

**Импутация данных РНК-белковых взаимодействий для поиска новых триад  
«ДНК-РНК-белок»**

**Научный руководитель – Миронов Андрей Александрович**

***Рыбаков Арсений Константинович***

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова, Факультет  
биоинженерии и биоинформатики, Москва, Россия

*E-mail: irisunikova@gmail.com*

При изучении контактов РНК с хроматином показывает себя полезным подход поиска триад взаимодействия «ДНК-РНК-белок», разработанный в нашей лаборатории. Он основан на пересечении результатов экспериментов по выявлению ДНК-РНК, ДНК-белковых и РНК-белковых взаимодействий. При этом количество найденных троек ограничивается самым "бедным" из трёх типов данных - РНК-белковыми контактами. Существуют различные лабораторные протоколы обнаружения взаимодействующих РНК-белковых пар (eCLIP [6], fRIP-seq [3] и другие), однако за один эксперимент можно получить данные только для одного белка, что и является причиной того, что накопленные к настоящему моменту результаты покрывают довольно мало белков (например, около 100 для клеточной линии K562) [4]. Поэтому в решении этой проблемы набирают актуальность вычислительные подходы [7]. В настоящей работе была разработана модель машинного обучения, позволяющая сделать вывод о наличии взаимодействия в произвольной паре РНК-белок.

Датасет для обучения модели был сконструирован из нескольких существующих датасетов, доступных публично [5]. Для векторизации последовательности РНК использовался подсчёт количества k-мер'ов, последовательность белка векторизовалась при помощи алгоритма UniRep [2]. Структуры белков были представлены как графы [1]. В качестве архитектуры модели была избрана полносвязная нейронная сеть с дополнительными слоями для обработки и векторизации графов.

Таким образом, в ходе работы была создана модель, принимающая на вход информацию о структуре и последовательности как белка, так и РНК, и проводящая бинарное предсказание о наличии контакта между ними. В дальнейшем планируется использовать модель для расширения данных РНК-белковых взаимодействий и увеличения количества найденных триад "ДНК-РНК-белок".

**Источники и литература**

- 1) Козлов М.В., Попова Н. Н. Оценка сходства 3D-структур белка с использованием нейросетевых подходов // Магистерская дипломная работа, ВМК МГУ. 2020.
- 2) Alley, Ethan C., et al. Unified rational protein engineering with sequence-based deep representation learning // Nature methods. No 16.12. 2019. С. 1315-1322.
- 3) G Hendrickson, David, et al. Widespread RNA binding by chromatin-associated proteins // Genome biology. No 17. 2016. С. 1-18.
- 4) Hu, Boqin, et al. POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins // Nucleic acids research. No 45.D1. 2017. С. D104-D114.
- 5) Sun, L., Xu, K., Huang, W. et al. Predicting dynamic cellular protein-RNA interactions by deep learning using in vivo RNA structures // Cell Res. No 31. 2021. С. 495-516.

- 6) Van Nostrand, Eric L., et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP) // Nature methods. No 13.6. 2016. С. 508-514.
- 7) Wei, Junkang, et al. Protein–RNA interaction prediction with deep learning: structure matters // Briefings in bioinformatics. No 23.1. 2022. С. 540.