

УДК 577.29

А.Ю. Пыркова \*, А.Т. Иващенко, О.А. Берилло

Казахского национального университета им.аль-Фараби, Республика Казахстан, г. Алматы

\* E-mail: Anna.Pyrkova@kaznu.kz

### Параллелизация сканирования генов на кластерной платформе: программа MirTarget

В настоящее время в геноме человека известно более 2500 miRNA, необходимо для каждой miRNA найти гены мишени среди 30 тысяч генов человека. Большой объём вычислений требует создания программы, позволяющей обрабатывать эти огромные массивы данных. В представленной статье предложено решение задачи сканирования генов с целью предсказания сайтов связывания miRNA с матричной РНК (mRNA). Разработанная программа была использована для проведения исследований при поиске сайтов связывания miRNA с матричной РНК (mRNA). В ходе проведённого исследования авторами были получены следующие результаты:

- разработан и проанализирован алгоритм сканирования генов с miRNA с учётом одного разрыва в miRNA и максимальной (в процентном соотношении) свободной энергии при совпадении miRNA и участка гена на основе свойств комплементарности;
- построенный алгоритм сканирования генов с miRNA распараллелен на кластерной платформе с использованием средств MPJ (Java MPI);
- проведена оценка эффективности работы распараллеленного алгоритма на кластерной платформе с последовательным алгоритмом при обработке больших объёмов данных.

**Ключевые слова:** miRNA и mRNA, алгоритм, Java MPI, кластерная платформа, распараллеленный алгоритм, сканирование, комплементарность.

A.Yu. Pyrkova, A.T. Ivashchenko, O.A. Berillo

#### Parallelization of scanning genes on the cluster platform: MirTarget program

Today in the human genome of the person more than 2500 miRNA are known, it is necessary for each miRNA to find target genes among 30 thousands genes of the human. Large amount of calculations demands creation of the program, allowing processing these huge data files. In presented article the solution of the problem of gene scanning for the purpose of prediction of sites of binding of miRNA with matrix RNA (mRNA) is proposed. The developed program was used for performing researches by search of sites of binding of miRNA from matrix RNA (mRNA). During the conducted research by authors the following results were received:

- the algorithm of scanning of genes with miRNA with a glance one gap in miRNA and maximum (in a percentage ratio) free energy is developed and analysed at coincidence of miRNA and a gene site on the basis of complementarity properties;
- the constructed algorithm of scanning of genes with miRNA is parallelized on the cluster platform with use of MPJ tools (Java MPI);
- the assessment of overall performance of the parallelized algorithm on the cluster platform with consecutive algorithm is performed when processing large volumes of data.

**Key words:** miRNA and mRNA, algorithm, Java MPI, the cluster platform, the parallelized algorithm, scanning, complementarity.

А.Ю. Пыркова, А.Т. Иващенко, О.А. Берилло

#### Кластерлі платформада гендерді сканерлеуді параллелдеу: MirTarget программасы

Осы уақытта адам геномында 2500 артық miRNA белгілі, 30 мың адам геномының арасынан әрбір miRNA үшін нысана генін табу қажет. Есептеулердің үлкен көлемі, ауқымды деректер массивін өңдеуге рұқсат ететін программа құруды талап етеді. Ұсынылған мақалада РНК (mRNA) матрицасымен miRNA байланысын гендерді сканерлеу есебінің шешімі сайттарды болжау мақсатымен ұсынылған. Құрылған программа РНК (mRNA) матрицасы miRNA байланыс сайттарын зерттеу жүргізу үшін іздеу кезінде қолданылған. Жұмыс барысында жүргізілген зерттеулер нәтижесінде авторлар төмендегідей нәтижелерге қол жекізді:

- miRNA гендерімен miRNA-дағы бір бөліктің үзілуі есебінде және максималды (пайыздық қатынаста) бос энергия miRNA-мен сәйкес келгенде және ген телімдерінің комплементарлық қасиеттердің негізінде сканерлеу алгоритмдері құрылған және талданған;
- MPJ (Java MPI) құралдарын қолдану арқылы кластерлі платформада параллелденген miRNA гендерін сканерлеу алгоритмі құрылған;
- үлкен көлемді деректерді өңдеу үшін кластерлі платформада тізбектелген алгоритмдерді параллелдеу жұмысының тиімділеу бағасын жүргізу.

**Түйін сөздер:** miRNA және mRNA, алгоритм, Java MPI, кластерлі платформа, параллелденген алгоритм, сканерлеу, комплементарлық.

## Введение

После открытия важной роли микроРНК (miRNA) в регуляции экспрессии генов возникла проблема предсказания сайтов связывания miRNA с матричной РНК (mRNA). Было создано несколько программ, которые предсказывали сайты связывания с miRNA. Однако многие из них имели необоснованные ограничения при поиске сайтов связывания. Ранее утверждалось, что сайты связывания локализованы исключительно в 3'UTR. Позже было установлено, что сайты связывания локализованы в 5'UTR и CDS. Другие программы были основаны на выявлении сайтов связывания с обязательным требованием иметь комплементарные взаимодействия гуанина (G) и аденина (A) в участке «seed», который соответствует 5'-концу miRNA. Многие такие программы предсказывали большое число ложно-положительных сайтов и не позволяли выявлять сайты связывания, расположенные в 5'UTR и CDS. По этой и другим причинам неточно устанавливалось начало сайтов связывания и некорректно составлялись схемы взаимодействия miRNA с mRNA. В настоящее время, в геноме человека известно более 2500 miRNA и необходимо для каждой из них найти гены мишени среди 30 тысяч генов человека. Большой объем вычислений требует создания программы, позволяющей обрабатывать эти огромные массивы данных. Нами создана программа MirTarget, которая не имеет приведенных выше недостатков и с большой достоверностью находит сайты связывания miRNA с mRNA.

## Постановка задачи сканирования генов

Сканирование генов [1, 2] – это процесс последовательного сопоставления участков гена с miRNA с возможностью внесения одного пробела в miRNA в позициях с 3 по n-2, где n – длина miRNA в нуклеотидах. При этом происходит оценка все возможных сопоставлений на одном участке mRNA с miRNA, которая определяется согласно величине свободной энергии сравниваемых последовательностей. Лучшим считается тот вариант, который ближе (в процентном соотношении) по свободной энергии для сов-

падения miRNA и участка гена на основе комплементарности. Сканирование генома позволяет выявить сотни возможных мишеней [3, 4] для терапии различных заболеваний. Такое сканирование важно потому, что знание взаимодействующих генов позволит определить то, за что отвечает тот или иной белок и, соответственно, какие внутриклеточные процессы могут нарушаться при данном заболевании.

### Алгоритм сканирования генов на кластере

1 шаг. Проверка наличия атрибутов в командной строке.

`mpjrun.bat -np количество_параллельных_процессов /путь/к/Base минимальный_процент_сходимости папка_с_генами папка_с_miRNA имя_файла_результата_mres имя_файла_результата_xls`

Например, `mpjrun.bat -np 10 Base 80 gene mir ResultsFull ResultsBrief`

2 шаг. Главный процесс (MASTER) вызывает функцию, которая формирует массив `geneName` файлов, содержащих гены.

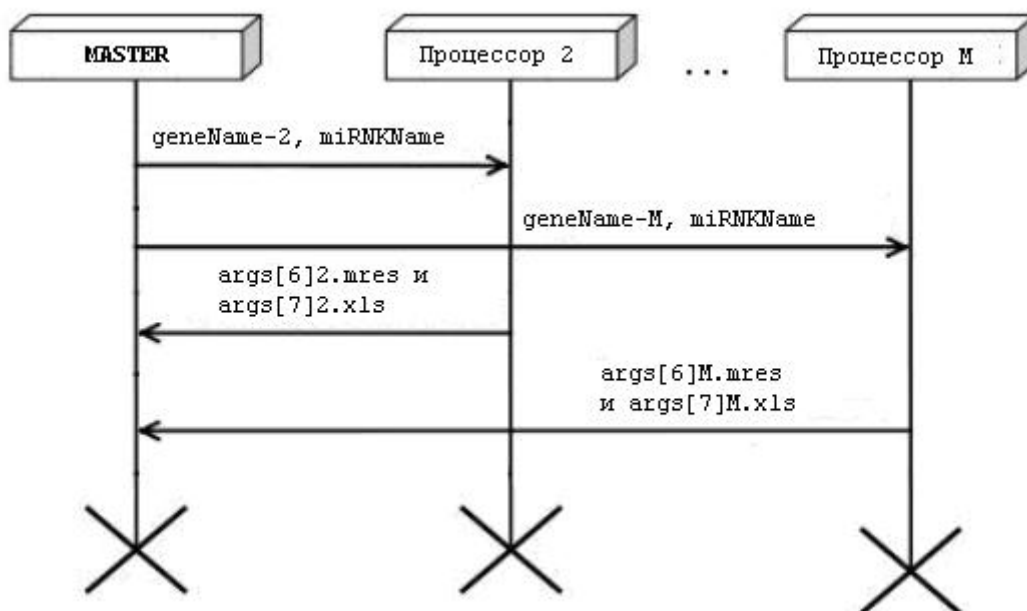


Рисунок 1 – Протокол обмена сообщениями между MPJ-процессами

3 шаг. Главный процесс (MASTER) вызывает функцию, которая формирует два массива по miRNA:

- 1) массив `miRNKName` имён miRNA;
- 2) массив `miRNK` самих miRNA последовательностей. При этом происходит проверка формата файла miRNA.

```
>let-7a-2-3p          MIRLET7A2-ex-cod  11
CUGUACAGCCUCCUAGCUUCC
```

Рисунок 2 – Структура файла с miRNA (.mir)

4 шаг. Главным и параллельными процессами создаются файлы с именами `args[6](task_id).mres` и `args[7](task_id).xls` для записи результата.

5 шаг. Главный процесс (MASTER) разделяет гены между параллельными и процессами и рассылает определённую часть массива geneName и массивы miRNKName и miRNK. Главный процесс также получает часть генов для сканирования.

6 шаг. Процессы считывают последовательно гены из файлов массива geneName и проверяют формат файла с геном.

```
>SC5DL-v-2 | 1-221 | 222-1121 | 1122-7020
GGGCCGGGGGGCGATGGAATAAAAAGAAAGATGGAGAGACTTCAGCGCCTGGGACTCGGGTGGGCGAGGCGGAAGGTGTCTC
CGCAGCACGGCTTTTCTCCGCGCCGCGGTTGGTTAGCGAGTGCSTCTGGGTGCTAGGCCTTGGGCGGATGGTAGGATCG
CGGTAGCATAACGGATCCGAGTCCSTGCGCCGAGTGAGAGGAGAGGCTGGCAGGGGCTAAGTGATGGATCTTGTACTCCGTG
TTGCAGATTACTATTTTTTTACACCATACGTGTATCCAGCCACATGGCCAGAAGATGACATCTTCCGACAAGCTATTAGT
CTTCTGATTGTAACAATAATGTTACACGCTTACATCCTTTATTTCTTCTGTGCAACACTGAGCTATTATTTGTCTTCGATCA
TGCATTAATGAAAACATCCACAATT.....ACATTTGTATACAA*
```

Рисунок 3 – Структура файла с геном (.gene)

7 шаг. Процессы сканируют выделенные им гены последовательно с miRNA, допуская один разрыв в miRNA в позициях с 3 по n-2, где n – длина miRNA и выбирая тот вариант, который ближе (в процентном соотношении) по свободной энергии для совпадения miRNA и участка гена на основе комплементарности.

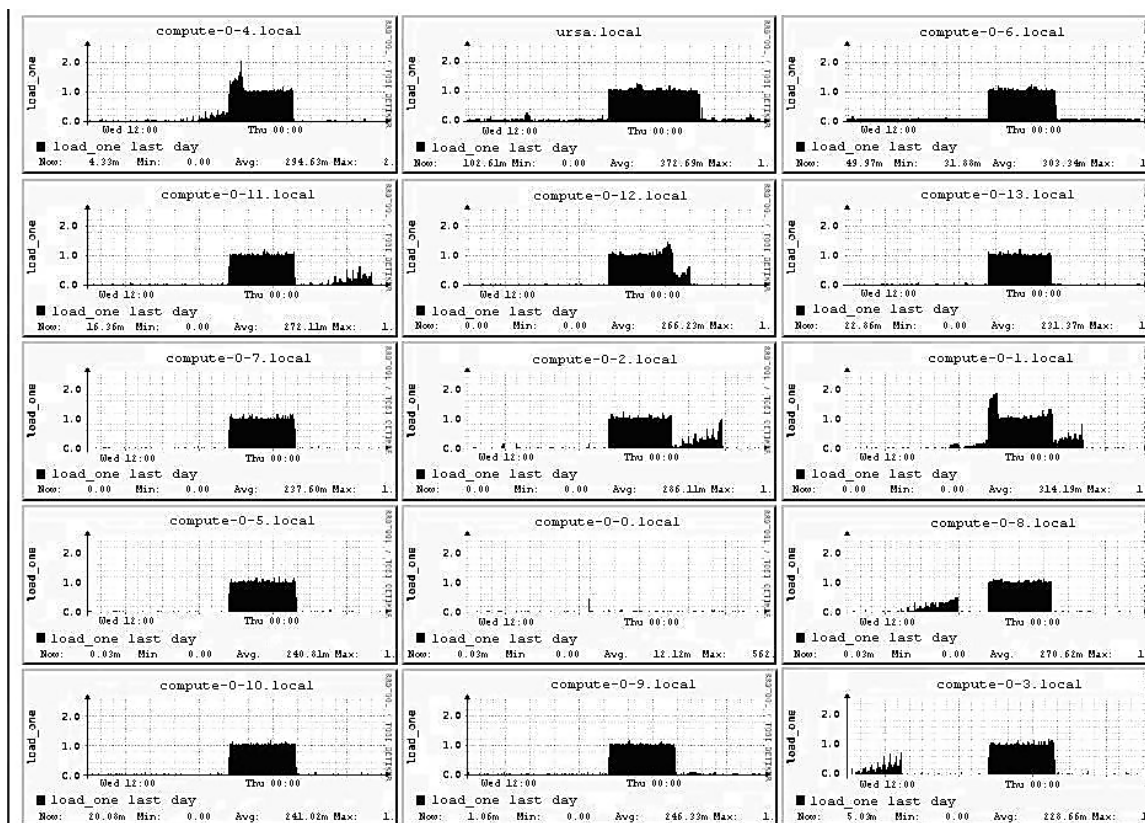


Рисунок 4 – Загруженность кластера (Для базы, содержащей 13016 генов и 100 miRNA, время работы составило 7 часов 27 минут 6 секунд)

8 шаг. Процессы формируют свои файлы *args[6]<task\_id>.mres* и *args[7]<task\_id>.xls* и по мере завершения обработки своей части генов отправляют сообщения о завершении главному процессу.

9 шаг. Главный процесс (MASTER) получает сообщения от параллельных процессов, копирует информацию из соответствующих файлов  $args[6]\langle task\_id \rangle.mres$  и  $args[7]\langle task\_id \rangle.xls$  в файлы  $args[6].mres$  и  $args[7].xls$ , а файлы  $args[6]\langle task\_id \rangle.mres$  и  $args[7]\langle task\_id \rangle.xls$  удаляет.

## Результаты и обсуждение

Программа MirTarget имеет достоинства, которых нет в известных программах предсказания сайтов связывания miRNA с mRNA. В литературе имеется много сведений о величине свободной энергии водородной связи между нуклеотидами в водном растворе [5]. Однако существует большой разброс величин свободной энергии этой связи и трудно отдать предпочтение определенным данным [6, 7]. Важно знать относительные отношения свободной энергии водородной связи между нуклеотидами, поскольку они необходимы при образовании RNA вторичной и третичной структур. Анализ свободной энергии водородной связи, возникающей между нуклеотидами при внутримолекулярном взаимодействии mRNA при образовании её вторичной структуры показал, что между нуклеотидами G-C образуется три, между A-U – две и между G-U и A-C – по одной водородной связи. Отношение свободной энергии водородной связи в парах G-C и A-U примерно соответствует отношению сил их взаимодействия 3:2 (0.188 nNewton и 0.125 nNewton) [8]. Величина свободной энергии одной водородной связи между нуклеотидами изменяется в диапазоне от -0.7 до -1.6 ккал/моль [9]. В программе MirTarget свободную энергию взаимодействия нуклеотидов за счет водородных связей считали равной 6.368 кДж/моль и 4.246 кДж/моль для пар G-C и A-U, и 2.123 кДж/моль для пар G-U и A-C, соответственно.

Расстояние между нуклеотидами пар G-C и A-U составляет 1.03 nm, между нуклеотидами пары G-U оно равно 1.02 nm и между нуклеотидами пары A-C равно 1.04 nm [10]. Следовательно, образование водородных связей между этими парами нуклеотидов позволяет двухцепочечной структуре mRNA иметь спиральную форму аналогичную регулярной структуре ДНК. Такая структура mRNA кроме водородных связей стабилизируется стэкинг-взаимодействиями между азотистыми основаниями [11]. Между нуклеотидами пар пурин-пурин и пиримидин-пиримидин расстояния существенно отличаются от 1.03 nm: расстояния A-A, G-A и G-G равны 1.23 nm, 1.25 nm и 1.25 nm, соответственно. В парах пиримидин-пиримидин расстояния между нуклеотидами тоже существенно отличаются от 1.03 nm: расстояния C-C, U-U и U-C равны 0.85 nm, 0.81 nm и 1.18 nm, соответственно. Следовательно, в таких парах водородные связи не образуются, и эти пары будут нарушать регулярную структуру двухцепочечной miRNA с mRNA уменьшая стабильность комплекса RISC (RNA-induced silencing complex). Поэтому, в программе такие пары водородной связи не учитывались.

При выравнивании нуклеотидных последовательностей miRNA с mRNA мы допускаем наличие только одного пропуска на miRNA (отсутствие комплементарной пары водородной связи), что позволяет учитывать сайты связывания на mRNA длиннее miRNA на один нуклеотид. В этом случае нарушается регулярная структура спирали и возникает ее изгиб (bulg).

Gene	miRNA	Number	Position	Where	Energy (kJ/m)	Score	Length
>SUPT3H-v-3	>miR-1 (C20...)	1	2000	3'UTR	-87.0443	81.99	22
5' - AUCCACCACUCCUGCACAUGCCA - 3'							
3' - UAUGUA-UGAAGAAUUGUAAGGU - 5'							
>SUPT3H-v-3	>miR-9-3p(...)	2	1904	3'UTR	-87.0451	80.39	22
5' - ACUUCUGCCCAUCGCCUUCU - 3'							
3' - UGAAAAGCCAAUAGAUCCAAA-UA - 5'							
>SUPT3H-v-3	>miR-24-2-5p(ig)	3	878	CDS	-93.4128	80.00	22
5' - UUAUGGAUUCAGCUCAAUUGCA - 3'							
3' - GACACA-AAGUCGAGUCAUCCGU - 5'							
>SUPT3H-v-3	>let-7c(...)	4	1215	CDS	-91.2866	81.12	22
5' - GGCCAUUCGACGCCUACGCCACA - 3'							
3' - UUGCUAUGUUG-CAUGAUGGAGU - 5'							
>SUPT3H-v-3	>let-7c-5p(LIN...)	5	1215	CDS	-91.2866	81.12	22
5' - GGCCAUUCGACGCCUACGCCACA - 3'							
3' - UUGCUAUGUUG-CAUGAUGGAGU - 5'							
>SUPT3H-v-3	>miR-15b-5p(SMC...)	6	1308	3'UTR	-93.4120	83.01	22
5' - UGACAAACUGUGAUGUGCCUGCAA - 3'							
3' - ACAUUUGGUACUACACG-ACGAU - 5'							
>ARMC8-v-1	>miR-15a-5p(DLE...)	7	4319	3'UTR	-93.4145	83.01	22
5' - CACAAAACACUAUAGUGUUGGUA - 3'							
3' - GUGUUUGGUAUA-CACGACGAU - 5'							

Рисунок 5 – Результирующий файл (.mres)

	A	B	C	D	E	F	G	H
	Gene	miRNA	Number	Position	Where	Energy (kJ/m)	Score	Length
1	>SUPT3H-v-3	>miR-1(C20orf166-5UTR-noncod)	1	2000	3'UTR	-87.0443	81.99	22
2	>SUPT3H-v-3	>miR-9-3p(C1orf61-MIR9-inex-noncod-cod)	2	1904	3'UTR	-87.0451	80.39	22
3	>SUPT3H-v-3	>miR-24-2-5p(ig)	3	878	CDS	-93.4128	80.00	22
4	>SUPT3H-v-3	>let-7c(C21orf34-MIRLET7C-LINC00478-inex-noncod-cod)	4	1215	CDS	-91.2866	81.12	22
5	>SUPT3H-v-3	>let-7c-5p(LINC00478-MIRLET7C-in-ex)	5	1215	CDS	-91.2866	81.12	22
6	>SUPT3H-v-3	>miR-15b-5p(SMC4-RP11-in-noncod)	6	1308	3'UTR	-93.4120	83.01	22
7	>ARMC8-v-1	>miR-15a-5p(DLEU2-DLEU2-in-ex-noncod-cod)	7	4319	3'UTR	-93.4145	83.01	22
8	>DNAJC3-v-1	>miR-19b-2-5p(MIR19B2-ex-cod)	8	673	CDS	-89.1662	80.76	22
9	>DNAJC3-v-1	>miR-30c-1-3p(NFYC-in-noncod)	9	1534	CDS	-97.6543	82.14	22
10	>DNAJC3-v-1	>miR-23b-3p(C9orf3-in-noncod)	10	120	CDS	-89.1612	80.76	21
11	>PIGF-v-2	>miR-21-3p(ig)	11	48	5'UTR	-95.5307	83.33	21
12	>SYVN1-v-1	>miR-23a-5p(ig)	12	925	CDS	-97.6551	80.70	22
13	>SYVN1-v-1	>miR-23a-5p(ig)	13	1784	CDS	-101.8991	84.20	22
14	>SYVN1-v-1	>miR-23a-5p(ig)	14	1785	CDS	-106.1448	87.71	22
15	>SYVN1-v-1	>let-7i-5p(RP11-MIRLET7I-C12orf61-in-ex-5UTR-noncod-cod-noncod)	15	148	CDS	-91.2874	81.12	22
16	>ICAM5-v-1	>miR-21-5p(ig)	16	2428	CDS	-93.4111	84.61	22
17	>ICAM5-v-1	>miR-31-5p(RP11-MIR31HG-in-noncod)	17	62	5'UTR	-91.2858	81.12	21
18	>ICAM5-v-1	>miR-15a-5p(DLEU2-DLEU2-in-ex-noncod-cod)	18	2228	CDS	-91.2874	81.12	22
19	>ICAM5-v-1	>let-7g-3p(WDR82-MIRLET7G-in-noncod)	19	954	CDS	-95.5315	81.81	21
20	>ICAM5-v-1	>let-7g-3p(WDR82-MIRLET7G-in-noncod)	20	2701	CDS	-99.7763	85.45	21
21	>CMTM4-v-2	>miR-10b-3p(HOXD3-HOXD3-in-3UTR-noncod-cod)	21	3114	3'UTR	-91.2883	81.13	22
22	>CMTM4-v-2	>miR-20a-3p(MIR17HG-MIR17HG-inex-noncod-cod)	22	2000	3'UTR	-89.1663	80.76	22
23	>CMTM4-v-2	>let-7i-3p(RP11-MIRLET7I-C12orf61-in-ex-5UTR-noncod-cod-noncod)	23	169	5'UTR	-97.6526	80.69	22
24	>GSG2-v-1	>miR-31-5p(RP11-MIR31HG-in-noncod)	24	181	CDS	-91.2858	81.12	21
25	>GSG2-v-1	>miR-31-5p(RP11-MIR31HG-in-noncod)	25	2386	CDS	-91.2875	81.13	21
26	>GSG2-v-1	>let-7a-2-3p(MIRLET7A2-ex-cod)	26	110	CDS	-95.5288	80.35	22
27	>GSG2-v-1	>miR-27a-3p(ig)	27	1707	CDS	-91.2875	81.13	21
28	>GSG2-v-1	>miR-19a-3p(MIR17HG-MIR17HG-exin-cod-noncod)	28	2384	CDS	-93.4111	81.47	23
29	>GSG2-v-1	>miR-27b-3p(C9orf3-in-noncod)	29	1707	CDS	-91.2875	82.69	21

Рисунок 6 – Результирующий файл (.xls)

Свободная энергия связывания miRNA с mRNA такой структуры меньше, чем в альтернативном варианте. Программа определяет свободную энергию гибридизации ( $\Delta G$ , 100 kJ/mole) miRNA с mRNA и схемы их взаимодействий, расчет отношений  $\Delta G/\Delta G_m$ , уровни достоверности ( $p$ ) и области mRNA, где расположен сайт (5'UTR, CDS или 3'UTR), начиная с первого нуклеотида 5'UTR.  $\Delta G_m$  равно свободной энергии связывания miRNA с полностью комплементарным ей участком нуклеотидной последовательности mRNA. Уровень достоверности ( $p$ ) определялся на основе величины  $\Delta G$  и её стандартного отклонения. Программа выдает схему взаимодействия miRNA с mRNA, позицию сайта в 5'UTR, CDS или 3'UTR, свободную энергию взаимодействия miRNA с mRNA, и её относительную величину от максимальной энергии связывания miRNA. В программе задается пороговая величина этого отношения, что позволяет не учитывать сайты со слабой свободной энергией связывания.

## Литература

- [1] *Lesk Arthur M.* Introduction to Bioinformatics. - Oxford: Oxford University Press, 2002. - 255 p.
- [2] *Jones Neil C., Pevzner Pavel A.* An Introduction to Bioinformatics Algorithms. - Massachusetts: Massachusetts Institute of Technology Press, 2004. - 435 p.
- [3] *Jonathan M. Keith* Methods in Molecular Biology. Bioinformatics: in 2 vols. - New York: Humana Press, 2008. - V. 2. - 502 p.
- [4] Bioinformatics and Biological Computing [Electronic resource]. - 2012. - URL: [http://bip.weizmann.ac.il/toolbox/overview/software\\_avail.html](http://bip.weizmann.ac.il/toolbox/overview/software_avail.html) (дата обращения: 07.09.2012)
- [5] *Guckian K.M., Schweitzer B.A., Ren RXF, Sheils C.J., Paris P.L., et al* Experimental measurement of aromatic stacking in the context of duplex DNA. // J.Am.Chem.Soc. - 1996. - 118: 8182–83.
- [6] *Turner D.H., Sugimoto N., Kierzek R., Dreiker S.D.* Free energy increments for hydrogen-bonds in nucleic acid base pairs. // J. Am. Chem. Soc. - 1987. - 109: 3783–85.
- [7] *Sugimoto N., Kierzek R., Turner D.H.* Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid. // Biochemistry. - 1987. - 14: 4554–58.
- [8] *Boland T., Ratner B.D.* Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy. // Proc. Natl. Acad. Sci. USA. - 1995. - V. 92. - Pp. 5297–5301.
- [9] *Kool E.T.* Hydrogen bonding, base stacking, and steric effects in DNA replication. // Annu. Rev. Biophys. Biomol. Struct. - 2001. - 30: 1–22.
- [10] *Leontis N.B., Stombaugh J., Westhof E.* The non-Watson–Crick base pairs and their associated isostericity matrices. // Nucleic Acids Res. - 2002. - 30(16): 3497–3531.
- [11] *Richard A. Friedman, Honig B.A.* Free Energy Analysis of Nucleic Acid Base Stacking in Aqueous Solution. // Biophysical Journal. - 1995. - V. 69. - Pp. 1528–1535.

## References

- [1] *Lesk Arthur M.* Introduction to Bioinformatics. - Oxford: Oxford University Press, 2002. - 255 p.
- [2] *Jones Neil C., Pevzner Pavel A.* An Introduction to Bioinformatics Algorithms. - Massachusetts: Massachusetts Institute of Technology Press, 2004. - 435 p.
- [3] *Jonathan M. Keith* Methods in Molecular Biology. Bioinformatics: in 2 vols. - New York: Humana Press, 2008. - V. 2. - 502 p.

- 
- [4] Bioinformatics and Biological Computing [Electronic resource]. - 2012. - URL: [http : //bip.weizmann.ac.il/toolbox/overview/software\\_avail.html](http://bip.weizmann.ac.il/toolbox/overview/software_avail.html) (дата обращения: 07.09.2012)
- [5] *Guckian K.M, Schweitzer B.A, Ren RXF, Sheils C.J, Paris P.L, et al* Experimental measurement of aromatic stacking in the context of duplex DNA. // *J.Am.Chem.Soc.* - 1996. - 118: 8182–83.
- [6] *Turner D.H, Sugimoto N, Kierzek R, Dreiker S.D.* Free energy increments for hydrogen-bonds in nucleic acid base pairs. // *J. Am. Chem. Soc.* - 1987. - 109: 3783–85.
- [7] *Sugimoto N, Kierzek R, Turner D.H.* Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid. // *Biochemistry.* - 1987. - 14: 4554–58.
- [8] *Boland T., Ratner B.D.* Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy. // *Proc. Natl. Acad. Sci. USA.* - 1995. - V. 92. - Pp. 5297-5301.
- [9] *Kool E.T.* Hydrogen bonding, base stacking, and steric effects in DNA replication. // *Annu. Rev. Biophys. Biomol. Struct.* - 2001. - 30: 1–22.
- [10] *Leontis N.B., Stombaugh J., Westhof E.* The non-Watson–Crick base pairs and their associated isostericity matrices. // *Nucleic Acids Res.* - 2002. - 30(16): 3497–3531.
- [11] *Richard A. Friedman, Honig B.A.* Free Energy Analysis of Nucleic Acid Base Stacking in Aqueous Solution. // *Biophysical Journal.* - 1995. - V. 69. - Pp. 1528-1535.