



Wrocław, 8.08.2023

Recenzja rozprawy doktorskiej Pana mgra Michała Własnowolskiego pt.
"Computational Modelling and Analysis of the Three-Dimensional Structure of Human
Genome at the Population Scale"

Odczyt informacji genetycznej zwłaszcza u organizmów eukariotycznych, a szczególności u zwierząt, jest procesem bardzo skomplikowanym i wciąż słabo poznanym. Wymaga on nie tylko samej sekwencji genomowego DNA oraz oddziaływania czynników regulacyjnych i epigenetycznych, ale również odpowiedniej przestrzennej organizacji chromatyny. Aby zrozumieć te zjawiska stosowane są techniki eksperymentalne, jak ChIA-PET i Hi-C, dostarczające danych o przestrzennych kontaktach w chromatynie. Jednakże są one czasochłonne i kosztowne, dlatego nie można ich stosować na szeroką skalę, np. w badaniach populacyjnych, które są niezbędne do poznania różnic i mechanizmów molekularnych wielu chorób. W związku z tym istnieje potrzeba rozwijania metod komputerowych do badania i interpretowania struktury i oddziaływań chromatyny.

Dlatego bardzo słusznie ambitnym przedmiotem pracy doktorskiej Pana mgra Michała Własnowolskiego stało się opracowanie zaawansowanych narzędzi informatycznych służących do modyfikacji wzorów kontaktów chromatynowych wynikających ze zmian sekwencji DNA. Dzięki temu możliwe jest generowanie i porównywanie różnych modeli przestrzennych chromatyny na poziomie populacyjnym.

Rozprawa doktorska została napisana w języku angielskim ze streszczeniem w języku polskim. Zawiera ona: spis treści, wykaz rycin, wstęp, główny rozdział przedstawiający najważniejsze osiągnięcia naukowe związane z rozprawą doktorską i przedstawione w załączonych czterech pracach, rozdział zawierający pozostałe osiągnięcia doktoranta, wnioski i przyszłe badania, bibliografię, kopie publikacji będące przedmiotem rozprawy, oświadczenia współautorów i kopie dodatkowych publikacji.

Wstęp stanowi zwięzłe, ale dobre wprowadzenie do idei określania struktury przestrzennej chromatyny. Przedstawiono w nim metody eksperymentalne (Hi-C i ChIA-PET) i sposób prezentowania oddziaływań w postaci diagramów łukowych. Najwięcej miejsca poświęcono podejściu obliczeniowemu 3D-GENOME, które w oparciu o eksperymenty ChIA-PET przedstawia trójwymiarowy model chromatyny. Zakłada ono interakcje między miejscami chromatyny, w których pośredniczą białka RNAPII, CTCF i kohezyny. Podejście to oparte o symulacje metodą wyżarzania metodą Monte Carlo i zakładające hierarchiczną organizację chromatyny zostało wykorzystane w doktoracie. W wystarczający sposób scharakteryzowano także regulację ekspresji genów na poziomie epigenetycznym poprzez metylację cytozyny i modyfikacje histonów oraz wiązanie się białek do chromatyny i elementów regulatorowych w DNA, tj. promotorów, wyciszaczy i wzmacniaczy. Wyjaśniono, że ich interakcje w przestrzeni trójwymiarowej wpływają na odpowiednią ekspresję informacji genetycznej, a zaburzenia w ich oddziaływaniu mogą być związane z chorobami.

W osobnym rozdziale jasno sformułowano cele pracy doktorskiej, którymi było opracowanie i zastosowanie narzędzi obliczeniowych służących do generowania i analizy trójwymiarowego modelu chromatyny oraz zbadania wpływu przestrzennej struktury chromatyny na aktywność genetyczną. Cele pracy są poprawnie sformułowane i wszystkie zostały właściwie zrealizowane.

Osiągnięcia związane z rozprawą doktorską zostały przedstawione w czterech pracach. W trzech z nich doktorant jest pierwszym autorem, a w jednej czwartym. Wyniki trzech prac zostały opublikowane w prestiżowych czasopismach (*Genome biology* i *Nucleic Acids Research*), a czwarta jest w trakcie recenzji w również prestiżowym czasopiśmie *Bioinformatics*. Wyniki zaprezentowane w tych pracach są spójne i układają się w jedną całość.

Załączone oświadczenia autorów nie pozostawiają wątpliwości, że Pan Michał Własnowolski miał wiodący i istotny wpływ na koncepcję prac i uzyskanie oraz opisanie wyników w publikacjach będących przedmiotem pracy doktorskiej. Przedstawiona rozprawa opisująca wyniki prac jest oryginalnym wkładem doktoranta.

W pierwszej publikacji doktorant przedstawił wpływ wariantów genetycznych na strukturę trójwymiarową chromatyny i opisał narzędzie do przewidywania zmiany tej struktury w oparciu o warianty strukturalne (SV, structural variants) w genomie człowieka, takie jak delecje, duplikacje, insercje i inwersje. Są one związane z wieloma chorobami i wadami rozwojowymi. Opracowane narzędzie opiera się o dane ChIA-PET w połączeniu z wariantami

strukturalnymi genomów uzyskanymi w ramach 1000 Genomes Consortium. Opracowane narzędzie umożliwia badania w skali populacyjnej, co jest istotne, aby powiązać zmiany genomowe z fenotypem w skali globalnej. Warto podkreślić, że opracowane narzędzie i badania są pierwszymi, które wiążą zmienność struktury przestrzennej genomu człowieka w skali całej populacji. W pracy tej określono tendencję wariantów strukturalnych do gromadzenia się w przestrzennie oddziałujących segmentach genomowych. Pokazano, że zróżnicowana transkrypcja genów jest ściśle związana ze zmiennością sieci interakcji chromatyny, w których pośredniczy polimeraza RNA II. Wykazano również, że interakcje, w których bierze udział białko CTCF, są konserwatywne w genomach człowieka, ale zawierają dużo polimorfizmów pojedynczych nukleotydów (SNP) związanych z chorobami. Ponadto stwierdzono, że granice domen topologicznych w chromatynie są stosunkowo częstymi miejscami duplikacji, co sugeruje, że tego typu zmiany mogą być ważnym mechanizmem wpływającym na przestrzenną organizację genomu.

W drugiej publikacji opisano integrację wcześniej utworzonego narzędzia z serwisem internetowym 3D-GNOME, który posiada przyjazny interfejs dla użytkownika oraz użyteczne narzędzia umożliwiające analizę różnic w trójwymiarowej strukturze chromatyny dla różnych wariantów strukturalnych. W początkowej wersji, narzędzie to generowało trójwymiarowe modele chromatyny oparte na danych z metody Chromatin Conformation Capture (3C). Serwis dostarcza również narzędzia do wizualizacji i analizy struktury przestrzennej. Użytkownicy mogą przewidzieć strukturę chromatyny dla wybranego regionu genomu człowieka określając jego współrzędne. Poza wariantami strukturalnymi uzyskanymi z 1000 Genomes Project można również podawać własne warianty. Podobnie, poza danymi dotyczącymi interakcji białek CTCF i RNAPII z chromatyną w oparciu o eksperymenty ChIA-PET dla linii limfoblastycznej komórek GM12878 można podawać własne dane tego typu. Usługa internetowa wykorzystuje framework webowy Flask z zapytaniami do bazy danych MySQL. Dane są przetwarzane w oparciu o skrypty Pythonie, natomiast główne oprogramowanie jest napisane w C++, PHP i R. Serwis 3D-GNOME generuje diagramy kontaktowe chromatyny, wykresy rozkładu długości miejsc kontaktowych, podsumowujące statystyki i modele przestrzenne chromatyny. Modele można oglądać poprzez interaktywną przeglądarkę zaimplementowaną w WebGL. Wygenerowane modele można pobrać w standardowych formatach i oglądać w innych programach.

W trzeciej publikacji Pan Michał Własnowolski zaprezentował narzędzie cudaMMC, czyli nową wersję 3D-GNOME, które wykorzystując kart graficzne (GPU) do akceleracji

obliczeniowej, potrafi 25-krotnie przyspieszyć obliczenia przewidywania struktury przestrzennej w porównaniu z wersją oryginalną. Jest to istotne podejście ze względu na intensywny przyrost liczby danych z eksperymentów ChiA-PET i Hi-C. Dzięki temu można przeprowadzać wiele przewidywań struktury chromatyny w krótkim czasie, co jest istotne dla analiz statystycznych w oparciu o dane populacyjne. Takie analizy są kluczowe dla zrozumienia wpływ trójwymiarowej struktury chromatyny na regulację transkrypcji i badanie zaburzeń w tej strukturze z powodu zmiany odległości między elementami regulatorowymi, takimi jak wzmacniacze i promotory. Zadanie to było dużym wyzwaniem, ponieważ w przypadku symulowanego wyżarzania metodą Monte Carlo trudno jest zoptymalizować obliczenia równoległe i sekwencyjne na GPU i CPU, gdyż energia globalna dla całej struktury chromatyny jest obliczana po poprawieniu lokalnych oddziaływań. Wymagało to napisania nowego oprogramowania. Obliczenia oparte o GPU charakteryzują się także większą stabilnością.

W czwartej publikacji opisano aktualizację serwisu internetowego 3D-GNOME. Udoskonalenie polegało na uwzględnieniu w modelach trójwymiarowych chromatyny narzędzia do analizy zmian odległości między wzmacniaczami i promotorami biorącymi udział w regulacji transkrypcji genów. W badaniach tych zastosowano większy zbiór danych interakcji chromatyny ChIA-PET o lepszej rozdzielczości oraz uaktualnione informacje o wariantach strukturalnych z danych genomowych o pokryciu 30-krotnym. Dodatkowo w tej aktualizacji narzędzie cudaMMC zintegrowano z serwerem WWW, a obliczenia optymalizujące proces generowania modeli 3D przerzucono na klaster obliczeniowy EdenN, który wykorzystuje karty graficzne Nvidia DGX A100. To znacznie przyspiesza obliczenia. W nowym narzędziu, po podaniu danych dotyczących wzorów interakcji chromatyny oraz wymodelowaniu struktury przestrzennej chromatyny, a także zmapowaniu promotorów i wzmacniaczy genów porównywane są rozkłady odległości między nimi dla sekwencji referencyjnej i zmienionego wariantu strukturalnego. Wyniki podawane są w formie tabelarycznej, wykresów łukowych, modeli przestrzennych i wykresów pudełkowych. Zmiany odległości między promotorami i wzmacniaczami mogą mieć istotne znaczenie w zmianie ekspresji genów. Dlatego to narzędzie jest bardzo pomocne i ważne, aby przewidywać takie zmiany.

Część odpowiadająca wnioskowi jest napisana przejrzysto i zawierają najważniejsze informacje uzyskane z przeprowadzonych badań i opracowanego narzędzia. Interesujące jest nawiązanie do przyszłych analiz dotyczących paleogenomów człowieka, które mogą wnieść wiele ciekawych informacji na temat zmiany ekspresji informacji genetycznej w ewolucji.

Doktorant włożył dużo trudu w opracowanie nowatorskiego narzędzia i przeprowadzone analizy, a przedstawione opisy wyników świadczą o dużej dojrzałości naukowej doktoranta i umiejętności wydobywania najważniejszych informacji z uzyskanych rezultatów. Nie mam zastrzeżeń do metodyki przeprowadzonych analiz. Opisy są dobrze przedstawione pod względem formalnym. Praca i artykuły są napisane poprawnym językiem i stylem.

Mam tylko drobne uwagi. W Streszczeniu polskim zamiast meatzoa powinno się użyć terminu zwierząt albo po prostu Metazoa z dużej litery lub Animalia. Abstract angielski nie odpowiada dokładnie Streszczeniu w języku polskim. Na przykład nie ma w nim informacji, że zastosowanie kart graficznych zwiększyło obliczenia do 25 razy. W opisie polskim zamiast "enhancerów" dałbym "wzmacniaczy". Zdanie: "In the case of metazoans, such as humans, the complete genetic information in each cell is identical (apart from mutations accumulated during development)." można zastosować dla wszelkich organizmów wielokomórkowych, nie tylko zwierząt. Na Fig. 2.4 w części Output jest napisane "gene – enhancer distances". Czy nie powinno być "promotor-enhancer distances"? W nowej wersji programu przy określaniu istotności różnic między rozkładami odległości między promotorami i wzmacniaczami zastosowano test Mann–Whitneya U. Test ten adekwatny, jeśli dane nie mają rozkładu normalnego. Nie zawsze musi tak być i w tym przypadku, gdyby dane miały taki rozkład, można by stosować test t-Studenta. Można to rozważyć w uaktualnionej wersji. Mam jeszcze pytanie: dlaczego narzędzie nazwano 3D-GNOME, a nie 3D-GENOME?

Stwierdzone przeze mnie zastrzeżenia nie rzutują jednak na bardzo pozytywną ocenę pracy, a przedstawione powyżej uwagi nie zmniejszają wartości ocenianej rozprawy. Tematyka pracy doktorskiej jest bardzo zasadna, ponieważ istnieje potrzeba tworzenia narzędzi i analizy struktury przestrzennej chromatyny w aspekcie całej populacji człowieka z uwzględnieniem zmian strukturalnych i elementów regulujących transkrypcję, co stało się przedmiotem rozprawy. Należy podkreślić, że opracowany program jest zaawansowany, została polepszona szybkość i skuteczność jego działania, a analizy zostały przeprowadzone skrupulatnie na dużym zbiorze danych dostarczając nowych informacji na temat ekspresji informacji genetycznej. Reasumując chciałbym stwierdzić, że recenzowana rozprawa z artykułami stanowi istotny wkład w przewidywanie i analizę struktury przestrzennej chromatyny.

Na uwagę zasługuje dodatkowy dorobek publikacyjny doktoranta obejmujący cztery prace, uczestnictwo w czterech projektach badawczych oraz trzy wizyty naukowe. Świadczą one o wszechstronności i dojrzałości doktoranta.

Uważam, więc, że przedstawiona do recenzji rozprawa doktorska spełnia wszystkie wymogi Ustawy o Stopniach Naukowych. Zgłaszam, zatem wniosek do Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o uznanie rozprawy Pana mgra Michała Własnowolskiego za odpowiadającą wymogom stawianym rozprawom doktorskim i o dopuszczenie doktoranta do dalszych etapów przewodu doktorskiego. W związku z tym, że doktorant miał postawiony trudny cel badawczy i go efektywnie rozwiązał, a wyniki zostały przedstawione w czterech bardzo dobrych pracach proponuję wyróżnić rozprawę.



Prof. dr hab. Paweł Mackiewicz