ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 8, pp. 951-957. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © D. A. Didych, N. A. Smirnov, E. S. Kotova, S. B. Akopov, L. G. Nikolaev, E. D. Sverdlov, 2011, published in Biokhimiya, 2011, Vol. 76, No. 8, pp. 1164-1172.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM10-399, July 3, 2011.

Functional Dissection of an Enhancer-Like Element Located within the Second Intron of the Human *U2AF1L4* Gene

D. A. Didych, N. A. Smirnov, E. S. Kotova, S. B. Akopov*, L. G. Nikolaev, and E. D. Sverdlov

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

Received December 29, 2010
Revision received March 21, 2011

Abstract—A detailed functional and evolutionary analysis of an enhancer element of the human genome (enhancer 12) located in the second intron of the *U2AF1L4* gene, which we identified earlier, is presented. Overlapping fragments of the studied genome region were analyzed for enhancer activity, and the site responsible for the activity of this element was identified using transient transfections of HeLa cells. Comparison of the enhancer 12 sequence with orthologous sequences from seven primate species revealed the existence of evolutionarily conserved sequences within this element. One of the identified conservative regions is likely responsible for the enhancer activity and is able to specifically interact *in vitro* with proteins of HeLa cell nuclear extract. The ability of orthologous primate sequences to compete with enhancer 12 for binding with HeLa cell nuclear extract proteins and to enhance the activity of the reporter gene in transient transfection of HeLa cells is demonstrated.

DOI: 10.1134/S0006297911080116

Key words: human genome, regulation of gene expression, functional analysis, enhancer

Enhancers ensure increased expression of target genes, and they are probably the main elements of the genome providing tissue-specific gene expression [1, 2]. Enhancers have modular organization, and they usually contain clusters of binding sites of regulatory proteins involved in nucleosome remodeling, specific post-translational modifications of histone proteins, recruiting mediator protein activators of transcription, as well as in establishing the local spatial organization of chromatin [3]. Enhancer mutations can cause genetic diseases [4-6]. The large-scale search for these elements in the genome and their functional characterization are necessary steps in the preparation of a complete genomic map of regulatory elements indicating their functions, specificity, target genes, spatial interactions, transcription factors that bind to regulatory elements sequence, and mediator proteins.

Despite achievements in the field of large-scale mapping of transcription factors binding sites [7], the findings do not allow to judge about the functions of the sites that have been found. Therefore, an alternative evidence of regulatory element activity is a necessary step for its characterization. The most often used method to analyze enhancer activity is transient expression of the reporter gene, which allows for relatively quick assessment of the ability of a DNA fragment to activate the heterologous promoter.

We identified earlier a number of short (200-500 bp) genomic sequences which possess enhancer activity in HeLa cells [8]. The detected sequences were mapped in 1 megabase polygenic region of human chromosome 19. In present study using transient expression of the reporter gene we carried out the functional analysis of one of the identified enhancer elements (enhancer 12) located in the second intron of the *U2AF1L4* gene. This gene encodes U2AF26 protein, which is a part of the heterodimeric complex that interacts with U2 snRNA and ensures recognition of 3′-splice sites in pre-mRNA by a spliceosome [9, 10]. *U2AF1L4* is expressed in tissue-specific manner [9, 11], but the regulation of its expression is largely unknown. A number of potential transcription factor binding sites – NFκB, NFκB1, CREB, SREBP-1b, SREBP-1a, SREBP-1c, deltaCREB, PPAR-gamma2, PPAR-gamma1, C/EBPalpha – were identified by the ENCODE project in the putative *U2AF1L4* gene promoter [12], but there are no data on enhancer elements that might be involved in regulation of expression of this gene. It is possible that the enhancer element

* To whom correspondence should be addressed.

under discussion is one of the enhancers involved in this regulation. Therefore, we subjected it to a more detailed functional analysis. Deletion analysis of enhancer 12 revealed the region responsible for its enhancer activity. The most conserved regions were found by comparative nucleotide sequence analysis of the enhancer 12-containing region and of orthologous sequences from genomes of seven primate species. The ability of orthologous sequences of primates to compete with enhancer 12 for binding with proteins from HeLa nuclear extract and to enhance the reporter gene activity in transient transfections of HeLa cells was shown.

MATERIALS AND METHODS

Standard protocols. Cultivation and transformation of *E. coli* strains, isolation of plasmid and genomic DNA, agarose gel electrophoresis, and other standard procedures were carried out as described previously [13].

Cell cultures. Cell line HeLa-B (human cervical epithelial carcinoma) was cultured at 37°C and 5% CO_2 in DMEM/F12 medium (1 : 1), containing 10% fetal calf serum.

Vector construction. Full-length sequence of enhancer 12 and its overlapping fragments (A, B, and C) were amplified using human genomic DNA as template with high-precision DNA polymerase PfuII Ultra (Stratagene, USA) under conditions recommended by the manufacturer. The following primer pairs were used in amplification: 12L (AGTTTGGGCTGTGTTCTGTG) and 12R (TTTTCTCATTGGACCTTTCTG), AL (CGGTACAGGTTGAGCAGCA) and AR (TCTGGTTTCTTAGGGGCAAA), BL (GGGCTGCATGAACAGAGAGT) and BR (TCATCTTTGCCAGTCCCTTC), CL (TACACCCCAGCCTAAGCATC) and CR (CGGTCTTCGTGCCGTTTT). Samples of genomic DNA kindly provided by Y. B. Lebedev were used for amplification of primate sequences orthologous to enhancer 12. Amplification products were separated by electrophoresis in 1% agarose gel, and target DNA fragments were eluted using a Wizard SV Gel kit and PCR Clean-Up System (Promega, USA). Purified fragments were cloned into the linearized and pretreated with Klenow fragment pGL3-PV plasmid (Promega) in the *Sal*I site. Plasmids with inserts of the studied fragments were isolated from *E. coli* clones using a Wizard Plus SV Minipreps kit (Promega) and sequenced.

DNA sequencing was performed at "Genome" center (Institute of Molecular Biology, Russian Academy of Sciences) using an ABI PRISM BigDye Terminator v.3.1 reagent kit (Applied Biosystems, USA) followed by analysis of reaction products on an ABI PRISM 3100 Avant automated DNA sequencer (Applied Biosystems).

Transfection of cells and analysis of luciferase activity. Transfection of HeLa cells was performed in 24-well plates using Lipofectamine 2000 reagent (Invitrogen, USA) according to the manufacturer's recommendations. One microgram of experimental plasmid DNA and 75 ng of normalization plasmid pRL-TK (Promega) per well were used for transfection. Forty-eight hours after transfection, cells were lysed in PLB buffer (Promega), and *Photinus pyralis* firefly luciferase activity and *Renilla reniformis* luciferase activity were measured with a GENios Pro luminometer (Tecan, Switzerland) using the Dual-Luciferase Reporter Assay System reagent kit (Promega). All samples of each transfection were presented in two repetitions needed for mean luciferase activity determination. The average firefly luciferase activity was normalized relative to the average values of *R. reniformis* luciferase activity. Three to four independent transfections were carried out for each experiment, and then the standard error of the mean was determined.

Electrophoretic mobility shift assay (EMSA). Enhancer element was labeled with [α-^{32}P]dATP by PCR and purified with gel electrophoresis as described previously [14]. Oligonucleotides (CAGAAAAGGGATGGAGGGGACTGGCAAGG and TCCTTGCCAGTCCCCTCCATCCCTTTTCTG) representing conservative region IV were phosphorylated using T4 phage polynucleotide kinase with the addition of labeled [γ-^{32}P]ATP according to standard methods [13]. To obtain the oligonucleotide duplex, labeled oligonucleotides were hybridized with each other and duplex was purified by gel electrophoresis. Nuclear extract from HeLa cells was prepared according to the procedure described previously [15]. Binding of labeled fragments with HeLa nuclear extract proteins was performed in a 20 μl final volume of the labeled fragment (10,000 cpm), 2 μg of poly(dI-dC), 4 μg of total nuclear extract protein and EMSA buffer (20 mM Hepes-KOH, pH 7.3, 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 8% glycerol, 1 mM 2-mercaptoethanol). Double-stranded oligonucleotides containing binding sites for transcription factors AP2 (GATCGAACTGACCGCCCGCGGCCCGT) and Sp1 (ATTCGATCGGGGCGGGGCGAGC), taken from [16], were used as competitors. Competitors were added to reaction mixture in 10- or 100-fold excess relative to labeled fragment. The reaction mixture was incubated for 20 min at 25°C, and then DNA–protein complexes were separated from free DNA by electrophoresis in 10% polyacrylamide gel in 25 mM Tris-borate buffer.

Bioinformatics analysis. Oligonucleotide primers were designed using the Primer 3 program (http://frodo.wi.mit.edu/primer3/). The sequences were aligned using the ClustalW program (www.ebi.ac.uk/clustalw/). The search for potential transcription factor binding sites was performed with the TESS program (www.cbil.upenn.edu/tess/), which uses databases TRANSFAC, JASPAR, IMD, and CBIL-GibbsMat. In this study we used the resources of the NCBI internet server (http://www.ncbi.

nlm.nih.gov/BLAST) and UCSC Genome Browser (http://genome.ucsc.edu/goldenPath/hgTracks.html).

RESULTS AND DISCUSSION

Analysis of activity of enhancer 12 fragments. Previously, we identified a number of DNA regions in the human genome that exhibit enhancer activity in retroviral construction integrated in the genome of HeLa cells [8]. Here enhancer activity of one of the identified sites (enhancer 12) and its fragments was studied in detail using transient expression of a reporter gene.

Enhancer 12 is located on human chromosome 19 in the second intron of the *U2AF1L4* gene (Fig. 1). According to the UCSC Genome Browser, one can observe clustering of binding sites for proteins such as insulator protein CTCF and the components of the SWI/SNF complex (BAF155, BAF170, and IniI) in this region [12]. In addition, whole-genome analysis of DNase I-sensitive sites, as well as mapping of regulatory elements, based on the identification of nucleosome-free DNA fragments (FAIRE), carried out by the ENCODE project, confirm the presence of DNA–protein complexes in the second intron of the *U2AF1L4* gene [17-22].

We amplified the region containing the fragment under study with a pair of primers, 12L and 12R. The amplified sequence with a length of 364 bp (coordinates according to UCSC Genome Browser, assembly of 2006: chr19:40927383-40927746) was cloned into the pGEM-T vector (Promega) and sequenced.

To determine which of the enhancer 12 regions is responsible for its enhancer activity, we cloned three overlapping fragments of the enhancer and its full-length sequence in pGL3-PV vector, which contains the firefly luciferase gene under the control of SV40 promoter (Fig. 2). Constructs with both orientations of the studied fragments relative to promoter were obtained.

To characterize the activity of enhancer 12 and its fragments, a series of transient transfections of HeLa cells was carried out with a set of plasmids, as well as with the initial vector pGL3-PV used as a control (Fig. 2). The maximal activity was demonstrated by the middle fragment B that increased the SV40 promoter activity twofold, which indicates the presence of the site responsible for enhancer activity in this fragment. Interestingly, fragment C and, therefore, the full-length enhancer contain the binding site of insulator protein CTCF [23, 24]. It is known that in some cases CTCF interacts with silencers, and therefore may reduce the activity of a promoter–enhancer pair [25], so its presence may lead to some reduction in the activity of full-length enhancer and its fragment C. Fragment A does not display enhancer activity in this system, so, taking into account the overlapping of this fragment with fragment B, we can assume that the core enhancer element is located in the 3′-region of fragment B. The observed low level of enhancer activity is possibly linked to the fact that strong SV40 promoter is weakly modulated by cellular

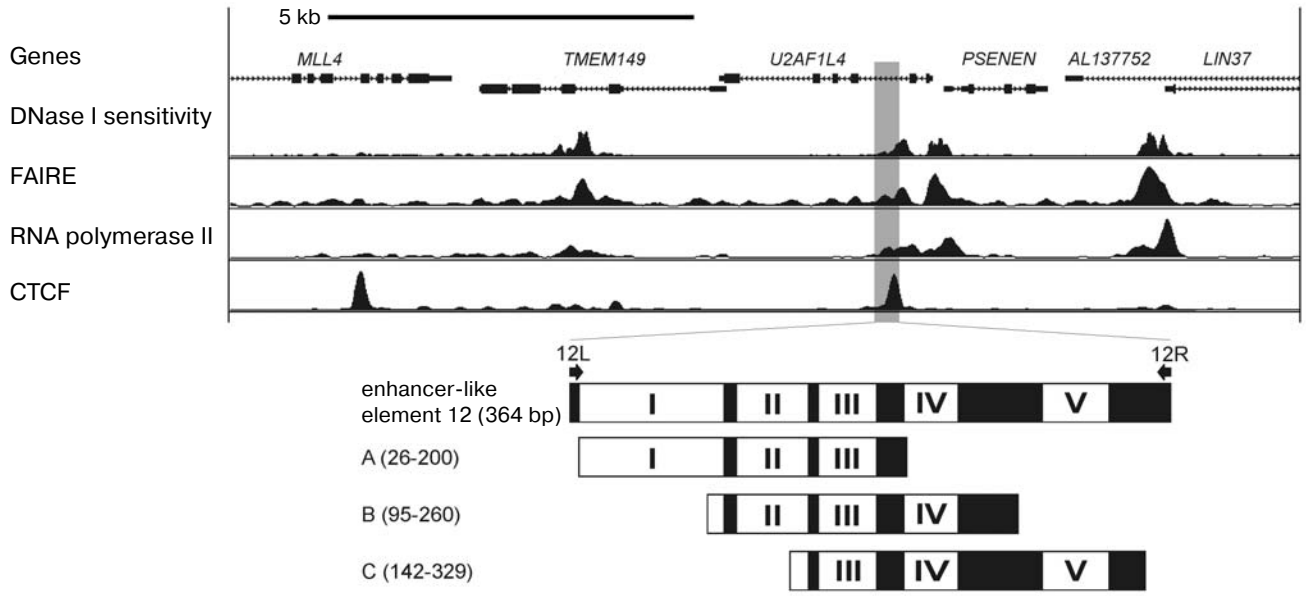

Fig. 1. Location of enhancer 12 relative to neighboring genes on human chromosome 19. Data from the UCSC genome browser (http://genome.ucsc.edu/goldenPath/hgTracks.html): distribution of density of mapped DNase I hypersensitive sites, regulatory sites found by nucleosome-free DNA fragment detection (FAIRE, Formaldehyde Assisted Isolation of Regulatory Elements), binding sites of RNA polymerase II and insulator protein CTCF in the genome of the HeLa cells. The enhancer 12-containing region is marked by gray shading. Full-length enhancer 12 and its overlapping deletion fragments (A, B, and C), used for functional analysis of enhancer activity in HeLa cell line (Fig. 2), are shown below. Roman numerals denote regions within the element that are conserved in primates.

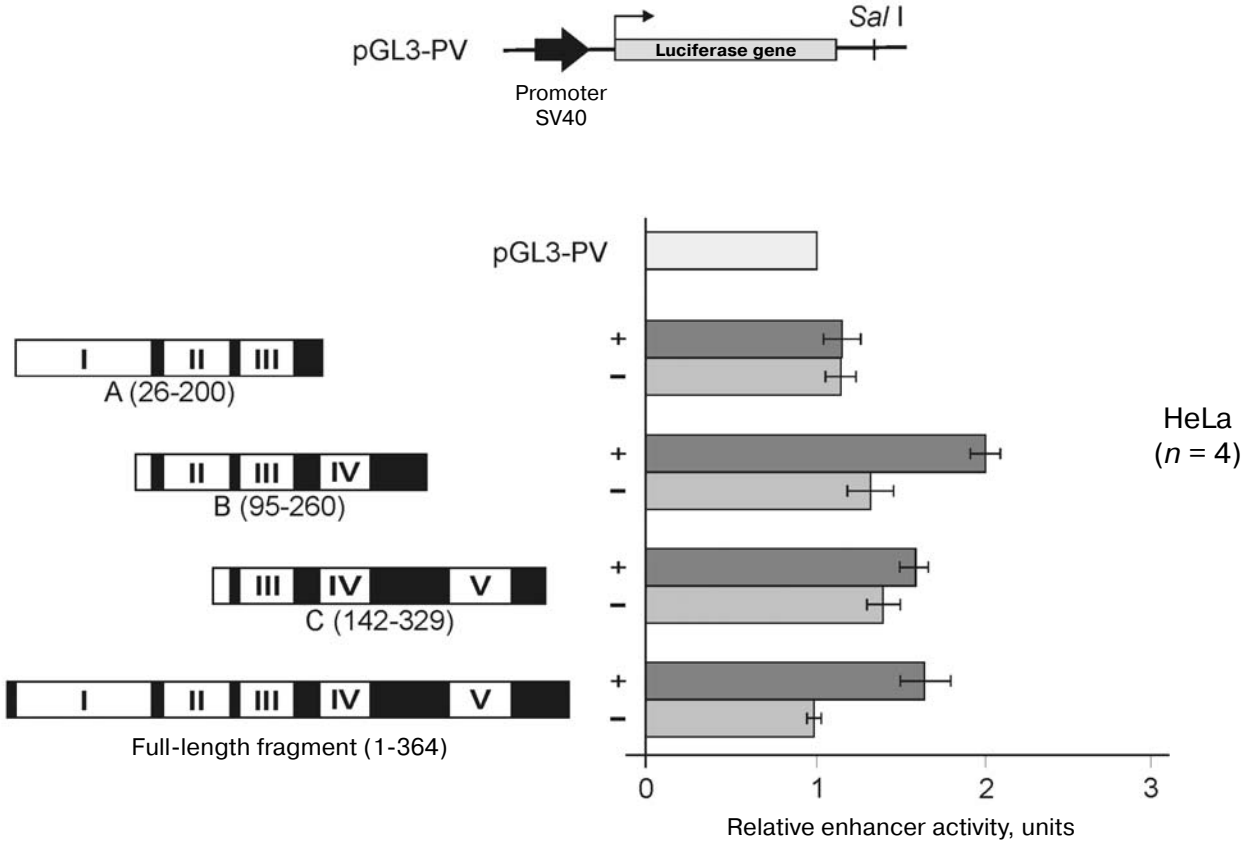

Fig. 2. Enhancer activity determination in HeLa cell line of full-lenght enhancer 12 and its fragments in forward (+) and reverse (–) orientations using the dual luciferase detection system (Promega). The fragments were cloned into pGL3-PV vector (scheme of linearized vector is shown at the top) in the *Sal*I site in both orientations. Activities of fragments are normalized to SV40 promoter activity (pGL3-PV). Standard error of the mean for four independent experiments ($n = 4$) is shown. Roman numerals denote regions that are conserved in primates.

enhancers in HeLa cells and only in the presence of strong enhancers, such as SV40 viral enhancer [26] or enhancers, located in LTR of endogenous retroviruses [27], its activity may greatly increase.

The activity of the sequences under study depends on their orientation relative to the promoter, which is not characteristic of classical enhancers. Full-length enhancer 12 and its fragments B and C were most active when in the forward orientation (Fig. 2).

Enhancer activity of primate sequences orthologous to enhancer 12. We amplified and cloned sequences orthologous to enhancer 12 from the genomes of seven primate species using a universal primer pair (12L and 12R). These sequences represent forms of enhancer caused by natural mutations during evolution. Enhancer activity measurements for three of them (*Hylobates lar*, *Macaca mulatta*, *Saimiri sciureus*) in forward and reverse orientations indicate that mutations only slightly affect the activity of the fragments (Fig. 3a), which reflects the conserved activity of enhancer 12 in human cells.

Binding of orthologous sequences with HeLa nuclear extract proteins. We aligned the human enhancer 12 sequence and the orthologous primate sequences, which revealed the five most conservative regions within enhancer 12 (I-V; Fig. 4). It is known that mutations affecting transcription factor binding sites may interfere with DNA–protein interactions and, consequently, may lead to modification or complete loss of activity of the regulatory element. We tested the ability of primate sequences to interact with proteins of HeLa nuclear extract using electrophoretic mobility shift. Addition of unlabeled sequences from primate genomes to the binding reaction of human enhancer 12 with HeLa nuclear extract proteins weakens the intensity of retarded bands, indicating the ability of the primate sequences to interact with the same proteins as does enhancer 12 (Fig. 3b).

Therefore, nucleotide substitutions, deletions, and insertions in the studied sequences of primates do not affect the DNA–protein interactions *in vitro*. The conservative regions of enhancer 12 and the regions with the fewest number of interspecies differences can be assumed to be responsible for binding with HeLa nuclear extract proteins and, probably, for the enhancer activity manifestation of the studied region of the human genome.

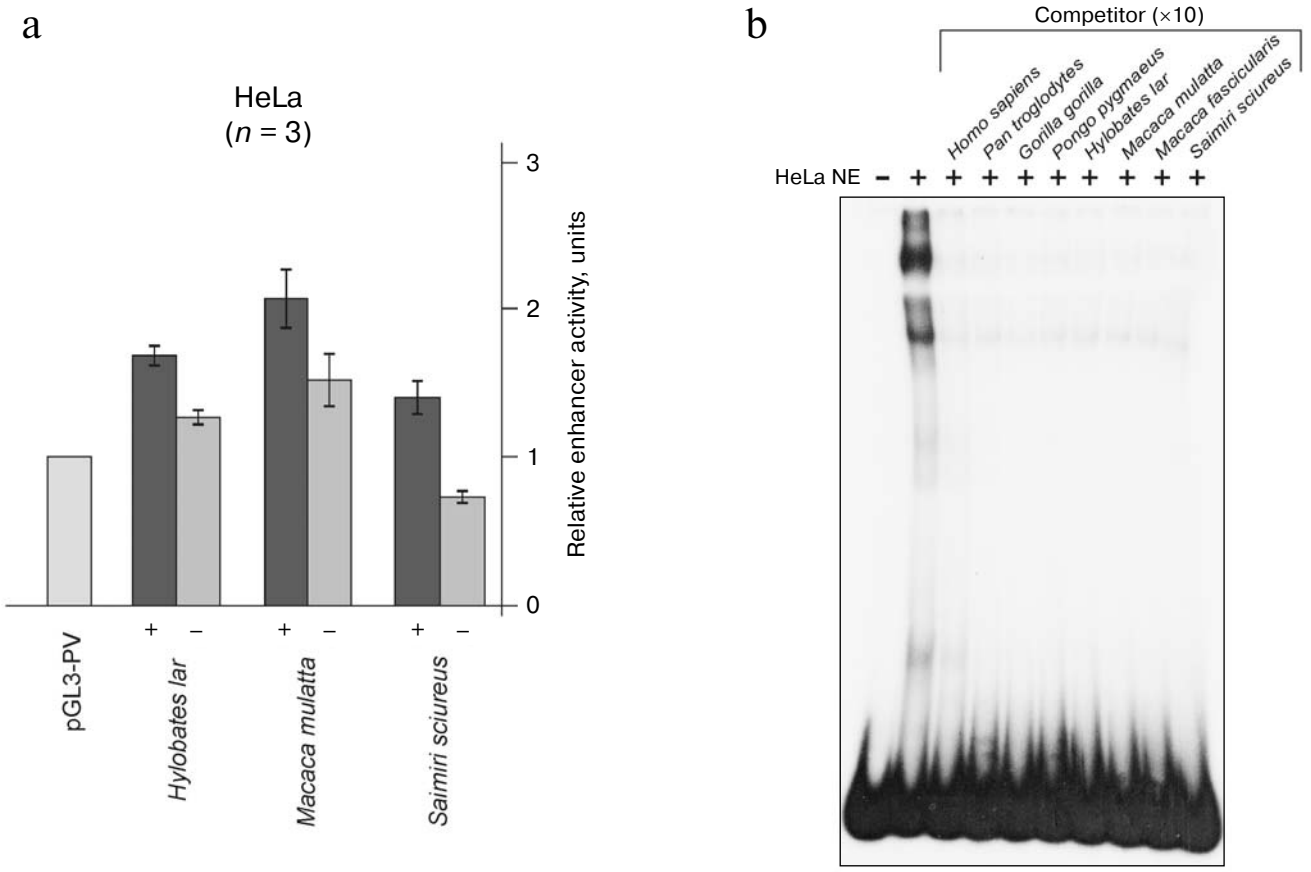

Fig. 3. Analysis of functional properties of primate sequences orthologous to enhancer 12. a) Enhancer activity of sequences of gibbon (*Hylobates lar*), rhesus macaque (*Macaca mulatta*), and squirrel monkey (*Saimiri sciureus*) in forward (+) and reverse (–) orientations in HeLa cells using a dual luciferase detection system. Activities are normalized to SV40 promoter activity (pGL3-PV). Standard error of the mean of three independent experiments ($n = 3$) is shown. b) EMSA analysis of the ability of primate sequences orthologous to enhancer 12 to compete with enhancer 12 for binding HeLa nuclear extract proteins (HeLa NE). Competing sequences were added to the reaction mixture in 10-fold excess relative to labeled fragment.

As seen in Fig. 2, fragment B, in contrast to fragment A, which exhibits no enhancer activity, contains 30 bp conservative region IV located in 3′-region of the fragment. We hypothesized that this particular region is responsible for enhancer activity. We identified the hypothetical binding sites of known transcription factors with the TESS program using the TRANSFAC, JASPAR, IMD, and CBIL-GibbsMat databases. The fragment was found to contain binding sites for TCF-1 transcription factor family members (T-cell factor 7) and lymphocyte enhancer factor LEF-1. However, this site has a mononucleotide substitution in four of the studied primate species and a dinucleotide substitution in two species, which may affect DNA–protein interactions. At the same time, the binding site of transcription activator Sp1 CCCCTCC has no interspecies differences. We tested the ability of conserved region IV to interact with factors Sp1 and AP2 using electrophoretic mobility shift with nuclear extract of HeLa cells. A double-stranded oligonucleotide containing the region IV sequence was used as a probe, and oligonucleotides with consensus binding site of Sp1 factor (GC-box) and transcription activator AP-2 were used as competitors (Fig. 5a). Addition of the competitors to the reaction did not reduce the intensity of the inhibition zone (Fig. 5b), which indicates that the conserved region IV does not interact with these factors *in vitro*. However, addition of unlabeled conserved region IV to the reaction mix in 100-fold excess leads to disappearance of the retarded band (Fig. 5b).

It is worth noting that sequences similar to the CCC-CTCC motif can interact besides Sp1 with other factors with different activities, such as c-Krox, ZBP-89, SMYD3, MAZ, CTCF, indicating the possible influence of specific genomic environment, which might determine which of the factors interact with this motif.

Identification of the short sequence responsible for the enhancer 12 functional activity will allow identification of the transcription factors that bind to it using DNA-affinity chromatography as well as determination,

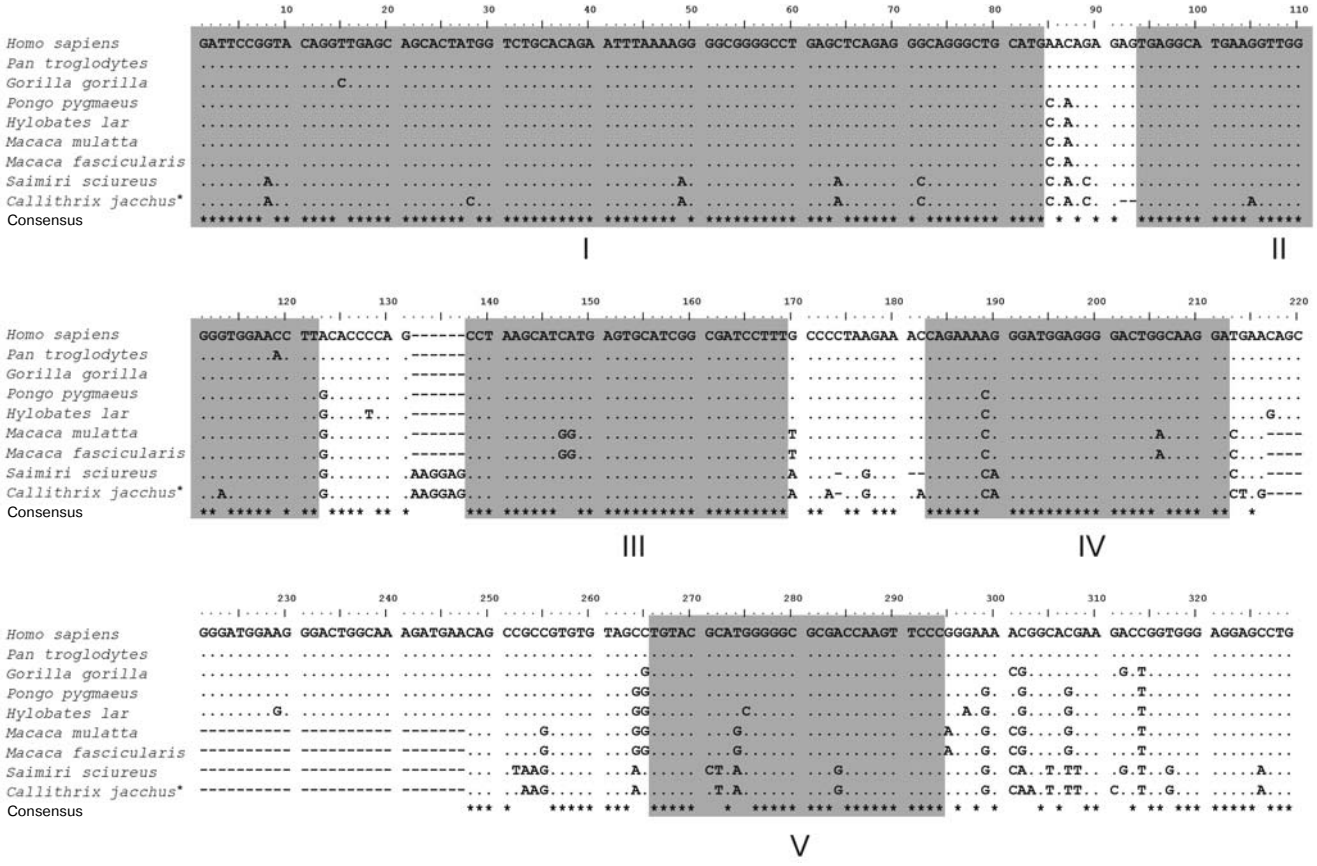

Fig. 4. Alignment of enhancer 12 sequence and orthologous sequences from genomes of eight primate species using the ClustalW program. Conserved regions with the least number of interspecies differences are marked in gray and with Roman numerals (I-V). The *Callithrix jacchus* sequence was obtained from the UCSC Genome Browser database.

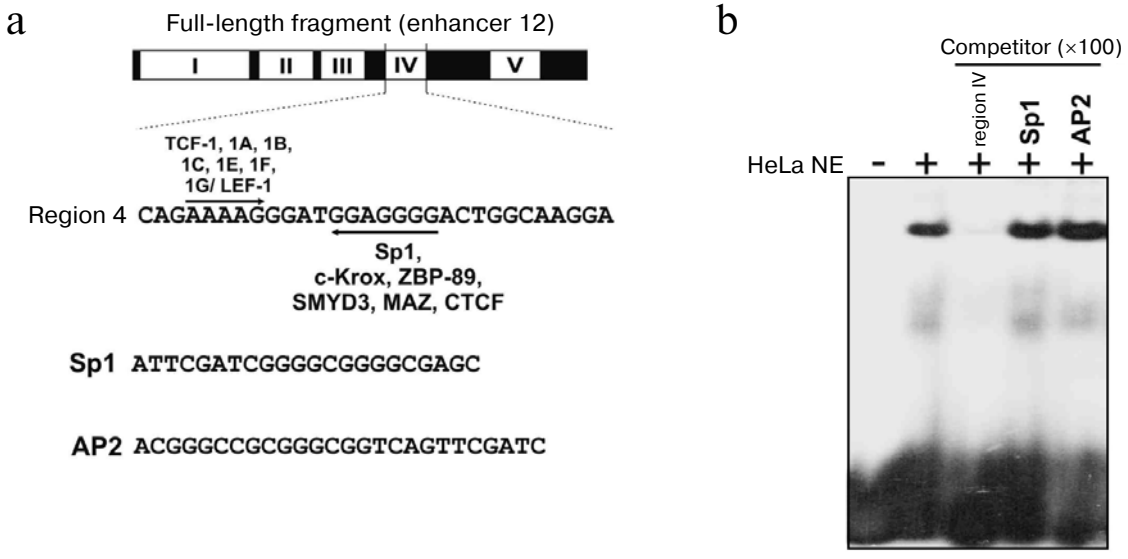

Fig. 5. The *in vitro* interaction of enhancer 12 conserved region IV with HeLa nuclear extracts. a) Conserved region IV sequence and its consensus binding motifs of different transcription factors. The sequences of oligonucleotide competitors containing binding sites for factors Sp1 and AP2 are shown below. b) EMSA analysis of conserved region IV binding with proteins from HeLa nuclear extract. Competing sequences containing binding sites for factors Sp1 and AP-2 and unlabeled conserved region IV were added to the reaction mixture in hundredfold excess relative to the labeled fragment.

using the 3C method, the gene whose promoter is activated by this enhancer.

The authors thank Yu. B. Lebedev for provided samples of primate genomic DNA as well as the ENCODE consortium for providing free access to the project data.

This work was financially supported by the Program of the Russian Academy of Sciences Presidium "Molecular and Cellular Biology" and the program "Leading Scientific Schools of the Russian Federation" (NSh-5638.2010.4).

REFERENCES

1. Maston, G. A., Evans, S. K., and Green, M. R. (2006) *Annu. Rev. Genomics Hum. Genet.*, **7**, 29-59.
2. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., et al. (2009) *Nature*, **459**, 108-112.
3. Atchison, M. L. (1988) *Annu. Rev. Cell Biol.*, **4**, 127-153.
4. De Kok, Y. J., Vossenaar, E. R., Cremers, C. W., Dahl, N., Laporte, J., Hu, L. J., Lacombe, D., Fischel-Ghodsian, N., Friedman, R. A., Parnes, L. S., et al. (1996) *Hum. Mol. Genet.*, **5**, 1229-1235.
5. Loots, G. G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N. M., Ovcharenko, D., Plajzer-Frick, I., and Rubin, E. M. (2005) *Genome Res.*, **15**, 928-935.
6. Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. (2003) *Hum. Mol. Genet.*, **12**, 1725-1735.
7. Heintzman, N. D., and Ren, B. (2009) *Curr. Opin. Genet. Dev.*, **19**, 541-549.
8. Chernov, I., Stukacheva, E., Akopov, S., Didych, D., Nikolaev, L., and Sverdlov, E. (2008) *Biotechniques*, **44**, 775-784.
9. Shepard, J., Reick, M., Olson, S., and Graveley, B. R. (2002) *Mol. Cell. Biol.*, **22**, 221-230.
10. Heyd, F., Carmo-Fonseca, M., and Moroy, T. (2008) *J. Biol. Chem.*, **283**, 19636-19645.
11. Chen, F., Ji, C., Dou, T., Zheng, N., Qiu, R., Peng, J., Fang, W., Feng, C., Xie, Y., and Mao, Y. (2006) *DNA Seq.*, **17**, 282-286.
12. Celniker, S. E., Dillon, L. A., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., et al. (2009) *Nature*, **459**, 927-930.
13. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
14. Nikolaev, L. G., Tsevegiyn, T., Akopov, S. B., Ashworth, L. K., and Sverdlov, E. D. (1996) *Nucleic Acids Res.*, **24**, 1330-1336.
15. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.*, **11**, 1475-1489.
16. Butta, N., Larrucea, S., Alonso, S., Rodriguez, R. B., Arias-Salgado, E. G., Ayuso, M. S., Gonzalez-Manchon, C., and Parrilla, R. (2006) *BMC Mol. Biol.*, **7**, 17.
17. Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006) *Nat. Methods*, **3**, 511-518.
18. Crawford, G. E., Davis, S., Scacheri, P. C., Renaud, G., Halawi, M. J., Erdos, M. R., Green, R., Meltzer, P. S., Wolfsberg, T. G., and Collins, F. S. (2006) *Nat. Methods*, **3**, 503-509.
19. Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., and Crawford, G. E. (2008) *Cell*, **132**, 311-322.
20. Crawford, G. E., Holt, I. E., Whittle, J., Webb, B. D., Tai, D., Davis, S., Margulies, E. H., Chen, Y., Bernat, J. A., Ginsburg, D., et al. (2006) *Genome Res.*, **16**, 123-131.
21. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007) *Genome Res.*, **17**, 877-885.
22. Giresi, P. G., and Lieb, J. D. (2009) *Methods*, **48**, 233-239.
23. Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., and Ren, B. (2007) *Cell*, **128**, 1231-1245.
24. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) *Cell*, **129**, 823-837.
25. Klochkov, D., Rincon-Arano, H., Ioudinkova, E. S., Valadez-Graham, V., Gavrilov, A., Recillas-Targa, F., and Razin, S. V. (2006) *Mol. Cell. Biol.*, **26**, 1589-1597.
26. 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.
27. Abrarova, N. D., Stukacheva, E. A., Pleshkan, V. V., Vinogradova, T. V., and Sverdlov, E. D. (2010) *Mol. Biol. (Moscow)*, **44**, 627-634.