

Analiza niekodujących sekwencji nukleotydowych końca 3' genów przy użyciu nowego, stworzonego do tego celu narzędzia analizy bioinformatycznej.

Roman Jaksik, Joanna Rzeszowska-Wolny

Instytut Automatyki, Wydział Automatyki, Elektroniki i Informatyki Politechniki Śląskiej

Struktura zapisu genetycznego fascynuje naukowców już od czasu odkrycia nośnika informacji genetycznej jakim jest DNA, jednak pomimo wielu lat badań nad znaczeniem sekwencji nukleotydowych szereg pytań pozostaje wciąż bez odpowiedzi. Przykładowo, do dziś nie wiadomo dokładnie, jaką rolę pełnią niekodujące sekwencje nukleotydowe końca 3' genu, tj. sekwencje znajdujące się za ostatnim egzonom. Wiadomo, że obecność niektórych motywów sekwencyjnych z tej części genu jest niezbędna do oddziaływania z białkami cytoplazmatycznymi i wpływa na okres półtrwania cząsteczek mRNA.

Celem przedstawianej pracy było stworzenie narzędzia bioinformatycznego, które umożliwiłoby analizę sekwencji nukleotydowej końca 3' dużej liczby genów charakteryzujących się podobną stabilnością w określonych warunkach i przeprowadzenie analizy dla genów, których transkrypty były destabilizowane przez promieniowanie jonizujące.

Wykonano aplikację, która umożliwia pobieranie bezpośrednio z ogólnodostępnej bazy danych serwera Europejskiego Laboratorium Biologii Molekularnej (EMBL) najnowszych informacji na temat sekwencji wybranych genów, wyodrębnianie z nich niekodującej sekwencji nukleotydowej końca 3', charakteryzowanie pod względem składu nukleotydowego, długości części niekodującej, obecności krótkich kilkunukleotydowych motywów sekwencyjnych oraz klasyfikację i analizę statystyczną wyników.

Narzędzie to wykorzystano do porównania końców 3' genów, których transkrypty wykazywały zmianę poziomu (mogły ulegać stabilizacji lub destabilizacji) pod wpływem promieniowania jonizującego. Komórki białaczki z linii K562 oraz czerniaka Me45 hodowane *in vitro*, były ekspozowane na działanie promieni jonizujących a następnie w czasie możliwie najkrótszym, badano poziom transkryptów wielu genów za pomocą metody mikromacierzy i porównywano z poziomem w komórkach kontrolnych. Na podstawie pomiaru ilości transkryptu odczytanego z sond umieszczonych na mikromacierzy, odpowiednio pogrupowano i poselekcjonowano geny, a następnie porównano ich sekwencje końca 3'.

Wyniki badań wykazały istotne różnice pomiędzy grupami ekspresyjnymi nie tylko w obrębie długości sekwencji końca 3', ale także składu nukleotydowego. Grupa genów, u których zaobserwowano wzrost poziomu transkryptu charakteryzowała się statystycznie większą ilością adeniny i tyminy oraz mniejszą guaniny i cytozyny w stosunku do genów o stałej ekspresji, podczas gdy w grupie genów, których ekspresja

uległa spadkowi było odwrotnie, a więc obserwowano większą ilość guaniny i cytozyny oraz mniejszą adeniny i tyminy.

Różnice występują także w przypadku powtórzeń określonych motywów sekwencyjnych. Jednym z dość charakterystycznych jest sekwencja TAAAT, odpowiedzialna w wielu przypadkach za destabilizację transkryptu. W grupie genów, których ekspresja wzrosła, tego typu powtórzeń jest o 20% więcej niż w grupie genów których ekspresja utrzymywała się na stałym poziomie, natomiast w grupie o zmniejszonej ekspresji, występuje ich z kolei średnio o 29% mniej co sugeruje że ten motyw może w niektórych komórkach pełnić również funkcję stabilizującą.

Przeprowadzone badania ukazują jak duże możliwości daje połączenie wiedzy z zakresu biologii i informatyki, bez której jednoczesna analiza ponad 10 tysięcy genów o łącznej długości sekwencji końca 3' przekraczającej 11 milionów nukleotydów, byłaby niemożliwa.