

Wrocław. 10.11.2023

Joanna Szyda

Uniwersytet Przyrodniczy we Wrocławiu

Katedra Genetyki, Pracownia Biostatystyki

Wiodący Zespół Badawczy: THETA - The Biostatistic Group

Recenzja pracy doktorskiej Pana mgr inż. Michała Antkowiaka
„Geny otyłości: analiza wariantów strukturalnych w genomie psa domowego
(*Canis lupus familiaris*)”
wykonanej pod kierunkiem Pana prof. dr hab. Macieja Szydłowskiego

Wykorzystanie informacji o polimorfizmach genomowych do opisu genetycznej determinacji cech fenotypowych jest podejściem stosowanym od dawna w odniesieniu do różnych populacji. Obecnie najbardziej rozpowszechnionym wariantem analizy jest wykorzystywanie do tego celu polimorfizmów pojedynczego nukleotydu, ang. *Single Nucleotide Polymorphism* (SNP) z uwagi na stosunkową łatwość identyfikacji tego typu polimorfizmów w genomach oraz ich powszechność, dochodzącą do kilku(nastu) milionów SNP w genomie pojedynczego osobnika. W recenzowanej pracy wykorzystano natomiast inny rodzaj polimorfizmu – polimorfizmy strukturalne, obejmujące delecje i duplikacje odcinków genomu składające się nawet z kilkuset par zasad. Co prawda liczba takich polimorfizmów w genomach ssaków jest o wiele niższa niż liczba SNP, jednakże z uwagi na to że obejmują one długie regiony chromosomów, ich potencjalny wpływ na fenotyp jest bardziej znaczący niż efekt pojedynczego SNP, szczególnie gdy polimorfizmy strukturalne pokrywają regiony genomu kodujące białka. Uważam więc, że wykorzystanie polimorfizmów strukturalnych do analizy genetycznego podłoża otyłości psów jest bardzo ważnym atutem pracy i stanowi nowatorski temat badawczy.

Sama rozprawa doktorska składa się z wprowadzenia do tematyki, opisu materiału oraz zastosowanej metodyki laboratoryjnej i bioinformatycznej, uzupełnionej dyskusją, jak również z trzech publikacji naukowych. Przedstawiona koncepcja badawcza jest spójna, dobrze zdefiniowana i stanowi logiczny ciąg analiz genetycznego uwarunkowania otyłości psów, prowadzonych od (i) wstępnej oceny powiązania zmienności liczby kopii w pojedynczym genie (*AMY2B*) z poziomem otyłości psów rasy labrador retriever, przez (ii) analizę zmienności liczby kopii w obrębie sekwencji tego samego genu u innych gatunków psowatych: lisa rudego (*Vulpes vulpes*), lisa polarnego (*Vulpes lagopus*) i jenota azjatyckiego (*Nyctereutes procyonoides*), aż do (iii) finalnego etapu analizy zmienności polimorfizmów liczby kopii w sześciu wybranych genach (*ALPL*, *KCTD8*, *SGSM1*, *SLC12A6*, *RYR3*, *VPS26C*). Te trzy etapy badawcze stanowią logicznie powiązaną koncepcję analityczną.

Tekst rozprawy doktorskiej jest bardzo płynnie napisany i przyjemnie się go czyta, co jest bardzo ważnym, „miękkim”, atutem pracy, jednakże Autor nie ustrzegł się od popełnienia kilku błędów językowych. Ponadto, w aspekcie technicznym uważam, że umieszczenie wszystkich tabel na końcu tekstu utrudnia korzystanie z informacji w nich zawartych w trakcie czytania.

Treść **Wstępu** stanowi adekwatne prowadzenie do tematu uwarunkowania otyłości psów, ze szczególnym naciskiem na jej genetyczną determinację. ① W części dotyczącej definicji wariantów strukturalnych, Autor definiuje SV jako mutacje obejmujące do kilku tysięcy par zasad, podczas gdy SV mogą „rozciągać się” nawet na kilkaset tysięcy par zasad.

Materiał pracy z jednej strony stanowi bardzo małą próbę danych, obejmującą, w zależności od analizy, od kilkudziesięciu do 109 osobników, z drugiej strony dotyczy przedstawicieli rodziny psowatych, które nie są tak częstym przedmiotem analiz genomicznych jak zwierzęta hodowlane, gatunki laboratoryjne, czy też ludzie. Uważam to za interesujący aspekt pracy. Jednakże w kontraście do bardzo płynnie napisanego wstępu, w rozdziale **Materiał i metody** nie wszystkie etapy analizy oraz wykorzystany w poszczególnych etapach materiał badawczy zostały jasno zdefiniowane w tekście rozprawy. Dla przykładu, podrozdział *Metody badawcze* nieprecyzyjnie opisuje jakie analizy statystyczne i dla jakiego materiału zostały przeprowadzone. Staje się to dopiero jasne po przeczytaniu poszczególnych publikacji. Uważam, że w tym rozdziale rozprawy powinien znaleźć się skrótowy opis zastosowanego potoku bioinformatycznego przetwarzania danych oraz ich wstępnej edycji. Również ani w tekście

rozprawy doktorskiej, ani też w samej publikacji opisującej analizy etapu III nie podano szczegółowych informacji o materiale. Dane te są kluczowe dla interpretacji jakości uzyskanych wariantów strukturalnych. Konkretnie, (i) nie podano czy / jakie były zasady edycji surowych danych oraz jakich parametrów i oprogramowania użyto do edycji, (ii) nie wiadomo jaka część surowych danych została wykorzystana na etapie detekcji polimorfizmów strukturalnych, czyli jaki % danych został odrzucony z powodu niewystarczającej jakości odczytów, (iii) nie podano również jakie jest średnie pokrycie genomu odczytami po przyrównaniu do genomu referencyjnego, a jest to bardzo ważna informacja w kontekście zastosowania programów do detekcji wariantów strukturalnych wykorzystujących podejście *split read*. Domyślam się, że większość tych informacji można znaleźć w publikacji Szydłowski i Antkowiak (2022), jednak z uwagi na ich ważność dla analizy będącej przedmiotem badań doktorskich, uważam, że powinny one zostać włączone w tekst rozprawy. Największy niedosyt informacyjny pozostawia podrozdział 5.2.5 (*Analiza statystyczna*), w którym przeprowadzone analizy zostały opisane bardzo skrótowo, bez wyraźnego podziału na to, które metody zastosowano do jakiego materiału. Specyficzne aspekty przeprowadzonej analizy, o których chciałabym podyskutować w trakcie obrony są następujące:

② W części rozprawy doktorskiej opisującej materiał zaskakuje zastosowanie różnych kryteriów podziału osobników na normalne i otyłe w zależności od badania – różnice między Analizą I i Analizą III. Co więcej, w tekście rozprawy nie podano uzasadnienia dla zmiany w klasyfikacji osobników.

③ Po przeczytaniu opisu materiału niejasnym pozostaje fakt wyboru jedynie kilku genów do dalszej analizy, pomimo dostępności sekwencji całych genomów dla osobników. Analiza wszystkich wykrytych wariantów strukturalnych stanowiłaby bardziej kompleksowe podejście do próby zrozumienia genetycznej determinacji otyłości psów. W rozprawie brakuje uzasadnienia dla braku zastosowania takiego podejścia.

④ W opisie analizy w rozdziale 5.2.1 podano, że w modelu wykorzystano również efekty składowych głównych. Jednakże poniżej, w rozdziale 5.2.5 te efekty nie znalazły się na liście podanych zmiennych niezależnych modelu. Co więcej, ani w tekście rozprawy, ani w publikacji opisującej III-ci etap badań nie zdefiniowano jakie dane zostały wykorzystane do przeprowadzenia

analizy składowych głównych. Można się jedynie domyślać, że wykorzystano w tym celu polimorfizmy pojedynczych nukleotydów (SNP), określone w publikacji jako *short variants*. Tutaj również nie podano potoku analizy bioinformatycznej wykorzystanego do detekcji SNP, ani też podstawowych statystyk opisujących np. liczbę SNP oraz zasady ich filtracji.

⑤ W rozdziale 5.2.5 podano iż „zastosowano ogólny model liniowy z masą ciała i modelem regresji logistycznej”. Jest to niejasne sformułowanie. Ponadto, regresja logistyczna jest już sama w sobie modelem nieliniowym ponieważ obejmuje transformację zmiennej zależnej przy użyciu funkcji logit.

⑥ W części III badania analizowano rozdzielnie powiązanie występowania pojedynczego wariantu strukturalnego w wybranych genach. Takie podejście traktujące każdy gen jako odrębną zmienną objaśniającą implikuje monogeniczną determinację otyłości i ignoruje ewentualny poligeniczny / oligogeniczny model genetycznej determinacji tej cechy.

Kolejny rozdział, **Dyskusja**, jest napisany ciekawie i czyta się równie płynnie jak Wstęp. Jednakże tekst tego rozdziału koncentruje się jedynie na aspekcie biologicznym uzyskanych wyników analizy. Natomiast brakuje mi w tym miejscu uwzględnienia następujących aspektów związanych ze statystyczną analizą danych, w szczególności:

⑥ dyskusji nad aspektem założonego modelu dziedziczenia otyłości, która jako cecha mierzona *de facto* w skali ciągłej, mogłaby być determinowana przez dużą liczbę genów posiadających dodatkowo oddziaływania epistatyczne;

⑦ dyskusji na temat ograniczeń metodycznych jakie wynikają z małych rozmiarów próby danych, a co za tym idzie (i) z małej mocy detekcji potencjalnie istniejącej asocjacji wariantu strukturalnego z klasą otyłości oraz (ii) niskiej dokładności uzyskanych estymatorów.

W treści mojej recenzji zawarłam wiele uwag krytycznych, o których chciałabym podyskutować z Autorem w czasie rozprawy. Dla ułatwienia każda uwaga została opatrzona odrębnym numerem ① - ⑦. Jednakże podkreślam, że przedstawione przeze mnie uwagi krytyczne w żaden sposób nie dyskredytują pracy i stanowią jedynie przyczynek do dyskusji akademickiej. Uważam pracę za bardzo ciekawy temat, który przez Autora został przeanalizowany w logicznie

zdefiniowanych etapach wskazujących na dojrzałą koncepcję naukową. Za szczególnie wartościowe aspekty uważam wykorzystanie polimorfizmów strukturalnych, które są trudniejsze w detekcji niż polimorfizmy punktowe oraz zmierzenie się Autora z problematyką analizy małej próby danych. Są to nie tylko trudne, ale i nowatorskie aspekty pracy. W podsumowaniu uważam, że przedstawiona do recenzji praca spełnia wszystkie wymagania stawiane dysertacji doktorskiej, w szczególności zasady zdefiniowane w **Ustawie o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki** z 03.07.2018. Jednocześnie stawiam wniosek o dopuszczenia Pana mgr inż. Michała Antkowiaka do dalszych etapów przewodu doktorskiego.



Joanna Szyda

