

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ ДЛЯ ГЕНЕРАТИВНОГО ДИЗАЙНА НАПРАВЛЯЮЩИХ РНК СИСТЕМЫ CRISPR-CAS9

Кечин А. А.¹, Вепрева А. С.¹

Научный руководитель – канд. хим. наук, доцент Серов Н. С.¹

¹Университет ИТМО

kechin.ars@icloud.com

Введение

Система CRISPR-Cas9 представляет собой программируемый механизм редактирования генома, в котором направляющая РНК (gRNA) обеспечивает распознавание целевой последовательности ДНК и определяет точность разреза. Качество gRNA напрямую влияет на эффективность редактирования (on-target активность) и вероятность возникновения нежелательных мутаций в других участках генома (off-target эффекты).

В литературе предложены модели машинного обучения для прогнозирования активности и специфичности CRISPR, включая работу [1], посвященную предсказанию on-target эффективности на основе позиционных нуклеотидных предпочтений, глубокую нейросетевую модель DeepSpCas9 [2], а также биофизическую модель CRISPRspec [3] для оценки специфичности gRNA на основе термодинамических параметров. Эти подходы эффективно решают задачу предсказания и ранжирования gRNA по отдельным критериям. Однако существующие методы ориентированы на оценку уже заданных последовательностей и не реализуют прямую оптимизацию gRNA. Они не обеспечивают многоцелевой генеративный поиск с одновременным учетом эффективности, специфичности и других ограничений. Таким образом, алгоритмы комплексной оптимизации gRNA фактически отсутствуют.

В данной работе мы впервые применяем обучение с подкреплением для решения задачи многоцелевой оптимизации gRNA, рассматривая генерацию последовательности как процесс направленного поиска с учетом совокупности критериев качества.

Основная часть

В работе задача оптимизации gRNA формализуется как задача обучения с подкреплением. Последовательность длиной 20 нуклеотидов рассматривается как состояние среды, а агент последовательно модифицирует её, формируя новую направляющую РНК. В качестве действий задаются точечные нуклеотидные замены в отдельных позициях последовательности. Такие мутации позволяют целенаправленно изменять профиль связывания gRNA с геномной ДНК, в частности снижать вероятность off-target взаимодействий при сохранении высокой on-target активности. При этом модификации ограничены биологически обоснованными правилами: агент не изменяет нуклеотиды в seed зоне, критически важной для распознавания мишени, что позволяет сохранить базовую комплементарность к целевому сайту.

Функция вознаграждения носит многоцелевой характер и включает предсказанную on-target эффективность, специфичность по отношению к потенциальным off-target сайтам, а также штрафы за отклонение в суммарном содержании гуанина и цитозина и за наличие гомополимерных участков. Такая постановка позволяет направленно искать последовательности с оптимальным совокупным профилем характеристик.

Выводы

Разработанный метод демонстрирует возможность применения обучения с подкреплением для многоцелевой оптимизации направляющих РНК системы CRISPR-Cas9. Полученные результаты подтверждают улучшение характеристик gRNA при сохранении биофизических ограничений и снижении потенциальных off-target эффектов, а также способность модели переноситься на ранее не использованные гены.

Практическая значимость работы заключается в возможности интеграции алгоритма в программные инструменты проектирования CRISPR для предварительного компьютерного скрининга gRNA перед экспериментальной валидацией. Перспективным направлением является расширение подхода на другие системы редактирования и включение дополнительных факторов, связанных с клеточным контекстом и механизмами репарации ДНК.

Литература

1. Doench J., Fusi N., Sullender M. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9 // Nature Biotechnology. 2016. Vol. 34, no. 2. P. 184–191. <https://doi.org/10.1038/nbt.3437>.
2. Wang D., Zhang C., Wang B. et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning // Nature Communications. 2019. Vol. 10. P. 4284. <https://doi.org/10.1038/s41467-019-12281-8>.
3. Alkan F., Wenzel A., Anthon C. et al. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters // Genome Biology. 2018. Vol. 19. P. 177. <https://doi.org/10.1186/s13059-018-1534-x>.