

Zajęcia 3 Różnicowa analiza ekspresji w Galaxy

Przygotowanie danych do analizy

1. Otwórz stronę Galaxy: <http://courses.combio.pl>
2. Znajdź pliki znajdujące się w **Shared data** → **Histories** → **RNA-seq** i skopiuj do swojej historii
 - a. Cd20_Rep1
 - b. Cd20_Rep2
 - c. Gene_Annotations_for_chr19
 - d. H1hesc_Rep1
 - e. H1hesc_Rep2

Analiza jakości

1. Przeprowadź analizę jakości za pomocą **FastQC**. Zapisz wnioski.
2. Usuń końce niskiej jakości i adaptory za pomocą **TrimGalore!**
 - a. Adapter sequence to be trimmed: Illumina universal
 - b. Trim Galore! advanced settings : Full parameter list
 - c. Trim low-quality ends from reads in addition to adapter removal (Enter phred quality score threshold): 20
 - d. Overlap with adapter sequence required to trim a sequence: 5
 - e. Generate a report file: Yes
3. Ponownie przeprowadź analizę jakości.

Mapowanie do genomu

1. Przejdź do RNA-seq → **RNA STAR**
2. Uruchom program na plikach FASTQ otrzymanych w poprzednim kroku z podanymi parametrami
 - a. Select reference genome: Human (Homo sapiens) GRCH38 chromosome 19 - training data
 - b. Gene annotation file: gene annotation GrCh38.95
 - c. Would you like to set output parameters (formatting and filtering)?: Yes
 - d. Filter alignments containing non-canonical junctions: Remove alignment with non-canonical junctions
3. Zapisz do protokołu statystyki mapowania
4. Ile miejsc splicingowych zostało zidentyfikowanych podczas mapowania?

Zliczenie odczytów mapujących do transkryptów

1. Przejdź do RNA-seq → **featureCounts**
2. Uruchom na plikach bam otrzymanych w wyniku mapowania
 - a. Gene annotation file: in your history
 - b. Gene annotation file: gene annotation GrCh38.95
 - c. Output format: Gene-ID "\t" read-count (DESeq2 IUC wrapper compatible)

Analiza różnicowa ekspresji

1. Przejdź do RNA-seq -> DESeq2
2. Parametry:
 - a. Specify a factor name, e.g. effects_drug_x or cancer_markers: cell line
 - b. 1. Factor: CD20 + pliki ze zliczeniami dla CD20
 - c. 2. Factor: Hihesc + pliki ze zliczeniami dla Hihesc
 - d. Visualising the analysis results: Yes
3. Użyj **Filter and Sort** -> **Filter** w celu przefiltrowania wyników pod względem istotności statystycznej: **P-adj <= 0.05 [c7 <= 0.05]**
4. Na pliku z punktu 3. uruchom ponownie **Filter: log2(FC) <= -1 [c3 <= -1]**
5. Na pliku z punktu 3. uruchom ponownie **Filter: log2(FC) <= -1 [c3 >= 1]**
6. Ile genów pozostało po filtrowaniu?
7. Przeprowadź analizę wzbogacenia ścieżek wyników w g:profiler'a