

Разработка подходов к анализу масс-спектрометрических данных методом главных компонент для поиска потенциальных биомаркеров клеточных субпопуляций

Научный руководитель – Клычников Олег Игоревич

Стрекаловских Вадим Владимирович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: vadimstrek@yandex.ru

В современной биомедицине масс-спектрометрические данные позволяют проводить качественный и количественный анализ молекулярного состава клеток и тканей. Известно, что клеточные субпопуляции различаются по уровню экспрессии генов и по представленности тех или иных белков. Для понимания различий между ними на клеточном и субклеточном уровне также важно исследование их секретомы с применением, в первую очередь, методов хроматографии и масс-спектрометрии^[1]. При этом важное место занимает поиск потенциальных биомаркеров, который требует эффективных методов анализа больших объёмов данных для выявления скрытых закономерностей. Один из таких подходов к анализу – метод главных компонент, который позволяет сжать информацию и выделить наиболее значимые компоненты в многомерных данных^[2]. В данном исследовании предполагается разработать алгоритм анализа экспериментальных масс-спектрометрических данных с целью выявления отдельных молекулярных маркеров клеточных субпопуляций с помощью исследования их секретомы.

Объект исследования – секретомы трёх разных популяций мезенхимных стромальных клеток (МСК) человека, культивированных в контрольных или в профибротических условиях (тотальная популяция и субпопуляция PDGFR α + МСК, выделенная методом клеточного сортирования). Для каждого образца проведено три повторности масс-спектрометрического анализа.

Методы анализа данных:

1. Предварительная фильтрация масс-спектрометрических данных.
2. Разметка пиков («peak picking») для преобразования трёхмерного набора данных в массив пиков, которые некоторым образом определяются как наиболее информативные.
3. Методы выравнивания хроматограмм масс-спектрометрических экспериментов.
4. Метод главных компонент для выявления молекулярных паттернов, позволяющих разделить группы образцов.
5. Методы анализа «без учителя» (unsupervised methods) для прямого сравнения паттернов масс-спектрометрических данных.

Мы составили два конвейера анализа данных с различными значениями параметров каждого используемого метода. Оба конвейера мы применили к имеющимся датасетам, и по результатам анализа главных компонент выбрали оптимальный алгоритм. Анализ главных компонент показал хорошую степень разделения и кластеризации имеющихся образцов.

Результаты исследования показывают, что секретомы трёх популяций МСК заметно различаются между собой. Более того, в секретоме PDGFR α + МСК, культивированных в профибротических условиях, наблюдается совершенно уникальный набор секретлируемых веществ, значимо отличающийся от секретомы двух других популяций. Применение анализа главных компонент параллельно с методами идентификации веществ в перспективе

может позволить определить химическую природу и биологические функции обнаруженных секретируемых биомаркеров.

Источники и литература

- 1) Kulebyakina M. et al., “Balance between Pro- and Antifibrotic Proteins in Mesenchymal Stromal Cell Secretome Fractions Revealed by Proteome and Cell Subpopulation Analysis”, *Int J Mol Sci.* 2023 Dec 25;25(1):290.
- 2) Melanie Hilario, Alexandros Kalousis, “Approaches to dimensionality reduction in proteomic biomarker studies”, *Briefings in Bioinformatics*, Volume 9, Issue 2, March 2008, Pages 102–118.