

Перевод на английский язык <https://vavilov.elpub.ru/jour>

## 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](mailto: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, измерена 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-1 $\alpha$*  китайского хомячка. Трансгены с укороченным вариантом этого же промотора, а также трансгены с промотором мышиного гена *PGK (mPGK)* 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.

**Для цитирования:** Яринич Л.А., Огиенко А.А., Пиндюрин А.В., Омелина Е.С. Анализ транскрипционной активности модельных piggyBac-трансгенов, стабильно интегрированных в разные локусы генома культивируемых клеток CHO при отсутствии селекционного давления. *Вавиловский журнал генетики и селекции*. 2023;27(7): 906-915. DOI 10.18699/VJGB-23-105

## 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](mailto: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-1 $\alpha$*  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 Selekcii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):906-915. DOI 10.18699/VJGB-23-105

## Введение

Метод TRIP (thousands of reporters integrated in parallel) позволяет выполнять масштабные исследования влияния локального окружения хроматина на активность трансгенов. Метод основан на использовании ДНК-штрихкодов (далее – просто штрихкодов) и исходно был апробирован на культивируемых эмбриональных стволовых клетках мыши с помощью транспозона piggyBac в качестве средства доставки трансгенов в геном (Akhtar et al., 2013). Штрихкод – это короткая последовательность ДНК (длиной 16–20 п. н.), являющаяся уникальной для каждой копии трансгена, используемой в исследовании. Крайне важно, что штрихкод располагается в пределах транскрибируемой части трансгенов, что обеспечивает его присутствие не только в ДНК, но также в составе молекул мРНК, синтезированных с трансгенов. Соответственно, штрихкод можно применять и для количественного измерения уровня транскрипционной активности трансгенов.

Транспозон piggyBac позволяет эффективно модифицировать различные клеточные линии и организмы (Wilson et al., 2007), в том числе протяженными конструкциями (Ding et al., 2005). Кроме того, транспозон piggyBac характеризуется относительно равномерным профилем встраивания в клеточный геном (Huang et al., 2010). В экспериментах TRIP система для трансгенеза исследуемых клеток состоит из двух плазмидных конструкций: конструкции для экспрессии транспозазы piggyBac, катализирующей встраивание трансгена в случайное место генома, и собственно трансгена – целевой конструкции (состоящей из промотора, репортерного гена, штрихкода и сигнала полиаденилирования), расположенной между обращенными повторами транспозона piggyBac (Akhtar et al., 2014; Lebedev et al., 2019). Котрансфекция клеток такими плазмидными конструкциями позволяет получить поликлональную популяцию трансгенных клеток, в которой каждая индивидуальная встройка трансгена в геноме маркирована уникальной последовательностью штрихкода. После размножения трансфицированных клеток из них выделяют геномную ДНК и тотальную РНК. С помощью образца геномной ДНК идентифицируют сайты локализации трансгенов в геноме и определяют представленность каждого штрихкода в популяции клеток. На основе образца тотальной РНК определяют представленность каждого штрихкода в общей массе транскриптов, синтезированных с трансгенов. Наконец, соотношение представленности каждого штрихкода в молекулах мРНК

и его представленности в геноме трансгенных клеток позволяет количественно измерить уровень транскрипционной активности каждого отдельного трансгена (Akhtar et al., 2014).

В данной работе с помощью метода TRIP исследовано влияние локального окружения хроматина на транскрипционную активность трансгенов в клетках яичника китайского хомячка CHO. Культивируемая клеточная линия CHO наиболее часто используется для наработки разнообразных белков (Xu et al., 2023). Несмотря на доступность ряда других культивируемых клеток млекопитающих, таких как клетки почки детеныша хомячка, клетки мышинной миеломы NS0, эмбриональные клетки почки человека (HEK293) и эмбриональные клетки сетчатки глаза человека PerC6, более 70 % всех рекомбинантных терапевтических белков производится в клетках яичника китайского хомячка (Kim et al., 2012; Ritacco et al., 2018; Gupta et al., 2021). Популярность клеток CHO можно объяснить следующими причинами. Во-первых, применение клеток CHO для производства рекомбинантных белков является безопасным, так как клетки CHO невосприимчивы к заражению вирусами человека (Lalonde, Durocher, 2017). Во-вторых, клетки CHO обладают способностью к эффективной посттрансляционной модификации и продуцируют рекомбинантные белки с совместимыми для человека гликоформами (Stach et al., 2019). Наконец, клетки линии CHO обладают высокой скоростью роста и относительно легко адаптируются к росту в суспензии, что является предпочтительной характеристикой для крупномасштабного культивирования в биореакторах (Ritacco et al., 2018; Dahodwala, Lee, 2019). В настоящее время биореакторы объемом более 10 тыс. литров используются для суспензионных культур рекомбинантных клеток CHO, продуцирующих терапевтические антитела (Kim et al., 2012).

Локализация в геноме оказывает большое влияние на уровень экспрессии рекомбинантного гена (явление, известное как эффект положения) (Gierman et al., 2007; Vabbenko et al., 2010; Ruf et al., 2011; Chen M. et al., 2013; Elgin, Reuter, 2013). Интеграция в неактивный гетерохроматин приводит к незначительной экспрессии трансгена или ее полному отсутствию, тогда как интеграция в активный эухроматин часто делает возможной экспрессию трансгена. Однако просто интеграции в эухроматин может быть недостаточно для обеспечения длительной экспрессии рекомбинантного гена. Экспрессия трансгена в клетках

млекопитающих во многих случаях быстро инактивируется (замалчивается), в частности, вероятно, из-за влияния соседнего конденсированного хроматина.

Интеграция трансгенов в транскрипционно-активные области генома является одной из стратегий, позволяющих избежать их инактивации. Данная работа направлена на анализ транскрипционной активности *riggy* в трансгенов, интегрированных в разные локусы генома культивируемых клеток CHO при отсутствии селекционного давления.

## Материалы и методы

**Приготовление конструкции pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242.** Плазмида pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242 была приготовлена на основе ранее описанной универсальной конструкции (Lebedev et al., 2019). Встройку амплифицировали с использованием праймеров

mPGK-EcoRI-F и eGFP-XbaI-R (табл. 1), используя в качестве матрицы плазмиду mPGK-Puro-IRES-eGFP-sNRP-pA (Akhtar et al., 2013). В 50 мкл реакционной смеси добавляли 1 нг матрицы плазмиды, 2,5 е. а. Phusion-полимеразы (ThermoFisher Scientific), по 1 мкл 10 мкМ праймеров и дНТФ до 0.2 мМ. Условия ПЦР: 98 °С в течение 30 с, 35 циклов: 98 °С 10 с, 62 °С 10 с, 72 °С 1 мин, инкубация 10 мин при 72 °С.

**Получение генно-инженерных конструкций с различными промоторами генов китайского хомячка.** Плазмиду pPB-mPGK-Puro-IRES-eGFP-PI.11-TR.242 гидролизовали по сайтам рестрикции EcoRI, BglIII, AgeI. Для получения встроок амплифицировали последовательности промоторов гена *PGK* китайского хомячка и длинного и короткого варианта промотора гена *EF-1α* с использованием праймеров hamP<sub>gk1</sub>-EcoRI-F и hamP<sub>gk1</sub>-BglIII-R, CHEF-1-v1-EcoRI-F и CHEF-1-v1-BclII-R, CHEF-

**Таблица 1.** Список праймеров, использованных в работе

Название праймера	Последовательность (5'→3')
mPGK-EcoRI-F	aaagaattctcgacaattctaccgggtagg
eGFP-XbaI-R	aatctagaccctccggattacttg
hamP <sub>gk1</sub> -EcoRI-F	aaagaattcaggtccctgggattcca
hamP <sub>gk1</sub> -BglIII-R	aaaagatctcgtaggataagaggctcag
CHEF-1-v1-EcoRI-F	aaagaattccagttgtgcatagaacagatgc
CHEF-1-v1-BclII-R	aatgatcatgtgtttcacaacaccttaaaaaaagtctg
CHEF-1-v2-EcoRI-F	aaagaattcaagcttctgtgatagaaatgattag
CHEF-1-v2-BclII-R	aatgatcactgcgttctgacggcaaac
Plasmid-1	ccgcttaattaatccagcttttggtc
pPB-eGFP-PI-6-R	ctcgagctctcgatcttagacc
pPB-eGFP-PI-11-R	ctcactagctcgatcttagacc
pPB-eGFP-PI-16-R	ctctgtactcgatcttagacc
pPB-eGFP-PI-28-R	ctcctcggctcgatcttagacc
PB-Barcode-PI-6-Gibson-F	gtctagagatcgagagctcaggN <sub>18</sub> gagttgtggccggccctgtg
PB-Barcode-PI-11-Gibson-F	gtctagagatcgagctagtgaggN <sub>18</sub> gagttgtggccggccctgtg
PB-Barcode-PI-16-Gibson-F	gtctagagatcgagtacaagaggN <sub>18</sub> gagttgtggccggccctgtg
PB-Barcode-PI-28-Gibson-F	gtctagagatcgagccgaggaggN <sub>18</sub> gagttgtggccggccctgtg
PB-Gibson-R1	aacaaagctggattaattaagcggccgcatagctactagattaacc
Libr-cDNA-for	gtctcgtgggctcggagatgtgtataagagacaggtcctgctggagttcgtgac
Libr-cDNA-A16-rev	tcgtcggcagcgtcagatgtgtataagagacagcctatggtcggcagggtttccagtcacaagg
Libr-cDNA-A23-rev	tcgtcggcagcgtcagatgtgtataagagacagtaattgctcggcagggtttccagtcacaagg
Libr-P5-for	aatgatacggcgaccaccgagatctacactcgtcggcagcgtc
Libr-P7-rev	caagcagaagcggcatacagagatgtctcgtgggctcgg
PB-outer-F-2	ttttacgatgatattcttaacgtacgtc
cDNA-ampl-R	cgccagggtttccagtcacaag
PB-cDNA-fwd-A7	tcgtcggcagcgtcagatgtgtataagagacagagcagctgtcacaaggccggccaca
InvPCR-F-Nextera2	gtctcgtgggctcggagatgtgtataagagacaggtacgtcacaatatgattatcttctag

Примечание. N<sub>18</sub> – случайная 18-буквенная последовательность ДНК-штрихкода.

1-v2-EcoRI-F и CHEF-1-v2-BclI-R (см. табл. 1) соответственно. В 50 мкл реакционной смеси добавляли 50 нг матрицы геномной ДНК, выделенной из культивируемых клеток CHO, 2.5 е. а. Phusion-полимеразы (ThermoFisher Scientific), по 1 мкл 10 мкМ праймеров и дНТФ до 0.2 мМ. Условия ПЦР: 98 °С 30 с, 35 циклов: 98 °С 10 с, 62 °С 10 с, 72 °С 1 мин, инкубация 10 мин при 72 °С.

**Штрихкодированные плазмидные библиотеки** были приготовлены в соответствии с ранее описанным протоколом (Lebedev et al., 2019) с помощью метода бесшовного клонирования по Гибсону. Для этого с помощью ПЦР были получены векторы и встройки, содержащие 18-буквенный ДНК-штрихкод и промоторный индекс. Для амплификации векторов использовали праймеры Plasmid-1 и pPB-eGFP-PI-6-R/pPB-eGFP-PI-11-R/pPB-eGFP-PI-16-R/pPB-eGFP-PI-28-R (см. табл. 1) соответственно для конструкций с промотором гена *PGK* китайского хомячка/промотором гена *mPGK*/коротким вариантом промотора гена *EF-1 $\alpha$* /длинным вариантом промотора гена *EF-1 $\alpha$* . Для амплификации штрихкодированных встроек для получения конструкций с промотором гена *PGK* китайского хомячка/промотором гена *mPGK*/коротким вариантом промотора гена *EF-1 $\alpha$* /длинным вариантом промотора гена *EF-1 $\alpha$*  использовали праймеры PB-Gibson-R1 и PB-Barcode-PI-6-Gibson-F/PB-Barcode-PI-11-Gibson-F/PB-Barcode-PI-16-Gibson-F/PB-Barcode-PI-28-Gibson-F (см. табл. 1) соответственно. В 50 мкл реакционной смеси добавляли 1 нг матрицы, 2.5 е. а. Phusion-полимеразы (ThermoFisher Scientific), по 1 мкл 10 мкМ праймеров и дНТФ до 0.2 мМ. Условия ПЦР: 98 °С 30 с, 35 циклов: 98 °С 10 с, 62 °С 10 с, 72 °С 1 мин, инкубация 10 мин при 72 °С. После очистки 200 нг «вектора» и 135 нг «встроек» смешивали с 10 мкл 2 $\times$  NEBuilder HiFi DNA Assembly Master в суммарном объеме 20 мкл. Лигирование ДНК и трансформацию бактерий выполняли в соответствии с ранее описанным протоколом (Lebedev et al., 2019). Штрихкодированные плазмидные библиотеки выделяли с помощью набора Mega Plasmid Kit (Qiagen).

**Получение поликлональной трансгенной популяции клеток CHO.** За 24 ч до трансфекции культивируемые клетки CHO-S (далее – клетки CHO; любезно предоставлены лабораторией иммуногенетики ИМКБ СО РАН) рассаживали в 12-луночный культуральный планшет в концентрации  $1.5 \cdot 10^5$  клеток на лунку в 1 мл среды IMDM с добавлением 10 % сыворотки крупного рогатого скота. Проводили котрансфекцию клеток смесью плазмидных штрихкодированных библиотек (3 мкг) и плазмидой pRP[Exp]-mCherry-CAG>hyPBase (VectorBuilder #VB160216-10057; любезно предоставлена проф. В.В. Верхушей, Медицинский колледж им. Альберта Эйнштейна, Бронкс, Нью-Йорк, США) (0.3 мкг) с использованием реагента X-tremeGENE HP DNA transfection reagent (Roche). Трансгенную популяцию клеток культивировали в течение месяца при отсутствии селекционного давления.

**Выделение геномной ДНК.** Геномную ДНК выделяли из  $5 \cdot 10^7$  клеток полученной поликлональной трансгенной популяции с помощью набора PureLink<sup>®</sup> Genomic DNA Kit (Invitrogen) в соответствии с рекомендациями производителя.

### **Выделение тотальной РНК, обратная транскрипция.**

Тотальную РНК выделяли из  $5 \cdot 10^7$  клеток полученной поликлональной трансгенной популяции с помощью реагента RNazol RT (Molecular Research Center) в соответствии с рекомендациями производителя. Выделенную РНК инкубировали с 20 е. а. эндонуклеазы DpnI (New England Biolabs) и 3 е. а. ДНКазы I (ThermoFisher Scientific) 30 мин при 37 °С. Для очистки РНК использовали набор CleanRNA Standard («Евроген»). 2 мкг тотальной РНК смешивали с 1 мкл 50 мМ олиго(dT) праймера в суммарном объеме 13.5 мкл, смесь инкубировали в течение 5 мин при 65 °С. Последующую реакцию обратной транскрипции проводили в объеме 20 мкл со следующими компонентами: 13.5 мкл матрицы РНК с отожженными праймерами, 4 мкл 5 $\times$  буфера RT (ThermoFisher Scientific), 1 мкл 10 мМ дНТФ, 1 мкл RNaseOUT (ThermoFisher Scientific), 100 е. а. обратной транскриптазы RevertAid (ThermoFisher Scientific). Смесь инкубировали 60 мин при 42 °С, инактивировали фермент в течение 10 мин при 70 °С.

**Приготовление образцов нормирования и экспрессии.** Проводили два раунда ПЦР. Для первого раунда амплификации использовали: 600 нг матрицы геномной ДНК (для образцов нормирования) или 3 мкл кДНК (для образцов экспрессии), 0.5 мкл 10 мкМ праймеров Libr-cDNA-for и Libr-cDNA-A16-rev/Libr-cDNA-A23-rev (см. табл. 1) соответственно для образцов нормирования/экспрессии, 2 мкл 2.5 мМ дНТФ, 1.25 е. а. Phusion HS II ДНК-полимеразы (ThermoFisher Scientific), 5 мкл 5 $\times$  буфера Phusion HF (ThermoFisher Scientific) и бидистиллированную воду до конечного объема 25 мкл. Условия первого раунда ПЦР: 98 °С 1 мин, 15 циклов: 98 °С 30 с, 70 °С 30 с, 72 °С 30 с, инкубация 5 мин при 72 °С. Второй раунд амплификации проводили в объеме 25 мкл со следующими компонентами: 0.5 мкл ПЦР-продукта первого раунда, по 0.25 мкл 10 мкМ праймеров Libr-P5-for и Libr-P7-rev (см. табл. 1), 2 мкл 2.5 мМ дНТФ, 1.25 е. а. Phusion HotStart II ДНК-полимеразы (ThermoFisher Scientific), 5 мкл 5 $\times$  буфера Phusion HighFidelity (ThermoFisher Scientific). Условия второго раунда ПЦР: 98 °С 1 мин, 23 цикла: 98 °С 30 с, 61 °С 30 с, 72 °С 30 с, инкубация 5 мин при 72 °С.

**Приготовление образцов картирования.** 2 мкг геномной ДНК инкубировали с 10 е. а. эндонуклеазы рестрикции DpnII (New England Biolabs) при 37 °С в течение 16 ч, затем очищали с помощью набора реагентов GeneJET PCR Purification Kit (ThermoFisher Scientific). 600 нг фрагментированной геномной ДНК смешивали с 4 мкл 100 мМ АТФ, 2.5 е. а. T4 ДНК-лигазы («Евроген») в суммарном объеме 400 мкл. Лигазную смесь инкубировали 2 ч при комнатной температуре и 16 ч при 4 °С, лигазу инактивировали при 65 °С в течение 10 мин. К реакции лигирования добавляли 100 мкл бидистиллированной воды и 500 мкл раствора фенол:хлороформ (в соотношении 1:1), перемешивали, центрифугировали при комнатной температуре 5 мин 10000 g, переносили верхнюю фазу в новую пробирку. К полученному раствору добавляли 1/10 объема 3М NaOAc (рН 5.5), 2.5 объема 96 % этилового спирта, инкубировали 2 ч при –70 °С, центрифугировали 30 мин при 4 °С, 14000 об/мин. Удаляли супернатант, осадок промывали 750 мкл охлажденного 70 % этилового спирта, центрифугировали 10 мин при 4 °С, 14000 об/мин.

Удаляли супернатант, осадок высушивали 15 мин при 37 °С, ДНК растворяли в 30 мкл бидистиллированной воды.

Для приготовления образцов картирования проводили три раунда ПЦР. Для первого раунда амплификации использовали: 5 мкл очищенной лигазной смеси, по 0.5 мкл 10 мкМ праймеров PB-outer-F-2 и cDNA-ampl-R (см. табл. 1), 2 мкл 2.5 мМ дНТФ, 1.25 е. а. Phusion HS II ДНК-полимеразы (ThermoFisher Scientific), 5 мкл 5× буфера Phusion HighFidelity (ThermoFisher Scientific) и бидистиллированную воду до конечного объема 25 мкл. Второй и третий раунды амплификации проводили в том же составе с использованием 10 мкМ праймеров PB-cDNA-fwd-A7 и InvPCR-F-Nextera2 (см. табл. 1) для второго раунда, Libr-P5-for и Libr-P7-rev (см. табл. 1) для третьего раунда, в качестве матрицы использовали 1 мкл смеси ПЦР первого и второго раундов соответственно. Условия первого раунда ПЦР: 98 °С 1 мин, 12 циклов: 98 °С 30 с, 65 °С 30 с, 72 °С 2 мин, инкубация 5 мин при 72 °С. Условия второго раунда ПЦР: 98 °С 1 мин, 12 циклов: 98 °С 30 с, 62 °С 30 с, 72 °С 2 мин, инкубация 5 мин при 72 °С. Условия третьего раунда ПЦР: 98 °С 1 мин, 16 циклов: 98 °С 30 с, 61 °С 30 с, 72 °С 2 мин, инкубация 5 мин при 72 °С. Образец картирования (5 мкг) обрабатывали 10 е. а. эндонуклеазы рестрикции NotI для удаления побочных продуктов в суммарном объеме 100 мкл при 37 °С в течение 2 ч.

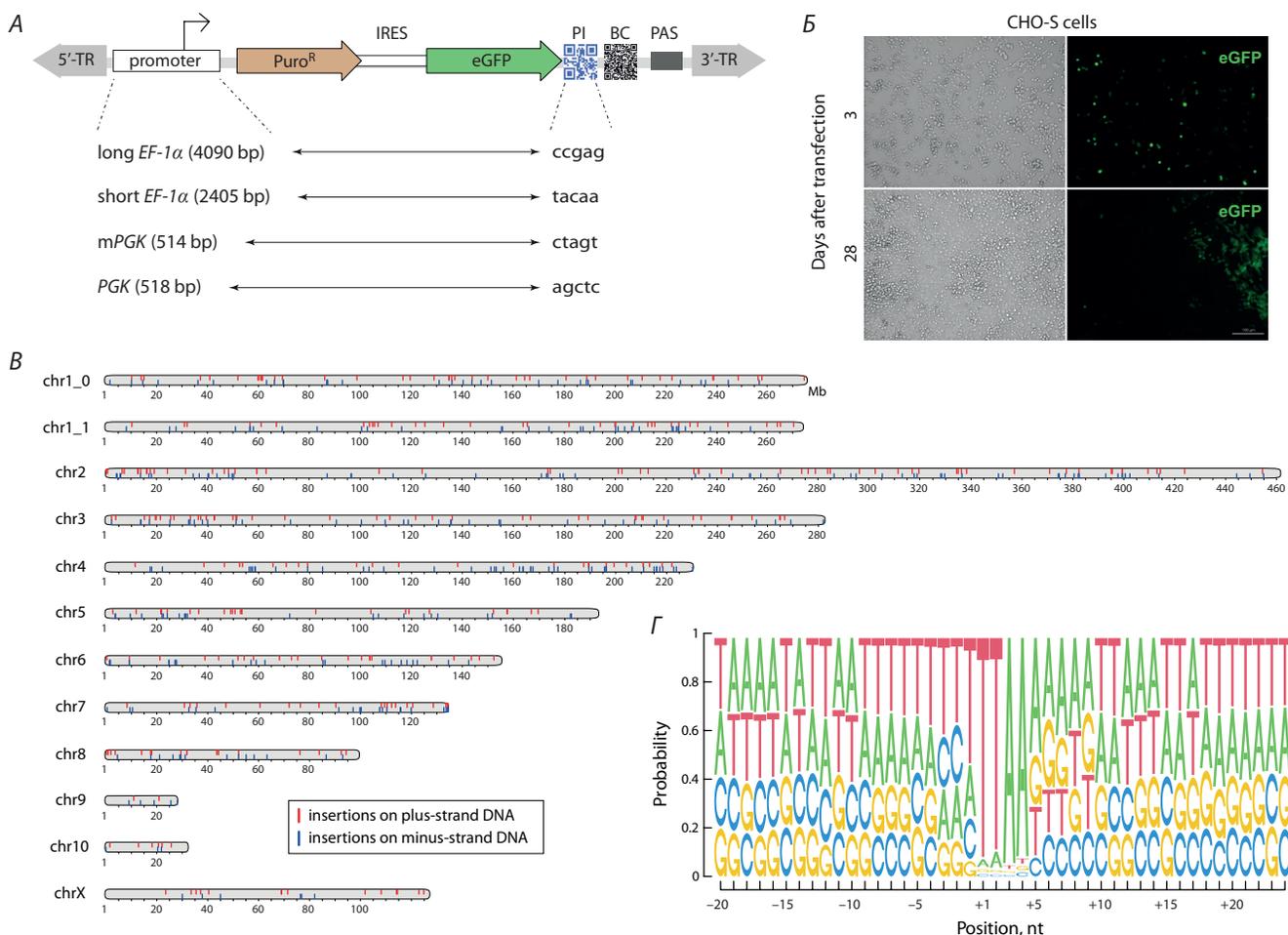
**Секвенирование и анализ данных.** Секвенирование образцов проводили на платформе Genolab 2×75 п. о. (<https://genomed.ru/>). Fastq-файлы были демультимплексированы с помощью инструмента sabre (<https://github.com/najoshi/sabre>). В результате получено 4.5 млн, 1.6 млн и около 1 млн прочтений для образцов картирования, нормирования и экспрессии соответственно. Анализ качества прочтений fastq файлов для каждого образца выполняли с помощью специального инструмента FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Далее с помощью инструмента TASK (The TRIP Analysis Software Kit, <https://trip.nki.nl/>) были установлены последовательности достоверно выявляемых штрихкодов, а также их нормированные уровни экспрессии и локализация в версиях генома китайского хомячка CriGri-PICRH-1.0 (GCA\_003668045.2) и Cgr1.0 (GCA\_000448345.1). Версия генома CriGri-PICRH-1.0 характеризуется наличием очень протяженных последовательностей, соответствующих всем ожидаемым хромосомам клеток CHO, и поэтому именно эта версия генома была использована в работе как основная, тогда как версия генома Cgr1.0 была ранее использована для картирования типов хроматина в клетках CHO-K1 (Feichtinger et al., 2016). Для определения наиболее «надежных» (далее – отфильтрованных) трансгенов дополнительно использовали следующие параметры инструмента TASK: norm >= 5, reads\_r >= 10, freq1\_r > 0.60. Данные по типам хроматина были взяты для временной точки Tr0, соответствующей 4 ч культивирования клеток (<https://cho-epigenome.boku.ac.at/JB/>). Позиционная весовая матрица для геномных последовательностей, перекрывающих сайты инсерции трансгенов, была построена с помощью специального приложения pLogo (<https://plogo.uconn.edu/>) (O'Shea et al., 2013).

## Результаты и обсуждение

Чтобы изучить активность нескольких промоторных элементов в разных локальных окружениях хроматина в культивируемых клетках CHO, на основе транспозона piggyBac были сконструированы штрихкодированные модельные трансгены, несущие ген устойчивости к пуромицину (далее – *Puro<sup>R</sup>*) и ген улучшенного зеленого флуоресцентного белка eGFP под контролем четырех следующих промоторов: 1) промотора гена *PGK* мыши (*mPGK*), использованного ранее для аналогичного исследования на культивируемых мышечных эмбриональных стволовых клетках (Akhtar et al., 2013); 2) промотора гена *PGK* китайского хомячка, гомологичного промотору *mPGK*; 3) полноразмерного (long) и 4) усеченного (short) вариантов промотора гена *EF-1 $\alpha$*  китайского хомячка (Running Deer, Allison, 2004; Orlova et al., 2014; Wang et al., 2017) (рис. 1, А). При этом в конструкциях с каждым отдельным промотором непосредственно перед 18-буквенным штрихкодом присутствовал также специфический 5-буквенный мотив (промоторный индекс), позволяющий одновременно использовать все четыре штрихкодированных модельных трансгена в одном эксперименте (Gisler et al., 2019) (см. рис. 1, А). Полученные штрихкодированные плазмидные библиотеки с промоторами long *EF-1 $\alpha$* , short *EF-1 $\alpha$* , *mPGK* и *PGK* были смешаны в молярных пропорциях 7:7:7:1. Меньшая доля конструкции с промотором *PGK* объясняется ее использованием в данном эксперименте в качестве контроля. Мы также использовали эту конструкцию для получения стабильных трансгенных популяций клеток CHO при селекции пуромицином (результаты исследования будут сообщены отдельно), и нам представлялось полезным иметь в будущем техническую возможность для корректного сравнения данных для таких разных трансгенных популяций.

Культивируемые клетки CHO (сублинии CHO-S) были котрансфицированы вышеописанной смесью модельных трансгенов, а также плазмидой, кодирующей транспозазу piggyBac. Спустя 72 ч после трансфекции экспрессия белка eGFP наблюдалась примерно в 40 % клеток (см. рис. 1, Б). После этого трансфицированные клетки культивировали в отсутствие какой-либо селекции еще 25 дней с целью размножить трансгенные клетки и параллельно избавиться от молекул плазмидной ДНК, которые могут загрязнить интересующие нас данные. Действительно, в результате в популяции наблюдались множественные клоны трансгенных клеток (см. рис. 1, Б).

Из полученной поликлональной популяции клеток были выделены геномная ДНК и тотальная РНК, на основе которых определены геномные локализации и нормированные уровни экспрессии штрихкодированных трансгенов. Всего в трансгенной популяции выявлен 641 уникально штрихкодированный и картированный в геноме трансген. Эти трансгены присутствовали в более-менее ожидаемых количествах на всех хромосомах клеток CHO (см. рис. 1, В). Анализ геномных последовательностей, перекрывающих сайты инсерции трансгенов, выявил их АТ-обогащенность, а также присутствие центрального мотива taa (см. рис. 1, Г) – черты, характерные для транспозона piggyBac (Fraser et al., 1996; Li et al., 2013; Chen Q. et al., 2020).



**Рис. 1.** Стабильная интеграция модельных *riggyBac*-трансгенов в геном культивируемых клеток CHO.

А – схема штрихкодированных репортерных конструкций, использованных в работе. 5'-TR и 3'-TR – обращенные концы транспозона *riggyBac*; IRES – участок внутренней посадки рибосомы; PI – промоторный индекс; BC – штрихкод; PAS – сигнал полиаденилирования. Б – клетки CHO спустя 3 дня и 28 дней после трансфекции. В – распределение всех уникально картированных трансгенов по хромосомам китайского хомячка. Красными и синими черточками обозначены встройки трансгенов по плюс- и минус-цепи ДНК соответственно. Г – анализ мотивов генома, по которым произошла интеграция всех уникально картированных трансгенов. Позиции от +1 до +4 соответствуют последовательности, которая при встройке транспозона *riggyBac* дублируется и фланкирует интегрированный трансген.

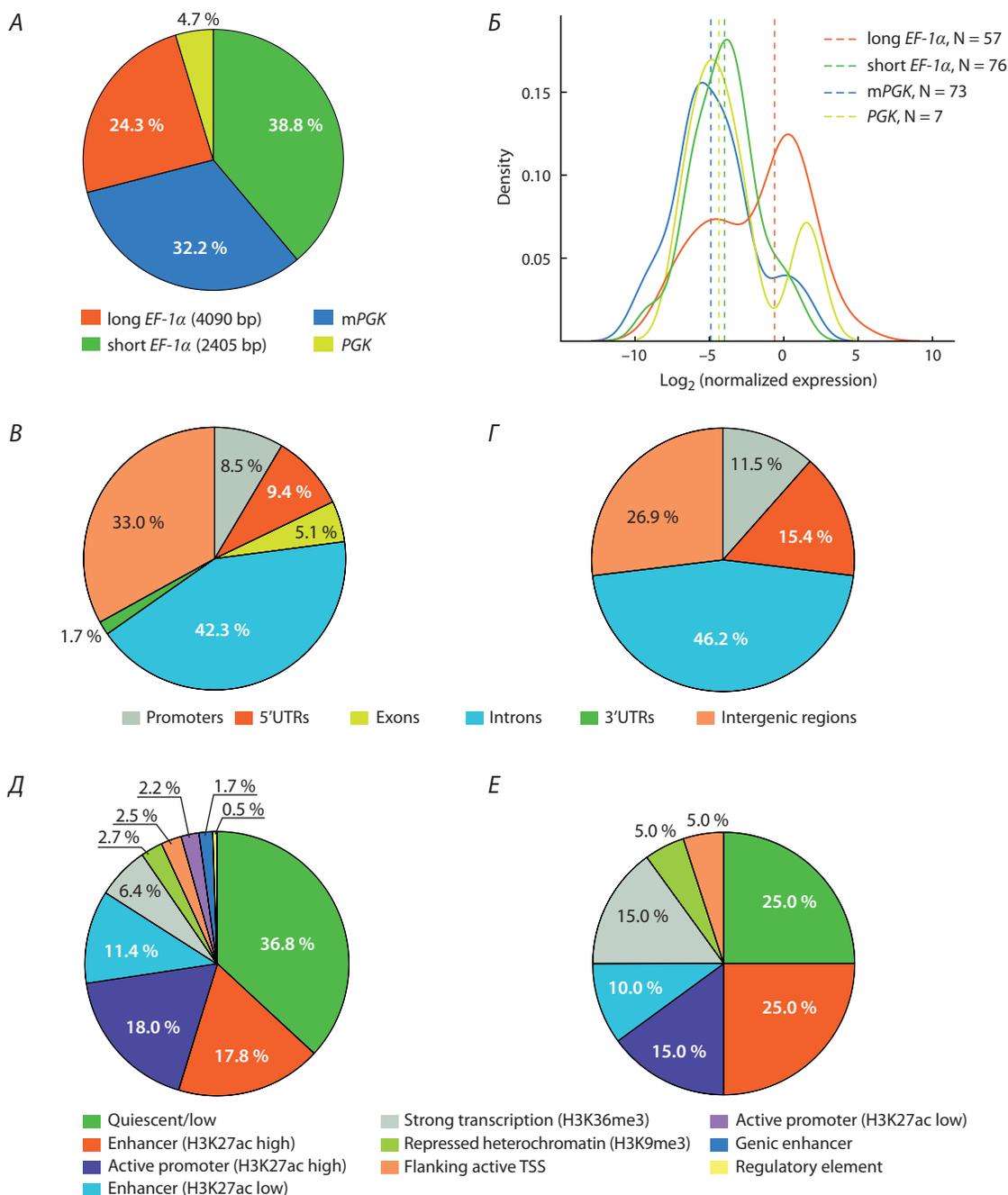
Среди выявленных трансгенов 38.8 % оказались с промоторным индексом *taca*a (соответствующим промотору *short EF-1 $\alpha$* ), 24.3 % – с промоторным индексом *ccgag* (соответствующим промотору *long EF-1 $\alpha$* ), 32.2 % – с промоторным индексом *ctagt* (соответствующим промотору *mPGK*), 4.7 % – с промоторным индексом *agctc* (соответствующим промотору *PGK* китайского хомячка) (рис. 2, А).

Анализ активности репортерных конструкций, находящихся под контролем четырех разных промоторов, выявил наличие большого числа молчащих (т. е. транскрипционно неактивных) трансгенов с каждым промотором (табл. 2), что, наиболее вероятно, связано с отсутствием селекции антибиотиком при получении поликлональной популяции трансгенных клеток CHO.

Сравнение активности промоторов среди отфильтрованных экспрессирующихся трансгенов (144 шт.) показало, что основная часть высокоактивных репортерных конструкций находится под контролем полноразмерного варианта промотора гена *EF-1 $\alpha$*  (см. рис. 2, Б, табл. 2). В

частности, среди 10 % наиболее активных отфильтрованных трансгенов доли промоторов распределяются следующим образом: *long EF-1 $\alpha$*  – 70 %, *mPGK* – 20 %, *short EF-1 $\alpha$*  – 10 %, *PGK* – 0 %. Надо отметить, что из-за малого количества исследованных трансгенов с промотором *PGK* результаты по его активности носят очень предварительный характер.

Две трети всех трансгенов встроились в геном клеток CHO внутри генов (которые были определены как –1000 п. н. от дистального сайта инициации транскрипции до сайта терминации транскрипции), причем преимущественно в интроны (42.3 %), промоторы (8.5 %) и 5'-некодирующие области (9.4 %) (см. рис. 2, В). Необходимо отметить, что промоторы были определены как районы от –1000 до +100 п. н. относительно сайтов инициации транскрипции. Схожие паттерны интеграции трансгенов на основе транспозона *riggyBac* были описаны ранее для культивируемых клеток других видов (Ding et al., 2005; Wilson et al., 2007; Galvan et al., 2009; Li et al., 2013). Анализ 10 % наиболее активных отфильтрованных



**Рис. 2.** Характеристика исследованных трансгенов.

А – распределение всех выявленных трансгенов по исследуемым промоторам *long EF-1α*, *short EF-1α*, *mPGK* и *PGK*. Б – сравнение активностей промоторов на основе данных для отфильтрованных 144 экспрессирующихся трансгенов (см. Материалы и методы). Штриховыми вертикальными линиями показаны медианные значения нормированной экспрессии для каждого промотора. В, Г – распределение всех выявленных трансгенов (В) и 10% наиболее активных отфильтрованных трансгенов (Г) по элементам генов (промоторам, 5'- и 3'-некодирующим районам, экзонам, интронам), а также межгенным промежуткам. Д, Е – распределение всех трансгенов (Д) и 10% наиболее активных отфильтрованных трансгенов (Е) по типам хроматина, определенным ранее для клеток CHO-K1 (Feichtinger et al., 2016).

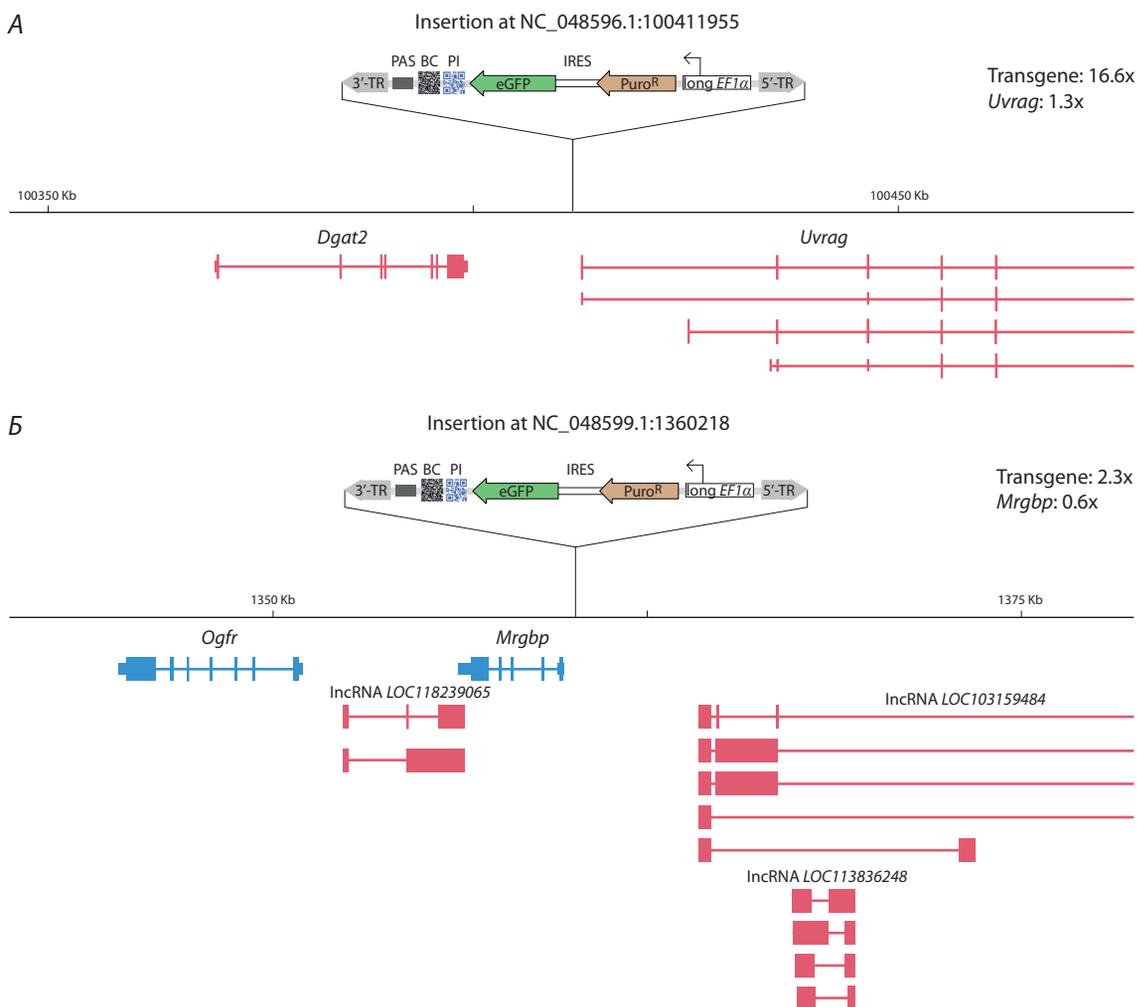
репортерных конструкций (21 шт.) выявил увеличение доли трансгенов как раз именно в 5'-некодирующих областях генов, промоторах и интронах (в 1.6, 1.4 и 1.1 раза соответственно) (см. рис. 2, Г). Интересно, что трансгены чаще локализовались ближе к началу, чем к концу генов. Значения медианных расстояний от позиции локализации трансгена в геноме до ближайших сайтов инициации и терминации транскрипции оказались равны соответствен-

но 11.4 и 20.2 т. п. н. для полного набора исследованных репортерных генов (641 шт.). При этом для 10% наиболее активных отфильтрованных трансгенов (21 шт.) такие значения были равны 6.6 и 17.8 т. п. н. соответственно.

Для изучения влияния локального окружения хроматина на активность репортерных генов были использованы ранее опубликованные данные по распределению 11 типов хроматина в геноме клеток CHO-K1 (Feichtinger

**Таблица 2.** Сравнительная активность исследованных промоторов

Промотор	Количество трансгенов	Доля трансгенов, %		Значение медианной активности промотора, отн. ед.
		молчащих	активных	
long <i>EF-1α</i>	156	42.31	57.69	18.93
short <i>EF-1α</i>	249	60.24	39.76	1.84
mPGK	206	53.88	46.12	1
PGK	30	66.67	33.33	1.42
Всего	641	54.13	45.87	



**Рис. 3.** Примеры геномной локализации трансгенов из числа наиболее активных (А) и со средней транскрипционной активностью (Б).

Схемы встроек трансгенов не масштабированы относительно геномной ДНК. Активности трансгенов и ближайших к ним генов указаны относительно среднего уровня экспрессии всех изученных штрихкодированных трансгенов и всех эндогенных генов соответственно.

et al., 2016). Изначально эти данные были определены для версии генома китайского хомячка, отличной от использованной для всех описанных выше анализов (подробности см. в разделе Материалы и методы), поэтому только для 595 из 641 трансгена удалось определить типы хроматина, перекрывающие позиции их локализации в геноме. При этом лишь 39.5 % трансгенов оказались расположены в неактивных типах хроматина “Quiescent/

low”, “Repressed heterochromatin (H3K9me3)” и “Polycomb repressed regions (H3K27me3)”, которые суммарно покрывают более 88 % генома клеток китайского хомячка (Feichtinger et al., 2016). Остальные 60.5 % трансгенов были выявлены в различных активных типах хроматина (см. рис. 2, Д).

Наиболее активные трансгены чаще обнаруживались в районах генома, ассоциированных с активными типами

хроматина “Enhancer (H3K27ac high)”, “Strong transcription (H3K36me3)” и “Flanking active TSS”, а также с неактивным типом хроматина “Repressed heterochromatin (H3K9me3)” (см. рис. 2, E). Последнее достаточно неожиданное наблюдение, возможно, связано с тем, что типы хроматина были определены для другой (отличной от использованной в данной работе) сублинии клеток СНО.

Поскольку, как отмечено выше, две трети всех трансгенов локализовались внутри генов (см. рис. 2, B), стоит отметить, что встройка трансгена даже в какой-то важный ген, вероятно, лишь незначительно сказывается на жизнеспособности клеток. Это связано как с тем, что далеко не каждая встройка чужеродной последовательности в пределах гена существенно нарушает его функцию, так и с наличием второй нативной копии этого гена в геноме. Вместе это обеспечивает выживание таких трансгенных клеток в поликлональной популяции. Шансы повредить сразу оба аллеля гена в использованном нами экспериментальном подходе ничтожно малы. Для этого сразу два трансгена должны встроиться в геном одной и той же клетки, причем в разные аллели одного и того же гена. Таким образом, выявленные в данной работе геномные позиции активных трансгенов вполне могут претендовать на рассмотрение в качестве перспективных сайтов для направленной интеграции целевых биотехнологических трансгенов, даже если они располагаются внутри активных генов (рис. 3).

## Заключение

В полученной при отсутствии селекционного давления поликлональной популяции культивируемых клеток СНО более половины модельных трансгенов, стабильно интегрированных в геном, оказались транскрипционно неактивными. По сравнению с полным набором трансгенов, наиболее активные трансгены локализовались в 1.6 и 1.4 раза чаще в районах промоторов и 5'-некодирующих областей генов соответственно. Также наиболее активные трансгены локализовались в 2.3 и 1.4 раза чаще в транскрипционно активных типах хроматина “Strong transcription (H3K36me3)” и “Enhancer (H3K27ac high)” соответственно. Трансгены, содержащие полноразмерный промотор гена *EF-1a* китайского хомячка, оказались в среднем наиболее активными. При этом медианная активность короткого варианта промотора гена *EF-1a* была в 10 раз ниже медианной активности полноразмерного промотора этого гена (см. табл. 2). Это можно объяснить наличием важных сайтов связывания транскрипционных факторов в полноразмерной версии промотора гена *EF-1a*. Геномные сайты локализации наиболее активных встроек модельных трансгенов могут представлять интерес для дальнейших экспериментов как перспективные позиции для направленной интеграции целевых биотехнологических конструкций.

## Список литературы

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., Kreml P.M., Schmid 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- $\alpha$ . *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 Seleksii = 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-1 $\alpha$  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-1 $\alpha$  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](https://orcid.org/0000-0003-0469-0371)  
A.A. Ogienko [orcid.org/0000-0002-0896-1899](https://orcid.org/0000-0002-0896-1899)  
A.V. Pindyurin [orcid.org/0000-0001-6959-0641](https://orcid.org/0000-0001-6959-0641)  
E.S. Omelina [orcid.org/0000-0002-2189-5101](https://orcid.org/0000-0002-2189-5101)

**Благодарности.** Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (Соглашение № 075-15-2021-1086, контракт RF---193021X0015, 15.ИП.21.0015).

Авторы благодарны А.В. Таранину (Институт молекулярной и клеточной биологии СО РАН), В.В. Верхуше (Медицинский колледж им. Альберта Эйнштейна, Бронкс, Нью-Йорк, США), В.С. Фишману (Институт цитологии и генетики СО РАН) за предоставление культивируемых клеток линии CHO-S, плазмиды rRP[Exp]-mCherry-CAG>hyPBase и помощь в извлечении данных по распределению типов хроматина в геноме китайского хомячка.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 13.09.2023. После доработки 21.09.2023. Принята к публикации 27.09.2023.