M2D8: Mod2 data analysis

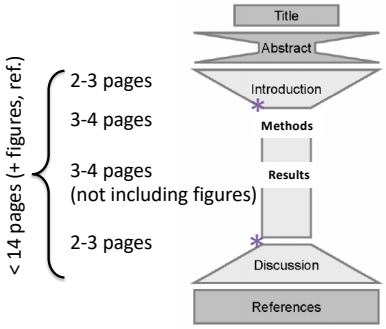
1. Start R.studio.cloud exercise 4
2. Complete Ex.3 and qPCR analysis

Extra Help for Mod2:

- Today we will give you code for Ex 3
- Thursday we will help you through all Ex4
- Outline of exact figures you should include on wiki

1

Mod2 Research Article



- Title, Abstract (10%)
- Introduction (10%)
- Methods (20%)
- Results, Figures and captions (50%)

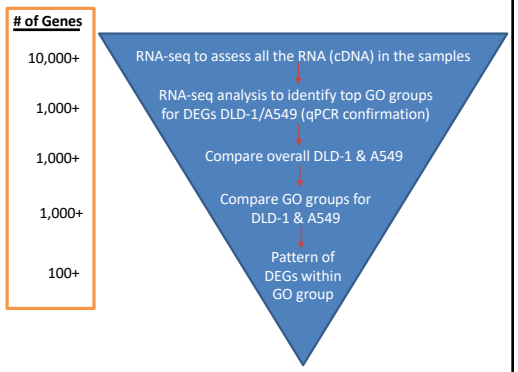
-Top 5 Up GO terms, top 5 Down GO terms (Ex2)
- qPCR
-Plots from Ex3: PCA and heatmap
-Plots from Ex4: GO heatmaps and scatterplots

- Discussion (10%)
- References (not included in page count)

2

Big data figures deserve a large scope introduction

- Figures are documenting the analysis of thousands of genes
- Focused on identifying patterns of differential gene expression instead of identifying individual genes
  - Only looking at individual genes to confirm data set
- Introduction should cover more than what etoposide does to a single cell type or a subset of DNA damage genes



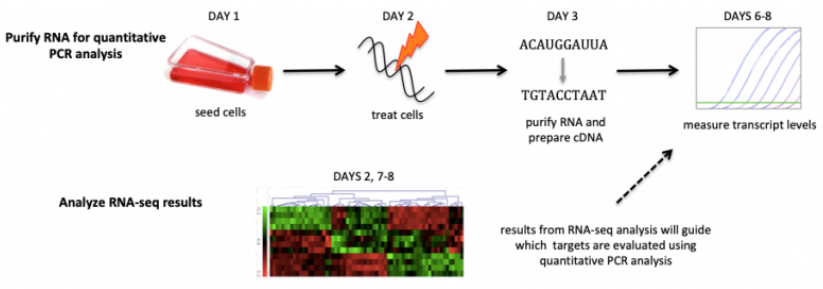
3

Assignments tab→ Research Article→ Results

1. Figure 1
   - experimental overview / schematic illustrating the work-flow (just the key steps!) used in your research project
2. Figure 2 (this figure should include three panels)
   - Panel A: tables with top 5 GO terms in DLD-1 and DLD-1 + etoposide
   - Panel B: histogram containing the qPCR results for the genes of interest, including statistics
   - Panel C: heatmap comparing genes of interest across DLD-1 qPCR data, DLD-1 RNA-seq data, and A549 data
3. Figure 3 (this figure should include two panels)
   - Panel A: plot of PCA data showing DLD-1 + etoposide and A549 + etoposide
   - Panel B: heatmap comparing DLD-1 + etoposide and A549 + etoposide
4. Figure 4
   - heatmap comparing 4 GO terms
5. Figure 5
   - scatterplots generated from the GO terms used in Fig. 3

4

Mod2: Experimental overview



5

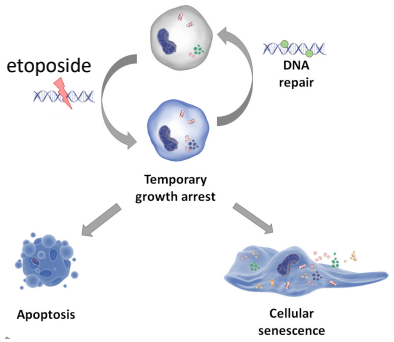
Overview of the purpose of RNA sequencing

- Understanding the sum of mRNA in a cell or organism (called transcriptome) is key if we are to connect the information about our genes with protein expression
- RNA-seq can suggest which genes are turned on or off in a cell by their level of expression
- This allows scientists to more deeply understand the biology of a cell and assess changes that may indicate disease.
- RNA-Seq has the potential to identify new disease biology
- These results could further highlight more effective prevention, diagnostics, and therapy
- RNA-Seq data can provide a unique snapshot of the transcriptomic status of a disease and look at an **unbiased** population of transcripts that allows the identification of novel transcripts that would not be detected through other technologies
- How would you determine the effect of an unknown drug?

technologynetworks.com, RNA-seq: Basics, Applications and Protocol, by Ruairi J Mackenzie

6

How does gene expression change upon etoposide treatment in DLD-1?



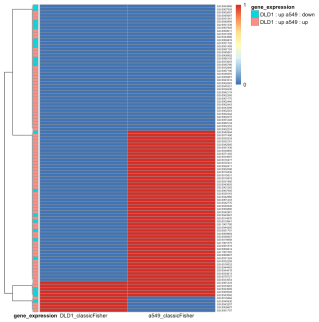
Top up and down gene ontology terms:

- response to stimulus
- cell communication
- Signaling
- Signal transduction
- cellular response to stimulus
- chromosome organization
- RNA splicing
- biogenesis/cellular component organization
- RNA splicing
- mRNA splicing

Soto Gamez et al. Regulation of Survival Networks in Senescent Cells: From Mechanisms to Interventions. JMB July 2019

7

Top 100 enriched GO terms across DLD-1 and A549



- Blue= significant by Fisher statistic
- Red= not significant
- Top 100 was a value we chose for analysis

- Fisher exact test looks at the observed number of significantly differentially expressed genes (p-value cutoff 0.1) assigned to each GO term and compares it to the expected number of significant genes at random
- **Read prompts on the wiki carefully and address questions in your benchling notebook**

8

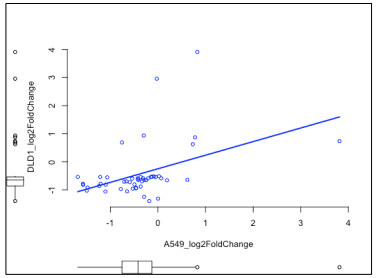
GO terms associated with qPCR gene choices and Ex. 4 analysis

Gene Ontology term	Abbreviation
RNA splicing	RNA_spl_genes
Cell adhesion	cell_adhesion_genes
Cell proliferation	cell_pro_genes
Regulation of mitotic cell cycle	Reg_mcc_genes

Note: GO terms on M2D6 gene list

9

Notes on Interpreting Scatterplots



- Comparing DLD-1 and A549 L2FC of genes in RNA splicing GO term
- Blue dots represent DEGs in this GO term
- Axis = box plot and black dots are DEG for one cell line. Black circles on the axis are the points which fall outside the quartiles (25-75%).
- The blue line is the correlation/regression line, and the slope tells us if it's a positive/negative correlation or if there is no correlation
- **NOTE**: just because one GO annotation/pathway is "upregulated/downregulated" doesn't mean that every gene is expressed in the same direction
- Some genes associated with the GO term are driving upregulation/downregulation of the pathway.
- Not all are similarly expressed, and some genes may even be expressed in the opposite direction.

10

M2D8 "Lab" Checklist

1. Work through the thought questions on the wiki introduction in your benchling notebook.
2. Ask questions and understand the RNA-seq analysis for Ex3 and start Ex4
3. Complete qPCR analysis with confidence interval and Student's t-test statistical analysis
4. Ask questions!!

11

M2D9HW: Outline of figures

- you don't need to draft actual figures
  - 1 sentence: describes the figure
  - 1 sentence: motivation
  - 1 sentence: transition
- Figure order can be found on under assignments tab -> Research Article

12

Details for methods RNA-sequencing and analysis

- **Sequencing:** HiSeq 2000 sequencing at the Massachusetts Institute of Technology BioMicro Center.
- **Data analysis:** performed according to a workflow developed by Amanda Kedaigle, Anne Shen and Ernest Fraenkel at the Massachusetts Institute of Technology using Rstudio.cloud.
- DESeq2 (v. 1.26.0)
- Transcriptomic data for A549 cell line was obtained from the Gene Expression Omnibus (Wang 2017).

13

Mod2 major assignments

- **Research Article (20%)**
  - individual, submit on Stellar
  - due Monday April 20th at 10pm
  - format: word document, figures can be submitted separately
- **Journal Club Presentation (17.5%)**
  - presentation **slides** due on Stellar April 11th 10pm
  - Presentation **video** due to Dropbox April 11th 10pm
- Lab quizzes ~~M2D7,~~ M2D9
- Homework and Notebook (10%)
- Blog (5%), 3 posts for full credit
  - ~~4/6 at 10 pm,~~ 4/13 at 10 pm, 4/21 at 10 pm, 5/12 at 10pm

14