Instytut Farmakologii
Polskiej Akademii Nauk
Zakład Chemii Leków

ul. Smętna 12
31- 343 Kraków
tel.: 12 662 32 20
fax: 12 637 45 00
www.if-pan.krakow.pl

Sekretariat Dyrekcji
tel.: 12 662 32 05
tel.: 12 662 32 49
e-mail: ifpan@if-pan.krakow.pl

Kraków, 28 January 2019

Recenzja

pracy doktorskiej mgr Przemysława Szałaja pod tytułem:

Modeling and analysis of spatial genome organization

wykonanej pod kierunkiem prof. dr hab. Dariusza Plewczyńskiego
na Wydziale Lekarskim Uniwersytetu Medycznego w Białymstoku.
Część badań została przeprowadzona na belgijskim uniwersytecie w Hasselt,
w ramach stypendium Etiuda przyznanego doktorantowi przez NCN w 2016 roku.

Już prawie 16 lat minęło odkąd międzynarodowe konsorcjum ogłosiło wyniki projektu poznania ludzkiego genomu udostępniając publicznie dane 3 miliardów par nukleotydów budujących DNA. Równolegle dokonał się ogromny postęp technologiczny w zakresie badań genetycznych i sekwencjonowanie DNA stało się rutynową częścią badań naukowych. Jednak sama sekwencja nukleotydów to dopiero początek. Biorąc pod uwagę, że ludzki materiał genetyczny zakodowany w nici DNA o długości 2 metrów, mieści się w jądrze komórkowym o średnicy ca. 10 mikrometrów, informacja dotycząca sposobu zwijania nici DNA wewnątrz jądra jest nie mniej istotna dla regulacji jego funkcji biologicznej.

Źródłami danych dla ustalenia architektury genomu są wysoko przepustowe metody eksperymentalne, takie jak ChIA-PET i Hi-C. Uzyskane dane pozwalają na ustalenie map interakcji chromatyny w jądrze, z których za pomocą odpowiednich bioinformatycznych narzędzi możliwa jest rekonstrukcja modelu 3-D genomu użytecznego w analizie jego właściwości.

Zagadnieniu modelowania przestrzennej struktury genomu jest poświęcona rozprawa doktorska mgr Przemysława Szałaja, obejmująca cykl trzech opublikowanych prac, z których dwie pierwsze to publikacje oryginalne, a trzecia ma charakter poglądowy. Praca doktorska została napisana w języku angielskim, zawiera także streszczenie w języku polskim.

Krótkie wprowadzenie przedstawia cel badań, tj. opracowanie oryginalnego algorytmu 3D-GNOME (akronim: 3-Dimensional Genome Modeling Engine) dla rekonstrukcji trójwymiarowej organizacji chromatyny. Wieloskalowość procesu modelowania odzwierciedla hierarchiczną strukturę biologiczną genomu. Algorytm wykorzystuje dane pochodzące z eksperymentów ChIA-PET i może być użyty do modelowania wybranego regionu genomu, pojedynczych chromosomów lub całego genomu w wysokiej rozdzielczości.

Kolejny rozdział to zarys metod eksperymentalnych powiązanych z pracą doktoranta, wraz z omówieniem poziomów organizacji genomu, głównie pętli chromatynowych domen i wpływających na architekturę specyficznych czynników białkowych, m.in. CTCF oraz krótka dyskusja korelacji pomiędzy strukturą genomu a występującymi chorobami.

W rozdziale 3. Autor przedstawia różne podejścia stosowane do analizy danych eksperymentalnych wykorzystywanych do ustalenia konfiguracji 3-D łańcucha chromatyny i identyfikacji domen topologicznych, wraz z przeglądem istniejących metod modelowania genomu.

Rozdział 4 zawiera ogólne omówienie algorytmu 3D-GNOME, jego zalety i ograniczenia. Autor nadmienia w tym miejscu, że algorytm został po raz pierwszy przedstawiony w pracy Tang i wsp., opublikowanej w 2015 roku w czasopiśmie Cell. Wskazano w niej na znaczenie interakcji specyficznych czynników (głównie białka wiążącego CTCF i polimerazy RNAPII) dla przestrzennej organizacji chromatyny i regulacji procesu transkrypcji. Pan Przemysław Szałaj, jako współautor tej publikacji (niewchodzącej w skład doktoratu), był członkiem czteroosobowego zespołu, który opracował program pod nazwą 3D-NOME (3D NucleoOme Modeling Engine) do wizualizacji danych uzyskanych z eksperymentów ChIA-PET i modelowania hierarchicznej struktury 3D genomu. W tym miejscu nasuwa się pytanie o powód zmiany nazwy algorytmu 3D-NOME na 3D-GNOME w kolejnej publikacji, będącej już zasadniczą częścią pracy doktorskiej. Czy był to zabieg tylko stylistyczny, czy też wiązał się z modyfikacjami opracowanego wcześniej podejścia?

Wśród najważniejszych cech, które wyróżniają 3D-GNOME na tle innych dostępnych metod, Autor wymienia mi.in.: ścisłe powiązanie z jednostkami strukturalnymi organizacji genomu (tj. z określonymi obszarami chromosomu, domenami topologicznymi, pętlami chromatyny) co ułatwia interpretację uzyskanych modeli pod względem ich znaczenia biologicznego; oparcie na danych ChIA-PET (w momencie opublikowania pracy stanowiło to o jego unikalności); ukierunkowanie na motywy pętli CFCT oraz wieloskalowość podejścia, które umożliwia symulacje trójwymiarowych modeli genomu odtwarzających zarówno struktury całego jądra komórkowego, pojedynczych chromosomów czy wybranych krótszych regionów, np. zawierających wybrany gen.

Na uwagę zasługuje fakt, że obok mocnych stron opracowanego algorytmu stanowiących o jego niezaprzeczalnej wartości, Autor analizuje także ograniczenia modelu. Są to głównie: trudności w modelowaniu gęsto upakowanych w pętle regionów chromatyny oraz całkowity brak pętli, który nawet uniemożliwia zastosowanie metody. Takie krytyczne podejście do własnych badań jest nieczęsto spotykane w pracach doktorskich i dowodzi dojrzałości naukowej, czego wyrazem jest rzetelna ocena realnych możliwości opracowanej metody.

Kopie trzech publikacji składających się na pracę doktorską zamieszczone są w rozdziale 5. Doktorant jest pierwszym autorem we wszystkich pracach, a każda z nich jest poprzedzona krótkim wstępem określającym wkład innych współautorów.

Pierwsza praca zawiera szczegółowy opis algorytmu 3D-GNOME, składającego się z trzech zasadniczych modułów, które w oparciu o dane eksperymentalne umożliwiają: 1) tworzenie tzw. map cieplnych (heat maps) kontaktów; 2) otrzymanie modeli 3D zarówno pojedynczych chromosomów, jak i całego genomu w niskiej rozdzielczości przy zastosowaniu skalowania wielowymiarowego, lub w różnych skalach za pomocą symulowanego wyżarzania; 3) wizualizację modeli 3D do analizy strukturalnej. W skład narzędzi 3D GNOME oprócz domyślnych funkcji oryginalnie opracowanych przez Autora, wchodzą także dodatkowe opcje, które można zastosować niezależnie. Obejmują one: normalizację macierzy częstotliwości interakcji pozwalającą na usunięcie niepożądanych efektów (szumów) z map kontaktów, metodę modelowania struktur na poziomie segmentów i interaktywną przeglądarkę modeli 3D.

Druga praca przedstawia organizację serwera sieciowego umożliwiającego korzystanie z platformy 3D-GNOME. Udostępnione narzędzia pozwalają użytkownikowi (nawet nieposiadjącemu zaawansowanej wiedzy komputerowej) na samodzielne przeprowadzenie symulacji z użyciem własnych albo dostępnych w dedykowanych bazach, danych eksperymentalnych.

Praca przeglądowa zatytułowana *'Three-dimensional organization and dynamics of the genome,'* opublikowana w zeszłym roku w Cell Biology and Toxicology podsumowuje dostępną wiedzę dotyczącą złożoności struktury genomu i badań eksperymentalnych służących poznaniu jego struktury.

Rozdział 6. Poświęcony jest dyskusji wyników i przedstawieniu dalszych planów związanych z rozwojem 3D-GNOME. Doktorant jeszcze raz podkreśla tutaj wyróżniające cechy stworzonego oprogramowania, które zostało z sukcesem zastosowane w wielu analizach komputerowych. Szkoda, że nie zostały wymienione i skrótowo scharakteryzowane przykłady jego wykorzystania choćby przez badaczy cytujących opublikowane w 2016 roku i wchodzące w skład dysertacji, prace Pana Przemysława Szałaja.

W celu zwiększenia uniwersalności opracowanego podejścia, planuje się wprowadzenie możliwości wykorzystania wyników badań prowadzonych z użyciem innych technik eksperymentalnych, np. danych pochodzących z badań mikroskopowych lub wyników eksperymentalnych uzyskanych z pojedynczych komórek. Ponadto, istotnym kierunkiem byłoby wprowadzenie możliwości symulacji dynamiki, jako że 3D-GNOME produkuje statyczne struktury reprezentujące pojedyncze konformacje organizacji chromatyny. Autor nadmienia ponadto, że standardową techniką eksperymentalną w badaniach organizacji genomu jest Hi-C i specyficzne narzędzia bioinformatyczne do modelowania 3D w oparciu o dane Hi-C są dostępne, tak więc metoda 3D-GNOME może być mniej popularna. W związku z tym, korzystne byłoby przeprowadzenie dodatkowych analiz i porównań z zastosowaniem danych Hi-C i ChIA-PET, w celu zachęcenia nowych użytkowników do wypróbowania algorytmu 3D-GNOME badaniach. Czy zatem, biorąc pod uwagę fakt, że dostęp do oprogramowania, jak należy sądzić po datach publikacji, został zapewniony w 2016 roku, wskazane modyfikacje i analizy porównawcze zostały przeprowadzone oraz czy opinie zwrotne użytkowników, (jeśli ewentualnie się pojawiły), przyczyniły się do wprowadzenia nowych funkcji w 3D-GNOME?

Końcowe części dysertacji obejmują: listę publikacji i doniesień konferencyjnych Doktoranta, deklaracje współautorów oraz wykaz literatury cytowanej w pracy doktorskiej. Lista publikacji Doktoranta zawiera także pozycje, które nie weszły w skład rozprawy doktorskiej.

Ze względu na charakter, przedstawiony przez Pana mgr Szałaja materiał nie zawiera jasno sformułowanej tezy/hipotezy naukowej, natomiast stanowi przykład rozwiązania technologicznego, którego zastosowanie powinno przyczynić się do lepszego zrozumienia czynników wpływających na organizację genomu oraz mechanizmów funkcjonowania genomu jako złożonej całości. Praca jest logicznie skonstruowana, bardzo starannie opracowana pod względem edycji i trudno w niej doszukać się większych usterek. Jedynymi mankamentami jakie dostrzegłem jest brak numeracji rozdziałów w rozprawie (choć w spisie treści są numerowane) oraz brak spisu używanych skrótów. Pojedynczych omyłek literowych, które zdarzają się nawet największym perfekcjonistom, nawet nie warto wymieniać. Nie mam też uwag natury merytorycznej. Badania przedstawione w pracach wchodzących w skład dysertacji zostały niezależnie poddane szczegółowej ocenie przez międzynarodowych recenzentów, tym samym uzyskując pozytywną opinię. Należy zwrócić uwagę na bardzo wysoki Impact Factor czasopism, w których zostały opublikowane wyniki pracy Doktoranta, który wynosi 24,417. Ponadto, z zamieszczonego podsumowania bibliometrycznego wynika, że łączny Impact Factor 6 prac opublikowanych z Jego udziałem wynosi 91,265 punktów, co świadczy o znaczeniu badań w jakich bierze udział.

Podsumowując, tematyka podjęta przez mgr Przemysława Szałaja w pracy doktorskiej wpisuje się w jedną z najszybciej rozwijających się obecnie dziedzin nauki jaką jest genomika. Opracowane oprogramowanie pozwala na przybliżenie trójwymiarowej struktury genomu i znalezienie powiązań pomiędzy wybranymi regionami a funkcją biologiczną, a w dalszej perspektywie może pomóc w ustalaniu prawdopodobieństwa wystąpienia określonych schorzeń.

Stwierdzam, że przedstawiona do recenzji praca doktorska spełnia ustawowe wymogi, w związku z czym wnioskuję do Wysokiej Rady Wydziału Lekarskiego Uniwersytetu Medycznego w Białymstoku o przyjęcie pracy i dopuszczenie mgr Przemysława Szałaja do dalszych etapów przewodu doktorskiego.