

Wrocław, 18.09.2020 r.

Prof. dr hab. inż. Małgorzata Kotulska  
Politechnika Wrocławska  
Wydział Podstawowych Problemów Techniki  
Katedra Inżynierii Biomedycznej  
E-mail: malgorzata.kotulska@pwr.edu.pl

**Recenzja rozprawy doktorskiej: *mgra Rafała Zaborowskiego***

zatytułowanej

***Computational methods for differentia analysis of chromatin contact matrices***

**1. Charakterystyka rozprawy**

Przedstawiona mi do oceny praca doktorska pana Rafała Zaborowskiego powstała na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego, pod kierunkiem dra hab. Bartosza Wilczyńskiego. Rozprawa została napisana w języku angielskim, liczy 94 strony i podzielona jest na 6 rozdziałów. Rozprawę otwiera wprowadzenie teoretyczne, przedstawiające problem biologiczny badań nad strukturą chromatyny, aktualnie stosowane metody eksperymentalne oraz obliczeniowe (rozdziały 1-3). W rozdziale 4 autor prezentuje zastosowanie własnej, autorskiej miary odległości BP do porównania segmentacji chromosomów. Również rozdział 5 poświęcony jest oryginalnym wynikom doktoranta i przedstawia jego własną metodę analizy różnicowej chromatynowych kontaktów dalekiego zasięgu - DiADeM. Pracę zamyka Podsumowanie i wykaz literatury. Dodatkowo, autor załączył oddzielne obszernie Autoreferaty w języku polskim i angielskim, streszczające główne wyniki jego pracy badawczej.

**2. Problem badawczy i jego znaczenie**

Bioinformatyczne i biostatystyczne metody badań genomu są obecnie bardzo intensywnie rozwijane. Ich zastosowanie znacząco skraca czas analizy problemu, a przede wszystkim koszty badań. Mimo, że skuteczność tych metod nadal nie jest idealna, to ich zastosowanie na początkowym etapie badań na ogół znacząco zawęża zakres koniecznych do przeprowadzenia eksperymentów, często wymagających dostępu do kosztownych odczynników i unikalnej aparatury pomiarowej. Z drugiej strony, metody obliczeniowe pomagają wydobyć sens z ogromu, nie zawsze czytelnych, wyników eksperymentalnych i wskazać konkretne zjawiska, które się pod tymi danymi ukrywają. Zauważalna jest rosnąca

skuteczność dostępnych algorytmów, które coraz częściej będą ograniczyć eksperyment do ostatecznej weryfikacji uzyskanego wyniku obliczeniowego. Stąd znaczenie rozwoju takich metod jest olbrzymie.

Autor przedstawionej mi do recenzji rozprawy zajął się rozwojem metod obliczeniowych do analizy kontaktów między nićmi chromatyny. Opracował dwie metody, które ułatwiają analizę krótkich odczytów uzyskanych z eksperymentów typu *Hi-C*, prowadzonych w celu rozwiązania struktury chromatyny. Wiedza na ten temat przyczynia się do lepszego zrozumienia mechanizmów odpowiedzialnych za regulację ekspresji genów, a w tym również różnicowanie komórkowe. Różnorodność funkcjonowania, a nawet budowy komórek, które mają identyczny materiał genetyczny, jest efektem oddziaływania elementów regulatorowych DNA z genami. W tych procesach, szczególnie u eukariontów, bardzo istotna jest struktura chromatyny i położenie odcinków wzmacniaczy i atenuatorów w stosunku do promotorów regulowanych genów. Rolę odgrywa tu fizyczna odległość między tymi dwoma obszarami, o której można wnioskować na podstawie macierzy kontaktów. Struktura chromatyny, a więc także występujące w niej miejsca kontaktowe, zależą od typu komórki, jej aktualnego stanu i warunków środowiskowych.

Problem, jakim zajął się doktorant w swojej rozprawie, wpisuje się w nurt ważnych i aktualnych badań naukowych prowadzonych na świecie. Zjawiska te, szczególnie regulacja genów za pomocą dystalnych (odległych) sekwencji regulatorowych jest dotychczas słabo poznana i jest przedmiotem wielu badań naukowych. Ich wyniki będą miały nie tylko wartość poznawczą, ale również praktyczną. W dobie rosnącej dostępności do badań genetycznych, medycyna spersonalizowana będzie odgrywać coraz większą rolę, korzystając z informacji zawartej w genotypach pacjentów, a z czasem też ich cech funkcjonalnych. Konieczne jest więc opracowanie metodyki wiarygodnej analizy danych takich danych i standardów do których będą odnoszone aktualne indywidualne wyniki pacjenta.

### **3. Ocena przedstawionej rozprawy, komentarze i pytania**

Twórczy wkład autora przedstawionej rozprawy można podzielić na dwie zasadnicze części: krytyczny przegląd i omówienie dostępnych metod obliczeniowych oraz wyniki własnych prac badawczych.

Przegląd metod stosowanych do analizy kontaktów chromatyny pokazuje szeroką wiedzę Autora w tematyce doktoratu, co jest wymogiem uzyskania stopnia doktora. Przegląd ten obejmuje zarówno podstawy biologiczne podjętej problematyki, jak i opis wybranych metod obliczeniowych. Autor zrobił też dość szerokie rozpoznanie literatury z zakresu swojej

tematyki, sięgające odległych lat. W tym przeglądzie zabrakło mi jednak kilku nowych prac z tej dziedziny, opublikowanych ostatnio w wiodących czasopismach. Przykładowo, należałoby wspomnieć o publikacjach prezentujących nowe metody obliczeniowe, takie jak: *FitHiChIP* (Bhattacharyya i in. Nat. Commun. 2019;10(1):4221), *MAPS* (Juric i in. PLoS Comput Biol. 2019;15(4):e1006982), czy *DeepTACT* (Li i in. Nucleic Acids Res. 2019; 47(10):e60). Warto byłoby też odnieść się do interesującej publikacji porównawczej, której autorzy testują skuteczność aż 13 różnych metod obliczeniowych stosowanych w analizie kontaktów chromatynowych (Forcato i in. Nat Methods. 2017; 14(7): 679–685) i do nowszej pracy przeglądowej z tego samego zespołu (Pal i in. Biophys Rev. 2019; 11(1): 67–78).

Część przeglądowna jest przygotowana starannie – zarówno pod względem ogólnej kompozycji, zawartości, jak i językowo oraz edycyjnie. Ujęło mnie to, że Autor zadbał o zgody wydawców, którzy są właścicielami praw autorskich do ilustracji zamieszczonych w doktoracie. Dotychczas nie spotkałam się z tym na poziomie pracy doktorskiej, która nie jest publikacją *sensu stricto*, ale uważam, że to jest bardzo eleganckie rozwiązanie. Dyskusyjny jest natomiast niestandardowy sposób budowania odnośników do pozycji literaturowych, za pomocą etykiet. Sądzę, że lepiej byłoby jednak używać nazwisk pierwszych autorów, co łatwiej pozwala zapamiętać powtarzające się w tekście odniesienia i kojarzyć je z konkretnymi zespołami badawczymi.

Wkład badawczy Autora rozprawy obejmuje dwa zagadnienia. Pierwsze z nich dotyczy miary odległości pomiędzy uzyskaną segmentacją chromosomów. Miara *BP* zaprojektowana przez Autora jest bardzo intuicyjna i łatwa do zrozumienia. Aż dziw bierze, że okazała się tak znacząco lepsza od metod konkurencji. Nie jestem matematykiem i trudno mi ocenić poprawność wszystkich wywodów Autora, jednak rozumiem, że *BP* spełnia wszystkie niezbędne własności metryk, a w konkretnym zastosowaniu daje wyniki lepsze od aktualnie stosowanych.

Drugie z tych zagadnień, to opracowanie metody wykrywania różnicowych kontaktów Hi-C. Autor skonstruował model o nazwie DiADeM. Zastosował pomysł sparowania zbiorów interakcji i uwzględnił zanik liczby kontaktów wraz ze wzrostem odległości genomowej. Zauważył i wykorzystał korelacje w profilach interakcji pomiędzy parami eksperymentów na chromatynie, przy ustalonych odległościach genomowych. Nasuwa mi się tu pewna wątpliwość, o której wyjaśnienie prosiłabym Autora. Jeżeli pod uwagę brana jest tylko odległość genomowa (5.5), to jaka jest pewność, że odstający punkt – różniący dwie macierze kontaktów – dotyczy dokładnie tego samego położenia bezwzględnego w chromosomie? Analizowane tu profile interakcji tego położenia nie uwzględniają. W końcu, Autor wskazuje,

że jego metoda wykazuje przewagę nad innymi i dokumentuje to na rys. 5.9. Porównanie to rzeczywiście wskazywałoby na przewagę metody DiADeM nad konkurencyjnymi metodami wybranymi do porównań. Na rysunku tym jednak nie do końca jest dla mnie jasne jak dokładnie rozumieć *krotność* zmiany i jak zdefiniowana jest dyspersja pomiędzy parami zestawów danych, wyrażona w procentach.

Poza uwagami zaprezentowanymi w omówieniu powyżej, chciałabym też zachęcić doktoranta do odpowiedzi na kilka ogólnych pytań, które nasunęły mi się w trakcie czytania pracy doktorskiej:

1. Czy został opracowany i udostępniony pakiet obliczeniowy, oparty na algorytmach zaproponowanych przez autora? Nie ma o tym wzmianki w pracy doktorskiej, ani w opublikowanych artykułach.
2. Czy zaproponowane metody obliczeniowe będą skuteczne tylko w odniesieniu do danych pochodzących z eksperymentów wykonanych za pomocą technologii Hi-C? Co z nowszymi metodami, które pojawiają się ostatnio – przykładowo *ChIA-Drop* (Zheng i in. Nature 2019, 566(7745):558-562, niecytowane w pracy doktorskiej)?
3. Czy autor widzi zastosowanie nowej (niewspomnianej jeszcze w pracy doktorskiej) metody eksperymentalnej służącej do badania różnicowych interakcji w chromatynie (Gryder i in. Nat Protoc 2020;15(3):1209-1236.), na przykład jako metody walidującej jego modele? Jak można byłoby zestawzić ze sobą wyniki metod obliczeniowych Autora i metody doświadczalnej?
4. Czy można byłoby porównać cele i wyniki metod zaprezentowanych w pracy doktorskiej ze wspomnianymi już wcześniej przeze mnie FitHiChIP, MAPS, DeepTACT? Jak metody zaproponowane przez Autora plasują się w rankingu metod opisanych w pracy Forcato i in. ?

#### **4. Podsumowanie**

Stwierdzam, że mgr Rafał Zaborowski zaprezentował rozprawę doktorską rozwiązującą ważny i interesujący problem naukowy. Rozprawa zawiera wyniki samodzielnych i oryginalnych badań naukowych, potwierdzonych publikacjami w bardzo dobrych czasopiśmie, jak również przyczyniła się do rozwoju reprezentowanej dyscypliny naukowej. Autor wykazał się umiejętnością prowadzenia samodzielnej pracy naukowej, jak również wystarczającą ogólną wiedzą teoretyczną, niezbędną do uzyskania stopnia doktora z w dyscyplinie informatyka.

Biorąc pod uwagę omówione powyżej elementy oceny, stwierdzam że rozprawa doktorska mgra Rafała Zaborowskiego spełnia wymagania określone w art. 13.1 Ustawy z dn. 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki. Wnioskuje o dopuszczenie autora do kolejnych etapów przewodu doktorskiego oraz do publicznej obrony przedstawionej rozprawy. Ponadto, ze względu na jej wysoki poziom *wnioskuje o wyróżnienie rozprawy.*

Małgorzata Kotulska

