



UNIwersytet Warszawski

Instytut Informatyki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
POLSKA

dr hab. Bartosz Wilczyński
profesor uczelni
Phone: +(48 22) 5544 577
Fax: +(48 22) 5544 400
e-mail: bartek@mimuw.edu.pl

Warszawa, 22. maja 2023 r.

Recenzja rozprawy doktorskiej pt. „Modeling of transcription factors influence on gene expression based on data obtained using next-generation sequencing methods ” przedstawionej przez mgr inż. Macieja Migdała

Recenzja niniejsza została sporządzona na zlecenie Rady Dyscypliny Nauk Medycznych, Warszawskiego Uniwersytetu Medycznego zgodnie z wymogami ustawy dotyczącej procedur nadawania stopnia doktora.

Opis rozprawy

Przedstawiona rozprawa napisana jest w języku angielskim i jest złożona z trzech publikacji, których współautorem jest doktorant, poprzedzonych krótkim wprowadzeniem i zakończonych podsumowaniem.

Artykuły wchodzące w skład rozprawy to:

1. Pawlak M, Kędzierska KZ, Migdał M, Nahia KA, Ramilowski JA, Bugajski Ł, Hashimoto K, Marconi A, Piwocka K, Carninci P, Winata CL. *Dynamics of cardiomyocyte transcriptome and chromatin landscape demarcates key events of heart development*. **Genome Res.** 2019 Mar;29(3):506-519.
2. Migdał M, Tralle E, Nahia KA, Bugajski Ł, Kędzierska KZ, Garbicz F, Piwocka K, Winata CL, Pawlak M. *Multi-omics analyses of early liver injury reveals cell-type-specific transcriptional and epigenomic shift*. **BMC Genomics.** 2021 Dec18;22(1):904
3. Migdał M, Arakawa T, Takizawa S, Furuno M, Suzuki H, Arner E, Winata CL, Kaczkowski B. *score: an R package for inference of gene expression regulators*. **BMC Bioinformatics.** 2023 Jan 11;24(1):14.

Według załączonych do rozprawy oświadczeń współautorów, procentowy wkład doktoranta w przedstawione prace, to odpowiednio 10 % (1), 40 % (2) i 45 % (3), jednak z mojego punktu widzenia istotniejszym niż deklarowane procentowe udziały jest fakt, że doktorant jest pierwszym autorem dwóch spośród przedstawionych prac (2 i 3).

Prace te są powiązane tematycznie - łączy je problem bioinformatycznej analizy i opisu procesu regulacji transkrypcyjnej genów. Wszystkie te prace powiązane są z nowoczesnymi badaniami regulacji genów na podstawie danych z sekwencjonowania RNA (profile ekspresji genów z RNA-Seq), sekwencjonowania DNA po immunoprecypitacji chromatyny (miejsca wiązania czynników transkrypcyjnych z ChIP-Seq), badania otwartości/dostępności chromatyny metodą ATAC-Seq a także różnych podejść do przewidywania wiązania czynników transkrypcyjnych na podstawie motywów sekwencyjnych. Wszystkie te metody należą do najnowocześniejszych obecnie stosowanych narzędzi w badaniach nad regulacją genów. W przedstawionych pracach są stosowane zgodnie ze stanem wiedzy w dziedzinie.

W pracy (1), udział mgra Migdała jest najmniejszy, wg opisu w podsumowaniu, głównie polegał na przygotowaniu narzędzi do analizy danych ATAC-Seq i wykonaniu dość podstawowych analiz bioinformatycznych zebranych danych eksperymentalnych. Nie należy jednak zapominać, że laboratorium genomiki rozwoju Danio Pręgowanego było jednym z pierwszych w Polsce miejsc, gdzie technika ATAC-Seq była wykorzystywana, więc udział doktoranta w tym projekcie na pewno był bardzo korzystny dla jego rozwoju naukowego i pozwolił mu rozwinąć swoje umiejętności.

W pracy (2), gdzie doktorant jest już pierwszym autorem, przeprowadzono kompleksową analizę danych eksperymentalnych dla kilku rodzajów komórek związanych z uszkodzeniami wątroby przez toksyny. Analizy te obejmowały między innymi różnicową analizę ekspresji genów w kilku typach komórek, klastrowanie metodą SOM (ang. self-organizing maps), analizę składowych głównych i analizę wzbogacenia motywów miejsc wiązania czynników transkrypcyjnych. Wszystkie te analizy, pozwoliły autorom na postawienie kilku hipotez dotyczących prawdopodobnych czynników transkrypcyjnych zaangażowanych w regulację tego procesu.

Praca (3) jest artykułem metodologicznym, opisującym narzędzia `xcore` i `ixcoredata`, które pozwala na wykorzystanie modelu liniowego do analizy ekspresji genów w kontekście wiązania czynników transkrypcyjnych w promotorach genów. Oprócz prezentacji samego narzędzia, autorzy prezentują wyniki zastosowania tego narzędzia do danych eksperymentalnych, zarówno przy użyciu informacji o wiązaniu pochodzących z eksperymentów ChIP-Seq jak i miejsc wiązania przewidywanych na podstawie motywów sekwencyjnych. Wyniki korelacji pomiędzy przewidywaną ekspresją a zaobserwowaną są lepsze dla danych ChIP-Seq, co jest raczej spodzewanym wynikiem.

Oprócz prac stanowiących rozprawę, widać, że doktorant angażuje się w inne projekty naukowe, co nie stanowi podstawy do oceny doktoratu, ale dobrze świadczy o jego dojrzałości naukowej.

Uwagi krytyczne

Rozprawa składa się z 3 prac, blisko związanych tematycznie i wszystkich opublikowanych w dobrych czasopismach międzynarodowych, co świadczy o tym, że ich rzetelność została

już zweryfikowana przez niezależnych recenzentów.

Nie mam wielu uwag krytycznych, ale zwróciłem uwagę na 2 kwestie, które być może można będzie przedyskutować podczas obrony:

Pierwsza, dość powierzchowna kwestia, wiąże się z tym, że autor stwierdza, że narzędzia wykonane przez niego na użytek pracy (1) są dostępne w repozytorium

https://gitlab.com/zdglab/atacseq_pipeline, jednak nie udało mi się dotrzeć do tych materiałów.

Druga kwestia wiąże się z modelem stosowanym w narzędziu xcore. Autorzy proponują tam model liniowy jako wyjaśnienie ekspresji RNA. Jakkolwiek model liniowy jest popularny w zastosowaniach, ze względu na prostotę stosowania, wydaje się, że dla modelowania ekspresji genów lepsze mogłyby być uogólnione modele liniowe (ang. generalized linear models), które przetwarzają sygnały wejściowe przy pomocy funkcji nieliniowej, co mogłoby dać zdecydowanie lepsze współczynniki korelacji z danymi eksperymentalnymi niż te prezentowane w pracy (3). Ciekawi mnie, czy autorzy rozważali zastosowanie uogólnionych modeli liniowych w metodzie xcore, czy może są jakieś obiektywne przesłanki, które w tym przeszkadzają.

Podsumowanie

Podsumowując, praca doktorska mgr inż. Macieja Migdała stanowi bardzo dobry przykład, jak praca bioinformatyka może stanowić istotny wkład w wiodące badania naukowe w naukach biomedycznych. Doktorant opanował niewątpliwie duży zasób wiedzy nt. metodologii badań w genomice funkcjonalnej i zastosował swoją wiedzę w kilku projektach, które z całą pewnością stanowią nowy wkład do naszego poznania procesów regulacji genów. W szczególności praca (2), gdzie autorzy zajmują się zmianami ekspresji genów powodowanymi przez uszkodzenia wątroby na pewno może stanowić przyczynek do poprawienia naszego rozumienia procesów biologicznych, które mogą mieć także znaczenie medyczne.

Prace, których współautorem jest doktorant zostały opublikowane w dobrych i bardzo dobrych czasopismach naukowych i obejmują duży zakres metod, które opanował doktorant.

W związku z tym uważam, że **rozprawa doktorska spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668)** i może zostać skierowana do kolejnych etapów postępowania w kierunku nadania stopnia doktora. Co więcej, w związku tym, że autor w sposób wyjątkowo przemyślany łączył rozwój nowych metod bioinformatycznych z ich zastosowaniami, proponuję rozważyć **wyróżnienie rozprawy doktorskiej mgra inż. Macieja Migdała.**

Bartosz Wilczyński