

Gdańsk, 2024-12-06

Recenzja

rozprawy doktorskiej autorstwa mgr Łukasza Prussa zatytułowanej „*Bioinformatic analysis in targeted metabolomics, automating the process of translating raw NMR spectrum signals into qualitative and quantitative data for use in the analysis of disease states*”

Największym wyzwaniem w badaniach metabolomicznych jest bioinformatyczna analiza danych, co związane jest z ogromną złożonością i różnorodnością danych metabolomicznych. Dane te najczęściej obejmują tysiące związków chemicznych o różnej strukturze, stężeniu i właściwościach fizykochemicznych, co utrudnia ich przetwarzanie, interpretację i integrację z innymi typami danych omicznych. W szczególności analiza widm NMR wymaga zaawansowanych algorytmów do przetwarzania surowych sygnałów, przypisywania metabolitów do odpowiednich pików oraz eliminacji zakłóceń i szumów.

Jednym z kluczowych problemów jest standaryzacja procesów analitycznych – brak uniwersalnych metod prowadzi do trudności w porównywaniu wyników między różnymi laboratoriami czy badaniami. Kolejnym wyzwaniem jest identyfikacja metabolitów w skomplikowanych matrycach biologicznych, szczególnie w obecności nieliniowych zależności oraz cech wzajemnie skorelowanych. Tradycyjne metody statystyczne często okazują się niewystarczające w radzeniu sobie z wysoką wymiarowością danych, co wymaga stosowania bardziej zaawansowanych podejść, takich jak uczenie maszynowe czy sztuczna inteligencja.

Równie istotne jest zapewnienie interpretowalności wyników, zwłaszcza przy integracji danych metabolomicznych z innymi typami danych, np. genomowymi czy proteomicznymi. Narzędzia bioinformatyczne muszą umożliwiać biologiczną interpretację danych, co wymaga nie tylko zaawansowanych algorytmów, ale także dostępnych i dobrze opracowanych baz danych, takich jak KEGG czy HMDB. Efektywna bioinformatyczna analiza metabolomiczna nie tylko przyspiesza badania, ale także umożliwia wyciąganie precyzyjnych wniosków dotyczących mechanizmów biologicznych, biomarkerów diagnostycznych czy potencjalnych celów terapeutycznych.

Przedmiot badań niniejszej pracy doktorskiej Pana mgr. Łukasza Prussa, wykonanej pod kierunkiem prof. dr hab. Piotra Młynarza z Zakładu Biochemii, Biologii Molekularnej i Biotechnologii, Wydziału Chemicznego Politechniki Wrocławskiej oraz dr Kaji Milanowskiej-Zabel, perfekcyjnie wpisuje się w wyżej wymienioną problematykę i szeroko pojęty obszar badawczy. Oboje Promotorzy pracy doktorskiej, to uznani eksperci chemii analitycznej w obszarze magnetycznego rezonansu jądrowego oraz bioinformatyki. Rozprawa doktorska potwierdza także interdyscyplinarny charakter wykonywanych badań metabolomicznych.

Rozprawa doktorska autorstwa Łukasza Prussa dotyczy potoku przetwarzania danych (*pipeline*) o akronimie NASQQ (*Nextflow Automatization and Standardization for Qualitative and Quantitative ¹H NMR Metabolomic Pipeline*) czyli otwarty i ogólnodostępny potok przetwarzania danych, zaprojektowany do analizy jednowymiarowych protonowych widm NMR, stanowi ważny wkład w rozwój bioinformatyki i metabolomiki, szczególnie w kontekście analizy spektroskopii magnetycznego rezonansu jądrowego (1D ¹H NMR). Praca odpowiada na kluczowe wyzwania współczesnej metabolomiki, takie jak potrzeba standaryzacji procesów analitycznych, automatyzacji oraz integracji z metodami uczenia maszynowego związane z wysoką zmiennością i złożonością danych metabolomicznych. Zaproponowana automatyzacja minimalizuje ryzyko błędów ludzkich i usprawnia analizę wyników, co stanowi istotny postęp w badaniach metabolomicznych. Podejście to ma istotne

znaczenie w badaniach mechanizmów chorób i monitorowania efektów terapii, a zaprezentowane rozwiązania mogą znaleźć zastosowanie w diagnostyce i badaniach translacyjnych.

W rozprawie doktorskiej Autor odnosi się do pięciu publikacjach naukowych swojego współautorstwa, w tym dwóch w trakcie recenzji, które powstały w trakcie wykonywania badań. Publikacje opublikowane ukazały się w czasopismach z listy *Journal Citation Reports* i obejmują zarówno prace o charakterze zastosowań klinicznych (np. badania nad reumatoidalnym zapaleniem stawów), jak i publikacje metodologiczne, co również podkreśla interdyscyplinarność rozprawy.

Praca jest logicznie podzielona na cztery główne rozdziały, które wyczerpująco przedstawiają wprowadzenie, metodykę, wyniki oraz dyskusję. Każda część jest spójna, a wykorzystanie zaawansowanych technologii bioinformatycznych i narzędzi analizy danych zostało opisane w sposób precyzyjny i zgodny z najlepszymi praktykami naukowymi.

We wprowadzeniu Autor w sposób przystępny przedstawia podstawy teoretyczne metabolomiki oraz techniki NMR, a także kontekst ich zastosowania w badaniach nad poszukiwaniem biomarkerów chorób. Na koniec przedstawia metody obliczeniowe, w tym bioinformatykę, jej kluczową rolę w analizie danych metabolomicznych oraz przegląd baz danych i nowoczesnych narzędzi. Część eksperymentalna w szczegółowy sposób opisuje zastosowaną metodykę. Przedstawia ogólny przebieg analizy metabolomicznej, z naciskiem na implementację NextFlow oraz wykorzystanie środowiska obliczeniowego. Omówiono proces przetwarzania surowych widm 1D oraz identyfikację metabolitów, a także moduł analizy danych obejmujący testy uniwersalne, wielowymiarowe oraz modele uczenia maszynowego. Rozdział kończy się opisem biologicznej interpretacji cech uzyskanych w module analizy danych. W rozdziale III przedstawione zostały wyniki zastosowania pipeline'u NASQQ do otwartego zbioru danych zawierających surowe widma z próbek surowicy pacjentów z dystonią rodzinną. Opisano przygotowanie surowych widm oraz odpowiadających im metadanych, wyniki modułów przetwarzania widm, a także testów w module uniwersalnym. Dodatkowo

przeanalizowano metabolity sklasyfikowane za pomocą modeli uczenia maszynowego i omówiono przecięcia szlaków metabolomicznych w oparciu o bazę KEGG. Ostatnia część pracy, Dyskusja, podsumowuje zastosowanie pipeline'u NASQQ w analizie metabolomicznej, omawiając jego zalety i ograniczenia. Autor wskazuje kierunki dalszego rozwoju i proponuje możliwe usprawnienia. Rozdział przedstawia zrównoważoną ocenę mocnych stron potoku danych i obszarów wymagających poprawy, otwierając możliwości dalszych postępów w metabolomice. Dodatkowo w formie załączników przedstawione zostały materiały uzupełniające, w tym dodatkowe dane, szczegóły metodologiczne oraz pełne wyniki, które wspierają główną treść rozprawy.

Za najbardziej wartościowe poznawczo wnioski z wykonanych badań przypadków dystonii rodzinnej (FD), uważam wykazanie w wyniku zastosowania zautomatyzowanej analizy sygnałów metabolitów z widm 1D ^1H NMR, że parametry definiowane przez użytkownika, takie jak m.in. *pulse program*, korekcja linii bazowej czy normalizacja, znacząco wpływają na wyniki analizy. Zastosowanie uczenia maszynowego poprawiło dokładność identyfikacji metabolitów, umożliwiając analizę złożonych zależności i integrację danych omicznych. Pipeline NASQQ wykazał wysoką wydajność czasową – analiza 101 próbek trwała około 5,5 godziny na standardowym komputerze, co znacznie przewyższa dotychczas stosowane podejście. Praca potwierdziła efektywność automatycznej analizy bioinformatycznej jako narzędzia dla dalszych badań i zastosowań klinicznych.

Czytając dysertację oraz załączone publikacje nasunęły mi się następujące pytania i komentarze. Choć potok (*pipeline*) NASQQ jest efektywnym narzędziem analitycznym, jego zastosowanie ogranicza się do danych NMR. Wskazane byłoby rozwinięcie metodologii o integrację z innymi technikami metabolomicznymi, takimi jak LC-MS. Brakuje walidacji proponowanych rozwiązań w szerszym kontekście klinicznym, np. poprzez testowanie ich skuteczności w diagnostyce różnych jednostek chorobowych. Rozdział metodyczny może być trudny do zrozumienia dla czytelników spoza wąskiej specjalizacji bioinformatycznej, co ogranicza jego przystępność.

Powyższe uwagi i komentarze nie wpływają na ogólną wartość merytoryczną pracy, którą oceniam bardzo wysoko. Praca doktorska mgr Łukasza Prussa ma znaczące walory nowości i oryginalności naukowej oraz reprezentuje wysoki poziom naukowy, wnosząc istotny wkład w rozwój metodologii analizy metabolomicznej. Praca wyróżnia się interdyscyplinarnym podejściem oraz integracją zaawansowanych technologii analizy danych. Pomimo drobnych ograniczeń, jest to solidna rozprawa, która spełnia wymagania stawiane przed pracami doktorskimi. Dysertacja potwierdza wysoką wiedzę i umiejętności Kandydata do stopnia naukowego doktora nauk chemicznych. Rozprawa doktorska mgr Łukasza Prussa spełnia warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (t.j. Dz.U. z 2022 r. poz. 574 z późn. zm.).

Biorąc powyższe pod uwagę wnioskuję do Rady Dyscypliny Naukowej Nauki Chemiczne Politechniki Wrocławskiej o dopuszczenie mgr Łukasza Prussa do dalszych etapów przewodu doktorskiego w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie nauki chemiczne. Jednocześnie mając na uwadze bardzo duże zaangażowanie Doktoranta w pracę badawczą, rangę problematyki i potencjał innowacyjny uzyskanych wyników oraz znaczący dotychczasowy dorobek naukometryczny Doktoranta, wnioskuję o wyróżnienie rozprawy doktorskiej.



Prof. Michał Markuszewski

Zakład Biofarmacji i Farmakokinetyki

Gdański Uniwersytet Medyczny