

АНАЛИЗ *IN SILICO* САЙТОВ СВЯЗЫВАНИЯ miRNAs ВОКРУГ ГЕНОВ ЛИПИДНОГО ОБМЕНА И ФЕРМЕНТОВ АНТИОКСИДАНТНОЙ ЗАЩИТЫ

М.А. Шкурат¹, М.А. Амелина^{1,2}, Н.А. Григорян², Н.С. Пономарева¹

¹Южный федеральный университет, 344090, Российская Федерация, г. Ростов-на-Дону, пр. Стачки, 194/1

²Ростовский государственный медицинский университет Минздрава России, 344022, Российская Федерация, г. Ростов-на-Дону, пер. Нахичеванский, 29

E-mail: mikhail@shkurat.com; amelinamascha@yandex.ru; narine_g69@mail.ru

Недавние исследования показали, что некодирующие РНК (ncRNAs) регулируют многие процессы, такие как транскрипция, трансляция, клеточная дифференцировка, регуляция экспрессии генов и регуляция клеточного цикла [1, 2]. Однако из многих тысяч выявленных генов и транскриптовнекодирующей ДНК функционально охарактеризовано незначительное количество. Функциональная значимость некодирующей РНК особенно очевидна для класса длинных некодирующих (lncRNA) и микроРНК (miRNAs). Полиморфизмы и нарушение профиля экспрессии генов miRNA, в последние годы связывают с различными мультифакторными заболеваниями. Выявлено, что многие miRNA кодируются в межгенных участках ДНК и обнаружены в интронах. Показана важная роль miRNAs в эпигенетических процессах, которые контролируют дифференциацию и развитие живого организма.

Поиск мотивов гомологичных микро РНК осуществлялся в *цис*-регуляторных районах изучаемых генов с помощью биоинформационного пакета MEME Suite. Известные miRNA взяли согласно данным баз данных miRBase (<http://mirbase.org/>). Нуклеотидные последовательности *цис*-регуляторных районов и интронов генов взяты нами из электронных баз данных NCBI (<http://www.ncbi.nlm.nih.gov/>) с помощью наборов скрипта, IFITCH, которые были разработаны специально для автоматизированного получения информации от последовательностей NCBI. Анализ *in silico* проводили при помощи автоматического поиска сайтов связывания с использованием программы «mscanner» [3]. На момент проведения биоинформационного поиска в базе данных MiRBase было зарегистрировано 28645 микроРНК (<http://www.mirbase.org/>). В таблице представлены гены липидного обмена и ферментов антиоксидантной защиты, а также длины окрестностей некодирующей ДНК, локализованной перед этими генами и после них в геноме человека. Таким образом, общая протяжённость исследуемых последовательностей составила 3 098 209 нуклеотидов. При этом было изучено 788 404 нуклеотидов перед генами и 631 332 – после генов. Мы отметили, что наименьшее число нуклеотидов в межгенном пространстве было локализовано перед геном *TNF* – 1 243 нуклеотида, а самое большое перед геном *LIPC* – 209 736. Самое малое число нуклеотидов было после гена *APOC3* – 3 046 нуклеотида, а самое большое после гена *PPARGCIA* – 210 553 нуклеотидов.

Далее в работе использовали программу dotolog [4], которая предназначена для построения точечной матрицы гомологий (дотплота) для двух и более последовательностей. Данная программа позволяет проводить множественное выравнивание последовательностей. Отличительной особенностью программы также является поддержка отображения аннотаций к последовательностям и возможность добавления и редактирования пользовательских аннотаций, полученных по результатам визуального анализа. Интерфейс программы реализует концепцию «как GoogleMaps», предоставляя удобство ориентирования по последовательностям. При отображении широко используется цветовое кодирование, в частности, мотивы (непрерывные последовательности совпадений) подсвечиваются зеленым цветом. Пользовательские аннотации подсвечиваются цветом, сгенерированным по уникальному идентификатору элемента. Реализована возможность фильтрации отображаемых элементов. Таким образом, предоставляемый программой dotolog функционал позволяет эффективно выявлять исследуемые элементы генома.

**Длины окрестностей некодирующей ДНК, локализованной
перед и после исследуемых генов в геноме человека**

Общепринятое название	Длина окрестности	Длина окрестности перед геном	Длина гена	Длина окрестности после гена
<i>FTO</i>	510505	50000	410505	50000
<i>LIPC</i>	485370	209736	158199	117435
<i>LPL</i>	97138	18949	28189	50000
<i>TNF</i>	11856	1243	2770	7843
<i>EDNI</i>	107420	44429	40732	22259
<i>PPARGC1A</i>	778172	47363	680809	210553
<i>PPARGC1B</i>	185490	7880	127610	50000
<i>PONI</i>	112197	35166	27031	50000
<i>SOD1</i>	96856	50000	9310	37546
<i>CAT</i>	85203	2067	33136	50000
<i>SOD2</i>	149435	15969	83466	50000
<i>NOS3</i>	87367	50000	23544	13823
<i>GPX4</i>	68959	16095	2864	50000

Результаты наших исследований показали, что наибольшее количество генов микро РНК локализовано в интронах генов *PPARGC1A* и *PPARGC1B*, а также в гене *FTO* (35 молекул микро РНК). Наименьшее количество сайтов связывания микро РНК локализовано в *цис*-регуляторных последовательностях гена *APOC1*. Согласно нашим данным биоинформационного анализа, в окрестностях гена *FTO* локализованы следующие мотивы: hsa-miR-619-5p (16), hsa-miR-5585-3p (2), hsa-miR-1285-3p (2), hsa-miR-3613-5p (1), hsa-miR-6788-5p (1), hsa-miR-3134 (1), hsa-miR-1256 (1), hsa-miR-320d (1), hsa-miR-6845-3p (1), hsa-miR-3133 (1), hsa-miR-649 (1), hsa-miR-1261 (1), hsa-miR-4496 (1), hsa-miR-4433b-5p (1), hsa-miR-297 (1), hsa-miR-183-5p (1), hsa-miR-3162-3p (1), hsa-miR-7110-5p (1), hsa-miR-4445-5p (1), hsa-miR-877-3p (1), hsa-miR-205-3p (1), hsa-miR-708-5p (1), hsa-miR-3140-5p (1), hsa-miR-2052 (1), hsa-miR-3661 (1), hsa-miR-1972 (1).

Типы зрелых мотивов микро РНК распределенных вокруг гена *PPARGC1B* hsa-miR-619-5p (5), hsa-miR-5585-3p (3), hsa-miR-1303 (2), hsa-miR-648 (1), hsa-miR-6125 (1), hsa-miR-552-3p (1), hsa-miR-2392 (1), hsa-miR-3074-5p (1), hsa-miR-6747-3p (1), hsa-miR-1271-3p (1), hsa-miR-1285-5p (1), hsa-miR-649 (1), hsa-miR-1304-5p (1), hsa-miR-4266 (1), hsa-miR-146a-3p (1), hsa-miR-6844 (1), hsa-miR-4261 (1), hsa-miR-378i (1), hsa-miR-527 (1), hsa-miR-378a-3p (1), hsa-miR-1285-3p (1), hsa-miR-345-5p (1), hsa-miR-133a-3p (1), hsa-miR-378g (1), hsa-miR-378a-5p (1), hsa-miR-378d (1)

Типы зрелых мотивов микро РНК, локализованных в окрестности гена *APOC1* – hsa-miR-619-5p (3), hsa-miR-629-3p (1), hsa-miR-4646-3p (1). В скобках указано количество копий.

Как видно из представленных результатов, наиболее часто в окрестностях изученных генов встречался мотив hsa-miR-619.

Нами отмечено, что количество локализованных сайтов связывания с микро РНК коррелирует с числом тканей в которых экспрессируются данные гены. Так ген *FTO* во взрослом организме экспрессируется в 23 из 23 исследованных тканях в проекте транскриптома человека, результаты которого представлены в NCBI (<https://www.ncbi.nlm.nih.gov/gene/79068>). Ген *APOC1* экспрессируется только в четырех тканях из 23 исследованных (<https://www.ncbi.nlm.nih.gov/gene/341>).

ЛИТЕРАТУРА

1. Mattick J. // Bioessays. 2003. Vol. 25 (10). P. 930–939.
2. Maracaja-Coutinho V., Paschoal A.R., Caris-Maldonado J.C. et al. // Methods Mol. Biol. 2019. Vol. 1912. P. 251–285. Doi: 10.1007/978-1-4939-8982-9_10.
3. Shkurat T., Romanov D., Pshenichnyy E. et al. // American Journal of Applied Sciences. 2015. Vol. 12, №. 1. P. 1.
4. Романов Д.Е., Шкурят Т.П. Программа ЭВМ «Dotolog: программа для автоматизации визуального анализа дотплот-изображений нуклеотидных последовательностей ДНК» №2016663454 от 7 декабря 2016.

Исследование выполнено в рамках базовой части госзадания Минобрнауки РФ по теме: «Исследования функциональной роли генетических полиморфизмов и микро РНК в геноме человека и животных», проект № 6.6762.2017 БЧ.