

УДК 575.112, 004.8

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ ГРАФОВЫХ И ПОСЛЕДОВАТЕЛЬНЫХ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕКЦИИ ХРОМОТРИПСИСА В ОНКОГЕНОМАХ

Карицкая П.Е. (НГУ)

Научный руководитель – доцент Вяткин Ю.В. (НГУ)

**Введение.** В онкогенезе ключевую роль играет геномная нестабильность, приводящая к накоплению крупномасштабных перестроек ДНК. Одним из наиболее разрушительных проявлений такой нестабильности является хромотрипсис – катастрофическое событие, при котором в пределах одной или нескольких хромосом за короткое время возникает множество локализованных разрывов и повторных соединений фрагментов, формируя характерные кластеры точек разрыва и осциллирующие состояния копияности [1]. Выявление хромотрипсиса имеет практическую ценность как для фундаментального понимания механизмов развития опухолей, так и для задач молекулярной диагностики и стратификации риска.

**Основная часть.** Существующие программные решения для детекции хромотрипсиса (например, ShatterProof и ShatterSeek) как правило опираются на набор эвристических статистических критериев и требуют ручной настройки порогов и параметров, что снижает воспроизводимость при анализе данных разного качества и происхождения. В качестве альтернативы рассматриваются модели машинного обучения, однако подходы, основанные только на агрегированных признаках профиля вариаций числа копий (CNV), могут терять информацию о пространственной организации и топологии перестроек.

В работе предложен воспроизводимый фреймворк глубокого обучения для автоматизированной классификации событий хромотрипсиса с учетом структурных вариаций (SV) и CNV. В качестве референсного набора использована вручную аннотированная база ChromothripsisDB, включающая 2428 пациентов, 37 типов опухолей и 19713 размеченных геномных локусов [2, 3]. Данные SV из формата BEDPE были унифицированы для совместной обработки (в том числе по источникам ICGC и TCGA), после чего каждое SV-событие в пределах клинического региона представлялось в виде 32-мерного вектора признаков, содержащего пространственные характеристики, статистические агрегаты и нелинейные взаимодействия. Для моделирования были использованы два формата представления регионов: последовательности упорядоченных по координате событий SV длиной до 200 элементов (с паддингом) для обучения BiLSTM и графы, где вершины соответствуют SV-событиям, а ребра задают пространственную близость и связь между событиями, что позволяет обучать графовые нейронные сети (GNN). Эксперименты проводились в модульном пайплайне с управлением конфигурациями через Hydra, что обеспечивает повторяемость сравнений архитектур и протоколов обучения.

В качестве базовой точки сравнения выполнен бенчмаркинг CNV-ориентированной модели GECNVNet [4]. Было показано, что исходная версия склонна к смещению в сторону положительного класса (высокий recall при низкой precision). Даже после оптимизации обучения модель по-прежнему ограничена отсутствием явного моделирования локальной кластеризации разрывов. Для устранения этого ограничения предложены событийно-ориентированные модели, работающие напрямую с паттернами SV на уровне региона: BiLSTM, учитывающая порядок разрывов, и Graph-GNN, учитывающая топологическую связанность и геометрию кластера перестроек.

**Выводы.** Проведенные эксперименты показали, что событийные представления на основе SV несут более информативный сигнал для детекции хромотрипсиса, чем глобальные CNV-сводки. На данных ChromothripsisDB модели BiLSTM и Graph-GNN продемонстрировали более сбалансированное качество при сильном дисбалансе классов;

наилучшие показатели достигнуты графовой моделью ( $F1=0.78$ ,  $precision=0.90$ ,  $recall=0.70$ ), тогда как BiLSTM показала  $F1=0.71$ . Практически результаты могут быть использованы как часть автоматизированного биоинформатического пайплайна первичного скрининга регионов с признаками хромотрипсиса и как основа для дальнейшего расширения модели (валидация на независимых когортах, интеграция CNV-осцилляций и других геномных контекстов).

#### **Список использованных источников:**

1. Stephens P. J. [и др.]. Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development // *Cell*. 2011. Т. 144. № 1. С. 27–40.
2. Cortés-Ciriano I. [и др.]. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing // *Nature Genetics*. 2020. Т. 52. № 3. С. 331–341.
3. Yang J., Deng G., Cai H. ChromothripsisDB: a curated database of chromothripsis // *Bioinformatics*. 2015. Т. 32. № 9. С. 1433–1435.
4. Yu J. [и др.]. Chromothripsis detection with multiple myeloma patients based on deep graph learning // *Bioinformatics*. 2023. Т. 39. № 7.