

Określenie źródła zmienności ekspresji genów w populacjach komórek – analiza zależności między sygnałami epigenetycznymi a trójwymiarową strukturą chromatyny

dr Michał Dąbrowski
Instytut Podstaw Informatyki Polskiej Akademii Nauk
m.dabrowski@ipipan.waw.pl

Opis tematyki

Dzięki analizom ekspresji genów na poziomie pojedynczych komórek (single-cell RNA-seq) możliwe jest wyodrębnienie specyficznych ich populacji, które różnią się między sobą wzorcem ekspresji genów. Celem tego tematu jest wykrycie grup genów o charakterystycznym profilu ekspresji dla poszczególnych populacji komórek, a następnie zbadanie mechanizmów modulujących ich ekspresję na poziomie zmian epigenetycznych (np. metylacja DNA, modyfikacje histonowe) oraz trójwymiarowej struktury DNA (Hi-C).

Realizacja projektu obejmuje wyszukanie oraz integrację zbiorów danych zawierających różne typy danych single-cell oraz danych opisujących trójwymiarową strukturę DNA. Spodziewanym efektem tego etapu będzie przypisanie genów oraz modyfikacji epigenetycznych do poszczególnych domen, jak również opis kontaktów pomiędzy enhancerami a promotorami. W kolejnym etapie porównane zostaną metody, które pozwalają na wyodrębnienie populacji komórek na podstawie analizy zmienności ekspresji genów z danych single-cell RNA. Wyodrębnione zostaną cechy markerowe (RNA) dla poszczególnych populacji komórek. Analizowana będzie ekspresja genów markerowych, charakterystyczna dla poszczególnych populacji komórek, w poszczególnych domenach jak również zmienność sygnałów epigenetycznych w regionach regulatorowych przypisanych do tych genów. Jednym ze szczegółowych zadań na tym etapie będzie wyszukanie promotorów genów mających potwierdzone kontakty z wieloma enhancerami, a następnie ustalenie asocjacji pomiędzy sygnałem ChIP-seq (H3K27ac) tych enhancerów a zmiennością ekspresji genu. Określimy w ten sposób zależność pomiędzy aktywnością enhancerów a poziomem ekspresji genu.

Zgromadzenie i analiza różnych danych single-cell np. ATAC-seq, metylacja DNA pozwoli na określenie mechanizmów regulacji ekspresji genów homogenicznych w obrębie domen, jak również specyficznych w poszczególnych regionach regulatorowych genów markerowych dla poszczególnych populacji komórek. W sytuacji braku interesujących nas danych pochodzących z analiz single-cell będziemy się posilkowali danymi Next Generation Sequencing uzyskanymi dla puli komórek. Na zaawansowanym etapie analiz bioinformatycznych przewidujemy przygotowanie wniosku grantowego, którego celem będzie laboratoryjna, *in vitro* weryfikacja uzyskanych wyników.

Wymagania

- ukończone studia drugiego stopnia z dziedzin biologiczno-, matematyczno- lub informatyczno-pokrewnych
- dobra wiedza z zakresu analizy danych, statystyki
- dobra znajomość zagadnień molekularnych związanych z regulacją ekspresji genów
- bardzo dobra znajomość języka angielskiego
- wysoka umiejętność programowania przynajmniej w jednym języku (np. R, Perl, Python)
- mile widziana znajomość metod sekwencjonowania NGS.