RNA-seq Differential Expression (DE) Toolkit

Alexis Norris[1], Mamatha Garige[2], Fu-Jyh Luo[3], Mike Mikailov[3], Nicholas Petrick[3], Heather Lombardi[1], Mayumi Miller[4]

[1]Center for Veterinary Medicine (CVM), Office of New Animal Drug Evaluation (ONADE), Division of Animal Bioengineering and Cellular Therapies (DABCT)
[2]Center for Drug Evaluation and Research (CDER), Office of Biotechnology Products (OBP), Division of Division of Biotechnology Review and Research I (DBRRI)
[3]Center for Devices and Radiological Health (CDRH), Office of Science and Engineering Laboratories (OSEL), Division of Imaging Diagnostics and Software Reliability (DIDSR)
[4]Center for Veterinary Medicine (CVM), Office of Research (OR), Division of Animal and Food Microbiology (DAFM)

Background

Biomarker discovery with RNA-seq

Bulk RNA sequencing (RNA-seq) is a powerful and widely used tool to identify biomarkers. The most common RNA-seq analysis is Differential Expression (DE) combined with Gene Set Enrichment Analysis (GSEA; or "pathway analysis"), for which there are well-benchmarked analysis tools available.

Quick, inexpensive analysis with DE toolkit

However, these tools are relatively inaccessible to scientists without bioinformatics expertise. Thus, we sought to create an end-to-end toolkit for DE analysis that could be quickly employed by all FDA scientists, using existing resources and with minimal cost to the user.

Method

Our toolkit makes use of two user-friendly, open-source interfaces: Galaxy and RStudio. First, CDRH's Galaxy[1] is used to trim the sequencing reads (trimmomatic), align (HISAT2), and generate gene expression counts (featureCounts). Then, RStudio software is used to identify DE genes (edgeR) and perform GSEA (clusterProfiler). Finally, RStudio Shiny Apps are used to explore and visualize the DE and GSEA results.

1. **Galaxy** – quantify gene expression.
2. **RStudio** – identify differentially expressed genes and pathways, as well as data visualization.

Demo Dataset[3]:

Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown

Mihaela Pertea, Daehwan Kim, Geo M Pertea, Jeffrey T Leek & Steven L Salzberg

Nature Protocols **11**, 1650–1667(2016) | Cite this article

Discover. Visualize.

Quickly get to the biology in your RNA-seq experiments!

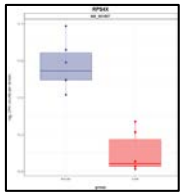
Join the discussion
9/30/20 11AM-1PM ET:
https://fda1.webex.com/fda1/j.php?MTID=mc12697abecabe26cd544262ea7f3085a

Contact:
Alexis.Norris@fda.hhs.gov

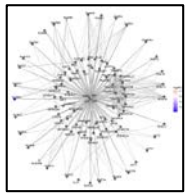
Results

The entire analysis can be performed on a laptop, and typical analyses (6-12 samples) can be completed within one day. Galaxy runs in a web browser without additional cost; RStudio[2] software is free and can be run on PC, Mac, or Linux computers. The Shiny Apps allow users to generate custom, publication quality figures and tables without costly bioinformatics support.

Dysregulated genes



Disrupted pathways



Conclusions & Next Steps

We have created a toolkit for FDA research and review scientists to perform RNA-seq analysis without the need for bioinformatics expertise. The analysis is low cost, quick, reproducible, and available to FDA scientists to support regulatory research and evaluation.

- ☑ Add tools to Galaxy
- ☑ Generate Rmd files for statistical analyses in RStudio
- ☑ Generate RShiny apps for visualization
- ☑ Test Rmds/RShiny on PC with beta users
- ☐ Test Rmds/RShiny on Mac/Linux (underway)
- ☑ Now also available on precisionFDA[4]

References

1. www.usegalaxy.org
2. https://rstudio.com/
3. ftp://ftp.ccb.jhu.edu/pub/RNAseq_protocol
4. https://precision.fda.gov/

Acknowledgments

- Susmita Ghosh (CDER)
- Carole Sourbier (CDER)
- Daniel Tadesse (CVM)
- Zhuoming Lui (CVM)