

УДК 577.214.6+616-006.88

Транскрипция специфичных для семенников генов *ZBTB32* и *ZNF473* подавлена в герминогенных опухолях

С. С. Буланенкова, О. Б. Филюкова, Е. В. Снежков, С. Б. Акопов*, Л. Г. Николаев

Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, 117997 Россия

* E-mail: sergeyakopov@mail.ru

Поступила в редакцию 28.10.2021

Принята к печати 22.08.2022

DOI: 10.32607/actanaturae.11620

РЕФЕРАТ Одним из самых многочисленных в геноме является семейство генов, содержащих домены «цинковый палец» типа C2H2, которое насчитывает более 700 членов. Особый интерес среди них представляют гены с предполагаемой тканеспецифичной транскрипцией, определяющие функциональные свойства отдельных типов клеток и обуславливающих их специфичный ответ на внешние и внутренние воздействия, в том числе при патологических процессах. Целью работы было выявление генов семейства C2H2 с тканеспецифичной транскрипцией и анализ изменений их активности при образовании опухолей. Для поиска таких генов мы воспользовались четырьмя базами данных, содержащими информацию о транскрипции генов в тканях человеческого организма, полученную в результате крупномасштабного секвенирования мРНК. Показано, что основная часть генов семейства C2H2 транскрибируется практически во всех тканях, однако для небольшой группы генов была характерна тканеспецифичная транскрипция, при этом основная часть транскриптов приходилась на семенники. При сопоставлении всех четырех баз данных выявлено 9 таких генов. Специфичную для семенников транскрипцию двух из них – *ZBTB32* и *ZNF473* – мы подтвердили методом количественной ПЦР на матрицах кДНК из разных органов. Показано снижение транскрипции *ZBTB32* и *ZNF473* в герминогенных опухолях. Эти гены рассматриваются как потенциальные маркеры ракового перерождения клеток зародышевого типа.

КЛЮЧЕВЫЕ СЛОВА белки с доменами цинковых пальцев, семенники, герминогенные опухоли, транскрипция.

ВВЕДЕНИЕ

Семейство генов, содержащих домены «цинковый палец» типа C2H2, насчитывает более 700 членов и является одним из самых многочисленных [1, 2]. Одна из основных функций структуры типа «цинковый палец» – связывание с ДНК, поэтому многие белки этого семейства обладают свойствами факторов транскрипции [1]. Помимо цинковых пальцев, разные члены семейства могут содержать дополнительные N-концевые домены, такие, как KRAB, SCAN, VTB/POZ, которые определяют регуляторные функции белков [3]. Несмотря на эволюционную распространенность этого семейства у млекопитающих [4], его значение для клеточной активности пока не вполне ясно. В целом, исследования сфокусированы на детальном анализе структуры и функций отдельных членов семейства, таких, как CTCF [5, 6]. Некоторые гены упоминаются в связи с рядом заболеваний [7]. Учитывая многочисленность семей-

ства, продолжаются попытки определения более общих и глобальных функций его членов. Например, рассматривается их участие в организации структуры хроматина [8], в частности предполагается, что они могут обеспечивать сближенность удаленных участков генома за счет связывания цинковых пальцев с ДНК и белок-белковых контактов [5]. Одна из функций белков семейства, несущих репрессорный домен KRAB, состоит в подавлении активности ретротранспозонов, большое количество которых объясняет широкое распространение семейства C2H2 в геноме человека [9, 10]. Дополнительную сложность в изучение семейства вносит тот факт, что возможности цинковых пальцев не ограничиваются связыванием с ДНК, и ряд доменов этого типа взаимодействует с РНК и белками [7, 11, 12]. Таким образом, изучение данного семейства представляется масштабной и многоступенчатой задачей, направленной на выяснение того, связаны ли разные чле-

ны семейства общим функциональным единством, выполняют узкие специфические функции или же являются многофункциональными белками.

Принимая во внимание разнообразие белков семейства, логично предположить, что они могут участвовать в регуляции различных биологических процессов как общих для всех типов клеток, так и характерных для отдельных типов клеток. Возникает естественный вопрос об активности членов семейства C2H2 при различных патологиях, включая злокачественную трансформацию клеток. Учитывая, что число тканеспецифичных генов на порядок меньше числа генов с повсеместной экспрессией и при этом большая их часть приходится на семенники [13, 14], можно предположить, что эта тенденция сохраняется и у генов семейства C2H2.

В настоящей работе мы провели анализ нескольких баз данных, содержащих информацию о транскрипции генов в разных тканях человеческого организма, и выявили несколько специфичных для семенников генов семейства C2H2. Мы экспериментально подтвердили высокоспецифичную транскрипцию двух из этих генов в семенниках по сравнению с другими тканями. Анализ опухолевых и нормальных тканей семенников показал подавление транскрипции этих генов в опухолях герминогенного происхождения.

Исследования специфичных для семенников генов в перспективе поможет пониманию процессов опухолеобразования и возможному практическому их применению для прогнозирования, диагностики и лечения онкологических заболеваний.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Источники данных

Усредненные данные по уровням транскрипции генов в разных тканях, представленные в TPM (Transcripts Per Million, см. [15]), получены из <https://proteinatlas.org> (Human Protein Atlas, HPA), <https://gtexportal.org/home/> (The Genotype-Tissue Expression, GTEx, версия файла GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz, без трансформированных лимфоцитов), <https://www.ebi.ac.uk/> (использовали данные E-MTAB-513 (Illumina Body Map), E-MTAB-4344 (ENCODE project), E-MTAB-2836 (HPA), E-MTAB-5214 (GTEx)). Полные данные по транскрипции генов во всех образцах разных тканей, представленные в TPM, получены с сайта GTEx (<https://gtexportal.org/home/>), версия файла GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.

Тканеспецифичные гены отбирали для каждой базы данных следующим образом. Сначала были

отобраны все гены каждой ткани с уровнем экспрессии не менее 5 TPM. Дальнейший анализ проведен для каждого выбранного гена. Уровень экспрессии гена в определенной ткани сравнивали с уровнем его экспрессии в каждой из оставшихся тканей. Для каждой ткани подсчитывали соотношение уровня экспрессии гена в TPM к уровням его экспрессии в других тканях. Затем из набора значений, полученных для каждой ткани, выбирали минимальное, и если это значение было не менее 3, то считали, что ген специфически экспрессируется в этой ткани.

Коллекция тканей

Образцы легкого, почки, толстого и тонкого кишечника, скелетной мышцы, лимфоузла, селезенки и передней коры головного мозга получены от здоровых взрослых пациентов, погибших вследствие травм, несовместимых с жизнью. Образцы опухолей семенников (31) получены в результате орхиэктомии, среди них 27 образцов герминогенного происхождения и 11 образцов прилегающих нормальных тканей, среди них 7 герминогенного происхождения (паренхима яичка). 18 образцов были парными (6 пар герминогенного происхождения и три пары негерминогенного происхождения). Два образца нормальной ткани семенников, полученные в ходе хирургической кастрации пациентов с раком простаты, далее использовали в качестве контроля (подробности см. в дополнительной таблице Table_S1, доступна по запросу). Все репрезентативные образцы были немедленно заморожены в жидком азоте. Забор образцов производился в соответствии с федеральным законом № 180 «О биомедицинских клеточных продуктах» (приказ Министерства здравоохранения Российской Федерации от 11.08.2017 № 517н, приложение 2, см. <http://publication.pravo.gov.ru/Document/View/0001201709290030>) и утверждался этически комитетами Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук и Национального медицинского исследовательского центра онкологии имени Н.Н. Блохина Министерства здравоохранения Российской Федерации.

Выделение РНК

Тотальную РНК выделяли с использованием гуанидинизотиоцианата согласно [16]. Все препараты РНК обрабатывали ДНКазой I (Promega, США) в соответствии с рекомендациями фирмы-производителя. Итоговые образцы очищали с использованием RNeasy MINI RNA kit (Qiagen, США). Качество и чистоту образцов РНК определяли при помощи электрофореза в 1% агарозном геле. Количество РНК определяли спектрофотометрически.

Таблица 1. Последовательности праймеров для количественной ПЦР

Ген	Прямой	Обратный	Длина продукта, п.н.
18S	TGAGAAACGGCTACCACATC	GCTATTGGAGCTGGAATTACC	203
<i>GAPDH</i>	ACTCCTCCACCTTTGACGCT	TCTTCCTCTTGTGCTCTTGCT	179
<i>ZBTB32</i>	GCCCTATGCGTGCTCTGTCT	GGTCATGGCCGAGAAGTCC	139
<i>ZNF473</i>	GGAAGCCCAGAAGCAACAAG	TTCTGGATCGCCTAGCAAAC	189
<i>ZNF446</i>	AATAGAGGGGTCTGTCCAGC	CCGTACTTCTCCAGCATCGC	231

Количественная ПЦР на матрице кДНК (ОТ-ПЦР)

Синтез первых цепей осуществляли с использованием рассеянной гексануклеотидной затравки (Promega) и обратной транскриптазы PowerScript (Clontech, США) согласно рекомендациям фирмы-производителя. Количество кДНК-матрицы в каждой реакции ПЦР было эквивалентно 10 нг тотальной РНК. Последовательности использованных в работе праймеров приведены в табл. 1. Подбор праймеров проводили с использованием программного обеспечения Primer-Blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>), при этом одним из критериев было расположение праймеров в разных экзонах. Реакцию проводили в буферной системе qPCR-HS SYBR («Евроген», Россия) на ПЦР-амплификаторе LightCycler 480 (Roche, США) с использованием следующего температурного режима: 95°C, 3 мин; затем 40 циклов: 95°C, 20 с; 65°C, 20 с; 72°C, 40 с. Уровень транскрипции оценивали относительно среднего геометрического представления транскриптов генов 18S рРНК и *GAPDH*. Все эксперименты ставили в трех технических повторях.

Статистическая обработка экспериментальных результатов

Сравнение двух групп данных проводили с использованием критерия Манна–Уитни для независимых выборок и с применением критерия Вилкоксона для связанных выборок. Для анализа согласованности транскрипции рассчитывали значение коэффициента корреляции Спирмена. Уровень значимости принимали равным 0.05. Все вычисления на основе баз данных производили с использованием Excel2010, программной среды R [17] и дополнительных пакетов stats, openxlsx (<https://github.com/ycphs/openxlsx>). При расчете тканеспецифичности транскрипции генов рассчитывали отношение уровня транскрипции каждого гена в ТРМ во всех тканях, кроме исследуемой, к его транскрипции в исследуемой ткани, что позволяло избежать случаев деления на ноль при отсутствии транскрипции в других органах. Для каждого гена определяли максимальное из по-

лученных отношений и, если оно не превышало 0.3, то этот ген отбирали для последующего анализа.

Для построения корреляционных матриц использовали пакеты stats, cluster (<https://CRAN.R-project.org/package=cluster>), corrplot (<https://github.com/taiyun/corrplot>). При проведении кластерного анализа, построении дендрограмм и расчете оптимального количества кластеров использовали пакеты stats, cluster, dendextend [18], NbClust [19], clValid [20]. В качестве меры расстояния между генами использовали величину, равную (1 – модуль коэффициента корреляции Спирмена). Кластеры выделяли согласно алгоритму иерархической классификации по методу полной связи. Скрипты доступны по запросу.

Дополнительные онлайн-ресурсы

Построение диаграмм Венна – <http://bioinformatics.psb.ugent.be/webtools/Venn/>,

GePIA (Gene Expression Profiling Interactive Analysis – <http://gepia.cancer-pku.cn/>) [21] – сопоставление уровней транскрипции генов в нормальных и опухолевых образцах, основанное на данных GTEx (<https://gtexportal.org/home/>) и TCGA (The Cancer Genome Atlas Program – <https://portal.gdc.cancer.gov/>),

HGNC (HUGO Gene Nomenclature Committee) – <https://www.genenames.org/>,

Ensembl – <https://www.ensembl.org/index.html>,

Онлайн-конвертер идентификационных номеров генов – <https://www.biotools.fr/>,

Human genome browser [22] – <https://genome.ucsc.edu/>

РЕЗУЛЬТАТЫ

Девять генов семейства C2H2 тканеспецифично транскрибируются в семенниках

В анализе использовали данные крупномасштабного секвенирования РНК, предоставленные проектами Illumina Body Map, Encyclopedia of DNA Elements (ENCODE) [23], The Genotype-Tissue Expression (GTEx) [24] и Human Protein Atlas (HPA) [25]. Приведенные источники данных различаются

Таблица 2. Представленность тканей и генов в базах данных

База данных	Ткани, число	Гены семейства C2H2, число
Illumina Body Map	16	717
ENCODE	13	718
GTE _x	53	718
HPA	43	709

по количеству тканей, количеству образцов, приходящихся на каждую ткань (от одного до нескольких сотен) и по способу получения биоматериала (в результате хирургической операции или при посмертном вскрытии). Более подробное описание баз данных приведено в работе [13].

Мы анализировали усредненные по нескольким образцам уровни транскрипции, представленные в TPM (Transcripts Per Million). Из каждого набора данных отбирали гены семейства C2H2; отбор проводили на основании принадлежности к группе 28 в базе генов человека HGNC [26], используя идентификационные номера Ensembl. Количество тканей и генов семейства C2H2, представленных в разных базах, приведено в табл. 2.

Далее мы оценили общее количество генов данного семейства, транскрибирующихся в разных тканях на уровне не менее 3 TPM. За исключением нескольких тканей, где количество генов, транскрибируемых на таком уровне, находится в диапазоне 200–300 (например, печень и скелетная мышца), в других тканях количество таких генов превышает 450, при этом наибольшее число генов транскрибируется на таком уровне в семенниках (более 600). Таким образом, можно заключить, что большая часть генов семейства активна практически во всех представленных тканях (см. дополнительную таблицу Table_S2, доступна по запросу).

Далее проводили поиск генов семейства C2H2, специфичных для каждой ткани. Сначала отбирали гены с уровнем транскрипции в данной ткани не менее 5 TPM. Ген считали активным преимущественно в данной ткани, если уровень его транскрипции в ней превышал минимум в 3 раза его транскрипцию в любой из других тканей или органов.

Результаты по тканеспецифичной транскрипции генов, представленные в каждой из баз данных, приведены в дополнительной таблице Table_S3 (доступна по запросу). Отметим, что во всех четырех базах приведены данные только для тканей надпочечников, семенников, яичников, печени и легкого. Во всех случаях наибольшее число генов с тканеспецифичной транскрипцией (16 и более) обнаруже-

но в семенниках. Также тканеспецифично транскрибирующиеся гены C2H2 (не более 10) обнаружены в яичниках, мозге, селезенке, коре головного мозга, костном мозге, предстательной железе. Однако в отличие от семенников сопоставление результатов, представленных в разных базах данных, практически не выявило общих генов. Это наблюдение согласуется с результатами работы [14], в которой найдено 35 генов, специфично транскрибирующихся в семенниках, тогда как в других органах и тканях число специфичных для них генов семейства C2H2 не превышало шести. Таким образом, в то время как в большинстве тканей основная часть генов семейства C2H2 транскрибируется неспецифично, в семенниках транскрибируется численно наибольшая тканеспецифичная фракция генов.

В дальнейшем мы сосредоточились на этой фракции генов. Суммарно во всех базах данных обнаружено 52 гена этого типа. Стоит отметить, что при использовании разных версий сборок баз данных количество и набор генов незначительно меняются. Это связано с тем, что при пополнении баз новыми данными усредненные значения уровней транскрипции генов изменяются, и гены, находящиеся на границе заданных нами условий, могут оказаться на любой из сторон. Стоит также отметить, что в таких базах данных, как Illumina и ENCODE, представлено незначительное количество тканей (и наименьшее количество образцов, приходящееся на ткань – от одного до трех) по сравнению с HPA и GTE_x, что повышает вероятность ложного выбора гена как семенник-специфичного в связи с отсутствием информации о его транскрипции в тканях, которые не представлены в базе. Поэтому далее мы анализировали 25 генов, уровень транскрипции которых в семенниках превышал максимальный уровень в других органах более чем в 1.5 раза во всех рассмотренных базах данных. Из них 13 генов совпали с данными [14] по 35 семенник-специфичным генам. Расхождения могут объясняться использованием разных баз данных (GTE_x, HPA в этой работе и TiGER в [14]) и различных алгоритмов поиска тканеспецифичных генов. Таким образом, видно, что отбор генов зависит от алгоритма поиска и состояния базы данных. Все это указывает на огромную значимость непосредственного экспериментального подтверждения анализируемых данных.

Из наборов генов, выделенных из каждой базы данных, мы отобрали общие для всех баз (рис. 1). Всего таким образом отобрано 9 генов: ZBTB32, CTCFL, ZNF560, ZNF541, ZNF473, ZNF165, PRDM9, ZSCAN5A, ZNF487. Все остальные гены, за исключением PRDM9, ZSCAN5A и ZNF487, выявляются как семенник-специфичные при использовании раз-

Таблица 3. Транскрипция отобранных генов в семенниках согласно четырем базам данных (в ТРМ)*

Ген	HPA	ENCODE	GTEX	Illumina
ZNF473	76.4 (11.1/4.25)	79 (8/2.5)	49.1 (8.3/3.25)	46 (8/4)
ZBTB32	43.3 (7.5/0.15)	84 (4/0)	109.3 (6.5/0.3)	32 (5/0.4)
ZNF541	19.2 (1.7/0.1)	40 (1/0.3)	45.7 (2.7/0.3)	18 (2/0.3)
ZSCAN5A	34.6 (11.3/3.4)	12 (2/0.95)	14.4 (2.3/1.2)	12 (3/1)
ZNF487	51.4 (9.9/2.95)	28 (7/2)	23.2 (4.4/2.1)	21 (5/2)
PRDM9	7 (1.7/0)	9 (0/0)	6.9 (0/0)	6 (0/0)
ZNF560	11.9 (1.2/0)	15 (0.2/0)	15.5 (0.6/0)	12 (0.3/0)
CTCF	20.4 (0.8/0.3)	14 (0.2/0)	7.5 (0.1/0)	17 (0.5/0.2)
ZNF165	35.8 (7.4/1.55)	46 (15/2)	45.6 (7.6/1.35)	49 (9/2)

*В скобках через косую черту также указаны максимальное и медианное значение транскрипции отобранных генов в других тканях (в ТРМ) согласно той же базе данных.

Таблица 4. Данные по белковым доменам и положению генов

Ген	Число С2H2	Другие домены	Расположение
CTCF	11	Нет	20q13.31
PRDM9	14	SET, KRAB	5p14.2
ZBTB32	3	BTB/POZ	19q13.12
ZNF165	6	SCAN	6p22.1
ZNF473	20	KRAB	19q13.33
ZNF541	5	ELM2, SANT	19q13.33
ZNF560	15	KRAB+KRAB	19p13.2
ZSCAN5A	5	SCAN	19q13.43
ZNF487	3	KRAB	10q11.21

личных сборок баз данных, а также представлены в группе генов, специфичных для семенников [14]. Данные по транскрипции 9 генов в семенниках согласно четырем рассмотренным базам данных, медианные и максимальные значения транскрипции генов в тканях, отличных от семенников, представлены в табл. 3, а структурные характеристики генов – в табл. 4 (аналогичные данные по всем 52 генам представлены в дополнительной таблице Table_S4, доступна по запросу).

В то же время существуют различия в профилях транскрипции отобранных генов. Гены *CTCF*,

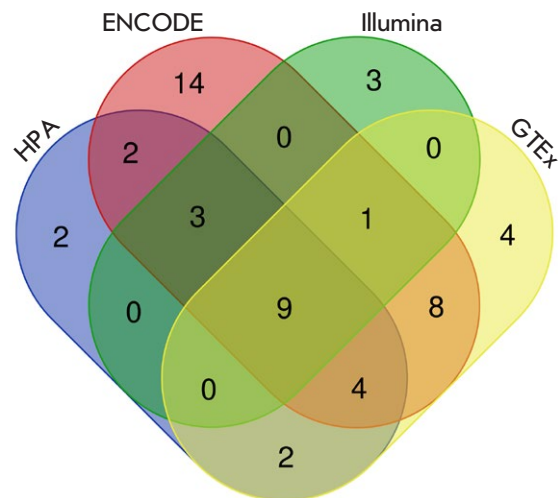


Рис. 1. Диаграмма Венна, показывающая отобранные по заданным критериям гены, общие для четырех баз данных – Human Protein Atlas, ENCODE, Illumina Body Map и GTEX. Диаграмма построена с помощью онлайн-программы (<http://bioinformatics.psb.ugent.be/webtools/Venn/>)

PRDM9, *ZNF560* – это сугубо специфичные гены семенников, уровень их транскрипции в других тканях не превышает 0.6 ТРМ. Кроме того, уровень транскрипции этих генов в семенниках приблизительно в 2–3 раза ниже, чем у генов *ZBTB32*, *ZNF165*, *ZNF473* и *ZNF541*. Гены *ZBTB32* и *ZNF541* имеют практически недетектируемый уровень транскрипции в большинстве тканей (медиана варьирует в пределах 0–0.6 ТРМ), однако в ряде органов и клеток, таких, как В-клетки, лимфоузлы, аппендикс, селезенка, миндалина, Пейеровы бляшки, транскрипция *ZBTB32* может достигать 7.5 ТРМ, а уровень транскрипции *ZNF541* в надпочечниках составляет 2–3 ТРМ. Гены *ZNF165*, *ZSCAN5A*, *ZNF487* и *ZNF473* на низком уровне транскрибируются практически во всех тканях, но уровень их транскрипции в семенниках минимум в 3 раза выше, чем в любой из остальных тканей.

Для дальнейшего анализа мы отобрали два гена – *ZNF473* и *ZBTB32*, важные для определения тканевой идентичности семенников [27].

Транскрипция генов *ZBTB32* и *ZNF473* понижена в герминогенных опухолях семенников

Мы экспериментально определили содержание транскриптов *ZBTB32* и *ZNF473* в образцах семенников, легкого, почки, толстого и тонкого кишечника, скелетной мышцы, лимфоузла, селезенки и передней коры головного мозга человека. Уровень транскрипции оценивали с помощью ПЦР в реальном времени

с использованием матрицы кДНК и среднего геометрического содержания транскриптов *GAPDH* и 18S рРНК для нормализации. Результаты представлены на *рис. 2А,Б*. Как видно, транскрипция *ZNF473* в семенниках превышает уровень транскрипции в других органах минимум в 5 раз, а *ZBTB32* – в 4 раза. Не высокий, но заметный уровень транскрипции этих генов наблюдался в лимфоидных тканях (селезенка и лимфоузлы), в остальных проверенных тканях уровень транскрипции этих генов незначителен, что согласуется с рассмотренными выше результатами анализа баз данных.

Для сравнения профилей транскрипции двух исследуемых генов и подтверждения их независимости от особенностей подготовки образцов, постановки реакций и т.п. мы параллельно провели анализ транскрипции случайно выбранного гена *ZNF446*, также принадлежащего семейству С2Н2. Этот ген не был отобран из баз данных согласно указанным выше критериям и транскрибировался в семенниках и других тканях на низком уровне (1–10 TPM) без выраженной тканевой специфичности. Полученные нами результаты подтверждают отсутствие тканевой специфичной транскрипции гена *ZNF446* (*рис. 2В*).

Мы определили содержание транскриптов тех же генов на панели кДНК, полученной из образцов опухолей и нормальных тканей семенников. Уровни транскрипции оценивали как описано выше. Результаты показаны на *рис. 3*. Панель представлена образцами паренхимы здоровых семенников (контроль – образцы 17N, 19N), образцами опухолей и прилегающих к ним нормальных тканей. Ряд образцов представлен парами опухоль/прилегающая условно нормальная ткань (норма), полученными от одного пациента. Опухоли (и прилегающие к ним нормы) представлены как образцами, имеющими герминогенное происхождение (семинома, тератома, опухоль желточного мешка, эмбриональный рак и смешанные типы), так и негерминогенными образцами, представленными стромальными и паратестикулярными опухолями (лейдигомы, рабдомиосаркомы, лейомиосаркомы).

В герминогенных опухолях транскрипция гена *ZBTB32* (*рис. 3А*) подавляется до практически недетектируемого уровня, а транскрипция гена *ZNF473* – до значений, характерных для транскрипции в других тканях (*рис. 3Б*). Четких закономерностей изменения транскрипции гена *ZNF446* при опухолеобразовании не выявлено (*рис. 3В*).

Образцы нормальных тканей, прилегающих к опухолям герминогенного происхождения, характеризуются большим разбросом уровней транскрипции генов *ZBTB32* и *ZNF473* от значений, сравни-

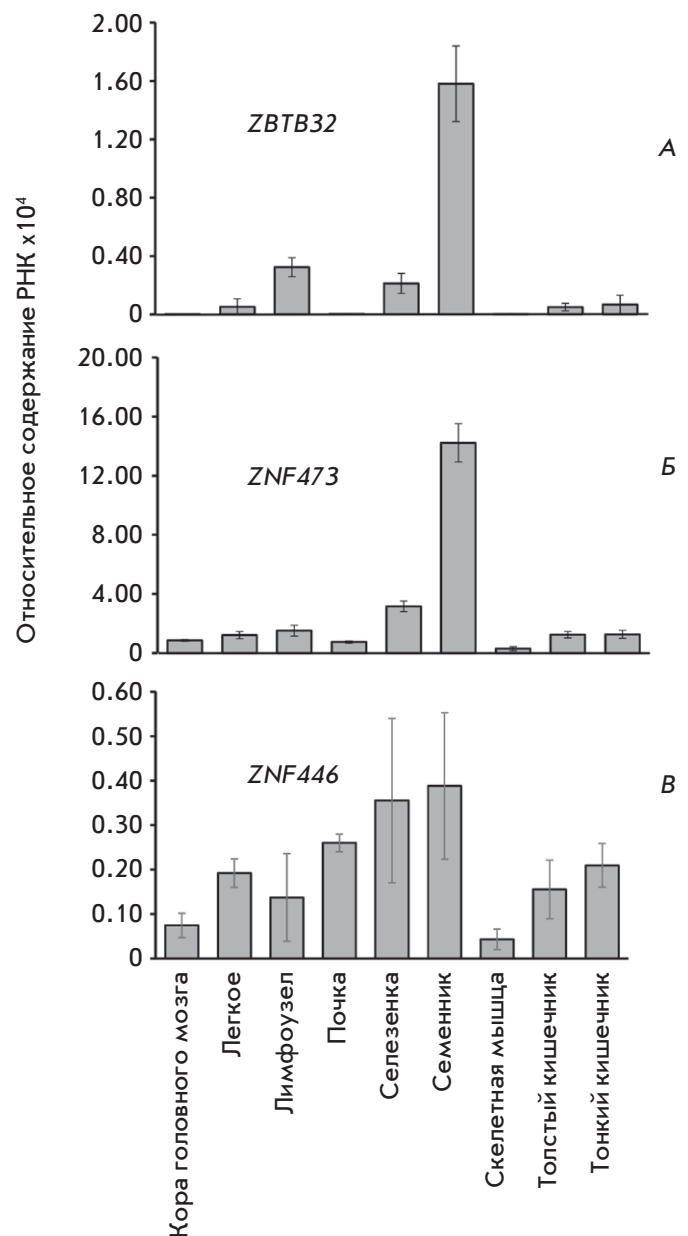


Рис. 2. Относительное содержание транскриптов генов *ZBTB32* (А), *ZNF473* (Б) и *ZNF446* (В) в различных тканях. Содержание РНК нормализовали относительно среднего геометрического содержания транскриптов генов домашнего хозяйства 18S рРНК и *GAPDH*

мых с контрольными образцами (образцы 5N, 8N и 14N), до значений, характерных для опухолей (образцы 7N, 21N). У парных образцов с высоким уровнем транскрипции в норме (пары № 8, 14 и 32) происходит снижение уровня транскрипции гена *ZBTB32* в опухоли минимум в 8 раз, а *ZNF473* – в 3 раза. Стоит отметить, что в этих же образцах транскрипция *ZNF446* может как повышаться, так и понижаться. Одной из причин разброса уров-

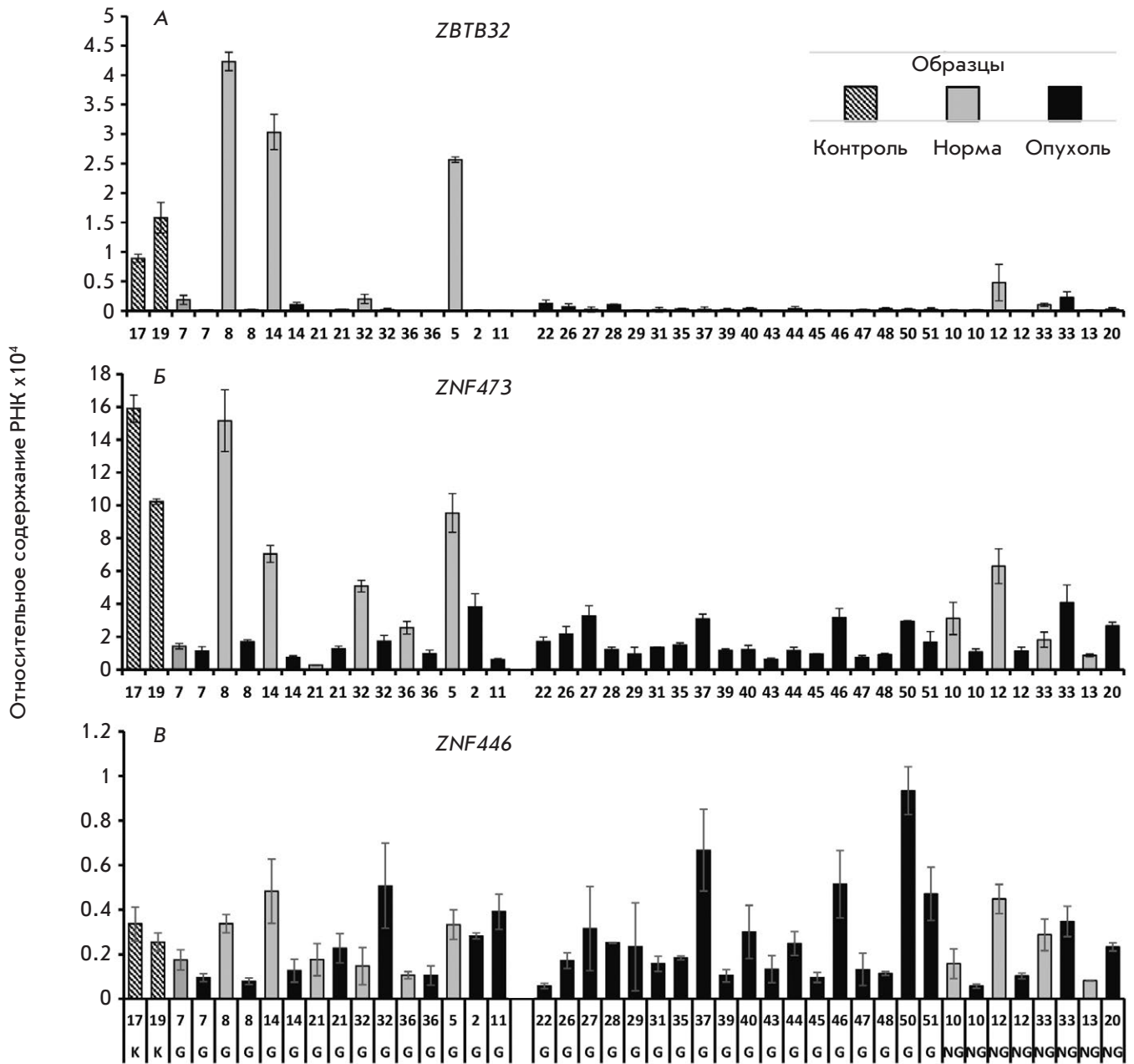


Рис. 3. Относительное содержание транскриптов генов *ZBTB32* (А), *ZNF473* (Б) и *ZNF446* (В) в образцах опухолей семенников герминогенного и негерминогенного происхождения. Контрольные образцы (здоровая паренхима яичка) отмечены штриховкой, образцы из здоровой ткани, прилегающей к опухоли, отмечены серым цветом, а образцы опухолевой ткани отмечены черным. Образцы здоровой и опухолевой ткани с одинаковым номером принадлежат одному и тому же пациенту. Буквы внизу обозначают: К для контрольных образцов, G для образцов опухолей зародышевых клеток и NG для образцов опухолей, не содержащих зародышевых клеток. Содержание РНК нормировали относительно среднего геометрического уровней экспрессии 18S рРНК и *GAPDH*

ней транскрипции *ZBTB32* и *ZNF473* в образцах, прилегающих к герминогенной опухоли, может быть начало злокачественного перерождения клеток в тканях, морфологически определяемых как норма. Нельзя также исключить возможность влияния некоторых типов опухолевых клеток на соседние тка-

ни. Например, показана зависимость транскрипции специфичных для семенников генов семейства *PIWI* в прилегающей к опухоли ткани от типа опухоли [28]. Однако в исследуемых нами образцах закономерностей между типом опухоли и транскрипцией генов в прилегающей нормальной ткани не выяв-

лено. Изучение влияния опухолей различного типа на свойства прилегающих тканей может быть перспективной задачей будущих исследований.

Наблюдается разброс в уровнях экспрессии генов *ZNF473*, *ZBTB32* и *ZNF446* в негерминогенных тканях. Сопоставление содержания транскриптов этих генов в негерминогенных опухолях (образцы 10Т, 12Т, 20Т, 33Т) и прилегающих нормальных тканях (образцы 10N, 12N, 13N, 33N) выявляет разнонаправленность изменений экспрессии всех трех генов при образовании негерминогенных опухолей и не обнаруживает четких закономерностей изменения уровней транскрипции всех трех генов.

В целом, наблюдается достоверное снижение уровня транскрипции генов *ZNF473* и *ZBTB32* в опухоли по сравнению с нормой (*p*-значение по Манну–Уитни менее 0.02 в обоих случаях и более 0.4 у *ZNF446*). Эти гены могут выступать в качестве маркеров опухолей, происходящих из клеток зародышевого пути.

Для подтверждения этих данных мы при помощи онлайн-ресурса GePIA [21] сопоставили уровни транскрипции 25 выявленных нами ранее генов в нормальных и опухолевых образцах семенников (табл. 5). Уровень транскрипции большинства генов (за исключением *ZNF728*, *ZNF560* и *ZFP42*) в опухоли снижен более чем в 3 раза до значений, сопоставимых с уровнями их транскрипции в других тканях. Уровень транскрипции гена *ZNF728* в опухоли снижается менее чем в 2 раза, уровень транскрипции гена *ZNF560* остается практически неизменным, а гена *ZFP42*, напротив, возрастает в 17 раз.

Результаты, полученные нами для генов *ZBTB32* и *ZNF473*, согласуются с данными GePIA. Это позволяет рассматривать другие гены со сходным, согласно GePIA, поведением в качестве потенциальных маркеров опухолеобразования в семенниках. Можно предположить, что большая часть рассматриваемых генов входит в состав сетей межгенных взаимодействий, которые отвечают за основные клеточные процессы, определяющие функции семенников, подавление этих процессов происходит при опухолеобразовании.

В связи с этим, важным является сходство профилей транскрипции генов *ZBTB32* и *ZNF473* в герминогенных образцах в норме и в опухоли, что предполагает согласованность их транскрипции. Согласованность транскрипции 25 отобранных семенник-специфичных генов мы проанализировали с использованием базы данных GTEx, содержащей данные крупномасштабного секвенирования РНК в 361 образце семенников в норме и построили для них и контрольного гена *ZNF446* корреляционную матрицу (рис. 4А).

Таблица 5. Уровни транскрипции генов семейства C2H2, специфически транскрибирующихся в семенниках, в нормальных и опухолевых образцах семенников согласно данным GePIA*

Ген	Норма, TPM	Опухоль, TPM	Соотношение норма/опухоль, раз	Приведенное <i>p</i> -значение
ZBTB32	93.0	0.93	100.0	1.78e-77
PRDM9	4.52	0.05	90.4	7.58e-78
ZNF541	27.4	0.36	76.2	1.19e-99
KLF17	15.8	0.31	51.0	5.38e-24
CTCFL	9.11	0.20	45.6	1.60e-77
ZNF479	3.34	0.13	25.7	6.61e-39
ZFHX2	17.7	1.09	16.2	4.19e-56
ZNF487	33.0	2.48	13.3	4.01e-63
ZNF433	32.9	2.61	12.6	1.77e-75
ZSCAN5B	2.25	0.18	12.5	7.78e-45
ZNF165	21.3	2.20	9.68	2.25e-44
ZNF563	15.3	1.98	7.71	2.93e-70
ZNF473	39.1	6.37	6.14	3.95e-47
ZSCAN5A	34.6	5.94	5.83	1.09e-59
ZNF628	31.7	5.45	5.82	7.35e-45
ZNF233	9.55	1.68	5.68	1.84e-32
ZNF829	12.6	3.28	3.85	5.86e-20
ZNF646	32.7	8.83	3.70	3.21e-38
ZNF689	22.4	6.40	3.50	1.54e-33
ZNF318	46.8	14.3	3.27	7.26e-34
BNC1	19.6	6.45	3.04	3.99e-9
ZNF574	37.0	15.7	2.36	2.47e-22
ZNF728	5.86	3.44	1.70	1.94e-7
ZNF560	8.89	7.53	1.18	4.09e-2
ZFP42	2.54	43.0	0.06	2.86e-34

*Приведены медианы транскрипции в TPM. Гены расположены в порядке уменьшения соотношения уровней их транскрипции в норме и в опухоли. Девять генов, отобранных одновременно из четырех баз данных, выделены жирным.

Используя коэффициенты корреляции в качестве меры расстояния между генами по формуле (1 – модуль коэффициента корреляции), мы построили иерархическое дерево, применив алгоритм иерархической классификации (рис. 4Б). Следует отметить, что при таком подходе наиболее «близко расположенными» генами оказываются гены с повышенным значением коэффициента корреляции независимо от знака. Оптимальное количество кластеров раз-

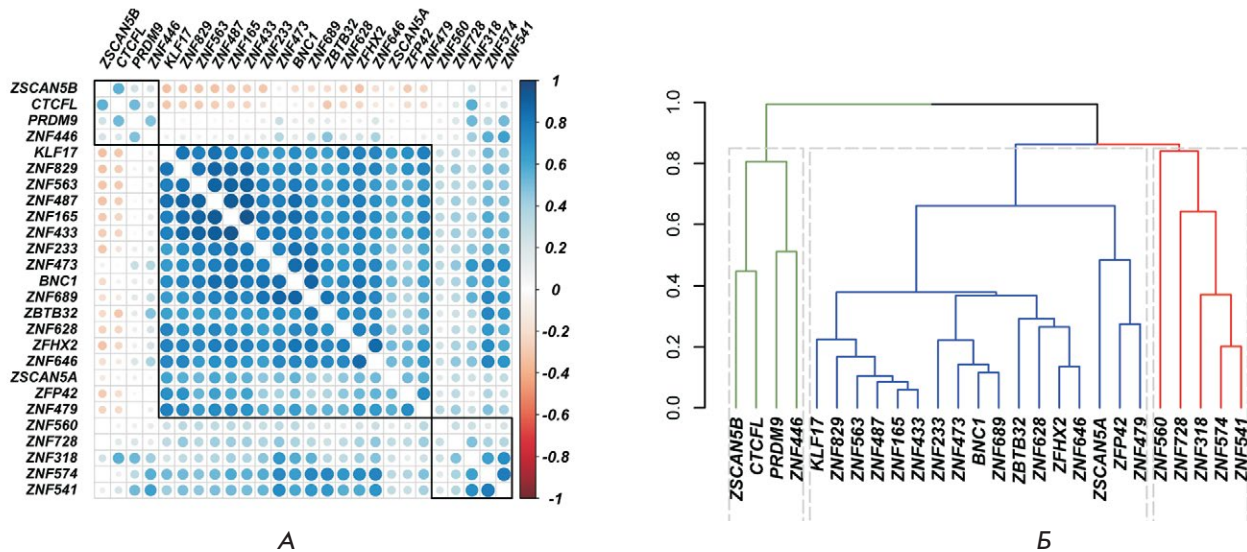


Рис. 4. Корреляционная матрица уровней транскрипции 25 генов, отобранных на основании их специфичной экспрессии в семенниках, и контрольного гена *ZNF446* в 361 образце семенников, содержащихся в коллекции GTEh (А). Гены расположены согласно иерархической кластеризации методом полной связи. Границы кластеров обведены черным. Рассчитаны коэффициенты корреляции Спирмена, значение которых обозначено цветом. Дендрограмма, отображающая согласованность экспрессии генов (Б). В качестве меры различия использована величина, равная $(1 - \text{модуль коэффициента корреляции Спирмена})$. Выделены три кластера, которые обозначены разными цветами, пунктирная линия серого цвета показывает границы кластеров

ными алгоритмами определяется числом от 2 до 3. Между разными способами иерархической классификации наблюдается хорошее соответствие (коэффициенты корреляции при кофенетическом анализе превышают 0.51). При кластеризации разными методами большая часть генов, включая *ZNF473* и *ZBTB32*, попадает в один кластер, и только четыре гена – *CTCFL*, *ZSCAN5B*, *PRDM9* и контрольный ген *ZNF446* – в другой. Таким образом, большая часть отобранных генов согласованно транскрибируется в семенниках. Уровни транскрипции генов в пределах одного кластера положительно коррелируют между собой. Гены, входящие в разные кластеры, могут входить в состав разных ветвей специфичной для семенников генной сети и, как следствие, быть участниками разных биологических процессов, происходящих в семенниках. Установление положения рассмотренных генов в иерархии межгенных взаимодействий и их взаимосвязи с внутриклеточными процессами в семенниках является масштабной и перспективной задачей будущих исследований.

ОБСУЖДЕНИЕ

Изучение экспрессии генов представляет особый интерес в случае патологических процессов, включая злокачественное перерождение клеток. Важным этапом является поиск генов, которые в дальнейшем предполагается использовать как диагностические маркеры или как объекты направленной

генной терапии. Особый интерес представляют гены с выраженной тканеспецифичной экспрессией, так как именно они обеспечивают специфический ответ клеток на внешние и внутренние воздействия. В данной работе в качестве объекта исследования мы выбрали семейство генов, содержащих домены цинковый палец типа C2H2, представляющее особый интерес в связи с тем, что, во-первых, большинство его членов из-за наличия ДНК-связывающего домена относятся к факторам транскрипции, т.е. к регуляторным генам, во-вторых, многочисленность членов его семейства может позволить выявить ряд любопытных закономерностей их экспрессии.

В настоящий момент в связи с развитием технологии массивированного секвенирования создано несколько крупных баз данных по экспрессии генов в разных тканях и органах. Применяя заданные пользователем алгоритмы поиска генов с тканеспецифичной экспрессией, можно производить отбор генов-кандидатов. Базы данных различаются по количеству и способу получения образцов, поэтому при поиске и дальнейшем анализе генов-кандидатов важно правильное сопоставление получаемых результатов. В данной работе мы провели одновременный анализ четырех баз данных, беря за основу средний уровень экспрессии гена в ткани/органе в каждой базе данных. В результате было отобрано 9 генов семейства C2H2 с потенциально тканеспецифичной транскрипцией в семенниках. Для даль-

нейшего анализа было выбрано два гена – *ZBTB32* и *ZNF473*, тканеспецифичность транскрипции которых в клетках паренхимы семенников мы подтвердили экспериментально.

К важным параметрам экспрессии генов относится ее изменение при злокачественном перерождении клеток. Такого рода данным посвящен онлайн-ресурс GePIA (Gene Expression Profiling Interactive Analysis). В его основе лежит алгоритм, позволяющий сопоставлять данные массивированного секвенирования, полученные из двух источников – коллекции нормальных тканей GTEx и коллекции опухолевых тканей TCGA. Согласно этому ресурсу, транскрипция генов *ZBTB32*, *ZNF473*, *PRDM9*, *CTCF*, *ZNF165*, *ZNF541* и ряда других понижена в герминогенных опухолях семенников, поэтому они могут рассматриваться как потенциальные маркеры онкотрансформации клеток зародышевого типа. Мы экспериментально подтвердили снижение уровня транскрипции двух выбранных генов (*ZBTB32* и *ZNF473*) в герминогенных опухолях. Четких закономерностей по экспрессии этих генов в негерминогенных опухолях и в прилегающих нормальных тканях выявить не удалось. Уровень экспрессии этих генов в негерминогенных нормальных тканях

изначально низок, таким он остается в негерминогенных опухолях. Уменьшение уровня транскрипции этих генов в клетках зародышевого пути может служить маркером риска развития герминогенных опухолей. Отсутствие транскрипции *ZBTB32* может также служить доказательством отсутствия примесей прилегающих нормальных тканей при получении опухолевого образца в экспериментах, где важна чистота опухолевого препарата. В последнем случае, однако, необходимы дополнительные маркеры, позволяющие отличить герминогенные опухоли от негерминогенных тканей.

Таким образом, накопленный в базах данных материал по экспрессии генов служит существенным подспорьем при поиске генов-кандидатов, отвечающих за развитие патологических процессов. Дальнейший анализ в виде экспериментального подтверждения выявленных *in silico* закономерностей, установление функций генов и их положения в иерархии генных сетей представляют собой интересную и масштабную задачу будущих исследований. ●

Авторы благодарны М.В. Зиновьевой за предоставленные препараты кДНК.

СПИСОК ЛИТЕРАТУРЫ

- Alba M.M. // *Genome Biol.* 2017. V. 18. P. 168.
- Tupler R., Perini G., Green M.R. // *Nature.* 2001. V. 409. P. 832–833.
- Collins T., Stone J.R., Williams A.J. // *Mol. Cell. Biol.* 2001. V. 21. P. 3609–3615.
- Tadepally H.D., Burger G., Aubry M. // *BMC Evol. Biol.* 2008. V. 8. P. 176.
- Kotova E.S., Akopov S.B., Sverdlov E.D., Nikolaev L.G. // *Biopol. Cell.* 2014. V. 30. P. 260–272.
- Kim S., Yu N.K., Kaang B.K. // *Exp. Mol. Med.* 2015. V. 47. P. e166.
- Cassandri M., Smirnov A., Novelli F., Pitolli C., Agostini M., Malewicz M., Melino G., Raschella G. // *Cell Death Discov.* 2017. V. 3. P. 17071.
- Fedotova A.A., Bonchuk A.N., Mogila V.A., Georgiev P.G. // *Acta Naturae.* 2017. V. 9. P. 47–58.
- Lukic S., Nicolas J.C., Levine A.J. // *Cell Death Differ.* 2014. V. 21. P. 381–387.
- Najafabadi H.S., Mnaimneh S., Schmitges F.W., Garton M., Lam K.N., Yang A., Albu M., Weirauch M.T., Radovani E., Kim P.M., et al. // *Nat. Biotechnol.* 2015. V. 33. P. 555–562.
- Brayer K.J., Segal D.J. // *Cell. Biochem. Biophys.* 2008. V. 50. P. 111–131.
- Burdach J., O'Connell M.R., Mackay J.P., Crossley M. // *Trends Biochem. Sci.* 2012. V. 37. P. 199–205.
- Aguet F., Ardlie K.G. // *Curr. Genet. Med. Repts.* 2016. V. 4. P. 163–169.
- Kim P., Park A., Han G., Sun H., Jia P., Zhao Z. // *Nucl. Acids Res.* 2018. V. 46. P. D1031–D1038.
- Wagner G.P., Kin K., Lynch V.J. // *Theory Biosci.* 2012. V. 131. P. 281–285.
- Sambrook J., Maniatis T., Fritsch E.F. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab. Press, 1987.
- R CoreTeam. R: A language and environment for statistical computing. Vienna, Austria: R. Fdn for Statistical Computing, 2014.
- Galili T. // *Bioinformatics.* 2015. V. 31. P. 3718–3720.
- Charrad M., Ghazzali N., Boiteau V., Niknafs A. // *J. Stat. Softw.* 2014. V. 61. P. 1–36.
- Brock G., Pihur V., Datta S., Datta S. // *J. Stat. Softw.* 2008. V. 25. P. 1–22.
- Tang Z., Li C., Kang B., Gao G., Zhang Z. // *Nucl. Acids Res.* 2017. V. 45. P. W98–W102.
- Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. // *Genome Res.* 2002. V. 12. P. 996–1006.
- ENCODE Project Consortium // *Nature.* 2012. V. 489. P. 57–74.
- GTEx Consortium // *Science.* 2015. V. 348. P. 648–660.
- Uhlen M., Hallstrom B.M., Lindskog C., Mardinoglu A., Ponten F., Nielsen J. // *Mol. Syst. Biol.* 2016. V. 12. P. 862.
- Gray K.A., Yates B., Seal R.L., Wright M.W., Bruford E.A. // *Nucl. Acids Res.* 2015. V. 43. P. D1079–D1085.
- Pierson E., Koller D., Battle A., Mostafavi S., Ardlie K.G., Getz G., Wright F.A., Kellis M., Volpi S., Dermitzakis E.T. // *PLoS Comput Biol.* 2015. V. 11. P. e1004220.
- Gainetdinov I.V., Kondratieva S.A., Skvortsova Y.V., Zinovyeva M.V., Stukacheva E.A., Klimov A., Tryakin A.A., Azhikina T.L. // *Oncotarget.* 2016. V. 7. P. 22439–22447.