

OpenPtmFinder: интеграция подхода открытого поиска и статистического моделирования для анализа пост-трансляционных модификаций белков в ТМТ-протеомике

Научный руководитель – Тарасова Ирина Алексеевна

Копейкина Анна Сергеевна

Аспирант

Московский физико-технический институт, Москва, Россия

E-mail: koreykina.00@bk.ru

Комплексный анализ пост-трансляционных модификаций (ПТМ) остаётся одной из сложнейших задач современной масс-спектрометрической протеомики из-за химического разнообразия и лабильности ПТМ. Существенный прогресс связан с подходом открытого поиска, позволяющим регистрировать широкий диапазон массовых сдвигов у пептидов без заранее заданного набора модификаций. В традиционном MS/MS-анализе «снизу-вверх» используется узкий допуск по массе ($\pm 0,02-0,05$ Da), что ограничивает выявление неизвестных Δ -масс. Открытый поиск расширяет допуск до сотен дальтон, увеличивая охват модифицированных пептидов. Идентификация ПТМ [U+2500] лишь начальный этап; для корректной интерпретации требуется статистический анализ с учётом стохастичности спектров, мультиплексирования ТМТ и межбатчевых эффектов. Большинство решений ограничены анализом пептидов или белков и редко переходят к уровню сайта модификации, где объединяются различные PSM (peptide-spectrum match). Между тем именно сайт-специфический анализ точнее отражает биологические изменения.

Мы представляем конвейер, включающий программы MSFragger [1], AA_stat [2] и OpenPtmFinder для визуализации и статистического анализа ПТМ на уровне пептидов и сайтов. OpenPtmFinder использует байесовскую агрегацию PSM одного сайта с последующим линейным моделированием по аналогии с limma [3], учитывая межбатчевые смещения и биологическую вариабельность. Метод применим к стандартным масс-спектрометрическим данным без дополнительного обогащения ПТМ.

Конвейер апробирован на глубоко фракционированных ТМТ-данных 190 образцов височной коры головного мозга человека (контроль, бессимптомная и симптоматическая болезнь Альцгеймера). Идентифицировано около 15 000 модифицированных пептидов в каждом наборе данных. Наиболее распространёнными ПТМ стали: дезамидирование, окисление, фосфорилирование, дегидратация и метилирование. Сайт-специфический анализ выявил 100 достоверно изменённых сайтов, тогда как на уровне пептидов — 77 значимых последовательностей, что демонстрирует преимущество агрегации PSM. Изменения представленности ПТМ отражают патогенез нейродегенерации и могут служить источником потенциальных биомаркеров. Исходный код доступен: <https://github.com/Anchovy-bio/OpenPtmFinder>.

Работа выполнена при поддержке РФФ, грант № 23-45-00012.

Источники и литература

- 1) Kong A.T. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics // Nat. Methods. 2017. V. 14. P. 513–520.
- 2) Levitsky L.I. AA_stat: Intelligent profiling of in vivo and in vitro modifications from open search results // J. Proteomics. 2021. V. 248: 104350.
- 3) Ritchie M.E. limma powers differential expression analyses for RNA-sequencing and microarray studies // Nucleic Acids Res. 2015. V. 43(7): e47.