M2 Mycologie

Analyses RNA-Seq

Université
Paris Cité

Principe des analyses RNA-Seq

- revues de bonnes pratiques : [2, 5]
- outils statistiques : Bioconductor ; [1, 3]

Etapes d'analyse

1. DNA extraction from a sample
2. DNA sequencing
3. [*] Alignment of sequencing reads to a reference genome
4. Basic exploratory data analysis
5. Identification of genomic variants (SNPs, small insertions and deletions)
6. Gene quantification (i.e., statistics on count data)

Principes de l'analyse

Basically, NGS analyses (RNA, CHIP, etc.) need to account for within-group variance estimates when analysing lot of genes, hence the need to pool information across genes. The DESeq approach detects and corrects dispersion estimates that are too low through modeling of the dependence of the dispersion on the average expression strength over all samples. In addition, it provides a novel method for gene ranking and the visualization of stable estimates of effect sizes [4]. The DESeq2 package further includes shrunken fold changes (with SE).

Kallisto (ultra-fast mapping)

SRA ou Fastq -> TPM

target_id	length	eff_length	est_counts		tpm
jgi|Podans1|100005|CE99640_2373	1994	1895	1407		51.6298
jgi|Podans1|100060|CE99695_1372	3146	3047	504		11.502
jgi|Podans1|10015|CE9650_8706	1582	1483	1156		54.2042

Attention : l'ID du transcrit doit correspondre exactement à l'ID du gène dans le fichier GFF3 d'annotation.

Fichier d'annotation (GFF3)

```
##gff-version 3
-%<-----
38763:scaffold_3^^Iprediction^^Igene^^I644807^^I646915^^I0^^I+^^
38764:scaffold_3^^Iprediction^^ImRNA^^I644807^^I646915^^I.^^I+^^
```

Références