

Streszczenie

W dobie ogromnej ilości danych pochodzących z wysoko przepustowego sekwencjonowania istnieje konieczność poszukiwania coraz to doskonalszych metod analizy regionów funkcjonalnych genomu. Szczególną uwagę zwracają niekodujące regiony regulatorowe, które mogą mieć kluczowe znaczenie w poszukiwaniu polimorfizmów powiązanych z zachorowalnością na wiele chorób.

Celem przedstawionej rozprawy doktorskiej było opracowanie metody oceny funkcjonalności sekwencji polimorficznych w regionach promotorowych w oparciu o oszacowanie lokalnego kształtu DNA do rozwiązywania problemu klasyfikacji pojedynczych zmian nukleotydowych.

W pracy wykorzystano nadzorowane techniki uczenia maszynowego. Zbiór wejściowy stanowiły zweryfikowane przykłady funkcjonalnych wariantów polimorficznych położonych w promotorach genów oraz zmiany neutralne wraz z ich charakterystyką: otoczeniem sekwencyjnym, zawartością par GC (guaniny i cytozyny), lokalnym kształtem DNA (szerokością mniejszego rowka, zwinięciem, skrętem „śmigła” oraz skrętem helisy), modyfikacjami histonowymi, wiązaniem czynników transkrypcyjnych, nadwrażliwością na DNazę I oraz zachowaniem ewolucyjnym. Wykorzystując powszechnie stosowane miary oceny jakości modeli takie jak dokładność, precyzja, czułość oraz wartość F1, dokonano wyboru najlepszego modelu, który następnie zwalidowano na zewnętrznym zbiorze oraz porównano z wybranymi, opublikowanymi narzędziami przewidywania konsekwencji funkcjonalnych pojedynczych zmian nukleotydowych w regionach niekodujących genomu.

Wyniki przedstawionej pracy wskazują, że zawartość par GC w otoczeniu sekwencyjnym wariantów nukleotydowych, a także lokalny kształt DNA stanowią istotne źródło informacji dotyczącej prawdopodobieństwa, z jakim wprowadzona zamiana niesie za sobą konsekwencje funkcjonalne. Na dzień dzisiejszy opracowany klasyfikator stanowi jedyny opublikowany model skoncentrowany na promotorach oraz wykorzystujący informacje dotyczące szerokości mniejszego rowka, zwinięcia, skrętu „śmigła” oraz skrętu helisy do oceny funkcjonalności pojedynczych zmian nukleotydowych.

Słowa kluczowe: polimorfizm pojedynczego nukleotydu, zmienność sekwencji DNA, promotor, kształt DNA, uczenie maszynowe, priorytetyzacja wariantów nukleotydowych