

**Тема работы:**

Исследование молекулярной эволюции регуляторных участков генов человека и гоминид.

**Состав коллектива:**

Гунбин К.В. к.б.н., с.н.с. ИЦИГ СО РАН

**Информация о гранте:**

Грант Министерство образования и науки РФ 14.В25.31.0033 (решение № 220). Руководитель: Рогаев Е.И. 2013 – 2017гг.

**Научное содержание работы:***1. Постановка задачи.*

Человек обладает по крайней мере одной характеристикой, сильно отличающей его от низших гоминид (шимпанзе и гориллы) – эврибионтностью [Manica et al., 2005], которая, в свою очередь, связана с широкой нормой реакции генотипа [Schmalhausen, 1986]. Следовательно, искомым, гоминидспецифичный вектор отбора, мог быть отбором на расширение нормы реакции, то есть быть связанным не столько с уровнем экспрессии генов, сколько со способностью варьировать этот уровень в широких пределах. Поиск следов такого отбора в полногеномных данных до сих пор не проводился.

*2. Современное состояние проблемы.*

Полногеномные сравнения белоккодирующих последовательностей между человеком и шимпанзе дают оценки различий в пределах 1-2%, что резко контрастирует как с морфологическими, так и с экологическими различиями [CSAC, 2005]. Близкие оценки получены для человека и гориллы [Sally et al., 2012]. Близкими оказались и функциональные наборы генов, наиболее подверженные адаптивной эволюции. У всех трех видов прежде всего эволюционировали сперматогенез, рецепторы, иммунная система, гены передачи сигналов (например, G-белки) и регуляции эмбриогенеза [CSAC, 2005, Arbiza et al., 2006, Sally et al., 2012]. Ожидалось, что в линии человека видоспецифическим будет повышенный темп эволюции генов мозга. Но эти надежды не оправдались. Удивительно, но генов, сочетающих строгую мозг-специфичную экспрессию и адаптивную эволюцию белок-кодирующих последовательностей, выявлено крайне мало [Nielsen et al., 2005, Shi et al., 2006, Sherwood et al., 2010], они эволюционируют не только у человека, но и низших гоминид [CSAC, 2005, Arbiza et al., 2006, Sally et al., 2012] и единственной человеческой спецификой в них оказался спектр мутаций. Кроме того, по клиническим данным изменения в этих генах ведут у человека к тяжелым патологиям [Evans et al., 2006, Harris, Meyer, 2006, Vamathevan et al., 2008, Sally et al., 2012], а это значит, что отбор не мог действовать на них напрямую: мутации в них могли быть зафиксированы только на фоне компенсирующего патологический эффект отбора. Если адаптивная эволюция нейроспецифичных генов и выявлялась, то это была недавняя эволюция [Huang et al., 2013], в отличие от дивергенции человека и шимпанзе [White et al., 2006, Kunitatsu et al., 2007, Sally et al., 2012]. Таким образом,

главные генетические детерминанты эволюции человека по-видимому локализованы вне области белок-кодирующих генов – в их регуляторных участках, отвечающих за норму реакции генотипа.

Для того чтобы обнаружить следы отбора на расширение нормы реакции генотипа в настоящей работе мы сосредоточили внимание на структурной и функциональной эволюции коровых промотеров генов человека, в особенности промотеров генов, экспрессирующихся в префронтальной зоне коры головного мозга. В 2011 году впервые, для 17181 гена человека была построена карта локализации ТАТА боксов [Yang et al., 2011]. В то время карта была основана только на данных CAGE. В настоящее время за основу детекции местоположений промотеров, принято использовать несколько независимых экспериментальных подходов [ENCODE Project Consortium, 2012; Kawaji et al., 2014].

1. Arbiza L., Dopazo J., Dopazo H. Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome // *PLoS Comput. Biol.* 2006; 2:e38.
2. CSAC - Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome // *Nature.* 2005; 437(7055):69-87.
3. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489(7414):57-74.
4. Evans P.D., Vallender E.J., Lahn B.T. Molecular evolution of the brain size regulator genes CDK5RAP2 and CENPJ // *Gene.* 2006; 375:75-79.
5. Harris E.E., Meyer D. The molecular signature of selection underlying human adaptations // *Am. J. Phys. Anthropol.* 2006; Suppl 43:89-130.
6. Huang Y., Xie C., Ye A.Y., Li C.Y., Gao G., Wei L. Recent adaptive events in human brain revealed by meta-analysis of positively selected genes // *PLoS One.* 2013; 8(4):e61280.
7. Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, Shin JW, Kojima-Ishiyama M, Kawano M, Murata M, Ninomiya-Fukuda N, Ishikawa-Kato S, Nagao-Sato S, Noma S, Hayashizaki Y, Forrest AR, Carninci P; FANTOM Consortium. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res.* 2014; 24(4):708-17.
8. Kunitatsu Y., Nakatsukasa M., Sawada Y., Sakai T., Hyodo M., Hyodo H., Itaya T., Nakaya H., Saegusa H., Mazurier A., Saneyoshi M., Tsujikawa H., Yamamoto A., Mbua E. A new Late Miocene great ape from Kenya and its implications for the origins of African great apes and humans // *Proc Natl Acad Sci U S A.* 2007; 104(49):19220-19225.
9. Manica A., Prugnolle F., Balloux F. Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* 2005; 118, 366–371.
10. Nielsen R., Bustamante C., Clark A.G., Glanowski S., Sackton T.B., Hubisz M.J., Fledel-Alon A., Tanenbaum D.M., Civello D., White T.J., J. Sninsky J., Adams M.D., Cargill M. A scan for positively selected genes in the genomes of humans and chimpanzees // *PLoS Biol.* 2005; 3(6):e170.
11. Scally A., Dutheil J.Y., Hillier L.W., et al. Insights into hominid evolution from the gorilla genome sequence // *Nature.* 2012; 483(7388):169-175.
12. Schmalhausen I.I. *Factors of Evolution: The Theory of Stabilizing Selection.* Chicago Univ. Press. 1986, 327 p.
13. Sherwood C.C., Raghanti M.A., Stimpson C.D., Spocter M.A., Uddin M., Boddy A.M., Wildman D.E., Bonar C.J., Lewandowski A.H., Phillips K.A., Erwin J.M., Hof P.R. Inhibitory interneurons of the human prefrontal cortex display conserved evolution of the phenotype and related genes // *Proc. Biol. Sci.* 2010; 277(1684):1011-1020.
14. Shi P, Bakewell MA, Zhang J. Did brain-specific genes evolve faster in humans than in chimpanzees? // *Trends Genet.* 2006; 22(11):608-613.
15. Vamathevan J.J., Hasan S., Emes R.D., Amrine-Madsen H., Rajagopalan D., Topp S.D., Kumar V., Word M., Simmons M.D., Foord S.M., Sanseau P., Yang Z., Holbrook J.D. The role of positive

- selection in determining the molecular cause of species differences in disease // BMC Evol Biol. 2008; 8:273.
16. White T.D., Wolde Gabriel G., Asfaw B., Ambrose S., Beyene Y., Bernor R.L., Boissarie J.R., Currie B., Gilbert H., Haile-Selassie Y., Hart W.K., Hlusko L.J., Howell F.C., Kono R.T., Lehmann T., Louchart A., Lovejoy C.O., Renne P.R., Saegusa H., Vrba E.S., Wesselman H., Suwa G. Asa Issie, Aramis and the origin of Australopithecus // Nature. 2006; 440(7086):883-889.
  17. Yang M.Q., Laflamme K., Gotea V., Joiner C.H., Seidel N.E., Wong C., Petrykowska H.M., Lichtenberg J., Lee S., Welch L., Gallagher P.G., Bodine D.M., Elnitski L. Genome-wide detection of a TFIIID localization element from an initial human disease mutation. // Nucleic Acids Res., 2011; 39:2175-2187.

### *3. Подробное описание работы, включая используемые алгоритмы.*

В нашей работе местоположения промоторов были основаны на двух независимых источниках экспериментальной информации: 1) данные CAGE, полученные в ходе многочисленных тканеспецифических экспериментов и 2) H3K4me3 ChIP-seq данные показывающие присутствие комплекса RNAPolIII на определенных участках генома. С использованием H3K4me3 ChIP-seq данных была проведен отбор функциональных промоторов генов среди набора потенциальных промоторов из данных CAGE. Отбирались только те промоторы, которые располагались вблизи H3K4me3 ChIP-seq координат пиков.

В работе проведен анализ эволюции трех характеристик корового промотора – CG пар нуклеотидов, степени нуклеосомной упаковки и сродства к белку TBP. Величина  $\ln(KD)$ , измеряющаяся в единицах натурального логарифма равновесной константы KD диссоциации комплекса ТАТА-связывающего белка (TBP) с ДНК, оценивалась в центре скользящего окна нуклеотидной последовательности длины 26 п.о. на основе использования общепринятой позиционно-весовой матрицы Бухера [Bucher, 1990] и корреляций сродства TBP к одно- и двунитевым олигоДНК с содержанием динуклеотидов TA, WR, TV и шириной малой бороздки спирали ДНК [Ponomarenko M et al., 1999], как это описано в статье [Suslov et al., 2010]. Величина ScoreNUCL оценивалась в центре скользящего окна нуклеотидной последовательности длины 199 п.о. на основе использования позиционно-частотной матрицы экспериментально доказанных сайтов на ДНК для связывания гистоновых октомеров и среднеарифметических величин 38 конформационных и физико-химических свойств спирали ДНК [Ponomarenko J et al., 1999] вокруг центров этих сайтов, включая среднюю частоту контактов динуклеотидов с октомером гистонов [Satchwell et al., 1986], как это описано в статье [Levitsky et al., 1999]. Изменчивость нуклеотидного контекста промоторов генов подсчитывалась с использованием двух районов разной длины [-200; -1] и [-600; -1] относительно TSS. Учитывался как общий уровень изменчивости (подсчитывалась доля всех однонуклеотидных замен и indels между предковой и потомковой последовательностями по отношению к длине всего рассматриваемого участка) так и уровень изменчивости только CG динуклеотидов. Оценка изменчивости CG динуклеотидов проводилась на обеих цепях ДНК, т.е. подсчитывалась доля измененных как CG (+ цепь), так и GC (- цепь) динуклеотидов. В случае анализа района [-200; -1] исследовались только те участки, которые содержали минимум 7 CG/GC динуклеотидов на 200 в среднем для прямой и обратной цепи, что соответствует комбинации критериев Gardiner-Garden и Frommer [Gardiner-Garden, Frommer, 1987; Takai, Jones, 2002].

1. Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. // J. Mol. Biol., 1990, V. 212, P. 563-578.

2. Ponomarenko M.P., Ponomarenko J.V., Frolov A.S., Podkolodny N.L., Savinkova L.K., Kolchanov N.A., Overton G.C. Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins // *Bioinformatics*, 1999, V. 15, P. 687-703.
3. Suslov V.V., Ponomarenko P.M., Efimov V.M., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A. SNPs in the HIV-1 TATA box and the AIDS pandemic // *J. Bioinform. Comput. Biol.*, 2010. V. 8, N. 3, P. 607-625.
4. Ponomarenko J.V., Ponomarenko M.P., Frolov A.S., Vorobyev D.G., Overton G.C., Kolchanov N.A. Conformational and physicochemical DNA features specific for transcription factor binding sites // *Bioinformatics*, 1999a, V. 15, P. 654-668.
5. Satchwell S.C., Drew H.R., Travers A.A. Sequence periodicities in chicken nucleosome core DNA. // *J. Mol. Biol.* 1986. V. 191. N. 4. P. 659-675.
6. Levitsky V.G., Ponomarenko M.P., Ponomarenko J.V., Frolov A.S., Kolchanov N.A. Nucleosomal DNA property database // *Bioinformatics*, 1999, V. 15, N. 7/8. P. 582-592.
7. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. // *J Mol Biol.* 1987, V. 196, N. 2. P. 261-282.
8. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. // *Proc Natl Acad Sci U S A.* 2002, V. 99. N. 6. P. 3740-3745.

#### *4. Полученные результаты*

Было показано, что в линии человека происходит сохранение спектра эпигенетических возможностей регуляции в коровых промотерах генов, накопленного на предшествующих этапах эволюции его двух предков: предка шимпанзе и человека и предка всех гоминид – человека, шимпанзе и гориллы. Напротив, в линиях человекообразных обезьян преобладали изменения, сужающие этот спектр. Это в свою очередь, свидетельствует: во-первых, происхождение человека связано с отбором на расширение нормы реакции регуляции экспрессии генов; во-вторых, напротив, происхождение трех видов человекообразных обезьян связано с видоспецифичным отбором на сужение нормы реакции регуляции экспрессии генов; в-третьих, тренд отбора на расширение нормы реакции регуляции генов у гоминид имеет глубокую древность, как минимум, превосходящую древность дивергенции “шимпанзе – человек”.

#### *5. Иллюстрации, визуализация результатов.*

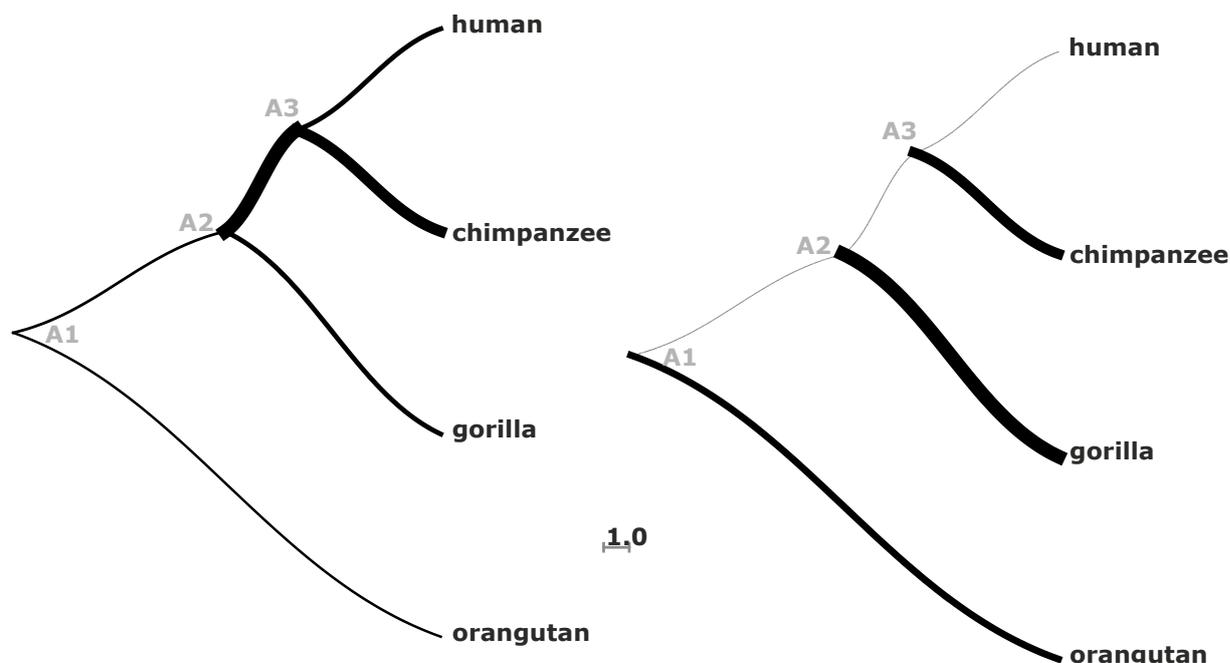


Рис. 1. Два взгляда на дерево приматов. Слева – распределение скоростей фиксации мутаций в коровых промотерах генов (толщина линии отображает скорость фиксации мутаций – жирная линия соответствует повышенной скорости, тонкая пониженной). Справа - распределение функциональных изменений нуклеосомного потенциала и потенциала связывания ТВР в коровых промотерах генов (толщина линии отображает степень изменения генов – жирная линия соответствует изменениям в большем количестве генов, тонкая в меньшем).

#### *6. Эффект от использования кластера в достижении целей работы.*

Проведение компьютерного анализа эволюции промотеров генов с использованием рабочих станций невозможно как вследствие большого объема требуемой оперативной памяти, так и по времени счета. Поэтому использование кластера является определяющим для успешного достижения целей работы.

#### **Перечень публикаций, содержащих результаты работы**

Gunbin KV, Ponomarenko MP, Suslov VV, Gusev F, Fedonin GG, Rogaev EI. Evolution of Brain Active Gene Promoters in Human Lineage Towards the Increased Plasticity of Gene Regulation. Mol Neurobiol. 2018 Mar;55(3):1871-1904. doi: 10.1007/s12035-017-0427-4. IF=5.076