

dr hab. Michał Dąbrowski
Pracownia Neurobiologii Molekularnej
Instytut Biologii Doświadczalnej
im. M. Nenckiego PAN

Recenzja rozprawy doktorskiej

Tytuł rozprawy: Modeling of transcription factors influence on gene expression based on data obtained using next-generation sequencing methods

Autor: mgr inż. Maciej Migdał

Rozprawa doktorska została przygotowana w Laboratorium Genomiki Rozwoju Danio Pręgowanego w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej. Promotorem rozprawy doktorskiej jest dr hab. Cecilia Winata.

Zgłoszona do oceny rozprawa doktorska ma postać zbioru opublikowanych i powiązanych tematycznie artykułów naukowych, wraz ze streszczeniami w języku polskim i angielskim oraz autoreferatem w języku angielskim, zgodnie z art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668).

Na rozprawę doktorską składają się trzy artykuły: Pawlak i wsp. "Dynamics of cardiomyocyte transcriptome and chromatin landscape demarcates key events of heart development", *Genome Res.* 2019; Migdał i wsp. "Multi-omics analyzes of early liver injury reveals cell-type specific transcriptional and epigenetic shift", *BMC Genomics*, 2021; oraz Migdał i wsp. "xcare: an R package for inference of gene expression regulators", *BMC Bioinformatics*, 2023.

Zgodnie z art. 192 ww. ustawy, dotyczącym publikacji wieloautorskich, w powiązaniu z Uchwałą Nr 68/2020 Senatu Warszawskiego Uniwersytetu Medycznego z dnia 14.12.2020 r., do każdej z prac włączonych do rozprawy dołączone są oświadczenia współautorów, które określają indywidualny wkład doktoranta w publikację oraz udział każdego z autorów (merytoryczny i procentowy).

Wspólnym tematem trzech publikacji zawartych w rozprawie doktorskiej jest analiza współregulacji transkrypcyjnej genów przez czynniki transkrypcyjne na podstawie danych transkryptomicznych i epigenomicznych.

W pierwszej pracy (Pawlak i wsp. 2019), opublikowanej w renomowanym czasopiśmie *Genome Research*, autorzy zbadali ekspresję genów i otwartość chromatyny w kardiomiocytach typu dzikiego wyselekcjonowanych z transgenicznym pod względem *nkx2.5:EGFP* i *myl7:EGFP* zarodków danio przegowanego w trzech punktach czasowych podczas rozwoju embrionalnego, a także w kardiomiocytach z nokautem *Gata5*, *Hand2* i *Tbx2*. Na potrzeby tej pracy doktorant przeprowadził analizę motywów wiążących czynniki transkrypcyjne w promotorach klastrów współeksprymowanych genów i w regionach chromatyny, które zmieniły otwartość podczas rozwoju. Umożliwiło to wnioskowanie o potencjalnych sieciach regulatorowych genów, a tym samym identyfikację prawdopodobnych kluczowych regionów cis-regulatorowych, czynników transkrypcyjnych i ich interakcji z genami docelowymi podczas rozwoju serca.

W drugiej publikacji (Migdał i wsp. 2021, *BMC Genomics*) autorzy zbadali u dorosłych osobników danio przegowanego odpowiedź komórek wątroby, na poziomie transkryptomu i zmian otwartości chromatyny, na przewlekłe podawanie hepatotoksycznego związku tioacetamidu (TAA). Autorzy zbadali odpowiedź trzech głównych populacji komórek wątroby: hepatocytów, komórek gwiazdzistych wątroby i komórek śródbłonka. Populacje te zostały wysortowane z transgenicznym dorosłych osobników danio przegowanego w oparciu o markery każdego z tych typów komórek. Wyniki pokazują, że populacją komórek w której doszło do największych zmian były komórki śródbłonka, a ponadto, że tylko w niektórych genach zmianom transkrypcyjnym towarzyszyły zmiany otwartości chromatyny regionów promotorowych. Doktorant jest wiodącym pierwszym autorem tej publikacji i głównym autorem odpowiedzialnym za bioinformatykę i analizy statystyczne, które obejmowały: wykorzystanie samoorganizujących się map (SOM) do analizy ekspresji genów, nowatorską analizę danych ATAC-seq oraz analizę zawartości i nadreprezentacji motywów wiążących czynniki transkrypcyjne.

W trzeciej publikacji (Migdał i wsp. 2023, *BMC Bioinformatics*) autorzy opracowali, zwalidowali i udostępnili środowisku naukowemu oprogramowanie *xcare* do analizy regulacji transkrypcyjnej genów, w postaci pakietu w języku statystycznym R. Pakiet *xcare* implementuje regularyzowany model regresji liniowej, a mianowicie regresję grzbietową, aby opisać związek pomiędzy, szacowaną na podstawie modelu, aktywnością czynników transkrypcyjnych a poziomami ekspresji wszystkich genów. Nowatorstwo podejścia zastosowanego w *xcare* polega na wykorzystaniu danych ChIP-seq z dużych publicznych repozytoriów (ReMap i ChIP-Atlas) na temat wiązania wszystkich zbadanych w dowolnym typie komórek czynników transkrypcyjnych, zamiast przeważnie stosowanej zawartości motywów wiążących czynniki transkrypcyjne, jako części danych wejściowych, których drugą część stanowią dane o ekspresji genów mierzone za pomocą RNA-seq, a także CAGE. Za pomocą walidacji krzyżowej autorzy pokazują, że model regresji dopasowany do danych ChIP-seq daje lepszą zgodność przewidywań z obserwacjami niż model

dopasowany do danych o motywach wiążących czynniki transkrypcyjne. Warto zauważyć, że dopasowanie było znacznie lepsze dla ekspresji mierzonej CAGE, co wskazuje, że wiązanie czynników transkrypcyjnych w promotorach reguluje przede wszystkim inicjację transkrypcji. Stosując *xcore* do własnych danych z dwóch linii komórek nowotworowych traktowanych TGFβ, autorzy wykazują, że metoda prawidłowo identyfikuje znane czynniki transkrypcyjne regulujące przejście nabłonkowo-mezenchymalne (EMT). Wyraźną zaletą *xcore* jest możliwość włączenia do modelu regulatorów, które tylko pośrednio wiążą się z DNA, takich jak enzymy epigenetyczne. Ogólnie rzecz biorąc, *xcore* jest obiecującym nowym narzędziem do analizy współregulacji transkrypcji u gatunków dokładnie zbadanych ChIP-seq. Doktorant jest wiodącym pierwszym autorem tej publikacji oraz autorem oprogramowania *xcore*.

Poprawność i oryginalność prac składających się na rozprawę została potwierdzona przez recenzentów na etapie ich publikacji, stąd też moje uwagi dotyczą głównie autoreferatu. Autoreferat podzielony jest na cztery części: „Wstęp” (8 stron), „Założenia i cel pracy” (1 strona), „Podsumowanie i wnioski” (2 strony), „Piśmiennictwo” (48 pozycji na 5 stronach). Autoreferat we właściwy sposób podsumowuje wyniki trzech artykułów składających się na rozprawę doktorską i precyzyjnie wskazuje indywidualny wkład doktoranta w powstanie każdego artykułu. Niektóre fragmenty autoreferatu są zbędne, jak np. podanie założeń rozprawy doktorskiej, zważywszy, że opisana praca ma charakter niededukcyjny. Zbędny jest również fragment dotyczący głębokich sieci neuronowych, ponieważ nie zostały one wykorzystane w rozprawie. Moje główne zastrzeżenie do autoreferatu dotyczy tego, że doktorant nie wykorzystał w pełni możliwości zaprezentowania w nim swojej ogólnej wiedzy teoretycznej w dyscyplinie. W szczególności, pisząc o mapach samoorganizujących się, doktorant nie przedstawia zasady ich działania. Podobnie, pisząc o regresji grzbietowej, doktorant nie wyjaśnia, dlaczego potrzebna jest regularyzacja, jak wygląda model regresji z karą L2, ani do czego służy walidacja krzyżowa. Braki te mogą zostać skorygowane w prezentacji podczas obrony doktoratu.

Mam również następujące pytanie do doktoranta:

Czy w dotychczasowym zastosowaniu do zbiorów danych analizowanych w pracy nr. 3, albo w zastosowaniach do innych zbiorów danych, algorytm *xcore* zwrócił (na szczycie list rankingowych) jakiegokolwiek regulator transkrypcji, który nie jest czynnikiem transkrypcyjnym?

Powyższe uwagi nie wpływają na moją ocenę wyników merytorycznych przedstawionych w rozprawie, którą oceniam jako wybitnie dobrą.

Z przyjemnością informuję, że zgłoszona do oceny rozprawa doktorska mgr inż. Macieja Migdała spełnia warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668). W związku z powyższym zwracam się do Rady Nauk Medycznych Warszawskiego Uniwersytetu Medycznego o dopuszczenie Macieja Migdała do dalszych etapów postępowania o nadanie stopnia doktora nauk medycznych i nauk o zdrowiu w dyscyplinie nauki medyczne.

Wnoszę również o nadanie tej rozprawie doktorskiej wyróżnienia, ze względu na bardzo wysoki poziom naukowy składających się na nią publikacji oraz przydatność opracowanego przez doktoranta oprogramowania dla środowiska naukowego.

Michał Dąbrowski

Warszawa, 15.05.2023.