

Studi Kasus 3

Adrian Nugraha Utama

2023-10-28

Pengenalan

Pada studi kasus ini, Anda akan mengeksplorasi dari DESeq2, beserta normalisasi dan tes statistiknya.

Pustaka

```
library(airway) # Dataset - cara akses: data(gse)
library(DESeq2)
# Dan library-library lainnya (dplyr, ggplot2) yang mungkin dibutuhkan
```

Task 1: Presentasikan step-step yang dilakukan oleh DESeq2

Carilah step-step yang diambil DESeq2 dalam melakukan analisis differential gene expression, dari bentuk raw data berupa count matrix, sampai pada hasil diferensiasi gen.

Task 2: Normalisasi pada DESeq2

Coba plotkan total count data untuk masing-masing sample pada DESeq, dan bandingkan sebelum dan sesudah proses normalisasi data!

Task 3: Grafik dispersi

Coba tampilkan kurva dispersi gen, beserta hasil shrinkage dan fitting, pada DESeq2. Petunjuk: Pada DESeq2, terdapat suatu in-built function untuk melakukan bagian ini secara otomatis.

Task 4: Koreksi Bonferroni dan Benjamini–Hochberg (BH)

Bandingkan jumlah total gen yang terseleksi untuk kasus-kasus berikut:

1. tidak ada koreksi pada p-value
2. koreksi Benjamini–Hochberg (BH) pada p-value (default DESeq2)
3. koreksi Bonferroni pada p-value
4. gen yang terseleksi dengan kasus 2 dan 3 secara bersamaan

Tampilkan juga dalam suatu bentuk grafik. Apa yang bisa kita simpulkan dari pengamatan tersebut!

Optional Task: Simulasi data count matrix beserta false discovery rate

Seyogianya, gen yang berasal dari distribusi yang sama tidak akan terseleksi. False discovery rate menunjukkan persentase gen yang salah terseleksi tersebut. Coba buat count matrix fiktif (dari angka acak), dengan gen yang terdistribusi secara negative binomial dan tingkat ekspresi (mean dan variance) yang berbeda-beda terhadap gen, namun sama terhadap sampel. Berapakah false discovery rate yang bisa diekspektasi dari kasus-kasus pada Task 4 di atas?