

УДК 577.21:577.29:004.42

КОМПЬЮТЕРНЫЙ АНАЛИЗ ДАННЫХ ЭКСПРЕССИИ ГЕНОВ В КЛЕТКАХ МОЗГА, ПОЛУЧЕННЫХ С ПОМОЩЬЮ МИКРОЧИПОВ И ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ

© 2013 г. **И.В. Медведева, О.В. Вишневский, Н.С. Сафронова,
О.С. Кожевникова, М.А. Генаев, Д.А. Афонников, А.В. Кочетов, Ю.Л. Орлов**

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики
Сибирского отделения Российской академии наук, e-mail: orlov@bionet.nsc.ru

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

В последние годы происходит стремительное расширение фронта нейробиологических исследований, сопровождающееся бурным ростом объема экспериментальных данных по структуре, функции и эволюции нервной системы на различных уровнях ее иерархической организации. Использование технологий высокопроизводительного секвенирования и микрочипов позволяет проводить сравнительный статистический анализ экспрессии тысяч генов одновременно, учитывая при этом пространственное расположение клеток в структурах мозга. Дан краткий обзор основных подходов анализа экспрессии генов в клетках мозга. Проанализированы особенности структуры генов, имеющих дифференциальную экспрессию в клетках мозга. Оценивалось число экзонов, альтернативных транскриптов и его соотношение с уровнем экспрессии. Показано статистическое различие числа альтернативных транскриптов для генов, активных в структурах головного мозга и других органов. Найдены гены, экспрессия которых повышена в структурах мозга и связана с нейродегенеративными заболеваниями.

Ключевые слова: биоинформатика, мозг, экспрессия генов, микрочипы, секвенирование.

ВВЕДЕНИЕ

Компьютерный анализ молекулярных механизмов деятельности высшей нервной системы имеет огромное фундаментальное значение для биологии, генетики и для исследования процессов познания. При этом компьютерные методы используются для изучения структуры генов, их взаимосвязи, координации их работы (экспрессии) в клетках мозга (Орлов и др., 2013). Использование технологий высокопроизводительного секвенирования и анализа данных экспрессии генов на микрочипах позволяет ставить задачи исследования на качественно более высоком уровне (Xie *et al.*, 2011; Lipovich *et al.*, 2012). Целью данной работы были обзор основных биоинформационных подходов и баз данных по анализу экспрессии генов в клетках мозга и статистический анализ распределения

параметров структуры гена по отношению к его экспрессии в структурах мозга при использовании разработанных ранее биоинформационных методов.

Одним из крупнейших достижений в области нейронаук является создание трехмерных атласов экспрессии генов в мозге мыши и мозге человека (рис. 1), среди которых наиболее детальным ресурсом является компьютерная база данных Allen Brain Atlas (Lein *et al.*, 2007; Hawrylycz *et al.*, 2012).

В настоящее время разработана серия баз данных по экспрессии генов в мозге: GENSAT (Gene Expression Nervous System Atlas) (<http://www.gensat.org>), MGI (Mouse Genome Informatics) (<http://www.informatics.jax.org/>), BGEM (Brain Gene Expression Map) (<http://www.stjude-bgem.org>). Разрабатываются и базы данных, описывающие активность структур мозга, ос-

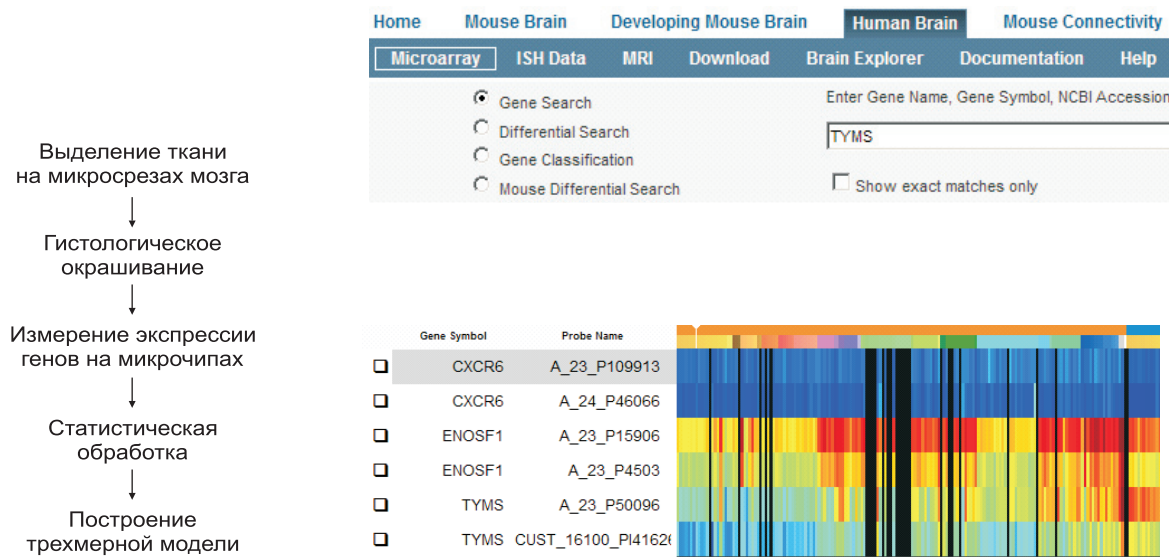


Рис. 1. Схема обработки данных по построению трехмерных карт экспрессии (левая панель) и пример поиска гена по имени (ген TYMS) с визуализацией экспрессии генов на микрочипах в структурах мозга в Allen Brain Atlas (<http://connectivity.brain-map.org/projection>) (правая панель).

нованные на методах магнитного резонанса и томографии. Методы электроэнцефалографии, топографического картирования электрической активности мозга и компьютерной томографии связаны с общим измерением активности структур мозга на более высоком уровне и выходят за рамки настоящей работы.

Опыт биоинформационных исследований в области анализа молекулярных механизмов регуляции экспрессии генов, использования технологий секвенирования и микрочипов, в том числе относящихся к экспрессии генов в тканях мозга, изучения роли серотониновой системы мозга в регуляции поведения, накопленный в ИЦиГ СО РАН (Ananko *et al.*, 2005; Витяев и др., 2001; Demenkov *et al.*, 2011; Nautenko *et al.*, 2011), используется при обработке новых полногеномных данных, полученных с помощью высокопроизводительных транскриптомных технологий.

Экспрессия генов в клетках зависит от внешних стимулов и внутренней генетической программы клеток (нейроны, клетки структур мозга, клетки крови, клетки внутренних органов). Регуляция проявления функции генов внутри клетки осуществляется на уровне транскрипции и трансляции.

Контроль экспрессии генов на уровне трансляции мРНК важен как для развития и

морфогенеза нейронов (Jung *et al.*, 2011), так и для функционирования специфических генных сетей в зрелых клетках различных разделов мозга. Нейроны характеризуются высокой степенью компартментализации (аксоны, дендриты, синапсы), при этом отдельные части клеток могут быть удалены на очень большое расстояние от ядра. Часть трансляционного аппарата локализована в удаленных районах клетки (вблизи синапсов) и мРНК генов, специфически задействованных в контроле передачи нервного импульса, транспортируется в эти районы (Liu-Yesucevitz *et al.*, 2011). Известны некоторые сигналы, локализованные в мРНК и опосредующие такой транспорт (Willis, Twiss, 2010; Wei, 2011).

Считается, что контроль экспрессии генов на уровне трансляционной активности специфических мРНК может иметь отношение к механизмам высшей нервной деятельности, такой, как физическая основа процесса запоминания (Darnell, 2011; Sidrauski *et al.*, 2013). Связь между трансляцией мРНК и процессами высшей нервной деятельности вызывает очень большой интерес в последние годы. Трансляция мРНК связана с механизмами, близкими к универсальным механизмам стрессового контроля экспрессии генов (фосфорилирование eIF2a, mTOR, eIF4E-BP и т. д.) (Gerashchenko

et al., 2012), адаптированными для решения специфических для клеток мозга задач (Sun *et al.*, 2013).

Трансляционный контроль используется как средство регуляции экспрессии генов у многих видов. Посттранскрипционные регуляторные механизмы играют важную роль в метаболических путях стрессового ответа и могут вести к нарушению физиологических функций при нарушении таких механизмов мутациями (Lohse *et al.*, 2011). Определенную роль в специфическом посттранскрипционном контроле экспрессии генов в клетках мозга также может играть цитоплазматическое полиаденилирование (Kundel *et al.*, 2009). Развитие методов высокопроизводительного секвенирования (Ribo-Seq) и протеомики (Menschaert *et al.*, 2013) существенно расширяет имеющиеся возможности для выявления молекулярных механизмов функционирования нейронов и мозга в целом.

Использование компьютерных технологий, таких, как GeneNet (Ananko *et al.*, 2005) и AndVisio (Demchenkova *et al.*, 2011), позволяет реконструировать на основе данных научных публикаций генные сети – ансамбли координированно функционирующих генов, контролирующих биохимические, молекулярно-генетические, физиологические процессы. С помощью GeneNet реконструированы генные сети, контролирующие различные системы и процессы, в том числе в мозговых тканях, включая генную сеть «Early long-term potentiation», отражающую белковые взаимодействия в дендритных шипиках зоны CA1 гиппокампа.

Конкретной целью нашей работы было выявление особенностей генов, активно экспрессирующихся в тканях мозга, с помощью комплексного компьютерного исследования, включая разработанные ранее авторским коллективом методы и программы, такие, как база данных качества проб микрочипов Affymetrix U133 (Orlov *et al.*, 2007), программный комплекс ICGenomics для функциональной аннотации генов (Орлов и др., 2012). Были подготовлены выборки генов, экспрессия которых повышена как в целом в структурах мозга, так и в отдельных районах головного мозга человека с использованием базы данных Allen Brain Atlas и BioGPS (Su *et al.*, 2009; Wu *et al.*, 2009). Из банка данных

UCSC (genome.ucsc.edu) были загружены данные нуклеотидных последовательностей, содержащих эти гены и их регуляторные районы. Исследовались встречаемость открытых рамок считывания (ОРС) в промоторных последовательностях, присутствие сигналов трансляции. Выполнен анализ контекстных особенностей нуклеотидных последовательностей таких генов, присутствие коротких некодирующих РНК, в том числе в противоположной ориентации. Для выборок генов, дифференциально экспрессирующихся в различных органах, оценивались число экзонов и его соотношение с уровнем экспрессии. Исследованы данные по экспрессии генов, консервативных для мыши, крысы и человека, на микрочипах; рассмотрены данные секвенирования полных транскриптом.

ИССЛЕДОВАНИЕ ЭКСПРЕССИИ ГЕНОВ В КЛЕТКАХ ГОЛОВНОГО МОЗГА

Экспрессия генов, т. е. проявление их функции в клетках мозга, является базисом работы нейрона. Обычно безотносительно к типу клеток активно транскрибируются в клетке не все гены одновременно, а какая-то небольшая их часть, около 5 %, что позволяет выделить органоспецифичные группы генов. Для поиска генов, экспрессия которых специфична для структур головного мозга, мы использовали базу данных Allen Brain Atlas и базу BioGPS (Wu *et al.*, 2009), содержащую данные генной экспрессии в широком круге тканей и органов. Экспрессия генов определялась на микрочипах Affymetrix U133 при использовании фильтрации по качеству (Orlov *et al.*, 2007).

Среди проб микрочипа Affymetrix U133, представленных в базе данных BioGPS, были выделены пробы с высокой экспрессией (по ранговым значениям проб всех генов, верхние 1 %) и гены, экспрессия которых представлена в структурах мозга (гипоталамус, префронтальный кортекс и др., всего 12 видов структур), но не в других органах (почки, печень, гладкие мышцы и т. д.). Всего после удаления дублирующих проб микрочипа было отобрано 11830 имен (уникальных идентификаторов) генов. Пример распределения экспрессии генов по органам дан на рис. 2.

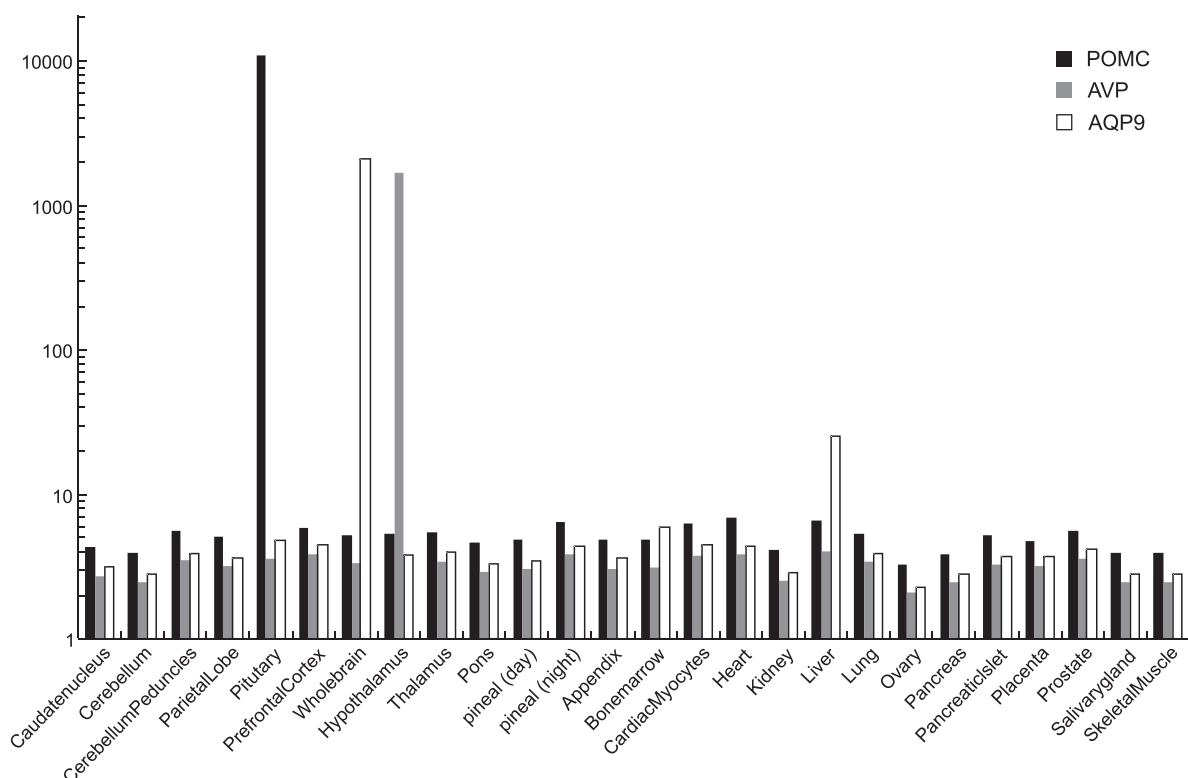


Рис. 2. Распределение экспрессии трех генов, наиболее высокоэкспрессирующихся в структурах головного мозга и других тканях.

По оси Y – уровень экспрессии на микрочипе Affymetrix U133 (база данных BioGPS). По оси X – исследованные ткани. Первые 12 групп слева соответствуют структурам головного мозга.

Из них 1801 ген (15 %) показывали высокую экспрессию хотя бы в одной из структур мозга. Большинство оставшихся генов (9253) не показывали значимую экспрессию (верхние 1 %) ни в одной из структур. Отметим, что использовались только белок-кодирующие гены, представленные на микрочипе.

Далее из выборки генов, входящих в верхние 1 % (1801), были выделены гены, показывающие высокую экспрессию в исследованных структурах мозга в среднем по сравнению со всеми остальными органами. Для полученных списков генов были проанализированы их геномное окружение, контекстная структура регуляторных районов, перекрывание с микроРНК и короткими некодирующими транскриптами в противоположной ориентации, число экзонов, эволюционная консервативность. Функциональная аннотация 1382 генов с высокой экспрессией только в структурах мозга выполнялась с помощью программы анализа генных онтологий DAVID (<http://david.abcc.ncifcrf.gov>). Результаты представлены в таблице.

Интересно отметить наличие категорий белкового транспорта, фосфопротеинов, нуклеотидного связывания, но не транскрипции. Присутствуют категории передачи нервного импульса, развития нейронов, что ожидаемо для структур головного мозга. Почти половина (45,7 %) генов из списка связана с альтернативным сплайсингом.

АНАЛИЗ ЧИСЛА ЭКЗОНОВ И ПЛАСТИЧНОСТИ ЭКСПРЕССИИ ГЕНОВ

На основе ранее полученных выборок генов мы проанализировали число экзонов, принадлежащих на группы генов с высокой экспрессией в структурах головного мозга, с повышенной экспрессией в других органах и группу оставшихся генов на микрочипе Affymetrix U133, экспрессия которых не была статистически значима ни в одном из исследованных органов, по данным BioGPS, представленным в предыдущем разделе. Гистограммы распределения по числу экзонов

Таблица

Функциональная аннотация генов с высокой экспрессией в клетках мозга

Описание белков и их функций	Процент генов	Значимость, (P-value)	Коррекция Бонферрони
Фосфопротеины	56,4	5,5E-50	3,5E-47
Альтернативный сплайсинг	45,7	2,7E-08	1,7E-05
Ацетилирование	24,6	2,2E-30	1,4E-27
Связывание нуклеотидов	14,5	1,4E-13	8,7E-11
Метаболизм фосфатов	9,9	1,1E-08	3,7E-05
Межклеточные сигналы	8,0	1,2E-14	4,1E-11
Передача нервного импульса	7,3	5,4E-28	1,9E-24
Киназы	7,3	6,8E-12	4,3E-09
Структурные молекулы	7,2	7,0E-09	8,2E-06
Рибонуклеиновый комплекс	6,5	8,5E-12	4,9E-09
Синаптическая передача	6,4	6,1E-26	2,2E-22
Развитие нейронов	6,3	1,6E-20	5,9E-17

и транскриптов представлены на рис. 3. Количество экзонов и транскриптов было подсчитано согласно базе данных Ensembl (<http://www.ensembl.org>). Таким образом, конечная выборка для обработки составила 8219 генов.

Сравнение последовательностей генов, высокоэкспрессирующихся в структурах мозга, по количеству экзонов в последовательности показало, что такие последовательности содержат меньшее количество экзонов (среднее 12, $t = 4,5$; $d.f. = 8210$; $p = 10^{-6}$), чем другие (среднее 13,5). Также показано, что для таких генов экспрессируется меньшее количество различных транскриптов ($t = 9,1$; $d.f. = 8176$; $p << 0,001$).

Среднее количество транскриптов, приходящееся на высокоэкспрессирующиеся гены в структурах мозга – 4,6, в то время как для низкоэкспрессирующихся генов – 5,7. Это согласуется с данными о том, что высокоэкспрессирующиеся гены, особенно экспрессирующиеся в различных тканях, обладают большей компактностью (Woody, Shoemaker, 2011; Park *et al.*, 2012).

Гены, имеющие высокий уровень экспрессии в широком круге органов, имеют высокий уровень экспрессии и в структурах головного мозга. На основе данных об экспрессии генов в различных органах был проведен следующий сравнительный анализ. Для каждого гена опре-

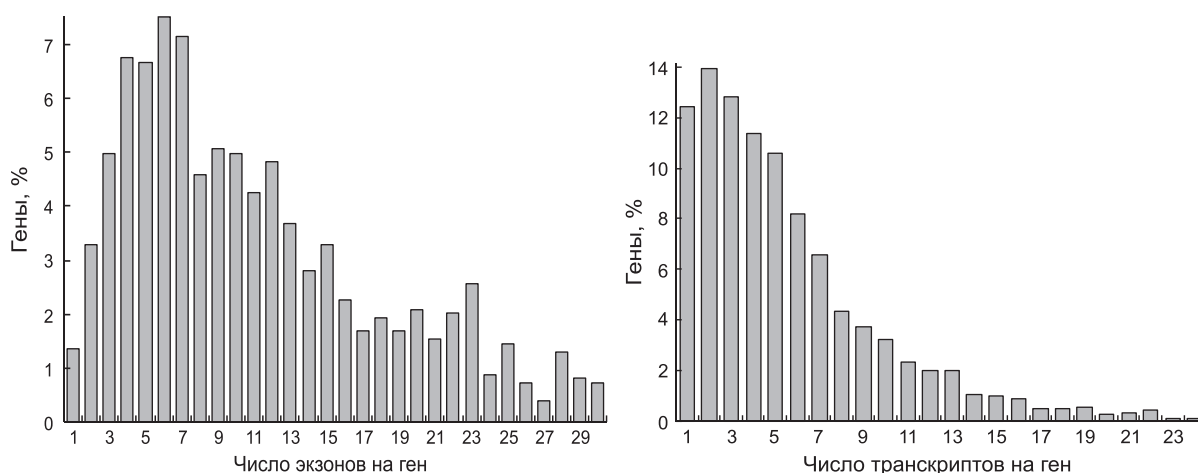


Рис. 3. Распределение последовательностей генов, высокоэкспрессирующихся в структурах мозга, по количеству экзонов в последовательности (слева) и количеству транскриптов (справа).

делялись наиболее высокие значения, лежащие вне доверительного интервала 99 %. Если они обнаруживались хотя бы для одной ткани головного мозга, то такие гены группировались. Всего в эту группу вошло 55 генов, и они имеют повышенную экспрессию во всех изучаемых тканях.

Оказалось, что по сравнению с другими генами количество транскриптов, соответствующих генам с повышенной экспрессией в структурах головного мозга, здесь больше (среднее 5,9, $t = 7,7$; $d.f. = 8176$; $p \ll 0,001$), чем в других органах (4,7). Также эта группа генов показала большее значение количества экзонов в последовательности (среднее 13,7 в отличие от среднего 11,9, $t = 5,5$; $d.f. = 8210$; $p \ll 0,001$). Ранее было показано, что более низкий уровень экспрессии легче поддается регуляции, в том числе и за счет увеличения количества экзонов и длины последовательности. Таким образом, повышенные значения для количества экзонов и количества транскриптов, вероятно, связаны с высокой специфичностью этих высокоэкспрессирующихся генов (Woody, Shoemaker, 2011).

КОМПЬЮТЕРНЫЙ АНАЛИЗ ВАРИАбельНОСТИ УРОВНЯ ЭКСПРЕССИИ ГЕНОВ В МОЗГЕ МЫШИ

В дополнение к анализу распределения экспрессии генов по тканям и структуры генов обрабатывалась информация о распределении экспрессии в отдельных структурах и участках мозга. Информация о генной экспрессии была получена из Allen Brain Atlas (ABA), содержащего данные коллометрической *in situ* гибридизации (ISH) об экспрессии генов мыши в $\sim 5 \times 10^4$ вокселях (кубических ячейках объемом $200 \mu\text{m}^3$) мозга. Мы провели компьютерную оценку зависимости варибельности уровня экспрессии 12932 генов от степени распространенности их экспрессии в различных районах мозга мыши.

Анализ взаимозависимостей между средним уровнем экспрессии генов E_{AVT} в вокселях мозга мыши с числом вокселей N_{Vox} на трехмерной карте мозга (см. рис. 1), в которых наблюдалась ненулевая экспрессия, и стандартным отклонением уровня экспрессии генов σ на микрочипе

показал, что наблюдается достоверная отрицательная корреляция между коэффициентом вариации экспрессии $CV = \sigma/E_{\text{AVT}}$ и N_{Vox} ($R = -0,69$, $p < 0,001$). То есть чем шире ген экспрессируется в различных участках мозга, тем ниже относительная варибельность его экспрессии (тренд представлен на рис. 4).

ГЕНОМНЫЙ КОНТЕКСТ: ЦИС-АНТИСЕНС ТРАНСКРИПТЫ И НЕКОДИРУЮЩИЕ РНК

Причиной большинства нейродегенеративных заболеваний является прогрессирующая гибель нейронов в определенных отделах головного мозга (Manfredsson *et al.*, 2012; Lazarev *et al.*, 2013). Традиционные инструменты для определения нейродегенеративных заболеваний, таких, как болезнь Альцгеймера, включают нейropsychологическое тестирование пациентов и специализированные технологии сканирования головного мозга. Поскольку нейродегенеративные изменения начинаются до проявления заметных клинических изменений, большое значение имеет поиск экспрессионных биомаркеров для ранней диагностики заболевания. Хорошими кандидатами для таких диагностик являются микроРНК – малые некодирующие РНК, вовлеченные в посттранскрипционную регуляцию генов (Cheng *et al.*, 2013). Они способны циркулировать в крови, тканеспецифические профили их экспрессии могут быть определены в жидкостях тела – крови, слюне,

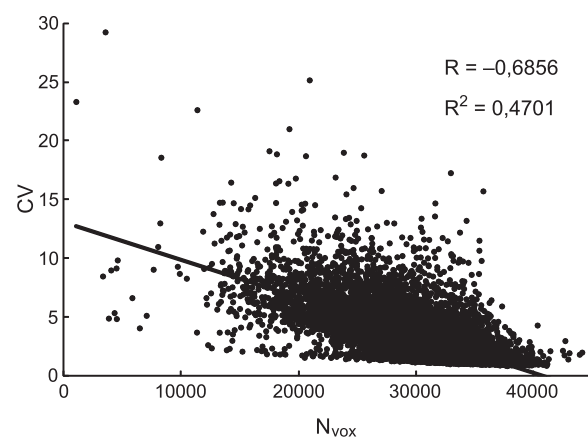


Рис. 4. Зависимость коэффициента вариации экспрессии CV (ось Y) от количества вокселей N_{Vox} (ось X).

моче. Для болезни Альцгеймера, в частности, кандидатами являются miR-9, miR-20a, и miR-132 (Cheng *et al.*, 2013).

По данным полного секвенирования транскриптома, до 50 % транскриптов не кодируют белки (Hung, Chang, 2010). Длинные некодирующие РНК (более 300 нуклеотидов) играют большую роль в развитии глиом мозга, их экспрессия также имеет предсказательное значение (Zhang *et al.*, 2012). Известен ряд некодирующих РНК генов (в частности, ген HULC, ген внутриядерной РНК MALAT-1), экспрессия которых значительно повышена при раке и может служить строгим маркером для диагностики (Lai *et al.*, 2012; Lipovich *et al.*, 2012).

По отношению к белок-кодирующим генам их можно классифицировать как межгенные,

внутригенные (в интронах) и антисенс-транскрипты (в противоположной ориентации к экзонам кодирующих генов). Цис-антисенс транскрипты – это кодируемые последовательности, которые транскрибируются в противоположных направлениях и частично перекрываются в геномных координатах. Интересно отметить, что для большого числа цис-антисенс транскриптов в геноме человека характерна активность в клетках мозга.

Для анализа мы использовали базу данных антисенс-транскриптов в геноме человека NATsDB (natsdb.cbi.pku.edu.cn/) и USAGDP (Grinchuk *et al.*, 2010). Была исследована экспрессия генов на микрочипах Affymetrix U133 на выборке здоровых тканей мозга (21 пациент, данные GEO NCBI) (рис. 5). Как видно из рис. 5,

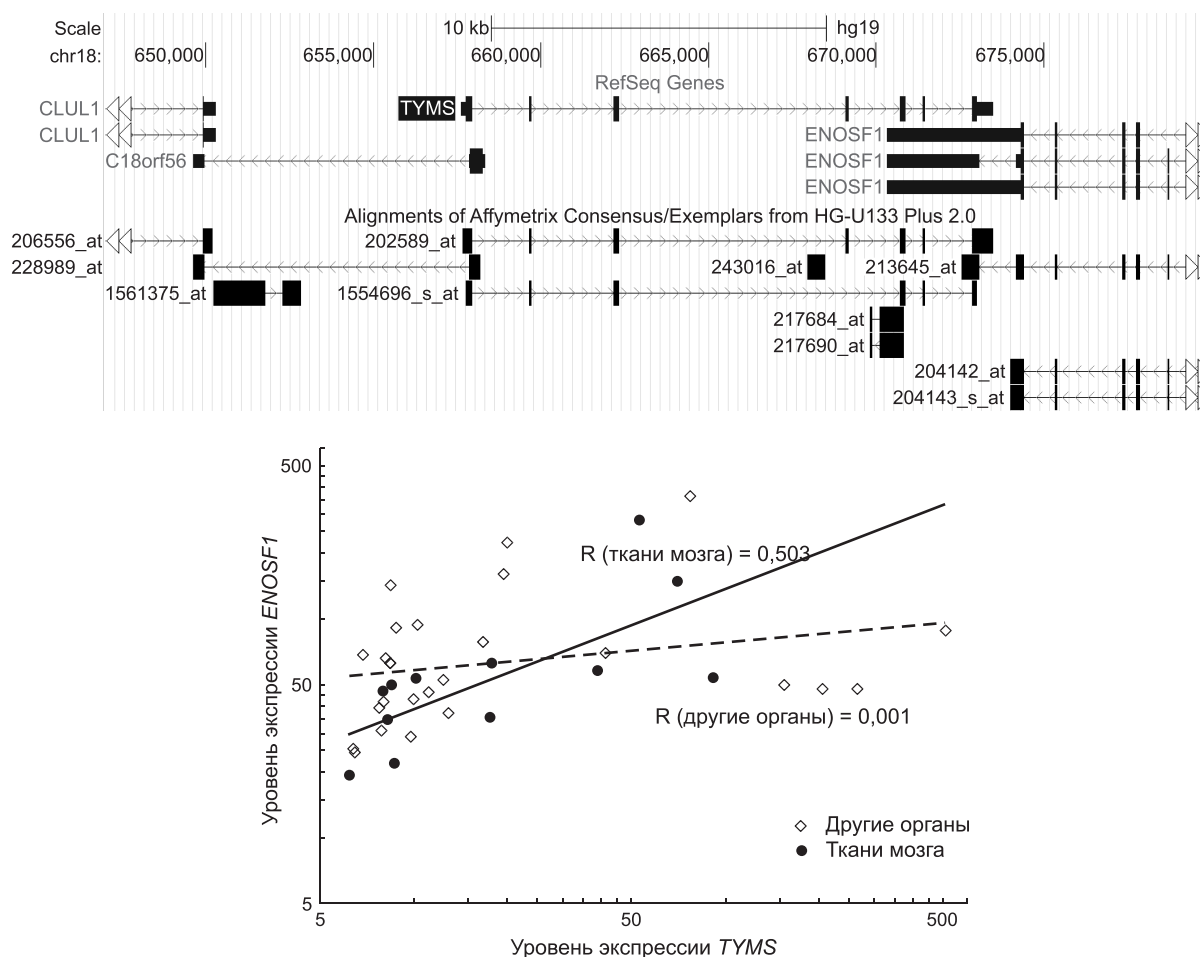


Рис. 5. Верхняя панель – гены *CLUL1*, *C18orf56*, *TYMS*, *ENOSF1* в противоположной ориентации в геноме человека (хромосома 18) и соответствующие пробы микрочипа Affymetrix (UCSC Genome Browser). Нижняя панель – корреляция экспрессии генов *ENOSF1* и *TYMS* в тканях мозга ($R = 0,5$) и в других органах ($R = 0,001$), по микрочиповым данным BioGPS.

существует несколько вариантов антисенс-расположения транскриптов (представленных наборами проб). 3 набора проб (*Affymetrix*) попадают в интроны гена *TUMS*, 2 набора проб находятся в противоположной ориентации (представлены EST).

Интересно отметить положительную корреляцию уровня экспрессии генов, расположенных в противоположной ориентации на клинической выборке микрочиповых данных (GEO NCBI GSE4290). Таким образом, это свидетельствует о повышенной транскрипционной активности всего геномного района в клетках мозга.

Иллюстрация корреляции экспрессии генов *TUMS* и *ENOSF1*, которые пересекаются по расположению транскриптов, на выборке тканей по базе BioGPS дана на рис. 5 (нижняя панель). Каждая точка соответствует ткани (всего 38). Корреляция экспрессии по 12 тканям мозга составляет около 0,5, в то же время по набору других тканей корреляция экспрессии практически не значима.

МОДЕЛИ ЭКСПРЕССИИ НА ЛАБОРАТОРНЫХ ЖИВОТНЫХ

Наибольший интерес представляет исследование экспрессии генов в клетках структур головного мозга человека, однако гораздо больший объем материала идет от лабораторных животных, в том числе в базе данных Allen Brain Atlas. Разработаны специальные модели животных, например, линии крыс, отличающихся по поведению, в частности крысы линии OXYS, выведенные в ИЦиГ СО РАН, (*Kolosova et al.*, 2006). Достаточное физиологическое сходство и эволюционная консервативность предполагают возможным исследовать экспрессию генов животных. Рассматривалась задача анализа экспрессии генов в тканях мозга лабораторных животных – крыс, селективированных по генетическим особенностям предрасположенности к стрессу и отличающихся, в том числе, по когнитивным функциям (способности к обучению). Для крыс линии OXYS был выделен набор генов, расположенных на хромосоме 1 и связанных с фенотипом животных (*Кожевникова и др.*, 2012; *Kozhevnikova et al.*, 2013). Среди биологических процессов, обогащенных в ка-

тегориях генных онтологий для генов данного локуса, интересно отметить передачу сигнала, неврологические процессы и визуальное восприятие. Присутствие генов из метаболических путей, связанных с болезнью Альцгеймера, может свидетельствовать об экспрессии генов данного локуса в нейронах и тканях мозга (*Kozhevnikova et al.*, 2013).

ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

Развитие новых экспериментальных методов секвенирования привело к стремительному росту объемов данных в геномике в целом и в области экспрессии генов в структурах мозга в частности. Обработка и анализ таких данных требуют разработки специализированных компьютерных средств. Нами использовались собственные компьютерные программы, базы данных и программный комплекс ICGenomics, предназначенный для анализа данных секвенирования и функциональной аннотации генов, компьютерной поддержки исследований в геномике и биомедицине (*Орлов и др.*, 2012).

Проведено компьютерное исследование генов человека, высокоэкспрессирующихся в тканях мозга по сравнению с другими тканями и в отдельных районах головного мозга (по данным BioGPS и Allen Brain Atlas). По выборкам генов, дифференциально экспрессирующихся в различных органах, рассчитаны число экзонов и его соотношение с уровнем экспрессии. Показано статистическое различие числа альтернативных транскриптов для генов, активных в тканях мозга и других органов. Компьютерный анализ вариабельности экспрессии генов показал, что чем шире ген экспрессируется в мозге мыши, тем меньшим коэффициентом вариации экспрессии он характеризуется. Рассмотрены особенности нетранслируемых районов генов, связанные с тканеспецифичной регуляцией экспрессии. Дан обзор особенностей геномной структуры генов, имеющих дифференциальную экспрессию в клетках мозга человека и лабораторных животных, показана связь с дополнительными механизмами регуляции экспрессии генов на уровне транскрипции и трансляции, включающими присутствие микрорНК, коротких некодирующих транскриптов. Системное исследование экспрессии генов

в клетках мозга с помощью взаимодополняющих экспериментальных подходов является необходимой основой междисциплинарных нейробиологических исследований и должно быть продолжено с использованием новых транскриптомных данных.

БЛАГОДАРНОСТИ

Работа поддержана Министерством образования и науки России (соглашение 8740), ИП СО РАН № 136.

ЛИТЕРАТУРА

- Витяев Е.Е., Орлов Ю.Л., Вишневский О.В. и др. Компьютерная система «GENE DISCOVERY» для поиска закономерностей организации регуляторных последовательностей эукариот // Молекуляр. биология. 2001. Т. 35. № 6. С. 952–960.
- Кожевникова О.С., Мартыщенко М.К., Генаев М.К. и др. RatDNA: база данных микрочиповых исследований на крысах для генов, ассоциированных с заболеваниями старения // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. P. 756–765.
- Орлов Ю.Л., Брагин А.О., Медведева И.В. и др. ICGenomics: программный комплекс анализа символьных последовательностей геномики // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. P. 732–741.
- Орлов Ю.Л., Вишневский О.В., Витяев Е.Е. и др. Биоинформационный анализ экспрессии генов в клетках мозга // Тр. XV Всерос. науч.-техн. конф. «Нейроинформатика-2013». 21–25 января 2013 г. М.: Национальный исследовательский ядерный ун-т «МИФИ», 2013. С. 74–85.
- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. 33(Database issue). D425–427.
- Cheng L., Quek C., Sun X. *et al.* Deep-sequencing of microRNA associated with Alzheimer's disease in biological fluids: From biomarker discovery to diagnostic practice // Frontiers in Genetics. 2013. V. 4. 00150.
- Darnell J.C. Defects in translational regulation contributing to human cognitive and behavioral disease // Curr. Opin. Genet. Dev. 2011. V. 21. No. 4. P. 465–473.
- Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem // In Silico Biol. 2011. V. 11. No. 3. P. 149–161.
- Gerashchenko M.V., Lobanov A.V., Gladyshev V.N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress // Proc. Natl Acad. Sci. USA. 2012. V. 109. No. 43. P. 17394–17399.
- Grinchuk O.V., Jenjaroenpun P., Orlov Y.L. *et al.* Integrative analysis of the human cis-antisense gene pairs, miRNAs and their transcription regulation patterns // Nucl. Acids Res. 2010. V. 38. No. 2. P. 534–547.
- Hawrylycz M.J., Lein E.S., Guillozet-Bongaarts A.L. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome // Nature. 2012. V. 489. No. 7416. P. 391–399.
- Hung T., Chang H.Y. Long noncoding RNA in genome regulation: prospects and mechanisms // RNA Biol. 2010. V. 7. No. 5. P. 582–585.
- Jung H., O'Hare C.M., Holt C.E. Translational regulation in growth cones // Curr. Opin. Genet. Dev. 2011. V. 21. No. 4. P. 458–464.
- Kolosova N.G., Trofimova N.A., Fursova A. Opposite effects of antioxidants on anxiety in Wistar and OXYS rats // Bull. Exp. Biol. Med. 2006. V. 141. P. 734–737.
- Kozhevnikova O.S., Korbolina E.E., Stefanova N.A. *et al.* Association of AMD-like retinopathy development with an Alzheimer's disease metabolic pathway in OXYS rats // Biogerontology. 2013. DOI 10.1007/s10522-013-9439-2 [Epub ahead of print].
- Kundel M., Jones K.J., Shin C.Y., Wells D.G. Cytoplasmic polyadenylation element-binding protein regulates neurotrophin-3-dependent beta-catenin mRNA translation in developing hippocampal neurons // J. Neurosci. 2009. V. 29. No. 43. P. 13630–13639.
- Lai M.C., Yang Z., Zhou L. *et al.* Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation // Med. Oncol. 2012. V. 29. No. 3. P. 1810–1816.
- Lazarev V.F., Sverchinskiy D.V., Ippolitova M.V. *et al.* Factors affecting aggregate formation in cell models of Huntington's disease and amyotrophic lateral sclerosis // Acta Naturae. 2013. V. 5. No. 2. P. 81–89.
- Lein E.S., Hawrylycz M.J., Ao N. *et al.* Genome-wide atlas of gene expression in the adult mouse brain // Nature. 2007. V. 445. No. 7124. P. 168–176.
- Lipovich L., Dacht F., Cai J. *et al.* Activity-dependent human brain coding/noncoding gene regulatory networks // Genetics. 2012. V. 192. No. 3. P. 1133–1148.
- Liu-Yesucevitz L., Bassell G.J., Gitler A.D. *et al.* Local RNA translation at the synapse and in disease // J. Neurosci. 2011. V. 31. No. 45. P. 16086–16093.
- Lohse I., Reilly P., Zaugg K. The CPT1C 5'UTR contains a repressing upstream open reading frame that is regulated by cellular energy availability and AMPK // PLoS One. 2011. V. 6. No. 9. e21486.
- Manfredsson F.P., Bloom D.C., Mandel R.J. Regulated protein expression for in vivo gene therapy for neurological disorders: progress, strategies, and issues // Neurobiol. Dis. 2012. V. 48. No. 2. P. 212–221.
- Menschaert G., Van Crieking W., Notelaers T. *et al.* Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events // Mol. Cell Proteomics. 2013. V. 12. No. 7. P. 1780–1790.
- Naumenko V.S., Kondurova E.M., Popova N.K. On the role of brain 5-HT7 receptor in the mechanism of hypothermia: comparison with hypothermia mediated via 5-HT1A and 5-HT3 receptor // Neuropharmacology. 2011. V. 61. No. 8. P. 1360–1365.

- Orlov Y.L., Zhou J., Lipovich L. *et al.* Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis // *In Silico Biol.* 2007. V. 7. No. 3. P. 241–260.
- Park J., Xu K., Park T., Yi S.V. What are the determinants of gene expression levels and breadths in the human genome? // *Hum. Mol. Genet.* 2012. V. 21. No. 1. P. 46–56.
- Savinkova L., Drachkova I., Arshinova T. *et al.* An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein // *PLoS One.* 2013. V. 8. No. 2. e54626.
- Sidrauski C., Acosta-Alvear D., Khoutorsky A. *et al.* Pharmacological brake-release of mRNA translation enhances cognitive memory // *eLife.* 2013. V. 28. e00498.
- Su A.I., Wiltshire T., Batalov S. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes // *Proc. Natl Acad. Sci. USA.* 2009. V. 101. No. 16. P. 6062–6067.
- Sun X., Liu J., Crary J.F. *et al.* ATF4 protects against neuronal death in cellular Parkinson's disease models by maintaining levels of parkin // *J. Neurosci.* 2013. V. 33. No. 6. P. 2398–2407.
- Wei L.N. The RNA superhighway: axonal RNA trafficking of kappa opioid receptor mRNA for neurite growth // *Integr. Biol. (Camb).* 2011. V. 3. No. 1. P. 10–16.
- Willis D.E., Twiss J.L. Regulation of protein levels in subcellular domains through mRNA transport and localized translation // *Mol. Cell Proteomics.* 2010. V. 9. No. 5. P. 952–962.
- Woody J.L., Shoemaker R.C. Gene expression: sizing it all up // *Front Genet.* 2011. V. 2. P. 70.
- Wu C., Orozco C., Boyer J. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources // *Genome Biol.* 2009. V. 10. No. 11. R130.
- Xie J., Zhao T., Lee T. *et al.* Anisotropic path searching for automatic neuron reconstruction // *Med. Image Anal.* 2011. V. 15. No. 5. P. 680–689.
- Zhang X., Sun S., Pu J.K. *et al.* Long non-coding RNA expression profiles predict clinical phenotypes in glioma // *Neurobiol. Dis.* 2012. V. 48. No. 1. P. 1–8.

COMPUTER ANALYSIS OF DATA ON GENE EXPRESSION IN BRAIN CELLS OBTAINED BY MICROARRAY TESTS AND HIGH-THROUGHPUT SEQUENCING

**I.V. Medvedeva, O.V. Vishnevsky, N.S. Safronova, O.S. Kozhevnikova,
M.A. Genaev, A.V. Kochetov, D.A. Afonnikov, Y.L. Orlov**

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: orlov@bionet.nsc.ru

Summary

The scope of neurobiological studies has been greatly expanded in the last years. It is accompanied by accumulation of a huge body of experimental data on the structure, function and evolution of the nervous system at different hierarchical levels of its organization. High-throughput sequencing technologies and microarray tests permit the expression of thousands of genes to be analyzed with regard to cell location in the brain. Methods of gene expression analysis are briefly reviewed in the context of brain research. We have analyzed specific features of genes differentially expressed in brain cells. Some genes overexpressed in brain tissues are associated with neurological diseases. The numbers of exons and active transcripts in genes differentially expressed in different organs are considered. Statistically significant difference in such parameters is shown for genes intensely expressed in the brain and other organs. Examples of such differentially expressed genes associated with neurological deceases are presented.

Key words: bioinformatics, brain, gene expression, microarray, sequencing.