CTCF As an Example of DNA-Binding Transcription Factors Containing Clusters of C2H2-Type Zinc Fingers

O. G. Maksimenko[1,2*], D. V. Fursenko[1], E. V. Belova[1,2], P. G. Georgiev[1*]
[1]Institute of Gene Biology RAS, Moscow, 119334 Russia
[2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology RAS, Moscow, 119334 Russia
[*]E-mail: maksog@mail.ru; georgiev_p@mail.ru
Received September 17, 2020; in final form, November 12, 2020
DOI: 10.32607/actanaturae.11206

ABSTRACT **In mammals, most of the boundaries of topologically associating domains and all well-studied insulators are rich in binding sites for the CTCF protein. According to existing experimental data, CTCF is a key factor in the organization of the architecture of mammalian chromosomes. A characteristic feature of the CTCF is that the central part of the protein contains a cluster consisting of eleven domains of C2H2-type zinc fingers, five of which specifically bind to a long DNA sequence conserved in most animals. The class of transcription factors that carry a cluster of C2H2-type zinc fingers consisting of five or more domains (C2H2 proteins) is widely represented in all groups of animals. The functions of most C2H2 proteins still remain unknown. This review presents data on the structure and possible functions of these proteins, using the example of the vertebrate CTCF protein and several well- characterized C2H2 proteins in Drosophila and mammals.**
KEYWORDS **C2H2-type zinc fingers, architectural proteins, transcription regulation, insulators, TAD, enhancers, promoters, CTCF.**
ABBREVIATIONS **kbp – kilobase pairs; TAD – topologically associating domain; CTCF – CCCTC-binding factor; C2H2 – a domain consisting of two cysteine and two histidine residues coordinated to a zinc ion.**

INTRODUCTION

Cell differentiation in higher eukaryotes has led to significant complication in the regulation of gene expression. Cell specialization is determined by transcription factor repertoires assembling on regulatory elements of the genome. The genes responsible for cell differentiation are usually regulated by enhancers, each of which activates a promoter in a particular group of cells for a specific time interval [1–3]. In some cases, transcription of the developmental genes is regulated by several dozens of enhancers; the distance between some of these enhancers and the regulated promoter can reach up to several hundred kilobase pairs.

The ability of enhancers to perform long-range stimulation of promoters has led to the assumption that there may be some specialized transcription domains within which contacts between enhancers and promoters occur more efficiently [4]. It was believed that at the boundaries of transcription domains there are special regulatory elements capable of blocking interactions between enhancers and promoters [5, 6]. The most common opinion was that domain boundaries interact either with each other or with the nuclear structures bound to the nuclear envelope. Indeed, regulatory elements with the predicted properties were found first in Drosophila and then in mammals; these elements became known as insulators [7]. The

two main properties of insulators have been described using the model systems in transgenic Drosophila lines: their ability to block the enhancer–promoter contacts and that to prevent repression of transgene expression during its integration into the heterochromatin regions within the genome [5, 6].

The emergence of methods for genome-wide identification of contacts between chromatin regions *in vivo*, and high-resolution microscopy [8–11], took the study of the spatial organization of the genome to a completely new level. It turns out that the chromosomes of all eukaryotes are organized into topologically associating domains (TADs), which are formed through predominant interaction between the ends or boundaries of the domains [12–15]. In this case, contacts within a TAD form much more efficiently than contacts between sequences located in adjacent TADs.

The discovery of TADs gave grounds to assume that their boundaries correspond to the insulators that restrict independent regulatory domains [16–18]. However, studies carried out on single cells have shown that TAD boundaries form as a set of preferred contacts and are not strict physical barriers blocking any trans-interactions between regulatory elements located in different TADs [12, 14, 19, 20]. Most of the characterized insulators are located within the same TAD. The improvement in the resolution of contact maps within the TAD has led to the discovery of subdomains, which usually correspond to local contacts between regulatory elements [19].

CTCF AS THE BEST STUDIED PROTEIN WITH A C2H2 ZINC FINGER CLUSTER IN MAMMALS

The vertebrate protein CTCF (CCCTC binding factor), which has been well-studied in humans and mice [21, 22], is expressed at all ontogenetic stages in all cell types and is required during embryogenesis. Depending on the context, CTCF can act as a transcriptional activator or repressor. It is involved in the inactivation of one of the X chromosomes in mammals, it regulates alternative splicing of pre-mRNA in some genes, controls imprinting, participates in recombination and repair, and is responsible for the activity of enhancers, promoters, and insulators. However, the key role played by vertebrate CTCF in the chromosomal architecture is what has been described most thoroughly [23–25]. Mammalian genomes contain from 40,000 to 80,000 CTCF binding sites, with over 5,000 sites being conserved in different species and cell lines [21, 26]. Approximately 50%, 15%, and 35% of the CTCF binding sites are located, respectively, in intergenic regions, near promoters, and within gene bodies (30% residing in introns and 5% residing in exons) [27]. Mammalian CTCF consists of non-structured terminal regions and eleven zinc fingers residing in the central part of the protein; the first ten zinc fingers are C2H2-type, and the last one is C2HC-type. It is worth noting that proteins containing one or, less frequently, several clusters of C2H2 zinc finger domains constitute a significant portion of all the C2H2 zinc finger proteins [28]. The classical C2H2 domain has the consensus sequence $CX_{2\text{-}4}CX_{12}HX_{2\text{-}8}H$. In the presence of a zinc ion, this sequence folds to form a ββα structure, where zinc is tetrahedrally coordinated by two cysteine residues at one end of the β-sheet and two histidine residues at the C-ends of α-helices. The structure is stabilized by hydrophobic bonds. In the canonical complex, the α-helical sections of tandem C2H2 zinc fingers are located in the major groove of DNA. The high-affinity specific binding is ensured by specific interactions with nitrogenous bases and nonspecific contacts with the phosphate backbone of DNA. For any DNA triplet, it is possible to choose a C2H2 domain carrying the desired amino acids at key positions of the α-helix and specifically recognizing this triplet [29–31]. Therefore, just within a few years after the first description of the structure of the C2H2 domains bound to DNA, chimeric proteins consisting of a C2H2 domain cluster and the FokI domain introducing double-strand breaks in the DNA sequence started being actively used as site-specific endonucleases for targeted genome editing [32, 33].

In proteins containing a C2H2-type zinc finger cluster, short 5-aa linkers residing between the domains possess the consensus sequence TGEKP and are a characteristic feature of DNA-binding C2H2 proteins [34]. The linkers are critical in terms of the affinity and specificity of DNA binding; mutations in them may cause a loss of the protein's function *in vivo* [35, 36]. It is believed that each amino acid residue within the linker plays its own role in the interaction with DNA. Flexible in its unbound state, the protein structure, consisting of several C2H2 domains, "latches itself" as soon as it binds to the correct DNA sequence. The OH group of the first threonine residue T1 (or serine residue) forms a hydrogen bond with the amide group of glutamic acid E3; glycine G2 ensures the flexibility of the main chain required for latching. Glutamic acid E3 can contribute to the stabilization of the contacts between the zinc fingers. The lysine residue K4 (or arginine residue) is in contact with the DNA phosphate backbone. Proline residue P5 probably strengthens the bond between the linker and the following zinc finger; it also immobilizes the following conserved phenylalanine or tyrosine residue, whose aromatic ring lies on the N-end of the α-helix [37]. TGEKP-like linkers also connect the DNA-binding C2H2 domains of human CTCF *(Fig. 1)*.

The conformational changes in the DNA structure introduced by the C2H2 domains during binding limit the potential number of C2H2 domains connected by short linkers and capable of cooperatively interacting with DNA, and, therefore, limit the length of the canonical binding site [37]. This is probably why only four or five C2H2 domains are involved in the interaction and specific recognition of a 12–15 bp long DNA site in most proteins. Studies with artificial C2H2 clusters have shown that the specificity of protein binding to DNA increases when several short DNA-recognizing C2H2-domain clusters are connected by longer non-canonical linkers [28]. Therefore, it can be assumed that proteins carrying a large number of C2H2 domains in a cluster can specifically recognize different DNA sequences.

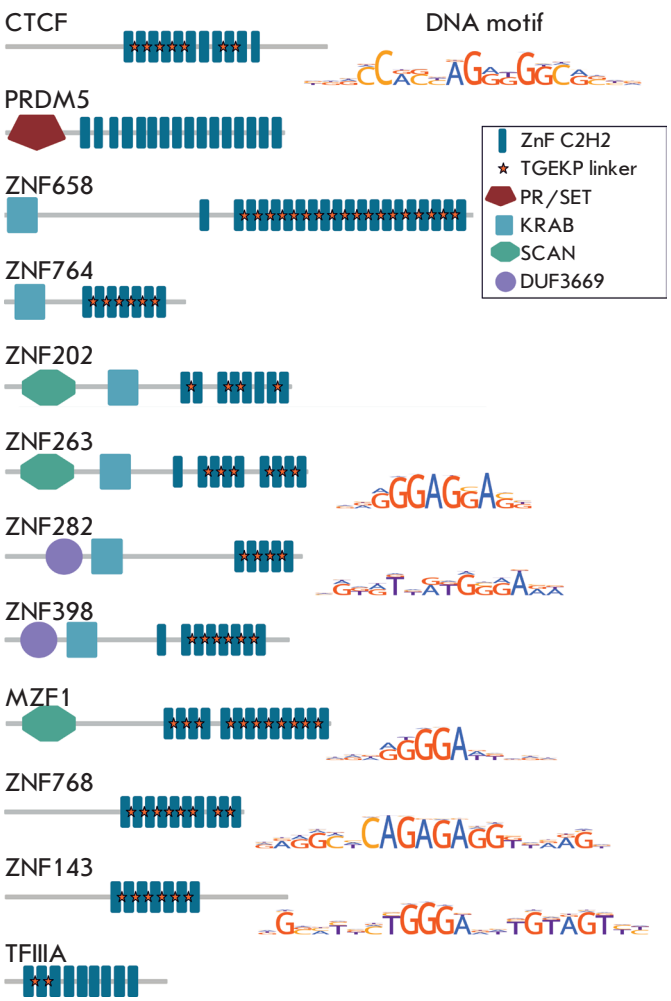

Fig. 1. C2H2 proteins of vertebrates with architectural functions. The domain organization of the described proteins and the known binding motifs are shown

In human CTCF, the C2H2 domains **3–7** are responsible for specific binding to the 15-bp consensus motif (*Fig. 1*) [38]. The C2H2 domain **8** lies outside the major groove and is not involved in the recognition of DNA nitrogenous bases; therefore, it can act as a bridge connecting the C2H2-domains **3–7** recognizing the key motif with the C2H2 domains **9–11**, which can specifically bind to an additional DNA motif that is found in approximately 15% of CTCF binding sites [39, 40]. The C2H2 domains **1–2** can also bind to a non-conserved DNA sequence [39]. Thus, different combinations of C2H2 domains of CTCF can bind to a broad range of motifs with different levels of efficiency [41, 42].

It has been shown *in vitro* that CTCF–DNA binding is inhibited by cytosine methylation at position 2 in the consensus site, whereas cytosine methylation at position 12 has almost no effect. The cytosine at position 2 is recognized by the aspartic acid residue, which prefers the unmethylated base. At position 12, the cytosine is recognized by a glutamic acid residue, with the binding affinity slightly increasing in the case of a methylated base. [38]. Moreover, an important role in methyl group recognition is played by the arginine residue that forms the 5-methylcytosine–arginine–guanine triad in a complex with DNA; this triad is found in all C2H2 protein complexes with methylated DNA [43, 44].

Cytosine methylation at binding sites can enhance, weaken, or completely inhibit the binding of C2H2 proteins to DNA; i.e., it is a global mechanism of regulation of the promoter, enhancer, and insulator activities [45]. The most striking example of the role played by the methylation of binding sites for C2H2 proteins is the participation of CTCF in genomic imprinting, an epigenetic mechanism for regulating the expression of alleles of the same gene depending on their parental origin (male or female) [46]. Imprinting occurs with the participation of special regulatory elements known as differentially methylated regions (DMRs), which often contain CTCF binding sites. Imprinting has been described most thoroughly for the *Igf2* and *H19* genes, which are activated by a group of adjacent enhancers. The DMR acting as an insulator resides between the *Igf2* gene and the enhancers; it consists of four CTCF binding sites carrying a cytosine residue at position 2. DMR methylation is maintained only in the paternal *Igf2/H19* locus, resulting in a loss of CTCF binding and activation of the *Igf2* gene. Meanwhile, in the maternal locus, CTCF binds to the corresponding sites in the DMR, thus inhibiting the interaction between the enhancers and the *Igf2* gene. Methyla-

tion of the binding sites for transcription factors (and C2H2 proteins in particular) can also be involved in the global inactivation of transcription in one of the two X chromosomes in mammals [47].

C2H2-domain clusters can participate in specific and non-specific interactions with RNA [48, 49]. Specific interaction between the TFIIIA protein and 5S RNA has been the one studied most thoroughly. It was shown that the C2H2 domains **1**–**3**, **5**, and **7**–**9** bind to DNA motifs in the promoter region of the 5S RNA gene, while the C2H2 domains **4**, **5**, and **6** interact with 5S RNA. Therefore, the C2H2 domains **4** and **6** act as linkers broadening the TFIIIA protein–DNA binding capacity. At the same time, specific interaction of these C2H2 domains with newly synthesized 5S RNA is necessary for its stabilization during export from the nucleus to the cytoplasm, prior to ribosome assembly.

Two C2H2 domains, **1** and **10**, are responsible for the nonspecific interaction between CTCF and a broad range of RNAs [50, 51]. Interestingly, the disruption of the C2H2-domain structure caused by mutation in histidine does not affect RNA binding. This finding suggests that individual amino acids in the C2H2 domains play an important role in RNA binding, rather than the structure of the zinc finger itself. There are experimental data showing that interaction between CTCF and RNA may cause protein multimerization, but the mechanism of this process remains unknown [50, 52]. Since a large number of CTCF sites reside inside the introns of genes, it can be expected that CTCF is involved in the regulation of pre-mRNA splicing and termination (the processes running concomitantly with transcription) by non-specifically binding to RNA. For example, CTCF can slow down the movement of RNA polymerase II, leading to selection of either an alternative exon during splicing [53, 54] or an alternative polyadenylation signal during transcription termination [55]. A domain capable of interacting with RNA polymerase II has been mapped to the C-terminal domain of the CTCF protein, which can also be involved in the slowing-down of the movement of RNA polymerase II when passing through the CTCF binding sites [56].

A large body of experimental data shows that individual C2H2 domains or their clusters are involved in protein–protein interactions [34]. However, the detailed mechanisms behind these processes and their specificity remain poorly studied. C2H2 domains often interact with the complexes that are involved in chromatin remodeling and histone modification. According to data obtained through mutational analysis, any amino acids within the C2H2 domains and linkers connecting them can participate in these interactions (unlike during DNA binding). Therefore, it can be assumed that, in some cases, even the C2H2 domains associated with DNA can participate in the recruitment of regulatory complexes to chromatin.

A cluster of C2H2 domains is the only conserved part of the CTCF protein that shares high homology in most vertebrates, insects, and some nematodes [57–59]. The CTCF protein is not found in plants, yeast, or roundworms. The distribution of CTCF binding sites in the genome is also characterized by a certain degree of conservatism. In particular, CTCF binding sites are found at the boundaries of the regulatory domains of the homeotic genes in mammals, fish, and Drosophila [60, 61], where CTCF performs an insulator function and delimits the regions where enhancers residing in the adjacent domains perform their function [62–66]. It is worth mentioning that the CTCF binding sites are located in the repetitive elements of mammalian genomes, which could be the starting point for the evolutionary expansion of the CTCF binding sites in the intergenic regions where TAD boundaries are located [26, 67].

Despite the absence of homologous regions, the N-terminal domains of the CTCF proteins in nine animal species belonging to different classes are represented by unstructured homodimerization domains [68]. Deletion of the dimerized domain within Drosophila CTCF significantly reduces the functional activity of the mutant CTCF [69]. It was discovered in mouse embryonic stem cells that the N-terminal domain is involved in the specific binding of CTCF to the respective sites [70]. A YxF motif was found between the N-terminal homodimerization domain and the C2H2 cluster, which is necessary for interaction with the SA2–Scc1 cohesin subcomplex [71]. A similar motif was also found in the CTCF of other animal species. Therefore, although there is no significant homology, the N-terminal domains of the CTCF proteins in different species share characteristic structural features. A region interacted with the SA2 subunit of the cohesin complex was previously mapped on the C-end of the CTCF protein *in vitro* [72]; however, a more recent study has failed to confirm this finding [71].

The roles played by C2H2 proteins largely depend on the proteins with which they interact. More than 90 potential CTCF partner proteins have been identified [73, 74]. However, the mechanisms and specificity of these interactions remain disputable. Most protein–protein interactions are found within the cluster of C2H2 domains and in the unstructured C-terminus of the CTCF protein. Many different C2H2 proteins can potentially interact with the same protein complexes through the C2H2 domains. CTCF was shown to interact directly with the catalytic subunit Brg1

of the SWI/SNF chromatin remodeling complex [74] and the general transcription factor II-I (TFII-I) [75]. Therefore, the most probable function of CTCF in the promoter regions of actively transcribed genes is participation in the formation of open chromatin regions through the recruitment of the SWI / SNF complex, which increases the mobility of nucleosomes. CTCF can also be involved in stabilization on promoters of the TFIID complex (TFII-I being a part of this complex). When CTCF is inactivated, the expression level drops significantly only in the genes whose promoter regions contain CTCF binding sites [76]. Thus, one of the key functions of CTCF consists in organizing active promoters. Interestingly, like many other C2H2 proteins, CTCF contains regions enriched in proline and acidic amino acids, which is typical of the transcription activators recruiting transcription complexes to chromatin.

A domain interacting with DEAD box RNA helicases was identified in the C-terminus of CTCF [74, 77], which may be related to the potential significant participation of CTCF in the regulation of splicing and transcription termination. To perform these functions, the found interaction of CTCF with topoisomerase II (Top2) is probably also needed [78]. Top2 regulates chromatin topology by introducing ATP-dependent double-strand breaks into DNA. The Top2 protein has been found in approximately half of all CTCF binding sites [78]. Top2 activity is most often observed in the close vicinity of CTCF binding sites [79]. It is thought that Top2 is recruited to the open chromatin regions that form at CTCF sites, and that direct protein-protein interactions enhance this process. Possibly, CTCF helps recruit Top2 to the introns and 3'-ends of genes, which might be required during gene transcription.

The activity of C2H2 proteins is regulated by various post-translational modifications. Phosphorylation of C2H2 proteins at the linkers between the C2H2 domains, which occurs during mitosis and reduces the efficiency of protein binding to chromatin, has been studied quite thoroughly [80–83]. C2H2 proteins can also undergo further modifications, such as ubiquitination, SUMOylation, and poly-ADP-ribosylation [84]. The ribosylation site resides at the N-terminus of CTCF [85]; this modification can affect protein dimerization and its binding to the cohesin complex. Poly-ADP-ribosylation affects the localization of the CTCF protein in nuclear compartments, chromatin binding, and transcription regulation [85–87]. Interestingly, the N-terminus of human CTCF interacts with the C-terminus of nucleophosmin 1 (NPM1), which can be responsible for CTCF localization within the cell [88]. Sites for covalent attachment of the SUMO protein through lysine were found in the C-terminal domains of the CTCF protein [89]. The Pc2 protein belonging to the Polycomb group of transcriptional repressors was identified as a SUMO E3 ligase for CTCF. Within cell nuclei, CTCF and Pc2 are found in bodies enriched in Polycomb-group proteins.

It is assumed that by interacting with various proteins and forming homopolymers, SUMO catalyzes the formation of dense intranuclear protein structures (bodies) that can perform many functions, including being a source of spare proteins during chromatin formation on newly synthesized DNA during replication [90, 91]. SUMOylation of CTCF on chromatin can also regulate the recruitment of transcriptional complexes to chromatin, thus changing the properties of CTCF during the activation or repression of gene transcription.

As a member of the C2H2 protein family, CTCF has typical structural features: it contains a cluster of zinc fingers that provides specific binding to genomic targets and interacts with RNA and proteins, as well as terminal domains that are involved in the organization of the architecture of chromosomes and the recruitment of various regulatory complexes to chromatin.

CTCF IN ORGANIZATION OF THE CHROMOSOME ARCHITECTURE AND INSULATION IN VERTEBRATES

The CTCF protein was initially considered as the main vertebrate insulator protein [92]. The first vertebrate insulator was reported to be located at the boundary of the heterochromatin region and the chicken β-globin gene cluster [93, 94]. The insulator, with its core being 275-bp long, was mapped in the DNase 1 hypersensitive site and was therefore named HS4 [95]. In transgenic cell model systems, two copies of the HS4 insulator can effectively block enhancer activity and protect transgene expression from repression by surrounding chromatin. In addition to the binding site for the CTCF protein, the HS4 insulator was found to contain binding sites for USF1/USF2 proteins [96] and three binding sites for the VEZF1 (vascular endothelial zinc finger 1) protein [97]. It has been demonstrated that CTCF is required to block enhancers and recruit USF1/USF2 proteins, which in turn, recruit the complexes responsible for chromatin remodeling and histone modification. As a result, the nucleosomes around the HS4 insulator are enriched by nucleosome modifications associated with active chromatin (histone H3 methylated at lysine 4 and acetylated histones H3 and H4).

The VEZF1 protein contains a cluster consisting of six C2H2 domains, and it predominantly binds to active promoters [98]. Inactivation of the VEZF1 binding sites on the HS4 insulator in transgenic cell

lines enhances DNA methylation at the promoter of a reporter gene [97]. It is assumed that VEZF1 recruits a complex that performs DNA demethylation, thereby facilitating the recruitment of transcription factors (which cannot efficiently bind to methylated sites) to the HS4 insulator and the adjacent regulatory elements. Thus, the HS4 insulator is a combination of the binding sites of at least two C2H2 proteins that function in close cooperation with each other.

Despite numerous examples illustrating the key role of CTCF sites in the organization of the boundaries of regulatory domains and the insulation of enhancers [23], the question remains as to what role is played by other unknown proteins whose binding to a particular regulatory element depends on the presence of CTCF. For example, in mammals, a large number of CTCF-dependent insulators block the spread of Polycomb-dependent heterochromatin, which is associated with H3K27me3 enrichment of long chromatin regions. However, inactivation of CTCF does not cause the propagation of the H3K27me3 modification into transcriptionally active regions, which suggests that other proteins are present at the boundaries that block the spreading of repressive chromatin and thereby mask the absence of CTCF [76]. Therefore, CTCF-dependent insulators, the boundaries of the regulatory domain, and TADs are most likely to consist of CTCF sites, in combination with the binding sites for other transcription factors (including C2H2 proteins that have not been described yet).

In our current understanding, which is supported by plenty of experimental data, mammalian CTCF forms chromatin loops, in cooperation with the cohesin complex, and defines the boundaries of most TADs [19, 99]. The cohesin complex is involved in mitosis, meiosis, and the regulation of gene expression [100, 101]. This complex consists of the SMC1, SMC3, and SCC1 (Rad21) proteins, forming a ring structure and binding to the fourth subunit that exists as two isoforms, STAG1 (SA1) and STAG2 (SA2), through SCC1. It has been hypothesized that SA1 and SA2 can determine the location of the cohesin complex at different chromatin sites. The NIPBL/MAU2 and WAPL complexes catalyze the ATP-dependent binding of the cohesin complex to chromatin and its subsequent dissociation, respectively [100].

Depending on the antibodies used and the cell line under study, the colocalization of CTCF and cohesin sites varies from 40% to 95% [102–104]. Inactivation of CTCF leads to a redistribution of the cohesin complexes from the CTCF binding sites to the promoters of active genes, accompanied by partial destruction of TADs [76]. Inactivation of the subunits of the cohesin complex or the Nipbl protein [105, 106], which ensures the recruitment of the cohesin complex to chromatin, leads to an almost complete disappearance of TADs. On the contrary, inactivation of the factors that negatively affect cohesin binding to chromatin stabilizes TADs and the long-range interactions in chromatin [106]. Finally, mutations and deletions in CTCF, disrupting its interaction with the cohesin complex, also significantly disturb the formation of long-range contacts and TADs [71, 104]. The Smc1 and Smc3 subunits contain ATPase domains, and the energy of ATP cleavage is required at the stages of binding and dissociation of the cohesin complex [107, 108]. Mutations in the subunits of the cohesin complex, which disrupt ATP hydrolysis, affect long-range contacts and the formation of TAD in chromosomes [109].

CTCF sites at the TAD boundaries usually have a convergent orientation [8, 110]. The convergent orientation of CTCF motifs was shown to define which pairs of CTCF sites preferentially stabilize DNA loops [8, 110–112]. A loop extrusion model has been proposed to explain why chromatin loops preferentially form between CTCF sites with a convergent orientation. According to this model, after being loaded onto chromatin, the cohesin complex triggers DNA extrusion and chromatin loop formation. CTCF can inhibit the movement of the cohesin complex only if its N-terminal domain, which interacts with the SA2–SCC1 subcomplex [71], is correctly oriented relative to the cohesin sliding complex.

The model postulates that the cohesin complex can induce chromatin extrusion and chromatin loop formation, either actively (using ATP energy) or passively. Indeed, *in vitro* studies have shown that in the presence of Nipbl and ATP molecules, the cohesin complex binds to DNA and slides along, causing loop formation [113], even if DNA is nucleosome-bound [114]. Cohesin can also bypass small nucleosome-sized protein complexes, but it is unable to overcome complexes larger than 13 nm in diameter; such complexes with a motor function can move cohesin themselves [115]. Therefore, the convergent CTCF sites limit the extrusion regions of chromatin loops, while the loop formation is performed by molecular motors.

According to the polymer model, TAD formation strongly depends on the physical properties of chromatin, which tends to form domains of the same type. This model has support in Drosophila, where TADs form through electrostatic inter-nucleosomal interactions. As a result, the TADs boundaries are predominantly composed of long, highly transcribed open chromatin regions and the inner regions of TADs are denser chromatin structures [13, 19, 116, 117]. In this model, the role of CTCF is to recruit cohesin complexes, stabilizing the interactions between chromatin

sites that are already in close vicinity to each other. However, this model does not explain why chromatin loops in mammals predominantly form only convergently oriented CTCF binding sites.

The experimental data [107, 113, 118] show that the size of a chromatin loop is independent of the time of cohesin–DNA binding but depends on the barriers limiting its sliding (similar to CTCF). CTCF binds dynamically to chromatin, which is consistent with the heterogeneity of the TAD boundaries observed in studies of single cells [20]. The CTCF binding sites at the TAD boundaries are usually represented by clusters, which probably ensure CTCF binding to the genomic targets for a longer time [119].

According to the loop extrusion model, the cohesin complexes are only transiently blocked at a certain CTCF site and can continue to pull chromatin after crossing the block created by CTCF or after CTCF leaves chromatin [20]. Inactivation of Wapl stabilizes the binding of the cohesin complexes to chromatin; the size of chromatin loops increases, which is explained by longer residence time of the cohesin complex on chromatin [106, 120, 121].

Mitosis is characterized by chromosome condensation associated with large-scale chromatin changes and the loss of binding of some transcription factors to DNA. During mitotic prophase, most of the cohesin leaves the chromosomes (except for cohesin associated with centromeres). In anaphase, cohesin dissociation caused by separase promotes the segregation of sister chromatids [101]. The structure of TADs on compact mitotic chromosomes is lost almost completely but is quickly restored by the mid-G1 stage [122]. The data on the binding of CTCFs to the respective sites on mitotic chromosomes are inconsistent. According to some estimates, CTCF remains on 18.6% of the sites [122]; however, the binding of CTCF to its sites is mostly lost, since there is phosphorylation of linkers between the C2H2 domains [123]. It is possible that by leaving its binding sites, CTCF contributes to a more efficient removal of the cohesin complexes from mitotic chromosomes. However, CTCF binding sites are rapidly restored after mitosis, which may result from the association between free CTCF and condensed chromosomes during mitosis [123]. It remains an open question how the effective restoration of CTCF binding to the corresponding sites after mitosis occurs. It is most likely that other transcription factors remain associated with mitotic chromosomes and maintain the partially open chromatin state (act as bookmarks), which facilitates CTCF binding to the respective sites after mitosis. As a result, both the CTCF binding profile and the structure of TADs on duplicated chromosomes are rapidly restored after DNA replication. It can be assumed that excess CTCF accumulates in specialized nuclear compartments (bodies) stabilized by SUMO [89]. During replication, the excess CTCF binds to an increasing number of sites on duplicating chromosomes.

ARCHITECTURAL FUNCTIONS OF OTHER VERTEBRATE C2H2 PROTEINS

Studies focused on designing artificial C2H2-type zinc fingers that ensure specific interaction with a particular genomic target have shown that the specificity of DNA binding increases dramatically for the cluster consisting of five properly organized zinc fingers. Therefore, in this section we would like to discuss proteins with this structural organization (with at least five C2H2-type domains separated by a typical 6-bp linker sequence) as the most promising architectural C2H2 proteins.

Other C2H2 proteins remain relatively less well studied than CTCF [124, 125]. The key problems in studying this class of proteins are associated with a significant overlap of functions between different C2H2 proteins and the lack of high-quality specific antibodies against these proteins, which would make it possible to perform genome-wide studies to identify the binding sites for C2H2 proteins and their role in the maintenance of long-range contacts between regulatory elements and the formation of the chromosome architecture. Two studies [126, 127] have focused on binding sites for the 60 and 221 C2H2 proteins with GFP or HA epitope tags in HEK293T cells. The binding sites for the same C2H2 proteins studied in both publications were found to overlap only slightly [128]. It should be noted that in these studies, expression of tagged C2H2 proteins occurred in the presence of the endogenous C2H2 proteins; so, most actual binding sites were occupied by the native protein, while the tagged protein was bound (mostly non-specifically) to the domains within open chromatin regions. In the near future, the use of the CRISPR/Cas9 genome editing tool could make it possible to replace endogenous genes with modified ones expressing tagged variants of C2H2 proteins, which will simplify genome-wide studies.

Mammalian genomes are enriched in various types of repetitive sequences of differing nature, including mobile elements and retroviruses [129]. Most of the studied C2H2 proteins, including CTCF, have binding sites in different repetitive sequences [130–133]. There are many examples of repeating sequences becoming part of genetic regulatory networks and TAD boundaries [134], thereby significantly expanding the possibilities of fine-tuning gene expression during evolution.

Approximately half of all C2H2 proteins carry another domain at their N-terminus. The two most evolutionarily ancient domains that are found in all eukaryotes include the PR/SET domain (e.g., Prdm5 protein (*Fig. 1*)), which typically exhibits methyltransferase activity [135], and the BTB domain that forms dimers and recruits transcription regulators to the genomic targets [136]. One of the most numerous groups of C2H2 proteins in mammals carries the KRAB domain at their N-terminus (e.g., ZNF658 and ZNF764 proteins (*Fig. 1*)). It is believed that this domain has become widespread in mammalian C2H2 proteins, due to its repressor function with respect to mobile elements. However, in parallel with the evolution of gene regulatory systems into which mobile elements are integrated, KRAB-C2H2 proteins acquire new functions in the regulation of host gene expression [130, 131]. Some of these C2H2 proteins containing the KRAB domain carry an additional domain, SCAN (e.g., ZNF202 and ZNF263 proteins) or DUF3669 (e.g., ZNF282 and ZNF398 proteins), at their N-terminus [137–139]. Some C2H2 proteins carry only the SCAN domain (e.g., MZF1) and derive from proteins that have lost their KRAB domain. It can be assumed that some functions of the SCAN-C2H2 and DUF3669-C2H2 proteins are associated with the ability of SCAN and DUF3669 to form homo- and heterodimers [131, 137].

The most thoroughly described functions of C2H2 proteins are the formation of an open chromatin region on promoters and recruitment of transcription complexes for transcriptional activation or repression. The ZNF658 protein binds to the regulatory element residing next to 3525 promoters and participates in the activation of the expression of the rRNA genes transcribed by RNA polymerase I [140, 141]. The ZNF764 protein is ubiquitously expressed; it is involved in the regulation of glucocorticoid, androgen, and thyroid hormones activity [142]. It is interesting that the binding sites, which predominantly reside in intergenic regions (60%) and introns (31%), significantly (37%) co-localize with the binding sites for glucocorticoid receptors (GRs) [143]. It has been experimentally proved that the KRAB domain of the ZNF764 protein directly interacts with the LBD domain of a GR, suggesting that these proteins bind cooperatively to the regulatory regions. The protein-binding sites of ZNF202 [126, 144], ZNF263 [145], MZF1 [146], ZNF768 [133], and Prdm5 [147] are predominantly located in the promoter regions of genes, indicating that these factors may contribute to the activation or repression of transcription, and promoter architecture.

The N-terminus of ZNF768 (*Fig. 1*) contains 15 heptad repeats that are similar to the C-terminal domain of RNA polymerase II [133] and are presumably involved in the recruitment of the transcription elongation complex to promoters.

It has been demonstrated that MZF1 can form heterodimers with other SCAN-containing proteins (ZNF24, ZNF174, and ZNF202) through the SCAN domain [148, 149]. The ZNF282 and ZNF398 proteins form homo- and heterodimers through the DUF3669 domain [150] and can bind to the promoters in a combinatorial manner [126]. The Prdm5 protein contains an N-terminal PR/SET domain that has lost its methylation ability and is possibly involved in protein–protein interaction [151, 152].

The protein ZNF143 (*Fig. 1*), which is crucial to the embryonic development of mammals, has been characterized thoroughly [153]. Its central part contains a cluster consisting of seven C2H2 domains. The N-terminal domain contains three 15-aa repeats separated by 10-aa to 12-aa spacers [154]. The C-terminal domain is enriched in acidic amino acids, which is typical for transcriptional activators. The ZNF143 binding sites reside within a region of approximately 2,000 promoters regulated by RNA polymerases II and III [155–158]. The functional activity of ZNF143 near the promoters is related to the formation of open chromatin regions and its involvement in the recruitment of transcription activation complexes [159–161]. The ZNF143 protein has two partially overlapping consensus binding sites with the same core CCCAGA sequence [155], which can be explained by the different contributions of individual C2H2 domains to the recognition of two site variants. Genome-wide studies have shown that the ZNF143 protein may be involved in the formation of chromatin loops between enhancers and promoters [155, 156, 162-164].

A relatively large percentage of the binding sites of the Prdm5 and ZNF143 proteins colocalize with CTCF [143, 152, 163]. The Prdm5 protein has been found in association with cohesin and CTCF [152]. It was shown for HEK293T cells that inactivation of ZNF143 disrupts some CTCF-dependent chromatin loops [163]. However, there are no experimental data proving that ZNF143 (unlike CTCF) can be involved in the localization of the cohesin complex on chromatin.

Another example of the structural function of C2H2 proteins was observed when studying the chromatin architecture organized by the TFIIIC complex. It was found that the binding sites of the TFIIIC complex colocalize with condensins and can act as boundaries between active chromatin and heterochromatin, as well as maintain distant interactions and participate in the formation of the chromosome architecture [165]. Interestingly, the binding sites for Prdm5, CTCF/cohesin, and ZNF143 proteins are adjacent to or colo-

calize with the TFIIIC binding regions [152, 155, 166], suggesting that these proteins are cooperatively involved in the organization of TFIIIC-dependent regulatory elements. Furthermore, the Prdm5 protein has been isolated in complex with TFIIIC, suggesting that Prdm5 participates in the recruitment of the TFIIIC complex to chromatin [152].

The above-mentioned TFIIIA is the second (after CTCF) best-studied C2H2 protein, which binds to Pol III-dependent promoters of the genes encoding 5S rRNA in all eukaryotes [167]. TFIIIA consists of nine C2H2 domains and a C-terminal activation domain called TAS (Transcription Activating Signal). The protein binds to a regulatory element called ICR, which is located in the transcribed parts of the genes. A structural analysis revealed that the C2H2 domains **1–3** and **7–9** bind to two regions (the C and A boxes) of the ICR element; the central C2H2 domains bind specifically to 5S RNA [167]. The lack of homology in the amino acid sequence of TFIIIA proteins from different species suggests a parallel evolution of the DNA sequences of the promoters, 5S rRNAs, and C2H2 domains, which are involved in the specific binding of DNA and RNA. The TFIIIA protein determines open chromatin on the promoter, while the TAS domain is involved in the recruitment and stable binding of the TFIIIB complex to the promoter [168].

The existing data show that many C2H2 proteins are involved in the formation of active promoters, as well as the recruitment of transcription factors and complexes to regulatory elements. It is obvious that many C2H2 proteins duplicate each other's functions, which makes it difficult to prove their role in the global organization of the chromosome architecture and regulation of transcription.

C2H2 PROTEINS IN DROSOPHILA: DIFFERENT STRUCTURES BUT SIMILAR PROPERTIES

Approximately 170 proteins with clusters consisting of at least five C2H2 domains have been found in the Drosophila genome. However, data on the distribution of C2H2 binding sites in the genome and their functional role in the regulation of gene transcription and the chromosomal architecture have been obtained for only a few of these proteins (*Fig. 2*). The best studied C2H2 proteins include the first protein with insulator properties described in higher eukaryotes, Su(Hw), and an homolog of mammalian CTCF, dCTCF [22, 24, 169, 170]. Both insulator proteins have a similar structure: they contain unstructured terminal domains and a central cluster consisting of 11 (dCTCF) or 12 (Su(Hw)) C2H2 domains. The N-terminus of the dCTCF protein contains an unstructured globular domain capable of forming tetrameric com-

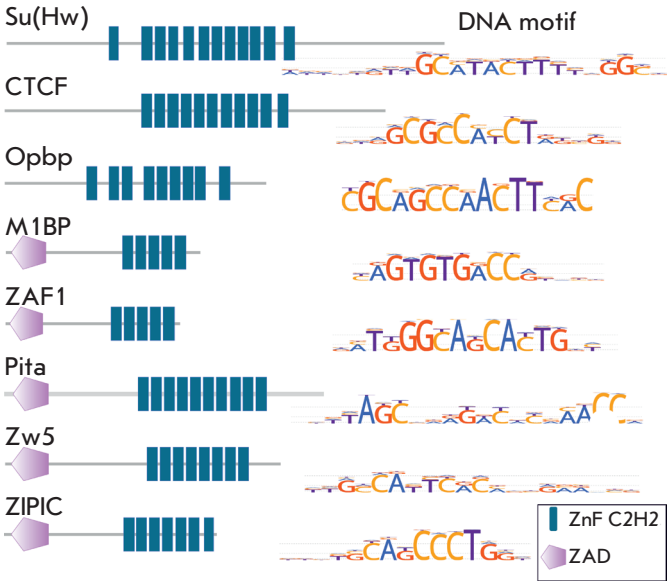

Fig.2. Drosophila C2H2 proteins with architectural functions. The domain organization of the known architectural proteins of Drosophila and their binding motifs are shown

plexes [68, 69], and a potential site of interaction with the cohesin complex, which has homology with the human YxF motif of CTCF that interacts with the SA2-SCC1 complex [71]. An interesting structural feature of another studied C2H2 protein, Opbp [171], is the presence at an N-terminus of an atypical zinc finger capable of homodimerization (*Fig. 2*). Opbp has also a cluster consisting of five C2H2 domains responsible for specific binding to DNA and an additional four C2H2 domains that can interact with RNA and proteins.

The remaining five C2H2 proteins (M1BP, ZAF1, Pita, Zw5, and ZIPIC) belong to a large group of ZAD-containing proteins. The ZAD (zinc-finger-associated domain) was found in 98 Drosophila proteins; approximately 70 of these proteins contain five or more C2H2 domains [172, 173]. The genes encoding ZAD-C2H2 proteins are typically arranged in clusters and, like mammalian KRAB-C2H2 proteins [174, 175], actively evolve as a result of multiple duplications of the original gene copies. The ZAD structure is formed by two pairs of cysteine residues coordinated to the zinc ion [176]. The N-terminal portion of the domain is a globular structure; the C-terminal stem structure is formed by a long α-helix. ZAD domains are capable of homodimerization with the formation of an antiparallel dimmer [176, 177]. Mutations in the genes

encoding the Pita and Zw5 proteins are lethal, which suggests an important role for some representatives of C2H2 proteins in the development of Drosophila [178, 179]. Inactivation of the Su(Hw) protein impairs gonad development, resulting in female sterility [180]. Like in mammals, the Drosophila CTCF protein is involved in the regulation of *hox* gene expression [181, 182]. Although the Drosophila genome was found to contain only ~ 40 Opbp binding sites, inactivation of this protein causes pupal mortality [171].

All the investigated C2H2 proteins bind to long (12–15 bp) DNA motifs via four or five C2H2 domains organized as a cluster (*Fig. 2*) [171, 183–186]. Except for Su(Hw), the binding sites of C2H2 proteins predominantly reside in the promoter regions of active genes and introns [177, 184–190]. The most illustrative example of a protein of this class is M1BP, which binds to the promoters of more than 2,000 genes [185] and, according to experimental data [191], participates in the formation of active promoters. The Opbp protein also binds exclusively to gene promoters, as it is colocalized with M1BP in about half of them [171]. Unlike M1BP and Opbp, which are involved in transcription activation, Su(Hw) binds to the promoters of a large group of neuronal genes and represses their transcription in female gonads during the early stages of Drosophila development [192, 193].

The role played by C2H2 proteins in the formation of long-range contacts and inhibiting enhancer activity was analyzed in transgenic Drosophila lines. *In vivo*, C2H2 proteins efficiently interact with artificial DNA fragments, each containing four to five binding sites [177, 184, 188, 194]. In transgenic lines, the activity of an enhancer surrounded by binding sites for the same C2H2 protein is substantially blocked. However, the enhancer activity is restored by the removal of either of the two binding sites of the C2H2 protein, which proves that the interaction between the C2H2 proteins plays a crucial role in the formation of the chromatin loop, resulting in steric isolation of the enhancer. In the transgenic model system, the C2H2 protein binding sites can bring the yeast GAL4 activator and the reporter gene promoter closer together, thus activating transcription [177, 184, 195]. At the same time, combinations of binding sites for different C2H2 proteins cannot bring the GAL4 activator closer to the promoter [177, 195], which can explain the importance of preferential homodimerization of C2H2 proteins in providing specific distant interactions between genomic elements. For example, it has been shown that the ability of the ZAF1 and ZIPIC proteins to maintain distant interactions is determined by their ZAD domains [177, 184]. Therefore, domains capable of forming homodimers seem to play an important role in the organization of specific long-range contacts between the regulatory elements in chromatin.

The role of C2H2 proteins in the organization of the boundaries of regulatory domains can be most clearly demonstrated by the example of the bithorax complex (BX-C), which includes three homeotic genes, Ubx, abd-A, and Abd-B [196, 197]. The regulatory region of the BX-C is divided into nine independent domains, each activating the transcription of one of the three homeotic genes during the development. Several domain boundaries are characterized in detail and mapped as minimal fragments that can function as effective insulators in transgenic model systems [198–202]. Each characterized boundary contains different combinations of the binding sites of the Pita, dCTCF, and Su(Hw) proteins required to ensure its functional activity [65, 66]. The boundaries can be replaced with repeats consisting of four to five binding sites for each C2H2 protein. Thus, despite the differences in their structural organization, the Su(Hw), Pita, and dCTCF proteins have similar functions and work in cooperation in such processes as the organization of regulatory domain boundaries [66, 203, 204].

Unlike in mammals, the boundaries of most TADs in Drosophila coincide with clusters of housekeeping genes [205, 206]. Thus, the M1BP protein (whose binding sites reside in many promoters of housekeeping genes) is most often found at the TAD boundaries, while the binding sites for other characterized C2H2 proteins usually reside inside the TADs. In embryos and embryonic cell lines, the dCTCF protein, despite its cohesin-binding motif, is rarely found at the boundaries of TADs, while from 40% to 60% of dCTCF sites colocalize with cohesin complexes on chromatin [205–207]. Binding of the bulk of cohesin is observed in the open chromatin zones of actively transcribed promoters [208]; therefore, it cannot be ruled out that C2H2 proteins play a direct or an indirect role (organization of open chromatin regions) in the recruitment of cohesin complexes. Interestingly, most TAD boundaries in a BG3 cell culture derived from Drosophila neural tissues coincide with dCTCF binding sites [207]. Therefore, the TAD boundaries in Drosophila can be changed during cell differentiation.

It is most likely that the TAD boundaries are fixed due to interaction between the protein complexes flanking TADs. In addition, for Drosophila, the existence of a mechanism of TAD formation has been demonstrated, and it is due to the electrostatic internucleosomal interactions that make transcriptionally active sites act as TAD boundaries [116].

CONCLUSIONS

At present, the C2H2 proteins of higher eukaryotes remain the least studied class of transcription factors. The well-studied mammalian CTCF protein provides general insight into the properties, partners, and functions of this class of transcription factors. CTCF is probably the ancestor of the entire class of C2H2 proteins, which in the course of evolution could acquire new domains and bind to new DNA sequences. In this context, it is interesting that CTCF in both Drosophila and mammals is involved in the organization of the boundaries of the transcriptional domains of homeotic genes. Drawing on existing information, it can be concluded that C2H2 proteins in mammals and Drosophila are often involved in the organization of active promoters. By interacting with nucleosome remodeling complexes, C2H2 proteins can form open chromatin and become simultaneously involved in the recruitment of major transcription factors to promoters. Many well-studied regulatory elements (promoters and insulators in particular) carry combinations of binding sites for C2H2 proteins which function cooperatively in their interaction with chromatin. Some C2H2 proteins, including CTCF, have been found to contain N-terminal homodimerization domains that may be involved in the organization of specific long-range contacts. The motif interacting with the cohesin complex has been identified only in the CTCF protein. However, C2H2 proteins can probably interact with other surfaces in cohesin and condensin complexes, which is consistent with the localization of these complexes on active promoters.

It is believed that different mechanisms are responsible for TAD boundary formation and long-range contacts in mammals and Drosophila. However, there still remains an open question as to whether the mammalian cohesin complex can cause intensive chromatin loop extrusion during the formation of TADs and long-range interactions between the regulatory elements. It is also unclear why other higher eukaryotes do not have a similar mechanism, although the cohesin complex is highly conserved. Interestingly, the genome of danio fish contains neither CTCF nor the cohesin complex at most TAD boundaries [209], despite the fact that CTCF in danio and humans shows 86% homology. On the other hand, CTCF is found at the TAD boundaries in Drosophila neural cells [207]. It can be assumed that the mechanisms of formation of TADs are actually much more universal than it appears at present. C2H2 proteins such as Prdm5 and ZNF143 can stabilize CTCF binding to mammalian TAD boundaries and be involved in long-range interactions. Drosophila C2H2 proteins, by binding in various combinations to insulators (for example, as part of BX-C), allow two identical copies of the insulator to maintain super-long-range interactions, which is similar to the formation of the boundaries of a new TAD. In mammals, the TAD boundaries usually contain the most evolutionarily conserved clusters of CTCF sites [119]. It can be assumed that at the early stages of vertebrate evolution, multiplied copies of one or several types of mobile elements containing CTCF binding sites, in combination with the sites of other C2H2 proteins, formed long-range interactions, and some of them have given rise to TADs. Therefore, despite the considerable progress achieved in studying the spatial organization of the genome and, in particular, the architectural role of CTCF, many questions remain unanswered due to the lack of data on the other participants necessary for the formation of the nucleus architecture. ●

This work was supported by the Russian Science Foundation (project No. 19-74-30026).

REFERENCES

1. Spitz F., Furlong E.E. // Nature Reviews Genetics. 2012. V. 13. № 9. P. 613–626.
2. Levine M., Cattoglio C., Tjian R. // Cell. 2014. V. 157. № 1. P. 13–25.
3. Zabidi M.A., Stark A. // Trends Genet. 2016. V. 32. № 12. P. 801–814.
4. Furlong E.E.M., Levine M. // Science. 2018. V. 361. № 6409. P. 1341–1345.
5. Geyer P.K., Clark I. // Cell. Mol. Life Sci. 2002. V. 59. № 12. P. 2112–2127.
6. West A.G., Gaszner M., Felsenfeld G. // Genes Dev. 2002. V. 16. № 3. P. 271–288.
7. Gerasimova T.I., Corces V.G. // Annu. Rev. Genet. 2001. V. 35. P. 193–208.
8. Rao S.S., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., et al. // Cell. 2014. V. 159. № 7. P. 1665–1680.
9. Boettiger A., Murphy S. // Trends Genet. 2020. V. 36. № 4. P. 273–287.
10. Boettiger A.N., Bintu B., Moffitt J.R., Wang S., Beliveau B.J., Fudenberg G., Imakaev M., Mirny L.A., Wu C.T., Zhuang X. // Nature. 2016. V. 529. № 7586. P. 418–422.
11. Dekker J., Misteli T. // Cold Spring Harbor Perspectives Biol. 2015. V. 7. № 10. P. a019356.
12. Sikorska N., Sexton T. // J. Mol. Biol. 2020. V. 432. № 3. P. 653–664.
13. Hansen A.S., Cattoglio C., Darzacq X., Tjian R. // Nucleus. 2018. V. 9. № 1. P. 20–32.
14. Luppino J.M., Park D.S., Nguyen S.C., Lan Y., Xu Z., Yunker R., Joyce E.F. // Nat. Genet. 2020. V. 52. № 8. P. 840–848.
15. Zheng H., Xie W. // Nat. Rev. Mol. Cell Biol. 2019. V. 20. № 9. P. 535–550.
16. Sexton T., Yaffe E., Kenigsberg E., Bantignies F., Leblanc B., Hoichman M., Parrinello H., Tanay A., Cavalli G. // Cell. 2012. V. 148. № 3. P. 458–472.
17. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. // Nature. 2012. V. 485. № 7398. P. 376–380.
18. Nora E.P., Lajoie B.R., Schulz E.G., Giorgetti L., Okamoto I., Servant N., Piolot T., van Berkum N.L., Meisig J., Sedat J., et al. // Nature. 2012. V. 485. № 7398. P. 381–385.
19. Szabo Q., Bantignies F., Cavalli G. // Science Advances. 2019. V. 5. № 4. P. eaaw1668.
20. Chang L.H., Ghosh S., Noordermeer D. // J. Mol. Biol. 2020. V. 432. № 3. P. 643–652.
21. Arzate-Mejia R.G., Recillas-Targa F., Corces V.G. // Development. 2018. V. 145. № 6. P. dev137729.
22. Ali T., Renkawitz R., Bartkuhn M. // Curr. Opin. Genetics Dev. 2016. V. 37. P. 17–26.
23. Braccioli L., de Wit E. // Essays Biochem. 2019. V. 63. № 1. P. 157–165.
24. Chen D., Lei E.P. // Curr. Opin. Cell Biol. 2019. V. 58. P. 61–68.
25. Merkenschlager M., Nora E.P. // Annu. Rev. Genomics Hum. Genet. 2016. V. 17. P. 17–43.
26. Schmidt D., Schwalie P.C., Wilson M.D., Ballester B., Goncalves A., Kutter C., Brown G.D., Marshall A., Flicek P., Odom D.T. // Cell. 2012. V. 148. № 1–2. P. 335–348.
27. Chen H., Tian Y., Shu W., Bo X., Wang S. // PLoS One. 2012. V. 7. № 7. P. e41374.
28. Klug A. // Quarterly Rev. Biophys. 2010. V. 43. № 1. P. 1–21.
29. Persikov A.V., Wetzel J.L., Rowland E.F., Oakes B.L., Xu D.J., Singh M., Noyes M.B. // Nucleic Acids Res. 2015. V. 43. № 3. P. 1965–1984.
30. Garton M., Najafabadi H.S., Schmitges F.W., Radovani E., Hughes T.R., Kim P.M. // Nucleic Acids Res. 2015. V. 43. № 19. P. 9147–9157.
31. Persikov A.V., Singh M. // Nucleic Acids Res. 2014. V. 42. № 1. P. 97–108.
32. Durai S., Mani M., Kandavelou K., Wu J., Porteus M.H., Chandrasegaran S. // Nucleic Acids Res. 2005. V. 33. № 18. P. 5978–5990.
33. Kim Y.G., Cha J., Chandrasegaran S. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. № 3. P. 1156–1160.
34. Brayer K.J., Segal D.J. // Cell Biochem. Biophys. 2008. V. 50. № 3. P. 111–131.
35. Ryan R.F., Darby M.K. // Nucleic Acids Res. 1998. V. 26. № 3. P. 703–709.
36. Crozatier M., Kongsuwan K., Ferrer P., Merriam J.R., Lengyel J.A., Vincent A. // Genetics. 1992. V. 131. № 4. P. 905–916.
37. Wolfe S.A., Nekludova L., Pabo C.O. // Annu. Rev. Biophys. Biomol. Structure. 2000. V. 29. P. 183–212.
38. Hashimoto H., Wang D., Horton J.R., Zhang X., Corces V.G., Cheng X. // Mol. Cell. 2017. V. 66. № 5. P. 711–720 e713.
39. Nakahashi H., Kwon K.R., Resch W., Vian L., Dose M., Stavreva D., Hakim O., Pruett N., Nelson S., Yamane A., et al. // Cell Reports. 2013. V. 3. № 5. P. 1678–1689.
40. Xiao T., Wongtrakoongate P., Trainor C., Felsenfeld G. // Cell Reports. 2015. V. 12. № 10. P. 1704–1714.
41. Xu D., Ma R., Zhang J., Liu Z., Wu B., Peng J., Zhai Y., Gong Q., Shi Y., Wu J., et al. // J. Phys. Chem. Lett. 2018. V. 9. № 14. P. 4020–4028.
42. Yin M., Wang J., Wang M., Li X., Zhang M., Wu Q., Wang Y. // Cell Research. 2017. V. 27. № 11. P. 1365–1377.
43. Liu Y., Zhang X., Blumenthal R.M., Cheng X. // Trends Biochem. Sci. 2013. V. 38. № 4. P. 177–183.
44. Hudson N.O., Buck-Koehntop B.A. // Molecules. 2018. V. 23. № 10. P. 2555.
45. Ren G., Zhao K. // Cell. Biosci. 2019. V. 9. P. 83.
46. Noordermeer D., Feil R. // Curr. Opin. Genet. Dev. 2020. V. 61. P. 17–24.
47. Heard E., Disteche C.M. // Genes Dev. 2006. V. 20. № 14. P. 1848–1867.
48. Iuchi S. // Cell. Mol. Life Sci. 2001. V. 58. № 4. P. 625–635.
49. Hall T.M. // Curr. Opin. Struct. Biol. 2005. V. 15. № 3. P. 367–373.
50. Saldana-Meyer R., Gonzalez-Buendia E., Guerrero G., Narendra V., Bonasio R., Recillas-Targa F., Reinberg D. // Genes Dev. 2014. V. 28. № 7. P. 723–734.
51. Kung J.T., Kesner B., An J.Y., Ahn J.Y., Cifuentes-Rojas C., Colognori D., Jeon Y., Szanto A., del Rosario B.C., Pinter S.F., et al. // Mol. Cell. 2015. V. 57. № 2. P. 361–375.
52. Hansen A.S., Hsieh T.S., Cattoglio C., Pustova I., Saldana-Meyer R., Reinberg D., Darzacq X., Tjian R. // Mol. Cell. 2019. V. 76. № 3. P. 395–411 e313.
53. Shukla S., Kavak E., Gregory M., Imashimizu M., Shutinoski B., Kashlev M., Oberdoerffer P., Sandberg R., Oberdoerffer S. // Nature. 2011. V. 479. № 7371. P. 74–79.
54. Marina R.J., Sturgill D., Bailly M.A., Thenoz M., Varma G., Prigge M.F., Nanan K.K., Shukla S., Haque N., Oberdoerffer S. // EMBO. 2016. V. 35. № 3. P. 335–355.

55. Nanavaty V., Abrash E.W., Hong C., Park S., Fink E.E., Li Z., Sweet T.J., Bhasin J.M., Singuri S., Lee B.H., et al. // Mol. Cell. 2020. V. 78. № 4. P. 752–764 e756.
56. Chernukhin I., Shamsuddin S., Kang S.Y., Bergstrom R., Kwon Y.W., Yu W., Whitehead J., Mukhopadhyay R., Docquier F., Farrar D., et al. // Mol. Cell Biol. 2007. V. 27. № 5. P. 1631–1648.
57. Heger P., Marin B., Schierenberg E. // BMC Mol. Biol. 2009. V. 10. P. 84.
58. Heger P., Marin B., Bartkuhn M., Schierenberg E., Wiehe T. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 43. P. 17507–17512.
59. Moon H., Filippova G., Loukinov D., Pugacheva E., Chen Q., Smith S.T., Munhall A., Grewe B., Bartkuhn M., Arnold R., et al. // EMBO Reports. 2005. V. 6. № 2. P. 165–170.
60. Holohan E.E., Kwong C., Adryan B., Bartkuhn M., Herold M., Renkawitz R., Russell S., White R. // PLoS Genet. 2007. V. 3. № 7. P. e112.
61. Kadota M., Hara Y., Tanaka K., Takagi W., Tanegashima C., Nishimura O., Kuraku S. // Sci. Rep. 2017. V. 7. № 1. P. 4957.
62. Narendra V., Rocha P.P., An D., Raviram R., Skok J.A., Mazzoni E.O., Reinberg D. // Science. 2015. V. 347. № 6225. P. 1017–1021.
63. Savitsky M., Kim M., Kravchuk O., Schwartz Y.B. // Genetics. 2016. V. 202. № 2. P. 601–617.
64. Luo H., Wang F., Zha J., Li H., Yan B., Du Q., Yang F., Sobh A., Vulpe C., Drusbosky L., et al. // Blood. 2018. V. 132. № 8. P. 837–848.
65. Kyrchanova O., Zolotarev N., Mogila V., Maksimenko O., Schedl P., Georgiev P. // Development. 2017. V. 144. № 14. P. 2663–2672.
66. Kyrchanova O., Maksimenko O., Ibragimov A., Sokolov V., Postika N., Lukyanova M., Schedl P., Georgiev P. // Sci. Adv. 2020. V. 6. № 13. P. eaaz3152.
67. Schwalie P.C., Ward M.C., Cain C.E., Faure A.J., Gilad Y., Odom D.T., Flicek P. // Genome Biol. 2013. V. 14. № 12. P. R148.
68. Bonchuk A., Kamalyan S., Mariasina S., Boyko K., Popov V., Maksimenko O., Georgiev P. // Sci. Rep. 2020. V. 10. № 1. P. 2677.
69. Bonchuk A., Maksimenko O., Kyrchanova O., Ivlieva T., Mogila V., Deshpande G., Wolle D., Schedl P., Georgiev P. // BMC Biol. 2015. V. 13. P. 63.
70. Nishana M., Ha C., Rodriguez-Hernaez J., Ranjbaran A., Chio E., Nora E.P., Badri S.B., Kloetgen A., Bruneau B.G., Tsirigos A., et al. // Genome Biol. 2020. V. 21. № 1. P. 108.
71. Li Y., Haarhuis J.H.I., Sedeno Cacciatore A., Oldenkamp R., van Ruiten M.S., Willems L., Teunissen H., Muir K.W., de Wit E., Rowland B.D., et al. // Nature. 2020. V. 578. № 7795. P. 472–476.
72. Xiao T., Wallace J., Felsenfeld G. // Mol. Cell Biol. 2011. V. 31. № 11. P. 2174–2183.
73. Zlatanova J., Caiafa P. // J. Cell Sci. 2009. V. 122. № Pt 9. P. 1275–1284.
74. Marino M.M., Rega C., Russo R., Valletta M., Gentile M.T., Esposito S., Baglivo I., De Feis I., Angelini C., Xiao T., et al. // J. Biol. Chem. 2019. V. 294. № 3. P. 861–873.
75. Pena-Hernandez R., Marques M., Hilmi K., Zhao T., Saad A., Alaoui-Jamali M.A., del Rincon S.V., Ashworth T., Roy A.L., Emerson B.M., et al. // Proc. Natl. Acad. Sci. USA. 2015. V. 112. № 7. P. E677–686.
76. Nora E.P., Goloborodko A., Valton A.L., Gibcus J.H., Uebersohn A., Abdennur N., Dekker J., Mirny L.A., Bruneau B.G. // Cell. 2017. V. 169. № 5. P. 930–944 e922.
77. Yao H., Brick K., Evrard Y., Xiao T., Camerini-Otero R.D., Felsenfeld G. // Genes Dev. 2010. V. 24. № 22. P. 2543–2555.
78. Uuskula-Reimand L., Hou H., Samavarchi-Tehrani P., Rudan M.V., Liang M., Medina-Rivera A., Mohammed H., Schmidt D., Schwalie P., Young E.J., et al. // Genome Biol. 2016. V. 17. № 1. P. 182.
79. Gittens W.H., Johnson D.J., Allison R.M., Cooper T.J., Thomas H., Neale M.J. // Nat. Commun. 2019. V. 10. № 1. P. 4846.
80. Jantz D., Berg J.M. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 20. P. 7589–7593.
81. Dovat S., Ronni T., Russell D., Ferrini R., Cobb B.S., Smale S.T. // Genes Dev. 2002. V. 16. № 23. P. 2985–2990.
82. Rizkallah R., Alexander K.E., Hurt M.M. // Cell Cycle. 2011. V. 10. № 19. P. 3327–3336.
83. Luo H., Yu Q., Liu Y., Tang M., Liang M., Zhang D., Xiao T.S., Wu L., Tan M., Ruan Y., et al. // Science Adv. 2020. V. 6. № 8. P. eaaw4651.
84. Caiafa P., Zlatanova J. // J. Cell. Physiol. 2009. V. 219. № 2. P. 265–270.
85. Farrar D., Rai S., Chernukhin I., Jagodic M., Ito Y., Yammine S., Ohlsson R., Murrell A., Klenova E. // Mol. Cell Biol. 2010. V. 30. № 5. P. 1199–1216.
86. Pavlaki I., Docquier F., Chernukhin I., Kita G., Gretton S., Clarkson C.T., Teif V.B., Klenova E. // Biochim. Biophys. Acta Gene Regul. Mech. 2018. V. 1861. № 8. P. 718–730.
87. Torrano V., Navascues J., Docquier F., Zhang R., Burke L.J., Chernukhin I., Farrar D., Leon J., Berciano M.T., Renkawitz R., et al. // J. Cell Sci. 2006. V. 119. Pt 9. P. 1746–1759.
88. Wang A.J., Han Y., Jia N., Chen P., Minden M.D. // Leukemia. 2020. V. 34. № 5. P. 1278–1290.
89. MacPherson M.J., Beatty L.G., Zhou W., Du M., Sadowski P.D. // Mol. Cell Biol. 2009. V. 29. № 3. P. 714–725.
90. Golovnin A., Volkov I., Georgiev P. // J. Cell Sci. 2012. V. 125. № Pt 8. P. 2064–2074.
91. Rosonina E., Akhter A., Dou Y., Babu J., Sri Theivakadadcham V.S. // Transcription. 2017. V. 8. № 4. P. 220–231.
92. Wallace J.A., Felsenfeld G. // Curr. Opin. Genet. Dev. 2007. V. 17. № 5. P. 400–407.
93. Barkess G., West A.G. // Epigenomics. 2012. V. 4. № 1. P. 67–80.
94. Ghirlando R., Felsenfeld G. // Genes Dev. 2016. V. 30. № 8. P. 881–891.
95. Farrell C.M., West A.G., Felsenfeld G. // Mol. Cell Biol. 2002. V. 22. № 11. P. 3820–3831.
96. West A.G., Huang S., Gaszner M., Litt M.D., Felsenfeld G. // Mol. Cell. 2004. V. 16. № 3. P. 453–463.
97. Dickson J., Gowher H., Strogantsev R., Gaszner M., Hair A., Felsenfeld G., West A.G. // PLoS Genet. 2010. V. 6. № 1. P. e1000804.
98. Gowher H., Brick K., Camerini-Otero R.D., Felsenfeld G. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 7. P. 2370–2375.
99. Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N., Mirny L.A. // Cell Reports. 2016. V. 15. № 9. P. 2038–2049.
100. Nishiyama T. // Curr. Opin. Cell Biol. 2019. V. 58. P. 8–14.
101. Morales C., Losada A. // Curr. Opin. Cell Biol. 2018. V. 52. P. 51–57.
102. Parelho V., Hadjur S., Spivakov M., Leleu M., Sauer S., Gregson H.C., Jarmuz A., Canzonetta C., Webster Z., Nesterova T., et al. // Cell. 2008. V. 132. № 3. P. 422–433.
103. Wendt K.S., Yoshida K., Itoh T., Bando M., Koch B., Schirghuber E., Tsutsumi S., Nagae G., Ishihara K., Mishiro T., et al. // Nature. 2008. V. 451. № 7180. P. 796–801.
104. Pugacheva E.M., Kubo N., Loukinov D., Tajmul M., Kang

S., Kovalchuk A.L., Strunnikov A.V., Zentner G.E., Ren B., Lobanenkov V.V. // Proc. Natl. Acad. Sci. USA. 2020. V. 117. № 4. P. 2020–2031.
105. Rao S.S.P., Huang S.C., Glenn St Hilaire B., Engreitz J.M., Perez E.M., Kieffer-Kwon K.R., Sanborn A.L., Johnstone S.E., Bascom G.D., Bochkov I.D., et al. // Cell. 2017. V. 171. № 2. P. 305–320 e324.
106. Wutz G., Varnai C., Nagasaka K., Cisneros D.A., Stocsits R.R., Tang W., Schoenfelder S., Jessberger G., Muhar M., Hossain M.J., et al. // EMBO. 2017. V. 36. № 24. P. 3573–3599.
107. Ladurner R., Bhaskara V., Huis in 't Veld P.J., Davidson I.F., Kreidl E., Petzold G., Peters J.M. // Curr. Biol. 2014. V. 24. № 19. P. 2228–2237.
108. Elbatsh A.M.O., Haarhuis J.H.I., Petela N., Chapard C., Fish A., Celie P.H., Stadnik M., Ristic D., Wyman C., Medema R.H., et al. // Mol. Cell. 2016. V. 61. № 4. P. 575–588.
109. Vian L., Pekowska A., Rao S.S.P., Kieffer-Kwon K.R., Jung S., Baranello L., Huang S.C., El Khattabi L., Dose M., Pruett N., et al. // Cell. 2018. V. 175. № 1. P. 292–294.
110. Vietri Rudan M., Barrington C., Henderson S., Ernst C., Odom D.T., Tanay A., Hadjur S. // Cell Rep. 2015. V. 10. № 8. P. 1297–1309.
111. de Wit E., Vos E.S., Holwerda S.J., Valdes-Quezada C., Verstegen M.J., Teunissen H., Splinter E., Wijchers P.J., Krijger P.H., de Laat W. // Mol. Cell. 2015. V. 60. № 4. P. 676–684.
112. Guo Y., Xu Q., Canzio D., Shou J., Li J., Gorkin D.U., Jung I., Wu H., Zhai Y., Tang Y., et al. // Cell. 2015. V. 162. № 4. P. 900–910.
113. Davidson I.F., Bauer B., Goetz D., Tang W., Wutz G., Peters J.M. // Science. 2019. V. 366. № 6471. P. 1338–1345.
114. Kim Y., Shi Z., Zhang H., Finkelstein I.J., Yu H. // Science. 2019. V. 366. № 6471. P. 1345–1349.
115. Stigler J., Camdere G.O., Koshland D.E., Greene E.C. // Cell Rep. 2016. V. 15. № 5. P. 988–998.
116. Ulianov S.V., Khrameeva E.E., Gavrilov A.A., Flyamer I.M., Kos P., Mikhaleva E.A., Penin A.A., Logacheva M.D., Imakaev M.V., Chertovich A., et al. // Genome Res. 2016. V. 26. № 1. P. 70–84.
117. Luzhin A.V., Flyamer I.M., Khrameeva E.E., Ulianov S.V., Razin S.V., Gavrilov A.A. // J. Cell. Biochem. 2019. V. 120. № 3. P. 4494–4503.
118. Holzmann J., Politi A.Z., Nagasaka K., Hantsche-Grininger M., Walther N., Koch B., Fuchs J., Durnberger G., Tang W., Ladurner R., et al. // eLife. 2019. V. 8. P. e46269.
119. Kentepozidou E., Aitken S.J., Feig C., Stefflova K., Ibarra-Soria X., Odom D.T., Roller M., Flicek P. // Genome Biol. 2020. V. 21. № 1. P. 5.
120. Haarhuis J.H.I., van der Weide R.H., Blomen V.A., Yanez-Cuna J.O., Amendola M., van Ruiten M.S., Krijger P.H.L., Teunissen H., Medema R.H., van Steensel B., et al. // Cell. 2017. V. 169. № 4. P. 693–707 e614.
121. Gassler J., Brandao H.B., Imakaev M., Flyamer I.M., Ladstatter S., Bickmore W.A., Peters J.M., Mirny L.A., Tachibana K. // EMBO J. 2017. V. 36. № 24. P. 3600–3618.
122. Zhang H., Emerson D.J., Gilgenast T.G., Titus K.R., Lan Y., Huang P., Zhang D., Wang H., Keller C.A., Giardine B., et al. // Nature. 2019. V. 576. № 7785. P. 158–162.
123. Owens N., Papadopoulou T., Festuccia N., Tachtsidi A., Gonzalez I., Dubois A., Vandormael-Pournin S., Nora E.P., Bruneau B.G., Cohen-Tannoudji M., et al. // eLife. 2019. V. 8. P. e47898.
124. Lambert S.A., Jolma A., Campitelli L.F., Das P.K., Yin Y., Albu M., Chen X., Taipale J., Hughes T.R., Weirauch M.T. // Cell. 2018. V. 172. № 4. P. 650–665.
125. Lambert S.A., Yang A.W.H., Sasse A., Cowley G., Albu M., Caddick M.X., Morris Q.D., Weirauch M.T., Hughes T.R. // Nat. Genet. 2019. V. 51. № 6. P. 981–989.
126. Imbeault M., Helleboid P.Y., Trono D. // Nature. 2017. V. 543. № 7646. P. 550–554.
127. Schmitges F.W., Radovani E., Najafabadi H.S., Barazandeh M., Campitelli L.F., Yin Y., Jolma A., Zhong G., Guo H., Kanagalingam T., et al. // Genome Res. 2016. V. 26. № 12. P. 1742–1752.
128. Barazandeh M., Lambert S.A., Albu M., Hughes T.R. // G3 (Bethesda). 2018. V. 8. № 1. P. 219–229.
129. Platt R.N., 2nd, Vandewege M.W., Ray D.A. // Chromosome Research. 2018. V. 26. № 1–2. P. 25–43.
130. Bruno M., Mahgoub M., Macfarlan T.S. // Annu. Rev. Genet. 2019. V. 53. P. 393–416.
131. Emerson R.O., Thomas J.H. // J. Virol. 2011. V. 85. № 22. P. 12043–12052.
132. Okumura K., Sakaguchi G., Naito K., Tamura T., Igarashi H. // Nucleic Acids Res. 1997. V. 25. № 24. P. 5025–5032.
133. Rohrmoser M., Kluge M., Yahia Y., Gruber-Eber A., Maqbool M.A., Forne I., Krebs S., Blum H., Greifenberg A.K., Geyer M., et al. // Nucleic Acids Res. 2019. V. 47. № 2. P. 700–715.
134. Diehl A.G., Ouyang N., Boyle A.P. // Nat. Commun. 2020. V. 11. № 1. P. 1796.
135. Herz H.M., Garruss A., Shilatifard A. // Trends in Biochemical Sciences. 2013. V. 38. № 12. P. 621–639.
136. Maeda T. // Int. J. Hematol. 2016. V. 104. № 3. P. 310–323.
137. Al Chiblak M., Steinbeck F., Thiesen H.J., Lorenz P. // BMC Molecular and Cell Biology. 2019. V. 20. № 1. P. 60.
138. Schumacher C., Wang H., Honer C., Ding W., Koehn J., Lawrence Q., Coulis C.M., Wang L.L., Ballinger D., Bowen B.R., et al. // J. Biol. Chem. 2000. V. 275. № 22. P. 17173–17179.
139. Yang P., Wang Y., Macfarlan T.S. // Trends Genet. 2017. V. 33. № 11. P. 871–881.
140. Francis M., Cheng H., Ma P., Grider A. // Biol. Trace Elem. Res. 2019. V. 192. № 2. P. 83–90.
141. Ogo O.A., Tyson J., Cockell S.J., Howard A., Valentine R.A., Ford D. // Mol. Cell Biol. 2015. V. 35. № 6. P. 977–987.
142. Kino T., Pavlatou M.G., Moraitis A.G., Nemery R.L., Raygada M., Stratakis C.A. // J. Clin. Endocrinol. Metab. 2012. V. 97. № 8. P. E1557–1566.
143. Fadda A., Syed N., Mackeh R., Papadopoulou A., Suzuki S., Jithesh P.V., Kino T. // Scientific Reports. 2017. V. 7. P. 41598.
144. Wagner S., Hess M.A., Ormonde-Hanson P., Malandro J., Hu H., Chen M., Kehrer R., Frodsham M., Schumacher C., Beluch M., et al. // J. Biol. Chem. 2000. V. 275. № 21. P. 15685–15690.
145. Frietze S., Lan X., Jin V.X., Farnham P.J. // J. Biol. Chem. 2010. V. 285. № 2. P. 1393–1403.
146. Brix D.M., Bundgaard Clemmensen K.K., Kallunki T. // Cells. 2020. V. 9. № 1. P. 223.
147. Galli G.G., Multhaupt H.A., Carrara M., de Lichtenberg K.H., Christensen I.B., Linnemann D., Santoni-Rugiu E., Calogero R.A., Lund A.H. // Oncogene. 2014. V. 33. № 25. P. 3342–3350.
148. Noll L., Peterson F.C., Hayes P.L., Volkman B.F., Sander T. // Leukemia Research. 2008. V. 32. № 10. P. 1582–1592.
149. Peterson F.C., Hayes P.L., Waltner J.K., Heisner A.K., Jensen D.R., Sander T.L., Volkman B.F. // J. Mol. Biol. 2006. V. 363. № 1. P. 137–147.
150. Helleboid P.Y., Heusel M., Duc J., Piot C., Thorball C.W., Coluccio A., Pontis J., Imbeault M., Turelli P., Aebersold R.,

et al. // EMBO. 2019. V. 38. № 18. P. e101220.
151. Duan Z., Person R.E., Lee H.H., Huang S., Donadieu J., Badolato R., Grimes H.L., Papayannopoulou T., Horwitz M.S. // Mol. Cell Biol. 2007. V. 27. № 19. P. 6889–6902.
152. Galli G.G., Carrara M., Francavilla C., de Lichtenberg K.H., Olsen J.V., Calogero R.A., Lund A.H. // Mol. Cell Biol. 2013. V. 33. № 22. P. 4504–4516.
153. Myslinski E., Krol A., Carbon P. // J. Biol. Chem. 1998. V. 273. № 34. P. 21998–22006.
154. Schuster C., Myslinski E., Krol A., Carbon P. // EMBO. 1995. V. 14. № 15. P. 3777–3787.
155. Bailey S.D., Zhang X., Desai K., Aid M., Corradin O., Cowper-Sal Lari R., Akhtar-Zaidi B., Scacheri P.C., Haibe-Kains B., Lupien M. // Nat. Commun. 2015. V. 2. P. 6186.
156. Heidari N., Phanstiel D.H., He C., Grubert F., Jahanbani F., Kasowski M., Zhang M.Q., Snyder M.P. // Genome Res. 2014. V. 24. № 12. P. 1905–1917.
157. Myslinski E., Gerard M.A., Krol A., Carbon P. // J. Biol. Chem. 2006. V. 281. № 52. P. 39953–39962.
158. Ngondo-Mbongo R.P., Myslinski E., Aster J.C., Carbon P. // Nucleic Acids Res. 2013. V. 41. № 7. P. 4000–4014.
159. Schaub M., Krol A., Carbon P. // Nucleic Acids Res. 2000. V. 28. № 10. P. 2114–2121.
160. Schaub M., Myslinski E., Krol A., Carbon P. // J. Biol. Chem. 1999. V. 274. № 35. P. 25042–25050.
161. Sathyan K.M., McKenna B.D., Anderson W.D., Duarte F.M., Core L., Guertin M.J. // Genes Dev. 2019. V. 33. № 19–20. P. 1441–1455.
162. Mourad R., Cuvier O. // Nucleic Acids Res. 2018. V. 46. № 5. P. e27.
163. Wen Z., Huang Z.T., Zhang R., Peng C. // Cell Biol. Toxicol. 2018. V. 34. № 6. P. 471–478.
164. Yang Y., Zhang R., Singh S., Ma J. // Bioinformatics. 2017. V. 33. № 14. P. i252–i260.
165. Raab J.R., Chiu J., Zhu J., Katzman S., Kurukuti S., Wade P.A., Haussler D., Kamakaka R.T. // EMBO. 2012. V. 31. № 2. P. 330–350.
166. van Bortle K., Phanstiel D.H., Snyder M.P. // Genome Biol. 2017. V. 18. № 1. P. 180.
167. Layat E., Probst A.V., Tourmente S. // Biochim. Biophys. Acta. 2013. V. 1829. № 3–4. P. 274–282.
168. Smith D.R., Jackson I.J., Brown D.D. // Cell. 1984. V. 37. № 2. P. 645–652.
169. Matthews N.E., White R. // BioEssays. 2019. P. e1900048.
170. Schwartz Y.B., Cavalli G. // Genetics. 2017. V. 205. № 1. P. 5–24.
171. Zolotarev N., Maksimenko O., Kyrchanova O., Sokolinskaya E., Osadchiy I., Girardot C., Bonchuk A., Ciglar L., Furlong E.E.M., Georgiev P. // Nucleic Acids Res. 2017. V. 45. № 21. P. 12285–12300.
172. Chung H.R., Schafer U., Jackle H., Bohm S. // EMBO Reports. 2002. V. 3. № 12. P. 1158–1162.
173. Chung H.R., Lohr U., Jackle H. // Mol. Biol. Evol. 2007. V. 24. № 9. P. 1934–1943.
174. Mackeh R., Marr A.K., Fadda A., Kino T. // Nuclear Receptor Signaling. 2018. V. 15. P. 1550762918801071.
175. Ecco G., Imbeault M., Trono D. // Development. 2017. V. 144. № 15. P. 2719–2729.
176. Jauch R., Bourenkov G.P., Chung H.R., Urlaub H., Reidt U., Jackle H., Wahl M.C. // Structure. 2003. V. 11. № 11. P. 1393–1402.
177. Zolotarev N., Fedotova A., Kyrchanova O., Bonchuk A., Penin A.A., Lando A.S., Eliseeva I.A., Kulakovskiy I.V., Maksimenko O., Georgiev P. // Nucleic Acids Res. 2016. V. 44. № 15. P. 7228–7241.
178. Gaszner M., Vazquez J., Schedl P. // Genes Dev. 1999. V. 13. № 16. P. 2098–2107.
179. Page A.R., Kovacs A., Deak P., Torok T., Kiss I., Dario P., Bastos C., Batista P., Gomes R., Ohkura H., et al. // EMBO. 2005. V. 24. № 24. P. 4304–4315.
180. Baxley R.M., Soshnev A.A., Koryakov D.E., Zhimulev I.F., Geyer P.K. // Dev. Biol. 2011. V. 356. № 2. P. 398–410.
181. Mohan M., Bartkuhn M., Herold M., Philippen A., Heinl N., Bardenhagen I., Leers J., White R.A., Renkawitz-Pohl R., Saumweber H., et al. // EMBO. 2007. V. 26. № 19. P. 4203–4214.
182. Gambetta M.C., Furlong E.E.M. // Genetics. 2018. V. 210. № 1. P. 129–136.
183. Baxley R.M., Bullard J.D., Klein M.W., Fell A.G., Morales-Rosado J.A., Duan T., Geyer P.K. // Nucleic Acids Res. 2017. V. 45. № 8. P. 4463–4478.
184. Maksimenko O., Kyrchanova O., Klimenko N., Zolotarev N., Elizarova A., Bonchuk A., Georgiev P. // Biochim. Biophys. Acta Gene Regul. Mech. 2020. V. 1863. № 1. P. 194446.
185. Li J., Gilmour D.S. // EMBO. 2013. V. 32. № 13. P. 1829–1841.
186. Schwartz Y.B., Linder-Basso D., Kharchenko P.V., Tolstorukov M.Y., Kim M., Li H.B., Gorchakov A.A., Minoda A., Shanower G., Alekseyenko A.A., et al. // Genome Res. 2012. V. 22. № 11. P. 2188–2198.
187. Soshnev A.A., He B., Baxley R.M., Jiang N., Hart C.M., Tan K., Geyer P.K. // Nucleic Acids Res. 2012. V. 40. № 12. P. 5415–5431.
188. Maksimenko O., Bartkuhn M., Stakhov V., Herold M., Zolotarev N., Jox T., Buxa M.K., Kirsch R., Bonchuk A., Fedotova A., et al. // Genome Res. 2015. V. 25. № 1. P. 89–99.
189. Negre N., Brown C.D., Shah P.K., Kheradpour P., Morrison C.A., Henikoff J.G., Feng X., Ahmad K., Russell S., White R.A., et al. // PLoS Genet. 2010. V. 6. № 1. P. e1000814.
190. Negre N., Brown C.D., Ma L., Bristow C.A., Miller S.W., Wagner U., Kheradpour P., Eaton M.L., Loriaux P., Sealfon R., et al. // Nature. 2011. V. 471. № 7339. P. 527–531.
191. Baumann D.G., Gilmour D.S. // Nucleic Acids Res. 2017. V. 45. № 18. P. 10481–10491.
192. Soshnev A.A., Baxley R.M., Manak J.R., Tan K., Geyer P.K. // Development. 2013. V. 140. № 17. P. 3613–3623.
193. Melnikova L., Elizar'ev P., Erokhin M., Molodina V., Chetverina D., Kostyuchenko M., Georgiev P., Golovnin A. // Sci. Rep. 2019. V. 9. № 1. P. 5314.
194. Kyrchanova O., Maksimenko O., Stakhov V., Ivlieva T., Parshikov A., Studitsky V.M., Georgiev P. // PLoS Genet. 2013. V. 9. № 7. P. e1003606.
195. Kyrchanova O., Chetverina D., Maksimenko O., Kullyev A., Georgiev P. // Nucleic Acids Res. 2008. V. 36. № 22. P. 7019–7028.
196. Maeda R.K., Karch F. // Chromosoma. 2015. V. 124. № 3. P. 293–307.
197. Kyrchanova O., Mogila V., Wolle D., Magbanua J.P., White R., Georgiev P., Schedl P. // Mech. Dev. 2015. V. 138. Pt 2. P. 122–132.
198. Gruzdeva N., Kyrchanova O., Parshikov A., Kullyev A., Georgiev P. // Mol. Cell Biol. 2005. V. 25. № 9. P. 3682–3689.
199. Barges S., Mihaly J., Galloni M., Hagstrom K., Muller M., Shanower G., Schedl P., Gyurkovics H., Karch F. // Development. 2000. V. 127. № 4. P. 779–790.
200. Iampietro C., Gummalla M., Mutero A., Karch F., Maeda R.K. // PLoS Genet. 2010. V. 6. № 12. P. e1001260.
201. Bender W., Lucas M. // Genetics. 2013. V. 193. № 4. P. 1135–1147.
202. Bowman S.K., Deaton A.M., Domingues H., Wang P.I.,

Sadreyev R.I., Kingston R.E., Bender W. // eLife. 2014. V. 3. P. e02833.
203. Kyrchanova O., Mogila V., Wolle D., Deshpande G., Parshikov A., Cleard F., Karch F., Schedl P., Georgiev P. // PLoS Genet. 2016. V. 12. № 7. P. e1006188.
204. Kyrchanova O., Sabirov M., Mogila V., Kurbidaeva A., Postika N., Maksimenko O., Schedl P., Georgiev P. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 27. P. 13462–13467.
205. Wang Q., Sun Q., Czajkowsky D.M., Shao Z. // Nat. Commun. 2018. V. 9. № 1. P. 188.
206. Ramirez F., Bhardwaj V., Arrigoni L., Lam K.C., Gruning B.A., Villaveces J., Habermann B., Akhtar A., Manke T. // Nat. Commun. 2018. V. 9. № 1. P. 189.
207. Chathoth K.T., Zabet N.R. // Genome Res. 2019. V. 29. № 4. P. 613–625.
208. Dorsett D. // Trends Genet. 2019. V. 35. № 7. P. 542–551.
209. Perez-Rico Y.A., Barillot E., Shkumatava A. // iScience. 2020. V. 23. № 5. P. 101046.