

## M2D6: Analyze RNA-seq data and select gene targets for quantitative PCR experiment

1. Prelab
2. Analyze RNA-seq data
  - R Clustering refresher
  - R Exercise #3
3. Choose genes from qPCR list



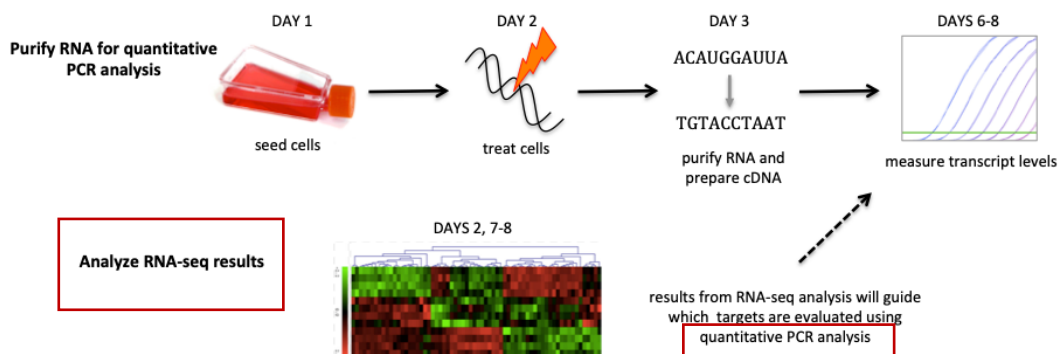
## The new normal: Class

- Prelab and office hours via Zoom
  - Can also be reached via email
- Each prelab will have slides posted 1-2 hours prior to the beginning of class
- Instructors and Joe will be available for entire class time to field questions
  - There will also be a Benchling notebook devoted to questions, especially for R
- Each prelab will be recorded and posted on the wiki for review purposes
  - I'd love to see you in video, but that is optional if you prefer privacy
- To ask or answer questions during class:
  - Raise hand or type something in the chat box and we'll stop to answer your question

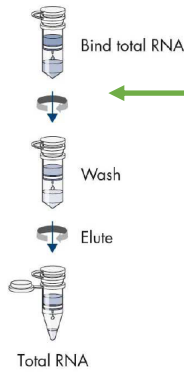
## The new normal: Homework/Quizzes

- Joe will be checking benchling notebooks 24hrs following the beginning of lab to see your progress (i.e. Wed. class is checked at 1pm EDT on Thurs.)
- Homework is due via Stellar by 10pm (EDT) the day of the lab session to be on time
- Homework will be returned via Stellar
  - See M1D7 and M2D2 for recent homework graded
- Quizzes will be emailed at the beginning of lab time and must be posted to Stellar by 10pm (EDT) on the same day to be on time
  - Still only a 10-15 minute quiz

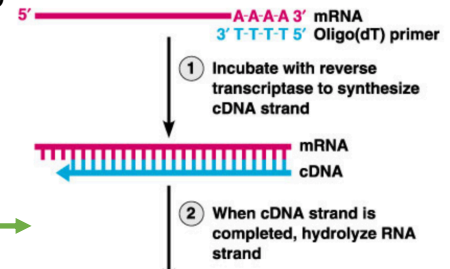
## Overview of Mod2 experiments



## Review of M2D3



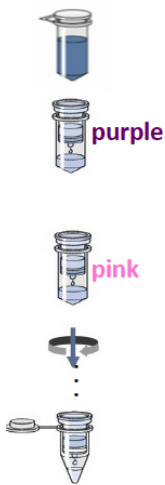
- Purification of RNA from etoposide-treated DLD1 cells
- Qiagen RNeasy kit (similar to miniprep)
- Generation of cDNA from RNA using reverse transcriptase
- Superscript III Reverse Transcriptase



- Wiki has prelab with notes added
- If you want to talk through the process during class or office hours, let me know

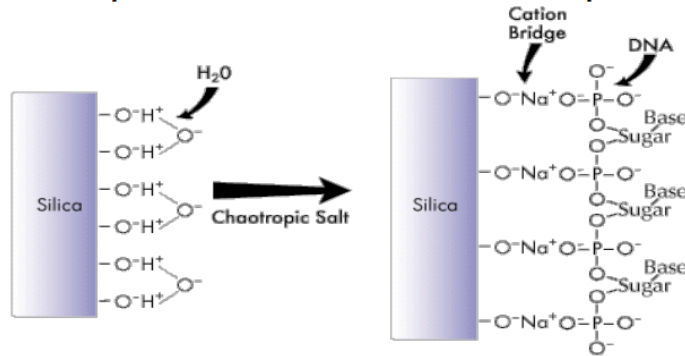
You will include methods through M2D3 in your research article (like you do for other instructor-based experiments)

## Isolate RNA: QIAshredder + Rneasy kit



steps	contents	purpose
lyse	RLT (with highly denaturing guanidine-thiocyanate salt)	Lysis buffer with chaotropic salts (i.e. guanidine-thiocyanate salt) disrupts hydrogen bonds of water and hydrophobic interactions of proteins, but not DNA or RNA
lyse	QIAshredder (purple column)	Shears membranes to release RNA and improve extraction
prepare	70% ethanol	Promotes binding of RNA to silica column (see next slides for details)
bind	silica membrane (pink column)	Binds RNA backbone to extract it from the solution (optimized for mRNA)
wash	RW1 RPE	Washes away molecules not bound to the silica ** after this wash, important to get rid of <u>all</u> ethanol (ethanol in the buffers)
elute	water, RNase-free	Elutes RNA from silica (see next slides for details)

## Chaotropic salts and Ethanol help DNA/RNA bind to column



EtOH and chaotropic salts disrupt hydrogen bonds of water to promote binding of the RNA backbone to the silica

## What genes are differentially expressed in response to DNA damage?

*How are we addressing this question?*

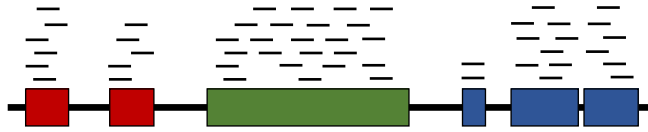
DLD-1 cells: Compare 2 groups 

- No Treatment
- 60min Etoposide

We are using 2 methods to compare mRNA expression in the drug-treated vs control cells

1. High throughput Illumina RNA-seq
  1. Assess all mRNA/cDNA present in sample
  2. Analyzed through DESeq2 in the R environment
2. qPCR with SYBR green
  1. Design primers against cDNA of gene of interest
  2. Determine expression compared to housekeeping gene

## Ex2 Review: RNA-seq data was pre-processed

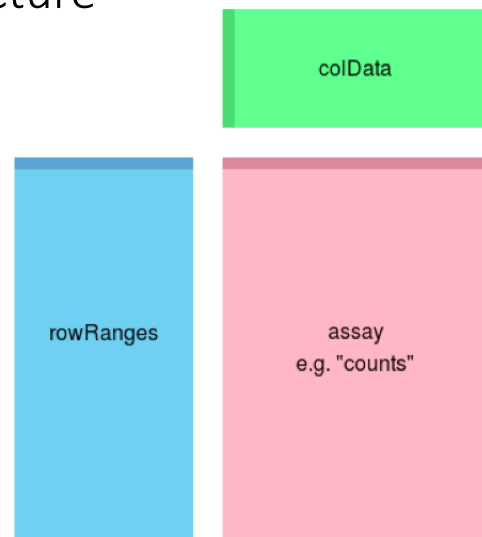


- Data from sequencer provided reads
- Reads were aligned to reference genome
- Aligned reads were counted
- RPKM corrects for differences in sequencing depth and gene length
 
$$\text{RPKM} = \frac{\text{number of reads mapped to gene sequence}}{\text{gene length (kb)} \times \text{total number of reads (millions)}}$$
- The counts of aligned RNA-seq reads were loaded into the DESeqDataSet

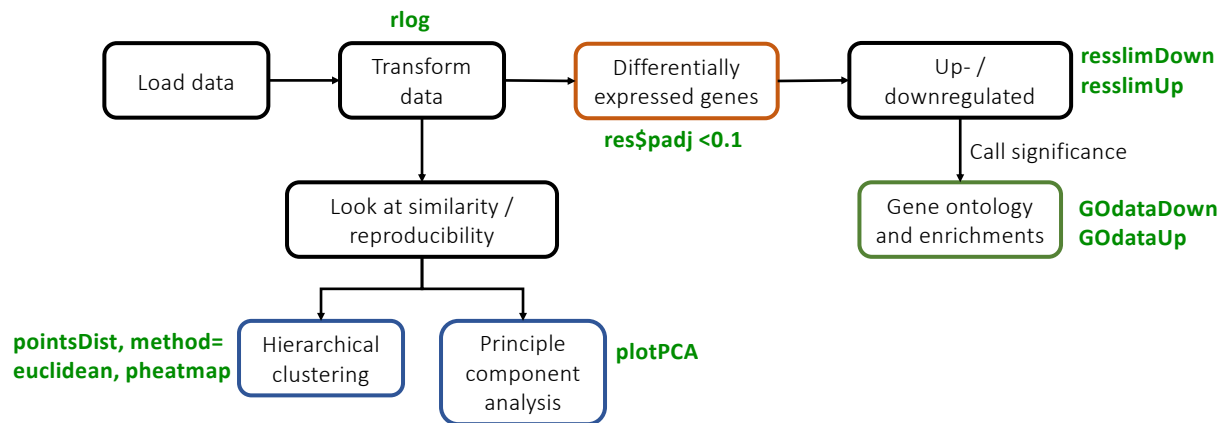
## Ex2 Review: DESeqDataSet structure

Reads aligned to genes were loaded into data structure called “DESeqDataSet”

- colData: sample information
- rowRanges: gene information
- assay: matrix of counts assigned to each gene for each sample



## Ex2 Review: Workflow RNA-seq analysis



R notes from Leslie

Image from Casper Enghuus, Sp17 TA 20.109

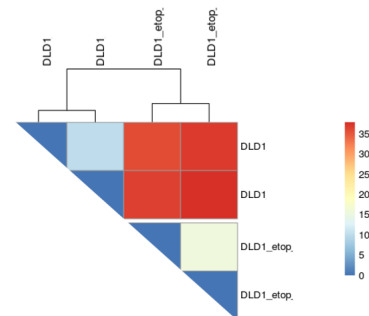
## Ex2 Review: Hierarchical clustering groups similar objects

How are samples clustered?

- Algorithm grouping similar objects into “clusters”
- Endpoint: set of clusters
  - Each cluster is distinct from another cluster
  - Objects in a cluster are broadly similar

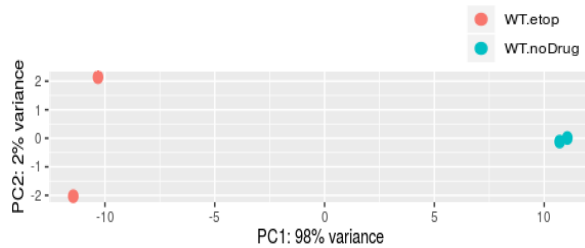
What do graph components represent?

- Rows/columns = samples
- Color (value) = Euclidean distance between samples



What is the scale?

## Ex2 Review: Principal component analysis (PCA) shows relatedness of objects



- Algorithm to reduce dimensionality of data while keeping variation
- Highlights patterns within data and aids interpretation
- PC1 captures the most variance, and so on...
- Which difference captures the most variance (98%)?

## Ex2 Review: Gene ontology (GO) terms based on gene product properties

GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	classicKS
GO:0031668	cellular response to extracellular stimu...	12	8	1.78	1	4.2e-05	0.00013
GO:0010389	regulation of G2/M transition of mitotic...	30	7	4.45	260	0.13535	0.00019

### • GO table terms:

- **GO ID:** Numerical identifiers of GO group
- **Term:** Describes broad properties of gene set
- **Annotated:** number of genes in our data set annotated with this term
- **Significant:** number of significantly differentially expressed genes (DEGs) annotated with that term
- **Expected:** under random chance, number of DEGs expected in that term
- **Classic Fisher:** p value determined with Fisher's test
- **Classic KS:** p value determined with Kolmogorov–Smirnov test

## Classic Fisher vs Classic KS

These tests answer the same question: Do we have gene enrichment for any GO terms?

### Fishers Exact Test

- compares the expected number of significant genes at random to the observed number of significant genes to arrive at a probability.
- In our scripts, this p-value is 0.1
  - `geneSel = function(allScore)`  
`{return(allScore<0.1)}`

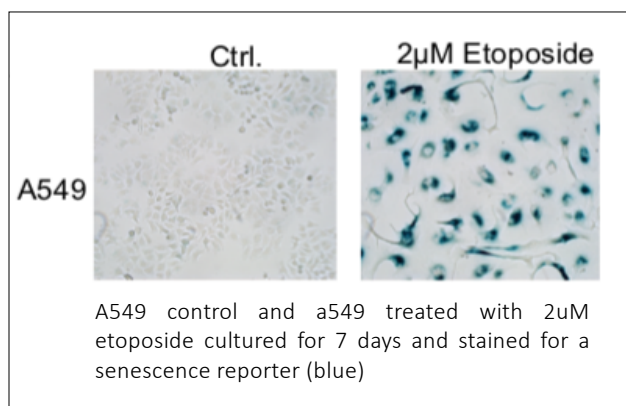
### KS Test

- compares the distribution of gene p-values expected at random to the observed distribution of the gene p-values to arrive at a probability
- KS is theoretically the better choice because it does not require an arbitrary p-value threshold.

Both methods are acceptable and used in RNA-seq analysis

Adapted from Leslie McClain and Anne Shen

## Apply R workflow from Ex2 to RNA-seq dataset in Ex3



- Authors studying senescence induction as an approach to cancer treatment
- A549, model cell line for lung cancer
- Treated with 2uM etoposide, harvested RNA for sequencing after 7 days
- RNA-seq read counts were made available as a public data set

Wang et al. High-Throughput Functional Genetic and Compound Screens Identify Targets for Senescence Induction in Cancer. Cell Reports 2017.



## Getting help with R:

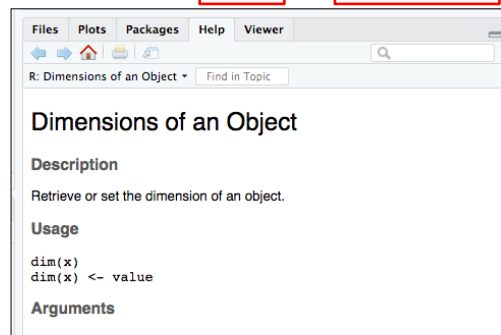
- Ask Anne questions during lab: on Zoom from 3:30-4:30EDT
- Review 20.109.Ex2.codeExplained.pdf under Ex2
- Ask questions on the Mod2 R.studio.cloud benchling page
- Make an appt with \*new\* BE data lab! mit.mywconline.net
- Use R help function

**Method 1**  
`?function`

or

**Method 2**  
`help(function)`

Example: Type `?dim` or `help(dim)`



## Exercise 3 Checklist

- Complete Exercise3\_clustering\_refresher.R
- Generate PCA plot of A549 data
- Generate GO tables of top upregulated and downregulated A549 genes in response to etoposide treatment, with statistical tests
- Generate PCA plot comparing etoposide treatment in DLD1 and A549 cells
- Create heatmap of DLD1 and A549 datasets

## M2D6 “Lab” Checklist

1. Work through the RNA-seq data analysis
  - this analysis will translate to figures in your research article
2. You must choose genes for qPCR analysis, note this in your Benchling notebook

## M2D7 Homework

- Write methods for M2D1-M2D3
  - Omit RNA-seq data analysis practice
- Introduction for Research article
  - Write first paragraph of introduction
  - Write topic sentences for each subsequent paragraph in introduction
    - Include references throughout!
  - Submit annotated reference list
    - Write brief summary of each referenced source and to support why you chose it



## Methods Notes:

- Include enough information to replicate the experiment
  - list manufacturers name, like (Qiagen)
- Organize methods into subsections with descriptive titles
  - Put in logical order
  - Begin with topic sentence to introduce purpose
  - R subsection, include package and version, **DESeq2 (v. 1.26.0)**
- Use clear and concise full sentences
  - NO tables and lists
  - Passive voice and past tense
- Use the most flexible units
  - Write concentrations (when known) rather than volumes
- Eliminate 20.109 specific details
  - Example “labeled Row A, Row B...”
  - Do not include details about tubes and water!
  - Assume reader has some biology experience
  - Include steps teaching faculty carried out for you

## Tissue Culture:

TK6 cells were grown in a flask with 12ml RPMI

supplemented with FBS. The cells were kept in an

incubator at 37°C. A stain was used to assess if the cells

were alive or dead.

## Mod2 Introduction Reminders

M2D7 homework should include:

- Draft the entire first big picture paragraph
- Topic sentence (first sentence) of each additional paragraph
- References in text and brief summary of each reference at the end

Impact statement/ Big picture	Motivation, why should the reader care?
Specific background	What does a scientist need to know to understand your research? What is your experimental approach?
Knowledge gap/ Statement of problem	What is unknown?
Hypothesis	What do you predict the result will be?
Here we show	What do you report in this research article?

## M2D2HW feedback: for journal club presentations

- edit the figures / data you are presenting. Take time to describe one or two plots or images rather than list many
- identify color coding on slide in text if space allows
- Verbally transition to next experiment, what did the result motivate the authors to do next?