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. 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.

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.

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; RStudio2 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.

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.

Contact: Alexis.Norris@fda.hhs.gov

References
1. www.usegalaxy.org
2. https://rstudio.com/