

# Streszczenie

Opracowanie techniki *chromosome conformation capture* (3C) oraz jej zaawansowanych odmian takich jak Hi-C czy ChIA-PET pozwoliło na wieloskalowe badanie struktury genomu oraz doprowadziło do wielu odkryć dotyczących mechanizmów funkcjonowania genomu. Jest już dobrze wiadomo, że organizacja genomu jest silnie związana z procesami komórkowymi takimi jak regulacja ekspresji genów, rozwój komórkowy czy różnicowanie komórkowe.

Co więcej, zmiany tej organizacji spowodowane przez mutacje genomiczne wpływające na strukturalne elementy genomu są bezpośrednią przyczyną powstawania wielu chorób, w tym także różnych typów nowotworów. Dodaje to dodatkowy poziom złożoności do analiz genomicznych, ponieważ nawet niewielka mutacja, np. polimorfizm pojedynczego nukleotydu może znacząco wpływać na ekspresję odległych genów. Aby w pełni zrozumieć funkcjonowanie genomu, zarówno w warunkach normalnych jak i w chorobie, konieczna jest więc wiedza o strukturze genomu i jej dynamice.

Dane z eksperymentów 3C są zazwyczaj przedstawiane za pomocą prostych reprezentacji jedno- i dwuwymiarowych, takich jak sieci interakcji czy mapy kontaktów. Pomimo użyteczności tych reprezentacji, modele trójwymiarowe zapewniają pełniejszy obraz struktury genomu i pozwalają na badanie cech takich jak dostępność czy stopień zagęszczenia chromatyny. Ograniczenia natury biologicznej, eksperymentalnej oraz obliczeniowej sprawiają jednak że tworzenie wiarygodnych trójwymiarowych modeli genomu nie jest proste.

W ciągu kilku minionych lat zaproponowano kilka podejść do modelowania przestrzennej organizacji genomu. Podejścia te są jednak zazwyczaj oparte na danych pochodzących z eksperymentów wykrywających niespecyficzne interakcje (np. Hi-C) oraz nie uwzględniają w modelowaniu naszej obecnej wiedzy o organizacji genomu, co sprawia że mogą one nie być w stanie odzwierciedlić istniejących biologicznych struktur. Ponadto, z powodów obliczeniowych rozmiary symulacji są często ograniczone, co wymusza wykorzystywanie danych o niskiej rozdzielczości lub modelowanie krótkich fragmentów genomu.

Niniejsza praca przedstawia 3D-GNOME (3-Dimensional Genome Modeling Engine), nowatorski wieloskalowy algorytm modelowania przestrzennej organizacji genomu. Algorytm jest w stanie wygenerować wysokorozdzielcze struktury całego genomu w rozsądnym czasie. Wyróżniającą cechą algorytmu jest jego bliskie powiązanie z biologicznym modelem topologii genomu. 3D-GNOME wykorzystuje podejście hierarchiczne, w którym każdy poziom odpowiada wybranej skali organizacji genomu (terytoriom chromosomowym, domenom topologicznym oraz pętlom chromatyny), a modelowanie jest oparte na punktach które ściśle odpowiadają strukturom biologicznym - w odróżnieniu od istniejących metod, w których modelowanie opiera się na arbitralnie wybranych fragmentach chromatyny nieposiadających specjalnego biologicznego znaczenia. Pozwala to na tworzenie modeli, które mogą być wizualizowane oraz analizowane na różnych biologicznych skalach, co zapewnia pełniejszy obraz topologii genomu.

3D-GNOME został zaprojektowany z myślą o korzystaniu z danych ChIA-PET, zaawansowanej techniki 3C która pozwala na wykrywanie kontaktów genomicznych

powiązanych z wybranym białkiem, ale, co do zasady, może być także zastosowany do innych typów danych, takich jak 5C czy Hi-C.

Algorytm został zaimplementowany jako niezależny program, co pozwala na proste wykorzystanie skryptów do automatycznego uruchamiania dużej liczby symulacji, oraz jako aplikacja webowa, co ułatwia korzystanie z algorytmu użytkownikom nieposiadającym wiedzy technicznej.