

Identification and Mapping of Ten New Potential Insulators in the *FXVD5-COX7A1* Region of Human Chromosome 19q13.12

D. A. Didych, S. B. Akopov*, E. V. Snezhkov, N. V. Skaptsova, L. G. Nikolaev, and E. D. Sverdlov

Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (495) 330-6538; E-mail: akser@humgen.siocb.ras.ru

Received December 12, 2008

Revision received February 24, 2009

Abstract—A positive–negative selection system revealed 10 potential insulators able to block enhancer interaction with promoter in the 10⁶ bp human chromosome 19 region between genes *FXVD5* and *COX7A1*. Relative positions of insulators and genes are in accord with the hypothesis that insulators subdivide genomic DNA into independently regulated loop domains.

DOI: 10.1134/S0006297909070049

Key words: human genome, functional mapping, annotations of DNA sequences, insulator

The human genome contains numerous functionally significant non-coding elements like promoters, enhancers, silencers, insulators, and binding sites for hormone receptors and transcription factors. Determination of complete nucleotide sequence of the genome does not give exhaustive information concerning the arrangement of functional elements, and as a result, about regulatory mechanisms of gene expression. Detection of such mechanisms is the task of functional genomics, studying both the mechanism of the activity of each gene and mutual influence of different functional units of the genome. Construction of a detailed map of functional elements for extended genome regions and whole genomes should bring us closer to understanding the functioning of the genome as a whole [1, 2].

Now great attention is paid to investigation of insulators, the boundary elements that form and maintain the loop–domain structure of interphase chromatin by protecting promoter against undesirable activation by outside enhancers (enhancer blocking effect) [3–5]. Owing to their ability to neutralize the position effect, insulators are also valuable tools in the design of genetic constructs [6, 7]. Nucleotide sequences of insulators have no clear similarity, and therefore computer search for these regulatory elements in eukaryotic genomes using conserved DNA regions is difficult.

We have developed a method for identification of insulators by their functional activity, which is based on positive–negative selection using the main property of insulators, their ability to block activation of promoter by enhancer. Earlier we applied this method for identification and mapping of eight insulators in the *FXVD5-COX7A1* region of human chromosome 19 [8]. This region contains a large number of characterized genes, which allows one to evaluate mutual position of insulators and genes and their possible relationship.

This work deals with mapping of 10 new insulators in the same locus and analysis of their arrangement relative to genes.

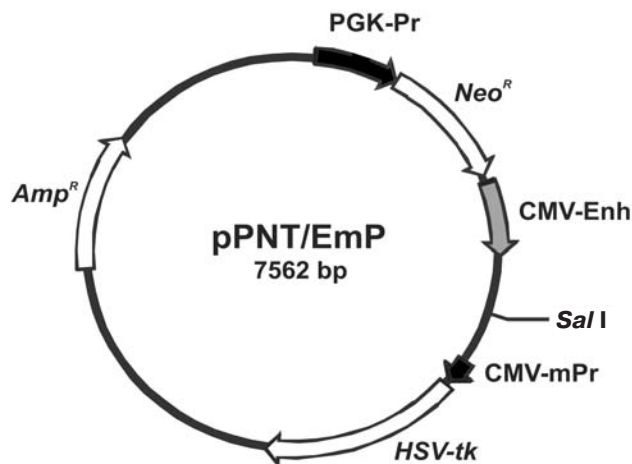
MATERIALS AND METHODS

Standard protocols. Growing and transformation of *Escherichia coli* strains, isolation of plasmid and genomic DNA, electrophoresis in agarose gel, and other standard procedures were carried out as described elsewhere [9].

Construction of library of potential insulators. Plasmid pPNT/EmP [8], earlier constructed by us on the basis of plasmid pPNT [10], was used for selection of potential insulators (Scheme). Later it was used as negative control in negative selection, thus conferring the GANC^S phenotype to the cells containing this plasmid. The plasmid pPNT/mP, containing in the regulatory region upstream from the thymidine kinase gene only minimal CMV (cytomegalovirus) promoter and conferring to the cells the GANC^R phenotype, was prepared as positive control [8].

Abbreviations: CMV, cytomegalovirus; GANC, 2-amino-1,9-dihydro-9-((2-hydroxy-1-(hydroxymethyl)ethoxy)methyl)-6H-purine-6-one (ganciclovir).

* To whom correspondence should be addressed.



Scheme of pPNT/EmP plasmid for positive–negative selection of insulators. mPGK1, promoter of mouse phosphoglycerate kinase gene; *Neo^R*, neomycin phosphotransferase gene; *HSV-tk*, herpes simplex virus thymidine kinase gene; CMV-mPr and CMV-Enh, cytomegalovirus minimal promoter and enhancer, respectively

Insulators were selected from the library of short (overlapping owing to the use of two restriction endonucleases) fragments of the human chromosome 19 region between genes *FXYD5* and *COX7A1*. The library was prepared as described earlier [8] and cloned into the *SalI* site of pPNT/EmP (Scheme).

The Chinese hamster ovary cell line CHO-K1 (CCL-61) was grown at 37°C in 5% CO₂ in a mixture of media DMEM/RPMI (1 : 1) supplemented with 10% fetal calf serum as described previously [11]. Cells were electroporated using a Gene Pulser Xcell apparatus (BioRad, USA) according to the manufacturer's recommendations. Twenty-four hours before transfection, 2·10⁶ cells were inoculated into a 75 cm² culture flask and incubated in 15 ml growth medium at 37°C in a CO₂-incubator. Immediately before transfection, cells were trypsinized, collected by centrifugation, and resuspended in RPMI-1640 medium without serum. Five micrograms of linearized plasmid and 10⁶ cells in 600–800 µl medium were introduced into the electroporation cell, and the

suspension was treated by a single pulse with exponential voltage decay (20 msec, 360 V, 800 µF). After incubation for 10–15 min on ice, transfected cells were transferred into a 25 cm² culture flask and incubated in 5 ml growth medium at 37°C in the CO₂ incubator.

After 24 h, the growth medium was replaced by another containing 500 µg/ml G418 (Geneticin; Gibco-BRL, USA), and after reaching a monolayer the cells were replated and their culture continued in the growth medium containing antibiotic. Two weeks later ganciclovir (Cymeven; Roche, Switzerland) was added to 4 µM to the growth medium during replating. After another two weeks, the cells resistant to G418 and ganciclovir were collected and frozen at –70°C.

The template of genomic DNA isolated from these cells was used for PCR amplification of the region located between CMV enhancer and promoter and containing sequences of supposed insulators; primers GGATTTC-CAAGTCTCCAGGGGAT (1L) and ACCTCCCACCG-TACACGCCT (1R) were used. After cleavage by *XhoI*, the resulting fragments were recloned at the *SalI* site into the pPNT/EmP vector. The *E. coli* clones obtained by transformation were arrayed in 96-well microtiter plates.

Library clone sequencing and mapping. Clones were sequenced on an ABI Prism 3100 automatic sequencer using the Big Dye Terminator Kit v.3.1 (Applied Biosystems, USA) and 1L primer (see above).

For analysis and mapping of the resulting sequences, programs BLAST [12] (<http://www.ncbi.nlm.nih.gov/BLAST/>) and Human Genome Browser [13] (<http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg12>) were used.

RESULTS AND DISCUSSION

Selection of putative insulators. To identify and map insulators within extended DNA regions, we elaborated a method based on protection of the reporter gene promoter against the effect of enhancer [14] and positive–negative selection [15]. The general scheme of the method is given in Fig. 1 of [8].

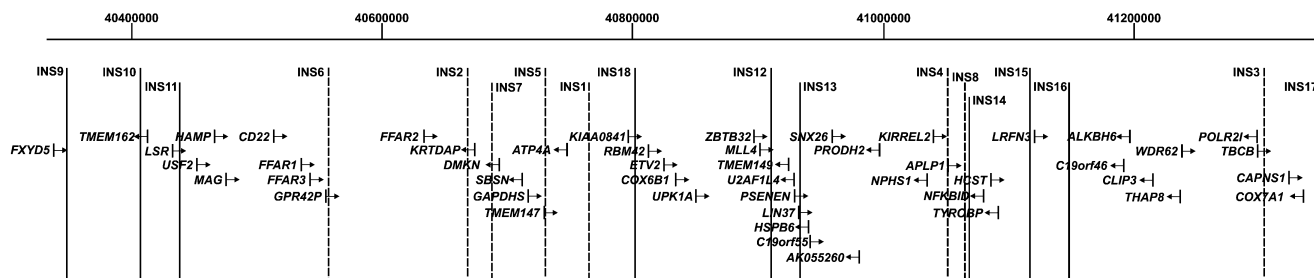


Fig. 1. Metric map of 10⁶ bp long *FXYD5-COX7A1* region of human chromosome 19. Solid vertical lines show positions of insulator sequences identified in this work (Nos. 11–18) and dashed lines point to sequences identified in [8]. Short vertical lines show positions of gene transcription start points and arrows show transcription direction. The map was designed using the Human Genome Browser [13], assembly of March, 2006.

For selection of insulator sequences, we constructed a vector called pPNT/EmP (see the Scheme). The vector contains the neomycin-resistance marker, enabling positive selection of cells with integrated plasmid, as well as the marker of negative selection, the herpes simplex virus thymidine kinase gene (*HSV-tk*), expression of which confers sensitivity to ganciclovir to cells with the integrated vector. The original vector efficiently expresses the *HSV-tk* gene, but insertion between this gene promoter and enhancer of the DNA sequence able to block the effect of enhancer (insulator) results in cessation or significant decrease in the *HSV-tk* gene expression in cells stably transfected by such vector, and cells containing this sequence become ganciclovir-resistant.

Sequences able to block the interaction between promoter and enhancer were selected within the 10^6 bp part of human chromosome 19 located in the 19q13.2 region between the *FXVD5* and *COX7A1* genes. The *FXVD5-COX7A1* region contains over 40 identified genes whose products (proteins) are characterized. The gene names, positions, and transcription direction are shown in Fig. 1. A library of potential insulators of this region was prepared as described elsewhere [8]. Briefly, DNA of 30 cosmids overlapping the *FXVD5-COX7A1* region was cleaved separately by two restriction endonucleases, *Sau3A* and *Csp6I*. Synthetic adapters for PCR amplification were attached to the resulting fragments (200–700 bp long). Adapters contained the *XhoI* cleavage site. Then

the two libraries were combined. The combined library was amplified by PCR using primers corresponding to synthetic adapters, cleaved by *XhoI* restriction endonuclease and cloned into plasmid pPNT/EmP at the *SaII* site.

To preserve the representativeness of the library, it was amplified on Petri dishes. Overall, 12,000 colonies were grown, which approximately corresponds to four-fold overlapping of the *FXVD5-COX7A1* region. All 12,000 colonies were washed away from Petri dishes and used for plasmid DNA isolation.

The resulting DNA was used to transfect the CHO cell line. Control transfections by plasmids pPNT/mP and pPNT/EmP were carried out in parallel. Cells with integrated vectors were selected by their resistance to geneticin (G-418). Then geneticin-resistant cells underwent negative selection using ganciclovir. The success of selection was confirmed by 100% death of cells containing pPNT/EmP, conferring to cells sensitivity to ganciclovir. At the same time most cells transfected by plasmid pPNT/mP were ganciclovir-resistant. Insignificant death of cells can be explained by the fact that in these cells pPNT/mP integrated into genome regions that were under control of endogenous enhancers.

It should be noted that cells can also become sensitive to ganciclovir if the cloned DNA fragment exhibits properties of a silencer [16, 17]. However, a sequence with silencer characteristics will also inhibit transcription

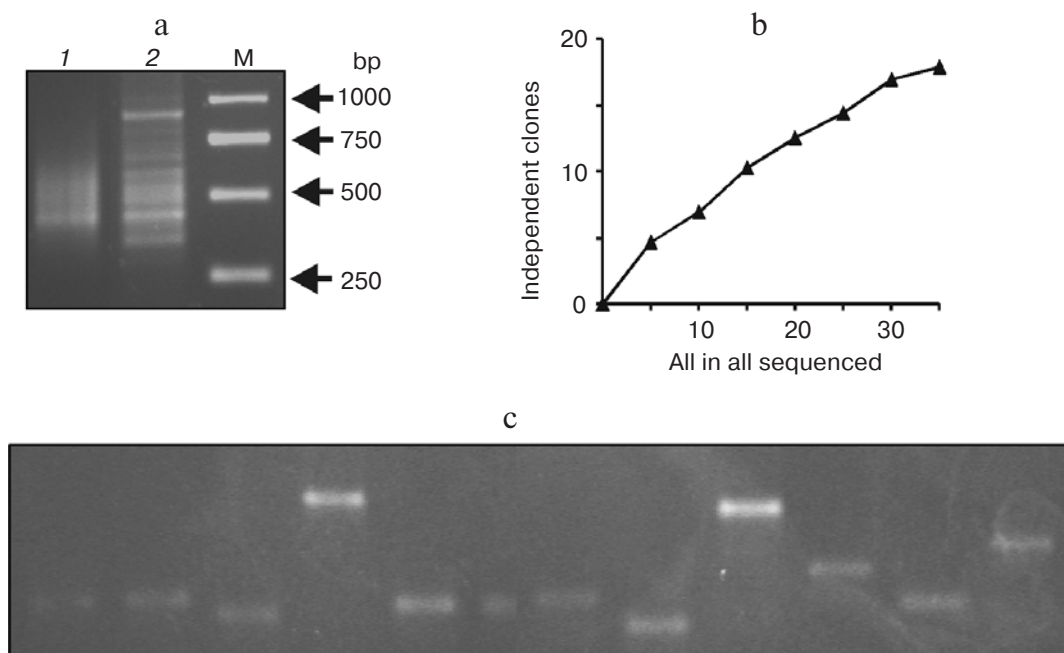


Fig. 2. Insulator selection and cloning. a) Results of insulator selection. PCR amplification with 1L and 1R primers on templates: 1) plasmid pool before transfection; 2) genomic DNA of cells transfected by the short fragment library cloned in the pPNT/EmP vector after positive (G418) and negative (GANC) selections; M, 1 kb DNA marker (Sibenzyme, Russia). b) Estimation of the number of independent clones in the library. The relationship between the number of clones with unique sequences and total number of sequenced clones. c) PCR amplification (with 1L and 1R primers) of randomly chosen clones from the library of potential insulators.

Table 1. Arrangement of potential insulators in the *FXYD5-COX7A1* region of human chromosome 19

Insulator number	Coordinates on chromosome 19*	Length, bp	Position relative to genes
9	40347492-40347685	194	6th intron of gene
10	40406076-40406590	515	~1 kb from 3' of gene <i>TMEM162</i>
11	40437627-40437699	73	2nd intron of gene <i>LSR</i>
12	40909389-40909561	173	14th intron of gene <i>MLL4</i>
13	40932935-40933049	114	1st intron of gene <i>LIN37</i>
14	41067647-41067745	99	~4 kb from 3' of gene <i>NFKBID</i>
15	41116085-41116309	225	~4 kb from 5' of gene <i>LRFN3</i>
16	41147417-41147594	178	intergenic**
17	41350295-41350437	143	—
18	40801090-40801286	197	10th and 11th introns, 11th exon of gene <i>KIAA0841</i>

* According to Human Genome Browser [13], assembly of March, 2006.

** At distance of over 10 kb from nearest genes.

of the neomycin-resistance gene, which will result in the death of such cells at the stage of positive selection. We checked this hypothesis earlier [8].

After negative selection, genomic DNA was isolated from ganciclovir-resistant cells. The region between the CMV enhancer and promoter was PCR amplified on this template using primers 1L and 1R. Amplification products were recloned back into vector pPNT/EmP for further analysis.

Figure 2a shows results of PCR of plasmid pool with primers 1L and 1R before transfection (lane 1) and of PCR with the same primers on the genomic DNA template from cells transfected by the library cloned in plasmid pPNT/EmP and after negative (G418) and positive (GANC) selection (lane 2). The figure shows that after selection the original library containing a set of numerous different length fragments (lane 1) is transformed into a “ladder” (lane 2) of a limited number of fragments. Previously [8], we analyzed a small part of the resulting fragments. In this work, we tried to clone all or the majority of potential insulators of the region under study. To do this, the selected DNA fragments (Fig. 2a, lane 2) were cloned into plasmid pPNT/EmP and the resulting clones were tested for the presence and length of PCR inserts with the same primers (Fig. 2c). The insert was observed in all 12 tested clones, and the insert lengths were different in most clones. We sequenced 20 clones, four of which did not belong to the region under study (they contained fragments of *E. coli* DNA and DNA of different human chromosomes) and obtained 10 new sequences. Then the number of independent (unique) sequences obtained in this and previous work [18] was plotted against the total number of sequenced clones (Fig. 2b). It is seen in the figure that identification of new insulators upon further library sequencing is unlikely. Ten new sequences were obtained and mapped on chromosome 19

between genes *FXYD5* and *COX7A1* (Fig. 1) and their properties are summarized in Table 1.

Arrangement of insulators within the genome. It is supposed that formation of functional domains of the genome is one of main functions of insulators [3-5]. It is important to note that unlike enhancers, activity of insulators can be both dependent on [19-21] and independent [22-24] of their orientation relative to a promoter. Moreover, if there are more than one insulator between enhancer and promoter, in some cases their enhancer-blocking activity is not observed [25-27]. The effect of mutual neutralization of insulators can also depend on their orientation relative to each other [28]. Thus, insulators do not simply subdivide the genome into domains containing genes expressed under similar conditions, but together with genes, promoters, and enhancers they establish a multilevel network of transcription activity.

It should be noted that the term “insulator” is used for designation of both the enhancer-blocking elements and boundary genome elements that separate chromatin regions with different structure or activity. In most cases, insulator elements exhibit both activities, but since sometimes these activities can be separated [29], it can be supposed that different (although neighboring) insulator sequences are responsible for these activities. Since insulators in this work have been identified only by the enhancer-blocking activity, in principle, they might not reveal the “barrier” activity, although it is most likely located near these elements. Nevertheless, in some cases hypothetical domains located between insulators can be identified.

The *FXYD5-COX7A1* region contains a cluster of genes encoding G-protein coupled receptors: *FFAR1*, *FFAR2*, *FFAR3*, and *GPR42P*. It was found earlier that genes *FFAR3* and *FFAR2* have different expression profiles [30]. Location of insulator 6 between genes *FFAR3* and

Table 2. Repetitive genome elements incorporated in potential insulators

Repetitive elements		Insulator number	Insulator overlapping with repetitive elements, %
SINE	AluSx	12	31
	AluSx	15	100
	AluSg	16	100
	AluSg	17	29
LINE	L2	9	98
LTR	ERVL	17	59
	MaLR	10	35
DNA	Tc2	14	100

FFAR2 and insulator 2 separating *FFAR2* from gene *KRT-DAP*, expressed mainly in epidermal keratinocytes [31], correlates with independent expression of these genes. Each of the genes *ATP4A* and *KIAA0841* are located in its own hypothetical domain (between insulators 5 and 1, 1 and 18, respectively). Gene *ATP4A* is transcribed mainly in stomach tissues [32, 33], and in this respect it strongly differs from surrounding genes.

Genes *HCST* and *TYROBP* are located between insulators 14 and 15 and are characterized by highly tissue-specific expression distinct from specificity of surrounding genes. Thus, these genes are expressed at a high level in myeloid cells [34, 35], whereas gene *APLP1*, separated from them by insulator 14, is mainly expressed in cerebral cortex [36], and gene *LRFN3* separated by insulator 15 is expressed in neuronal cells [37]. These data support the concept that sequences exhibiting insulator activity separate the DNA into domains, each containing genes characterized by a specific expression profile [38].

Eight out of 10 identified sequences contain fragments of repetitive genome elements (Table 2). Insulator 9 is a LINE (Long Interspersed Element) fragment, while insulators 10 and 17 are LTR of endogenous human retroviruses MaLR and ERVL, respectively. Other potential insulators are overlapped with Alu elements (Nos. 12, 15-17). Sequence No. 14 is a repetitive element of the Tc2 transposon family.

It is known that some Alu elements can exhibit insulator properties [39, 40]. However, the *FXVD5-COX7A1* region contains over 700 Alu elements, and detection of only two of them among supposed insulators shows that only a few Alu have such activity.

Thus, retrotransposition of Alu and LINE can be an element of rearrangement of the chromatin domain structure [41]. It can be supposed that one of the regulatory functions of retro-elements is their participation in formation of functional domains. We suggested a similar idea previously for retro-elements containing S/MARs [42].

It should be noted that the studied region might contain insulators that were not identified due to limitations

of the approach used. Insulators identified by this approach should be active in CHO cells towards the CMV promoter–enhancer pair, and therefore some tissue- or enhancer-specific insulators could be missed. However, these drawbacks can be overcome by using other cell lines and promoter–enhancer constructs.

The authors are grateful to V. K. Potapov for synthesis of oligonucleotides used in this study.

This work was supported by the Presidium of the Russian Academy of Sciences (program “Molecular and Cell Biology”) and by the program “Leading Scientific Schools of Russian Federation” (NSH-2395.2008.4).

REFERENCES

1. Akopov, S. B., Chernov, I. P., Bulanenkova, S. S., Skvortsova, Yu. V., Vetchinova, A. S., and Nikolaev, L. G. (2007) *Biochemistry (Moscow)*, **72**, 589-594.
2. Akopov, S. B., Chernov, I. P., Vetchinova, A. S., Bulanenkova, S. S., and Nikolaev, L. G. (2007) *Mol. Biol. (Moscow)*, **41**, 717-722.
3. Maximenko, O. G., Chetverina, D. A., and Georgiev, P. G. (2006) *Russ. J. Genet.*, **42**, 845-857.
4. Dorman, E. R., Bushey, A. M., and Corces, V. G. (2007) *Semin. Cell. Dev. Biol.*, **18**, 682-690.
5. Wallace, J. A., and Felsenfeld, G. (2007) *Curr. Opin. Genet. Dev.*, **17**, 400-407.
6. Emery, D. W., Yannaki, E., Tubb, J., Nishino, T., Li, Q., and Stamatoyannopoulos, G. (2002) *Blood*, **100**, 2012-2019.
7. Nagaya, S., Yoshida, K., Kato, K., Akasaka, K., and Shinmyo, A. (2001) *Mol. Genet. Genom.*, **265**, 405-413.
8. Akopov, S. B., Ruda, V. M., Batrak, V. V., Vetchinova, A. S., Chernov, I. P., Nikolaev, L. G., Bode, J., and Sverdlov, E. D. (2006) *Mamm. Genome*, **17**, 1042-1049.
9. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning. A Laboratory Manual*, 3rd Edn., CSHL Press, Cold Spring Harbor.
10. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T., and Mulligan, R. C. (1991) *Cell*, **65**, 1153-1163.

11. Ruda, V. M., Akopov, S. B., Trubetskoy, D. O., Manuylov, N. L., Vetchinova, A. S., Zavalova, L. L., Nikolaev, L. G., and Sverdlov, E. D. (2004) *Virus Res.*, **104**, 11-16.
12. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.*, **25**, 3389-3402.
13. Kuhn, R. M., Karolchik, D., Zweig, A. S., Trumbower, H., Thomas, D. J., Thakkapallayil, A., Sugnet, C. W., Stanke, M., Smith, K. E., Siepel, A., et al. (2007) *Nucleic Acids Res.*, **35**, D668-673.
14. Kellum, R., and Schedl, P. (1992) *Mol. Cell. Biol.*, **12**, 2424-2431.
15. Borrelli, E., Heyman, R., Hsi, M., and Evans, R. M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7572-7576.
16. Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985) *Cell*, **41**, 41-48.
17. Laimins, L., Holmgren-Konig, M., and Khoury, G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3151-3155.
18. Chernov, I., Stukacheva, E., Akopov, S., Didych, D., Nikolaev, L., and Sverdlov, E. (2008) *Biotechniques*, **44**, 775-784.
19. Akasaka, K., Nishimura, A., Takata, K., Mitsunaga, K., Mibuka, F., Ueda, H., Hirose, S., Tsutsui, K., and Shimada, H. (1999) *Cell. Mol. Biol. (Noisy-le-grand)*, **45**, 555-565.
20. Bi, X., and Broach, J. R. (1999) *Genes Dev.*, **13**, 1089-1101.
21. Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A., and Georgiev, P. (2005) *Mol. Cell. Biol.*, **25**, 3682-3689.
22. Belozero, V. E., Majumder, P., Shen, P., and Cai, H. N. (2003) *EMBO J.*, **22**, 3113-3121.
23. Yannaki, E., Tubb, J., Aker, M., Stamatoyannopoulos, G., and Emery, D. W. (2002) *Mol. Ther.*, **5**, 589-598.
24. Zhou, J., Barolo, S., Szymanski, P., and Levine, M. (1996) *Genes Dev.*, **10**, 3195-3201.
25. Kuhn, E. J., Viering, M. M., Rhodes, K. M., and Geyer, P. K. (2003) *EMBO J.*, **22**, 2463-2471.
26. Maksimenko, O., Golovnin, A., and Georgiev, P. (2008) *Mol. Cell. Biol.*, **28**, 5469-5477.
27. Mongelard, F., and Corces, V. G. (2001) *Nat. Struct. Biol.*, **8**, 192-194.
28. Kyrchanova, O., Toshchakov, S., Parshikov, A., and Georgiev, P. (2007) *Mol. Cell. Biol.*, **27**, 3035-3043.
29. Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M., and Felsenfeld, G. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 6883-6888.
30. Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., Kinghorn, I., Fraser, N. J., et al. (2003) *J. Biol. Chem.*, **278**, 11312-11319.
31. Bazzi, H., Fantauzzo, K. A., Richardson, G. D., Jahoda, C. A., and Christiano, A. M. (2007) *Dev. Dyn.*, **236**, 961-970.
32. Herrmann, M., Selige, J., Raffael, S., Sachs, G., Brambilla, A., and Klein, T. (2007) *Scand. J. Gastroenterol.*, **42**, 1275-1288.
33. Chalaya, T. V., Akopov, S. B., Nikolaev, L. G., and Sverdlov, E. D. (2006) *Biochemistry (Moscow)*, **71**, 294-299.
34. Takaki, R., Watson, S. R., and Lanier, L. L. (2006) *Immunol. Rev.*, **214**, 118-129.
35. Wu, J., Song, Y., Bakker, A. B., Bauer, S., Spies, T., Lanier, L. L., and Phillips, J. H. (1999) *Science*, **285**, 730-732.
36. Kim, T. W., Wu, K., Xu, J. L., McAuliffe, G., Tanzi, R. E., Wasco, W., and Black, I. B. (1995) *Brain Res. Mol. Brain Res.*, **32**, 36-44.
37. Morimura, N., Inoue, T., Katayama, K., and Aruga, J. (2006) *Gene*, **380**, 72-83.
38. Bell, A. C., West, A. G., and Felsenfeld, G. (2001) *Science*, **291**, 447-450.
39. Willoughby, D. A., Vilalta, A., and Oshima, R. G. (2000) *J. Biol. Chem.*, **275**, 759-768.
40. Ruiz-Narvaez, E. A., and Campos, H. (2008) *J. Hum. Genet.*, **53**, 247-253.
41. Arnold, R., Maueier, W., Bassili, G., Lutz, M., Burke, L., Epplen, T. J., and Renkawitz, R. (2000) *Gene*, **253**, 209-214.
42. Chernov, I., Akopov, S., and Nikolaev, L. (2004) *Russ. J. Bioorg. Chem.*, **30**, 1-11.