

С. С. Ковалев^{1,2}, Е. Ю. Леберфарб^{1,2}, Н. В. Губанова¹, А. О. Брагин¹, А. Г. Галиева³
А. В. Цуканов^{1,3}, В. Н. Бабенко^{1,3}, Ю. Л. Орлов^{1,3}

¹ Институт цитологии и генетики СО РАН
пр. Академика Лаврентьева, 10, Новосибирск, 630090, Россия

² Новосибирский государственный медицинский университет
Красный проспект, 52, Новосибирск, 630091, Россия

³ Новосибирский государственный университет
ул. Пирогова, 1, Новосибирск, 630090, Россия

sergey.kovalev.1994@list.ru, orlov@bionet.nsc.ru

КОМПЬЮТЕРНЫЙ АНАЛИЗ АЛЬТЕРНАТИВНОГО СПЛАЙСИНГА ГЕНОВ В КУЛЬТУРАХ КЛЕТОК ГЛИОМ ПО ДАННЫМ RNA-SEQ *

Современные постгеномные методы изучения экспрессии генов с помощью транскриптомного профилирования имеют большое значение для фундаментальных биомедицинских исследований в онкологии, поиска новых маркеров развития опухолей на культурах клеток глиом. Такие эксперименты требуют разработки новых компьютерных инструментов анализа объемных данных секвенирования. Цель представленного исследования – компьютерный поиск генов и их изоформ, нарушение экспрессии которых связано с развитием глиобластом, с помощью современных высокопроизводительных технологий секвенирования транскриптом и международных биомедицинских банков данных. Поиск генов – кандидатов для терапевтического воздействия в опухолях, в том числе отдельных изоформ генов, актуален в здравоохранении и современной высокотехнологичной медицине. В данной работе представлены задачи биоинформатики, связанные с разработкой компьютерных конвейеров обработки транскриптомных данных, определения дифференциально экспрессирующихся генов, анализа альтернативного сплайсинга, описания категорий генных онтологий для найденных групп генов. Рассмотрены задачи автоматического поиска и описания функций генов в связи с раковыми заболеваниями, визуализации результатов и разработки биомедицинских баз данных. Представлен прототип базы данных дифференциального альтернативного сплайсинга генов – «Дифференциальный альтернативный сплайсинг генов человека при вторичной глиобластоме (ДАСГГ)», с возможностью работы через веб-сайт, поиска уровней экспрессии отдельных изоформ в глиальной опухоли.

Ключевые слова: биоинформатика, транскриптомика, биомедицинская информатика, глиобластома, альтернативный сплайсинг, базы данных.

Введение

Исследование экспрессии генов в клетках глиобластомы, поиск генов – кандидатов для терапевтического воздействия имеют несомненную актуальность в здравоохранении, совре-

* Работа была частично поддержана Министерством образования и науки РФ (28.12487.2018/12.1), разработка базы данных поддержана бюджетным проектом ИЦиГ СО РАН (№ 0324-2018-0019).

Первичный послеоперационный материал был любезно предоставлен А. Л. Кривошапкиным и А. С. Гайтаном (Национальный медицинский исследовательский центр им. академика Е. Н. Мешалкина, Новосибирск).

Авторы благодарны Геннадию Владимировичу Васильеву за организацию транскриптомного секвенирования, Ирине Вадимовне Медведевой, Роману Олеговичу Бабенко и Антону Геннадьевичу Богомолу за помощь в разработке компьютерной базы данных.

Ковалев С. С., Леберфарб Е. Ю., Губанова Н. В., Брагин А. О., Галиева А. Г., Цуканов А. В., Бабенко В. Н., Орлов Ю. Л. Компьютерный анализ альтернативного сплайсинга генов в культурах клеток глиом по данным RNA-seq // Вестн. НГУ. Серия: Информационные технологии. 2018. Т. 16, № 3. С. 22–36.

менной высокотехнологичной медицине [1]. Для такого сложного объекта, как опухоли мозга, требуется проведение новых исследований, опирающихся на современные клеточные технологии, технологии высокопроизводительного секвенирования, методы современной биоинформатики, использующие интеграцию имеющейся информации из международных баз и банков данных.

Глиальные опухоли составляют большинство первичных опухолей центральной нервной системы у взрослых и включают целый спектр опухолей, различающихся по уровню клеточной дифференциации и злокачественности. Глиобластома является наиболее распространенной (60 % от всех первичных опухолей) и злокачественной (выживаемость около 1 года после постановки диагноза) первичной опухолью центральной нервной системы у взрослых [1; 2]. Глиобластома может развиваться *de novo* (первичная) или как конечный этап трансформации фибриллярных астроцитом (II степень злокачественности согласно классификации ВОЗ) либо анапластических астроцитом (III степень злокачественности согласно классификации ВОЗ) (вторичная глиобластома) [2]. Первичная глиобластома в большинстве случаев встречается у лиц в возрасте старше 50 лет, и для нее характерен, как правило, короткий анамнез заболевания. Вторичная глиобластома чаще развивается в возрасте до 45 лет, трансформация в глиобластому может длиться от 1 до 10 лет. Злокачественные глиальные опухоли характеризуются ярким инвазивным фенотипом, отсутствием четких границ распространения опухоли и способностью к продолжению роста после хирургического удаления.

Несмотря на анализ отдельных мутаций и генов, далеко не все транскрипты, часто встречающиеся в глиальных опухолях, функционально аннотированы. Таким образом, большую актуальность имеет поиск новых маркеров развития глиом, разработка биоинформационных методов анализа глиом, в том числе разработка компьютерных баз данных по экспериментам транскриптомного профилирования [3]. Основные пути активации генов в опухолевых клетках были изучены ранее, в настоящее время необходимо более детальное исследование отдельных изоформ транскриптов, некодирующей РНК, которое может быть сделано с помощью современных компьютерных программ на основе технологий глубокого секвенирования [3; 4].

Секвенирование полного транскриптома (RNA-seq) с целью определения генов, ответственных за рост опухоли, на культурах клеток является современным молекулярно-биологическим методом исследования, послужившим основой данной работы. Биоинформационная часть исследования связана с использованием наиболее полных международных транскриптомных баз данных и аннотаций транскриптов, анализом отдельных изоформ генов для исследования глиобластомы. В данной работе выполнен перерасчет дифференциального альтернативного сплайсинга, с помощью конвейеров анализа данных [1]. Рассмотрены отдельные изоформы генов с дифференциальной экспрессией в культурах клеток глиом и культурах здоровой ткани.

Создан прототип компьютерной базы данных с возможностью поиска уровней экспрессии отдельных изоформ в глиальной опухоли. База содержит ссылки на международные ресурсы и базы данных генетической информации (в частности OMIM). Подготовлена заявка на госрегистрацию базы данных «Дифференциальный альтернативный сплайсинг генов человека при вторичной глиобластоме (ДАСТГ)» / Database «Differential Alternative Splicing of human Genes in secondary Glioblastome (DASGG)», ИЦиГ СО РАН, Новосибирск, 2018. База данных доступна по запросу к авторам. Представленные в базе данные по дифференциальному альтернативному сплайсингу могут быть использованы в фундаментальных исследованиях по стволовым клеткам рака, а также в разработке диагностики глиом.

Материалы и методы

Используемые базы данных

Использовалась аннотация генома и транскриптома человека Ensembl¹. Отметим, что, помимо общей аннотации генома, существуют специализированные международные базы данных по экспрессии генов в различных тканях и органах (включая микрочипы и транс-

¹ <http://ensembl.org/>

криптомные данные GEO NCBI², в том числе содержащие данные по экспрессии в клетках опухолей различных типов (The Cancer Gene Atlas³), экспрессии генов в компартментах мозга (Allen Brain Atlas), а также по транскриптомному профилированию опухолей, в том числе глиом и глиобластом⁴. В работе использовались общие базы данных белковых взаимодействий, такие как HPRD⁵, биохимических реакций KEGG⁶, Interactome⁷.

Институтом Аллена разработана база Ivy Glioblastoma Atlas Project⁸ по данным пациентов, страдающих глиомой. Авторами этой базы данных ранее выделено 343 гена, специфичных для глиом в различных структурах мозга. По запросу 'Glioma' в базе данных генов и фенотипов заболеваний OMIM (Online Mendelian Inheritance in Man) получен список генов, ассоциированных с глиомой по литературным данным⁹. Таким образом, были получены базовые списки генов человека для анализа.

Культуры клеток

В работе использовались данные транскриптомного секвенирования на клетках глиом и нормального мозга, полученные в ИЦиГ СО РАН в рамках завершеного проекта РФФИ в сотрудничестве с Национальным медицинским исследовательским центром имени академика Е. Н. Мешалкина, г. Новосибирск (см. прилож.).

Компьютерный конвейер анализа сплайсинга

Определение альтернативных вариантов изоформ генов по данным секвенирования представляет важную задачу биоинформатики. Альтернативный сплайсинг – процесс, в ходе которого экзоны генов, вырезаемые из пре-мРНК, объединяются в различных комбинациях (альтернативно), что порождает различные формы зрелой мРНК. Один ген может порождать не одну, а множество форм белка. Экзон одного варианта сплайсинга может оказаться интроном в альтернативном варианте. Поэтому молекулы мРНК, образованные в результате альтернативного сплайсинга, различаются набором экзонов, что приводит к образованию разных мРНК и, соответственно, разных белков одного первичного транскрипта. Сплайсинг гена может приводить к образованию разных изоформ белка, кодируемого этим геном [5]. На рис. 1 схематически представлены варианты сплайсинга гена и образующиеся изоформы.



Рис. 1. Варианты сплайсинга и возможные мутации. Адаптировано из [5]

Мутации генов в опухолевых клетках могут приводить к распространению нефункциональных изоформ белка, нарушению клеточной функции и пролиферации опухоли. Ранее считалось, что только мутации в кодирующей ДНК вызывают рак. В настоящее время показано, что изменения в альтернативном сплайсинге также провоцируют заболевание [4]; развиваются соответствующие научные направления поиска специфических паттернов сплайсинга [3], исследования процессинга РНК [6].

² <http://www.ncbi.nlm.nih.gov/geoprofiles/>.

³ cancergenome.nih.gov/.

⁴ <https://cghub.ucsc.edu/>.

⁵ <http://hprd.org/>.

⁶ <http://www.genome.jp/kegg/>.

⁷ <http://interactome.org/>.

⁸ <http://glioblastoma.alleninstitute.org/>.

⁹ <http://omim.org/>.

Существует несколько программ и конвейеров анализа данных сплайсинга. Нами разработан и применен конвейер по обработке данных РНК-секвенирования образцов глиом и здоровой ткани человека. Конвейер основан на использовании таких программ, как TopHat2 [7], Cufflinks [8], rMATS [9], Samtools [10], VCFtools [11], GSeq [12], и скриптов на языках программирования Perl, R и Bash. На первом этапе проводится картирование программой TopHat2 данных РНК-секвенирования на референсный геном человека GRCh38 с учетом экзон-интронной структуры (рис. 2).

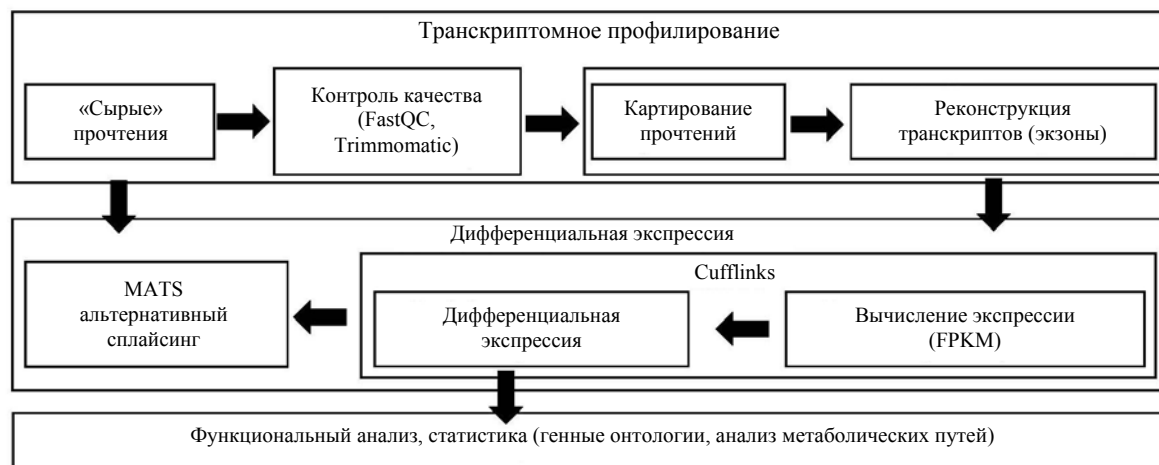


Рис. 2. Компьютерный конвейер обработки транскриптомных данных с учетом контроля качества прочтений [1]

На рис. 2 показаны этапы обработки данных – оценка качества прочтений, картирование с помощью TopHat. Далее проводится оценка экспрессии генов и выявление случаев дифференциальной экспрессии между анализируемыми образцами при помощи программ Cufflinks [8].

Составление списков геномных мутаций проводилось Samtools с последующим нахождением мутаций, встречающихся только в образцах глиомы, программами VCFtools. Найденные мутации сравнивались с мутациями глиомы из базы данных COSMIC [13]. Случаи альтернативного сплайсинга предсказывались программой rMATS согласно инструкции, приведенной на сайте программы.

Проводился поиск сверхпредставленных терминов генных онтологий (GO) интернет-ресурсом DAVID¹⁰ и GSeq по спискам дифференциально экспрессирующихся генов. Полученные списки генов, ассоциированных с глиомой, сравнивались со списками генов, опосредующих развитие глиомы, из литературы.

Конвейер реализован на языке программирования Bash под ОС Linux Redhat, использует многопоточность. На вход конвейера подаются данные секвенирования в формате fastq. Результат представляет собой набор файлов со следующей информацией: экспрессия генов в каждом из анализируемых образцов, дифференциально экспрессирующиеся гены, случаи альтернативного сплайсинга, сверхпредставленные GO термины, полиморфизмы в каждом из анализируемых образцов, подтвержденные случаи ассоциации с глиомой дифференциально экспрессирующихся генов и полиморфизмов.

Результаты

Список дифференциально экспрессирующихся генов в глиоме

В результате работы при сравнении тканей глиомы со здоровой тканью были найдены 8 284 случая дифференциальной экспрессии (ДЭ). ДЭ гены оказались сверхпредставлены

¹⁰ <https://david.ncifcrf.gov/>.

в процессах, связанных с межклеточным взаимодействием, клеточной гибелью, метаболизмом в клетке, генной экспрессией, что подтверждает связь выявленных случаев ДЭ с раковыми заболеваниями. На основе анализа базы данных OMIM и литературы выявлено 73 ДЭ гена, которые, как ранее показано, опосредуют развитие глиомы. Найдено 38 500 мутаций, специфичных глиоме, из которых более 3 000 представлены в базе данных COSMIC. Выявлено более 18 000 случаев альтернативного сплайсинга.

При сравнении тканей глиомы со здоровой тканью были найдены более восьми тысяч случаев дифференциальной экспрессии. Выявлено 73 дифференциально экспрессирующихся гена, которые опосредуют развитие глиомы.

Построена таблица категорий генных онтологий для генов со значимым дифференциальным сплайсингом в выборках клеток глиобластом и нормального мозга.

Категории генных онтологий генов с дифференциальным альтернативным сплайсингом в исследованных выборках глиом, рассчитанные с помощью инструмента DAVID

Категория	Термин	Число генов	<i>p</i> -value	Корректированное значение
GOTERM_CC_DIRECT	Extracellular exosome	26	8,3E-7	1,0E-4
GOTERM_CC_DIRECT	Focal adhesion	10	4,8E-6	2,9E-4
UP_KEYWORDS	Coiled coil	23	3,8E-5	2,4E-3
GOTERM_BP_DIRECT	Intracellular protein transport	6	7,4E-4	1,8E-1
GOTERM_CC_DIRECT	Brush border	4	9,3E-4	3,7E-2
INTERPRO	Proteinase inhibitor I2, Kunitz, conserved site	3	2,2E-3	3,4E-1
INTERPRO	Proteinase inhibitor I2, Kunitz metazoa	3	2,5E-3	2,1E-1
GOTERM_CC_DIRECT	Membrane	11	3,2E-3	9,2E-2
SMART	KU	3	3,3E-3	1,5E-1
KEGG_PATHWAY	Focal adhesion	5	4,0E-3	2,8E-1
INTERPRO	E F-Hand 1, calcium-binding site	5	4,5E-3	2,5E-1

Примечание: корректированное значение статистической значимости дано по критерию Бенджамини – Хохберга.

Таблица показывает связь категорий генных онтологий (рассчитано по DAVID для 73 генов из списка) с клеточной адгезией, мембраной, внеклеточными белками, что указывает на свойства пролиферации опухолей.

Анализ изоформ генов в глиобластоме

Для анализа участия изоформ этих генов в развитии глиом был выполнен поиск опубликованной информации в GenBank, OMIM и PubMed. Подтверждение в литературе об их роли в развитии глиомы имеют 123 гена.

В процессе анализа дифференциального сплайсинга были выявлены достоверные различия профилей сплайсинга в трех генах, связанных с возможной пролиферацией, между клетками нормального мозга и глиобластомы: белок-прекурсор амилоида бета APP (amyloid beta precursor protein), ген предрасположенности к раку CASC4 (cancer susceptibility candidate 4) и известный онкоген – транскрипционный фактор TP53. В частности, в гене TP53 наблюдалась некодирующая изоформа NR_015381 с достоверно большей частотой в клетках глиобластомы.

Короткая изоформа гена APP (NM_201413) высоко экспрессировалась в клетках глиобластомы, тогда как наиболее его длинная изоформа (NM_000484) специфически экспрессировалась в клетках нормального мозга. Для гена CASC4 его короткая изоформа (NM_177974) экспрессировалась в 6 раз выше в клетках глиобластомы, в то же время наиболее длинная изоформа (NM_138423) также имела высокий уровень экспрессии. Сверхэкспрессия CASC4 связана с повышенной экспрессией протоонкогена Her2 при раке яичника и молочной железы [11; 12].

В последние годы определены важные генетические мутации в глиомах. Ведущими мутациями в патогенезе злокачественных глиальных опухолей являются: потеря гетерозиготности (loss of heterozygosity – LOH) в длинном плече хромосомы 10 (LOH 10q), мутация гена PTEN (10q23.3), мутации в различных экзонах гена опухолевого супрессора p53, амплификация гена EGFR, делеция или инактивирующие мутации гена p16, а также гиперметилирование промотора гена MGMT. Эти мутации могут служить новым прогностическим фактором наряду с клиническими факторами прогноза и открывают новые перспективы и подходы в лечении ГО [16]. Последовательное изменение генов EGFR/PTEN/Akt/mTOR является основным патогенетическим путем развития первичной глиобластомы [17]. Амплификация гена EGFR встречается в 40% всех случаев первичных глиобластом и тесно связана с возрастом пациентов [18]. Мутация гена TP53 (p53) является основным событием, играющим роль в развитии вторичной глиобластомы [19].

Анализ роли изоформ в развитии опухолей

В ходе анализа дифференциальной экспрессии изоформ выявлены изоформы со статистически значимой разницей в экспрессии между культурой клеток нормального мозга и глиобластом. Ранее с помощью статистических подходов были выделены главные компоненты экспрессии изоформ гена APP в клетках нормального мозга и глиобластомы [1]. N-APP связывает TNFRSF21, запускающую активацию каспазы и дегенерацию тел нейрональных клеток (через каспазу-3) и аксонов (через каспазу-6). На рис. 3 представлены варианты сплайсинга гена APP.

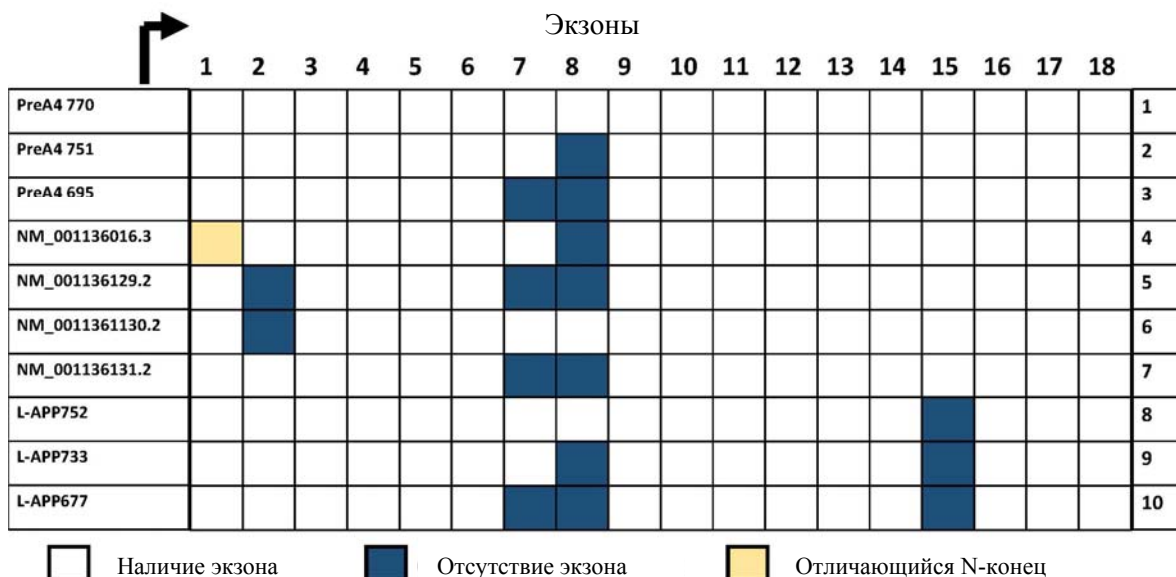


Рис. 3. Альтернативный сплайсинг гена APP

Известно десять вариантов транскрипции (см. рис. 3). Вариант транскрипции (1) представляет собой самый длинный транскрипт и кодирует самую длинную изоформу *a*, также известную как PreA4 770 (NM_000484).

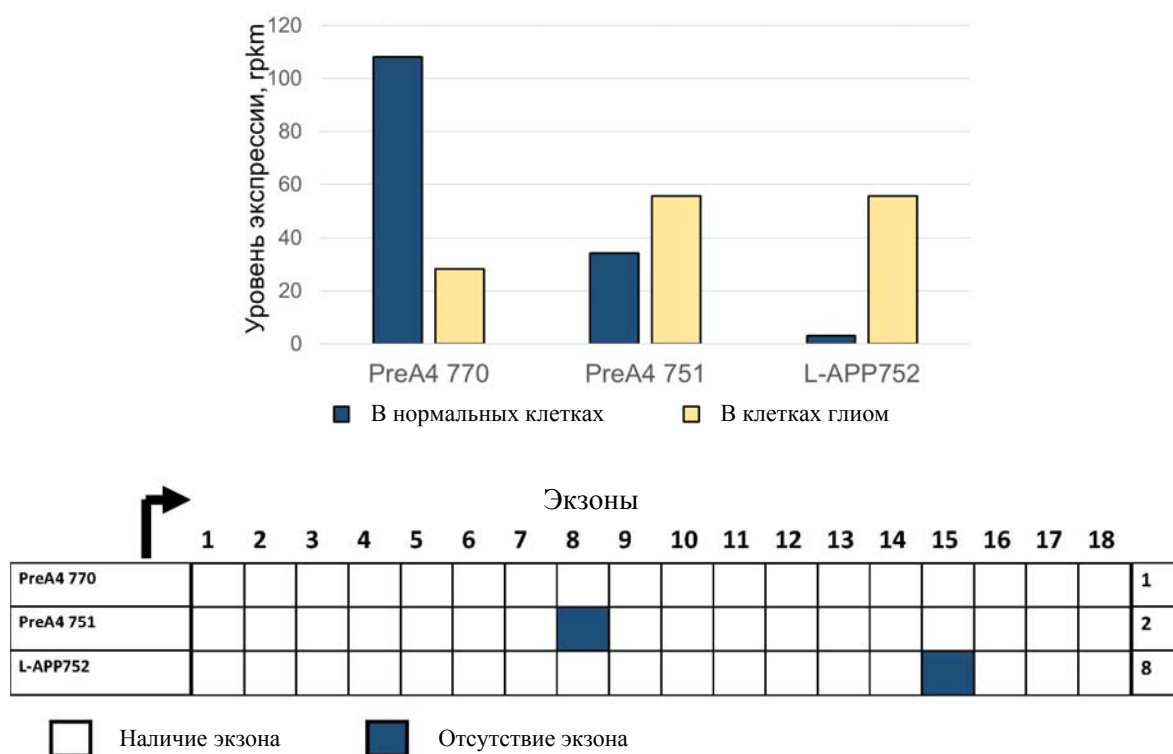


Рис. 4. Экспрессия изоформ гена APP в клетках глиом и нормальных клетках мозга (верхняя панель). Структура изоформ гена APP (нижняя панель). PreA4 751 и L-APP 752 имеют более короткие N- и C-концы и не имеют альтернативного экзона по сравнению с самой длинной изоформой PreA4 770

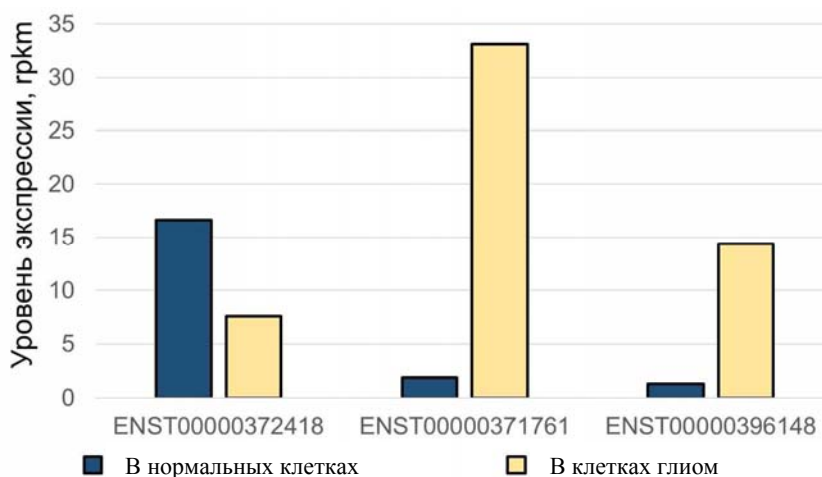


Рис. 5. Экспрессия изоформ гена CDKN2A в нормальных клетках и клетках глиом

Вариант транскрипции (2) не имеет альтернативного экзона в рамке считывания по сравнению с вариантом 1. Данная изоформа *b*, также известная как PreA4 751, имеет те же N- и C-концы, но короче по сравнению с изоформой *a*.

Вариант транскрипции (3) не имеет альтернативного внутрикадрового сегмента по сравнению с вариантом 1. Полученная изоформа *c*, также известная как PreA4 695, имеет те же N- и C-концы, но короче по сравнению с изоформой *a*.

Вариант транскрипции (4) отличается в 5' UTR и кодирующей последовательности и не имеет альтернативного экзона в кадре по сравнению с вариантом 1. Данная изоформа *d* имеет

более короткий и отличный N-конец и не имеет внутреннего сегмента по сравнению с изоформой *a*.

Вариант транскрипции (5) не имеет трех альтернативных экзонов в той же рамке считывания по сравнению с вариантом 1. Полученная изоформа *e* имеет те же N- и C-концы, но короче по сравнению с изоформой *a*.

У варианта транскрипции (6) отсутствует альтернативный экзон в той же рамке считывания по сравнению с вариантом 1. Полученная изоформа *f* имеет те же N- и C-концы, но короче по сравнению с изоформой *a*.

Вариант транскрипции (7) отличается в 5'UTR и кодирующей последовательности и не имеет двух альтернативных экзонов в той же рамке считывания по сравнению с вариантом 1. Полученная изоформа *g* короче на N-конце и не имеет внутреннего сегмента по сравнению с изоформой *a*.

Вариант транскрипции (8) не имеет альтернативного экзона в кадре по сравнению с вариантом 1. Полученная изоформа *h*, также известная как L-APP752, имеет те же N- и C-концы, но короче по сравнению с изоформой *a*. Для этой транскрипции нет полной расшифровки стенограммы.

Вариант транскрипции (9) не имеет двух альтернативных экзонов в той же рамке считывания по сравнению с вариантом 1. Полученная изоформа *i*, также известная как L-APP733, имеет те же N- и C-концы, но короче по сравнению с изоформой *a*. Для этой транскрипции нет полной расшифровки стенограммы.

Вариант транскрипции (10) не имеет трех альтернативных экзонов в кадре по сравнению с вариантом 1. Полученная изоформа *j*, также известная как L-APP677, имеет те же N- и C-концы, но короче по сравнению с изоформой *a*. Для этой транскрипции нет полной расшифровки стенограммы.

В качестве примера представлены диаграммы дифференциальной экспрессии изоформ APP и CDKN2A (рис. 4 и 5).

В ходе анализа экспрессии изоформ APP выявлено, что более короткие изоформы PreA4 751 и L-APP 752 больше экспрессируются в клетках глиом, нежели каноничная изоформа PreA4 770, которая больше экспрессируется в нормальных клетках мозга.

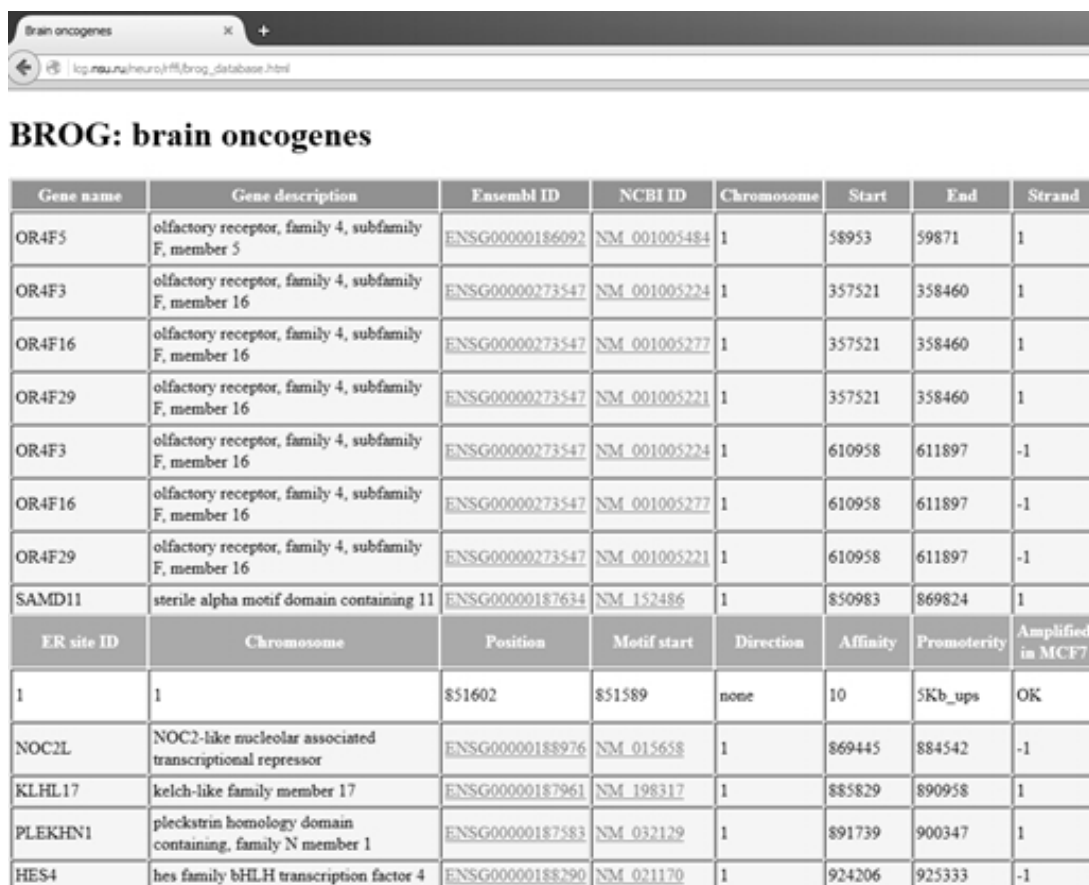
Для гена CDKN2A характерна такая же картина, что для гена APP. Более короткие изоформы ENST00000371761 и ENST00000396148 экспрессируются выше в клетках глиом, в то время как самая длинная изоформа ENST00000372418 экспрессируется выше в нормальных клетках мозга.

Отметим, что для TP53 весь спектр (37 видимых на наших данных изоформ из 43 известных) изоформ, наблюдаемых в клетках глиобластом, отличался по уровню экспрессии от изоформ в нормальных клетках мозга. В частности, некодирующая изоформа NR_015381 относится только к клеткам глиобластом.

База данных экспрессии генов и альтернативного сплайсинга на культурах клеток глиом

В настоящее время в открытом доступе находятся данные экспериментов секвенирования в опухолевых клетках, в том числе по различным типам глиом. Опубликованы статьи и материалы, содержащие информацию о генах, специфично экспрессирующихся в раковых клетках, о профилях метилирования и т. д. Эта информация нуждается в аккумулировании для дальнейшего использования в медицине. Ранее авторским коллективом была разработана компьютерная база данных BROG¹¹ генов-мишеней онкогенов, определенных по данным экспериментов ChIP-seq (включая транскрипционные факторы ER, MYC, TP53). В базе была представлена разметка сайтов связывания транскрипционных факторов; данные о сайтах связывания были получены по открытым публикациям и ресурсам GEO NCBI. На рис. 6 представлен фрагмент интерфейса этой базы данных.

¹¹ База представлена на сайте http://lcg.nsu.ru/neuro/rffi/brog_database.html.



The screenshot shows a web browser window with the address log.msu.ru/heuroinf/brog_database.html. The page title is "BROG: brain oncogenes". Below the title is a table with 8 columns: Gene name, Gene description, Ensembl ID, NCBI ID, Chromosome, Start, End, and Strand. The table lists several genes, including OR4F5, OR4F3, OR4F16, OR4F29, SAMD11, NOC2L, KLHL17, PLEKHN1, and HES4. Below the main table is a smaller table with 8 columns: ER site ID, Chromosome, Position, Motif start, Direction, Affinity, Promoterity, and Amplified in MCF7. The first row of this smaller table shows ER site ID 1, Chromosome 1, Position 851602, Motif start 851589, Direction none, Affinity 10, Promoterity 5Kb_ups, and Amplified in MCF7 OK.

Gene name	Gene description	Ensembl ID	NCBI ID	Chromosome	Start	End	Strand
OR4F5	olfactory receptor, family 4, subfamily F, member 5	ENSG00000186092	NM_001005484	1	58953	59871	1
OR4F3	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005224	1	357521	358460	1
OR4F16	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005277	1	357521	358460	1
OR4F29	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005221	1	357521	358460	1
OR4F3	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005224	1	610958	611897	-1
OR4F16	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005277	1	610958	611897	-1
OR4F29	olfactory receptor, family 4, subfamily F, member 16	ENSG00000273547	NM_001005221	1	610958	611897	-1
SAMD11	sterile alpha motif domain containing 11	ENSG00000187634	NM_152486	1	850983	869824	1
ER site ID	Chromosome	Position	Motif start	Direction	Affinity	Promoterity	Amplified in MCF7
1	1	851602	851589	none	10	5Kb_ups	OK
NOC2L	NOC2-like nucleolar associated transcriptional repressor	ENSG00000188976	NM_015658	1	869445	884542	-1
KLHL17	kelch-like family member 17	ENSG00000187961	NM_198317	1	885829	890958	1
PLEKHN1	pleckstrin homology domain containing, family N member 1	ENSG00000187583	NM_032129	1	891739	900347	1
HES4	hes family bHLH transcription factor 4	ENSG00000188290	NM_021170	1	924206	925333	-1

Рис. 6. Фрагмент разработанной ранее базы данных генов-мишеней транскрипционных факторов – BROG (Brain Oncogenes)

- Научная задача (Главная страница)
 - Цели Проекта
 - Методы и подходы, использованные в ходе выполнения Проекта
 - Актуальность проблемы
 - Результаты
- Демонстрация работы
- Базы данных
- Публикации

Рис. 7. Структура сайта по проекту анализа глиом, содержащего обновленную базу данных

Для написания собственного сайта была применена среда разработки WordPress. Сайт располагается по адресу: <http://gliomaicigsbras.ru>. База пополнена данными по альтернативному сплайсингу, ссылками на международные ресурсы экспрессии генов в глиомах, системой навигации. В настоящее время идет патентование (Заявка на получение свидетельства регистрации Базы данных «Дифференциальный альтернативный сплайсинг генов человека при вто-

ричной глиобластоме (ДАСГГ)» / Database «Differential Alternative Splicing of human Genes in secondary Glioblastome (DASGG)»). На рис. 7 показана структура сайта, представлены блоки содержащие информацию о научной работе, базы данных, публикации и контакты участников проекта.

Структура сайта состоит из трех блоков. Первый блок посвящен информации о научной работе. Здесь можно ознакомиться с целями и задачами научного проекта. В «Методах и подходах, использованных в ходе выполнения проекта» подробно описана методика получения первичных культур, использованных в дальнейшем для RNA-Seq. Там же есть информация о компьютерном конвейере анализа данных секвенирования транскриптом, комплексе программ на языке Java для статистического анализа расположения генов на пространственных топологических доменах и известных частях хромосом, данных по экспрессии генов. На странице «Демонстрация работы» размещены изображения, сделанные в ИЦиГ СО РАН, часть которых была опубликована. В разделе «Публикации» представлен перечень статей, опубликованных в журналах и сборниках.

В следующем блоке даны контакты участников проекта и ссылки на международные базы данных.

Далее располагается база данных, находящаяся в свободном доступе. Здесь предоставлена информация о дифференциальной экспрессии генов.

На рис. 8 представлен скриншот разработанной базы, показывающий уровни экспрессии генов в глиоме и здоровых клетках, и статистические параметры дифференциальной экспрессии. Также на фрагменте таблицы базы данных указаны идентификаторы Ensembl транскрипта, название гена, уровни экспрессии в выборках, статистическая достоверность различий.

test_id	gene_id	gene	locus	NB	NGB	log2(fold_change)	test_stat	p_value	q_value
ENST00000318602	ENSG00000175899	A2M	12:906517	11,1173	30,0908	1,43652	7,19784	5,00E-05	0,000802281
ENST00000316519	ENSG00000081760	AACS	12:125065	3,66967	1,46472	-1,32503	-3,49974	5,00E-05	0,000802281
ENST00000261772	ENSG00000090861	AARS	16:702522	20,116	13,6244	-0,562153	-2,89239	5,00E-05	0,000802281
ENST00000473553	ENSG00000008311	AASS	7:1220756	1,69523	4,99477	1,55894	3,39916	5,00E-05	0,000802281
ENST00000619387	ENSG00000275700	AATF	17:369488	18,0664	6,95821	-1,37652	-6,72055	5,00E-05	0,000802281
ENST00000629058	ENSG00000281376	ABALON	20:316644	18,0576	9,74291	-0,890185	-3,08979	5,00E-05	0,000802281
ENST00000374736	ENSG00000165029	ABCA1	9:1047810	5,54539	1,81719	-1,60958	-8,18991	5,00E-05	0,000802281
ENST00000435803	ENSG00000179869	ABCA13	7:4817145	0,69576	0,03087	-4,49433	-6,08813	5,00E-05	0,000802281
ENST00000295750	ENSG00000115657	ABCB6	2:2192097	4,61558	0,981341	-2,23369	-3,51466	5,00E-05	0,000802281
ENST00000503337	ENSG00000108846	ABCC3	17:506347	0,201968	3,07076	3,9264	3,93899	5,00E-05	0,000802281
ENST00000376887	ENSG00000125257	ABCC4	13:950198	1,26256	4,29679	1,76691	5,92342	5,00E-05	0,000802281
ENST00000427120	ENSG00000114770	ABCC5	3:1839195	0,232827	1,24171	2,41499	3,51256	5,00E-05	0,000802281
ENST00000296577	ENSG00000164163	ABCE1	4:1449671	17,9483	31,5973	0,815955	2,86213	5,00E-05	0,000802281
ENST00000222388	ENSG00000033050	ABCF2	7:1512078	4,58501	9,74052	1,08707	2,55459	5,00E-05	0,000802281

Страница 1 из 151

Рис. 8. Фрагмент данных (<http://gliomaicgsbras.ru/базы-данных/>)

Выводы и обсуждение

При помощи анализа данных современных высокопроизводительных технологий секвенирования транскриптом был выполнен компьютерный поиск генов, нарушение экспрессии которых связано с развитием глиобластом. Показана роль изоформ генов при развитии глиобластомы, дана функциональная аннотация [17–23]. Разработанный компьютерный конвейер может быть использован для решения аналогичных биомедицинских задач, основанных на обработке данных RNA-Seq [3; 4].

Исследование роли альтернативного сплайсинга при глиоме мозга на культурах клеток проводилось по данным RNA-seq, полученным в ИЦиГ СО РАН [1]. Высокопроизводитель-

ное секвенирование полного транскриптома культуры клеток (RNA-seq) с целью определения генов, ответственных за рост опухоли, является современным методом исследования, который должен шире использоваться в медицинской практике [22; 23].

Работа по поиску маркеров развития глиом имеет большую практическую значимость для медицины. База данных ДАСГГ предназначена для медиков и исследователей, которые заинтересованы в получении информации об альтернативном сплайсинге при опухолях на первичных культурах клеток. Представленные данные могут быть использованы в фундаментальных исследованиях по стволовым клеткам глиом и в разработке диагностик.

Список литературы

1. Babenko V. N., Gubanova N. V., Bragin A. O., Chadaeva I. V., Vasiliev G. V., Medvedeva I. V., Gaytan A. S., Krivoschapkin A. L., Orlov Yu. L. Computer Analysis of Glioma Transcriptome Profiling: Alternative Splicing Events // *J. Integr. Bioinform.* 2017. Vol. 14. No. 3. PII: /j/jib.2017.14.issue-3/jib-2017-0022/jib-2017-0022.xml. DOI: 10.1515/jib-2017-0022.
2. Ohgaki H., Kleihues P. Genetic Pathways to Primary and Secondary Glioblastoma // *Am. J. Pathol.* 2007. Vol. 170. No. 5. P. 1445–1453.
3. Ferrarese R., Harsh G. R. 4th, Yadav A. K., Bug E., Maticzka D., Reichardt W., Dombrowski S. M., Miller T. E., Masilamani A. P., Dai F., Kim H., Hadler M., Scholtens D. M., Yu I. L., Beck J., Srinivasasainagendra V., Costa F., Baxan N., Pfeifer D., von Elverfeldt D., Backofen R., Weyerbrock A., Duarte C. W., He X., Prinz M., Chandler J. P., Vogel H., Chakravarti A., Rich J. N., Carro M. S., Bredel M. Lineage-specific splicing of a brain-enriched alternative exon promotes glioblastoma progression // *J. Clin. Invest.* 2014. Vol. 124. No. 7. P. 2861–2876.
4. Correa B. R., de Araujo P. R., Qiao M., Burns S. C., Chen C., Schlegel R., Agarwal S., Galante P. A., Penalva L. O. Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma // *Genome Biol.* 2016. Vol. 17. No. 1. P. 125.
5. Thorsen F., Tysnes B. B. Brain tumor cell invasion, anatomical and biological considerations // *Anticancer Res.* 1997. Vol. 17. No. 6B. P. 4121–4126.
6. Marcelino Meliso F., Hubert C. G., Favoretto Galante P. A., Penalva L. O. RNA processing as an alternative route to attack glioblastoma // *Hum. Genet.* 2017. Vol. 136. No. 9. P. 1129–1141.
7. Kim N. K., Jayatillake R. V., Spouge J. L. NEXT-peak: A normal-exponential two-peak model for peak-calling in ChIP-seq data // *BMC Genomics.* 2013. Vol. 14. No. 1. P. 349.
8. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. R., Pimentel H., Salzberg S. L., Rinn J. L., Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks // *Nat. Protoc.* 2012. Vol. 7. No. 3. P. 562–578.
9. Shen S., Park J. W., Lu Z. X., Lin L., Henry M. D., Wu Y. N., Zhou Q., Xing Y. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data // *Proc. Natl. Acad. Sci.* 2014. Vol. 111. No. 51. P. E5593–E5601.
10. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment / Map format and SAMtools // *Bioinformatics.* 2009. Vol. 25. No. 16. P. 2078–2079.
11. Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T., Sherry S. T., McVean G., Durbin R. 1000 Genomes Project Analysis Group. The variant call format and VCFtools // *Bioinformatics.* 2011. Vol. 27. No. 15. P. 2156–2158.
12. Chen C. M., Lu Y. L., Sio C. P., Wu G. C., Tzou W. S., Pai T. W. Gene Ontology based housekeeping gene selection for RNA-seq normalization // *Methods.* 2014. Vol. 67. No. 3. P. 354–363.
13. Bamford S., Dawson E., Forbes S., Clements J., Pettett R., Dogan A., Flanagan A., Teague J., Futreal P. A., Stratton M. R., Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website // *Br. J. Cancer.* 2004. Vol. 91. No. 2. P. 355–358.
14. Inoue K., Fry E. A. Aberrant splicing of estrogen receptor, HER2, and CD44 genes in breast cancer // *Genet. Epigenetics.* 2015. Vol. 1. No. 7. P. 19–32.

15. *Oh J. J., Grosshans D. R., Wong S. G., Slamon D. J.* Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells // *Nucleic Acids Res.* 1999. Vol. 27. No. 20. P. 4008–4017.
16. *Van Meir E. G., Hadjipanayis C. G., Norden A. D., Shu H. K., Wen P. Y., Olson J. J.* Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma // *CA. Cancer J. Clin.* 2010. Vol. 60. No. 3. P. 166–193.
17. *Kita D., Yonekawa Y., Weller M., Ohgaki H.* PIK3CA alterations in primary (de novo) and secondary glioblastomas // *Acta Neuropathol.* 2007. Vol. 113. No. 3. P. 295–302.
18. *Ekstrand A. J., Sugawa N., James C. D., Collins V. P.* Amplified and rearranged epidermal growth factor receptor genes in human glioblastomas reveal deletions of sequences encoding portions of the N- and/or C-terminal tails // *Proc. Natl. Acad. Sci.* 1992. Vol. 89. No. 10. P. 4309–4313.
19. *Hulleman E., Helin K.* Molecular mechanisms in gliomagenesis // *Adv. Cancer Res.* 2005. Vol. 94. No. 1 (Suppl). P. 1–27.
20. *Gubanova N. V., Orlov Yu. L., Bragin A. O., Vasiliev G. V., Babenko V. N., Kovalev S. S., Gaytan A. S.* Transcriptome profiling of primary glioma cell cultures // *Proc. of MCCMB-2017, Moscow.* ISBN 978-5-901158-30-2.
21. *Брагин А. О., Губанова Н. В., Ковалев С. С., Бабенко В. Н., Орлов Ю. Л.* Анализ альтернативного сплайсинга в глиомах с помощью секвенирования транскриптом // II Междунар. науч.-практ. конф. «NGS в медицинской генетике 2017»: Тез. докл. Суздаль, 2017. С. 25.
22. *Gubanova N. V., Bragin A. O., Kovalev S. S., Medvedeva I. V., Babenko V. N., Gaytan A. S., Krivoshapkin A. L., Orlov Yu. L.* Analysis of nuclear pore complex genes in glioblastoma by transcriptome profiling // *Abstracts of Symposium «Cognitive Sciences, Genomics and Bioinformatics (CSGB-2016)» / Scientific Research Institute of Physiology and Basic Medicine.* Novosibirsk, 2016. P. 21. ISBN 978-5-91291-029-6.
23. *Ковалев С. С., Орлов Ю. Л., Леберфарб Е. Ю.* Дифференциальная экспрессия и альтернативный сплайсинг в культурах клеток глиом по данным RNA-seq // *Авиценна-2018: Сб. тез. докл.* Новосибирск: Изд-во НГМУ, 2018.

Приложение

РАБОТА С КУЛЬТУРАМИ КЛЕТОК

В работе использовались данные по первичным культурам клеток глиом и здоровых клеток мозга, полученные в ИЦиГ СО РАН (послеоперационный клеточный материал был получен в сотрудничестве с Национальным медицинским исследовательским центром имени академика Е. Н. Мешалкина, Новосибирск). Впервые сравнивались культуры злокачественной и здоровой ткани мозга, полученные в одинаковых условиях [1]. Первичный послеоперационный материал для культур клеток был любезно предоставлен А. Л. Кривошапкиным и А. С. Гайтаном (получено разрешение этического комитета Национального медицинского исследовательского центра им. академика Е. Н. Мешалкина).

Клетки культур были проанализированы методом иммуноокрашивания на маркеры: астроцитов – GFAP (Glial fibrillary acidic protein) и нейронов – beta III tubulin. Вторичная глиобластома была получена от пациента, перенесшего операцию по удалению первичной глиобластомы, а затем прошедшего курс химио- и радиотерапии. Иммуноцитохимическое окрашивание клеток культуры выявило, что все клетки культуры позитивны по нейральному и астроцитарному маркерам. Отбор материала для секвенирования производился, когда количество клеток достигало 8–10 млн, что соответствовало 2–3 пассажу в зависимости количества исходного материала.

Суммарная РНК из клеточных культур выделялась с использованием реагента Trizol (Ambion), для выделения РНК с культурального планшета удалялась среда, немедленно без промывки PBS добавлялось 2 мл Trizol, и клетки лизировались, дальнейшие процедуры проводились согласно протоколу изготовителя, для каждой культуры (здоровых клеток мозга и глиомы) использовались по 3 отдельных культуральных планшета. После выделения проводилась оценка качества РНК на биоанализаторе BA2100 набором RNA Nano.

Для создания бар-кодированных RNA-Seq библиотек было взято по 30 нг РНК, использован набор ScriptSeq™ v2 RNASeq Library Preparation Kit (Epicentre) согласно протоколу изготовителя, при амплификации библиотек использовано 12 циклов ПЦР, финальная очистка проводилась на магнитных шариках AMPure XP. Качество полученных библиотек и их молярность проверены на биоанализаторе BA2100 набором DNA High Sensitivity, перед нанесением библиотеки разбавлялись 1 : 10. Молярность 6-ти полученных библиотек находилась в пределах 78 000–157 000 pMol/l. Аликвоты полученных библиотек переданы в ЗАО «Гено-аналитика», где было выполнено секвенирование на приборе Illumina HiSeq (одностороннее прочтения размером 50 нт). Создано по 3 библиотеки транскриптомов глиом и здорового мозга. Полученные библиотеки секвенированы с глубиной прочтения по 20 млн прочтений ДНК каждая. Анализ внутреннего контроля ERCC Spike-In Mix показал отсутствие искажений представленности при приготовлении библиотек и секвенировании.

Материал поступил в редколлегию 18.07.2018

S. S. Kovalev^{1,2}, **E. Yu. Lieberfarb**^{1,2}, **N. V. Gubanova**¹, **A. O. Bragin**¹
A. G. Galieva¹, **A. V. Tsukanov**^{1,3}, **V. N. Babenko**^{1,3}, **Yu. L. Orlov**^{1,3}

¹*Institute of Cytology and Genetics SB RAS
10 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation*

²*Novosibirsk State Medical University
52 Krasnyi prospekt, Novosibirsk, 630091, Russian Federation*

³*Novosibirsk State University
1 Pirogov Str., Novosibirsk, 630090, Russian Federation*

sergey.kovalev.1994@list.ru, orlov@bionet.nsc.ru

COMPUTER ANALYSIS OF GENE ALTERNATIVE SPLICING IN GLIOMA CELL CULTURES BY RNA-seq DATA

Fundamental biomedical research in oncology, the search for new markers of tumor development, modern post-genomic studies of gene expression on cell cultures need glioma transcriptome profiling and analysis of individual gene isoforms. Such experiments, in turn, require development of new computer tools and database for analysis of bulk sequencing data. The aim of our study is a computer search for genes and gene isoforms, the difference of their expression is associated with the development of glioblastoma. The work is based on modern high-throughput sequencing technologies and international biomedical data banks analysis. The search for candidate genes in tumors for therapeutic treatment, including individual gene isoforms, is very relevant in healthcare and modern high-tech medicine. This work presents the bioinformatics problems related to the development of computer pipelines for the processing of transcriptomic data, the revealing of the differentially expressed genes, the analysis of alternative splicing, and the description of the gene ontologies categories for the genes sets found. The tasks of automatic search and description of gene functions in connection with cancer diseases, visualization of results and development of biomedical databases are considered. A prototype database of differential alternative splicing of genes is presented, «Differential Alternative Splicing of Human Genes in Secondary Glioblastoma (DASGG)», with the ability to work through a website, to search for expression levels of individual isoforms in tumor cells.

Keywords: bioinformatics, transcriptomics, biomedical informatics, glioblastoma, alternative splicing, databases.

References

1. Babenko V. N., Gubanova N. V., Bragin A. O., Chadaeva I. V., Vasiliev G. V., Medvedeva I. V., Gaytan A. S., Krivoshepkin A. L., Orlov Yu. L. Computer Analysis of Glioma Trans-

criptome Profiling: Alternative Splicing Events. *J. Integr. Bioinform.*, 2017, vol. 14, no. 3. PII: /j/jib.2017.14.issue-3/jib-2017-0022/jib-2017-0022.xml. DOI: 10.1515/jib-2017-0022.

2. Ohgaki H., Kleihues P. Genetic Pathways to Primary and Secondary Glioblastoma. *Am. J. Pathol.*, 2007, vol. 170, no. 5, p. 1445–1453.

3. Ferrarese R., Harsh G. R. 4th, Yadav A. K., Bug E., Maticzka D., Reichardt W., Dombrowski S. M., Miller T. E., Masilamani A. P., Dai F., Kim H., Hadler M., Scholtens D. M., Yu I. L., Beck J., Srinivasasainagendra V., Costa F., Baxan N., Pfeifer D., von Elverfeldt D., Backofen R., Weyerbrock A., Duarte C. W., He X., Prinz M., Chandler J. P., Vogel H., Chakravarti A., Rich J. N., Carro M. S., Bredel M. Lineage-specific splicing of a brain-enriched alternative exon promotes glioblastoma progression. *J. Clin. Invest.*, 2014, vol. 124, no. 7, p. 2861–2876.

4. Correa B. R., de Araujo P. R., Qiao M., Burns S. C., Chen C., Schlegel R., Agarwal S., Galante P. A., Penalva L. O. Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRNPB as an oncogenic candidate in glioblastoma. *Genome Biol.*, 2016, vol. 17, no. 1, p. 125.

5. Thorsen F., Tysnes B. B. Brain tumor cell invasion, anatomical and biological considerations. *Anticancer Res.*, 1997, vol. 17, no. 6B, p. 4121–4126.

6. Marcelino Meliso F., Hubert C. G., Favoretto Galante P. A., Penalva L. O. RNA processing as an alternative route to attack glioblastoma. *Hum. Genet.*, 2017, vol. 136, no. 9, p. 1129–1141.

7. Kim N. K., Jayatilake R. V., Spouge J. L. NEXT-peak: A normal-exponential two-peak model for peak-calling in ChIP-seq data. *BMC Genomics*, 2013, vol. 14, no. 1, p. 349.

8. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. R., Pimentel H., Salzberg S. L., Rinn J. L., Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, 2012, vol. 7, no. 3, p. 562–578.

9. Shen S., Park J. W., Lu Z. X., Lin L., Henry M. D., Wu Y. N., Zhou Q., Xing Y. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.*, 2014, vol. 111, no. 51, p. E5593–E5601.

10. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment / Map format and SAMtools. *Bioinformatics*, 2009, vol. 25, no. 16, p. 2078–2079.

11. Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T., Sherry S. T., McVean G., Durbin R. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 2011, vol. 27, no. 15, p. 2156–2158.

12. Chen C. M., Lu Y. L., Sio C. P., Wu G. C., Tzou W. S., Pai T. W. Gene Ontology based housekeeping gene selection for RNA-seq normalization. *Methods*, 2014, vol. 67, no. 3, p. 354–363.

13. Bamford S., Dawson E., Forbes S., Clements J., Pettett R., Dogan A., Flanagan A., Teague J., Futreal P. A., Stratton M. R., Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer*, 2004, vol. 91, no. 2, p. 355–358.

14. Inoue K., Fry E. A. Aberrant splicing of estrogen receptor, HER2, and CD44 genes in breast cancer. *Genet. Epigenetics*, 2015, vol. 1, no. 7, p. 19–32.

15. Oh J. J., Grosshans D. R., Wong S. G., Slamon D. J. Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells. *Nucleic Acids Res.*, 1999, vol. 27, no. 20, p. 4008–4017.

16. Van Meir E. G., Hadjipanayis C. G., Norden A. D., Shu H. K., Wen P. Y., Olson J. J. Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA. Cancer J. Clin.*, 2010, vol. 60, no. 3, p. 166–193.

17. Kita D., Yonekawa Y., Weller M., Ohgaki H. PIK3CA alterations in primary (de novo) and secondary glioblastomas. *Acta Neuropathol.*, 2007, vol. 113, no. 3, p. 295–302.

18. Ekstrand A. J., Sugawa N., James C. D., Collins V. P. Amplified and rearranged epidermal growth factor receptor genes in human glioblastomas reveal deletions of sequences encoding portions of the N- and/or C-terminal tails. *Proc. Natl. Acad. Sci.*, 1992, vol. 89, no. 10, p. 4309–4313.

19. Hulleman E., Helin K. Molecular mechanisms in gliomagenesis. *Adv. Cancer Res.*, 2005, vol. 94, no. 1 (suppl.), p. 1–27.

20. Gubanov N. V., Orlov Yu. L., Bragin A. O., Vasiliev G. V., Babenko V. N., Kovalev S. S., Gaytan A. S. Transcriptome profiling of primary glioma cell cultures. *Proc. of MCCMB-2017. Moscow*, 2017. ISBN 978-5-901158-30-2.

21. Bragin A. O., Gubanova N. V., Kovalev S. S., Babenko V. N., Orlov Yu. L. Analysis of alternative splicing in glioma using transcriptome sequencing. II International Science-Practical Conference «NGS in medical genetics 2017». Proc. Suzdal, 2017, p. 25. (in Russ.)

22. Gubanova N. V., Bragin A. O., Kovalev S. S., Medvedeva I. V., Babenko V. N., Gaytan A. S., Krivoshapkin A. L., Orlov Yu. L. Analysis of nuclear pore complex genes in glioblastoma by transcriptome profiling. *Abstracts of Symposium «Cognitive Sciences, Genomics and Bioinformatics (CSGB-2016)»*. Scientific Research Institute of Physiology and Basic Medicine. Novosibirsk, 2016, p. 21. ISBN 978-5-91291-029-6.

23. Kovalev S. S., Orlov Yu. L., Lieberfarb E. Yu. Differential expression and alternative splicing in glioma cell culture by RNA-seq data. *Abstract book of students conference "Avicenna-2018"*. Novosibirsk, NSMU Press, 2018. (in Russ.)

For citation:

Kovalev S. S., Lieberfarb E. Yu., Gubanova N. V., Bragin A. O., Galieva A. G., Tsukanov A. V., Babenko V. N., Orlov Yu. L. Computer Analysis of Gene Alternative Splicing in Glioma Cell Cultures by RNA-seq Data. *Vestnik NSU. Series: Information Technologies*, 2018, vol. 16, no. 3, p. 22–36. (in Russ.)

DOI 10.25205/1818-7900-2018-16-3-22-36