

Biologia Systemów 2024/25: Projekt zaliczeniowy z analizy danych czasoprzestrzennych

Termin oddania: 21 kwietnia 2025, godz. 18:00

Forma oddania: repozytorium na GitHub zawierające kod oraz plik PDF z raportem

Liczba punktów: maks. 10 punktów

Wprowadzenie

Dysponujemy eksperymentalnymi danymi z mikroskopii żywych komórek pochodzących z eksperymentu, w którym monitorowano aktywność dwóch ważnych szlaków sygnałowych: **ERK** oraz **AKT**. W tym celu wykorzystano linię komórkową transfekowaną biosensorami fluorescencyjnymi:

- H2B-miRFP703 (znacznik jądra komórkowego),
- ErkKTR-mTurq.2 (do pomiaru aktywności szlaku ERK, kolumna ERKKTR_ratio),
- FoxO-mNeonGreen (do pomiaru aktywności szlaku AKT, kolumna FoxO3A_ratio).

Eksperyment został przeprowadzony w następujących warunkach:

- Komórki były pozbawione czynników wzrostu przez 48 godzin przed rozpoczęciem akwizycji,
- Hodowla na płytkach 96-dołkowych, powlekanych fibronectiną (0,25 $\mu\text{g}/\text{cm}^2$),
- Akwizycja obrazu następowała co 5 minut przez 24 godziny.

Każda komórka (oznaczona unikalnym `track_id`) posiada serię czasową obserwacji, obejmującą:

- Zmiany aktywności ERK i AKT (`ERKKTR_ratio`, `FoxO3A_ratio`),
- Rozmiar jądra komórkowego (`Nuclear_size`),
- Położenie przestrzenne (`objNuclei_Location_Center_X/Y`),
- Moment czasowy obserwacji (`Image_Metadata_T`).

Dodatkowo, każda obserwacja pochodzi z konkretnego eksperymentu (`Exp_ID`) i miejsca na płycie (`Image_Metadata_Site`). Wartości `Image_Metadata_Site` traktujemy jako wskaźniki powtórzeń eksperymentu. Zbiór danych obejmuje obserwacje dla linii komórkowych z różnymi mutacjami:

- WT (dziki typ), W eksperymencie wykorzystano linie komórkowe o zróżnicowanym profilu mutacyjnym. Oprócz komórek dzikiego typu (WT, ang. *wild type*), analizowano także linie z mutacjami istotnymi dla szlaków sygnałowych PI3K–AKT–mTOR oraz MAPK–ERK:
 - **WT** (ang. *Wild Type*) - typ dziki, brak ingerencji.
 - **AKT1_E17K** – mutacja punktowa w genie *AKT1*, prowadząca do substytucji glutaminianu (E) na lizynę (K) w pozycji 17; skutkuje konstytutywną aktywacją białka AKT1 niezależnie od sygnałów zewnętrznych.
 - **PIK3CA_E545K** – mutacja w domenie helicalnej podjednostki katalitycznej PI3K (gen *PIK3CA*); zamiana kwasu glutaminowego (E) na lizynę (K) w pozycji 545 zwiększa aktywność enzymatyczną PI3K i pobudza szlak AKT.
 - **PIK3CA_H1047R** – mutacja w domenie kinazowej PI3K, gdzie histydyna (H) zostaje zastąpiona przez argininę (R) w pozycji 1047; również prowadzi do nadaktywacji szlaku PI3K–AKT.
 - **PTEN_del** – delecja genu *PTEN*, który pełni funkcję supresora nowotworowego i negatywnego regulatora szlaku PI3K–AKT; jego utrata skutkuje podwyższoną aktywnością AKT oraz zaburzoną homeostazą komórkową.

Waszym zadaniem jest eksploracyjna analiza tego zbioru danych w kontekście biologii systemów, z wykorzystaniem metod analizy czasoprzestrzennej aktywności komórkowej, poprzez odpowiedzi na trzy pytania badawcze – dwa zdefiniowane, jedno swobodne.

Zasady oceniania

- 3 punkty – **jakość kodu w repozytorium**: przejrzystość, użycie funkcji, podział na pliki, README.
- 1 punkt – **czytelność raportu**: tytuły, opisy wykresów, wnioski, przejrzysta struktura.
- 6 punktów – **eksploracje biologiczne zawarte w raporcie** (3 pytania/zadania badawcze po 2 punkty każde):
 - 2 pkt – pytanie 1: opisane poniżej,
 - 2 pkt – pytanie 2: opisane poniżej,
 - 2 pkt – pytanie 3: swobodnie zdefiniowane przez osobę rozwiązującą zadanie.

Aby otrzymać 2 pkt za eksplorację: należy zdefiniować pytanie, przygotować dane, zaprojektować analizę i zwizualizować wnioski. 1 pkt za pracę niekompletną (brak wniosków, kiepska (np. nieinformatywna) wizualizacja, słaba lub brak interpretacji).

Zadanie 1 (2 pkt) – Porównanie aktywności szlaków sygnałowych między mutacjami

Celem tego zadania jest porównanie dynamiki aktywności szlaków ERK i AKT między liniami komórkowymi o różnych mutacjach. Na podstawie wartości `ERKKTR_ratio` i `FoxO3A_ratio`:

- a) Oblicz średnie przebiegi aktywności dla każdej mutacji w czasie i przedstaw je na wspólnych wykresach (z przedziałami niepewności, np. 90% percentylami lub odchyleniem standardowym).
- b) Zinterpretuj zaobserwowane różnice – w jakich momentach i dla których mutacji aktywność sygnału różni się najbardziej?
- c) Przeprowadź testy istotności statystycznej porównujące poziomy aktywności między WT a każdą z mutacji. Możesz użyć np. testu Manna–Whitneya (dla dwóch grup) lub testu Kruskala–Wallisa (dla wielu grup), analizując średnią aktywność w wybranym przedziale czasowym (np. 60–180 minut). W przypadku wielu porównań zastosuj korektę (np. Bonferroniego). Wyniki przedstaw w tabeli i krótko zinterpretuj.

Wizualizacja: wykresy przebiegów sygnału w czasie (osobno dla ERK i AKT) oraz tabela wyników testów statystycznych.

Zadanie 2 (2 pkt) – Analiza koordynacji przestrzennej i dynamiki sygnału

W tym zadaniu skupiamy się na przestrzennym kontekście aktywności sygnałowej oraz możliwej propagacji sygnału w czasie.

Na podstawie wybranego wskaźnika (ERKKTR_ratio lub Fox03A_ratio):

- a) Zdefiniuj stan aktywny komórki (np. przez przekroczenie wartości względem baseline lub lokalnego maksimum).
- b) Dla każdego momentu czasu zidentyfikuj aktywne komórki, a następnie dla każdej z nich oblicz liczbę aktywnych sąsiadów w zadanym promieniu (np. przy użyciu KDTree). Oceń, jak średnia liczba aktywnych sąsiadów zmienia się w czasie i czy zależy od typu mutacji.
- c) (element czasoprzestrzenny) Zastanów się, czy aktywność sygnałowa ma charakter lokalny i czy może rozprzestrzeniać się w sposób skoordynowany w czasie i przestrzeni. Możesz to ocenić np. poprzez wizualizacje aktywności na płaszczyźnie XY w kolejnych przedziałach czasowych, analizę korelacji przestrzennej aktywności w kolejnych klatkach lub inne podejście. Nawet proste obserwacje (np. występowanie „ognisk” aktywacji) będą wartościowe.

Wizualizacja: wykres zmian średniej liczby aktywnych sąsiadów w czasie (dla różnych mutacji), przykładowe mapy aktywności na płaszczyźnie XY w kolejnych momentach czasu (np. zabarwione komórki aktywne).

Zadanie 3 (2 pkt) – Otwarte pytanie badawcze na podstawie danych czasoprzestrzennych

W tym zadaniu masz pełną swobodę w zaprojektowaniu własnej eksploracji danych i postawieniu biologicznego pytania badawczego związanego z czasem, przestrzenią lub mutacjami. Przykładowe inspiracje:

- Czy mutacje wpływają na częstość lub siłę pików aktywności sygnałowej?
- Czy aktywacja zachodzi falowo lub grupowo w przestrzeni?
- Czy aktywność komórki zależy od lokalnej gęstości otaczających ją komórek?

Twoim zadaniem jest:

- a) Sformułować hipotezę lub pytanie badawcze oraz uzasadnić je na podstawie danych.
- b) Przeprowadzić odpowiednią transformację danych (np. wykrywanie pików, agregacja po trajektoriach, obliczanie odległości przestrzennych itp.).
- c) Wykonać analizę oraz przygotować przynajmniej jedną wizualizację wspierającą Twoje wnioski.
- d) Na końcu krótko podsumować uzyskane wyniki oraz ich możliwe znaczenie biologiczne.

Kryteria oceny dla zadania 3:

- **1 punkt** – poprawnie sformułowane pytanie i przeprowadzona analiza o charakterze eksploracyjnym. Przykłady:
 - porównanie liczby pików aktywności między typami mutacji,
 - analiza zmian średniego rozmiaru jądra w czasie,
 - ocena, czy w danych eksperymentach następuje „wzrost aktywności” po określonym czasie.
- **2 punkty** – zaawansowana analiza z wykorzystaniem cech czasoprzestrzennych lub kreatywnego podejścia. Przykłady:
 - wykrycie propagacji aktywacji w przestrzeni (np. poprzez analizę prędkości „fal” aktywności między sąsiadującymi komórkami),
 - klasyfikacja trajektorii czasowych na typy (np. impulsywne, stałe, ciche) i porównanie częstości ich występowania między mutacjami.

Powodzenia i nie wahajcie się eksplorować kreatywnie danych!