 11. P. 171.
- Nishiyama A., Nakanishi M. // *Trends Genet.* 2021. V. 37. № 11. P. 1012–1027.
- Hwang J.-Y., Zukin R.S. // *Curr. Opin. Neurobiol.* 2018. V. 48. P. 193–200.
- Day J.J., Sweatt D.J. // *Nat. Neurosci.* 2010. P. 1319–1323.
- Rollins R.A., Haghghi F., Edwards J.R., Das R., Zhang M.Q., Ju J., Bestor T.H. // *Genome Res.* 2006. V. 16. № 2. P. 157–163.
- Toubiana S., Larom G., Smoom R., Duszynski R.J., Godley L.A., Francastel C., Velasco G., Selig S. // *Hum. Mol. Genet.* 2020. V. 29. № 19. P. 3197–3210.
- Toubiana S., Velasco G., Chityat A., Kaindl A.M., Hershtig N., Tzur-Gilat A., Francastel C., Selig S. // *Hum. Mol. Genet.* 2018. V. 27. № 20. P. 3568–3581.
- Rajshekar S., Yao J., Arnold P.K., Payne S.G., Zhang Y., Bowman T.V., Schmitz R.J., Edwards J.R., Goll M. // *eLife*. 2018. V. 7. e39658. doi: 10.7554/elife.39658
- Scelfo A., Fachinetti D. // *Cells*. 2019. V. 8. № 8. P. 912.
- Guenther M.G., Levine S.S., Boyer L.A., Jaenisch R., Young R.A. // *Cell*. 2007. V. 130. № 1. P. 77–88.
- Mohn F., Weber M., Rebhan M., Roloff T.C., Richter J., Stadler M.B., Bibel M., Schübeler D. // *Mol. Cell*. 2008. V. 30. № 6. P. 755–766.
- Sproul D., Nestor C., Culley J., Dickson J.H., Dixon M., Harrison D.J., Meehan R.R., Sims A.H., Ramsahoye B.H. // *Proc. Natl. Acad. Sci. USA*. 2011. V. 108. P. 4364–4369.
- Wan J., Oliver V.F., Wang G., Zhu H., Zack D.J., Merbs S.L., Qian J. // *BMC Genomics*. 2015. V. 16. № 1. P. 49.
- Illingworth R.S., Gruenewald-Schneider U., Webb S., Kerr A.R.W., James K.D., Turner D.J., Smith C., Harrison D.J., Andrews R., Bird A.P., et al. // *PLoS Genet.* 2010. V. 6. № 9. e1001134.
- Jeziorska D.M., Murray R.J.S., De Gobbi M., Gaentzsch R., Garrick D., Ayyub H., Chen T., Li E., Telenius J., Lynch M., et al. // *Proc. Natl. Acad. Sci. USA*. 2017. V. 114. № 36. P. E7526–E7535.
- Wu H., Coskun V., Tao J., Xie W., Ge W., Yoshikawa K., Li E., Zhang Y., Sun Y.E. // *Science*. 2010. V. 329. № 5990. P. 444–448.
- Bell J.S.K., Vertino P.M. // *Epigenetics*. 2017. V. 12. № 6.

- P. 449–464.
27. Pachano T, Sánchez-Gaya V, Ealo T, Mariner-Faulí M, Bleckwehl T, Asenjo H.G., Respuela P, Cruz-Molina S, Muñoz-San Martín M, Haro E., et al. // *Nat. Genet.* 2021. V. 53. № 7. P. 1036–1049.
 28. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R., Fischle W., et al. // *Nucl. Acids Res.* 2010. V. 38. № 13. P. 4246–4253.
 29. Otani J, Nankumo T, Arita K., Inamoto S., Ariyoshi M., Shirakawa M. // *EMBO Rep.* 2009. V. 10. № 11. P. 1235–1241.
 30. Zhang Z.-M., Liu S., Lin K., Luo Y., Perry J.J., Wang Y., Song J. // *J. Mol. Biol.* 2015. V. 427. № 15. P. 2520–2531.
 31. Long H.K., Blackledge N.P., Klose R.J. // *Biochem. Soc. Trans.* 2013. V. 41. № 3. P. 727–740.
 32. Brown D.A., Di Cerbo V., Feldmann A., Ahn J., Ito S., Blackledge N.P., Nakayama M., McClellan M., Dimitrova E., Turberfield A.H., et al. // *Cell Rep.* 2017. V. 20. № 10. P. 2313–2327.
 33. Clouaire T, Webb S, Skene P, Illingworth R, Kerr A., Andrews R., Lee J.-H., Skalnik D., Bird A. // *Genes Dev.* 2012. V. 26. № 15. P. 1714–1728.
 34. Costa Y, Ding J, Theunissen T.W., Faiola F, Hore T.A., Shliaha P.V., Fidalgo M., Saunders A., Lawrence M., Dietmann S., et al. // *Nature.* 2013. V. 495. P. 370–374.
 35. Deplus R., Delatte B., Schwinn M.K., Defrance M., Méndez J., Murphy N., Dawson M.A., Volkmar M., Putmans P., Calonne E., et al. // *EMBO J.* 2013. V. 32. № 5. P. 645–655.
 36. Sardina J.L., Collombet S., Tian T.V., Gómez A., Di Stefano B., Berenguer C., Brumbaugh J., Stadhouders R., Segura-Morales C., Gut M., et al. // *Cell Stem Cell.* 2018. V. 23. № 6. P. 727–741.e9.
 37. Fernandez Garcia M., Moore C.D., Schulz K.N., Alberto O., Donague G., Harrison M.M., Zhu H., Zaret K.S. // *Mol. Cell.* 2019. V. 75. № 5. P. 921–932.e6.
 38. Michael A.K., Grand R.S., Isbel L., Cavadini S., Koziacka Z., Kempf G., Bunker R. D., Schenk A.D., Graff-Meyer A., Pathare G.R., et al. // *Science.* 2020. V. 368. № 6498. P. 1460–1465.
 39. Donovan B.T., Chen H., Jipa C., Bai L., Poirier M.G. // *Elife.* 2019. V. 8. № 7. e43008.
 40. VandenBosch L.S., Wohl S.G., Wilken M.S., Hooper M., Finkbeiner C., Cox K., Chipman L., Reh T.A. // *Sci. Rep.* 2020. V. 1. P. 13615.
 41. Metzakopian E., Bouhali K., Alvarez-Saavedra M., Whitsett J.A., Picketts D.J., Ang S.-L. // *Development.* 2015. V. 142. № 7. P. 1315–1324.
 42. Yin Y., Morgunova E., Jolma A., Kaasinen E., Sahu B., Khund-Sayeed S., Das P.K., Kivioja T., Dave K., Zhong F., et al. // *Science.* 2017. V. 356. № 6337. aaj2239.
 43. Lu F., Liu Y., Jiang L., Yamaguchi S., Zhang Y. // *Genes Dev.* 2014. V. 28. № 19. P. 2103–2119.
 44. Arab K., Karaulanov E., Musheev M., Trnka P., Schäfer A., Grummt I., Niehrs C. // *Nat. Genet.* 2019. V. 51. P. 217–223.
 45. Marsico G., Chambers V.S., Sahakyan A.B., McCauley P., Boutell J.M., Antonio M.D., Balasubramanian S. // *Nucl. Acids Res.* 2019. V. 47. № 8. P. 3862–3874.
 46. Chashchina G.V., Beniaminov A.D., Kaluzhny D.N. // *Biochemistry (Moscow).* 2019. V. 84. № 5. P. 562–569.
 47. Chambers V.S., Marsico G., Boutell J.M., Di Antonio M., Smith G.P., Balasubramanian S. // *Nat. Biotechnol.* 2015. V. 33. № 8. P. 877–881.
 48. Isaakova E., Varizhuk A., Pozmogova G. // *Signif. Bio- eng. Biosci.* 2018. V. 1. № 3. P. 1–7.
 49. Chashchina G.V., Shchuyolkina A.K., Kolosov S.V., Beniaminov A.D., Kaluzhny D.N. // *Front. Microbiol.* 2021. V. 12. P. 647851.
 50. Wu F., Niu K., Cui Y., Li C., Lyu M., Ren Y., Chen Y., Deng H., Huang L., Zheng S., et al. // *Commun. Biol.* 2021. V. 4. № 1. P. 98.
 51. Stevens A.J., de Jong L., Kennedy M.A. // *Int. J. Mol. Sci.* 2022. V. 23. № 5. P. 2407.
 52. Hasegawa H., Sasaki I., Tsukakoshi K., Ma Y., Nagasawa K., Numata S., Inoue Y., Kim Y., Ikebukuro K. // *Int. J. Mol. Sci.* 2021. V. 22. № 23. P. 13159.
 53. Mao S.-Q., Ghanbarian A.T., Spiegel J., Martínez Cuesta S., Beraldi D., Di Antonio M., Marsico G., Hänsel-Hertsch R., Tannahill D., Balasubramanian S. // *Nat. Struct. Mol. Biol.* 2018. V. 25. P. 951–957.
 54. Cree S.L., Fredericks R., Miller A., Pearce F.G., Filichev V., Fee C., Kennedy M. A. // *FEBS Lett.* 2016. V. 590. № 17. P. 2870–2883.
 55. Jara-Espejo M., Line S.R. // *FEBS J.* 2020. V. 287. № 3. P. 483–495.
 56. Haggerty C., Kretzmer H., Riemenschneider C., Kumar A.S., Mattei A.L., Bailly N., Gottfreund J., Giesselmann P., Weigert R., Brändl B., et al. // *Nat. Struct. Mol. Biol.* 2021. V. 28. № 7. P. 594–603.
 57. Tian H.-P., Lun S.-M., Huang H.-J., He R., Kong P.-Z., Wang Q.-S., Li X.-Q., Feng Y.-M. // *J. Biol. Chem.* 2015. V. 290. № 31. P. 19173–19183.
 58. Lienert F., Wirbelauer C., Som I., Dean A., Mohn F., Schübeler D. // *Nat. Genet.* 2011. V. 43. № 11. P. 1091–1097.
 59. Braccioli L., de Wit E. // *Essays Biochem.* 2019. V. 63. № 1. P. 157–165.
 60. Schoenherr C.J., Levorse J.M., Tilghman S.M. // *Nat. Genet.* 2003. V. 33. № 1. P. 66–69.
 61. Fedoriw A.M., Stein P., Svoboda P., Schultz R.M., Bartolomei M.S. // *Science.* 2004. V. 303. № 5655. P. 238–240.
 62. Freschi A., Del Prete R., Pignata L., Cecere F., Manfredola F., Mattia M., Cobellis G., Sparago A., Bartolomei M.S., Riccio A., et al. // *Hum. Mol. Genet.* 2021. V. 30. № 16. P. 1509–1520.
 63. Damaschke N.A., Gawdzik J., Avilla M., Yang B., Svaren J., Roopra A., Luo J.-H., Yu Y.P., Keles S., Jarrard D.F., et al. // *Clin. Epigenetics.* 2020. V. 12. № 1. P. 80.
 64. Luo X., Zhang T., Zhai Y., Wang F., Zhang S., Wang G. // *Front. Genet.* 2021. V. 12. P. 639461.
 65. Hashimoto H., Wang D., Horton J.R., Zhang X., Corces V.G., Cheng X. // *Mol. Cell.* 2017. V. 66. № 5. P. 711–720.
 66. Stadler M.B., Murr R., Burger L., Ivanek R., Lienert F., Schöler A., van Nimwegen E., Wirbelauer C., Oakeley E.J., Gaidatzis D., et al. // *Nature.* 2011. V. 480. P. 490–495.
 67. Reshef Y.A., Finucane H.K., Kelley D.R., Gusev A., Kotliar D., Ulirsch J.C., Hormozdiari F., Nasser J., O'Connor L., van de Geijn B., et al. // *Nat. Genet.* 2018. V. 50. № 10. P. 1483–1493.
 68. Kurukuti S., Tiwari V.K., Tavoosidana G., Pugacheva E., Murrell A., Zhao Z., Lobanenko V., Reik W., Ohlsson R. // *Proc. Natl. Acad. Sci. USA.* 2006. V. 103. № 28. P. 10684–10689.
 69. Schmitz K.-M., Mayer C., Postepska A., Grummt I. // *Genes Dev.* 2010. V. 24. № 20. P. 2264–2269.
 70. Baubec T., Colombo D.F., Wirbelauer C., Schmidt J., Burger L., Krebs A.R., Akalin A., Schubeler D. // *Nature.* 2015. V. 520. P. 243–247.
 71. Weinberg D.N., Papillon-Cavanagh S., Chen H., Yue Y., Chen X., Rajagopalan K.N., Horth C., McGuire J.T., Xu X.,

- Nikbakht H. // *Nature*. 2019. V. 573. P. 281–286.
72. Li Y., Chen X., Lu C. // *EMBO Rep*. 2021. V. 22. № 5. e51803.
73. Hervouet E., Vallette F.M., Cartron P.-F. // *Epigenetics*. 2009. V. 4. № 7. P. 487–499.
74. Sato N., Kondo M., Arai K.-I. // *Biochem. Biophys. Res. Commun.* 2006. V. 344. № 3. P. 845–851.
75. Gu P., Le Menuet D., Chung A.C.-K., Cooney A.J. // *Mol. Cell. Biol.* 2006. V. 26. № 24. P. 9471–9483.
76. Tullius T., Parker S. // *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. 2013. V. 152. № 1–2.
77. Hudson N.O., Buck-Koehntop B.A. // *Molecules*. 2018. V. 23. № 10. P. 2555.
78. Prokhortchouk A., Hendrich B., Jørgensen H., Ruzov A., Wilm M., Georgiev G., Bird A., Prokhortchouk E. // *Genes Dev.* 2001. V. 15. № 13. P. 1613–1618.
79. Filion G.J.P., Zhenilo S., Salozhin S., Yamada D., Prokhortchouk E., Defossez P.-A. // *Mol. Cell Biol.* 2006. V. 26. № 1. P. 169–181.
80. Daniel J.M., Spring C.M., Crawford H.C., Reynolds A.B., Baig A. // *Nucl. Acids Res.* 2002. V. 30. № 13. P. 2911–2919.
81. Nikolova E.N., Stanfield R.L., Dyson H.J., Wright P.E. // *Biochemistry*. 2018. V. 57. № 14. P. 2109–2120.
82. Liu Y., Toh H., Sasaki H., Zhang X., Cheng X. // *Genes Dev.* 2012. V. 26. № 21. P. 2374–2379.
83. Zhenilo S., Deyev I., Litvinova E., Zhigalova N., Kaplun D., Sokolov A., Mazur A., Prokhortchouk E. // *Cell Death Differ.* 2018. V. 25. № 11. P. 1938–1951.
84. Yoon H.-G., Chan D.W., Reynolds A.B., Qin J., Wong J. // *Mol. Cell.* 2003. V. 12. № 3. P. 723–734.
85. Raghav S.K., Waszak S.M., Krier I., Gubelmann C., Isakova A., Mikkelsen T.S., Deplancke B. // *Mol. Cell.* 2012. V. 46. № 3. P. 335–350.
86. Bohne F., Langer D., Martiné U., Eider C.S., Cencic R., Begemann M., Elbracht M., Bülow L., Eggermann T., Zechner U., et al. // *Clin. Epigenet.* 2016. V. 8. P. 47.
87. Prokhortchouk A., Sansom O., Selfridge J., Caballero I.M., Salozhin S., Aithozhina D., Cerchietti L., Guo Meng F., Augenlicht L.H., Mariadason J.M., et al. // *Mol. Cell. Biol.* 2006. V. 26. № 1. P. 199–208.
88. Kaplun D.S., Fok R.E., Korostina V.S., Prokhortchouk E.B., Zhenilo S.V. // *Biochemistry (Moscow)*. 2019. V. 84. № 3. P. 283–290.
89. Kaplun D., Starshin A., Sharko F., Gainova K., Filonova G., Zhigalova N., Mazur A., Prokhortchouk E., Zhenilo S., et al. // *Int. J. Mol. Sci.* 2021. V. 22. № 14. P. 7587.
90. Shi X., Du Y., Li S., Wu H. // *Int. J. Mol. Sci.* 2022. V. 23. № 7. P. 3639.
91. Kagey M.H., Melhuish T.A., Wotton D. // *Cell*. 2003. V. 113. № 1. P. 127–137.
92. Zhenilo S., Kaplun D., Prokhortchouk E. // *FEBS Open Bio*. 2018. V. 8 (S1). P. 134.
93. Lee B., Muller M.T. // *Biochem. J.* 2009. V. 421. № 3. P. 449–461.
94. Steinacher R., Berekati Z., Botev P., Kuśnierczyk A., Slupphaug G., Schär P. // *EMBO J.* 2019. V. 38. № 1. e99242.
95. Quenneville S., Verde G., Corsinotti A., Kapopoulou A., Jakobsson J., Offner S., Baglivo I., Pedone P.V., Grimaldi G., Riccio A., et al. // *Mol. Cell.* 2011. V. 44. № 3. P. 361–372.
96. Riso V., Cammisa M., Kukreja H., Anvar Z., Verde G., Sparago A., Acurzio B., Lad S., Lonardo E., Sankar A., et al. // *Nucl. Acids Res.* 2016. V. 44. № 17. P. 8165–8178.
97. Shi H., Strogantsev R., Takahashi N., Kazachenka A., Lorincz M.C., Hemberger M., Ferguson-Smith A.C. // *Epigenetics Chromatin*. 2019. V. 12. № 1. P. 49.
98. Mancini M., Magnani E., Macchi F., Bonapace I.M. // *Nucl. Acids Res.* 2021. V. 49. № 11. P. 6053–6068.
99. Qin W., Wolf P., Liu N., Link S., Smets M., La Mastra F., Forné I., Pichler G., Hörl D., Fellingner K., et al. // *Cell Res.* 2015. V. 25. № 8. P. 911–929.
100. Beck A., Trippel F., Wagner A., Joppien S., Felle M., Vokuhl C., Schwarzmayr T., Strom T.M., von Schweinitz D., Längst G., et al. // *Clin. Epigenetics*. 2018. V. 10. P. 27.
101. Tian Y., Paramasivam M., Ghosal G., Chen D., Shen X., Huang Y., Akhter S., Legerski R., Chen J., Seidman M.M., et al. // *Cell Rep.* 2015. V. 10. № 12. P. 1957–1966.
102. Baubec T., Ivánek R., Lienert F., Schübeler D. // *Cell*. 2013. V. 153. P. 480–492.
103. Jin Y., Su K., Kong H.E., Ma W., Wang Z., Li Y., R. Li, Allen E. G., Wu H., Jin P. // *Hum. Mol. Genet.* 2022. ddac189.
104. Allan A.M., Liang X., Luo Y., Pak C., Li X., Szulwach K.E., Chen D., Jin P., Zhao X. // *Hum Mol. Genet.* 2008. V. 17. № 13. P. 2047–2057.
105. Li X., Barkho B.Z., Luo Y., Smrt R.D., Santistevan N.J., Liu C., Kuwabara T., Gage F. H., Zhao X. // *J. Biol. Chem.* 2008. V. 283. № 41. P. 27644–27652.
106. Katayama S., Andou M. // *Biochem. Biophys. Res. Commun.* 2021. V. 581. P. 20–24.
107. O’Geen H., Bates S.L., Carter S.S., Nisson K.A., Halmaj J., Fink K.D., Rhie S.K., Farnham P.J., Segal D.J. // *Epigenetics Chromatin*. 2019. V. 12. P. 26.
108. Choudhury S.R., Cui Y., Lubecka K., Stefanska B., Iru-dayaraj J. // *Oncotarget*. 2016. V. 7. № 29. P. 46545–46556.
109. Kang J.G., Park J.S., Ko J.-H., Kim Y.-S. // *Sci. Rep.* 2019. V. 9. № 1. P. 11960.
110. Devesa-Guerra I., Morales-Ruiz T., Pérez-Roldán J., Parrilla-Doblas J.T., Dorado-León M., García-Ortiz M.V., et al. // *J. Mol. Biol.* 2020. V. 432. № 7. P. 2204–2216.
111. Ciechomska M., Roszkowski L., Maslinski W. // *Cells*. 2019. V. 8. № 9. P. 953.
112. Jung G., Hernández-Illán E., Moreira L., Balaguer F., Goel A. // *Nat. Rev. Gastroenterol. Hepatol.* 2020. V. 17. № 2. P. 111–130.
113. Müller D., Gyórfy B. // *Biochim. Biophys. Acta Rev. Cancer*. 2022. V. 1877. № 3. 188722.
114. Taryma-Leśniak O., Sokolowska K.E., Wojdacz T.K. // *Clin Epigenetics*. 2020. V. 12. № 1. P. 107.
115. Brocks D., Schmidt C.R., Daskalakis M., Jang H.S., Shah N.M., Li D., Li J., Zhang B., Hou Y., Laudato S., et al. // *Nat. Genet.* 2017. V. 49. № 7. P. 1052–1060.
116. Graça I., Sousa E.J., Baptista T., Almeida M., Ramalho-Carvalho J., Palmeira C., Henrique R., Jerónimo C. // *Curr. Pharm. Des.* 2014. V. 20. № 11. P. 1803–1811.
117. Segura-Pacheco B., Perez-Cardenas E., Taja-Chayeb L., Chavez-Blanco A., Revilla-Vazquez A., Benitez-Bribiesca L., Duenas-González A. // *J. Transl. Med.* 2006. V. 4. P. 32.
118. Ansari I., Chaturvedi A., Chitkara D., Singh S. // *Semin. Cancer Biol.* 2022. V. 83. P. 570–583.
119. Hanzawa N., Hashimoto K., Yuan X., Kawahori K., Tsujimoto K., Hamaguchi M., Tanaka T., Nagaoka Y., Nishina H., Morita S., et al. // *Sci. Rep.* 2020. V. 10. № 1. P. 5181.
120. Wood K.H., Johnson B.S., Welsh S.A., Lee J.Y., Cui Y., Krizman E., Brodtkin E.S., Blendy J.A., Robinson M.B., Bartolomei M.S., et al. // *Epigenomics*. 2016. V. 8. № 4. P. 455–473.

REVIEWS

121. Kulikov A.V., Korostina V.S., Kulikova E.A., Fursenko D.V., Akulov A.E., Moshkin M.P., Prokhortchouk E.B. // *Behav. Brain Res.* 2016. V. 297. P. 76–83.
122. Gnanapragasam M.N., Scarsdale J.N., Amaya M.L., Webb H.D., Desai M.A., Walavalkar N.M., Wang S.Z., Zhu S.Z., Ginder G.D., Williams Jr., D.C. // *Proc. Natl. Acad. Sci. USA.* 2011. V. 108. № 18. P. 7487–7492.
123. Yu X., Azzo A., Bilinovich S.M., Li X., Dozmorov M., Kurita R., Nakamura Y., Williams Jr., D.C., Ginder G.D. // *Haematologica.* 2019. V. 104. № 12. P. 2361–2371.
124. Hendrich B., Guy J., Ramsahoye B., Wilson V.A., Bird A. // *Genes Dev.* 2001. V. 15. № 6. P. 710–723.
125. Robinson L., Guy J., McKay L., Brockett E., Spike R.C., Selfridge J., De Sousa D., Merusi C., Riedel G., Bird A., et al. // *Brain.* 2012. V. 135. P. 9. P. 2699–2710.
126. Scarsdale J.N., Webb H.D., Ginder G.D., Williams D.C., Jr. // *Nucl. Acids Res.* 2011. V. 39. № 15. P. 6741–6752.
127. Ghosh R.P., Horowitz-Scherer R.A., Nikitina T., Gierasch L.M., Woodcock C.L. // *J. Biol. Chem.* 2008. V. 283. № 29. P. 20523–20534.