

Poszukiwanie poliformizmów za pomocą sekwencera genomu

Streszczenie

Rozwój nowego pokolenia sekwencerów genomu daje nam szansę lepiej poznać genom człowieka oraz innych gatunków. Wyznaczenie lokusów, gdzie są różne warianty w danej populacji jest ważnym zagadnieniem. Praca ta rozważa testy dla obecności rzadkiego wariantu. Problem polega na rozróżnieniu między odczytami rzadkiego wariantu a błędnymi odczytami. Przedstawiamy prosty test, który wnioskuje, że wariant jest obecny, gdy maksymalna liczba odczytów tego wariantu u jednego osobnika jest większa niż odpowiedni próg. Wartość tego progu zależy od liczebności, parametrów sekwencera i przyjętego poziomu istotności. Drugi test jest oparty na ilorazie wiarygodności i bierze pod uwagę liczbę odczytów potencjalnego wariantu u każdego osobnika. Skoro powtarzamy ten test dla dużej liczby lokusów, korzystamy z procedury Benjamini'ego-Hochberga aby skontrolować FDR (intensywność fałszywych odkryć - false discovery rate).

Symulacje pokazują, że testy te kontrolują FDR i mają wysoką moc. Natomiast, statystyka testowa w drugim teście nie ma ładnego rozkładu asymptotycznego, więc korzystamy z podejścia Bayesowskiego aby oszacować prawdopodobieństwo "a posteriori" tego, że u danego lokusu są różne warianty.