

Dynamiczny rozwój współczesnej medycyny i biologii molekularnej opiera się na skomplikowanych technikach pomiarowych takich jak sekwencjonowanie, mikromacierze, cytometria przepływowa. Umożliwiają one poznanie mechanizmów molekularnych zachodzących w komórce, zaangażowanych w nie cząsteczek oraz scharakteryzować sekwencje genetyczne kodujące te cząsteczki. Genom człowieka oprócz sekwencji kodujących zawiera w sobie informację o wzajemnych powiązaniach, czyli regulacji poszczególnych genów, która zależy bezpośrednio od organizacji chromatyny. Rodziny czynników transkrypcyjnych odpowiadają za tkankowo specyficzną regulację zarówno podczas rozwoju organizmu oraz w procesach chorobowych.

Dostępność zsekwencjonowanych genomów wielu organizmów pozwala na zrozumienie ewolucji molekularnej samego procesu regulacji. Okazuje się, że niektóre cechy powstają w wyniku tak zwanej *filogenetycznej egzaptacji* gdzie pewna wcześniejsza ewolucyjna adaptacja zaczyna pełnić nową funkcję. Jako przykład rozważmy region regulatorowy dla genów FOXF1 oraz TBX4, które u wszystkich gatunków związane są z rozwojem płuc, a zaburzenia ich działania u człowieka prowadzą do śmiertelnych zaburzeń rozwojowych u noworodków. Zawiera on bardzo konserwowany *region niekodujący*, który ewolucyjnie po raz pierwszy pojawił się w genomach ryb *celakantokształtnych* oraz *dwudysznych*. Bardzo ciekawa jest rola tej regulacji w adaptacji ewolucyjnej przy wyjściu zwierząt wodnych na ląd. Postulujemy, że właśnie poprzez egzaptację te sekwencje mogły zostać dokooptowane w genach ryb, aby umożliwić krytyczne etapy formowania się płuc, które były potrzebne do wielkiego ewolucyjnego skoku 390-360 milionów lat temu.

Ogromna presja selekcyjna związana w takim skoku powoduje, że korzyści ewolucyjne mogą być związane z niestabilnością genomu. Zdarzeniami drastycznie przyspieszającymi ewolucję są np. wybuchy aktywności transpozonów w genomach ryb *dwudysznych*. Wyzwaniem jakie podejmujemy w projekcie jest zbudowanie modelu ewolucji regulacji genetycznej, badanej za pomocą interakcji chromatyny technologią HiC, w odpowiedzi na zmiany strukturalne genomu. Chcemy wykorzystać opracowane wcześniej narzędzie do analizy zaburzeń regulacji (<https://tadeus2.mimuw.edu.pl/>) i stworzyć bazę danych HiC dla genomów zmienionych przez warianty strukturalne, która zostanie wykorzystana jako zbiór uczący dla modelu. Oparty o głębokie uczenie model pozwoli na predykcję zmian w organizacji chromatyny indukowanych przez warianty strukturalne genomu oraz na zrozumienie procesu filogenetycznej egzaptacji tkankowo-swoistych niekodujących regionów regulatorowych.

Wnioskowanie z gigabajtów informacji opisujących architekturę genomów, dynamikę i ewolucję regulacji wymaga zastosowania modeli matematycznych i metod obliczeniowych. Wierzymy, że nasze wcześniejsze doświadczenie w analizie proliferacji sekwencji transpozonowych, stabilności ludzkiego genomu oraz dynamiki regulacji w procesach chorobowych, pozwoli na zaproponowanie algorytmów, które będą pomocne w analizach biomedycznych stanowiąc jednocześnie nietrywialny wkład w rozwój metod uczenia maszynowego.