

Gliwice, 04.11.2024

## Recenzja rozprawy doktorskiej mgr Zofii Tojek

*Modelowanie komputerowe struktury trójwymiarowej chromatyny  
na podstawie danych genomicznych i mikroskopii wysokorozdzielczej*

Ukończona na Wydziale Matematyki i Nauk Informacyjnych  
Politechniki Warszawskiej

Pod opieką promotora Profesora Dariusza Plewczyńskiego

### **Tematyka i cel pracy, problem badawczy i jego znaczenie**

Przedstawiona mi do recenzji rozprawa doktorska powstała na Wydziale Matematyki i Nauk Informacyjnych Politechniki Warszawskiej pod kierunkiem Profesora Dariusza Plewczyńskiego.

Tematyka rozprawy obejmuje wykorzystanie metod komputerowych do modelowania trójwymiarowej struktury chromatyny. W swoich badaniach Doktorantka koncentruje się na tworzeniu algorytmów oraz potoków analizy danych, które na podstawie danych eksperymentalnych umożliwiają modelowanie architektury chromatyny w przestrzeni 3D.

Rozwój metody analizy trójwymiarowej struktury chromatyny jest dynamicznie rozwijającą się dziedziną nauki. Dzięki zaawansowanym technikom mikroskopii i sekwencjonowania, możliwe jest badanie struktury chromatyny z niespotykaną wcześniej rozdzielczością, co otwiera nowe możliwości w zrozumieniu jej funkcji i znaczenia w organizmach. Badanie przestrzennej struktury chromatyny jest kluczowe z kilku powodów. Pętle chromatynowe odgrywają istotną rolę w przestrzennej organizacji chromatyny i umożliwiają interakcje odległych elementów genomu, takich jak *promotory* i *enhancery*, co jest istotne w procesie regulacji ekspresji. Ponadto, chromatyna odgrywa istotną rolę w procesie replikacji oraz naprawy DNA, a zmiany w jej strukturze mogą prowadzić do różnych chorób, w tym nowotworów. Badanie struktury chromatyny pomaga więc nie tylko w wyjaśnieniu procesu regulacji genów oraz funkcjonowania genomu, ale także może również prowadzić do zrozumienia mechanizmów chorobowych i powstania nowych terapii.

**Politechnika Śląska**  
Wydział Automatyki, Elektroniki i Informatyki  
Katedra Sieci i Systemów Komputerowych

ul. Akademicka 16, pok. 414, 44-100 Gliwice  
+48 32 237 21 51 / +48 237 27 33  
Aleksandra.gruca@polsl.pl

NIP 631 020 07 36  
ING Bank Śląski S.A. o/Gliwice 60 1050 1230 1000  
0002 0211 3056

W niniejszej rozprawie Doktorantka podejmuje próbę stworzenia metodyki analizy komputerowej zdjęć otrzymanych technologią iPALM w celu modelowania struktury 3D chromatyny, a także przedstawia narzędzie do automatycznej analizy danych HiChIP. Przedstawione w rozprawie technologie eksperymentalne analizy struktury przestrzennej 3D genomu, stanowiące podstawę wykonanych analiz, są relatywnie nowe. Protokoły eksperymentalne wciąż są ulepszone, a co za tym idzie, rozwijane są również metody analizy i przetwarzania danych eksperymentalnych. Analiza struktury przestrzennej chromatyny z wykorzystaniem opisanej w pracy technologii znakowania i mikroskopii wysokiej rozdzielczości iPALM to podejście nowatorskie, w związku z czym brakuje standardowych metod analizy tego typu danych eksperymentalnych.

Podsumowując tematykę badawczą, którą podjęła Doktorantka w przedstawionej pracy, należy stwierdzić, że wybrany problem badawczy jest istotny i aktualny oraz bez wątplenia wpisuje się w obecne trendy w dziedzinie bioinformatyki oraz biologii obliczeniowej.

## Charakterystyka rozprawy

Przedstawiona praca doktorska składa się z czterech rozdziałów.

Pierwszy rozdział stanowi wprowadzenie do rozprawy i koncentruje się na wprowadzeniu do tematyki pracy, wyjaśnieniu podstaw biologii komórkowej w kontekście tematu pracy. Opisano w nim techniki eksperymentalne oraz obrazowania struktury trójwymiarowej genomu, a także wybrane metody obliczeniowe, które służą za podstawę stworzonych bądź wykorzystywanych w rozprawie algorytmów.

Rozdział drugi zatytułowany jest „**Modelowanie struktury chromatyny na podstawie trójwymiarowych zdjęć mikroskopowych**”. W tej części pracy doktorskiej, autorka opisuje wyniki swoich badań dotyczących opracowania komputerowej metody rekonstrukcji struktury 3D chromatyny z danych mikroskopii super rozdzielczej. Celem opisywanych badań było zaobserwowanie pojedynczej pętli chromatynowej wraz z regionami flankującymi. Długość analizowanej pętli to 13k par zasad, region przedłużono o regiony flankujące na początku i końcu struktury co finalnie dało fragment o długości 23k par zasad. Do zobrazowania próbek wykorzystana została metoda iPALM. Przeprowadzono dziesięć przebiegów eksperymentów, a każdy przebieg obejmował obrazowanie od kilku do kilkunastu wybarwionych jąder komórkowych. Finalnie uzyskano 25 trójwymiarowych zdjęć sond świetlnych w regionie zawierającym badaną pętlę chromatynową. Zdjęcia te w dalszej części pracy przetwarzane i analizowane były przez Doktorantkę

Rolą Doktorantki w opisywanym projekcie była analiza komputerowa otrzymanych zdjęć oraz zaprojektowanie i napisanie narzędzi niezbędnych do modelowania struktury 3D pętli chromatynowej z danych mikroskopowych. Po przefiltrowaniu wyników, w celu uzyskania zdjęć regionów, sygnały z różnych klatek zostały najpierw pogrupowane za pomocą narzędzia PeakSelector i na każdym zdjęciu wyznaczono otoczkę wypukłą. W kolejnym kroku, na bazie wyników analiz symulacyjnych, wybrano próg stosunku liczby punktów wejściowych do liczby wprowadzonych sond oligonukleotydowych. Po zastosowaniu progu, pozostało 13 obrazów, które w celu rekonstrukcji pętli poddano analizie za pomocą heurystycznego algorytmu problemu komiwojażera, a następnie otrzymane struktury zostały wygładzone metodą interpolacji kubicznej. Uzyskane w wyniku modeli pętli porównano między sobą uzyskując dość duży poziom zmienności,

co potwierdziło znaną z literatury heterogeniczność pętli pomiędzy pojedynczymi komórkami. W dalszej części projektu Doktorantka porównuje rezultaty modelowania z danymi Hi-C, ChIA-PET, Micro-C oraz CHIA-Drop. Ostatnia część rozdziału to analiza możliwości zastosowania narzędzia ChromoLooping i jego modyfikacji do innych technik mikroskopowych, w tym mikroskopii elektronowej, co wymagało zmodyfikowania oryginalnej metody w sposób umożliwiający analizę pól gęstości. Wyniki prac przedstawionych w tym rozdziale zostały opublikowane pod postacią artykułu naukowego w czasopiśmie *Scientific Reports*, gdzie Doktorantka jest pierwszą autorką.

Trzeci rozdział pracy zatytułowany jest „**Analiza struktury trójwymiarowej chromatyny na podstawie danych sekwencyjnych**” i zawiera opis prac Doktorantki związanych z projektowaniem oraz implementacją narzędzia do automatycznej analizy danych HiChIP. Na kroki analizy zaimplementowane w narzędziu składają się mapowanie sekwencji do genomu referencyjnego (algorytm BWA-MEM), filtrowanie odczytów o słabej jakości, oraz usuwanie duplikatów, identyfikacja miejsc wiązania białko-DNA (oprogramowanie MACS3), identyfikacja interakcji chromatyny dalekiego zasięgu (program MAPS). W celu połączenia wszystkich komponentów w postaci serii zależnych kroków oraz umożliwienia równoległego przetwarzania danych, wykorzystano początkowo narzędzie Lugi, a w późniejszym czasie przepisano projekt to narzędzia Nextflow. W dalszej części rozdziału Autorka wykorzystuje zaprojektowane przez siebie narzędzie do analizy rzeczywistych zbiorów danych. W tej części szczególnie interesujące są analizy porównawcze wykonane dla wyników eksperymentalnych pochodzących ze zmodyfikowanego protokołu dcHiChIP. Wyniki uzyskane dla zmodyfikowanego protokołu zostały opublikowane na platformie bioRxiv oraz złożone do recenzji w czasopiśmie *Nature Communications*.

Ostatni rozdział pracy zwięźle opisuje wnioski z wykonanych analiz oraz kierunki dalszych prac.

## Opinia o rozprawie

Pierwsza część rozprawy prezentuje analizę danych pochodzących z nowatorskiego eksperymentu wizualizacji pętli chromatynowych za pomocą zdjęć mikroskopii świetlnej. Zadaniem Doktorantki było tu opracowanie całościowego podejścia metodologicznego do analizy przekazanych jej wyników eksperymentalnych. W szczególności na użytek projektu stworzone zostało narzędzie ChromoLooping umożliwiające przetwarzanie plików, filtrowanie danych, analizę jakości zdjęć oraz modelowanie pętli. Dane eksperymentalne analizowane przez Doktorantkę pochodzą z eksperymentów wykonanych najnowszymi metodami, walidacja wyników jest więc trudna z uwagi na brak danych referencyjnych, które mogłyby zostać wykorzystane do weryfikacji poprawności zastosowanych w trakcie modelowania metod. W tym celu w pracy wykonano analizy symulacyjne oraz wykonano porównanie macierzy dystansów uzyskanych z danych populacyjnych dla innych metod badania interakcji chromatyny.

W kolejnej części rozprawy, Doktorantka przedstawia stworzony przez siebie potok analizy danych nf-HiChIP. O ile doceniam wysiłek związany z zaprojektowaniem złożonego potoku danych umożliwiającego równoległe przetwarzanie wielu próbek, a także zaangażowanie w projekty, w których zostało to narzędzie wykorzystane, o tyle jednak tę część pracy oceniam nieco słabiej niż rozdział wcześniejszy, z uwagi na jej implementacyjny charakter polegający na połączeniu istniejących narzędzi.

Spis literatury składa się ze 203 pozycji i oddaje aktualny stan wiedzy w zakresie, którego dotyczy rozprawa, chociaż spośród cytowanych pozycji tylko 36 są artykułami cytującymi prace opublikowane na przestrzeni ostatnich 5-ciu lat. Pozostałe artykuły cytują prace starsze. Biorąc pod uwagę, iż tematyka, którą zajmuje się Doktorantka to dziedzina bardzo nowa, wydaje się, że przedstawiony spis literatury powinien zostać uaktualniony o najnowsze pozycje literaturowe w zakresie tematyki rozprawy.

Pozytywnie też oceniam fakt, iż kod źródłowy metody ChromoLooping, algorytmu DensityDots oraz a także metody nf-HiChIP zostały udostępnione środowisku naukowemu na platformie GitHub. W przypadku narzędzia nf-HiChIP dość zaskakującym jest fakt, że na platformie GitHub Doktorantka nie jest współtwórcą projektu (aczkolwiek zauważam, że repozytorium zawiera odniesienie do poprzedniego projektu luigi\_seq).

Podsumowując, uważam, iż w swojej pracy Doktorantka zastosowała odpowiednie metody badawcze w celu rozwiązania problemu komputerowego modelowania struktury chromatyny na bazie danych genomicznych oraz pochodzących z mikroskopii wysokiej rozdzielczości. Zastosowane w pracy podejścia do przetwarzania danych wskazują na dobrą znajomość przez Doktorantki nowoczesnych i efektywnych metod stosowanych w dziedzinie biologii obliczeniowej.

## Uwagi krytyczne i dyskusyjne

Należy podkreślić, że problem, który podjęła się rozwiązać Doktorantka jest niewątpliwie ważny i trudny i na początku tej części chciałbym podkreślić, że nie znalazłam w przedstawionych wynikach żadnych zasadniczych błędów merytorycznych. Wszystkie poniższe uwagi wynikają z chęci podjęcia dyskusji i dialogu na temat niektórych aspektów pracy. Uwagi te nie obniżają mojej pozytywnej oceny pracy.

- Typowym elementem pracy doktorskiej zwykle zawierającym we wstępie jest hipoteza badawcza lub cel badań zrealizowanych w pracy oraz motywacja przeprowadzonych badań. Hipoteza badawcza lub jasno określony cel pracy doktorskiej stanowią zasadniczy fundament przeprowadzonych badań naukowych. Hipoteza kieruje procesem badawczym, umożliwiając skoncentrowanie się na konkretnych pytaniach i problemach, które mają zostać rozwiązane. Jasno określony cel natomiast pozwala na ocenę czy badania osiągnęły zamierzone rezultaty, co jest istotne dla oceny wartości naukowej pracy. Przedstawiona mi do oceny praca nie zawiera ani hipotezy badawczej, ani jasno zdefiniowanego celu. Brak też motywacji badań, która również jest elementem istotnym, ponieważ wyjaśnia, dlaczego podjęto dane badania i jakie znaczenie mają one w szerszym kontekście naukowym. Brak tych elementów utrudnia ocenę pracy, prowadzi do trudności w interpretacji wyników i obniżeniu przejrzystości pracy badawczej, a także sprawia wrażenie, że doktorat stanowi zbiór wybranych projektów, w których Doktorantka brała udział, złożonych w jedną całość.
- Rozdziały 2 oraz 3 stanowią opisy odrębnych projektów i na końcu każdego z nich brakuje podsumowania oraz wniosków. Dla analiz opisywanych w Rozdziale 2 przydałby się schemat blokowy elementów składających się na metodę ChromoLooping ułatwiający zrozumienie poszczególnych kroków analizy.

- W sekcji 2.1.5 rozdziału drugiego Doktorantka dokonuje porównania wyników uzyskanych pętli chromatynowych w postaci mapy integracji z wynikami z danych populacyjnych uzyskanych metodami CHiA-Pet, Hi-C oraz Micro-C. Pewien niedosyt budzi w tej części brak konkluzji wynikających z porównań, gdyż przedstawione rysunki pozwalają jedynie na dokonanie porównań wizualnych. Skale przedstawionych map dystansu oraz interakcji na rysunkach 17 i 18 są różne co znacznie utrudnia stwierdzenie na ile uzyskane wyniki są podobne. W związku z tym nasuwa się pytanie czy istnieją metody normalizacji wyników pomiędzy eksperymentami? Jeśli tak, to czy możliwe byłoby ich zastosowanie w tym przypadku. Jeśli nie, to czy możliwe byłoby zaproponowanie takich metod w celu porównania wyników uzyskanych za pomocą różnych metod, w szczególności, jeśli porównujemy ze sobą mapy dystansów oraz interakcji. Dodatkowo, czy poza porównaniem wizualnym, w celu porównania i walidacji uzyskanych wyników, zasadne byłoby zastosowanie miar podobieństwa map kontaktów?
- W przypadku metody Micro-C, na rysunku 18, przedstawiono jedynie mapy interakcji Micro-C w różnych rozdzielczościach, porównując wyniki z rezultatami Hi-C. W tej części pojawił się komentarz, iż potencjalnie metoda Micro-C będzie w stanie zapewnić odpowiednią rozdzielczość umożliwiającą dokonywanie porównań danych mikroskopowych z populacyjnymi danymi genomycznymi. Jakiego warunku musiałby zostać więc spełnione, aby wyniki metody Micro-C można było wykorzystać w celu walidacji uzyskanych wyników modelowania?
- Pokazana na rysunku 16 mapa dystansów uzyskana z uśrednienia symulacji ze spaceru losowego wykazuje uśrednione wartości wzdłuż przekątnej macierzy. Czy w związku z tym jest to faktycznie najlepszy sposób przedstawienia wyników symulacji? W jaki sposób należy interpretować takie rozłożenie odległości dla uśrednionych wartości? Czy w przypadku uśrednienia danych populacyjnych wynik byłby podobny?
- W przypadku narzędzia nf-HiChIP, we wstępie Doktorantka podkreśla problem związany z przetwarzaniem wielu próbek oraz łączeniem replikatów technicznych jak i biologicznych. W jaki sposób łączone są ze sobą analizowane próbki? Czy narzędzie umożliwia weryfikację wyników odstających i, jeśli takie dane występują, jaki jest protokół postępowania zarówno w przypadku replikatów technicznych jak i biologicznych?

## Podsumowanie

Pan mgr Zofia Tojek przedstawiła rozprawę doktorską rozwiązującą aktualny problem naukowy, która przyczyni się do rozwoju reprezentowanej dyscypliny naukowej. Rozprawa zawiera oryginalne rozwiązanie problemu naukowego, a Doktorantka wykazała, że zarówno posiada ogólną wiedzę teoretyczną w dyscyplinie informatyka techniczna i telekomunikacja jak i umiejętność prowadzenia pracy naukowej.

Doktorantka jest pierwszą autorką publikacji w czasopiśmie Scientific Reports, współautorką publikacji w Nature Communications, a także współautorką publikacji, która została złożona do recenzji. W trakcie

swojego doktoratu zatrudniona była jako stypendystka w projektach naukowych, co świadczy o jej zaangażowaniu w prace naukowe.

Biorąc pod uwagę ocenę przedstawioną w niniejszej recenzji, stwierdzam, że przedstawiona do oceny praca doktorska w pełni odpowiada warunkom określonym w Art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (tekst jednolity Dz. U. z 2023 r. poz. 742 z późn. zm.) i na tej podstawie wnoszę do Wysockiej Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o dopuszczenie mgr Zofii Tojek do dalszych etapów przewodu doktorskiego.

dr inż. hab. Aleksandra Gruca, prof. PŚ.