

PRIMED: предсказание сайтов связывания ДНК с использованием предобученных белковых языковых моделей

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1763313>

биоинформатика

машинное обучение

молекулярная биология

предсказание структуры белка

терапевтический дизайн

Введение

Взаимодействия белок-ДНК лежат в основе регуляции генов, стабильности генома и механизмов заболеваний. Идентификация ДНК-связывающих остатков (DBRs) критически важна для структурного моделирования, белковой инженерии и разработки терапевтических средств. Хотя экспериментальные подходы предоставляют ценные данные, они остаются низкопроизводительными и ресурсозатратными. Вычислительные методы предлагают масштабируемые альтернативы, используя последовательную и структурную информацию о белках для предсказания DBRs.

Методы

Мы представляем PRIMED (Protein Residue Inference using Multilayer perceptron for Enhanced DNA-binding predictions), фреймворк машинного обучения, который интегрирует представления белков с различными биохимическими и структурными свойствами из трёх моделей языка белков: ESM-2, ESM-3 и ESM-C. Эти представления объединяются и обрабатываются многослойным перцептроном для выполнения предсказаний DBRs.

Результаты

PRIMED продемонстрировал высокую производительность на трёх наборах данных для тестирования: Test-46 и Test-129 из предыдущего исследования, CLAPE-DB и Test-10 К, которые мы собрали из UniProtKB/Swiss-Prot. Модель достигает площади под кривой рабочей характеристики приёмника (AUC) 0,92 и коэффициента корреляции Мэтьюза (MCC) 0,64 на Test-46, а также AUC 0,93 и MCC 0,45 на Test-129. На Test-10 К PRIMED демонстрирует обобщаемость для белков с различными процентами DBRs, сохраняя конкурентоспособную производительность по сравнению с методом-соперником, CLAPE-DB.

Обсуждение

Эти результаты подчёркивают эффективность интеграции разнообразных представлений моделей языка белков для точных и переносимых предсказаний DBRs.

Перевод выполнен: 25.03.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.