

Разработка компьютерных инструментов для поиска генов-мишеней микроРНК

Радочин П.А.¹, Воронина А.А.², Орёл Д.Ю.², Смирнова О.А.¹, Голубчиков Д.О.¹,
Мак Д.В.¹, Воронова С.С.¹, Волков И.А.¹, Савина Е.А.^{1,3}, Орлов Ю.Л.^{1,4*}

¹ Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия

² Санкт-Петербургский политехнический университет им. Петра Великого, Санкт-Петербург, Россия

³ Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия

⁴ Институт цитологии и генетики СО РАН, Новосибирск, Россия

* orlov@bionet.nsc.ru

Ключевые слова: биоинформатика; микроРНК; базы данных; генные сети; комплексные заболевания; генные онтологии; образование

Мотивация и цель: Исследование молекулярных механизмов комплексных заболеваний является важной задачей биомедицины, требующей интеграции данных, развития новых биоинформационных инструментов [1]. микроРНК – это класс некодирующих молекул РНК, которые регулируют трансляцию и деградацию своих МРНК-мишеней посредством взаимодействия, что делает изучение их функций при заболеваниях необходимым для выяснения их патогенетических механизмов [2]. микроРНК играют важную роль в регуляции экспрессии генов. Появляются все новые литературные данные о роли микроРНК в прогрессии опухолей, развитии ментальных расстройств, заболеваниях печени, хронических заболеваниях [1].

Представлен научный проект исследования генов-мишеней микроРНК в рамках Цифровой кафедры Сеченовского Университета в 2023/2024 учебном году [3]. Освоение использования существующих онлайн-программ для компьютерного поиска генов-мишеней микроРНК в геноме и разработка собственных программ для предсказания участков связывания микроРНК в генах.

Решаемая проблема в рамках представленного проекта – обучение по проблемам использования инструментов биоинформатики для анализа структуры РНК и микроРНК, механизмам связывания РНК с транскриптами генов человека.

Методы и алгоритмы: Собраны имеющиеся программные инструменты по поиску генов-мишеней микроРНК, подготовлен конвейер обработки данных. Использовались обзоры (Список программ и баз –

<https://www.tamirna.com/micrnas-web-based-tools/>), базы данных MIRDB (<https://mirdb.org/>), TargetScan (https://www.targetscan.org/vert_80/), MiRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php).

микроРНК – это РНК, состоящая из ~22 нуклеотидов, обработанных из шпильчатых структур РНК. микроРНК слишком короткие, что создает технические трудности их поиска. Определение генов мишеней по частичной комплементарности также осложняется возможностью несовпадения нуклеотидов [4].

Составление списков генов и микроРНК, связанных с заболеваниями, выполнялось на основе запросов к базам данных OMIM (Online Mendelian Inheritance in Man), GeneCards, MalaCards. Анализ генных онтологий для списка

генов был выполнен с помощью ресурса PANTHER (Protein ANalysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) и DAVID. Для реконструкции сети использовались инструменты STRING-DB, приложения Cytoscape. С помощью онлайн-инструментов биоинформатики был проанализирован актуальный массив данных.

Результаты: TargetScan предсказывает биологические мишени микроРНК путем поиска консервативных 8-, 7- и 6-мерных сайтов, которые соответствуют начальной области каждой микроРНК. В качестве вариантов также предлагаются предсказания только со слабоконсервативными сайтами и предсказания с несохраняемыми микроРНК [5]. Также идентифицированы сайты с несоответствиями в начальной области, которые компенсируются консервативным 3'-спариванием. У млекопитающих прогнозы ранжируются на основе прогнозируемой эффективности нацеливания, рассчитанной с использованием биохимической модели репрессии, опосредованной микроРНК. Выявлены общие черты, связанные со связыванием микроРНК и снижением регуляции мишеней, которые были использованы для прогнозирования мишеней для микроРНК с помощью методов машинного обучения.

Для поиска генов-мишеней терапии необходима реконструкция генной сети – комплекса взаимодействующих макромолекул, включая микроРНК, функционально связанных с заболеванием. Дальнейший анализ предполагает кластеризацию генов в сети, выявление ключевых генов, обладающих наибольшим числом контактов в сети, нахождение категорий генных онтологий.

Представлен конвейер использования программных инструментов для поиска генов-мишеней микроРНК для медицины. Возможно дальнейшее использование таких программ в качестве учебных ресурсов по биоинформатике для медиков.

Работы были представлены на студенческой конференции в Сеченовском Университете, подготовлены студенческие дипломы по теме анализа микроРНК с помощью онлайн-инструментов биоинформатики.

Выводы: Представлена методика биоинформационного исследования. Использование микроРНК как регуляторных макромолекул представляет перспективную научную тему в биотехнологии и медицине. Анализ предсказанных генов-мишеней позволит найти новые таргеты (мишени) для терапии. Определены гены мишени микроРНК, известных из научной литературы, для ряда заболеваний. Практическое использование найденных микроРНК как маркеров для диагностики заболеваний, в практических проектах по высокотехнологической медицине остается сложной задачей, требующей дальнейшей интеграции ресурсов.

Благодарности: Исследование поддержано проектом Цифровая кафедра Сеченовского Университета.

Development of computer tools to search for microRNA targets

Radochin P.A.¹, Voronina A.A.², Orel D.Y.², Smirnova O.A.¹, Golubchikov D.O.¹,

Mak D.V.¹, Voronova S.S.¹, Volkov I.A.¹, Savina E.A.^{1,3}, Orlov Y.L.^{1,4*}

¹ Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

² St. Petersburg Polytechnical University, St. Petersburg, Russia

³ Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

⁴ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* orlov@bionet.nsc.ru

Key words: bioinformatics; microRNAs; databases; gene networks; complex diseases; gene ontologies; education

Motivation and Aim: The study of the molecular mechanisms of complex diseases is an important task of biomedicine, requiring data integration and the development of new bioinformatic tools [1]. microRNAs are a class of non-coding RNA molecules that regulate the translation and degradation of their mRNA targets through interaction, which makes the study of their functions in diseases necessary to elucidate their pathogenetic mechanisms [2]. miRNAs play an important role in regulating gene expression. New literature data are emerging on the role of microRNAs in tumor progression, the development of mental disorders, liver diseases, and chronic diseases [1].

A scientific project for the study of microRNA target genes within the Digital Department of Sechenov University in the 2023–2024 academic year is presented [3]. Mastering the use of existing online programs for computer search of microRNA target genes in the genome, and developing proprietary programs for predicting microRNA binding sites in genes.

The problem being solved within the framework of the presented project is training on the problems of using bioinformatics tools to analyze the structure of RNA and microRNA, the mechanisms of binding RNA to human gene transcripts.

Methods and Algorithms: The available software tools for searching for microRNA target genes have been collected, and a data processing pipeline has been prepared.

Reviews were used (List of programs and databases –

<https://www.tamirna.com/micromas-web-based-tools/>), MIRDB databases (<https://mirdb.org/>), TargetScan (https://www.targetscan.org/vert_80/), MiRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php), microRNA is an RNA consisting of ~22 nucleotides processed from the hairpin structures of RNA. microRNAs are too short, which creates technical difficulties in finding them. The determination of target genes by partial complementarity is also complicated by the possibility of nucleotide mismatch [4].

Lists of genes and microRNAs associated with diseases were compiled based on queries to the OMIM (Online Mendelian Inheritance in Man), GeneCards, and MalaCards databases. The analysis of gene ontologies for the list of genes was performed using the PANTHER (Protein ANalysis THrough Evolutionary Relationships) resource (<http://pantherdb.org/>) and DAVID. STRING-DB tools and Cytoscape applications were used to reconstruct the network. An up-to-date data set was analyzed using online bioinformatics tools.

Results: We compared work of the tools for miRNA prediction. TargetScan predicts biological microRNA targets by searching for conserved 8-, 7-, and 6-dimensional sites that correspond to the initial region of each microRNA. Predictions with only weakly conservative sites and predictions with unsaved microRNAs are also offered as options [5]. Sites with inconsistencies in the initial region have also been identified, which are compensated by conservative 3'-pairing. In mammals, forecasts are ranked based on the predicted targeting efficiency calculated using a biochemical model of microRNA-mediated repression, common features associated with microRNA binding and decreased regulation of targets were identified, which were used to predict targets for microRNAs using machine learning methods.

To search for target genes for therapy, it is necessary to reconstruct the gene network – a complex of interacting macromolecules, including miRNAs, functionally related to the disease. Further analysis involves clustering genes in the network, identifying key genes with the largest number of contacts in the network, and finding categories of gene ontologies.

A pipeline of using software tools to search for microRNA target genes for medicine is presented. It is possible to further use such programs as educational resources on bioinformatics for physicians.

The works were presented at a student conference at Sechenov University, and student diplomas were prepared on the topic of microRNA analysis using online bioinformatics tools.

Conclusion: The technique of bioinformatics research is presented. The use of microRNAs as regulatory macromolecules is a promising scientific topic in biotechnology and medicine. The analysis of predicted target genes will allow us to find new targets for therapy. microRNA target genes known from scientific literature have been identified for a number of diseases. The practical use of the found microRNAs as markers for the diagnosis of diseases in practical projects on high-tech medicine remains a difficult task requiring further integration of resources.

Acknowledgements: The study was supported by the Digital Chair project of Sechenov University.

Список литературы/References

1. Beylerli O., Ilyasova T., Shi H., Sufianov A. MicroRNAs in meningiomas: Potential biomarkers and therapeutic targets. *Noncoding RNA Res.* 2024;9(3):641-648. doi 10.1016/j.ncrna.2024.02.011
2. McGeary S.E., Lin K.S., Shi C.Y., et al. The biochemical basis of microRNA targeting efficacy. *Science.* 2019;366(6472):eaav1741. doi 10.1126/science.aav1741
3. Turkina V.A., Orlova N.G., Orlov Y.L. Biophysics education section and computational training discussion at VII Congress of Russian Biophysicists. *Biophys Rev.* 2023;15:807-809. doi 10.1007/s12551-023-01147-5
4. Liu W., Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* 2019;20:18. doi 10.1186/s13059-019-1629-z
5. Peterson S.M., Thompson J.A., Ufkin M.L., Sathyanarayana P., Liaw L., Congdon C.B. Common features of microRNA target prediction tools. *Front. Genet.* 2014;5:23. doi 10.3389/fgene.2014.00023