

Kierunek: Bioinformatyka
Przedmiot: Analiza danych pochodzących z sekwencjonowania następnej generacji
Semestr: Zimowy 2025/2026

PROJEKT ZALICZENIOWY

- Wybierz i pobierz dane do analiz. Dane muszą spełniać następujące kryteria:
 - dane całogenomowe (Whole Genome Sequencing; WGS)
 - wygenerowane przez platformę Illumina*
 - wygenerowane przez nowsze maszyny (do 5 lat wstecz)*
 - biblioteka paired-end
 - przynajmniej 2 osobniki/próbki
 - dla wybranego gatunku musi być dostępny genom referencyjny
 - dla wybranego gatunku musi być dostępna możliwość przeprowadzenia anotacji polimorfizmów genetycznych*
 - najlepiej gdyby dla danego gatunku były dostępne polimorfizmy genetyczne (vcf ensembl)*
 - sugerowany maksymalny rozmiar 1 pliku (.fastq.gz) to 500MB.
 - każdy zespół studentów analizuje dane z różnych projektów.
- Scharakteryzuj projekt w ramach którego dane zostały uzyskane. Opisz krótko problem badawczy.
- Dokonaj kontroli jakości pobranych odczytów, dokonaj filtracji odczytów i ponownie sprawdź jakość odczytów po filtrowaniu.
- Pobierz odpowiedni genom referencyjny oraz przygotuj go do mapowania.
- Dokonaj mapowania pobranych odczytów względem genomu referencyjnego. Pamiętaj o wszystkich procesach, które należy wykonać po mapowaniu oraz o punktach kontrolnych.
- Dokonaj detekcji wariantów typu SNP za pomocą dwóch programów*
- Dokonaj anotacji i analizy funkcjonalnej zidentyfikowanych wariantów.
- Porównaj wyniki pomiędzy programami do detekcji polimorfizmów.
- Porównaj wyniki pomiędzy próbkami (kontekst biologiczny)
- Zwizualizuj najciekawsze wyniki będące rezultatem detekcji polimorfizmów i ich analizy funkcjonalnej.

Wykonanie projektu jest dowolne – nie są narzucane programy do analizy danych NGS. Konieczne jest zastosowanie automatyzacji procesu przetwarzania danych. Automatyzację należy jednak przeprowadzić samodzielnie, nie używamy gotowych „wrapperów” dla całych pipeline’ów.

Prezentacja dotycząca projektu musi zawierać:

- cel badań
- opis poszczególnych etapów analiz
- podsumowanie każdego z etapów
- interpretację wyników (zwłaszcza w kontekście biologicznym)

Skrypty oraz plik z prezentacją należy przesłać w wiadomości email w dniu zaliczenia.

Powyższe elementy będą podstawą do dyskusji. Każda osoba z grupy jest zobowiązana do prezentowania i wzięcia udziału w dyskusji. Na podstawie prezentacji i dyskusji zostanie wystawiona ocena z przedmiotu.

01.10.2025

Dr Magda Mielczarek

*Możliwa zmiana po uzgodnieniu z prowadzącym.