

УДК 004.85:575.2

Метод для поиска фенотипически обогащенных кластеров Т-клеточных рецепторов

Власова Е. К. (ИТМО)

Научный руководитель – Шугай М. А. (РНИМУ им. Пирогова)

Введение. Т-клеточные рецепторы (TCR) обеспечивают распознавание антигенов и являются ключевым элементом адаптивного иммунитета. Совокупность TCR-клонотипов образует иммунный репертуар, отражающий историю иммунных ответов организма. Анализ репертуаров Т-клеточных рецепторов рассматривается как перспективный подход к выявлению молекулярных биомаркеров инфекционных, аутоиммунных и онкологических заболеваний.

Однако высокая вариабельность TCR и отсутствие аннотаций для большинства клонотипов существенно усложняют поиск значимых иммунных сигналов. Существующие методы часто опираются на точные совпадения последовательностей или простые меры сходства, что ограничивает чувствительность анализа и затрудняет выявление функционально сходных рецепторов, распознающих один и тот же антиген [1, 2].

В работе рассматривается метод TCReMPNet для выявления фенотипически обогащенных групп TCR-клонотипов без априорного знания эпитопов, основанный на эмбедингах последовательностей и статистическом анализе обогащения кластеров относительно контрольного фона.

Основная часть. На первом этапе клонотипы TCR представляются в виде векторных эмбедингов методом TCReMP, основанным на similarity-подходе с использованием прототипных последовательностей [5, 6]. Такой подход позволяет перейти от сравнения отдельных последовательностей к анализу структуры пространства клонотипов и обеспечивает масштабируемость для больших наборов данных.

Далее выполняется совместная кластеризация клонотипов исследуемого образца и контрольного фона. В качестве фона могут использоваться репертуары здоровых доноров, образцы до изменения иммунного статуса, а также синтетические последовательности, полученные моделированием V(D)J-рекомбинации (например, с помощью OLGA) [3]. Для кластеризации используются графовые и плотностные подходы, включая алгоритм Leiden и модифицированные варианты DBSCAN, что позволяет учитывать неоднородную плотность распределения клонотипов в пространстве эмбедингов [4].

После кластеризации для каждого кластера оценивается статистическая значимость обогащения клонотипами исследуемого образца относительно контроля. Для этого сравнивается наблюдаемое число последовательностей образца с ожидаемым распределением на фоне; далее проводится коррекция на множественные проверки гипотез. Такой подход позволяет выделять кластеры, ассоциированные с фенотипом, даже при отсутствии предварительной информации об антигенной специфичности.

Метод масштабируется на когортные данные и допускает объединение обогащенных кластеров между пациентами для поиска воспроизводимых иммунных сигналов. По предварительным результатам, подход применим как к задачам анализа аутоиммунных заболеваний, так и к данным вакцинации и валидации аннотаций в специализированных базах TCR [1].

Выводы. Разрабатываемый метод TCReMPNet позволяет выявлять фенотипически ассоциированные группы TCR-клонотипов без предварительного знания эпитопов за счет комбинации эмбедингов, кластеризации относительно контрольного фона и статистического анализа обогащения.

Подход ориентирован на анализ крупных когорт и может использоваться для поиска иммунных биомаркеров инфекционных и аутоиммунных заболеваний, а также для повышения качества аннотаций TCR в специализированных базах данных.

Список использованных источников:

1. Shugay M., Bagaev D. V., Zvyagin I. V., et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity // *Nucleic Acids Research*. 2018. Vol. 46, no. D1. P. D419–D427. DOI: 10.1093/nar/gkx760.
2. Dash P., Fiore-Gartland A. J., Hertz T., et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires // *Nature*. 2017. Vol. 547, no. 7661. P. 89–93. DOI: 10.1038/nature22383.
3. Sethna Z., Elhanati Y., Callan C. G. Jr., Walczak A. M., Mora T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs // *Bioinformatics*. 2019. Vol. 35, no. 17. P. 2974–2981. DOI: 10.1093/bioinformatics/btz035.
4. Traag V. A., Waltman L., van Eck N. J. From Louvain to Leiden: guaranteeing well-connected communities // *Scientific Reports*. 2019. Vol. 9. Art. 5233. DOI: 10.1038/s41598-019-41695-z.
5. Kremlyakova Y., Vlasova E. K., Luppov D., Shugay M. TCReMP: a bioinformatic pipeline for efficient embedding of T-cell receptor sequences from immune repertoire and single-cell sequencing data // *Journal of Molecular Biology*. 2025. Art. 169205. DOI: 10.1016/j.jmb.2025.169205.
6. antigenomics/tcremp: T-cell receptor sequence embedding pipeline (GitHub repository).

Власова Е. К.

Шугай М. А.