

УДК 577.113.3:57.087.1

## ВЫЯВЛЕНИЕ СВЯЗИ ВАРИАБЕЛЬНОСТИ ЭКСПРЕССИИ ГЕНОВ ПУТЕЙ ПЕРЕДАЧИ СИГНАЛОВ В МОЗГЕ ЧЕЛОВЕКА СО СРОДСТВОМ ТАТА-СВЯЗЫВАЮЩЕГО БЕЛКА К ПРОМОТОРАМ ЭТИХ ГЕНОВ

© 2014 г. М.П. Пономаренко<sup>1,2</sup>, В.В. Суслов<sup>1</sup>, К.В. Гунбин<sup>1,2</sup>,  
П.М. Пономаренко<sup>3</sup>, О.В. Вишневский<sup>1,2</sup>, Н.А. Колчанов<sup>1,2</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики  
Сибирского отделения Российской академии наук, Новосибирск, Россия,  
e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Федеральное государственное автономное образовательное учреждение высшего образования  
«Новосибирский национальный исследовательский государственный университет»,  
Новосибирск, Россия;

<sup>3</sup> Children's Hospital Los Angeles, Los Angeles, CA, USA

Поступила в редакцию 27 сентября 2014 г. Принята к публикации 5 ноября 2014 г.

Вариабельность экспрессии генов всегда была в самом центре внимания весьма широкого круга исследований, которые были направлены на реконструкцию сайтов на ДНК для связывания регуляторных белков, регуляторных районов генов, генных сетей, путей передачи сигналов в них, а также всего разнообразия других геномных объектов, которые могли бы изменять экспрессию генов. Коэффициент вариации Пирсона ( $C_v$ , отношение стандартного отклонения к среднему) был чаще всего использован в качестве меры вариабельности экспрессии генов в таких областях интенсивных исследований, как биомедицина и селекция. В свою очередь, только единственный универсальный общий регуляторный сигнал был идентифицирован в геномах у всех эукариот, а именно: ТАТА-бокс вместе с сопровождающими его тремя оставшимися менее консервативными облигатными элементами, DPE, BRE и INR, в составе коровых промоторов. С помощью нашего экспериментально подтвержденного уравнения для оценки величин сродства ТАТА-бокс связывающего белка (TBP) к ТАТА-боксу мы проанализировали данные высокопроизводительного секвенирования для 35 609 мРНК в 946 сегментах мозга человека, которые были взяты нами из базы данных Allen Brain Atlas. В результате нами была обнаружена статистически достоверная линейная корреляция между экспериментально измеренными *in vivo* величинами коэффициента вариации экспрессии генов путей передачи сигналов в мозге человека и компьютерными *in silico* оценками для величин сродства ТАТА-бокс связывающего белка к промоторам этих генов. Этот результат может помочь в анализе данных высокопроизводительного секвенирования, которые были получены с целью идентификации того, что еще могло бы изменять экспрессию генов.

**Ключевые слова:** ген, промотор, ТАТА-связывающий белок (TBP), сродство TBP/ДНК, экспрессия, коэффициент вариации ( $C_v$ ), корреляция, мозг, человек.

### ВВЕДЕНИЕ

Одной из важнейших характеристик жизнедеятельности организма является вариабельность экспрессии его генов (дифференциальная, циклическая, стресс- и стимул-зависимая, эпигенетическая и т. п.) на различных уровнях

организации живых систем (организменном – внутриклеточном, тканевом, орган- и стадий-специфическом и др.; популяционном – возрастном, пол-специфическом, видовом и т. д.) (Rausenberger *et al.*, 2009). Анализ вариабельности экспрессии генов позволяет реконструировать генные сети (Chalancon *et al.*, 2012), пути

передачи сигналов регуляции генов (Gilbert *et al.*, 2006) и интерференцию между ними (Balazsi *et al.*, 2011). Также на основе анализа вариабельности экспрессии генов выявляют регуляторные районы этих генов (Munsky *et al.*, 2012), кластеры множественных элементов ответа на регуляторные сигналы (Mironova *et al.*, 2013), композиционные элементы (Stepanova, Alonso, 2009) и эпигенетические факторы (Whitelaw *et al.*, 2010), которые способны изменять экспрессию генов.

В опытах вариабельность экспрессии генов характеризуют соответствующими статистиками – главные компоненты, фактор Фано, коэффициенты Хилла, детерминации, кластеризации, причинности Грэнджера, авторегрессии, а также корреляциями Пирсона, Фурье, Спирмана и др. (см., например, Rosa *et al.*, 2012). Каждая статистика была лучше или хуже для некоторых типов вариабельности, проявлявшихся на определенном уровне организации живых систем.

Коэффициент вариации ( $C_v$ , отношение стандартного отклонения к среднему) применялся наиболее часто и показал себя удобным для широкого круга типов вариабельности и уровней организации жизни (Ghosh *et al.*, 2011). Его широко используют для оценки эффективности селекции на продуктивность, например, животных – на технологические свойства мяса (Михайлов и др., 2011) и на пряжильность шерсти (Кулаков и др., 1999), растений – на содержание витаминов (Расторгуев, 2008), качество зерна (Кислинский, 1997) и количество зеленой массы (Куркина, 2004). С помощью  $C_v$  также оценивают степень генетического дрейфа (Pardo *et al.*, 2005) и степень фенотипического проявления – экспрессивность и пенетрантность – патологий человека (Ou *et al.*, 2011), а также успешность их диагностики (Кравцов и др., 2011), мониторинга (Willems *et al.*, 2000) и лечения (Georger *et al.*, 2009).

С вариабельностью экспрессии признаков, с которыми работают исследователи, напрямую связаны и представления об адаптивной эволюции. Синтетическая теория эволюции (СТЭ) была построена преимущественно на качественных, т. е. с низкой степенью вариабельности, признаках. Для таких признаков отбор может непосредственно менять частоту

аллелей (Левонтин, 1978). Ч. Дарвин (1991), имевший дело, прежде всего, с морфологическими признаками с широкой вариабельностью, рассматривал эволюцию как адаптивный переход между изменчивостью неопределенной (признаки с широкой вариабельностью, с широкой нормой реакции (Йогансен, 1933)) и определенной – с низкой вариабельностью. При таком подходе  $C_v$  будет являться количественной мерой нормы реакции экспрессии гена, так как он характеризует степень различия признаков фенотипа у генетически однородной группы особей, прежде всего из-за различия факторов внешней среды (например, по доступности пищи, температуре). Ожидалось, что у группы генетически идентичных особей в идентичных условиях это различие будет нивелировано. Однако В.А. Струнников (1989) на популяции партеногенетических, т. е. генетически идентичных *Bombyx mori*, выращенных в среде высококонтролируемой идентичности, показал, что это не так. Природу открытой им реализационной изменчивости он связывал с шумовой вариабельностью – стохастической несогласованностью в многоступенчатых процессах регуляции онтогенеза.

Близкую точку зрения высказал В.А. Черданцев (2003), который связывал открытую им внутрииндивидуальную изменчивость сериальных органов с ограниченностью по ресурсу. Поскольку эмбриональная закладка имеет ограниченные размеры, лишь первые органы серии формируются в оптимальных условиях. Условия работы генетической программы при формировании каждого последующего органа серии будут деформированы предыдущим. Поэтому изменчивость в серии органов нарастает со временем их появления, хотя все они – продукты одной и той же генетической программы. В работе И.А. Аршавского (1982) была показана важность активации такой изменчивости для нормального онтогенеза. Вариабельность, связанная с интерференцией разных генетических программ и/или генных сетей на надгенном онтогенетическом уровне, была выявлена в экспериментах М.М. Камшилова (1939), а также при дестабилизирующем отборе (Belyaev, 1978). Однако возможность контекстного сигнала в геноме, который мог бы вносить достоверный вклад в вариабельность

экспрессии генов, до сих пор не рассматривалась и поэтому никто не искал в геноме такой сигнал. Лишь в узком диапазоне условий опыта (Blake *et al.*, 2006) наблюдали *in vivo* изменение  $C_v$  экспрессии искусственного гена при мутациях в ТАТА-боксе его химерного промотора, но экспрессия этого гена была стационарной при всех иных условиях.

В этой работе мы проверили эмпирическое наблюдение (Blake *et al.*, 2006) путем анализа экспериментальных данных об экспрессии генов в мозге человека и мыши. В результате мы обнаружили связь между  $C_v$  экспрессии генов путей передачи сигналов в мозге человека и средством ТВР к их промоторам. Эта связь означает, что ТАТА-бокс – один из контекстных сигналов, которые вносят вклад в вариабельность экспрессии генов. В этом состоит главный результат нашей работы.

## МАТЕРИАЛЫ И МЕТОДЫ

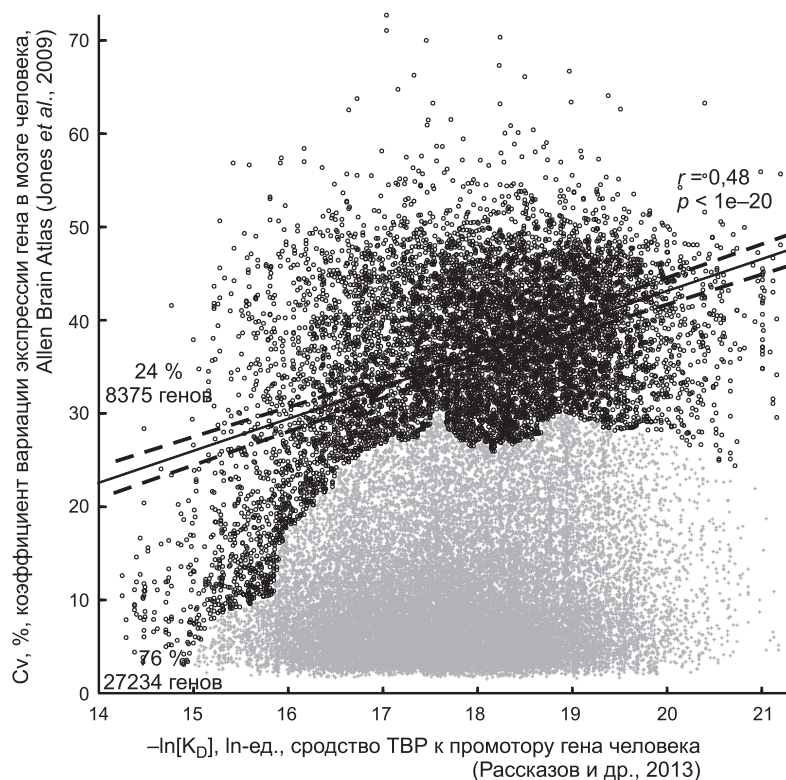
**Транскриптом мозга.** Были проанализированы последовательности ДНК районов  $[-70; -20]$  относительно стартов транскрипции (район экспериментально доказанных ТАТА-боксов (Ponomarenko *et al.*, 2013)) для 35 609 генов человека и для 12 931 гена мыши. Последовательности генов человека были взяты из референсного генома человека (hg19) из базы данных RefSeq, вып. 52, (Pruitt *et al.*, 2012). Последовательности генов мыши были взяты из референсного генома мыши, вып. 69, базы данных Ensembl (Flicek *et al.*, 2011) с помощью UCSC Genome Browser (Dreszer *et al.*, 2012). Каждая из этих последовательностей ДНК была охарактеризована величиной  $C_v$ . Эти величины отложены по вертикальным осям на рис. 1 и 2. Они были оценены по экспериментальным данным об экспрессии каждого гена. Величины экспрессии генов были в пределах от 0 до 255 условных единиц. Эти величины уровня экспрессии 35 609 генов в 946 сегментах мозга человека были взяты из базы данных Allen Brain Atlas (Jones *et al.*, 2009). Аналогично величины уровня экспрессии 12 931 гена мыши в 51 533 вокселях (элементарный образец мозга объемом  $200 \mu\text{m}^3$ , «voxel, volumetric picture element») были взяты из транскриптома мозга мыши (Lein *et al.*, 2007).

**Анализ транскриптома мозга.** Для каждой последовательности ДНК был найден максимум величины  $-\ln(K_D)$  средства ТВР/ДНК в скользящем окне длиной 26 п.о., как описано в статье Рассказова с соавт. (2013). Эти величины в натуральных логарифмических единицах ( $\ln$ ) отложены по горизонтальной оси на рис. 1 и 2.

Пары величин  $C_v$  (%) и  $-\ln(K_D)$  в  $\ln$  (ед.) для всех 35 609 гена человека и отдельно для всех 12 931 гена мыши были масштабированы в шкалу  $[0; 1]$  относительных единиц и сгруппированы в два кластера с помощью пакета Statistica (StatsoftTM, Tulsa, USA) при предустановленных параметрах без какой-либо иной обработки. Главному и минорному кластерам природных генов соответствуют серые «+» и черные «o» на рис. 1 для человека и на рис. 2 – для мыши. Эти рисунки представляют результаты Statistica для комбинации метода ближайших соседей (Single linkage) с расстоянием Евклида. Результаты Statistica для всех  $42 = 7 \times 6$  возможных комбинаций 7 методов кластеризации и 6 мер сходства этого пакета различались в пределах 2 % от объема данных, чаще всего вблизи границы между кластерами (данные не показаны). Эти кластеры генов были аннотированы с помощью пакета DAVID, вып. 6.7 (Huang *et al.*, 2009) в форме 10 терминов Gene Ontology (GO) с наивысшими оценками достоверности с поправкой Бонферрони (табл. 1, 2).

## РЕЗУЛЬТАТЫ

**Транскриптом мозга человека.** Результат бинарной кластеризации 35 609 генов человека по оценкам  $C_v$  экспрессии в мозге (Jones *et al.*, 2009) в сравнении с оценками  $-\ln(K_D)$  средства ТВР к их промоторам (Рассказов и др., 2013) показан на рис. 1. В табл. 1 можно видеть функциональную GO-аннотацию этого результата с использованием пакета DAVID, вып. 6.7 (Huang *et al.*, 2009). Главный кластер (серые «+») включал 27 234 гена (76 %), которые имели низкую  $10 \pm 6$  % среднеарифметическую оценку  $C_v$ . Как можно видеть в табл. 1, большинство генов человека из главного кластера с конститутивной (стационарной) экспрессией в мозге соответствуют ферментам внутриклеточного метаболизма. Это, чаще всего, – гены



**Рис. 1.** Достоверная ( $r=0,48$ ,  $p < 10^{-20}$ ) корреляция между коэффициентом вариации  $C_v$  экспрессии генов путей передачи сигналов в мозге человека и сродством ТАТА-связывающего белка к их промоторам,  $-\ln(K_D)$ .

Главный кластер (серые «+»): 27 234 гена (76 %) с низкой  $10 \pm 6$  % средней оценкой  $C_v$ , которые ассоциированы с «домашним хозяйством» внутри клеток мозга человека (табл. 1). Минорный кластер (черные «○»): 8 375 генов (24 %) с высокой  $36 \pm 9$  % средней оценкой  $C_v$ , которые ассоциированы с рецепторами путей передачи сигналов на внешних мембранах клеток мозга человека (табл. 1). Пунктир: 95 % – доверительный интервал регрессии, Statistica (Statsoft™, USA).

**Fig. 1.** The significant correlation ( $r=0,48$ ,  $p < 10^{-20}$ ) between  $C_v$ -values for mRNAs of signaling pathway genes in the human brain and  $-\ln[K_D]$  values of TATA-binding protein affinity for promoters of these genes.

The main cluster (+, gray), of 27 234 genes (76 %), clustered by Statistica (Statsoft™, USA) had the lowest mean  $C_v$  value,  $10 \pm 6$  %. These genes were associated with the «housekeeping» inside cells of the human brain (Table 1). The minor cluster of 8 375 genes (○, black), (24 %), had the highest mean  $C_v$  value,  $36 \pm 9$  %. These genes were associated by the same way with the signaling pathway receptors on outer membranes of human brain cells (Table 1). Dashed curves show 95 % confidence intervals for the linear regression (solid lines) built with Statistica.

«домашнего хозяйства», которые обеспечивают базовый уровень жизнеспособности любой живой клетки.

Минорный кластер (черные «○») включал 8 375 генов человека (24 %) с высокой  $36 \pm 9$  % среднеарифметической оценкой  $C_v$ . Эти гены кодировали рецепторы путей передачи сигналов на внешних мембранах клеток мозга человека. Была обнаружена достоверная связь между  $C_v$  экспрессии этих генов в мозге (Jones *et al.*, 2009) и оценкой  $-\ln(K_D)$  сродства ТВР к их промоторам,  $r = 0,48$  ( $p < 10^{-20}$ ).

Поскольку стало общепринятым требование (Encode project Consortium ..., 2012) сопостав-

лять результаты идентичного анализа нескольких независимых транскриптомов, то результат анализа данных RNA-Seq уровня экспрессии генов в мозге человека (рис. 1 и табл. 1) был сопоставлен с результатом идентичного анализа для независимых микрочип-данных уровня экспрессии генов в мозге мыши.

**Транскриптом мозга мыши.** Как можно видеть на рис. 2, в главном кластере (серые, «+») было 11 805 генов мыши (91 %) с низким  $189 \pm 98$  % средним  $C_v$ , тогда как в минорном кластере (черные «○») было 1126 генов (9 %) с высоким  $551 \pm 166$  % средним  $C_v$ . При этом минорный кластер был охарактеризован

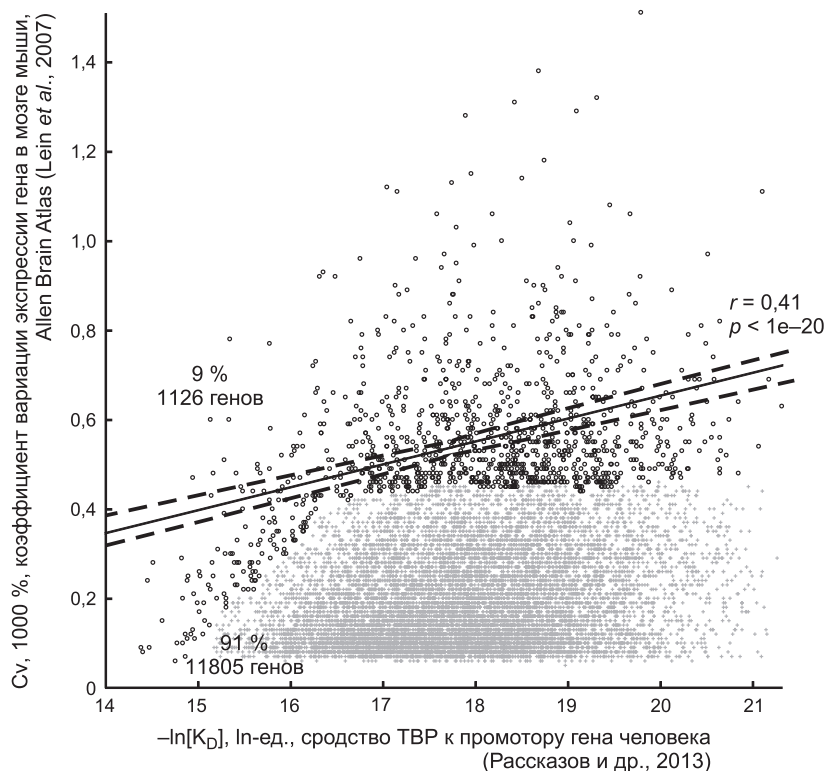


Таблица 1

Аннотация главного и минорного кластеров генов человека в терминах Gene Ontology,  
\*GO:№ (достоверность), с помощью DAVID, вып. 6.7 (Huang *et al.*, 2009)

| Главный кластер: 27 234 гена (76 %) с низкой вариабельностью экспрессии                      | Минорный кластер: 8 375 генов (24 %) с высокой вариабельностью экспрессии   |
|----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Биологический процесс</b>                                                                 |                                                                             |
| 0044237~cellular metabolic process ( $10^{-33}$ )                                            | 0032501~multicellular organismal process ( $10^{-78}$ )                     |
| 0009987~cellular process ( $10^{-29}$ )                                                      | 0007186~G-protein coupled receptor protein signaling pathway ( $10^{-61}$ ) |
| 0044260~cellular macromolecule metabolic process ( $10^{-24}$ )                              | 0007600~sensory perception ( $10^{-60}$ )                                   |
| 0008152~metabolic process ( $10^{-20}$ )                                                     | 0007606~sensory perception of chemical stimulus ( $10^{-59}$ )              |
| 0044267~cellular protein metabolic process ( $10^{-16}$ )                                    | 0007608~sensory perception of smell ( $10^{-58}$ )                          |
| 0044238~primary metabolic process ( $10^{-16}$ )                                             | 0050890~cognition ( $10^{-48}$ )                                            |
| 0043170~macromolecule metabolic process ( $10^{-12}$ )                                       | 0007166~cell surface receptor linked signal transduction ( $10^{-48}$ )     |
| 0046907~intracellular transport ( $10^{-12}$ )                                               | 0003008~system process ( $10^{-41}$ )                                       |
| 0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process ( $10^{-11}$ ) | 0050877~neurological system process ( $10^{-35}$ )                          |
| 0034641~cellular nitrogen compound metabolic process ( $10^{-11}$ )                          | 0050896~response to stimulus ( $10^{-32}$ )                                 |
| <b>Внутриклеточный компартмент</b>                                                           |                                                                             |
| 0005622~intracellular ( $10^{-90}$ )                                                         | 0005576~extracellular region ( $10^{-90}$ )                                 |
| 0044424~intracellular part ( $10^{-85}$ )                                                    | 0044421~extracellular region part ( $10^{-41}$ )                            |
| 0043231~intracellular membrane-bounded organelle ( $10^{-66}$ )                              | 0005886~plasma membrane ( $10^{-37}$ )                                      |
| 0043227~membrane-bounded organelle ( $10^{-66}$ )                                            | 0031224~intrinsic to membrane ( $10^{-28}$ )                                |
| 0043229~intracellular organelle ( $10^{-63}$ )                                               | 0005615~extracellular space ( $10^{-28}$ )                                  |
| 0043226~organelle ( $10^{-62}$ )                                                             | 0016021~integral to membrane ( $10^{-27}$ )                                 |
| 0005737~cytoplasm ( $10^{-54}$ )                                                             | 0031226~intrinsic to plasma membrane ( $10^{-17}$ )                         |
| 0005623~cell ( $10^{-44}$ )                                                                  | 0005887~integral to plasma membrane ( $10^{-16}$ )                          |
| 0044444~cytoplasmic part ( $10^{-44}$ )                                                      | 0005578~proteinaceous extracellular matrix ( $10^{-11}$ )                   |
| 0044464~cell part ( $10^{-44}$ )                                                             | 0031012~extracellular matrix ( $10^{-10}$ )                                 |
| <b>Молекулярная функция</b>                                                                  |                                                                             |
| 0005488~binding ( $10^{-21}$ )                                                               | 0004930~G-protein coupled receptor activity ( $10^{-74}$ )                  |
| 0005515~protein binding ( $10^{-15}$ )                                                       | 0004888~transmembrane receptor activity ( $10^{-74}$ )                      |
| 0003824~catalytic activity ( $10^{-10}$ )                                                    | 0004872~receptor activity ( $10^{-68}$ )                                    |
| 0003723~RNA binding ( $10^{-7}$ )                                                            | 0004984~olfactory receptor activity ( $10^{-62}$ )                          |
| 0000166~nucleotide binding ( $10^{-6}$ )                                                     | 0004871~signal transducer activity ( $10^{-48}$ )                           |
| 0046914~transition metal ion binding ( $10^{-4}$ )                                           | 0060089~molecular transducer activity ( $10^{-48}$ )                        |
| 0030695~GTPase regulator activity ( $10^{-3}$ )                                              | 0043565~sequence-specific DNA binding ( $10^{-16}$ )                        |
| 0017076~purine nucleotide binding ( $10^{-3}$ )                                              | 0005125~cytokine activity ( $10^{-13}$ )                                    |
| 0060589~nucleoside-triphosphatase regulator activity ( $10^{-3}$ )                           | 0001653~peptide receptor activity ( $10^{-13}$ )                            |
| 0016740~transferase activity ( $10^{-3}$ )                                                   | 0008528~peptide receptor activity, G-protein coupled ( $10^{-13}$ )         |

\* 10 лучших терминов по убыванию достоверности с поправкой Бонферрони.



**Рис. 2.** Достоверная ( $r = 0,41$ ,  $p < 10^{-20}$ ) корреляция между коэффициентом вариации  $C_v$  экспрессии генов из минорного кластера в мозге мыши и сродством ТАТА-связывающего белка к их промоторам,  $-\ln(K_D)$ . Главный кластер (серые «+»): 11 805 генов (91 %) с низким  $189 \pm 98$  % средним  $C_v$ . Минорный кластер (черные «○»): 1 126 генов (9 %) с высоким  $551 \pm 166$  % средним  $C_v$ . Пунктир – 95 %-й доверительный интервал, Statistica (Statsoft™, USA).

**Fig. 2.** The significant correlation ( $r = 0,41$ ,  $p < 10^{-20}$ ) between  $C_v$  values for mRNAs of the minor cluster of genes in the mouse brain and « $-\ln(K_D)$ »-values of TATA-binding protein affinity for promoters of these genes.

The main cluster of genes (+, colored gray), 11 805 out of 12 931 (91 %), clustered by Statistica had the lowest mean  $C_v$ -value,  $189 \pm 98$  %. The minor cluster of genes (○, colored black), 1 126 out of 12 931 (9 %), had the highest mean  $C_v$ -value,  $551 \pm 166$  %. Dashed curves show 95 % confidence intervals for the linear regression (solid lines) built with Statistica.

достоверной корреляцией между  $C_v$  экспрессии генов в мозге мыши и оценкой *in silico*  $-\ln(K_D)$  сродства TBP к их промоторам,  $r = 0,41$  ( $p < 10^{-20}$ ).

Однако пакет DAVID, вып. 6.7 (Huang *et al.*, 2009) не нашел достоверных GO-терминов ни для главного, ни для минорного кластеров генов мыши (данные не показаны). Поскольку главный кластер включал 91 % генов мыши, а минорный всего лишь 9 % генов из опыта (Lein *et al.*, 2007), то с помощью пакета DAVID были дополнительно аннотированы все исследованные гены мыши из опыта (Lein *et al.*, 2007) относительно референсного генома мыши, вып. 69 (Flicek *et al.*, 2011).

Результат GO-аннотации 12 931 гена мыши (Lein *et al.*, 2007) показан в табл. 2. Ни один из

30 терминов GO-аннотации генов мыши не совпал с GO-терминами минорного кластера генов человека, тогда как 22 термина (73 %) совпали с GO-терминами главного кластера генов человека. Это означает достоверное соответствие генов мыши из опыта (Lein *et al.*, 2007) главному кластеру генов человека ( $p < 10^{-17}$ , закон бинома).

## ОБСУЖДЕНИЕ

Сопоставление результатов анализа независимых данных об экспрессии генов в мозге человека (Jones *et al.*, 2009) и в мозге мыши (Lein *et al.*, 2007) показало хорошее качественное согласие одних с другими. Прежде всего, абсолютное большинство генов как у человека, так и у мыши имело конститутивную экс-

Таблица 2

Аннотация всех исследованных генов мыши  
из экспериментальных данных  
(Lein *et al.*, 2007) с помощью пакета DAVID,  
вып. 6.7 (Huang *et al.*, 2009)

| Термин Gene Ontology (GO)*                              | Достоверность<br>с поправкой Бонферрони |
|---------------------------------------------------------|-----------------------------------------|
| <b>Биологический процесс</b>                            |                                         |
| GO:0009987~cellular process                             | $1,80 \times 10^{-107}$                 |
| GO:0008152~metabolic process                            | $8,56 \times 10^{-57}$                  |
| GO:0044237~cellular metabolic process                   | $2,85 \times 10^{-47}$                  |
| GO:0044238~primary metabolic process                    | $5,29 \times 10^{-47}$                  |
| GO:0043170~macromolecule metabolic process              | $1,51 \times 10^{-29}$                  |
| GO:0044260~cellular macromolecule metabolic process     | $1,27 \times 10^{-27}$                  |
| <i>GO:0019538~protein metabolic process</i>             | $7,32 \times 10^{-15}$                  |
| GO:0044267~cellular protein metabolic process           | $2,86 \times 10^{-14}$                  |
| <i>GO:0006807~nitrogen compound metabolic process</i>   | $5,63 \times 10^{-14}$                  |
| GO:0034641~cellular nitrogen compound metabolic process | $2,72 \times 10^{-12}$                  |
| <b>Внутриклеточный компартмент</b>                      |                                         |
| GO:0044424~intracellular part                           | $8,11 \times 10^{-89}$                  |
| GO:0005737~cytoplasm                                    | $3,45 \times 10^{-79}$                  |
| GO:0005622~intracellular                                | $5,29 \times 10^{-74}$                  |
| GO:0043226~organelle                                    | $9,40 \times 10^{-62}$                  |
| GO:0043229~intracellular organelle                      | $2,03 \times 10^{-61}$                  |
| GO:0043231~intracellular membrane-bounded organelle     | $6,97 \times 10^{-59}$                  |
| GO:0043227~membrane-bounded organelle                   | $7,03 \times 10^{-59}$                  |
| GO:0044444~cytoplasmic part                             | $7,35 \times 10^{-51}$                  |
| <i>GO:0044422~organelle part</i>                        | $2,40 \times 10^{-16}$                  |
| <i>GO:0044446~intracellular organelle part</i>          | $4,89 \times 10^{-16}$                  |
| <b>Молекулярная функция</b>                             |                                         |
| GO:0005488~binding                                      | $1,01 \times 10^{-75}$                  |
| GO:0003824~catalytic activity                           | $6,99 \times 10^{-42}$                  |
| GO:0005515~protein binding                              | $2,26 \times 10^{-39}$                  |
| GO:0000166~nucleotide binding                           | $6,99 \times 10^{-21}$                  |
| GO:0017076~purine nucleotide binding                    | $5,03 \times 10^{-20}$                  |
| <i>GO:0032553~ribonucleotide binding</i>                | $2,84 \times 10^{-17}$                  |
| <i>GO:0032555~purine ribonucleotide binding</i>         | $2,84 \times 10^{-17}$                  |
| GO:0016740~transferase activity                         | $6,76 \times 10^{-16}$                  |
| <i>GO:0001883~purine nucleoside binding</i>             | $2,28 \times 10^{-15}$                  |
| <i>GO:0001882~nucleoside binding</i>                    | $3,64 \times 10^{-15}$                  |

\* 10 лучших терминов по убыванию их достоверности; курсив – 8 GO-терминов, отсутствующих в GO-аннотации главного кластера генов человека, табл. 1.

прессию для функций «домашнего хозяйства» внутри клеток мозга. При этом меньшинство генов имело вариабельную экспрессию в мозге. Наконец, для генов путей передачи сигналов в мозге человека была обнаружена достоверная корреляция между  $C_v$  экспрессии этих генов и оценкой  $-\ln(K_D)$  сродства ТВР к их промоторам. Эта корреляция была в согласии с такой корреляцией для минорного кластера генов мыши. Она указывает на контекстный сигнал генома – ТАТА бокс, который вносит достоверный вклад в вариабельность экспрессии генов путей передачи сигналов в мозге человека. Это главный результат нашей работы.

Прежде всего, этот результат согласуется с выводами ряда независимых полногеномных исследований. Например, Nakamura с соавт. (2002), Yang с соавт. (2007), Dong с соавт. (2011) достоверно ассоциировали пути передачи сигналов с ТАТА-содержащими промоторами генов, тогда как у генов «домашнего хозяйства» достоверно более частыми были ТАТА-несодержащие промоторы.

Аналогично гены с ТАТА-содержащими промоторами были ассоциированы со стохастическим ответом на индукторы в отличие от детерминированного ответа в случае генов с ТАТА-несодержащими промоторами (Raser, O'Shea, 2004). Также индуцибельная транскрипция была ассоциирована с ТАТА-содержащими промоторами генов, в то время как конститутивная – с ТАТА-несодержащими промоторами (Toia, Timmers, 2010). Наконец, пластичная регуляция экспрессии была достоверно ассоциирована с ТАТА-содержащими промоторами генов, тогда как постоянная – с ТАТА-несодержащими промоторами генов (Lin *et al.*, 2010).

В свою очередь, обнаружение в геноме контекстного сигнала – ТАТА бокса, который может вносить достоверный вклад в вариабельность экспрессии генов, развивает существующие представления о реализационной (Струнников, 1989), внутрииндивидуальной (Черданцев, 2003) и интерференционной (Камшилов, 1939; Belyaev, 1979; Аршавский, 1982) изменчивости признаков организмов без необходимости соответствующей вариабельности условий внешней среды.

Обнаруженная связь  $C_v$  экспрессии генов со сродством ТВР/промотор также может иметь

практическую ценность в анализе транскриптомов по крайней мере с целью предварительной группировки генов со сходными паттернами экспрессии – взаимно коррелирующими, циклическими, импульсными, запаздывающими, пороговыми и т. п. для последующего анализа каждой такой группы генов отдельно с помощью соответствующих статистических методов (Oh *et al.*, 2014).

Целесообразность подобной группировки генов на основе сравнения оценок *in vivo* для  $C_v$  экспрессии генов с оценками *in silico* для сродства ТВР к промоторам этих генов показана на рис. 1. Например, мысленно проводя горизонтальную линию, можно разделить все гены на две группы с низкими и с высокими оценками  $C_v$  *in vivo*, как и с помощью такой вертикальной линии – с низкими и высокими оценками *in silico* сродства ТВР к промоторам генов. В случаях каждой из этих ортогональных линий раздела большинство генов может быть сгруппировано правильно, хотя и с очень большой долей ошибочно сгруппированных генов из-за взаимосвязи этих оценок *in vivo* и *in silico*, а также вследствие соответствующей этой взаимосвязи нелинейной границы между группами генов. Поиск взаимосвязи между оценками *in vivo* и *in silico* способствует биологической интерпретации транскриптомов (для транскриптома мозга человека, см. табл. 1).

В свою очередь, опыт *in vivo* (Mogno *et al.*, 2010) доказал значимую корреляцию экспрессии генов со сродством ТВР к их промоторам. В приближении «лимитирующей стадии» сайт на ДНК для связывания ТВР – ТАТА-бокс, – может быть оптимизирован лишь для «лимитирующей» функции полифункционального гена. Тогда для всех остальных «нелимитирующих» функций экспрессия этого гена будет тем избыточнее, чем выше сродство ТВР к оптимизированному ТАТА-боксу этого гена. Эта избыточная экспрессия полифункциональных генов для их «нелимитирующих» функций может быть одним из возможных молекулярных механизмов вклада ТАТА-боксов в вариабельность экспрессии некоторых генов.

Наконец, в опыте Godde с соавт. (1995) было показано, что связывание ТВР с промотором ослабляет его нуклеосомную упаковку до уровня, необходимого для сборки преинициаторного



комплекса транскрипции. В этой связи представляется возможным предложить гипотезу о том, что нуклеосомный потенциал и CpG-острова для метилирования промотора также могут быть контекстными сигналами-кандидатами в геноме, способными вносить вклад в вариабельность экспрессии генов. Эта гипотеза согласуется с наблюдениями в опытах Gao с соавт. (2005), Mottis с соавт. (2013) фактов влияния нуклеосомного потенциала ДНК и CpG-островов на эпигенетическую изменчивость определенных признаков некоторых организмов.

## ЗАКЛЮЧЕНИЕ

В этой работе на примере транскриптома мозга человека была обнаружена достоверная связь вариабельности экспрессии генов путей передачи сигналов со средством ТВР к промоторам этих генов. Эта связь означает наличие в геноме контекстного сигнала – ТАТА-бокса, который вносит вклад в вариабельность экспрессии генов. Эта ранее не известная биологическая функция ТАТА-боксов дополняет их экспериментально доказанный линейно-аддитивный вклад в уровни экспрессии генов (Mogno *et al.*, 2010) и развивает существующие представления о некоторых формах изменчивости, которые не требуют необходимости изменений условий среды вокруг организмов. Показано также, что учет обнаруженной связи вариабельности экспрессии генов со средством ТВР к промоторам этих генов может способствовать улучшению биологической интерпретации транскриптома.

## БЛАГОДАРНОСТИ

Работа была поддержана грантами РФФИ 12-04-33112 и 14-04-00485, Интеграционным проектом СО РАН № 136 и проектом VI.61.1.2.

## ЛИТЕРАТУРА

- Аршавский И.А. Физиологические механизмы и закономерности индивидуального развития. М.: Наука, 1982. 270 с.
- Arshavsky I.A. Physiological mechanisms and patterns of individual development. Moscow: Nauka, 1982. 270 s. (In Russian).
- Дарвин Ч. Происхождение видов. СПб.: Наука, 1991. 540 с.
- Darwin C.R. On the origin of species. SPb.: Nauka, 1991. 540 p. (In Russian).
- Иогансен В. Элементы точного учения об изменчивости и наследственности с основами биологической вариационной статистики. М.: Сельхозгиз, 1933. 410 с.
- Johannsen W.L. The elements of an exact theory of variability and heredity based on biological variation statistics, Moscow: Sel'khozgiz, 1933. 410 s. (In Russian).
- Камшилов М.М. Отбор как фактор, меняющий зависимость признака от изменений внешних условий // Докл. АН СССР. 1939. Т. 23. № 4. С. 361–364.
- Kamshilov M.M. Selection as a factor that changes the trait dependence on changes in environmental conditions // Dokl. Akad. Nauk. 1939. V. 23. No. 4. P. 361–364. (In Russian).
- Кислинский К.Н. Оценка устойчивости к загущению различных гибридов кукурузы по величине коэффициента вариации морфоанатомических параметров // Кукуруза и сорго. 1997. № 5. С. 9–10.
- Kislinsky K.N. Evaluation of resistance to thickening of corn hybrids by the coefficient of variation of morphoanatomic measures // Kukuruza i sorgo. 1997. No. 5. P. 9–10. (In Russian).
- Кравцов А.Л., Ляпин М.Н., Шмелькова Т.П., Головки Е.М., Милукова Т.А., Костюкова Т.А., Езов И.Н. Определение содержания ДНК в отдельных клетках *Vibrio cholerae* методом проточной цитофлуориметрии: сравнительный анализ неоднородности в клетках штаммов с различными биологическими свойствами // Журн. микробиол., эпидемиол. и иммунобиологии. 2011. С. 7–11.
- Kravtsov A.L., Liapin M.N., Shmel'kova T.P., Golovko E.M., Maliukova T.A., Kostiukova T.A., Ezhov I.N. Determination of DNA content in individual *Vibrio cholerae* cells by the flow cytofluorometry method: comparative analysis of inhomogeneity in cells of strains with different biological properties // Zh. Mikrobiol. Epidemiol. Immunobiol. 2011. No. 5. P. 7–11. (In Russian).
- Кулаков Б.С., Ибрагимов Ю.Н., Завгородняя Г.В., Остроухов Н.А., Санькова О.Б. Характеристика тонкой шерсти некоторых пород овец Северного Кавказа // Овцы, козы, шерстяное дело. 1999. № 3. С. 10.
- Kulakov B.S., Ibragimov Yu.N., Zavgorodnyaya G.V., Ostroukhov N.A., Sankova O.B. Some characteristics of fine wool sheep breeds of North Caucasus // Ovtzy, kozy, sherstyanoie delo. 1999. No. 3. P. 10. (In Russian).
- Куркина Ю.Н. Мутанты с детерминантным типом роста // Информ. вестн. ВОГиС. 2004. № 30. С. 146–148.
- Kurkina Yu.N. Mutants with the determinant type of growth // VOGiS Herald. 2004. V. 8. No. 3 (30). P. 146–148. (In Russian).
- Левонтин Р. С. Генетические основы эволюции. М.: Мир, 1978. 338 с.
- Lewontin R.C. The genetic basis of evolutionary change. Moscow: Mir, 1978. 338 p. (In Russian).
- Михайлов Н., Свинарев И., Гончаров А. Мясные качества трехпородных гибридов // Животноводство России. 2011. № 3. С. 25–26.
- Mikhailov N., Svinarev I., Goncharov A. Meat quality of three-breed hybrids // Zhivotnovodstvo Rossii. 2011. No.

3. P. 25–26. (In Russian).
- Рассказов Д.А., Гунбин К.В., Пономаренко М.П., Вишне-  
невский О.В., Пономаренко М.П., Афонников Д.А.  
SNP\_TATA\_COMPARATOR: Web-сервис применения  
уравнения равновесия TBP/TATA-комплекса в сравни-  
тельной оценке SNPs промоторов генов, связанных с  
болезнями человека // Вавилов. журн. генет. и селек-  
ции. 2013. Т. 17. № 4/1. С. 599–606.
- Rasskazov D.A., Gunbin K.V., Ponomarenko M.P.,  
Vishnevsky O.V., Ponomarenko M.P., Afonnikov D.A.  
SNP\_TATA\_COMPARATOR: Web-service for comparison  
of SNPs within gene promoters associated with human  
diseases using the equilibrium equation of the TBP/TATA  
complex // Vavilovskii Zhurn. Genetiki i Selekcii. 2013.  
V. 17. No. 4/1. P. 599–606. (In Russian).
- Расторгуйев С.Л. Изменчивость растений-регенерантов  
земляники, полученных методом тканевых культур  
// Вопр. соврем. науки и практики. 2008. Т. 2. № 1.  
С. 46–51.
- Rastorgouev S.L. The variability of regenerated strawberry  
plants grown by tissue culture // Voprosy sovremennoi nauki  
i praktiki. 2008. V. 2. No. 1. P. 46–51. (In Russian).
- Струнников В.А. Третья изменчивость // Природа. 1989.  
№ 2. С. 17–27.
- Strunnikov V.A. Third variability // Priroda. 1989. No. 2.  
P. 17–27. (In Russian).
- Черданцев В.А. Морфогенез и эволюция. М.: Тов-во на-  
учных изданий КМК, 2003. 360 с.
- Cherdantsev V.G. Morphogenesis and evolution. Moscow:  
KMK, 2003. 360 p. (In Russian).
- Balazsi G., van Oudenaarden A., Collins J.J. Cellular decision  
making and biological noise: from microbes to mammals  
// Cell. 2011. V. 144. No. 6. P. 910–925.
- Belyaev D.K. Destabilizing selection as a factor in domestica-  
tion / The Wilhelmine E. Key 1978 Invitational Lecture //  
J. Heredity. 1979. V. 70. P. 301–308.
- Blake W.J., Balazsi G., Kohanski M.A., Isaacs F.J., Mur-  
phy K.F., Kuang Y., Cantor C.R., Walt D.R., Collins J.J.  
Phenotypic consequences of promoter-mediated transcrip-  
tional noise // Mol. Cell. 2006. V. 24. No. 6. P. 853–865.
- Chalancon G., Ravarani C.N., Balaji S., Martinez-Arias A.,  
Aravind L., Jothi R., Babu M.M. Interplay between gene  
expression noise and regulatory network architecture //  
Trends Genet. 2012. V. 28. No. 5. P. 221–232.
- Dong D., Yuan Z., Zhang Z. Evidences for increased expres-  
sion variation of duplicate genes in budding yeast: from  
cis- to trans-regulation effects // Nucl. Acids Res. 2011.  
V. 39. No. 3. P. 837–847.
- Dreszer T.R., Karolchik D., Zweig A.S., Hinrichs A.S., Ra-  
ney B.J., Kuhn R.M., Meyer L.R., Wong M., Sloan C.A.,  
Rosenbloom K.R., Roe G., Rhead B., Pohl A., Malladi V.S.,  
Li C.H., Learned K., Kirkup V., Hsu F., Harte R.A., Gu-  
ravadoo L., Goldman M., Giardine B.M., Fujita P.A., Die-  
khans M., Cline M.S., Clawson H., Barber G.P., Haussler D.,  
James Kent W. The UCSC Genome Browser database: ex-  
tensions and updates 2011 // Nucl. Acids Res. 2012. V. 40.  
Database issue. P. D918–D923.
- ENCODE Project Consortium. An integrated encyclopedia  
of DNA elements in the human genome // Nature. 2012.  
V. 489. No. 7414. P. 57–74.
- Flicke P., Amode M.R., Barrell D., Beal K., Brent S., Chen Y.,  
Clapham P., Coates G., Fairley S., Fitzgerald S., Gordon L.,  
Hendrix M., Hourlier T., Johnson N., Kahari A., Keefe D.,  
Keenan S., Kinsella R., Kokocinski F., Kulesha E., Lars-  
son P., Longden I., McLaren W., Overduin B., Pritchard B.,  
Riat H.S., Rios D., Ritchie G.R., Ruffier M., Schuster M.,  
Sobral D., Spudich G., Tang Y.A., Trevanion S., Vandro-  
cova J., Vilella A.J., White S., Wilder S.P., Zadissa A.,  
Zamora J., Aken B.L., Birney E., Cunningham F., Dunham I.,  
Durbin R., Fernandez-Suarez X.M., Herrero J., Hubbard T.J.,  
Parker A., Proctor G., Vogel J., Searle S.M. Ensembl  
2011 // Nucl. Acids Res. 2011. V. 39. Database issue.  
P. D800–D806.
- Gao S., Skeldal S., Kroghdal A., Sorensen J.A., Andreassen P.A.  
CpG methylation of the PAI-1 gene 5'-flanking region  
is inversely correlated with PAI-1 mRNA levels in hu-  
man cell lines // Thromb Haemost. 2005. V. 94. No. 3.  
P. 651–660.
- Georger B., Morland B., Ndiaye A., Doz F., Kalifa G., Ge-  
offray A., Pichon F., Frappaz D., Chatelut E., Opolon P.,  
Hain S., Boderet F., Bosq J., Emile J.F., Le Deley M.C.,  
Capdeville R., Vassal G. (Innovative Therapies for Chil-  
dren with Cancer European Consortium). Target-driven  
exploratory study of imatinib mesylate in children with  
solid malignancies by the Innovative Therapies for Chil-  
dren with Cancer (ITCC) European Consortium // Eur. J.  
Cancer. 2009. V. 45. No. 13. P. 2342–2351.
- Ghosh S., Sureka K., Ghosh B., Bose I., Basu J., Kundu M.  
Phenotypic heterogeneity in mycobacterial stringent re-  
sponse // BMC Syst. Biol. 2011. V. 5. P. 18.
- Gilbert D., Fuss H., Gu X., Orton R., Robinson S., Vyshe-  
mirsky V., Kurth M.J., Downes C.S., Dubitzky W. Computa-  
tional methodologies for modelling, analysis and simulation of  
signalling networks // Brief. Bioinform. 2006. V. 7. No. 4.  
P. 339–353.
- Godde J.S., Nakatani Y., Wolffe A.P. The amino-terminal  
tails of the core histones and the translational position  
of the TATA box determine TBP/TFIIA association with  
nucleosomal DNA // Nucl. Acids Res. 1995. V. 23. No. 22.  
P. 4557–4564.
- Huang da W., Sherman B.T., Lempicki R.A. Systematic and  
integrative analysis of large gene lists using DAVID bio-  
informatics resources // Nat. Protoc. 2009. V. 4. No. 1.  
P. 44–57.
- Jones A., Overly C., Sunkin S. The Allen Brain Atlas: 5 years  
and beyond // Nat. Rev. Neurosci. 2009. V. 10. No. 11.  
P. 821–828.
- Lein E.S., Hawrylycz M.J., Ao N., Ayres M., Bensinger A.,  
Bernard A., Boe A.F., Boguski M.S., Brockway K.S.,  
Byrnes E.J., Chen L., Chen L., Chen T.M., Chin M.C.,  
Chong J., Crook B.E., Czaplinska A., Dang C.N., Datta S.,  
Dee N.R., Desaki A.L., Desta T., Diep E., Dolbeare T.A.,  
Donelan M.J., Dong H.W., Dougherty J.G., Duncan B.J.,  
Ebbert A.J., Eichele G., Estin L.K., Faber C., Facer B.A.,  
Fields R., Fischer S.R., Fliss T.P., Frensley C., Gates S.N.,  
Glatfelter K.J., Halverson K.R., Hart M.R., Hohmann J.G.,  
Howell M.P., Jeung D.P., Johnson R.A., Karr P.T., Kaval R.,  
Kidney J.M., Knapik R.H., Kuan C.L., Lake J.H., Lara-  
mee A.R., Larsen K.D., Lau C., Lemon T.A., Liang A.J.,  
Liu Y., Luong L.T., Michaels J., Morgan J.J., Morgan R.J.,

- Mortrud M.T., Mosqueda N.F., Ng L.L., Ng R., Orta G.J., Overly C.C., Pak T.H., Parry S.E., Pathak S.D., Pearson O.C., Puchalski R.B., Riley Z.L., Rockett H.R., Rowland S.A., Royall J.J., Ruiz M.J., Sarno N.R., Schaffnit K., Shapovalova N.V., Sivasay T., Slaughterbeck C.R., Smith S.C., Smith K.A., Smith B.I., Sodt A.J., Stewart N.N., Stumpf K.R., Sunkin S.M., Sutram M., Tam A., Teemer C.D., Thaller C., Thompson C.L., Varnam L.R., Visel A., Whitlock R.M., Wohnoutka P.E., Wolkey C.K., Wong V.Y., Wood M., Yaylaoglu M.B., Young R.C., Youngstrom B.L., Yuan X.F., Zhang B., Zwingman T.A., Jones A.R. Genome-wide atlas of gene expression in the adult mouse brain // *Nature*. 2007. V. 445. P. 168–176.
- Lin Z., Wu W.S., Liang H., Woo Y., Li W.H. The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation // *BMC Genomics*. 2010. V. 11. P. 581.
- Mironova V.V., Omelyanchuk N.A., Savina M.S., Ponomarenko P.M., Ponomarenko M.P., Likhoshvai V.A., Kolchanov N.A. How multiple auxin responsive elements interact in plant promoters: evidences from a reverse problem solution // *J. Bioinform. Comput. Biol.* 2013. V. 11. No. 1. P. 1340011.
- Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters // *Genome Res*. 2010. V. 20. No. 10. P. 1391–1397.
- Mottis A., Mouchiroud L., Auwerx J. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis // *Genes Dev*. 2013. V. 27. No. 8. P. 819–835.
- Munsky B., Neuert G., van Oudenaarden A. Using gene expression noise to understand gene regulation // *Science*. 2012. V. 336. No. 6078. P. 183–187.
- Nakamura M., Tsunoda T., Obokata J. Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator // *Plant J*. 2002. V. 29. No. 1. P. 1–10.
- Oh S., Song S., Dasgupta N., Grabowski G. The analytical landscape of static and temporal dynamics in transcriptome data // *Front. Genet.* 2014. V. 5. P. 35.
- Ou Q.S., Cheng Z.J., Yang B., Jiang L., Chen J. Analysis of the ratio of mitochondrial DNA with A1555G mutant to wild type in deaf patients of Fujian province in China by a new method and its relationship with the severity of hearing loss // *Chin. Med. J. (Engl)*. 2011. V. 124. No. 20. P. 3347–3352.
- Pardo L.M., MacKay I., Oostra B., van Duijn C.M., Aulchenko Y.S. The effect of genetic drift in a young genetically isolated population // *Ann. Hum. Genet.* 2005. V. 69. Pt. 3. P. 288–295.
- Ponomarenko M., Mironova V., Gunbin K., Savinkova L. Hogness Box // *Brenner's Encyclopedia of Genetics*: 2nd ed. / Eds S. Maloy, K. Hughes. San Diego: Academic Press, Elsevier Inc. 2013. V. 3. P. 491–494.
- Pruitt K., Tatusova T., Brown G., Maglott D. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy // *Nucl. Acids Res.* 2012. V. 40. Database issue. P. D130–D135.
- Raser J.M., O'Shea E.K. Control of stochasticity in eukaryotic gene expression // *Science*. 2004. V. 304. No. 5678. P. 1811–1814.
- Rausenberger J., Fleck C., Timmer J., Kollmann M. Signatures of gene expression noise in cellular systems // *Prog. Biophys. Mol. Biol.* 2009. V. 100. No. 1/3. P. 57–66.
- Rosa B.A., Jiao Y., Oh S., Montgomery B.L., Qin W., Chen J. Frequency-based time-series gene expression recomposition using PRIISM // *BMC Syst. Biol.* 2012. V. 6. P. 69.
- Stepanova A.N., Alonso J.M. Ethylene signaling and response: where different regulatory modules meet // *Curr. Opin. Plant Biol.* 2009. V. 12. No. 5. P. 548–555.
- Tora L., Timmers H.T. The TATA box regulates TATA-binding protein (TBP) dynamics *in vivo* // *Trends Biochem. Sci.* 2010. V. 35. No. 6. P. 309–314.
- Whitelaw N.C., Chong S., Whitelaw E. Tuning in to noise: epigenetics and intangible variation // *Cell*. 2010. V. 19. No. 5. P. 649–650.
- Willems P., Verhagen O., Segeren C., Veenhuizen P., Guikema J., Wiemer E., Groothuis L., Jong T.B., Kok H., Bloem A., Bos N., Vellenga E., Mensink E., Sonneveld P., Lokhorst H., van Der Schoot E., Raymakers R. (Belgium-Dutch Hematology-Oncology Group) Consensus strategy to quantitate malignant cells in myeloma patients is validated in a multicenter study // *Blood*. 2000. V. 96. No. 1. P. 63–70.
- Yang C., Bolotin E., Jiang T., Sladek F.M., Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters // *Gene*. 2007. V. 389. No. 1. P. 52–65.

**IDENTIFICATION OF THE RELATIONSHIP  
BETWEEN THE VARIABILITY OF THE EXPRESSION  
OF SIGNALING PATHWAY GENES IN THE HUMAN BRAIN  
AND AFFINITY OF TATA-BINDING PROTEIN TO THEIR PROMOTERS**

**M.P. Ponomarenko<sup>1,2</sup>, V.V. Suslov<sup>1</sup>, K.V. Gunbin<sup>1,2</sup>, P.M. Ponomarenko<sup>3</sup>,  
O.V. Vishnevsky<sup>1,2</sup>, N.A. Kolchanov<sup>1,2</sup>**

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia;

<sup>3</sup> Children's Hospital Los Angeles, Los Angeles, CA, USA

**Summary**

Variations in gene expression are the subject of a wide range of research aimed to reconstruct regulatory protein binding sites on DNA, regulatory gene regions, gene networks, signal transduction pathways, and many other entities able to modify gene expression. Pearson's coefficient of variation (Cv, the ratio of the standard deviation to the mean) has been the most often used measure of variations in gene expression within the fields of intense research in biomedicine and breeding. In turn, only one common genomic regulatory signal has been identified in all eukaryotes, namely, the TATA-box together with three other less conservative obligatory elements of promoter cores accompanying it: DPE, BPE, and INR. Using our experimentally proven equation for TATA-box Binding Protein (TBP) affinity for TATA-boxes, we analyzed high-throughput sequencing data on 35609 mRNAs in 946 human brain segments taken from The Allen Brain Atlas. We found a significant correlation between the affinity of the TATA-box binding protein (TBP) binding to promoters of genes of signaling pathways and Pearson's coefficient of variation of the expression of these genes was found. This finding may help scientists analyze high-throughput sequencing data in order to identify more candidate factors modifying gene expression.

**Key words:** gene, promoter, TATA-binding protein (TBP), of TBP/DNA-affinity, gene expression, coefficient of variation (Cv), correlation, brain, human.