

Studi Kasus 2

Adrian Nugraha Utama

2023-10-28

Pengenalan

Pada studi kasus ini, Anda akan mengeksplorasi reference genome data dari RefSeq dan Ensemble, dan melihat distribusi entri dan gen pada data anotasi tersebut.

Pustaka

```
library(rtracklayer)
# Dan library-library lainnya (dplyr, ggplot2) yang mungkin dibutuhkan
```

Task 1: Cari dan load data RefSeq dan Ensemble

Cari dan unduh data anotasi RefSeq dan Ensemble terbatas yang tersedia secara publik. Petunjuk: Data anotasi tersebut tidak terlalu besar, dan masing-masing hanya sekitar 50 MB.

Coba impor data anotasi yang telah diunduh tersebut ke dalam R, dan bandingkan kolom-kolom meta-datanya!

Task 2: Bar chart distribusi kromosom

Coba buat bar chart distribusi kromosom (seqnames) untuk data anotasi RefSeq dan Ensemble!

Task 3: Perbandingan tipe entri

Coba buat bar chart untuk perbandingan tipe entri (gene, transcript, exon, CDS) untuk data anotasi RefSeq dan Ensemble! Apabila memungkinkan, bandingkan dalam satu grafik yang sama!

Task 4: Gen ENSG yang belum memiliki nama HGNC

Hitunglah berapa jumlah gen ENSG unik, dan juga gen HGNC unik. Berapa persen gen ENSG yang belum memiliki nama HGNCnya?

Optional Task: Biomart dan AnnotationDbi

Coba gunakan package R (Biomart dan atau AnnotationDbi) untuk mendownload data anotasi secara otomatis!