Instructor Guide: Mini-Mouse Tutorial

Purpose: This tutorial is designed for the RNA-seq analysis novice who has a basic conceptual understanding of Illumina paired-end sequencing. We structured Mini-Mouse as a prequel to the large-scale Jax RNA-seq project.

Data: The publicly-available data in the tutorial were retrieved from the European Nucleotide Archive (see Exercise 2 for accession information) but have not been published. The experimental underpinning for the tutorial is a mouse Eya2-knockout versus a control, with RNA extracted from retinas at midnight (mid-scotophase) in – presumably – adult mice. We excerpted 14 contiguous genes on chromosome 5 that represent a range of potential results one would find in a large-scale project: up-regulation, down-regulation, high and low expression, significant and nonsignificant differences, splice variants, protein-coding and non-coding genes, well-studied and less-known genes, and genes with interesting potential relationships with the Eya2 knockout (but not necessarily known, published associations).

Galaxy: We made a pedagogical decision to use Galaxy's GUI-interfaced tools instead of command-line tools. Our goal was to remove all possible barriers to the novice student (and potentially novice instructor) and to maximize accessibility. The tutorial could readily be reworked for command-line or any other RNA-seq analysis pipeline; the data are simple fastq files. Within Galaxy, we selected the Tuxedo suite of tools (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3334321/).

Visualization: We consider this to be a critically important component to learning as it removes the potential abstraction of working with genomics data. The Mini-Mouse small dataset simplifies this visualization process. Furthermore, the use of Excel to work with some aspects of data analysis provides students with a sense of familiarity – the idea that big-data genomics, when scaled down, has readily-identifiable components.

Hardware, Software, Handouts: If students are using their own computers, we recommend having them download software ahead of time in case of hardware/software issues. We find that providing students with printed handouts allows for maximum use of the computer screen for analysis, and for easier navigation back and forth through steps. An accompanying electronic copy allows students to click on live links.

Scalable Format - Recommendations:

Short Version: Steps 1-23 – the Galaxy data processing sequence - can potentially be completed in a typical 3-hour lab period. Note, however, that the public instance of Galaxy may not always run at optimum speed.

Long Version: 2 lab sessions or full-day workshop: The entire tutorial, including analyses downstream of Galaxy, can more feasibly be completed in two lab periods or a day-long session.

Exercise 2: If running this on the public Galaxy server, we recommend allowing students a week in case of slow running times, errors, trouble-shooting, etc. A typical run-time for TopHat, the most time-consuming step, is 12-24 hours, but this can be variable.

Part 1: Mini-Mouse Tutorial:

Preparation for the skills needed to run the full-scale Jax experiment:

- Mechanics and navigation of Galaxy
- Navigation of tools and web resources
- Comprehension of purpose of each step in the analysis pipeline
- Troubleshooting
- Critical thinking about experimental design
- Data management
- Interpretation of results – quantitatively, conceptually
- Downstream pathway analysis

Part 2: Bigger Dataset Assignment:

- Reinforce skills from Mini-Mouse
- Reinforce experimental design concepts
- Build confidence in ability to run a full analysis of a big dataset

Additional Pathway Analysis Tools and Information:

While Mini-Mouse uses only two pathway analysis tools, numerous tools have been developed for differential gene expression downstream analysis. Each offers different ways to visualize relationships among differentially-expressed genes. The best approach is to try an assortment of these.

See this website for insights:

http://www.gettinggeneticsdone.com/2012/03/pathway-analysis-for-high-throughput.html

Also see:

Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002375

Tools:

GeneCards: Basic Gene-by-Gene information http://www.genecards.org/

GO – Gene Ontology Enrichment Analysis: http://www.geneontology.org/

GO PANTHER: http://pantherdb.org/

STRING: http://string-db.org/

ENRICHr: http://amp.pharm.mssm.edu/Enrichr/ (meta-database)

GORILLA: http://cbl-gorilla.cs.technion.ac.il/

MSigDB: http://software.broadinstitute.org/gsea/msigdb

DAVID: https://david.ncifcrf.gov/

Consensus Path DataBase – CPDB: http://cpdb.molgen.mpg.de (meta-database)

iPathwayGuide (free for a few days): https://www.advaitabio.com/ipathwayguide.html

CPDB: http://cpdb.molgen.mpg.de/

GeneMANIA: http://genemania.org/ (good for smaller subset of genes or a single gene of interest – high-density information)

Reactome: http://www.reactome.org/ (an especially rich/dense pathway visualization tool)

WikiPathways: http://www.wikipathways.org/index.php/WikiPathways (a good way to view particular pathways once you've identified a pathway you're interested in – not for analysis of list of genes)