

## Предсказание субклеточной локализации РНК

Научный руководитель – Миронов Андрей Александрович

**Тюкаев Артём Алексеевич**

*Студент (специалист)*

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

*E-mail: artyomtyukaev@gmail.com*

РНК могут функционировать в различных клеточных компартментах, при этом их функция зачастую тесно связана с местом локализации [1, 3]. При этом речь идет не только про мРНК, но и про малоизученные некодирующие РНК, которых в клетке достаточно много, и они принимают участие во всех фундаментальных процессах работы клетки. Предсказание локализации РНК может пролить свет на их потенциальные функции и помочь изучить общую картину распределения транскриптов в клетке.

Существующие модели машинного обучения для предсказания локализации РНК имеют ряд недостатков. Целью работы является создание усовершенствованной модели для предсказания локализации РНК в клетке.

Мы провели анализ существующих моделей и обработали сырые данные из эксперимента APEx-RIP [2], который предоставляет библиотеки прочтений из ядра, цитоплазмы, матрикса митохондрий и люмена ЭПР. Эти данные послужили материалом для обучения и валидации модели. Первым этапом мы проверили, возможно ли предсказывать локализацию РНК, основываясь только на k-мерном составе последовательности. Для этой цели мы составили обучающую выборку из 8338 РНК со значимым обогащением в ядерном или цитоплазматическом компартментах. В качестве модели был выбран градиентный бустинг LightGBM. После подбора гиперпараметров и обучения метрика качества ROC-AUC составила 0.78 на тестовом наборе данных из 2780 РНК (Рис.). Следующим этапом мы решили проверить, можно ли предсказывать субклеточную локализацию для участка гена определенного размера, чтобы учесть его экзонную или интронную природу. Для этого мы разбили геном на окна по 200 нуклеотидов, для каждого такого участка определили его принадлежность к экзону или интрону и картировали на эти участки прочтения из APEx-RIP. Далее для предсказания локализации этих участков была подобрана модель машинного обучения.

## Источники и литература

- 1) Marina Chekulaeva. Mechanistic insights into the basis of widespread RNA localization // Nature Cell Biology volume 26, pages 1037–1046 (2024).
- 2) Pornchai Kaewsapsak, et al. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking // eLife, 6:e29224 (2017).
- 3) Sebastiaan van Heesch, et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes // Genome Biology volume 15, Article number: R6 (2014).

## Иллюстрации

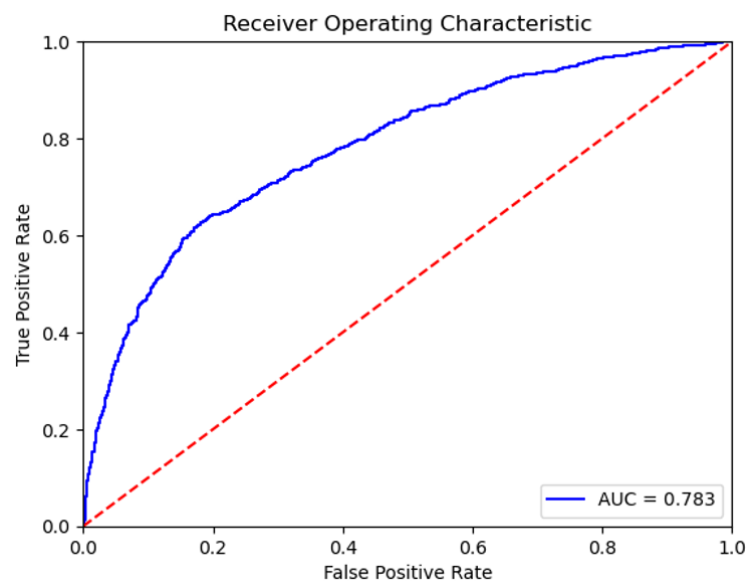


Рис. : Метрика качества ROC-AUC обученной модели.