

Особенности распределения мотивов РНК узнающих белков для протяженных генов 15 хромосомы *Homo sapiens*

Федосеева В.Б., Жаринова И.А., Александров А.А.

Институт молекулярной генетики РАН, Москва, Российская Федерация

fvb@img.ras.ru

Работа была посвящена выяснению особенностей распределения сайтов РНК-связывающихся белков. В качестве объекта использовали серию генов, локализованных на правом плече хромосомы 15 генома человека. Отдельные области длинных интронов потенциальных мест связывания белков SRSF1 и PTBP1 имеют волнообразный профиль преимущественно в противофазе друг к другу для этих белков. Для потенциальных сайтов связывания для пары белков SRSF2 и CTCF встречается близкая пиковая локализация, в особенности в первом длинном интроне. Гены с одним длинным интроном имеют более выраженные особенности, чем гены с множественными длинными интронами. Так плотность локализации потенциальных сайтов связывания белка SRSF2 с пре-мРНК первого длинного интрона для выборки генов, экспрессирующихся преимущественно в тканях мозга, заметно превосходит аналогичную плотность для выборки генов, транскрипционно активных в других тканях, тогда как плотность, усредненная для общей геновой последовательности из двух ткане-специфичных выборок не слишком разнятся друг от друга.

Ключевые слова: РНК-связывающиеся белки, длинные интроны, сплайсинг, аргенин-сериновые белки, пиримидиновые тракты.

The distribution of RNA recognizing protein sites for long genes on chromosome 15 of *Homo sapiens*

Fedoseyeva V.B., Zharinova I.A., Alexandrov A.A.

Institute of Molecular Genetics RAS, Moscow, Russian Federation

This work was devoted to elucidation of the distribution peculiarities of sites for RNA binding proteins. As objects we used a series of genes with long introns encoded on the 15th chromosome of *Homo sapiens*. Separate regions of long introns of potential binding sites of SRSF1 and PTBP1 proteins have a wave-like profile mainly in antiphase to each other for these proteins. For potential binding sites, closepike localization occurs for a pair of SRSF2 and CTCF proteins, especially in the first long intron. The genes with one long intron have more pronounced peculiarities than genes with multiple long introns. Thus, the density of localization of potential binding sites of SRSF2 protein with RNA of the first long intron for the sample of genes expressed mainly in brain tissues significantly exceeds the similar density for the sample of genes transcriptionally active in other tissues, while the density averaged over the total gene sequence of two tissue-specific samples is not too different from each other.

Key words: RNA-binding proteins, long intron, splicing, arginine-serine proteins, pyrimidine-tracts.

1. Введение

Ввиду существования огромного массива некодирующих областей в геноме человека особый интерес представляет выяснение их структурно-функциональных свойств. Много исследований посвящены изучению регуляторных сигналов (промоторы, энхансеры, сайты сплайсинга), остается неизученным огромный объем областей, в том числе, в составе генов с длинными интронами. Так, к примеру, гены с длинными интронами (40–

200 тн) присутствуют в значительной степени именно на 15-ой хромосоме человека, многие из них проявляют экспрессию в мозге человека, причем вариативность числа значительных фрагментов в них связана с наследственными болезнями. В рамках модели сплайсинга экзонов, соседствующих с длинными интронами, предполагается предборка сплайсомы [1] в добавление к броуновскому процессу при ко-транскрипционном сплайсинге. По данным статистики длинные интроны, имеющие первые порядковые номера, в большей степени, чем

прочие, участвуют в пост-транскрипционном сплайсинге [2, 3], что может сказываться на предборке сплайсома, они также имеют повышенную скорость транскрипции по сравнению с основной частью генов [4–6]. Эти свойства могут быть обусловлены многими причинами, в частности, влиянием самих интронов и их взаимодействия с компонентами процессов транскрипции и сплайсинга. Поэтому особое значение приобретает выяснение особенностей распределения сайтов РНК связывающихся белков на нуклеотидной последовательности. Из списка белков мы остановились на наиболее значимых активных участниках, а именно, на белках аргенин-серинового семейства SRSF1,2, полипиримидин-связывающихся белках, а также распространенных мотивах, входящих в состав сайтов; сайты ДНК связывающихся белков CTCF и сайты поли-аденилирования как паузо-образующие также использованы при анализе. Для распределения сайтов указанных белков на геномных последовательностях рецепторных субъединиц для гамма-аминомасляной кислоты типа A(β 3) и (α 5) (в печати) выявлены определенные предпочтения, для подтверждения особенностей такого распределения предлагается исследовать эти свойства на примере других протяженных генов с целью возможного обобщения.

2. Результаты

2.1. Особенности локализации мест связывания белка SRSF1

Анализ распределения сайтов связывания белков серин-аргенинового семейства SRSF1, 2 проводили на примере протяженных генов за счет длинных интронов для правого плеча 15 хромосомы человека. Мы ограничились серией из 15, для каждого из которых было выполнено картирование. Для белка SRSF1 (альтернативные обозначения ASR/SF2, SRp30a, SFRS1) последовательности сайтов связывания с РНК были использованы из данных [7–9]. Для белка SRSF1 повышенная плотность потенциальных мест связывания имеет тенденцию локализоваться в разных частях длинных интронов и с обязательным связыванием на 3' конце генов, включающем более дробную экзон-интронную структуру. График плотности имеет волнообразный характер, как правило, ассиметричный, причем, повышенная плотность удалена от сайтов сплайсинга в длинных интронах.

2.2. Особенности локализации мест связывания белка SRSF2

Для белка SRSF2 (альтернативные обозначения SC-35, SRp30b, SFRS2) последовательности сайтов связывания с РНК были использованы из данных [10, 11].

Выявлены тенденции для белка SRSF2: в значительном количестве генов первый интрон имеет преимущество (особенно первая его часть, примыкающая к 5' концу и центр) по плотности мест потенциального связывания, в других случаях есть места связывания и в других длинных интронах. Закономерности более четко прослеживаются для генов с одним длинным интроном. Во многих случаях отмечена корреляция с местами связывания белка CTCF, который связывается с последовательностью ДНК. Это вполне объяснимо в связи с тем, что места связывания CTCF [12], также сильные места нуклеосомного позиционирования, второстепенные места поли-аденилирования и др. вызывают паузы транскрипции, а SRSF2 входит в состав комплекса, помогающего высвободить РНК-полимеразу II из подобных пауз [13].

2.3. Особенности локализации мест связывания с пиримидиновыми мотивами белков РТВ

Для сайтов белков, РТВР1 и РТВР2 (альтернативное обозначение brРТВ) использовали данные из [14–16].

Для сайтов белков, связывающихся с полипиримидиновыми мотивами РТВР1,2 характерна повышенная плотность в разных частях длинных интронов, причем по характеру волнообразное распределение, в основном, находится в противофазе к местам локализации белка SRSF1, хотя есть и пересечения отдельных областей.

2.4. Ткане-специфичность исследуемых белков

Отмеченные особенности для набора исследуемых генов могут проявляться в разных тканях в зависимости от уровня экспрессии исследованных белков SRSF1, SRSF2, РТВР1 и CTCF. На основании данных протеомики [17], представленных на сайте [18], видно, что концентрации для разных тканей сильно варьируют, так для белков SRSF1 и 2 в тканях мозга они находятся на среднем уровне, тогда как для белков РТВР1 и Р2 значительно ниже среднего уровня. Из этого следует значимость сайтов связывания при взаимодействии с белками SRSF1 и 2 в тканях мозга, тогда как полипиримидиновые тракты могут быть значимыми сами по себе в тканях мозга в силу низкого уровня белков РТВР1 и Р2. Более сложная комбинация реализуется для белков РТВР1 и РТВР2 на разных стадиях развития нейронов.

3. Благодарности

Выражаем благодарность Суздальной М.В. за техническую помощь и полезные советы в оформлении и Рязанскому С.С. – за техническую помощь.

4. Список литературы

1. Hollander D., Naftelberg S., Lev-Maor G., Kornblihtt A.R., Ast G. How are short exons flanked by long introns defined and committed to splicing? *Trends Genet.* 2016. V. 32. P. 596–606. doi: [10.1016/j.tig.2016.07.003](https://doi.org/10.1016/j.tig.2016.07.003).
2. De la Mata M., Lafaille C., Kornblihtt A.R. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA.* 2010. V. 16. P. 904–912. doi: [10.1261/rna.1993510](https://doi.org/10.1261/rna.1993510).
3. Bhatt D.M., Pandya-Jones A., Tong J., Barozzi I., Lissner M.M., Natoli G., Black D.L., Smale S.T. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell.* 2012. V. 150. P. 279–290. doi: [10.1016/j.cell.2012.05.043](https://doi.org/10.1016/j.cell.2012.05.043).
4. Singh J., Padgett R.A. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 2009. V. 16. P. 1128–1133. doi: [10.1038/nsmb.1666](https://doi.org/10.1038/nsmb.1666).
5. Shaul O. How introns enhance gene expression. *Int. J. Biochem. Cell. Biol.* 2017. V. 91(B). P. 145–155. doi: [10.1016/j.biocel.2017.06.016](https://doi.org/10.1016/j.biocel.2017.06.016).
6. Veloso A., Kirkconnell K.S., Magnuson B., Biewen B., Paulsen M.T., Wilson T.E., Ljungman M. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 2014. V. 24. P. 896–905. doi: [10.1101/gr.171405.113](https://doi.org/10.1101/gr.171405.113).
7. Liu H.X., Zhang M., Krainer A.R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* 1998. V. 12. P. 1998–2012.
8. Sikora D., Zhang D., Bojic T., Beeharry Y., Tanara A., Pelchat M. Identification of a binding site for ASF/SF2 on an RNA fragment derived from the hepatitis delta virus genome. *PLoS One.* 2013. V. 8. № 1. P. e54832. doi: [10.1371/journal.pone.0054832](https://doi.org/10.1371/journal.pone.0054832).
9. Tacke R., Manley J.L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* 1995. V. 14. P. 3540–3551.
10. Cavaloc Y., Bourgeois C.F., Kister L., Stévenin J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA.* 1999. V. 5. P. 468–483.
11. Liu H.X., Chew S.L., Cartegni L., Zhang M.Q., Krainer A.R. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* 2000. V. 20. P. 1063–1071.
12. Wada Y., Ohta Y., Xu M., Tsutsumi S., Minami T., Inoue K., Komura D., Kitakami J., Oshida N., Papanonis A. et al. A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. USA.* 2009. V. 106. P. 18357–18361. doi: [10.1073/pnas.0902573106](https://doi.org/10.1073/pnas.0902573106).
13. Lui X., Zhou Y., Pandit S., Huang J., Li H., Lin C.Y., Xiao R., Burge C.B., Fu X.D. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell.* 2013. V. 153. P. 855–868. doi: [10.1016/j.cell.2013.04.028](https://doi.org/10.1016/j.cell.2013.04.028).
14. Wagner E.J., Garcia-Blanco M.A. Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol.* 2001. V. 21. P. 3281–3288. doi: [10.1128/MCB.21.10.3281-3288.2001](https://doi.org/10.1128/MCB.21.10.3281-3288.2001).
15. Le Guiner C., Plet A., Galiana D., Gesnel M.C., Del Gatto-Konczak F., Breathnach R. Polypyrimidine tract-binding protein represses splicing of a fibroblast growth factor receptor-2 gene alternative exon through exon sequences. *J. Biol. Chem.* 2001. V. 276. P. 43677–43687. doi: [10.1074/jbc.M107381200](https://doi.org/10.1074/jbc.M107381200).
16. Spellman R., Rideau A., Matlin A., Gooding C., Robinson F., McGlincy N., Grellscheid S.N., Southby J., Wollerton M., Smith C.W. Regulation of alternative splicing by PTB and associated factors. *Biochem. Soc. Trans.* 2005. V. 33. № 3. P. 457–460. doi: [10.1042/BST0330457](https://doi.org/10.1042/BST0330457).
17. Fagerberg L., Hallström B.M., Oksvold P., Kampf C., Djureinovic D., Odeberg J., Habuka M., Tahmasebpour S., Danielsson A., Edlund K., et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics.* 2014. V. 13. P. 397–406. doi: [10.1074/mcp.M113.035600](https://doi.org/10.1074/mcp.M113.035600).
18. <https://www.ncbi.nlm.nih.gov/gene/?term=ID+of+protein> (дата обращения: 09.08.2018).