

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

УДК 577.218

ДРОЖЖЕВОЙ БЕЛОК NHP6A СВЯЗЫВАЕТСЯ
С КОРОТКИМИ GC-БОГАТЫМИ ГЕНАМИЕ.С. Герасимов¹, Н.С. Герасимова², А.Л. Козлова¹, В.М. Студитский^{2,3,*}¹Кафедра молекулярной биологии и ²кафедра биоинженерии, биологический факультет, Московский государственный университет имени М.В. Ломоносова, Россия, 119234, г. Москва, Ленинские горы, д. 1, стр. 12;³Cancer Epigenetics Program, Fox Chase Cancer Center, Cottman avenue 333, Philadelphia, 19111 Pennsylvania, USA

*e-mail: vasily.studitsky@fccc.edu

Nhp6A — это небольшой негистоновый хромосомный белок дрожжей, неспецифично связывающий ДНК. Этот белок присутствует во многих промоторных и транскрибируемых областях генома и участвует в регуляции транскрипции. Недавно была показана роль Nhp6A в процессе дестабилизации структуры нуклеосом, что может объяснить его расположение в регуляторных участках. Тем не менее, его функция в кодирующих областях остается неизвестной. В настоящей работе с целью поиска механизма действия Nhp6A нами были изучены гены, связанные с белком по всей длине, включая область открытой рамки считывания. Мы показали, что Nhp6A преимущественно связывается с кодирующими областями коротких GC-богатых генов дрожжей. Наблюдаемое взаимодействие не обусловлено непосредственно высоким содержанием GC-пар в данных локусах ДНК, что позволяет предположить существование специфического для данной группы регуляторного механизма с участием Nhp6A. Так как достаточно много изученных генов сохраняют ряд характерных для предкового бактериального генома черт, мы полагаем, что данная группа скорее относится к “древним”. Таким образом, возможно, обнаруженная особенность распределения Nhp6A связана с сохранением механизмов регуляции транскрипции генов, возникших рано в ходе эволюции.

Ключевые слова: Nhp6A, белки HMG, *Saccharomyces cerevisiae*, связывание ДНК, регуляция транскрипции, эволюционная геномика, GC-состав, предпочтение кодонов

Хранение и реализация заложенной в ДНК генетической информации обеспечиваются действием многочисленных белков. В наибольшем количестве в ядре эукариотических организмов содержатся гистоны — небольшие белки, обеспечивающие компактную укладку ДНК в хроматин [1]. Большое влияние на структуру хроматина и функционирование генов оказывают также различные негистоновые белки. К ним относят внутриядерные ферменты и белковые факторы, необходимые для работы генетического аппарата клетки. В высоких концентрациях в ядрах присутствуют негистоновые белки, которые относят к так называемой группе белков с высокой подвижностью (HMG — high mobility group) [2]. Они быстро перемещаются при электрофорезе в полиакриламидном геле за счёт небольшой молекулярной массы и обладают характерной структурой, называемой HMG-box (HMGB). Суммарное содержание HMG-белков в ядрах клеток примерно в десять раз меньше, чем содержание гистонов.

К таким белкам относится и дрожжевой негистоновый хромосомный белок 6A (non-histone chromosomal protein 6A — Nhp6A) — небольшой полипептид с молекулярной массой 10810 Да, состоящий из 93 аминокислотных остатков [3]. Nhp6A связывается с участками ДНК длиной около 11 п.н. и не

имеет выраженной специфичности к ее последовательности и составу [4–6]. Nhp6A имеет типичную L-образную структуру HMGB-домена: три положительно заряженные альфа-спирали укладываются в малую бороздку ДНК, а неструктурированный N-конец, огибая двойную спираль, укладывается в большую [7].

Ген *NHP6A* имеет один паралог *NHP6B*, кодирующий идентичный ему на 87% белок длиной в 101 аминокислотный остаток [8]. Он представлен в клетках в заметно меньшем количестве, поэтому большинство исследований сфокусировано на основной изоформе.

Nhp6A обнаруживается в составе некоторых мультимолекулярных белковых комплексов и участвует в регуляции транскрипции. Одновременное нокаутирование генов *NHP6A* и *NHP6B* приводит к замедленному росту дрожжей при температуре 30°C и невозможности культивирования клеток при 38°C, что свидетельствует о функциональной значимости изоформ белка для выживания дрожжевых клеток в стрессовых условиях [9]. Важная роль Nhp6A в регуляции экспрессии генов была продемонстрирована в работах по индукции генов, транскрибируемых РНК-полимеразой 2. Так, в штаммах дрожжей с мутациями в *NHP6A* и *NHP6B* наблюдается подавление экспрессии некоторых

генов [10]. Тем не менее, механизмы действия Nhrp6 мало изучены.

Полногеномный анализ распределения Nhrp6A по дрожжевому геному демонстрирует, что белок присутствует на многих промоторах и транскрибируемых участках генов *in vivo* [11]. Исследования *in vitro* показали, что Nhrp6A участвует в процессе дестабилизации структуры нуклеосом. Это может объяснить его расположение в регуляторных участках генома [12]. Нами было показано, что белок может облегчать транскрипцию хроматина *in vitro* [13], однако его функция в кодирующих областях *in vivo* остается неизвестной. С целью поиска механизма влияния Nhrp6A на транскрипцию кодирующих областей генома нами были изучены гены, связанные с белком на всей протяженности открытой рамки считывания.

Методы

В работе были использованы данные о связывании белка Nhrp6A с участками генома *Saccharomyces cerevisiae*, полученные Дуэлл и соавт. [11], а также геном и аннотация генов *S. cerevisiae* версии R64-1-1, доступные в базе данных Ensembl. Все этапы анализа проводились только для нуклеотидных последовательностей 16 хромосом ядерного генома, содержащих 6692 белок-кодирующих гена. Список генов, с которыми Nhrp6A связывается в кодирующей области (далее — Nhrp6A⁺-гены), состоял из 176 наименований.

Контрольная выборка генов, распределение длины кодирующей части которых совпадает с таковым для Nhrp6A⁺-генов, была получена следующим образом. Гистограмма распределения длины кодирующих участков Nhrp6A⁺-генов делилась на равные интервалы длиной 50 п.н., затем определялось число генов k в каждом интервале длин i . Для каждого интервала длин i из всех генов дрожжевого генома, попадающих в заданный интервал по длине, случайным образом отбиралось ровно k_i генов. Размер выборки при таком способе генерации совпадает с количеством Nhrp6A⁺-генов. Распределение длины кодирующей части генов из одной такой выборки приведено на рисунке (А). Для исключения влияния отдельных генов было сделано 100 независимых случайных выборок. Результаты были усреднены для 100 выборок.

Исходный код программы на языке Python, использованной для получения выборок, проведения анализа и построения рисунков, доступен на веб-сервисе GitHub по ссылке <https://github.com/jalgard/MUBSB>.

Результаты

Дрожжевой белок Nhrp6A связывается с различными участками генома, преимущественно с промоторами и транскрибируемыми участками. Полногеномный анализ связывания Nhrp6A в геноме дрожжей показал, что белок связывается с кодиру-

ющими участками некоторых генов (Nhrp6A⁺-генов). В работе Дуэлл и соавт. [11] было обнаружено 176 таких открытых рамок считывания.

Нами было показано, что распределение длин кодирующих участков Nhrp6A⁺-генов существенно отличается от такового для всех генов ядерного дрожжевого генома (рисунок, А) (р-значение U-критерия Манна-Уитни $7,4 \cdot 10^{-58}$). При этом Nhrp6A⁺-гены представляют собой фракцию коротких генов. Так, средняя длина дрожжевого гена (кодирующей области) составляет 1377 п.н. (медиана распределения 1107 п.н.), а средняя длина гена из выборки Nhrp6A⁺-генов — 408 п.н. (медиана 345 п.н.).

Для проверки гипотезы о том, что данные гены обладают и другими отличительными свойствами, из всего ядерного генома дрожжей мы получили контрольную выборку коротких генов, похожую по распределению длины на исследуемую группу (производили 100 независимых случайных повторностей, чтобы устранить возможное влияние отдельных случайно взятых генов).

Оказалось, что в отличие от случайно взятых коротких генов дрожжевого генома, Nhrp6A⁺-гены имеют более высокое содержание GC-пар. Распределение GC-состава генов для сравниваемых выборок и для генов всего генома приведено на рисунке (Б) из которого хорошо видно, что случайно выбранные короткие гены не сильно отличаются по GC-составу от всех генов генома дрожжей, в отличие от Nhrp6A⁺-генов (р-значение U-критерия Манна-Уитни $9,0 \cdot 10^{-16}$ для сравнения Nhrp6A⁺-генов со случайной выборкой коротких генов). Среднее значение GC-содержания генома дрожжей — 38% (сплошная горизонтальная линия), тогда как медианное GC-содержание у генов ядерного генома несколько выше и составляет 40%. При этом для генов, связанных с Nhrp6A, данное значение существенно выше — 46%. Стоит отметить, что данное свойство не является само по себе фактором, определяющим связывание Nhrp6A с кодирующей областью гена. Анализ GC-содержания участков генома, с которыми связывается белок Nhrp6A, показал, что среднее значение GC-содержания в них составляет 40,7%, что почти не отличается от такового для генов ядерного генома дрожжей в целом. Кроме того, исследования *in vitro* свидетельствуют в пользу того, что Nhrp6A не имеет выраженной специфичности к последовательности и GC-составу ДНК [6].

Интересно, что Nhrp6A⁺-гены также находятся в существенно более GC-богатом окружении. Анализ GC-состава прилегающих к генам последовательностей длиной 200 п.н. приведен на рисунке (В). Из рисунка видно, что среднее содержание GC-пар в этих участках соответствует среднему значению для локуса генома, с которым связывается Nhrp6A, тогда как среднее содержание GC-пар в областях, непосредственно прилегающих к генам дрожжевого генома, в общем случае существенно ниже.

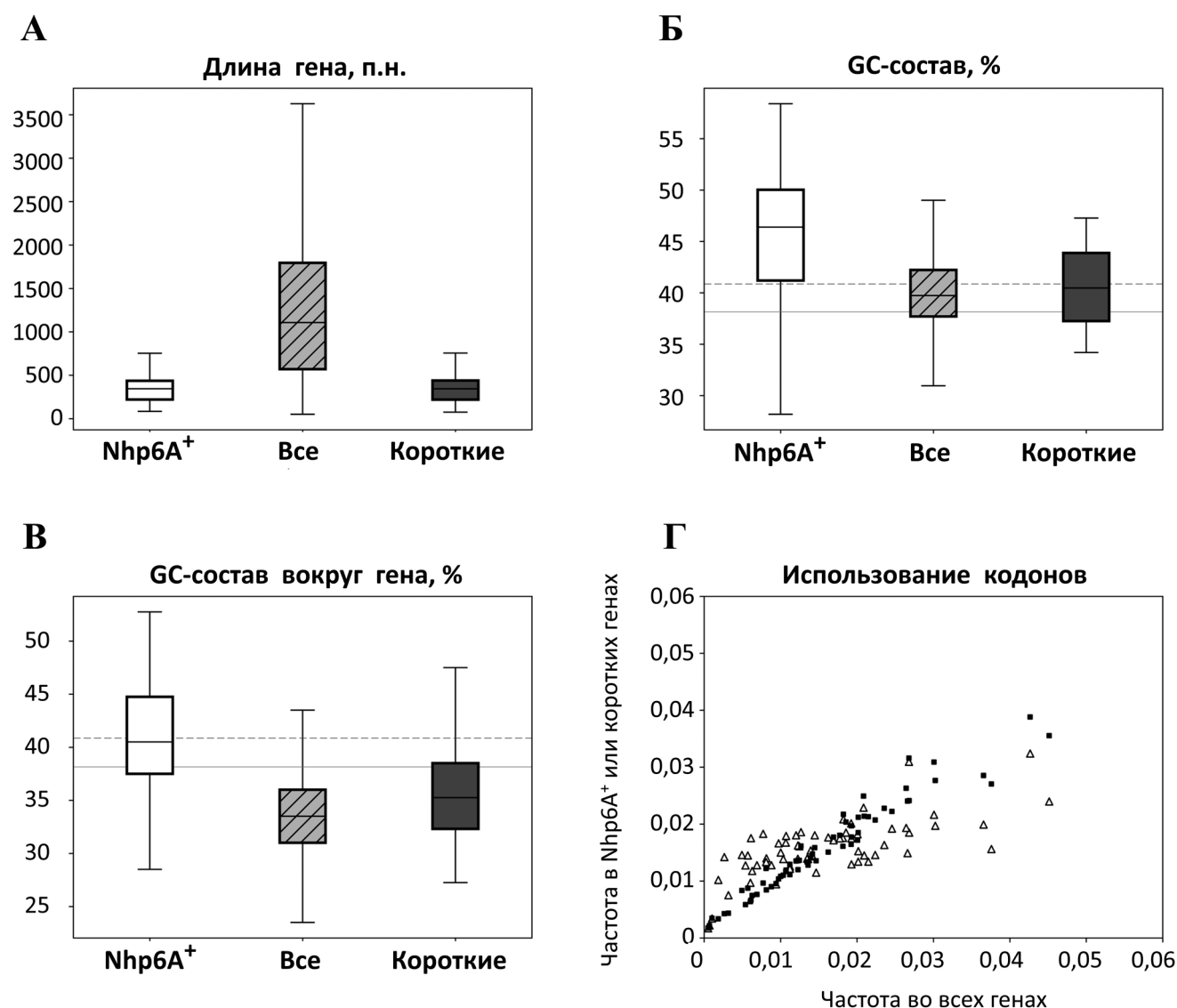


Рисунок. Особенности генов, связанных с дрожжевым белком Nhr6A. Анализ различных параметров генов, для которых показано взаимодействие с Nhr6A на протяжении всей кодирующей области (Nhr6A⁺-гены) [11], всех генов ядерного генома дрожжей ("Все") и соответствующие усредненные данные по 100 независимым выборкам коротких генов ("Короткие"). **А.** Распределение длин генов. **Б.** Распределение содержания GC-пар в кодирующих областях генов. Сплошная горизонтальная линия показывает средний GC-состав 16 хромосом ядерного генома *Saccharomyces cerevisiae*, пунктирная — участков генома, с которыми связывается белок Nhr6A (по данным о связывании белка Nhr6A из [11]). **В.** Распределение содержания GC-пар на участках генома длиной 200 п.н. вокруг генов. **Г.** Частота использования кодонов для Nhr6A⁺-генов (треугольники) и усредненная частота использования кодонов для 100 случайных выборок коротких генов (квадраты). Каждый символ соответствует кодону, координата по оси абсцисс соответствует частоте использования данного кодона во всех генах ядерного генома дрожжей, а координата по оси ординат — частоте использования данного кодона для генов из соответствующей выборки

Одним из наиболее интересных свойств генов, связанных с Nhr6A, является то, что они сильно отличаются по частоте использования кодонов от средних значений для всего генома дрожжей (рисунок, Г). На графике кодоны, которые относительно чаще используются во взятой выборке генов, чем в целом по генам ядерного генома дрожжей, оказываются над диагональю, а кодоны, которые используются относительно реже — под диагональю. Из графика видно, что кодоны для выборки случайных коротких генов расположены в основном на диагонали, тогда как кодоны для выборки Nhr6A⁺-генов — нет. При этом относительно реже

используемые кодоны чаще используются в выборке Nhr6A⁺-генов, а более часто используемые — реже. Аналогичной зависимости не наблюдается, если разделить кодирующую область на триплеты со сдвигом рамки +1 или +2, в таком случае триплеты обеих выборок ведут себя случайным образом (данные не приведены).

Обсуждение

Таким образом, гены, связанные с Nhr6A в кодирующей области, являются короткими, GC-богатыми и имеют отличную от среднего статистику использования кодонов. Ряд исследований пока-

зывает, что гены, имеющие сходный эволюционный “возраст” (например, древние или новые) имеют сходный GC-состав, длину белкового продукта и статистику использования кодонов [14]. Мы предполагаем, что данный сценарий реализуется и с Nhp6A⁺-генами, по-видимому, объединенными также единым механизмом регуляции их экспрессии с участием белка Nhp6A.

Анализ распределения Nhp6A по геному дрожжей показал, что участки генома, связанные с данным белком, имеют лишь немного более высокое содержание GC-пар, тогда как содержание GC-пар в кодирующих областях Nhp6A⁺-генов, а также в прилегающих к ним участках значительно выше. Существуют данные о том, что *in vitro* белок Nhp6A предпочтительно связывается с последовательностями, обогащенными динуклеотидом GT [5], однако в нашем исследовании не было выявлено данного эффекта. По-видимому, наблюдаемое взаимодействие с кодирующей областью генов само по себе не обусловлено высоким содержанием GC-пар в данных локусах ДНК. Можно предположить, что имеет место специфический для данной группы генов регуляторный механизм с участием Nhp6A.

К сожалению, анализ функциональной роли данных генов крайне затруднителен ввиду того,

что больше половины из них аннотированы в базе данных *Saccharomyces* Genome Database (SGD) как рамки считывания с неизвестной функцией. Однако среди генов, для которых известна функция, достаточно много митохондриальных, что может указывать на сохранение этими генами ряда черт, характерных для предкового бактериального генома. Поэтому мы предполагаем, что данная группа генов скорее относится к “древним”.

Нужно отметить, что у самого гена *NHP6A* содержание GC-пар в кодирующей области составляет 46% и он также окружен участками с повышенным GC-содержанием (42%), таким образом, вероятнее всего, относится к той же эволюционной группе генов. При этом его паралог *NHP6B*, определенно относящийся к более “новым” генам, возникшим во время полногеномной дупликации [15], имеет содержание GC-пар в кодирующей области 44% и окружен участками с GC-содержанием 41%. Это наблюдение хорошо согласуется с гипотезой о “древности” группы генов, связывающих Nhp6A в кодирующей области.

Список генов, связанных с Nhp6A в кодирующей области, любезно предоставлен Ноа Доуэлл. Работа выполнена при финансовой поддержке Российского научного фонда (проект № 14-24-00031).

СПИСОК ЛИТЕРАТУРЫ

1. Kornberg R.D., Thomas J.O. Chromatin structure; oligomers of the histones // *Science*. 1974. Vol. 184. N 4139. P. 865–868.
2. Reeves R. HMG nuclear proteins: linking chromatin structure to cellular phenotype // *Biochim. Biophys. Acta*. 2010. Vol. 1799. N 1–2. P. 3–14.
3. Stillman D.J. Nhp6: a small but powerful effector of chromatin structure in *Saccharomyces cerevisiae* // *Biochim. Biophys. Acta*. 2010. Vol. 1799. N 1–2. P. 175–180.
4. Ruone S., Rhoades A.R., Formosa T. Multiple Nhp6 molecules are required to recruit Spt16-Pob3 to form yFACT complexes and to reorganize nucleosomes // *J. Biol. Chem*. 2003. Vol. 278. N 46. P. 45288–45295.
5. Allain F.H., Yen Y.M., Masse J.E., Schultze P., Dieckmann T., Johnson R.C., Feigon J. Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding // *EMBO J*. 1999. Vol. 18. N 9. P. 2563–2579.
6. Coats J.E., Lin Y., Rueter E., Maher L.J., III, Rasnik I. Single-molecule FRET analysis of DNA binding and bending by yeast HMGB protein Nhp6A // *Nucleic Acids Res*. 2013. Vol. 41. N 2. P. 1372–1381.
7. Masse J.E., Wong B., Yen Y.M., Allain F.H., Johnson R.C., Feigon J. The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding // *J. Mol. Biol*. 2002. Vol. 323. N 2. P. 263–284.
8. Bustin M. Revised nomenclature for high mobility group (HMG) chromosomal proteins // *Trends Biochem. Sci*. 2001. Vol. 26. N 3. P. 152–153.
9. Costigan C., Kolodrubetz D., Snyder M. NHP6A and NHP6B, which encode HMG1-like proteins, are candidates for downstream components of the yeast SLT2 mitogen-activated protein kinase pathway // *Mol. Cell Biol*. 1994. Vol. 14. N 4. P. 2391–2403.
10. Yu Y., Eriksson P., Stillman D.J. Architectural transcription factors and the SAGA complex function in parallel pathways to activate transcription // *Mol. Cell Biol*. 2000. Vol. 20. N 7. P. 2350–2357.
11. Dowell N.L., Sperling A.S., Mason M.J., Johnson R.C. Chromatin-dependent binding of the *S. cerevisiae* HMGB protein Nhp6A affects nucleosome dynamics and transcription // *Genes Dev*. 2010. Vol. 24. N 18. P. 2031–2042.
12. Xin H., Takahata S., Blanksma M., McCullough L., Stillman D.J., Formosa T. yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement // *Mol. Cell*. 2009. Vol. 35. N 3. P. 365–376.
13. Hsieh F.K., Kozlova A.L., Gerasimova N.S., Kotova E.Yu., Formosa T., Studitsky V.M. Role of the Nhp6 protein in *in vitro* transcription through the nucleosome // *Moscow Univ. Biol. Sci. Bull*. 2017. Vol. 72. N 4. P. 218–221.
14. Prat Y., Fromer M., Linial N., Linial M. Codon usage is associated with the evolutionary age of genes in metazoan genomes // *BMC Evol. Biol*. 2009. Vol. 9:285.
15. Wolfe K.H., Shields D.C. Molecular evidence for an ancient duplication of the entire yeast genome // *Nature*. 1997. Vol. 387. N 6634. P. 708–713.

Поступила в редакцию
16.01.2018

Принята к печати
15.03.2018

MOLECULAR BIOLOGY

YEAST PROTEIN NHP6A BINDS WITH SHORT GC-RICH GENES

E.S. Gerasimov¹, N.S. Gerasimova², A.L. Kozlova¹, V.M. Studitsky^{2,3,}*¹*Department of Molecular Biology and* ²*Department of Bioengineering, School of Biology, Lomonosov Moscow State University, 1–12 Leninskiye Gory, Moscow, 119234, Russia;*³*Cancer Epigenetics Program, Fox Chase Cancer Center, Cottman Avenue 333, Philadelphia, 19111 Pennsylvania, USA***e-mail: vasily.studitsky@fccc.edu*

Nhp6A is a small non-histone chromosomal yeast protein that binds DNA nonspecifically. This protein is present at many promoters and transcribed regions of genome and is involved in regulation of transcription. Recently, Nhp6A was shown to participate in destabilization of the nucleosomal structure. This may explain its location in regulatory sites, but its function in the coding regions remains unknown. In the present work, in order to reveal the mechanism of action of Nhp6A, we have studied genes associated with the protein along the entire length, including the open reading frame. We have shown that Nhp6A predominantly binds to the coding regions of short GC-rich yeast genes. The observed interaction is not associated directly with the high content of GC-pairs in these DNA loci, so we can propose a specific regulatory mechanism involving Nhp6A for this group. Since a part of genes retain features of the ancestral bacterial genome, we suggest this group as an “ancient”. Presumably, this genomic distribution of Nhp6 is related to the mechanisms of regulation of gene transcription that appeared early in the course of evolution.

Keywords: *Nhp6A, HMG protein, Saccharomyces cerevisiae, DNA binding, regulation of transcription, evolutionary genomics, GC-content, codon usage bias*

Сведения об авторах

Герасимов Евгений Сергеевич — канд. биол. наук, науч. сотр. кафедры молекулярной биологии биологического факультета МГУ. Тел. 8-495-939-46-58; e-mail: jalgard@gmail.com

Герасимова Надежда Сергеевна — канд. биол. наук, ассистент кафедры биоинженерии биологического факультета МГУ. Тел. 8-495-938-22-91; e-mail: gerasimova@mail.bio.msu.ru

Козлова Анастасия Львовна — студентка кафедры молекулярной биологии биологического факультета МГУ. Тел.: 8-495-938-22-91; e-mail: mika.lorens@yandex.ru

Студитский Василий Михайлович — докт. биол. наук, гл. науч. сотр. кафедры биоинженерии биологического факультета МГУ; руководитель лаборатории эпигенетики рака Центра исследований рака Фокс Чейз (Филадельфия, США). Тел. 8-495-938-22-91; e-mail: vasily.studitsky@fccc.edu