УДК: 575
DOI: 10.18522/2308-9709-2025-53-7

Исследование CpG-островков в промоторных областях тканеспецифичных генов и генов домашнего хозяйства человека

Слюсарев С.С., Шкурат Т.П.

Южный федеральный университет
344090 Ростов-на-Дону, пр. Стачки 194/1 Академия биологии и медицины, кафедра генетики
slusarev-sergey@yandex.ru, tshkurat@yandex.ru

Аннотация

Представлен сравнительный анализ структуры и распределения CpG-островков в промоторных областях двух крупных групп генов человека — тканеспецифичных и генов домашнего хозяйства. CpG-островки, представляющие собой GC-обогащённые и неметилированные участки ДНК, играют ключевую роль в регуляции транскрипции и поддержании открытой хроматиновой структуры. Проанализировано 740 тканеспецифичных генов (разделённых по уровню Tau-специфичности) и 500 генов домашнего хозяйства. Для областей TSS200 и TSS1500 были определены наличие и количество CpG-островков, их длина и расстояние до сайта начала транскрипции. Поиск островков выполнялся с использованием Cpgplot, а статистические различия оценивались χ^2-критерием, тестом Краскела–Уоллиса и Post-hoc анализом. Исследование выявило глубокие и статистически значимые различия между группами генов. Домашние гены значительно чаще содержали CpG-островки: 92% — в TSS200 и 33.4% — в TSS1500, тогда как у высокоспецифичных тканевых генов эти показатели составляли 15.1% и 7.9% соответственно. Длина CpG-островков также оказалась максимальной у генов домашнего хозяйства (в среднем 723 п.н. в TSS200), уменьшаясь по мере роста тканевой специфичности. Найдена выраженная закономерность: чем более специфично экспрессируется ген, тем реже в его промоторных областях обнаруживаются CpG-островки и тем они короче. Это подтверждает модель, согласно которой стабильная экспрессия генов домашнего хозяйства поддерживается через CpG-островки и связанную с ними открытую хроматиновую структуру, тогда как тканеспецифичные гены опираются на более жёсткие и локализованные механизмы эпигенетической регуляции. В целом статья

демонстрирует, что распределение CpG-островков тесно связано с профилем экспрессии генов и отражает фундаментальные различия в организации регуляторных регионов разных функциональных групп.

Ключевые слова: CpG-островки, тканеспецифичные гены, промотор

Введение.

Все гены классифицируются в соответствии с их экспрессией во всех типах клеток и тканей организмов. Гены домашнего хозяйства (housekeeping genes) представляют собой гены, которые экспрессируются почти во всех клетках или тканях организма. Эти гены имеют большое значение для поддержания базовых клеточных функций, таких как обменные процессы и поддержание структуры клетки. Они кодируют белки, необходимые для основных метаболических путей и функций клетки. Гены домашнего хозяйства имеют высокую степень консервативности и представлены во всех организмах независимо от их эволюционной дальности (Joshi C.J. et al., 2022; Fedorova, L., et al., 2022.). В то же время исследования показывают, что индивидуальный набор генов домашнего хозяйства, даже при условии их стабильной экспрессии и важности для клетки, может значительно различаться у разных организмов.

Тканеспецифичные гены представляют собой гены, которые экспрессируются только в определенных типах тканей или клеток организма. Эти гены играют решающую роль в формировании и поддержании специализированных функций различных тканей, включая сердце, печень, мозг и другие органы. Экспрессия тканеспецифичных генов тесно связана с развитием и функционированием определенных клеток или тканей.

Различия между тканеспецифичными генами и генами домашнего хозяйства можно наблюдать в их промоторных регионах, которые играют важную роль в регуляции экспрессии генов. Одним из изучаемых аспектов является различия в присутствие CpG-островков в промоторах генов домашнего хозяйства и тканеспецифичных генов (J.Zhu et al., 2008; F.Antequera, 2003). У людей около 60% промоторов, расположенных вблизи сайта начала транскрипции гена, содержат CpG-островок (S.Saxonov et al., 2006; Vavouri T., Lehner B.,2012).

Метилирование CpG-островков может влиять на доступность факторов транскрипции к промотору гена, а также на связывание других регуляторных молекул, таких как белки связывания ДНК. Таким образом, метилирование CpG-островков может являться одним из ключевых факторов, определяющих активность гена (Lanata C. M. et al.,2018). Более того, изменения в метилировании CpG-островков могут иметь широкие физиологические последствия. Например,

гиперметилирование CpG-островков в промоторах определенных генов может приводить к их сильному подавлению или полному отсутствию экспрессии, что может быть связано с различными патологическими состояниями, включая рак и неврологические заболевания (A.Bird, 2002)

Исследования также показали, что CpG-островки часто находятся в районах эухроматина и представляют собой "островки" низкого метилирования ДНК. Высокий уровень метилирования CpG-островков обычно связан с гетерохроматином и подавлением транскрипции (Rose N. R., Klose R. J., 2014).

Важно отметить, что CpG-островки могут быть расположены как в ближней (TSS200) области от начала транскрипции, так и в дальней (TSS1500) области. Исследования показывают, что CpG-островки в TSS200 и TSS1500 областях имеют свою уникальную роль в регуляции генной экспрессии. Например, CpG-островки в TSS200 областях могут быть связаны с инициацией транскрипции, в то время как CpG-островки в TSS1500 областях могут влиять на уровень экспрессии гена (Deaton A. M., Bird A., 2011; Chen Y. et al.,2020)

Известно, что CpG-островки могут привлекать белки, которые обладают способностью изменять хроматиновую структуру. Некоторые из этих белков называются метилированными ДНК-связывающими белками (MBD), которые могут связываться с метилированной ДНК и привлекать другие факторы, чтобы изменить структуру хроматина (Springer N. M., Kaeppler S. M., 2005; Du Q. et al., 2015). Это изменение структуры хроматина может влиять на доступность генов для транскрипции и, следовательно, регулировать генную активность.

Исследование наличия CpG-островков в ближних и дальних областях от начала транскрипции тканеспецифичных генов и генов домашнего хозяйства может помочь понять молекулярные основы их регуляции и различия в их экспрессии. Это имеет важное значение для понимания процессов развития, функционирования и патологии клеток и тканей организма.

В данной статье рассмотрены основные различия между тканеспецифичными генами и генами домашнего хозяйства в контексте присутствия, размера и положения CpG-островков в TSS200 и TSS1500 областях генов.

Материал и методы.

Набор тканеспецифичных генов был отобран из базы данных The Human Protein Atlas (https://www.proteinatlas.org/). Для работы были выбраны 740 тканеспецифичных гена с индексом специфичности Tau от 1.00 до 0.93. Гены

домашнего хозяйства в количестве 500 были отобраны из работы (Eisenberg E., Levanon, 2003).

Для поиска CpG-островков была использована программа Cpgplot (https://www.bioinformatics.nl/cgi-bin/emboss/cpgplot). CpG-островком считается регион генома длиной не менее 200 пар нуклеотидов и процентным содержанием GC более 50% и наблюдаемым отношением CpG к ожидаемому более 60%. Ожидаемое CpG определялось как $\frac{\text{количествоC+количествоG}}{\text{длинапоследовательности}}$.

Поиск островков был осуществлен для двух областей выше сайта инициации транскрипции. TSS200-область представляет собой область последовательности от 200 нуклеотидов выше по течению до самого сайта начала транскрипции гена. TSS1500-область - область генома от 200 нуклеотидов до 1500 нуклеотидов выше по течению от TSS200-области. Все последовательности были извлечены из базы данных NCBI (http://www.ncbi.nlm.nih.gov/), сборка генома человека GRCh38.

Результаты исследований и их обсуждение.

В результате анализа тканеспецифичные гены были разбиты на три группы: высокоспецифичные гены (индекс специфичности Tau от 1.00 до 0.99), гены со средне-высокой специфичностью (индекс специфичности Tau от 0.98 до 0.97) и умеренно специфичные гены (индекс специфичности Tau от 0.96 до 0.93). В таблице 1 представлены результаты распределения CpG-островков по кластерам тканевой экспрессии. Гены домашнего хозяйства имели индекс специфичности Tau в диапазоне от 0.08 до 0.40.

Таблица 1. Распределение CpG-островков по кластерам тканевой экспрессии

Тканеспецифичные гены	**Кластер тканевой экспрессии**	**Количество генов**	**Количество CpG-островков в TSS200**	**Количество CpG-островков в TSS1500**
Высокоспецифичные гены	Яичко	299	62	27
	Кожа	61	0	1
	Придаток яичка	25	0	2
	Печень	16	0	1
	Сетчатка	12	1	0
	Головной мозг	10	5	4
	Плацента	7	0	0
	Кишечник	5	0	0
	Почка	5	0	0
	Гипофиз	4	0	0
	Лимфоидная ткань	2	0	0

	Поджелудочная железа	2	0	0
	Слюнная железа	2	0	0
	Костный мозг	1	0	0
	Молочная железа	1	0	0
	Надпочечник	1	0	0
	Паращитовидная железа	1	1	1
	Простата	1	0	0
	Скелетные мышцы	1	0	0
Гены со средне-высокой специфичностью	Яичко	92	25	9
	Головной мозг	21	13	6
	Печень	13	1	1
	Лимфоидная ткань	3	0	0
	Кожа	8	0	0
	Сетчатка	8	0	1
	Кишечник	4	0	0
	Костный мозг	4	1	0
	Придаток яичка	4	1	0
	Гипофиз	3	1	1
	Плацента	3	0	0
	Почка	3	0	0
	Слюнная железа	3	1	1
	Желудок	1	0	0
	Легкие	1	0	0
	Молочная железа	1	0	0
	Надпочечник	1	0	0
	Паращитовидная железа	1	0	0
	Простата	1	0	1
	Семенной пузырек	1	0	0
	Сердечная мышца	1	0	0
	Эндометрий	1	0	0
	Яичник	1	0	0
Умеренно-специфичные гены	Яичко	50	20	9
	Головной мозг	17	8	3
	Лимфоидная ткань	10	0	0
	Сетчатка	6	0	0
	Костный мозг	4	0	0
	Поджелудочная железа	2	0	0
	Скелетные мышцы	2	0	0
	Сосудистое сплетение	2	1	0
	Гипофиз	1	0	0
	Кишечник	3	0	0
	Кожа	1	0	0
	Мочевой пузырь	1	0	0

	Надпочечник	1	1	0
	Паращитовидная железа	1	1	1
	Плацента	1	0	0
	Почка	1	1	0
	Щитовидная железа	1	0	0
Общая сумма		740	144	69

В результате биоинформатического анализа было получено распределение CpG-островков в TSS200 и TSS1500 областях 740 тканеспецифичных генов и 500 генов домашнего хозяйства человека (Таблица 2 и 3). Как можно заметить, процент генов с CpG-островками в исследуемых областях значительно выше у генов домашнего хозяйства по сравнению с тканеспецифичными генами, особенно в области TSS200 (Рисунок 1). Были получены средние длины CpG-островков, а также расстояния от начала островка в TSS200 и конца островка в TSS1500 до сайта начала транскрипции (Таблица 4). Отмечено статистически значимое увеличение средней длины CpG-островков при снижении тканевой специфичности генов.

Таблица 2. Частота встречаемости CpG-островков в TSS200 в зависимости от категории специфичности

Категории генов	**Всего генов, n**	**Гены с островком, n(%)**	**χ^2 (попарно с дом.хоз.)***	**p-value***	**Отношение шансов (OR)[95% ДИ]****
Домашнее хозяйство	500	460 (92.0%)	-		1.00 (референс)
Умеренная специфичность	105	32 (30.5%)	191.8	<0.001	0.038 [0.024, 0.061]
Средне-высокая специфичность	179	43 (24.0%)	293.8	<0.001	0.028 [0.019 0.040]
Высокая специфичность	456	69 (15.1%)	583.7	<0.001	0.015 [0.011 0.022]
p-value общего χ^2-теста			<0.001		

*Примечания: * — Попарные сравнения с группой "Домашнее хозяйство" с поправкой Бонферрони. ** — OR рассчитано относительно группы "Домашнее хозяйство". ДИ — доверительный интервал.*

Для оценки связи между профилем экспрессии генов и наличием CpG-островков в регионах исследуемых регионов был проведен анализ таблиц сопряженности с использованием критерия χ^2 Пирсона. Наблюдалась статистически значимая зависимость между категорией гена и частотой встречаемости островков в области TSS200 ($\chi^2[3, N=1240] = 638.5$, $p< 0.001$) и в области TSS1500 соответственно ($\chi^2[3, N=1240] = 113.9$, $p< 0.001$).

Анализ выявил четкий градиент: гены домашнего хозяйства демонстрировали значительно более высокую частоту наличия CpG-островков в исследуемых регионах по сравнению с генами, обладающими тканеспецифичной экспрессией. Среди последних доля генов с островками снижалась по мере увеличения специфичности причем как в TSS200, так и в TSS1500.

Таблица 3. Частота встречаемости CpG-островков в TSS1500 в зависимости от категории специфичности

Категории генов	**Всего генов, n**	**Гены с островком, n(%)**	**χ^2 (попарно с дом.хоз.)***	**p-value***	**Отношение шансов (OR)[95% ДИ]****
Домашнее хозяйство	500	167 (33.4%)	-		1.00 (референс)
Умеренная специфичность	105	13 (12.4%)	24.5	<0.001	0.282 [0.153, 0.521]
Средне-высокая специфичность	179	20 (11.2%)	34.8	<0.001	0.251 [0.153 0.411]
Высокая специфичность	456	36 (7.9%)	101.9	<0.001	0.171 [0.117 0.250]
p-value общего χ^2-теста			<0.001		

*Примечания: * — Попарные сравнения с группой "Домашнее хозяйство" с поправкой Бонферрони. ** — OR рассчитано относительно группы "Домашнее хозяйство". ДИ — доверительный интервал*

Для уточнения различий между отдельными группами были проведены попарные сравнения с поправкой Бонферрони на множественное тестирование. В обеих проанализированных областях гены домашнего хозяйства достоверно отличались ($p< 0.001$) от всех трех групп тканеспецифичных генов. Кроме того, значимые различия были обнаружены в области TSS200 между высокоспецифичными генами и обеими группами умеренно ($p< 0.001$) и средне-высоко ($p = 0.036$) специфичных генов. При этом разница между группами

умеренно и средне-высоко специфичных генов в TSS200 не достигла статистической значимости (p = 0.19). В области TSS1500 между всеми группами тканеспецифичных генов (умеренные, средне-высокие и высокоспецифичные) не было выявлено статистически значимых различий (p> 0.1 для всех попарных сравнений).

Для количественной оценки силы связи было рассчитано отношение шансов (OR), где за референсную категорию были приняты гены домашнего хозяйства. Анализ показал, что у высокоспецифичных генов шанс иметь CpG-островок в TSS200 был в 65 раз ниже (OR = 0.015, 95% ДИ [0.011, 0.022]), чем у генов домашнего хозяйства. Аналогичным образом, шансы в TSS200 для генов со средне-высокой и умеренной специфичностью были ниже в 36 (OR = 0.028, 95% ДИ [0.019, 0.040]) и 26 раз (OR = 0.038, 95% ДИ [0.024, 0.061]) соответственно. В области TSS1500 у высокоспецифичных генов шанс иметь CpG-островок был в 5.8 раз ниже (OR = 0.171, 95% ДИ [0.117, 0.250]), чем у генов домашнего хозяйства. Шансы для генов со средне-высокой и умеренной специфичностью в TSS1500 были ниже в 4.0 (OR = 0.251, 95% ДИ [0.153, 0.411]) и 3.5 раза (OR = 0.282, 95% ДИ [0.153, 0.521]) соответственно.

Анализ показал что CpG-островки в области TSS200 в среднем длиннее у генов домашнего хозяйства, чем у тканеспецифичных генов [Таблица 4]. Это согласуется с ролью CpG-островков в поддержании открытой хроматиновой структуры и стабильной экспрессии.

Таблица 4. Параметры CpG-островков в областях TSS200 и TSS1500 для групп генов различной специфичности экспрессии.

Группа генов	**Средняя длина островка в TSS200 (п.н.)**	**Средняя длина островка в TSS1500 (п.н.)**	**Среднее расстояние от начала островка (TSS200) до TSS (п.н.)**	**Среднее расстояние от конца островка (TSS1500) до TSS (п.н.)**
Высокая специфичность	491	291	242	675
Средне-высокая специфичность	582	314	274	734
Умеренная специфичность	596	320	263	423
Гены домашнего хозяйства	723	357	322	544

Примечание: значения округлены до целых пар нуклеотидов

Критерий Краскела-Уоллиса для длины островков в TSS200 выявил статистически значимые различия между группами ($H(3) = 43.91$, $p < 0.001$). Размер эффекта был средним ($\eta^2 = 0.068$). Для определения конкретных различий был проведен постхок-анализ с использованием теста Данна с поправкой Бонферрони. Результаты показали, что гены домашнего хозяйства (средний ранг = 328.22) значимо превосходит по длине островков все остальные группы (Умеренно-специфичные гены, средний ранг = 242; Средне-высоко специфичные гены, средний ранг = 238.67; и Высокоспецифичные гены, средний ранг = 198.89; $p < 0.0083$ для всех сравнений). Достоверных различий между умеренно-специфичными генами, средне-высоко специфичными и высокоспецифичными гены выявлено не было.

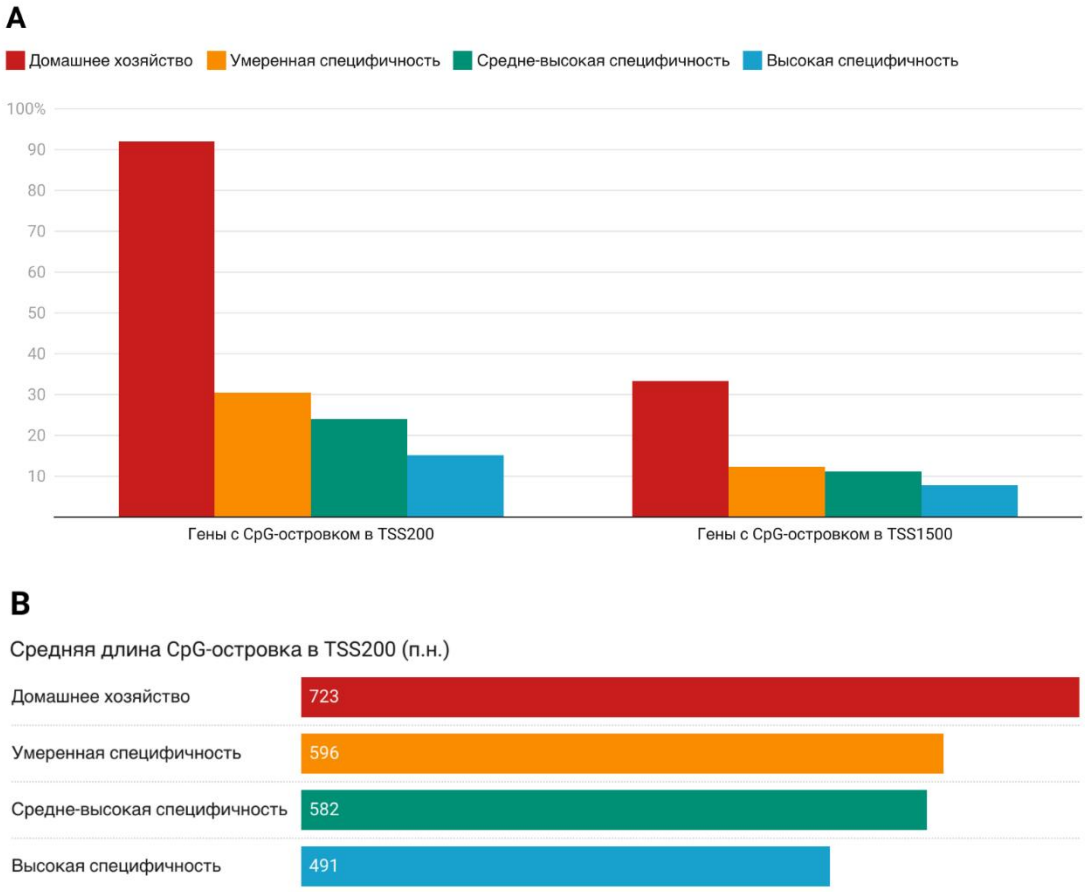

Рисунок 1. Результаты биоинформатического поиска CpG-островков в четырех исследованных кластерах тканеспецифичности. (**A**) Процент генов

содержащих CpG-островок в TSS200 и TSS1500 областях. (**B**) Средняя длина CpG-островка в TSS200 в парах нуклеотидов (значения округлены до целых).

Для длины островков в TSS1500 критерий Краскела-Уоллиса не выявил статистически значимых различий между четырьмя группами ($H(3) = 5.47$, $p = 0.140$). Наблюдаемый размер эффекта был незначительным ($\eta^2 = 0.011$). Полученные данные позволяют сделать вывод о том, что длина островков в TSS1500 статистически не различается между исследуемыми группами.

Для положения начала островка в TSS200 относительно TSS критерий Краскела-Уоллиса показал наличие статистически значимых различий между группами ($H(3) = 19.26$, $p < 0.001$). Однако размер эффекта был небольшим ($\eta^2 = 0.027$). Постхок-анализ с использованием теста Данна с поправкой Бонферрони выявил лишь одну пару групп с достоверным различием: гены домашнего хощяйства (средний ранг = 286.17) и высокоспецифичные гены (средний ранг = 376.91) ($p < 0.0083$). Другие попарные сравнения статистически значимых результатов не показали.

Для положения конца островка в TSS1500 относительно TSS критерий Краскела-Уоллиса показал статистически значимые различия между группами ($H(3) = 8.73$, $p = 0.033$). Тем не менее, последующий постхок-анализ с поправкой Бонферрони не выявил ни одной конкретной пары групп, где различия достигали бы порога статистической значимости. Наблюдаемый размер эффекта был небольшим ($\eta^2 = 0.025$). Это позволяет сделать вывод о наличии общего различия в распределении позиций островков между группами, однако для идентификации конкретных пар, вносящих основной вклад в этот эффект, может потребоваться больший объем данных.

Выводы

Найдена выраженная закономерность: чем более специфично экспрессируется ген, тем реже в его промоторных областях обнаруживаются CpG-островки и тем они короче. Это подтверждает модель, согласно которой стабильная экспрессия генов домашнего хозяйства поддерживается через CpG-островки и связанную с ними открытую хроматиновую структуру, тогда как тканеспецифичные гены опираются на более жёсткие и локализованные механизмы эпигенетической регуляции. Показано, что распределение CpG-островков тесно связано с профилем экспрессии генов и отражает фундаментальные различия в организации регуляторных регионов разных функциональных групп.

Финансовая поддержка

Работа была выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания в сфере научной деятельности № FENW-2023-0018.

Список литературы (References)

1. Antequera F. Structure, function and evolution of CpG island promoters // Cellular and Molecular Life Sciences. 2003. Т. 60. С. 1647–1658. DOI: 10.1007/s00018-003-2374-0
2. Bird A. DNA methylation patterns and epigenetic memory // Genes & Development. 2002. Т. 16, № 1. С. 6–21. DOI: 10.1101/gad.947102
3. Chen Y., Zhang J., Zhang X., Wang Y., Wang Z. Identification of key genes by integrating DNA methylation and next-generation transcriptome sequencing for esophageal squamous cell carcinoma // Aging. 2020. Т. 12, № 2. С. 1332. DOI: 10.18632/aging.102695
4. Deaton A. M., Bird A. CpG islands and the regulation of transcription // Genes & Development. 2011. Т. 25, № 10. С. 1010–1022. DOI: 10.1101/gad.2037511
5. Du Q., Luu P.-L., Stirzaker C., Clark S. J. Methyl-CpG-binding domain proteins: readers of the epigenome // Epigenomics. 2015. Т. 7, № 6. С. 1051–1073. DOI: 10.2217/epi.15.39
6. Eisenberg E., Levanon E. Y. Human housekeeping genes are compact // Trends in Genetics. 2003. Т. 19, № 7. С. 362–365. DOI: 10.1016/S0168-9525(03)00140-9
7. Fedorova L., Mulyar O. A., Lim J., Fedorov A. Nucleotide composition of ultra-conserved elements shows excess of GpC and depletion of GG and CC dinucleotides // Genes. 2022. Т. 13, № 11. Ст. 2053. DOI: 10.3390/genes13112053
8. Joshi C. J., Lee H. J., Lee Y. C., Choi H. I. What are housekeeping genes? // PLoS Computational Biology. 2022. Т. 18, № 7. С. e1010295. DOI: 10.1371/journal.pcbi.1010295
9. Rose N. R., Klose R. J. Understanding the relationship between DNA methylation and histone lysine methylation // Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms. 2014. Т. 1839, № 12. С. 1362–1372. DOI: 10.1016/j.bbagrm.2014.02.007

10. Saxonov S., Berg P., Brutlag D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters // PNAS. 2006. Т. 103, № 5. С. 1412–1417. DOI: 10.1073/pnas.0510310103
11. Springer N. M., Kaeppler S. M. Evolutionary divergence of monocot and dicot methyl-CpG-binding domain proteins // Plant Physiology. 2005. Т. 138, № 1. С. 92–104. DOI: 10.1104/pp.105.060434
12. Vavouri T., Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization // Genome Biology. 2012. Т. 13, № 11. С. R110. DOI: 10.1186/gb-2012-13-11-r110
13. Zhang Y., Rohde C., Tierling S., Stamerjohanns H., Walter J. Genome-wide differential-based analysis of the relationship between DNA methylation and gene expression in cancer // Current Bioinformatics. 2019. Т. 14, № 8. С. 783–792. DOI: 10.2174/1574893614666190220153449
14. Zhu J., He F., Hu S., Yu J. On the nature of human housekeeping genes // Trends in Genetics. 2008. Т. 24, № 10. С. 481–484. DOI: 10.1016/j.tig.2008.08.004

Статья поступила в редакцию 5 сентября 2025 г.
Поступила после доработки 19 сентября 2025 г.
Принята к печати 26 сентября 2025 г.
Received September 5, 2025
Revised September 19, 2025
Accepted September 26, 2025