



Dec 15, 2023

FASTQ alignment, gene counts, and differential expression analysis and functional annotation of DEGs

DOI

dx.doi.org/10.17504/protocols.io.dm6gp3mzdvp/v1

Rebecca Wallings¹

¹University of Florida



rebeccawallings

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.dm6gp3mzdvp/v1>

Protocol Citation: Rebecca Wallings 2023. FASTQ alignment, gene counts, and differential expression analysis and functional annotation of DEGs. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.dm6gp3mzdvp/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: December 15, 2023

Last Modified: December 15, 2023

Protocol Integer ID: 92337

Keywords: functional annotation of degs fastq alignment, degs fastq alignment, differential expression analysis, gene count, gene

Funders Acknowledgements:

ASAP

Grant ID: ASAP-020527

Abstract

FASTQ alignment, gene counts, and differential expression analysis and functional annotation of DEGs

Troubleshooting

FASTQ alignment, gene counts, and differential expression analysis

- 1 FASTQ files were aligned against the mouse genome (GRCm39) and GRCm39.107 annotation using STAR to generate BAM files. Gene counts were generated from BAM files using Rsamtools (Bioconductor package) and the summarizeOverlaps function with the GenomicAlignments package v1.36.0. Differential gene expression analysis was performed with DESeq2 package v1.40.2 using the “DESeq” function with default settings which fits a generalized linear model for each gene. Subsequent Wald test P-values are adjusted for multiple comparisons using the Benjamini–Hochberg method (adjusted P-value). Pair-wise changes in gene expression levels between groups were used to identify DEGs. DEGs will be defined as an absolute log2 fold change ≥ 0.5 and an adjusted P-value ≤ 0.05 .

Functional annotation of DEGs

- 2 Gene Ontology (GO) enrichment analysis was performed with goseq to identify enrichment in gene ontology categories and KEGG pathways. For DEGs, up- and down-regulated gene lists were analyzed separately. Over-represented P-values were adjusted for multiple comparisons using the Benjamini–Hochberg (BH) adjustments for controlling false-discovery rates. An enrichment score was calculated using an observed-over-expected ratio for each gene list. GO-BP categories and KEGG pathways were plotted if their BH adjusted FDR reached ≥ 0.05 and the number of DEGs within each category/pathway was greater than 4. To eliminate larger, broad GO-BP parent categories, categories with more than 250 total genes were not plotted. To eliminate redundancy, pathways with significant overlap regarding genes and semantics were merged.