

## **Recenzja rozprawy doktorskiej**

**mgr Jana Majty**

**pt: „Analiza różnic w przebiegu metabolizmu węglowodanów na podstawie badań mikrobiomu jelitowego niemowląt”**

Rozprawa doktorska dotyczy mikrobiomu człowieka, wpisuje się zatem w jeden z najbardziej aktualnych kierunków badawczych mikrobiologii i biologii medycznej. Badania mikrobiomu, oprócz technik biologii molekularnej, wymagają znajomości podstawowych technik obliczeniowych i statystycznych, określanych jako narzędzia bioinformatyczne. Nie jest to dziedzina łatwa do uprawiania; połączenie umiejętności mikrobiologa i bioinformatyka wymaga zdobycia dodatkowych kwalifikacji i doświadczenia w obu obszarach badawczych. Jest to też dziedzina niezwykle interesująca dla mikrobiologa i wyzywająca dla bioinformatyka; prowadzi do stawiania nowych hipotez o etiologii chorób człowieka i często wymaga własnej inwencji w rozwijaniu nowych narzędzi obliczeniowych.

Oceniana rozprawa doktorska, wykonana w Zakładzie Biofizyki Obliczeniowej i Bioinformatyki Wydziału Biochemii, Biofizyki i Biotechnologii Uniwersytetu Jagiellońskiego nie jest pracą eksperymentalną. Autor dla celów własnych analiz korzysta z wyników eksperymentalnych zespołów badawczych z Finlandii, Estonii, Rosji i Stanów Zjednoczonych, współpracujących ze sobą w ramach projektu Unii Europejskiej pt. DIABIMMUNE Microbiome Project. Celem projektu DIABIMMUNE była weryfikacja tzw. hipotezy higienicznej, tj. określenia wpływu higieny na rozwój układu immunologicznego u niemowląt wraz z oceną ryzyka rozwoju chorób autoimmunologicznych (w tym cukrzycy typu I) w późniejszym okresie życia. W projekcie tym starannie dobrano badaną kohortę, rekrutując niemowlęta z Rosji, Finlandii i Estonii, krajów o różnym poziomie higieny. Niemowlęta te miały zwiększone ryzyko wystąpienia chorób z autoagresji ponieważ dziedziczyły określone haplotypy układu HLA (ludzkie antygeny leukocytarne). Na podstawie badań mikrobiomu układu pokarmowego niemowląt (74 dzieci z każdego kraju) oraz badań na modelach zwierzęcych wykazano słuszność weryfikowanej hipotezy, wskazując inny skład i funkcjonalności mikrobiomów niemowląt mieszkających w rozwiniętej gospodarczo Finlandii w porównaniu do dzieci pochodzących z rozwijającej się gospodarczo Rosji. Prowadziło to do niewystarczającej stymulacji układu immunologicznego przez lipopolisacharydy bakterii z rodzaju *Bacteroides* we wczesnym okresie rozwoju dzieci w Finlandii w porównaniu z silną stymulacją immunologiczną przez lipopolisacharydy pałeczki *Escherichia coli*, dominującego drobnoustroju w drogach pokarmowych dzieci zamieszkałych w Rosji.

Zgodnie z tytułem rozprawy Autor za cel swoich analiz postawił funkcjonalną ocenę metabolizmu wybranych węglowodanów w oparciu o dane sekwencjonowania metagenomów próbek kału niemowląt. Dla realizacji tego celu zastosował zestaw technik obliczeniowych prowadzących do postawienia wniosków nieopisanych dotychczas przez zespół badawczy projektu DIABIMMUNE. Doktorant do swoich analiz dobrał grupę niemowląt w wieku od 3 do 6 miesięcy; po 38 dzieci z Rosji i Finlandii. Jednakże badana kohorta, w mojej opinii, została niewystarczająco opisana w ocenianej rozprawie. W bazie NCBI pod numerem projektu PRJNA290380 (dane projektu DIABIMMUNE) zdeponowano sekwencje uzyskane w wyniku rozproszonego sekwencjonowania 785 próbek kału niemowląt z rejonu Karelii. Zdaniem recenzenta należało podać w tej rozprawie konkretne numery próbek, które uwzględniono w analizach, tym bardziej, że stanowiły one niecałe 5% sekwencji uzyskanych w trakcie realizacji projektu DIABIMMUNE. Bez znajomości konkretnych numerów próbek trudne jest powtórzenie badań wykonanych w ocenianej rozprawie. Najlepiej byłoby przedstawić te informacje w tabeli, która powinna zawierać też krótką charakterystykę tych próbek, np. wyniki analizy taksonomicznej, przedstawione w zakładce „Analysis” w Archiwum Odczytów Sekwencji (SRA) i/lub wartość wskaźnika Phred.

Sekwencje metagenomów drobnoustrojów z próbek kału były następnie poddawane wnikliwej analizie, w której Doktorant początkowo stosował algorytmy opisane już w publikacjach, a następnie rozwinął dobrane przez siebie algorytmy uczenia maszynowego. Dla szeregu taksonów występujących w próbkach kału niemowląt po analizie składu drobnoustrojów za pomocą oprogramowania Kraken 2 i Bracken zidentyfikował bakterie do poziomu rodzaju, zgodnie z metodyką opublikowaną w czasopiśmie *Scientific Reports* w 2022 roku ([doi.org/10.1038/s41598-022-12037-3](https://doi.org/10.1038/s41598-022-12037-3)) przez zespół badawczy, którego był aktywnym członkiem. Analizę liczebności genów w metagenomach bakterii z próbek kału niemowląt wykonał zgodnie z opisanymi algorytmami (HUMAnN2 lub JAMS) lub stosując własne oprogramowanie z ramach narzędzia BiomeScout, opracowanego przez Doktoranta i zespół firmy Ardigen z Krakowa. Następnie wykorzystując narzędzie do predykcyjnej analizy danych, czyli pakiet *scikit-learn*, Doktorant stosował dwa typy uczenia maszynowego: nadzorowane i nienadzorowane do analiz sekwencji metagenomów. Do przewidywania klas zastosował tu metody regresji logistycznej i k-krotnej walidacji krzyżowej. Dla sprawdzenia, jak dobrze model klasyfikujący został przeszkolony do przewidywania cech w analizach tej rozprawy Doktorant wykorzystał krzywe ROC (ang. *Receiver Operating Characteristics*) i wskaźnik AUC (ang. *area under curve*).

Analiza taksonomiczna bakterii w próbkach niemowląt z badanej kohorty w tej rozprawie potwierdziła przy zastosowaniu oprogramowania BRACKEN obserwacje projektu



DIABIMMUNE, w którym używano innego narzędzia MetaPhlAn do klasyfikacji odczytów sekwencjonowania metagenomowego. Doktorant wykazał, że w próbkach kału niemowląt rosyjskich dominował rodzaj *Bifidobacterium*, a u niemowląt fińskich skład taksonomiczny mikrobiomu wykazywał się większą różnorodnością, co potwierdził też obliczając wskaźnik różnorodności Shannona. Wykazał istotny udział rodzaju *Bacteroides* w próbkach kału od fińskich niemowląt. Analiza głównych składowych (PCA) pozwoliła na częściowe wyróżnienie próbek pochodzących z Rosji i Finlandii, jednakże populacje te nie były całkowicie od siebie oddzielone. W oparciu o powyższe różnice taksonomiczne Doktorant zbudował modele uczenia maszynowego potwierdzające możliwość przeprowadzenia klasyfikacji próbek w zależności od kraju zamieszkania niemowląt na podstawie metagenomów bakterii z próbek kału.

Analiza funkcjonalna mikrobiomów dróg pokarmowych niemowląt z rejonu Karelii przedstawiona w rozprawie pokazała, że różnice w składzie taksonomicznym pozwalają na wyosobnienie z użyciem metod uczenia maszynowego w oparciu o wskaźnik poziomu liczebności BiomeScout score czterech cech – genów w najwyższym stopniu wpływających na wielkość wskaźnika ROC AUC. Jednakże wyodrębnianie tylko tych cech do ponownej analizy mocy predykcyjnej obniżyło wartości wskaźnika ROC AUC poniżej akceptowalnego poziomu.

Metodyka zastosowana w tych obliczeniach została następnie wykorzystana przez Doktoranta do testowania hipotezy o związku pomiędzy określonym składem taksonomicznym mikrobiomu dróg pokarmowych a zdolnością niemowląt do metabolizowania mleka matki. Podobną hipotezę testowano już w projekcie DIABIMMUNE ale w oparciu o inne algorytmy obliczeniowe. W rozprawie Doktorant zidentyfikował trzy geny jako występujące w istotnie różnej liczebności pomiędzy dziećmi z grupy rosyjskiej i fińskiej, które mogą być użyte jako klasyfikatory. Są to geny kodujące beta galaktozydazę, sialidazę i alfa fukozydazę. Analogicznymi metodami przeprowadził unikalne analizy dla genów kodujących lakto-n-biozydazę i beta galaktozydazę 42A, które, jak pokazały uzyskane wyniki, zarówno osobno, jak i łącznie są bardzo dobrymi klasyfikatorami do zastosowania w uczeniu maszynowym dla osiągnięcia wyraźnego rozróżnienia mikrobiomów niemowląt z Rosji i Finlandii.

W oparciu o przedstawiony powyżej wgląd w opis metod i wyników rozprawy doktorskiej stwierdzam, że stanowi ona oryginalne rozwiązanie problemu naukowego dotyczącego optymalnego sposobu analizy wyników sekwencjonowania rozproszonego w badaniach mikrobiomów człowieka. Doktorant zademonstrował użyteczność metod uczenia maszynowego dla znalezienia cech charakterystycznych dla mikrobiomów określonych populacji niemowląt. Prowadząc własne poszukiwania optymalnych metod i algorytmów dla osiągnięcia celu pokazał też umiejętność samodzielnego prowadzenia pracy naukowej w trudnej interdyscyplinarnej dziedzinie badań mikrobiomu człowieka. Samodzielność pracy naukowej

potwierdza też rozdział Dyskusja, w którym Doktorant prawidłowo interpretuje uzyskane w rozprawie wyniki. W rozprawie odnosi się do 154 publikacji i innych źródeł naukowych. Ponad 35% stanowią pozycje piśmiennictwa z ostatnich pięciu lat.

Istotnym elementem recenzji rozprawy doktorskiej powinna być ocena ogólnej wiedzy teoretycznej Doktoranta zaprezentowanej w ocenianej pracy. Na podstawie rozdziału Wstęp oraz Dyskusja stwierdzam, że Pan mgr Majka ma wiedzę na temat metod biologii molekularnej, mikrobiomów i ich roli w zdrowiu człowieka, złożoności badań tych konsorcjów drobnoustrojów a także swobodnie wykorzystuje swoje doświadczenie w stosowaniu metod obliczeniowych do analizy wyników sekwencjonowania pełnogenomowego. Zatem spełnia wspomnianą powyżej przesłankę wymaganą od kandydatów ubiegających się o nadanie stopnia doktora.

Uwagi krytyczne:

1. W wielu miejscach rozdziałów Wstęp i Dyskusja styl wypowiedzi budzi wątpliwości. Recenzent zdaje sobie sprawę, że bezpośrednie wyrażenie terminów angielskich w języku polskim jest trudne, ale pewne zwroty i zdania są niezrozumiałe dla czytelnika, np.

Str. 7 – „niskim ogólnym ładunkiem mikroorganizmów”,

Str. 10 – „relacja bakterii przewodu pokarmowego z innymi elementami obszaru analiz, którego popularność obecnie dynamicznie rośnie”,

Str. 14 – „sto lat temu mikrobiom niemowlęcy podlegał innej kolonizacji w okresie niemowlęcym niż obecnie”,

Str. 16 – „bakterie rodzaju *Bifidobacterium* ko-ewoluowały w kontekście ludzkich oligosacharydów mleka matki”,

2. W mojej opinii sformułowanie celu pracy nie jest wystarczająco precyzyjne. Autor rozprawy rzeczywiście swoją aktywność badawczą poświęcił poszukiwaniom wydajnych sposobów analizy olbrzymiego zbioru danych metagenomicznych; natomiast analizy prowadzi się w celu potwierdzenia lub wykluczenia hipotezy badawczej. Najbliższa testowanej w tej rozprawie hipoteza badawcza przedstawiona jest dopiero na stronie 79 w rozdziale Dyskusja i brzmi, tu cytuję: „Czy mimo pokrewieństwa taksonomicznego, szczepy *Bifidobacterium* typowe dla obu populacji różnią się między sobą na poziomie funkcjonalnym?”. Niestety w dalszej części rozdziału Dyskusja, jak i we wnioskach czytelnik nie znajduje udzielonej wprost odpowiedzi na to pytanie. Prosiłabym zatem Doktoranta o uwzględnienie weryfikacji tej hipotezy w prezentacji pracy w trakcie obrony.



3. W rozdziale Wnioski zawarte jest raczej podsumowanie wyników. Proszę o przedstawienie wniosku/ wniosków, jaki można wyprowadzić z uzyskanych wyników w trakcie obrony pracy doktorskiej.

4. Spis Rysunków i Spis Tabel odnosi się raczej do innej wersji pracy i nie pokrywa się z numeracją stron rozprawy otrzymanej do oceny przez recenzentów.

5. Autor rozprawy błędnie odmienia nazwy łacińskie taksonów bakterii w wielu miejscach rozprawy. Np.

Str. 48 – „Poza *Bifidobakteriami* w mikrobiomie jelitowym rosyjskich niemowląt identyfikowane są istotne populacje *Escherichii* i *Enterococcusa*..”,

Str. 77 – „rodzaj *Bifidobacteria* z typu *Aktynobacteria*” – ponadto nazwy obu taksonów napisane są błędnie.

Proszę zatem Doktoranta o przedstawienie zasad pisowni łacińskich nazw taksonów bakterii w języku polskim w trakcie obrony.

Powyższe uwagi nie zmieniają pozytywnej oceny tej rozprawy doktorskiej. Spełnia ona warunki stawiane rozprawom doktorskim zgodnie z artykułem 13 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz. U. z 2003 r. Nr 65, poz. 595 z późniejszymi zmianami).

W związku w tym wnoszę do Wysokiej Rady Dyscypliny Nauki Biologiczne Uniwersytetu Jagiellońskiego o dopuszczenie Pana mgr Jana Majty do dalszych etapów postępowania o nadanie stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie nauki biologiczne.

Prof. dr. hab. Elżbieta Anna Trafny

Centrum Inżynierii Biomedycznej IOE WAT

Warszawa, 2 sierpnia 2023 r.