

Münster, 8/30/16

**Recenzja pracy doktorskiej Pani mgr Magdy Mielczarek pt. “The genome-wide distribution of copy number variations in various breeds of domestic cattle (*Bos taurus* Linnaeus, 1758) based on the next-generation sequencing data”**

Praca doktorska Pani mgr Magdy Mielczarek poświęcona jest analizie szczególnego rodzaju polimorfizmu DNA w genomach bydła domowego, a mianowicie zmienności liczby kopii pewnych odcinków DNA (CNV - copy number variation) w analizowanych genomach. Praca została wykonana w Zakładzie Genetyki Uniwersytetu Przyrodniczego we Wrocławiu pod kierunkiem prof. dr hab. Joanny Szydy.

Przełom XX i XXI wieku to okres niezwykłego rozwoju nauk biologicznych. W niesłychanym tempie pojawiają się nie tylko nowe odkrycia, ale także nowe metody, szczególnie te w biologii molekularnej, jak i innowacyjne technologie pozwalające na prowadzenie badań w zupełnie nowym wymiarze. W zalewającym nas tsunami danych molekularnych, możemy mówić o “złotej epoce” dla bioinformatyków. Niestety, skupiając się na przetwarzaniu gromadzących się danych bioinformatyk bardzo często staje się kimś w rodzaju naukowego rzemieślnika. Na szczęście pani mgr Magda Mielczarek nie wpadła w pułapkę “bioinformatycznej rutyny” i nie zapomina o biologii, co jest bliskie mojemu sercu. Zazwyczaj na pierwszym wykładzie z bioinformatyki, dla każdej grupy studentów, czy to biologii czy informatyki, przedstawiam swoje osobiste bioinformatyczne “credo” i “pamiętaj o biologii” jest jego głównym punktem.

Mgr Mielczarek zaprezentowała swoje badania w formie “klasycznej” rozprawy, a nie zbioru opublikowanych artykułów, i w związku z tym mamy przed sobą pracę składającą się z kilku rozdziałów. Pierwszy rozdział zwięźle, acz elegancko wprowadza nas do tematu. W tym rozdziale czytelnik może się dowiedzieć jak wyglądają dane otrzymane bezpośrednio z sekwenatorów nowej generacji i jakie się z tym wiążą wyzwania. Następnie przedstawiony zostaje główny bohater rozprawy, a mianowicie zmienność liczby kopii odcinków genomowego DNA. Rozdział ten kończy przedstawienie zakresu i motywacji badań związanych z prezentowaną pracą dokorską.

W następnych dwóch rozdziałach Pani mgr Mielczarek prezentuje materiały i metody wykorzystane w pracy. Zarówno dane dotyczące materiału (sekwencji DNA) jak i metod użytych przez doktorantkę są opisane w sposób wystarczający aby ewentualnie móc

powtórzyć wszystkie analizy, a więc zgodnie z naukowymi standardami. Coś co może się wydawać sprawą trywialną wcale taką nie jest, gdyż w ostatnich latach obserwujemy niepokojący trend opisywania metod, zwłaszcza tych bioinformatycznych, w sposób nieadekwatny i nie pozwalający na powtórzenie eksperymentów. Z radością spostrzegam, że doktorantka nie ulega temu trendowi i opisuje metody bardzo precyzyjnie, zwłaszcza te najtrudniejsze, np. testy statystyczne.

Następnie Pani mgr Mielczarek prezentuje wyniki swoich analiz. Okazuje się, że nie wszystkie zsekwencjonowane genomy nadawały się do przedstawionych analiz, ze względu na zbyt małe pokrycie genomu odczytami sekwencyjnymi. W ten sposób doktorantka "straciła" dziewięć genomów. Niemniej, pozostałe 146 genomów w zupełności wystarczyły do solidnej analizy CNV. Doktorantka użyła dwóch różnych programów do detekcji wariantów. Ponieważ każdy z nich używa innego algorytmu, nie jest niespodzianką, że otrzymane wyniki nie są zgodne ze sobą, aczkolwiek niespodzianką, dla mnie osobiście jest fakt, że wyniki są znacząco różne w każdym aspekcie. Prawdę powiedziawszy, zabrakło mi w rozprawie opinii doktorantki na ten temat i mam nadzieję, że ją usłyszę na obronie pracy. Doktorantka, jako zweryfikowane uznała te regiony CNV, które zostały wykryte przez oba programy. Wydaje mi się, że dla większości biologów nie jest to idealna metoda. Szkoda, że doktorantka nie pokusiła się o prostą weryfikację wybranych CNV z użyciem metod molekularnych, np. PCR lub sekwencjonowanie długich cząsteczek, choć zdają sobie sprawę, że taka weryfikacja mogła być niedostępna dla doktorantki ze względu na ograniczenia laboratoryjne. Chciałbym też zwrócić uwagę, że statystyka odkrytych CNV jest trochę zafałszowana ze względu na arbitralne wybranie minimalnych i maksymalnych długości analizowanych CNV. Chętnie bym usłyszał od doktorantki, co nią kierowało przy wyborze konkretnych wartości. Dla mnie jako biologa ewolucyjnego najciekawsza jest część poświęcona analizie funkcjonalnej. Na przykład jeden z odcinków oznaczonych jako mający potencjalnie duży wpływ na białka jest odcinek zlokalizowany w rejonie BTA28 w obrębie genum 5S rRNA. Niemniej, znając organizację genów 5S rRNA w genomach ssaków, duży wpływ tego polimorfizmu na białka wydaje się być mało prawdopodobny, gdyż genom ssaków zawiera kilkaset kopii tychże genów i jedna kopia nie powinna stanowić istotnej różnicy.

Po przedstawieniu wyników doktorantka dokonała dyskusji tychże i muszę przyznać, że jestem pod wrażeniem tego rozdziału. Pani Mielczarek wykazała się w nim znajomością tematu z uwzględnieniem bieżącej literatury. Umiejętnie dyskutuje uzyskane wyniki w świetle rezultatów pochodzących z innych badań. Według mnie jest to najmocniejsza część zaprezentowanej rozprawy. Po dyskusji doktorantka w krótkim acz adekwatnym paragrafie podsumowuje swoją rozprawę. W tej części rozprawy zabrakło mi rozdziału "Perspektywy". Dlatego też, chciałbym usłyszeć od doktorantki podczas obrony, co ona myśli na ten temat. Czy uważa ten temat za zakończony? Jeśli tak, to dlaczego? A jeśli nie, to co można by jeszcze zrobić aby przedstawione analizy były pełniejsze i w jakim kierunku powinny pójść badania genomów była.

Poniżej przedstawiam kilka punktów do dyskusji/wyjaśnienia lub po prostu zaspokojenia czystej naukowej ciekawości.

1. Doktorantka zanalizowała genomy 13 ras bydła. Które z tych ras są hodowane w Polsce i jak często?
2. Dopasowanie (alignment) odczytów sekwencyjnych do genomu jest chyba najważniejszą częścią całej analizy. Istnieje wiele metod i programów opracowanych w tym celu. Doktorantka wybrała program BWA-MEM. Dlaczego? Czy przetestowała pani również inne metody?
3. Na stronie 24 stwierdza pani, że ponad 95% odczytów sekwencyjnych dla 144 jednostek mogły być dopasowane do genomów i tylko w jednym przypadku ten odsetek był niższy. W sumie daje nam to 145 prób sekwencyjnych. Natomiast w innych miejscach wspomina pani o 155 próbach. Co się stało z 10 próbami?

Podsumowując stwierdzam, że rozprawa pani mgr Magdy Mielczarek jest istotna nie tylko z punktu widzenia poznawczego, ale jej wyniki mogą także być wykorzystane praktycznie. Praca spełnia wszystkie ustawowe i zwyczajowe wymogi i stawiam wniosek o dopuszczenie doktorantki do dalszych etapów przewodu doktorskiego.



Prof. Dr. Wojciech Makałowski