
BIOCHEMISTRY, BIOPHYSICS,
AND MOLECULAR BIOLOGY

Architecture of Promoters of House-Keeping Genes in Polytene Chromosome Interbands of *Drosophila melanogaster*

T. Yu. Zykova^{a,*}, V. G. Levitsky^{b,c}, and Academician I. F. Zhimulev^{a,c}

Received September 25, 2018

Abstract—This is the first study to investigate the molecular-genetic organization of polytene chromosome interbands located on both molecular and cytological maps of *Drosophila* genome. The majority of the studied interbands contained one gene with a single transcription initiation site; the remaining interbands contained one gene with several alternative promoters, two or more unidirectional genes, and “head-to-head” arranged genes. In addition, intricately arranged interbands containing three or more genes in both unidirectional and bidirectional orientation were found. Insulator proteins, ORC, *P*-insertions, DNase I hypersensitive sites, and other open chromatin structures were situated in the promoter region of the genes located in the interbands. This area is critical for the formation of the interband, an open chromatin region in which gene transcription and replication are combined.

DOI: 10.1134/S1607672919020029

As a result of recent large-scale studies of chromatin states in the entire *D. melanogaster* genome [1], the idea that genes with ubiquitous activity are located in the interbands of polytene chromosomes was formulated [2]. It was found that the genes that are specifically active at certain stages and in individual tissues (developmental genes) are located in tightly packed bands, which were previously named the intercalary heterochromatin bands [3]. Earlier, we showed that, at the whole-genome level, the chromatin that corresponds to the interbands has a complex structure in terms of genetic organization [2]. In particular, it was found that interbands contain unidirectional and bidirectional promoters, genes entirely located in the interbands, and genes with different number of transcription initiation sites. In view of this, the question arose of a comprehensive study of all interbands located on the molecular and cytological maps of the *D. melanogaster* genome in order to determine the organization of promoter types in them.

In this study, we obtained new data on the topology and location of proteins and different characteristics of chromatin within interband DNA sequences. We found that interbands containing bidirectional and unidirectional genes have different organization of the promoter region, which is apparently crucial for the formation of the open-chromatin region that we describe as an interband in polytene chromosomes of *D. melanogaster*.

On the basis of data of the international project modENCODE and the results of our studies [4], we developed a model of four chromatin states [2]. Using information about the location of proteins that are characteristic of the genetically active chromatin, we developed the stereotype of its classification by the degree of compaction from the minimum (aquamarine, corresponds to interbands) to the maximum (ruby, corresponds to the black dense bands). Another two states—lazurite and malachite—correspond primarily to the coding gene regions [2, 4] and the borders of blocks of developmental genes [5].

According to calculations, there are five thousand interbands in *D. melanogaster* chromosomes, which contain approximately 5% DNA of the euchromatin part of the genome [6, 7]. Only 33 of them were precisely mapped on the molecular and cytological maps of *D. melanogaster* genome. Since all interbands have common properties (such as the location of 5'-UTRs of the genes with ubiquitous activity, the enrichment in genes with bidirectional orientation of regulatory

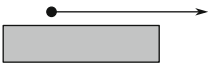

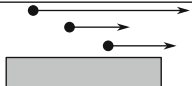
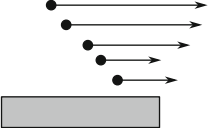


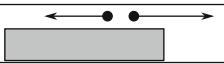


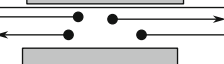
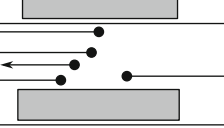
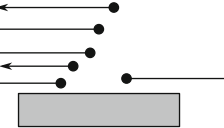
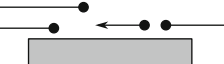
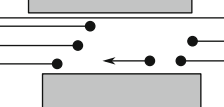
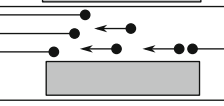
^a Institute of Molecular and Cellular Biology, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

^b Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

^c Novosibirsk State University, Novosibirsk, 630090 Russia

*e-mail: vatolina@mcb.nsc.ru

Table 1. Gene location in interbands

Group	No.	Scheme of genetic organization of interband	Number of genes in interband	Number of transcription initiation sites	Number of interbands
I	1		1	1	12
II	2		1	2	4
	3		1	3	2
	4		1	5	1
	Total				19
III	1		2	2	1
	2		2	3	2
	Total				3
IV	1		2	2	1
	2		2	2	2
	3		2	3	2
	4		2	4	2
	5		2	5	1
	6		2	6	1
	Total				9
V	1		3	3	3
	2		3	6	1
	3		4	7	1
	Total				5

regions, and the location of insulator proteins (CHRIZ, BEAF-32, and others) and correspond to the aquamarine chromatin [2], the analysis of their genetic structure and architecture is of primary interest.

Previously, we compared the organization and functions of the genes that are fully located in the open-chromatin region (interband) and the genes that occupy two structures (interband and gray band) [8].

In the present study, we investigated in detail the molecular-genetic organization of the sequences of 33 interbands localized on the molecular map of *D. melanogaster*. For analysis, we used the FlyBase version r.5.57 for *D. melanogaster*, which contains 29761 transcripts of 13753 protein-coding genes as well as interbands with 58 genes and 97 transcripts with unique starts.

In this study, we found that interbands had different genetic organization (Table 1). Most of the studied interbands contained one gene. The majority of these interbands contained genes with one transcription initiation site, and the remaining interbands contained one gene with several alternative promoters (Table 1, groups I and II). Other interband sequences contained two or more genes. Group III includes the interbands with two unidirectional genes and several transcription initiation sites. Group IV includes the interbands with bidirectional genes, and group V includes the intricately organized interbands with three or more genes in both unidirectional and bidirectional orientation.

Previously, we showed [9] that the transcription initiation sites of genes are located primarily in the center of interband sequences, and the border of the aquamarine chromatin, which corresponds to interbands on the molecular map, is almost always located in the first intron of an interband gene.

For the topology analysis, we chose the interbands from two contrasting groups I and IV, which contained the genes with one start and the bidirectional genes, respectively (Table 1). We evaluated the size of the interbands whose borders were localized using the four chromatin state model [2]. The length of the interbands that contained transcripts with one start was, on average, 3.5 kb, and their size ranged from 1 to 15.4 kb (Fig. 1a). The size of the interbands that contained bidirectional genes averaged 3.4 kb (1.6–7 kb). The distance between the promoters in these interbands averaged 0.7 kb (Fig. 1b). Such a large size of the four of the 33 interbands studied may be due to the fact that the used model identifies the first long intron of the active gene in the aquamarine chromatin. However, all genomic elements and proteins characteristic of the interbands are located in these domains in the promoter regions of the genes.

Recently we have shown [10] that insulator proteins CTCF and CHRIZ artificially recruited to a chromosomal site form a new interband. We studied [11] the distribution of insulator proteins CHRIZ and BEAF-32 as

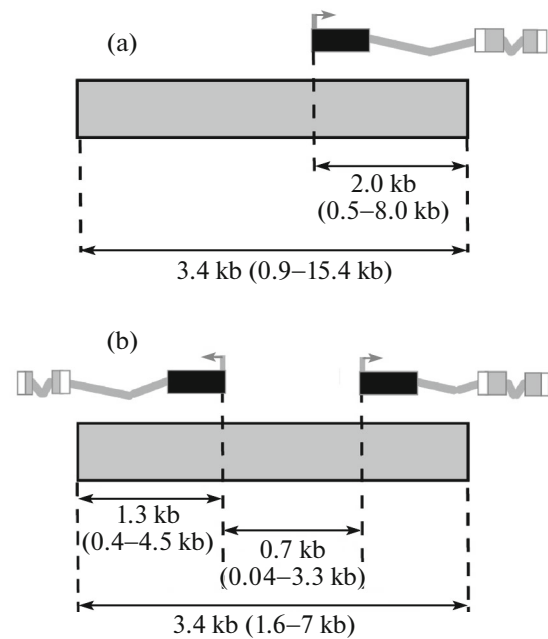


Fig. 1. Location of genes in interbands. Here and in Figs. 2 and 3, (a) are interbands containing the genes with one transcription initiation site (group I in Table 1) and (b) are the interbands containing bidirectional genes (group IV).

well as RNA polymerase II in the sequences of interbands containing one gene (group I) and bidirectional genes (group IV). The aquamarine chromatin sequences corresponding to the interbands were aligned with respect to the gene transcription initiation sites for group I interbands and with respect to the center between the transcription initiation sites for group IV interbands. Despite the fact that the average size of interbands was 3.4 kb (Figs. 1a, 1b), the proteins were located locally in the promoter region of the genes rather than occupied the entire sequence (Fig. 2). It can be seen that, in the sequences of the interbands that contained one gene, the insulator proteins CHRIZ and BEAF-32 were located immediately at the gene transcription initiation site, and RNA polymerase II was shifted 200 bp upstream of the transcription initiation site (Fig. 2a). In the interbands containing two bidirectional genes, BEAF-32 and CHRIZ proteins were located between the gene transcription initiation sites, and RNA polymerase II had two peaks in the promoter regions of the genes (Fig. 2b).

The promoters of many genes contain paused polymerase II, which generates short noncoding RNAs 25–60 bp long. The presence of this paused form of polymerase contributes to creating an accessible chromatin conformation in the regulatory gene region and binding the transcription factors [12].

In [13], short noncoding RNAs generated by the paused RNA polymerase II were mapped in the *Dro-*

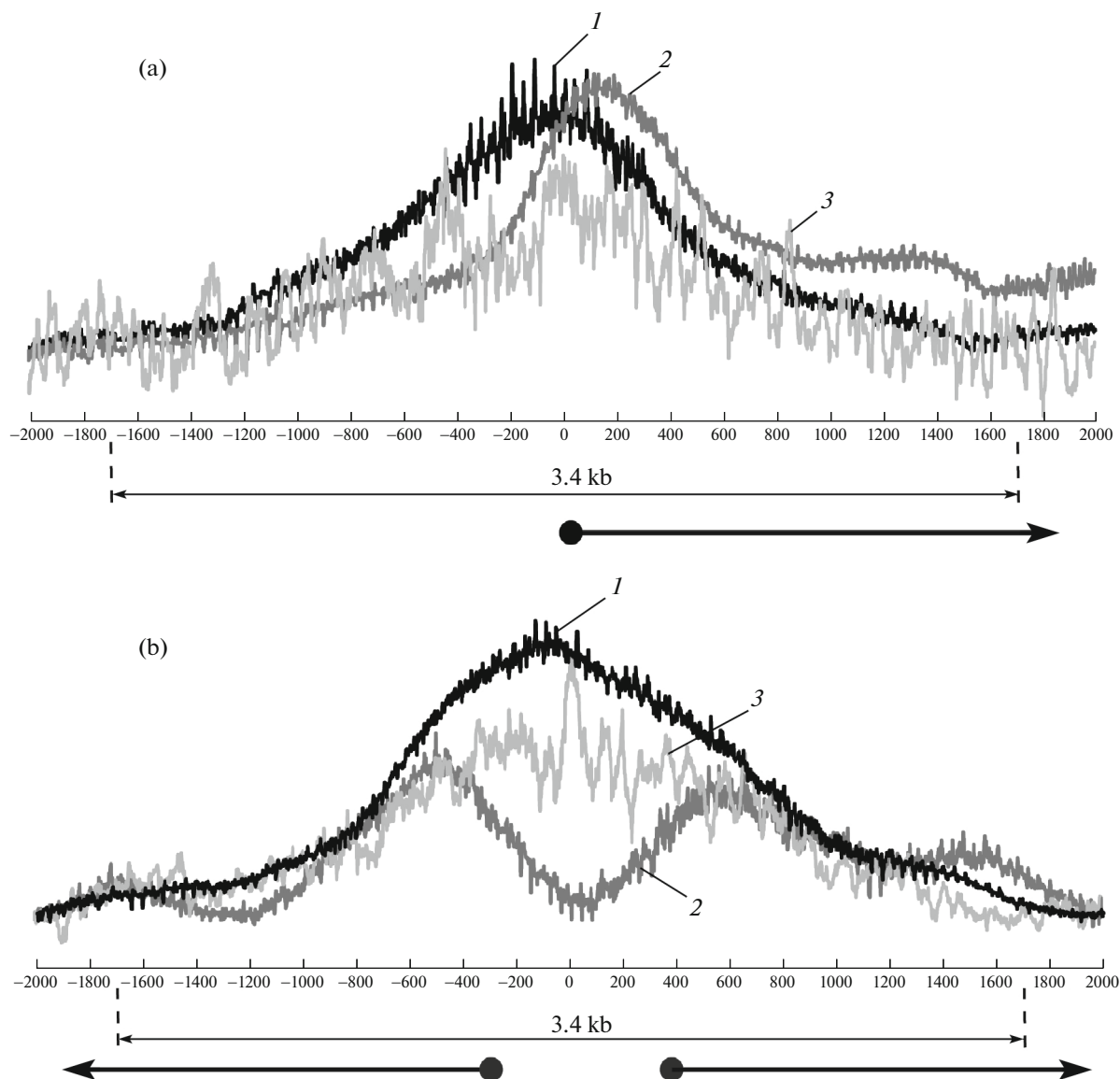


Fig. 2. Location of proteins in the DNA sequences corresponding to the interbands of polytene chromosomes. Here and in Fig. 3, point 0 on the abscissa axis corresponds to the gene transcription initiation site. Numbers designate the average binding intensity of proteins in S2 cells (modENCODE data): (1) CHRIZ (CHROMATOR), (2) RNA polymerase II, (3) BEAF-32.

sophila genome, and their 5' and 3' ends were determined. Therefore, their exact sequences are known. It was shown that such RNAs are characteristic of the open aquamarine chromatin domains (interbands) of *Drosophila* [9]. We investigated the topology of these noncoding RNAs, the proteins of the replicative complex (ORC2) [14], DNase I hypersensitive sites [1], P-element insertions (FlyBase r.5.57), and “broad” promoters [15] in the studied interbands.

All of these structures were located in the region of the transcription initiation sites of the interbands that contained the genes with one transcription initiation

site and had two independent peaks in the interbands contained bidirectional genes (Figs. 3a, 3b).

Thus, despite the fact that the size of individual interbands is, on average, 2 kb [6, 7] (according to our estimates, approximately 3.4 kb), the insulator proteins and other previously described structures characteristic of the interbands [2] had a narrow localization in the regulatory region of the genes located in the interbands. The enrichment of the aquamarine chromatin in ORC proteins and 5'-noncoding regions of the genes with mostly ubiquitous activity [2, 9] confirmed that fact that interbands combine the processes

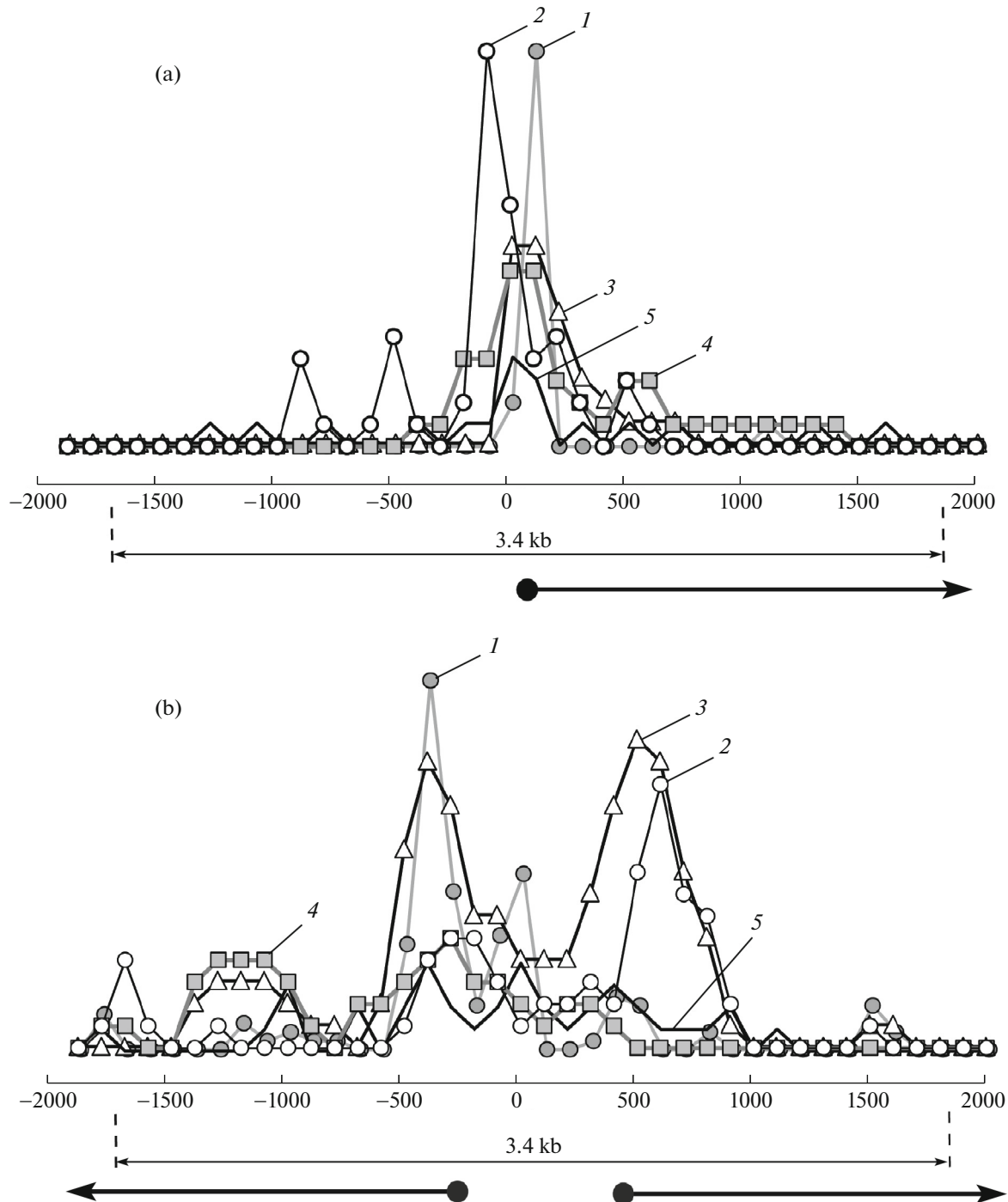


Fig. 3. Location of genomic elements in DNA sequences corresponding to the interbands of polytene chromosomes. Numbers designate the average number of annotations of characteristics in the interbands studied: (1) short RNAs generated by the paused form of RNA polymerase II in *D. melanogaster* embryos [13]; (2) P-element insertions (FlyBase r. 5.57); (3) broad promoters [15]; (4) replicative complex proteins (ORC2) in S2 cells [14]; (5) DNase I hypersensitive sites (DHS) in S2 cells [1].

of transcription initiation and replication of genes. Apparently, the gene regulatory region plays the decisive role in the formation of the domain of open decompacted chromatin of interbands.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (project no. 14-14-00934).

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

REFERENCES

1. Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., et al., Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*, *Nature*, 2011, vol. 471, no. 7339, pp. 480–485.
2. Zhimulev, I.F., Zykova, T.Y., Goncharov, F.P., Khoroshko, V.A., Demakova, O.V., Semeshin, V.F., Pokholkova, G.V., Boldyreva, L.V., Demidova, D.S., Babenko, V.N., Demakov, S.A., and Belyaeva, E.S., Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*, *PLoS One*, 2014, vol. 9, no. 7. e101631.
3. Zhimulev, I., Belyaeva, E., Makunin, I., Pirrotta, V., Volkova, E., and Alekseyenko, A., Influence of the *suur* gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes, *Chromosoma*, 2003, vol. 111, no. 6, pp. 377–398.
4. Vatolina, T.Yu., Boldyreva, L.V., Demakova, O.V., Demakov, S.A., Kokoza, E.B., Semeshin, V.F., Babenko, V.N., Goncharov, F.P., Belyaeva, E.S., and Zhimulev, I.F., Identical functional organization of cell line interphase and polytene chromosomes in *Drosophila melanogaster*, *PLoS One*, 2011, vol. 6, no. 10. e25960.
5. Khoroshko, V.A., Levitsky, V.G., Zykova, T.Y., Antonenko, O.V., Belyaeva, E.S., and Zhimulev, I.F., Chromatin heterogeneity and distribution of regulatory elements in the late-replicating intercalary heterochromatin domains of *Drosophila melanogaster* chromosomes, *PLoS One*, 2016, vol. 11, no. 6. e0157147.
6. Zhimulev, I.F., Belyaeva, E.S., Semeshin, V.F., Koryakov, D.E., Demakov, S.A., Demakova, O.V., Pokholkova, G.V., and Andreyeva, E.N., Polytene chromosomes: 70 years of genetic research, *Int. Rev. Cytol.*, 2004, vol. 241, pp. 203–275.
7. Beermann, W., Chromomeres and genes, *Results Probl. Cell Diff.*, 1972, vol. 4, pp. 1–33.
8. Zykova, T., Popova, O., Khoroshko, V., Levitskii, V., Lavrov, S., and Zhimulev, I. Genetic organization of open chromatin domains situated in polytene chromosome interbands in *Drosophila*, *Dokl. Biochem. Biophys.*, 2018, vol. 483, pp. 297–301.
9. Zykova, T., Levitsky, V., Belyaeva, E., and Zhimulev, I., Polytene chromosomes—a portrait of functional organization of the *Drosophila* genome, *Curr. Genomics*, 2018, vol. 19, no. 3, pp. 179–191.
10. Pokholkova, G., Demakov, S., Andreenkov, O., Andreenkova, N., Volkova, E., Belyaeva, E., and Zhimulev, I., Tethering of CHROMATOR and dCTCF proteins results in decompaction of condensed bands in the *Drosophila melanogaster* polytene chromosomes but does not affect their transcription and replication timing, *PLoS One*, 2018, vol. 13, no. 4. e0192634.
11. Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., Cuvier, O., Nègre, N., Cohen-Gonsaud, M., Margeat, E., and Nöllmann, M., Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome, *PLoS Genet.*, 2014, vol. 10, no. 8. e1004544.
12. Henriques, T., Gilchrist, D., Nechaev, S., Bern, M., Muse, G., Burkholder, A., Fargo, D., and Adelman, K., Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals, *Mol. Cell*, 2013, vol. 52, no. 4, pp. 517–528.
13. Nechaev, S., Fargo, D., Santos, G., Liu, L., Gao, Y., and Adelman, K., Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*, *Science*, 2010, vol. 327, no. 5963, pp. 335–338.
14. Eaton, M.L., Prinz, J.A., MacAlpine, H.K., Tretyakov, G., Kharchenko, P.V., and MacAlpine, D.M., Chromatin signatures of the *Drosophila* replication program, *Genome Res.*, 2011, vol. 21, no. 2, pp. 164–174.
15. Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., et al., Genome-wide analysis of promoter architecture in *Drosophila melanogaster*, *Genome Res.*, 2011, vol. 21, no. 2, pp. 182–192.

Translated by M. Batrukova