

СРАВНИТЕЛЬНАЯ ГЕНОМИКА РИБОЗИМОВ И РИБОСВИТЧЕЙ

Сорокина Н.В. (Университет ИТМО), **Зилов Д.С.** (Университет ИТМО)
Научный руководитель – в.н.с, к.б.н Комиссаров А.С. (Университет ИТМО)

Рибозимы и рибосвитчи перспективные инструменты для редактирования генома. С помощью машинного обучения возможно сгенерировать новые молекулы с подобными свойствами и функциями, они могут стать полезными инструментами в генной инженерии. Для обучения нейросети необходима обширная база данных рибозимов и рибосвитчей, ее создание описано в данной работе.

Введение. Развитие генной инженерии требует новых инструментов. Молекулы рибонуклеиновой кислоты (РНК), участвующие в регуляторных процессах в клетке, способны стать полезными и необходимыми средствами как в лабораторной практике, так и при разработке генной терапии. Одни из таких молекул — рибозимы и рибосвитчи. Первые способны осуществлять щелочной катализ, наподобие ряда белковых ферментов. Вторые могут регулировать экспрессию генов за счет конформационных изменений в ответ на присоединение специфичных метаболитов. Идея состоит в том, чтобы сгенерировать новые, отсутствующие в природе регуляторные молекулы РНК с использованием машинного обучения. Для этого необходима большая база данных рибозимов и рибосвитчей, а среди существующих нет удовлетворяющей этому условию. В данной работе будет рассмотрен способ создания базы данных рибозимов и рибосвитчей, найденных во всех доступных бактериальных геномах.

Основная часть. Для аннотации были отобраны 16 семейств рибозимов и 45 семейств рибосвитчей, существующих в прокариотических организмах. Для каждого из них вычислена ковариационная модель (КМ) — статистическая модель, описывающая пространственную структуру РНК. Используя эти модели как шаблоны, инструмент Infernal способен осуществлять поиск семейств регуляторных молекул в прокариотических геномах. Бактериальные геномы в количестве 216 118 были взяты из базы данных NCBI, ковариационные модели семейств — из базы данных Rfam. Принцип работы программы Infernal: считывает анализируемый геном в определенном участке, вычисляет удельный счет каждого участка и сравнивает его с шаблоном — ковариационной моделью. Если они совпадают, программа аннотирует регион в качестве рибозима или рибосвитча. В результате была получена база данных, состоящая из 43 118 589 элементов, среди которых 10 626 962 рибозимы и 32 491 627 рибосвитчи. Помимо основных, были получены косвенные результаты. Изучая элементы созданной базы данных, мы установили наличие коротких нуклеотидных последовательностей РНК длиной меньше, чем 16 нуклеотидов, расположенных рядом друг с другом. Вероятность того, что такие элементы могут быть независимыми регуляторными молекулами очень мала. С помощью инструмента R2DT были сгенерированы 2D структуры этих элементов. Некоторые из них способны образовывать вторичную структуру. Есть предположение, что эти элементы являются фрагментами новых рибозимов или рибосвитчей. Данная гипотеза сейчас проверяется.

Выводы. Полученная база данных намного больше других, содержащих информацию о рибозимах и рибосвитчах. Во-первых, потому что мы впервые выполнили полногеномный поиск. Во-вторых, были исследованы все доступные геномы бактерий. Используя эту базу данных, мы планируем построить генеративно-состязательную нейросеть (GAN), способную генерировать новые рибозимы и рибосвитчи. Помимо поиска новых рибозимов и рибосвитчей, полученная база данных, после ее публикации, может быть использована

учеными для поиска нуклеотидных последовательностей для синтеза регуляторных молекул РНК. Также такая обширная база данных может быть использована для изучения эволюции рибозимов и рибосвитчей.

Сорокина Н.В. (автор)

Подпись

Комиссаров А.С. (научный руководитель)

Подпись