

Тонко настроенные базовые модели на основе GPT эффективно реконструируют транскрипционные регуляторные сети бактерий из литературы

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1788196>

LLM

NLP

бактериология

биоинформатика

регуляторные сети

Введение

Жизнь обладает свойством производить из одного генома (совокупности молекул ДНК) различные типы клеток, а также механизмы, позволяющие бактериям адаптироваться к изменениям окружающей среды. Хотя регуляция может происходить на разных уровнях, регуляция инициации транскрипции — начала процесса копирования ДНК в РНК — является наиболее изученным уровнем у бактерий. Совокупность регуляторов и их регулируемых элементов определяет **транскрипционные регуляторные сети (TRNs)**, изучение которых стимулирует развитие таких важных областей, как антимикробная резистентность. Их анализ и понимание зависят от нескольких немногих баз данных, прошедших тщательную ручную курацию. Традиционный способ реконструкции этих сетей заключается в ручном анализе литературы, что является точным, но трудоемким и длительным процессом. Эти ограничения привели к дефициту и неполноте данных о бактериальных TRNs.

Методы

В данной работе мы представляем новый ансамблевый модельный подход с использованием двух базовых моделей на основе GPT (**LLaMA-3** и **GPT-4o mini**) для эффективной реконструкции TRNs на основе литературы. Мы применили стратегию контролируемого тонкого настраивания (**supervised fine-tuning**) с использованием предложений из литературы по *Escherichia coli*, чтобы обучить модели предсказывать тип регуляторного эффекта между фактором транскрипции и регулируемым элементом (геном/опероном). Для оценки эффективности реконструкции курируемой TRN мы использовали 264 полнотекстовые статьи по *Salmonella Typhimurium*, патогену, представляющему клинический интерес.

Результаты

На тестовых данных обе модели продемонстрировали значительную производительность (**F1-Score** > 0,87, коэффициент корреляции Мэтьюса > 0,82). При реконструкции курируемой TRN ансамблевый подход, использующий согласованность моделей, правильно реконструировал 80% TRN (**Recall**: 0,80, **F1-score**: 0,64). Мы применили этот подход для реконструкции обширной TRN *Salmonella*, используя литературу по транскрипционной регуляции этой бактерии, доступную на тот момент (2 278 статей). Эта сеть была описана с помощью сетевых метрик, анализа сверхпредставленности и сравнена с существующими биологическими знаниями.

Обсуждение

Наш подход превзошел результаты предыдущих работ по предсказанию эффекта взаимодействия. Анализ TRN, полученной на основе 2 278 статей, показал эффективность нашего метода для реконструкции TRNs различных бактерий, так как сеть соответствует биологическим знаниям. Таким образом, наша работа может способствовать изучению бактерий, представляющих биологический и клинический интерес, особенно тех, для которых TRN еще не была реконструирована.

Перевод выполнен: 13.04.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.