Вавиловский журнал генетики и селекции. 2023;27(7):906-915
DOI 10.18699/VJGB-23-105

Original Russian text https://vavilovj-icg.ru/

Analysis of the transcriptional activity of model piggyBac transgenes stably integrated into different loci of the genome of CHO cells in the absence of selection pressure

L.A. Yarinich, A.A. Ogienko, A.V. Pindyurin, E.S. Omelina

Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
omelina@mcb.nsc.ru

Abstract. CHO cells are most commonly used for the synthesis of recombinant proteins in biopharmaceutical production. When stable producer cell lines are obtained, the locus of transgene integration into the genome has a great influence on the level of its expression. Therefore, the identification of genomic loci ensuring a high level of protein production is very important. Here, we used the TRIP assay to study the influence of the local chromatin environment on the activity of transgenes in CHO cells. For this purpose, reporter constructs encoding eGFP under the control of four promoters were stably integrated into the genome of CHO cells using the piggyBac transposon. Each individual transgene contained a unique tag, a DNA barcode, and the resulting polyclonal cell population was cultured for almost a month without any selection. Next, using the high-throughput sequencing, genomic localizations of barcodes, as well as their abundances in the population and transcriptional activities were identified. In total, ~640 transgenes more or less evenly distributed across all chromosomes of CHO cells were characterized. More than half of the transgenes were completely silent. The most active transgenes were identified to be inserted in gene promoters and 5′ UTRs. Transgenes carrying Chinese hamster full-length promoter of the *EF-1α* gene showed the highest activity. Transgenes with a truncated version of the same promoter and with the mouse *PGK* gene promoter were on average 10 and 19 times less active, respectively. In total, combinations of genomic loci of CHO cells and transgene promoters that together provide different levels of transcriptional activity of the model reporter construct were described.
Key words: TRIP; barcode; chromatin position effect; transgene; chromatin; transcription.

For citation: Yarinich L.A., Ogienko A.A., Pindyurin A.V., Omelina E.S. Analysis of the transcriptional activity of model piggyBac transgenes stably integrated into different loci of the genome of CHO cells in the absence of selection pressure. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):906-915. DOI 10.18699/VJGB-23-105

Анализ транскрипционной активности модельных piggyBac-трансгенов, стабильно интегрированных в разные локусы генома культивируемых клеток CHO при отсутствии селекционного давления

Л.А. Яринич, А.А. Огиенко, А.В. Пиндюрин, Е.С. Омелина

Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
omelina@mcb.nsc.ru

Аннотация. Культивируемые клетки яичника китайского хомячка (CHO) наиболее часто используются для синтеза рекомбинантных белков в биофармацевтическом производстве. При получении стабильных клеточных линий-продуцентов локус интеграции трансгена в геном оказывает большое влияние на уровень его экспрессии (явление, известное как эффект положения гена). Соответственно, поиск локусов генома, обеспечивающих высокий уровень продукции белков, является актуальной практической задачей. В данной работе мы использовали метод TRIP для исследования влияния локального окружения хроматина на активность трансгенов, встроенных в разные локусы генома культивируемых клеток CHO. С этой целью репортерные конструкции, кодирующие белок eGFP под контролем четырех разных промоторов, были стабильно встроены в геном клеток CHO при помощи транспозона piggyBac. При этом каждый отдельный трансген содержал уникальную метку – ДНК-штрихкод. Полученная трансгенная поликлональная популяция клеток была культивирована в течение месяца без какой-либо селекции. Далее при помощи присутствующих в конструкциях штрихкодов и высокопроизводительного секвенирования были определены сайты локализации трансгенов в геноме, из-

мерена их представленность в популяции, а также транскрипционная активность. Всего удалось полностью охарактеризовать около 640 трансгенов, более-менее равномерно распределенных по всем хромосомам клеток CHO. Более половины трансгенов оказались полностью молчащими. Наиболее активные трансгены выявлены в окрестностях геномных сайтов инициации транскрипции – в промоторных и 5′-некодирующих районах генов. Наибольшей активностью обладали трансгены, несущие полноразмерный промотор гена *EF-1α* китайского хомячка. Трансгены с укороченным вариантом этого же промотора, а также трансгены с промотором мышиного гена *PGK* (m*PGK*) были соответственно в среднем в 10 и 19 раз менее активны. В целом в результате данной работы выявлены сочетания локусов генома культивируемых клеток CHO и промоторных элементов, которые обеспечивают разные уровни транскрипционной активности модельной репортерной конструкции.
Ключевые слова: TRIP; штрихкод; эффект положения гена; трансген; хроматин; транскрипция.

Introduction

The TRIP (thousands of reporters integrated in parallel) assay enables large-scale studies of the influence of the chromatin position effects on the transgene activity. It is based on DNA barcodes (hereafter, barcodes) and was originally performed on the mouse embryonic stem cells using the piggyBac transposon system to deliver reporter transgenes into the genome (Akhtar et al., 2013). A barcode is a short DNA sequence (16–20 bp) that is unique to each individual transgene in the study. It should be noticed that the barcode is located within the transcribed region of the transgenes ensuring its presence not only in DNA, but also in mRNA molecules. Therefore, the barcode can be used for quantitative measurements of the level of transcriptional activity of transgenes.

The piggyBac transposon system has been previously used to effectively modify various cell lines and organisms (Wilson et al., 2007) even with large transgenic constructs (Ding et al., 2005). In addition, the piggyBac transposon is characterized by a relatively uniform distribution of insertions across chromosomes (Huang et al., 2010). In TRIP experiments, the system for transgenesis consists of two plasmid constructs: a construct for the expression of piggyBac transposase, which catalyzes the insertion of a transgene into a random genome locus, and the transgene itself – a target construct (consisting of a promoter, a reporter gene, a barcode, and a polyadenylation signal) located between the inverted terminal repeats of the piggyBac transposon (Akhtar et al., 2014; Lebedev et al., 2019). Co-transfection of cells with such plasmid constructs enables obtaining a polyclonal population of transgenic cells, in which each individual transgene insertion in the genome is marked with a unique barcode sequence. After cultivation of transfected cells, genomic DNA and total RNA are isolated from them. Using the genomic DNA sample, the genomic localization of each transgene is identified and the abundance of each barcode in the cell population is measured. Based on the total RNA sample, the abundance of each barcode in the total pool of transcripts synthesized from transgenes is measured. Finally, the ratio of the abundance of each barcode in mRNA molecules to its abundance in the cell population allows quantitative estimation of transcriptional activity of all transgenes (Akhtar et al., 2014).

In this study, we used the TRIP assay to investigate the chromatin position effects on the transcriptional activity of stably integrated transgenes in the Chinese hamster ovary cells (CHO). The CHO line is the most commonly used cell line to produce a variety of proteins (Xu et al., 2023). Despite the availability of other mammalian cells, such as baby hamster kidney cells, murine myeloma NS0 cells, human embryonic kidney cells (HEK293), human embryonic retinal PerC6 cells, more than 70 % of all recombinant therapeutic proteins are produced in CHO cells (Kim et al., 2012; Ritacco et al., 2018; Gupta et al., 2021). The popularity of CHO cells is explained by the following reasons. First, the use of CHO cells for the production of recombinant proteins is safe, since CHO cells are insensitive to infection by human viruses (Lalonde, Durocher, 2017). Second, CHO cells have the ability for efficient post-translational modification and produce recombinant proteins in human-compatible glycoforms (Stache et al., 2019). Third, CHO cells have a high growth rate and are relatively easily adapted to growth in suspension, which is a preferred characteristic for large-scale cultivation in bioreactors (Ritacco et al., 2018; Dahodwala, Lee, 2019). Currently, bioreactors with a volume of more than 10 thousand liters are used for suspension cultures of recombinant CHO cells producing therapeutic antibodies (Kim et al., 2012).

Localization in the genome has a great influence on the expression level of the recombinant gene (a phenomenon known as the chromatin position effect) (Gierman et al., 2007; Babenko et al., 2010; Ruf et al., 2011; Chen M. et al., 2013; Elgin, Reuter, 2013). Integration into inactive heterochromatin results in low or no transgene expression, whereas integration into active euchromatin often allows moderate to high transgene expression. However, simple integration into euchromatin may not be sufficient to ensure long-term expression of the recombinant gene. The phenomenon of transgene expression silencing is well known in mammalian cells, it occurs in part likely due to the influence of adjacent condensed chromatin.

Thus, integration of transgenes into transcriptionally highly active regions of the genome is a reasonable strategy to avoid position effects. This study was aimed at analyzing the transcriptional activity of the piggyBac transgenes integrated into different loci of the genome of CHO cells in the absence of selection pressure.

Materials and methods

Generation of the pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242 construct. Plasmid pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242 was made based on a previously described "universal" construct (Lebedev et al., 2019). The insertion was amplified using the primers mPGK-EcoRI-F and eGFP-XbaI-R (Table 1) and the plasmid pPTK-Gal4-mPGK-Puro-IRES-eGFP-sNRP-pA as a template (Akhtar et al., 2013). Fifty μl of the reaction mixture contained 1 ng of plasmid template, 10 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 1 μl of each 10 μM primer, 0.2 mM dNTPs, and 2.5 U of Phusion polymerase (Thermo Fisher Scientific). The PCR

Table 1. Primers used in the study

Primer name	Primer sequence (5′→3′)
mPGK-EcoRI-F	aaagaattctcgacaattctaccgggtagg
eGFP-XbaI-R	aaatctagaccctccggattacttg
hamPgk1-EcoRI-F	aaagaattcaggtccctggggattcca
hamPgk1-BglII-R	aaaagatctcggtaggatcaagaggctcag
CHEF-1-v1-EcoRI-F	aaagaattccacgttgtgcatagaaacagatgc
CHEF-1-v1-BclI-R	aaatgatcatggttttcacaacaccttaaaaaaaaagttcg
CHEF-1-v2-EcoRI-F	aaagaattcaagcttctgtggatagaaaatgattag
CHEF-1-v2-BclI-R	aaatgatcactgcgttctgacggcaaac
Plasmid-1	ccgcttaattaatccagcttttgttc
pPB-eGFP-PI-6-R	ctcgagctctcgatctctagacc
pPB-eGFP-PI-11-R	ctcactagctcgatctctagacc
pPB-eGFP-PI-16-R	ctcttgtactcgatctctagacc
pPB-eGFP-PI-28-R	ctcctcggctcgatctctagacc
PB-Barcode-PI-6-Gibson-F	gtctagagatcgagagctcgaggN_{18}gagttgtggccggcccttgtg
PB-Barcode-PI-11-Gibson-F	gtctagagatcgagctagtgaggN$_{18}$gagttgtggccggcccttgtg
PB-Barcode-PI-16-Gibson-F	gtctagagatcgagtacaagaggN$_{18}$gagttgtggccggcccttgtg
PB-Barcode-PI-28-Gibson-F	gtctagagatcgagccgaggaggN$_{18}$gagttgtggccggcccttgtg
PB-Gibson-R1	aacaaaagctggattaattaagcggccgcatacgcgtatactagattaaccc
Libr-cDNA-for	gtctcgtgggctcggagatgtgtataagagacaggtcctgctggagttcgtgac
Libr-cDNA-A16-rev	tcgtcggcagcgtcagatgtgtataagagacagcctatggtcgccagggttttcccagtcacaagg
Libr-cDNA-A23-rev	tcgtcggcagcgtcagatgtgtataagagacagtaattgcgcgccagggttttcccagtcacaagg
Libr-P5-for	aatgatacggcgaccaccgagatctacactcgtcggcagcgtc
Libr-P7-rev	caagcagaagacggcatacgagatgtctcgtgggctcgg
PB-outer-F-2	ttttacgcatgattatctttaacgtacgtc
cDNA-ampl-R	cgccagggttttcccagtcacaag
PB-cDNA-fwd-A7	tcgtcggcagcgtcagatgtgtataagagacagagcgagctgtcacaagggccggccacaa
InvPCR-F-Nextera2	gtctcgtgggctcggagatgtgtataagagacaggtacgtcacaatatgattatctttctag

Note. N_{18} – random 18-nt barcode sequence.

thermal cycle conditions were as follows: 98 °C for 30 sec, 35 cycles of 98 °C for 10 sec, 62 °C for 10 sec, 72 °C for 1 min, and a final incubation for 10 min at 72 °C.

Cloning of constructs with various Chinese hamster gene promoters. The plasmid pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242 was digested with EcoRI, BglII, and AgeI restriction enzymes. To obtain inserts, the sequences of the Chinese hamster *PGK* gene promoter and the long and short variants of the *EF-1α* gene promoter were amplified using the primers hamPgk1-EcoRI-F and hamPgk1-BglII-R, CHEF-1-v1-EcoRI-F and CHEF-1-v1-BclI-R, and CHEF-1-v2-EcoRI-F and CHEF-1-v2-BclI-R (see Table 1), respectively. Fifty μl of the reaction mixture contained 50 ng of genomic DNA template isolated from CHO cells, 10 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 1 μl of each 10 μM primer, 0.2 mM dNTPs, and 2.5 U of Phusion polymerase (Thermo Fisher Scientific). The PCR thermal cycle conditions were as follows: 98 °C for 30 sec, 35 cycles of 98 °C for 10 sec, 62 °C for 10 sec, 72 °C for 1 min, and a final incubation for 10 min at 72 °C.

Generation of barcoded plasmid libraries. Barcoded plasmid libraries were made according to a previously described protocol (Lebedev et al., 2019) using the Gibson cloning method. For this purpose, vectors and inserts containing an

18-nt DNA barcode and a promoter index were prepared using PCR. For vector amplification, the primers Plasmid-1 and pPB-eGFP-PI-6-R / pPB-eGFP-PI-11-R / pPB-eGFP-PI-16-R / pPB-eGFP-PI-28-R were used (see Table 1) for the constructs with the Chinese hamster *PGK* gene promoter / m*PGK* gene promoter / short variant of the *EF-1α* gene promoter / long variant of the *EF-1α* gene promoter, respectively. The primers PB-Barcode-PI-6-Gibson-F / PB-Barcode-PI-11-Gibson-F / PB-Barcode-PI-16-Gibson-F / PB-Barcode-PI-28-Gibson-F and PB-Gibson-R1 (see Table 1) were used to amplify the barcoded inserts for constructs with the Chinese hamster *PGK* gene promoter / the *mPGK* gene promoter / short variant of the *EF-1α* gene promoter / long variant of the *EF-1α* gene promoter, respectively. Fifty μl of the reaction mixture contained 1 ng of template, 10 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 1 μl of each 10 μM primer, 0.2 mM dNTPs, and 2.5 U of Phusion polymerase (Thermo Fisher Scientific). The PCR thermal cycle conditions were as follows: 98 °C for 30 sec, 35 cycles of 98 °C for 10 sec, 62 °C for 10 sec, 72 °C for 1 min, and a final incubation for 10 min at 72 °C. After purification, 200 ng of "vector" and 135 ng of "inserts" were mixed with 10 μl of 2× NEBuilder HiFi DNA Assembly Master in a total volume of 20 μl. DNA ligation and bacterial transformation were performed as described previously (Lebedev et al., 2019). Barcoded plasmid libraries were isolated using the Mega Plasmid Kit (Qiagen).

Generation of polyclonal transgenic population of CHO cells. Twenty-four h before transfection, CHO-S cells (hereafter CHO cells; kindly provided by Dr. A.V. Taranin, Institute of Molecular and Cellular Biology, Novosibirsk, Russia) were seeded into a 12-well culture plate at a concentration of 1.5×10^5 cells per ml in IMDM medium supplemented with 10 % bovine serum. Cells were co-transfected with a mixture of barcoded plasmid libraries (3 μg) and the pRP[Exp]-mCherry-CAG>hyPBase plasmid (VectorBuilder #VB160216-10057; kindly provided by Prof. V.V. Verkhusha, Albert Einstein College of Medicine, Bronx, NY, USA) (0.3 μg) using the X-tremeGENE HP DNA transfection reagent (Roche). The transfected cells were cultured for a month in the absence of selection pressure.

Isolation of genomic DNA. Genomic DNA was isolated from $5 \cdot 10^7$ cells of the resulting polyclonal transgenic population using the PureLink® Genomic DNA Kit (Invitrogen) according to the manufacturer's recommendations.

Isolation of total RNA, reverse transcription. Total RNA was isolated from $5 \cdot 10^7$ cells of the resulting polyclonal transgenic population using RNAzol RT (Molecular Research Center) according to the manufacturer's recommendations. The isolated RNA was incubated with 20 U of DpnI restriction endonuclease (New England Biolabs) and 3 U of DNase I (Thermo Fisher Scientific) for 30 min at 37 °C. The Clean RNA Standard kit (Evrogen) was used to purify RNA. Two μg of purified total RNA was mixed with 1 μl of the 50 mM oligo(dT) primer in a total volume of 13.5 μl, and the mixture was incubated for 5 min at 65 °C. The subsequent reverse transcription reaction was carried out in a volume of 20 μl with the following components: 13.5 μl of RNA with annealed primer, 4 μl of 5× RT buffer (Thermo Fisher Scientific), 1 μl of 10 mM dNTPs, 1 μl of RNaseOUT (Thermo Fisher Scientific), 100 U of RevertAid reverse transcriptase (Thermo Fisher Scientific). The mixture was incubated for 60 min at 42 °C, and the enzyme was inactivated for 10 min at 70 °C.

Preparation of the normalization and expression samples. To prepare each sample, two rounds of PCR were performed. For the first round of amplification, we used 600 ng of genomic DNA template (for the normalization sample) or 3 μl of cDNA (for the expression sample), 0.5 μl of 10 μM primers Libr-cDNA-for and Libr-cDNA-A16-rev / Libr-cDNA-A23-rev (see Table 1) for the normalization / expression sample, respectively, 5 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 2 μl of 2.5 mM dNTPs, and 1.25 U of Phusion HS II DNA polymerase (Thermo Fisher Scientific) in a total volume of 25 μl. The thermal cycle conditions of the first round of PCR were as follows: 98 °C for 1 min, 15 cycles of 98 °C for 30 sec, 70 °C for 30 sec, 72 °C for 30 sec, and a final incubation for 5 min at 72 °C. The second round of amplification was carried out in a volume of 25 μl with the following components: 0.5 μl of the PCR products of the first round, 0.25 μl of 10 μM primers Libr-P5-for and Libr-P7-rev (see Table 1), 5 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 2 μl of 2.5 mM dNTPs, and 1.25 U of Phusion HotStart II DNA polymerase (Thermo Fisher Scientific). The thermal cycle conditions of the second round of PCR were as follows: 98 °C for 1 min, 23 cycles of 98 °C for 30 sec, 61 °C for 30 sec, 72 °C for 30 sec, and a final incubation for 5 min at 72 °C.

Preparation of the mapping sample. Two μg of genomic DNA was incubated with 10 U of DpnII restriction endonuclease (New England Biolabs) at 37 °C for 16 h, then purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific). 600 ng of the digested genomic DNA was mixed with 4 μl of 100 mM ATP, 2.5 U of T4 DNA ligase (Evrogen) in a total volume of 400 μl. The ligase mixture was incubated for 2 h at room temperature and then for 16 h at 4 °C, followed by enzyme inactivation at 65 °C for 10 min. 100 μl of double-distilled water and 500 μl of a phenol:chloroform solution (1:1 ratio) were added to the ligation reaction, mixed, centrifuged at room temperature for 5 min at 10,000×*g*, and the upper phase was transferred to a new tube. One-tenth volume of 3M NaOAc (pH 5.5) and 2.5 volumes of 96 % ethyl alcohol were added to the resulting solution, the mixture was incubated for 2 h at –70 °C, then centrifuged for 30 min at 4 °C, 14,000 rpm. The supernatant was removed, the pellet was washed with 750 μl of chilled 70 % ethyl alcohol, and centrifuged for 10 min at 4 °C, 14,000 rpm. The supernatant was removed, the pellet was dried for 15 min at 37 °C and then dissolved in 30 μl of double-distilled water.

Three rounds of PCR were used to prepare the mapping sample. For the first round of amplification, we used 5 μl of purified ligase mixture, 0.5 μl of 10 μM primers PB-outer-F-2 and cDNA-ampl-R (see Table 1), 5 μl of 5× Phusion HF buffer (Thermo Fisher Scientific), 2 μl of 2.5 mM dNTPs, and 1.25 U of Phusion HS II DNA polymerase (Thermo Fisher Scientific) in a total volume of 25 μl. The subsequent rounds of amplification were carried out similarly using (i) primers PB-cDNA-fwd-A7 and InvPCR-F-Nextera2 (see Table 1) and 1 μl of the PCR products of the first round for the second round

of amplification and (ii) primers Libr-P5-for and Libr-P7-rev (see Table 1) and 1 μl of the PCR products of the second round for the third round of amplification. The thermal cycle conditions for the first round of PCR were as follows: 98 °C for 1 min, 12 cycles of 98 °C for 30 sec, 65 °C for 30 sec, 72 °C for 2 min, and a final incubation for 5 min at 72 °C. The thermal cycle conditions for the second round of PCR were as follows: 98 °C for 1 min, 12 cycles of 98 °C for 30 sec, 62 °C for 30 sec, 72 °C for 2 min, and a final incubation for 5 min at 72 °C. The thermal cycle conditions for the third round of PCR were as follows: 98 °C for 1 min, 16 cycles of 98 °C for 30 sec, 61 °C for 30 sec, 72 °C for 2 min, and a final incubation for 5 min at 72 °C. Finally, 5 μg of the mapping sample was treated with 10 U of NotI restriction endonuclease (New England Biolabs) in a total volume of 100 μl at 37 °C for 2 h to remove byproducts.

High-throughput sequencing and data analysis. Sequencing of the samples was performed on the Genolab 2 × 75 bp platform (https://genomed.ru/). Demultiplexing of the obtained fastq files using the sabre tool (https://github.com/najoshi/sabre) resulted in ~4.5, ~1.6 and ~1 million reads for the mapping, normalization, and expression samples, respectively. The quality analysis of the raw data was carried out using the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, using the TASK tool (The TRIP Analysis Software Kit, https://trip.nki.nl/), the sequences of reliably identified barcodes, as well as their normalized expression levels and genomic locations in the Chinese hamster genome assemblies CriGri-PICRH-1.0 (GCA_003668045.2) and Cgr1.0 (GCA_000448345.1) were identified. The genome assembly CriGri-PICRH-1.0 is characterized by the presence of extremely long sequences corresponding to all expected chromosomes, and therefore it was used as the "default" genome of CHO cells, whereas the genome assembly Cgr1.0 was previously used to map chromatin types in CHO-K1 cells (Feichtinger et al., 2016). Then, to determine the most reliable (hereafter referred to as filtered) transgenes, the following additional parameters were applied to the TASK output: norm >= 5, reads_r >= 10, freq1_r > 0.60. Data on chromatin types were taken for the Tp0 time point corresponding to the culturing of CHO-K1 cells for 4 h (https://cho-epigenome.boku.ac.at/JB/). A positional weight matrix for genomic sequences overlapping transgene insertion sites was generated using the pLogo application (https://plogo.uconn.edu/) (O'Shea et al., 2013).

Results and discussion

To study the activity of several promoters in different local chromatin environments in CHO cells in parallel, barcoded model transgenes were constructed based on the piggyBac transposon. The transgenes contained the puromycin-*N*-acetyltransferase (*pac*) resistance gene (hereafter $Puro^R$) and the improved green fluorescent protein (e*GFP*) gene (separated by an IRES element) under the control of the following four promoters: (1) the promoter of the mouse *PGK* (m*PGK*) gene, previously used for a similar study on mouse embryonic stem cells (Akhtar et al., 2013), (2) the promoter of the Chinese hamster *PGK* gene, homologous to the m*PGK* promoter, (3) full-length ("long") and (4) truncated ("short") variants of the Chinese hamster *EF-1α* gene promoter (Running Deer, Allison, 2004; Orlova et al., 2014; Wang et al., 2017) (Fig. 1, *A*). Constructs with each individual promoter also contained a specific 5-bp motif (promoter index) immediately before the 18-bp barcode located in the 3′UTR region. The presence of promoter indexes allowed the simultaneous use of all 4 barcoded model transgenes in a single experiment (Gisler et al., 2019) (see Fig. 1, *A*). Thus, the resulting barcoded plasmid libraries of the constructs with the long *EF-1α,* short *EF-1α*, m*PGK* and *PGK* promoters were mixed in a molar ratio of 7:7:7:1. The smaller proportion of the latter construct was due to its use in this experiment as a control; we also used the *PGK* promoter-containing construct to obtain stable transgenic populations of CHO cells upon puromycin selection (the results of that study will be reported elsewhere) and it seemed useful to us to have the technical ability to correctly compare data for such different transgenic populations in the future.

CHO cells (CHO-S subline) were co-transfected with the above-described mixture of model transgenes, as well as a plasmid encoding the piggyBac transposase. Seventy-two h after transfection, the eGFP protein expression was observed in approximately 40 % of the cells (see Fig. 1, *B*). After that, the cells were cultivated in the absence of any selection for additional 25 days in order to multiply the transgenic cells and to get rid of plasmid DNA molecules that could ultimately contaminate the data. As a result, multiple clones of transgenic cells were observed in the population (see Fig. 1, *B*).

From the resulting polyclonal population of cells, genomic DNA and total RNA were isolated, on the basis of which the genomic localizations and normalized expression levels of barcoded transgenes were determined. A total of 641 uniquely barcoded and genome-mapped transgenes were identified in the transgenic population. These transgenes were present in more or less expected numbers on all chromosomes of CHO cells (see Fig. 1, *C*). Analysis of genomic sequences overlapping transgene insertion sites revealed their at-richness, as well as the presence of a central ttaa motif (see Fig. 1, *D*) specific for the piggyBac transposon (Frase et al., 1996; Li et al., 2013; Chen Q. et al., 2020).

Among the identified transgenes, 38.8 % were with the tacaa promoter index (corresponding to the short *EF-1α* promoter), 24.3 %, with the ccgag promoter index (corresponding to the long *EF-1α* promoter), 32.2 %, with the ctagt promoter index (corresponding to the m*PGK* promoter), and 4.7 %, with the promoter index agctc (corresponding to the Chinese hamster *PGK* promoter) (Fig. 2, *A*).

Analysis of the activity of reporter constructs under the control of four different promoters revealed the presence of a large number of silent (i. e., transcriptionally inactive) transgenes with each promoter (Table 2), which is most likely due to the lack of antibiotic selection during generation of the population of transgenic cells.

Comparison of promoter activities among the filtered expressed transgenes (144 cases) showed that the main part of highly active reporter constructs is under the control of the full-length variant of the *EF-1α* gene promoter (see Fig. 2, *B,* Table 2). Particularly, among 10 % of the most active filtered transgenes, the promoters were distributed as follows: long *EF-1α* – 70 %, m*PGK* – 20 %, short *EF-1α* – 10 %, *PGK* – 0 %.

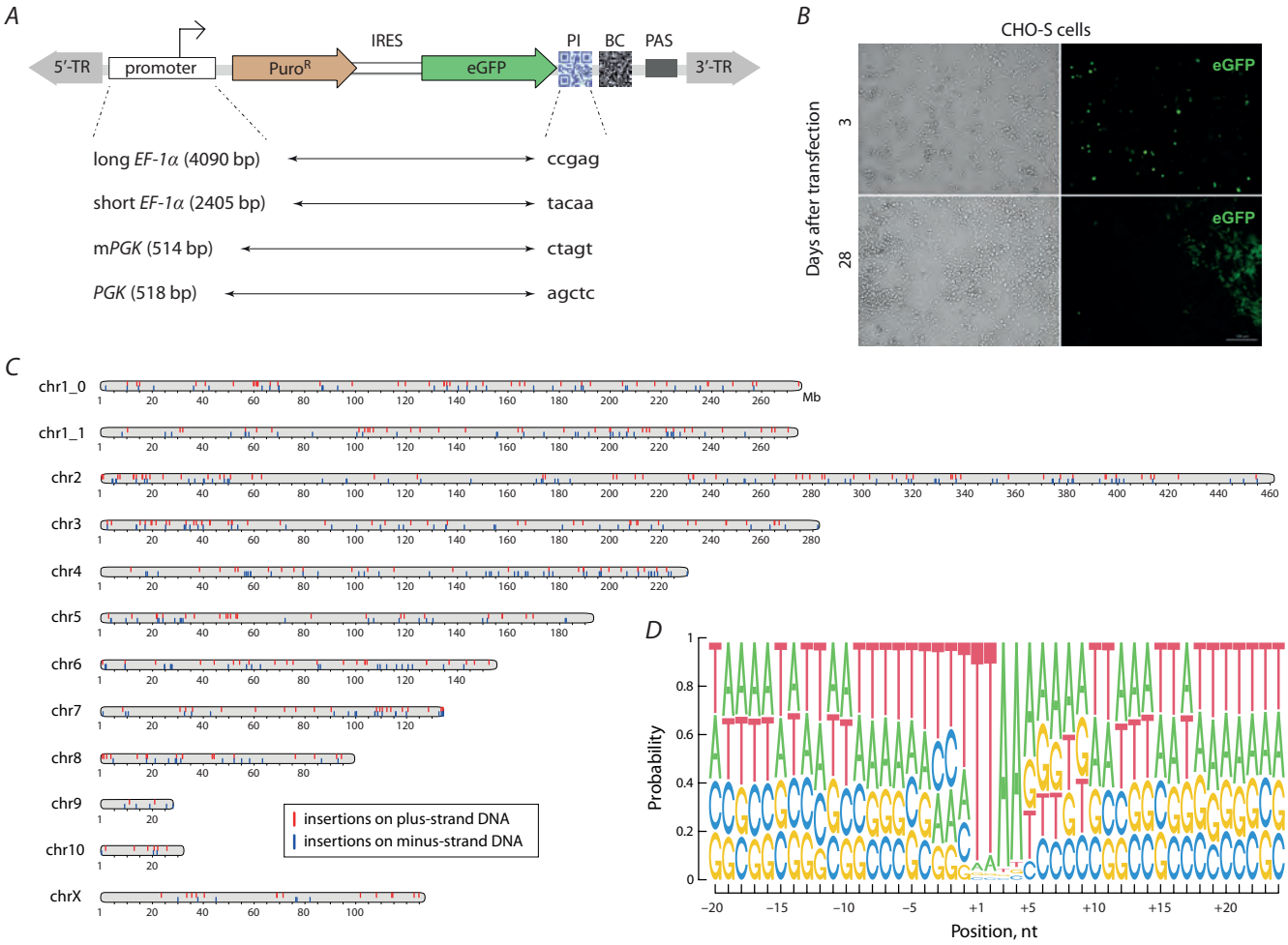

Fig. 1. Stable integration of model piggyBac transgenes into the genome of cultured CHO cells.

A, Schematic of the barcoded reporter constructs with four different promoters used in the study. 5′-TR and 3′-TR are inverse terminal repeats of the piggyBac transposon; IRES – internal ribosome entry site, PI – promoter index, BC – barcode, PAS – polyadenylation signal. *B*, CHO cells 3 and 28 days after transfection. *C*, Distribution of all uniquely mapped transgenes across Chinese hamster chromosomes. Red and blue vertical lines indicate transgene insertions on the plus and minus strand of the genome, respectively. *D*, Analysis of the genomic motifs, at which integration of all uniquely mapped transgenes occurred. Positions +1…+4 correspond to the sequence that is duplicated upon the piggyBac transposon insertion and flanks the integrated transgene.

It is worth noting that due to the small number of transgenes with the *PGK* promoter, the results on its activity are rather preliminary.

Two thirds of all transgenes were inserted into the genome of CHO cells within genes (which were defined as –1000 bp from the distal transcription initiation site to the transcription termination site), predominantly in introns (42.3 %), promoters (8.5 %) and 5′UTRs (9.4 %) (see Fig. 2, *C*). It should be noted that promoters were defined as –1000…+100 bp from the transcription start site. Similar patterns of transgene integration based on the piggyBac transposon have been described previously for cell cultures of other species (Ding et al., 2005; Wilson et al., 2007; Galvan et al., 2009; Li et al., 2013). Analysis of 10 % of the most active filtered reporter constructs (21 cases) revealed a ~1.6-fold, ~1.4-fold and ~1.1-fold increase in the proportion of transgenes in 5′UTRs, promoters and introns, respectively (see Fig. 2, *D*). In addition, it is interesting to note that transgenes were more often located closer to the gene starts than to the gene ends. The median distances from the transgene localization position in the genome to the nearest transcription initiation and termination sites were 11.4 and 20.2 kb, respectively, for the complete set of reporter genes (641 cases). At the same time, for 10 % of the most active filtered transgenes (21 cases), these values were equal to 6.6 and 17.8 kb, respectively.

To study the influence of the local chromatin environment on the activity of reporter genes, previously published data on the distribution of 11 chromatin types in the genome of CHO-K1 cells were used (Feichtinger et al., 2016). Because these data were originally associated with a version of the Chinese hamster genome that is different from that used for the analysis described above (see Materials and methods for details), it was possible to extract chromatin types overlapping positions of transgenes in the genome only for 595 out of 641 transgenes. Among these 595 transgenes, only 39.5 % were localized in inactive chromatin types “Quiescent/low”, “Repressed heterochromatin (H3K9me3)” and “Polycomb repressed regions (H3K27me3)”, which together cover more

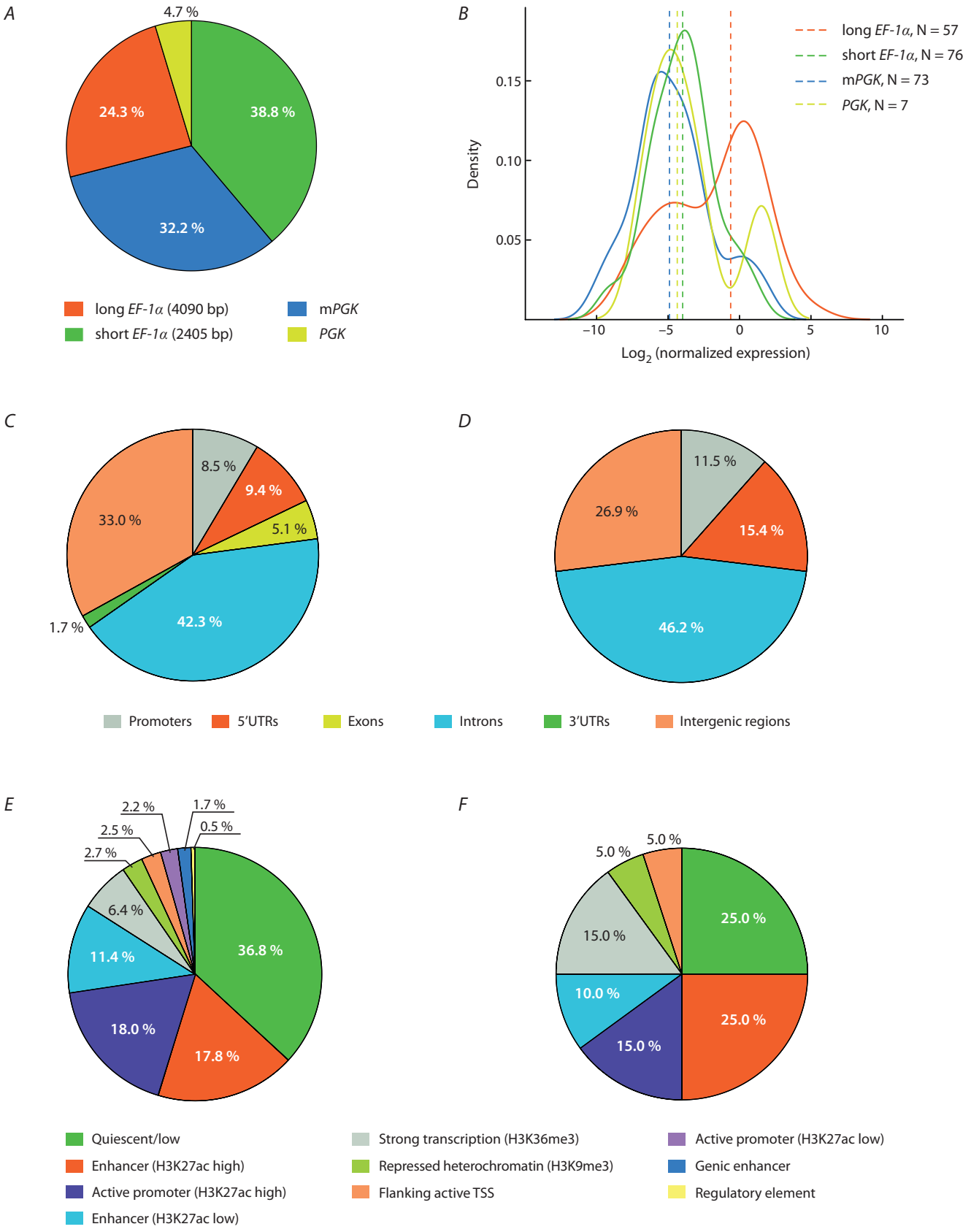

Fig. 2. Features of studied transgenes.

A, Distribution of all identified transgenes across the studied promoters. *B*, Comparison of promoter activities for the filtered 144 expressed transgenes (see Materials and methods). Dashed vertical lines indicate median normalized expression values. *C*, *D*, Distribution of all identified transgenes (*C*) and 10 % of the most active filtered transgenes (*D*) in gene elements (promoters, 5'UTRs, exons, introns, 3'UTRs) as well as intergenic regions. *E*, *F*, Distribution of all transgenes (*E*) and 10 % of the most active filtered transgenes (*F*) in chromatin types of CHO-K1 cells described previously (Feichtinger et al., 2016).

Table 2. Comparative activity of the studied promoters

Promoter	Transgene number	Proportion of silent transgenes, %	Proportion of active transgenes, %	Median promoter activity value, a. u.*
long *EF-1α*	156	42.31	57.69	18.93
short *EF-1α*	249	60.24	39.76	1.84
m*PGK*	206	53.88	46.12	1
PGK	30	66.67	33.33	1.42
Total	641	54.13	45.87	

* a. u. – arbitrary units.

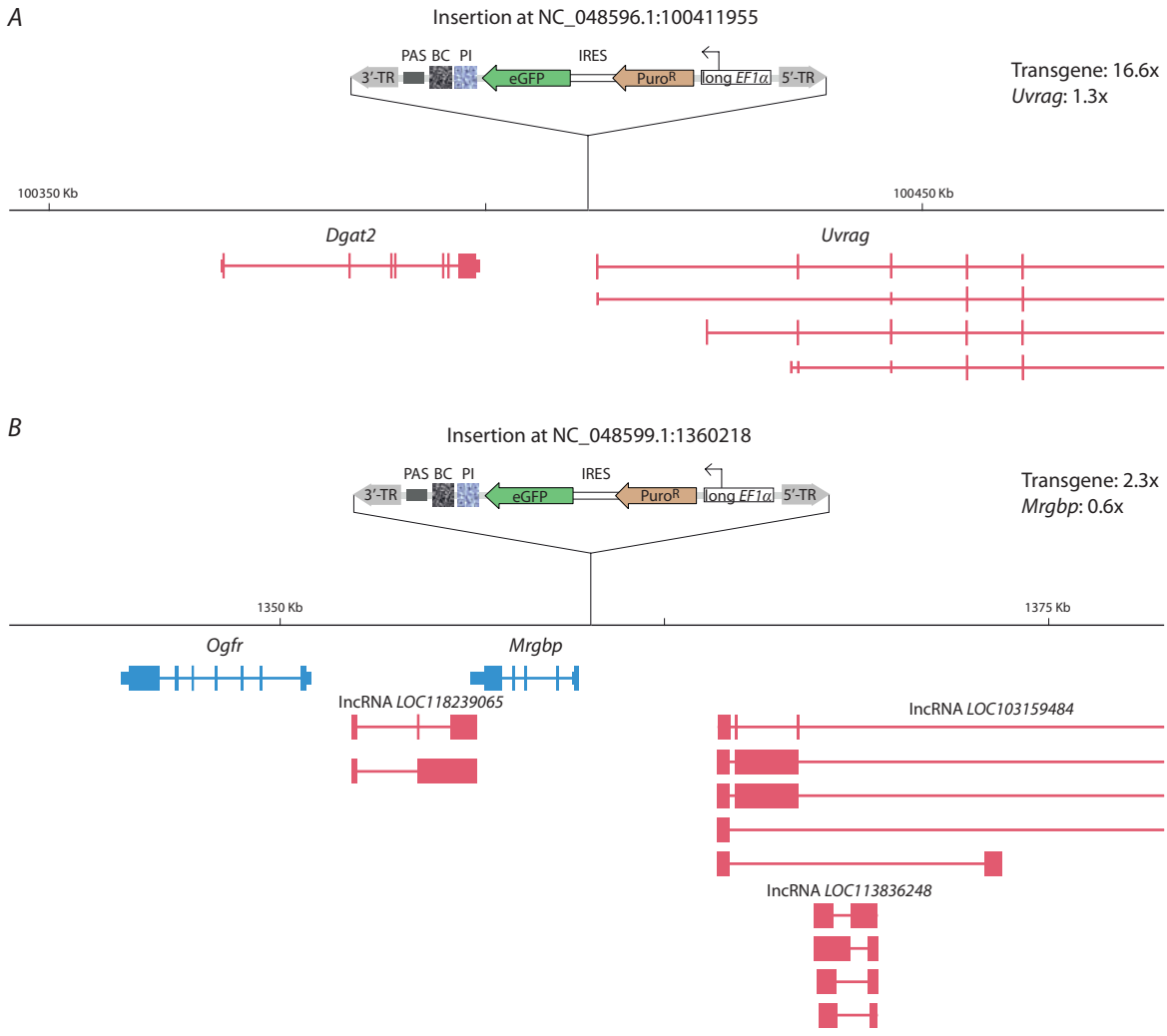

Fig. 3. Examples of genomic localization of transgenes with high (*A*) and medium (*B*) transcriptional activity.

The transgenes are not shown in scale. The activities of the transgenes and the genes nearest to them are indicated relative to the average expression levels of all reporter genes (641 cases) and all endogenous genes, respectively.

than 88 % of the genome of Chinese hamster cells (Feichtinger et al., 2016). The remaining 60.5 % of transgenes were localized in various active chromatin types (see Fig. 2, *E*).

The most active transgenes were more often localized in regions of the genome associated with the active chromatin types “Enhancer (H3K27ac high)”, “Strong transcription (H3K36me3)” and “Flanking active TSS” as well as with the inactive chromatin type “Repressed heterochromatin (H3K9me3)” (see Fig. 2, *F*). The latter rather unexpected observation may be due to the fact that the chromatin types were determined for a different subline of CHO cells.

Since, as noted above, two thirds of all transgenes were localized within genes (see Fig. 2, *C*), it is worth noting that the insertion of a transgene even into an important gene most

likely has only a minor effect on cell viability. This is supported by the following two considerations. First, not every insertion of a foreign sequence within a gene significantly disrupts its function. Second, typically, there is another native copy(ies) of the gene in the genome of cultured cells. Together, this ensures the successful survival of transgenic cells in a polyclonal population among non-modified (wild-type) cells. The chances of damaging both alleles of a gene in the experimental setup used are negligible: to achieve that, two transgenes must be integrated into both alleles of the same gene in the same cell. Accordingly, the genomic positions of active transgenes identified in this study may qualify for consideration as promising sites for targeted integration of biotechnological transgenes, even if they are located inside active genes (Fig. 3).

Conclusion

In a polyclonal population of transgenic CHO cells cultured in the absence of selection pressure, more than half of the model reporter constructs stably integrated into the genome were transcriptionally inactive. Compared to the complete set of transgenes, the most active transgenes were about 1.6 and 1.4 times more frequently localized in promoters and 5′UTRs of genes, respectively. Also, compared to the complete set of transgenes, the most active transgenes were about 2.3 and 1.4 times more frequently localized in the transcriptionally active chromatin types “Strong transcription (H3K36me3)” and “Enhancer (H3K27ac high)”, respectively. Transgenes containing the full-length promoter of the Chinese hamster *EF-1α* gene were on average the most active ones. At the same time, the median activity of the short variant of the *EF-1α* gene promoter was about 10 times lower than the median activity of the full-length promoter of this gene. This can be explained by the presence of important binding sites for transcription factors in the full-length version of the *EF-1α* gene promoter. Genomic sites of the most active insertions of model transgenes may be of interest for further experiments as promising positions for targeted integration of biotechnological constructs.

References

Akhtar W., de Jong J., Pindyurin A.V., Pagie L., Meuleman W., de Ridder J., Berns A., Wessels L.F.A., van Lohuizen M., van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*. 2013;154(4):914-927. DOI 10.1016/j.cell.2013.07.018

Akhtar W., Pindyurin A.V., de Jong J., Pagie L., ten Hoeve J., Berns A., Wessels L.F.A., van Steensel B., van Lohuizen M. Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc*. 2014;9(6):1255-1281. DOI 10.1038/nprot.2014.072

Babenko V.N., Makunin I.V., Brusentsova I.V., Belyaeva E.S., Maksimov D.A., Belyakin S.N., Maroy P., Vasil′eva L.A., Zhimulev I.F. Paucity and preferential suppression of transgenes in late replication domains of the *D. melanogaster* genome. *BMC Genomics*. 2010;11:318. DOI 10.1186/1471-2164-11-318

Chen M., Licon K., Otsuka R., Pillus L., Ideker T. Decoupling epigenetic and genetic effects through systematic analysis of gene position. *Cell Rep.* 2013;3(1):128-137. DOI 10.1016/j.celrep.2012.12.003

Chen Q., Luo W., Veach R.A., Hickman A.B., Wilson M.H., Dyda F. Structural basis of seamless excision and specific targeting by *piggyBac* transposase. *Nat. Commun*. 2020;11(1):3446. DOI 10.1038/s41467-020-17128-1

Dahodwala H., Lee K.H. The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Curr. Opin. Biotechnol*. 2019;60:128-137. DOI 10.1016/j.copbio.2019.01.011

Ding S., Wu X., Li G., Han M., Zhuang Y., Xu T. Efficient transposition of the *piggyBac* (*PB*) transposon in mammalian cells and mice. *Cell*. 2005;122(3):473-483. DOI 10.1016/j.cell.2005.07.013

Elgin S.C.R., Reuter G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol*. 2013;5(8):a017780. DOI 10.1101/cshperspect.a017780

Feichtinger J., Hernández I., Fischer C., Hanscho M., Auer N., Hackl M., Jadhav V., Baumann M., Krempl P.M., Schmidl C., Farlik M., Schuster M., Merkel A., Sommer A., Heath S., Rico D., Bock C., Thallinger G.G., Borth N. Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnol. Bioeng*. 2016;113(10):2241-2253. DOI 10.1002/bit.25990

Fraser M.J., Ciszczon T., Elick T., Bauser C. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol. Biol*. 1996;5(2):141-151. DOI 10.1111/j.1365-2583.1996.tb00048.x

Galvan D.L., Nakazawa Y., Kaja A., Kettlun C., Cooper L.J.N., Rooney C.M., Wilson M.H. Genome-wide mapping of *PiggyBac* transposon integrations in primary human T cells. *J. Immunother*. 2009;32(8):837-844. DOI 10.1097/CJI.0b013e3181b2914c

Gierman H.J., Indemans M.H.G., Koster J., Goetze S., Seppen J., Geerts D., van Driel R., Versteeg R. Domain-wide regulation of gene expression in the human genome. *Genome Res*. 2007;17(9):1286-1295. DOI 10.1101/gr.6276007

Gisler S., Gonçalves J.P., Akhtar W., de Jong J., Pindyurin A.V., Wessels L.F.A., van Lohuizen M. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat. Commun*. 2019;10(1):1598. DOI 10.1038/s41467-019-09551-w

Gupta K., Modi D., Jain R., Dandekar P. A stable CHO K1 cell line for producing recombinant monoclonal antibody against TNF-α. *Mol. Biotechnol*. 2021;63(9):828-839. DOI 10.1007/s12033-021-00329-4

Huang X., Guo H., Tammana S., Jung Y.-C., Mellgren E., Bassi P., Cao Q., Tu Z.J., Kim Y.C., Ekker S.C., Wu X., Wang S.M., Zhou X. Gene transfer efficiency and genome-wide integration profiling of *Sleeping Beauty*, *Tol2*, and *piggyBac* transposons in human primary T cells. *Mol. Ther*. 2010;18(10):1803-1813. DOI 10.1038/mt.2010.141

Kim J.Y., Kim Y.-G., Lee G.M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol*. 2012;93(3):917-930. DOI 10.1007/s00253-011-3758-5

Lalonde M.-E., Durocher Y. Therapeutic glycoprotein production in mammalian cells. *J. Biotechnol*. 2017;251:128-140. DOI 10.1016/j.jbiotec.2017.04.028

Lebedev M.O., Yarinich L.A., Ivankin A.V., Pindyurin A.V. Generation of barcoded plasmid libraries for massively parallel analysis of chromatin position effects. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2019;23(2):203-211. DOI 10.18699/VJ19.483

Li M.A., Pettitt S.J., Eckert S., Ning Z., Rice S., Cadiñanos J., Yusa K., Conte N., Bradley A. The *piggyBac* transposon displays local and distant reintegration preferences and can cause mutations at non-canonical integration sites. *Mol. Cell. Biol.* 2013;33(7):1317-1330. DOI 10.1128/MCB.00670-12

Orlova N.A., Kovnir S.V., Hodak J.A., Vorobiev I.I., Gabibov A.G., Skryabin K.G. Improved elongation factor-1 alpha-based vectors for stable high-level expression of heterologous proteins in Chinese hamster ovary cells. *BMC Biotechnol.* 2014;14:56. DOI 10.1186/1472-6750-14-56

O'Shea J.P., Chou M.F., Quader S.A., Ryan J.K., Church G.M., Schwartz D. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods*. 2013;10(12):1211-1212. DOI 10.1038/nmeth.2646

Ritacco F.V., Wu Y., Khetan A. Cell culture media for recombinant protein expression in Chinese hamster ovary (CHO) cells: history, key components, and optimization strategies. *Biotechnol. Prog*. 2018;34(6):1407-1426. DOI 10.1002/btpr.2706

Ruf S., Symmons O., Uslu V.V., Dolle D., Hot C., Ettwiller L., Spitz F. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* 2011;43(4):379-386. DOI 10.1038/ng.790

Running Deer J., Allison D.S. High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1α gene. *Biotechnol. Prog*. 2004;20(3):880-889. DOI 10.1021/bp034383r

Stach C.S., McCann M.G., O'Brien C.M., Le T.S., Somia N., Chen X., Lee K., Fu H.Y., Daoutidis P., Zhao L., Hu W.S., Smanski M. Model-driven engineering of N-linked glycosylation in Chinese hamster ovary cells. *ACS Synth. Biol*. 2019;8(11):2524-2535. DOI 10.1021/acssynbio.9b00215

Wang X., Xu Z., Tian Z., Zhang X., Xu D., Li Q., Zhang J., Wang T. The EF-1α promoter maintains high-level transgene expression from episomal vectors in transfected CHO-K1 cells. *J. Cell. Mol. Med*. 2017;21(11):3044-3054. DOI 10.1111/jcmm.13216

Wilson M.H., Coates C.J., George A.L., Jr. *PiggyBac* transposon-mediated gene transfer in human cells. *Mol. Ther*. 2007;15(1):139-145. DOI 10.1038/sj.mt.6300028

Xu W.-J., Lin Y., Mi C.-L., Pang J.-Y., Wang T.-Y. Progress in fed-batch culture for recombinant protein production in CHO cells. *Appl. Microbiol. Biotechnol*. 2023;107(4):1063-1075. DOI 10.1007/s00253-022-12342-x

ORCID ID

L.A. Yarinich orcid.org/0000-0003-0469-0371
A.A. Ogienko orcid.org/0000-0002-0896-1899
A.V. Pindyurin orcid.org/0000-0001-6959-0641
E.S. Omelina orcid.org/0000-0002-2189-5101

Acknowledgements. The work was carried out with financial support from the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1086, contract RF----193021X0015, 15.IP.21.0015).
We are grateful to A.V. Taranin (Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia), V.V. Verkhusha (Albert Einstein College of Medicine, Bronx, NY, USA) and V.S. Fishman (Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia) for providing CHO-S cells, the pRP[Exp]-mCherry-CAG>hyPBase plasmid and assistance in extracting data on the chromatin types of CHO cells, respectively.

Conflict of interest. The authors declare no conflict of interest.

Received September 13, 2023. Revised September 21, 2023. Accepted September 27, 2023.