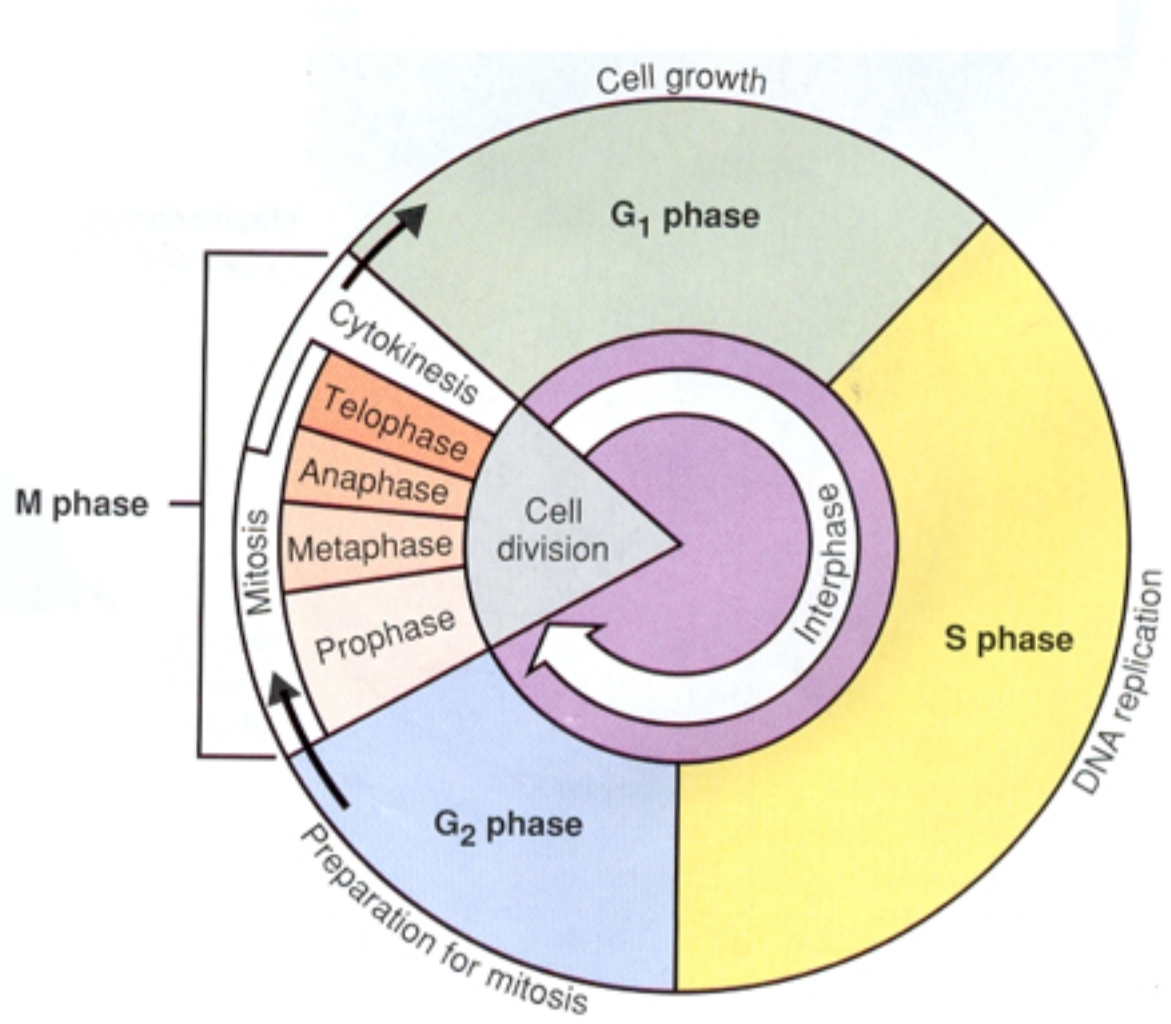
Analyze RNA-seq data and prepare for quantitative PCR experiment

M2D4

qPCR of p21 and GAPDH

- p21 (also CDKN1A)
  - Regulator of cell cycle progression at G1; arrest in G1 in response to stress
  - Tightly controlled by tumor suppressor p53

- GAPDH (glyceraldehyde 3-phosphate dehydrogenase)
  - Catalyzes glycolysis
  - housekeeping gene: constitutively and highly expressed in cells



Primer design guidelines

- Length ≈100 bp
- GC-content = 50-60%
- 3’ base preferentially a G or C
- T_m ≈ 60°C

- Must span exon-exon junction

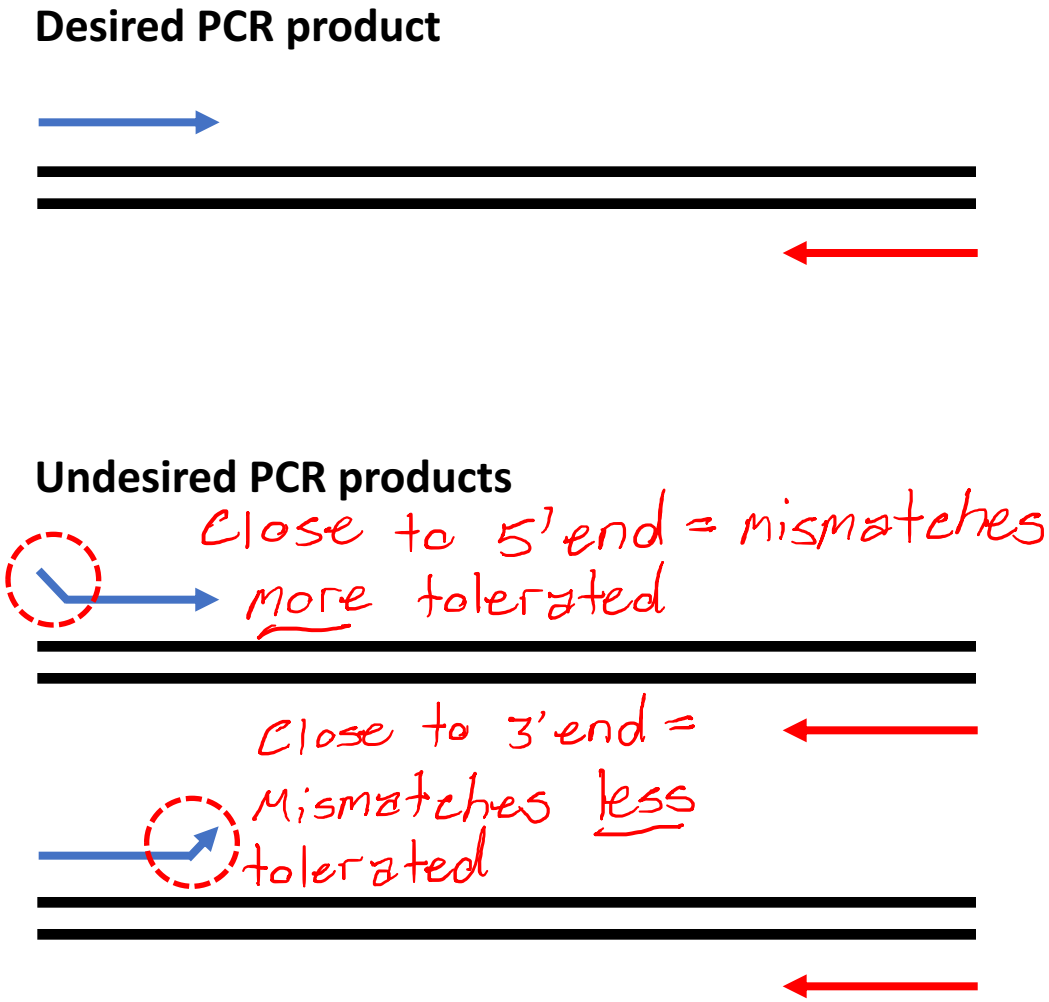
We can ignore DNA contaminants

- Does the location of mismatches matter?

- Does the length of alternative products matter?

Short products preferentially amplified.

Cycles are too short for long amplicons



Our RNA-seq data

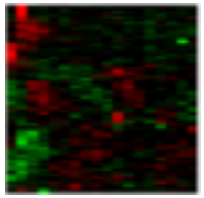
Seed cells

Induce DNA damage (etoposide)

ACAUGGAUUA

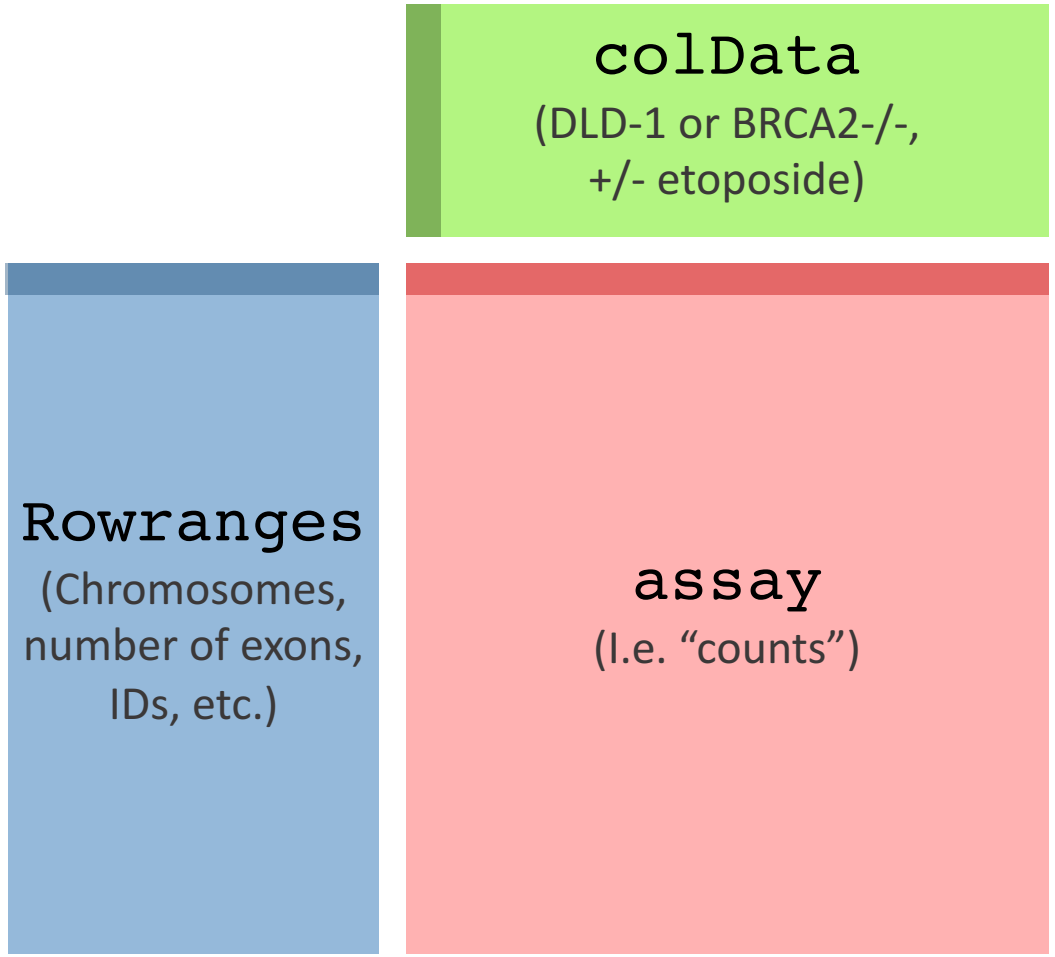
TGTACCTAAT

Purify RNA and prepare cDNA



RNA-seq

Our data structure – preprocessed by Amanda



What does raw RNA-seq data look like?

The data:

- Forward (+ Reverse) read
- Quality score
- Barcode to identify sample
- Label to map sample to flow cell

Example .fastq file

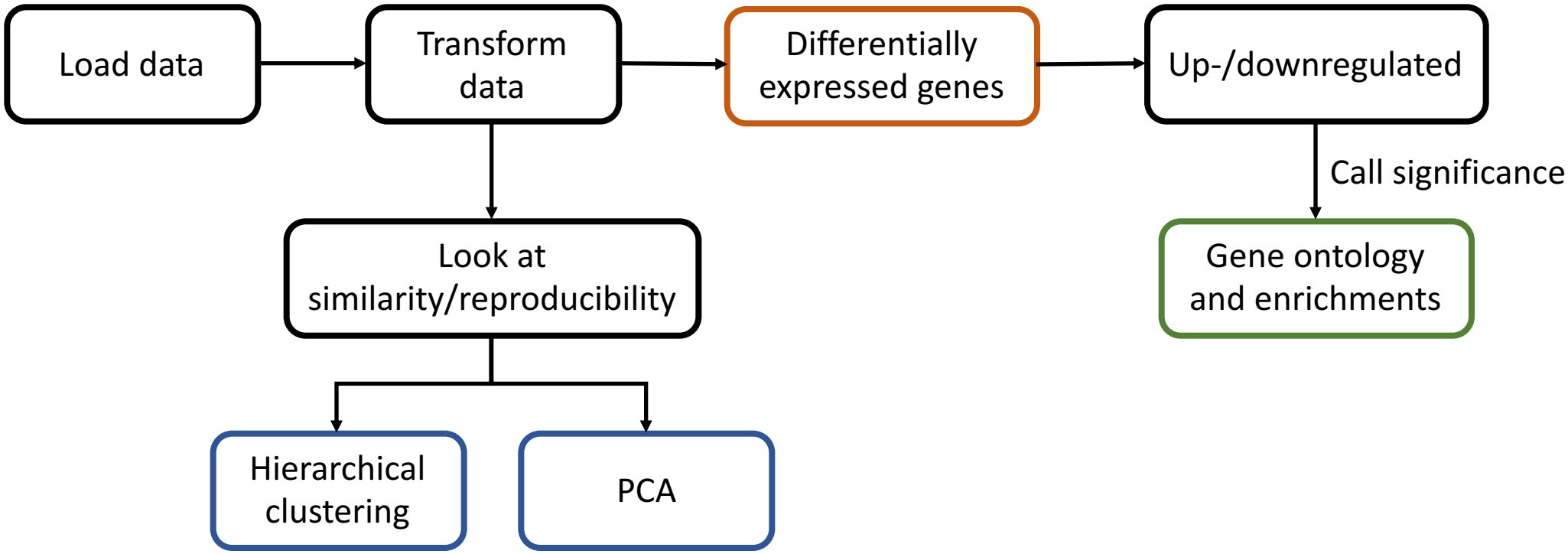
From RNA-seq data to transcriptomics

We can have from 8 million to 1.5 billion *reads!*

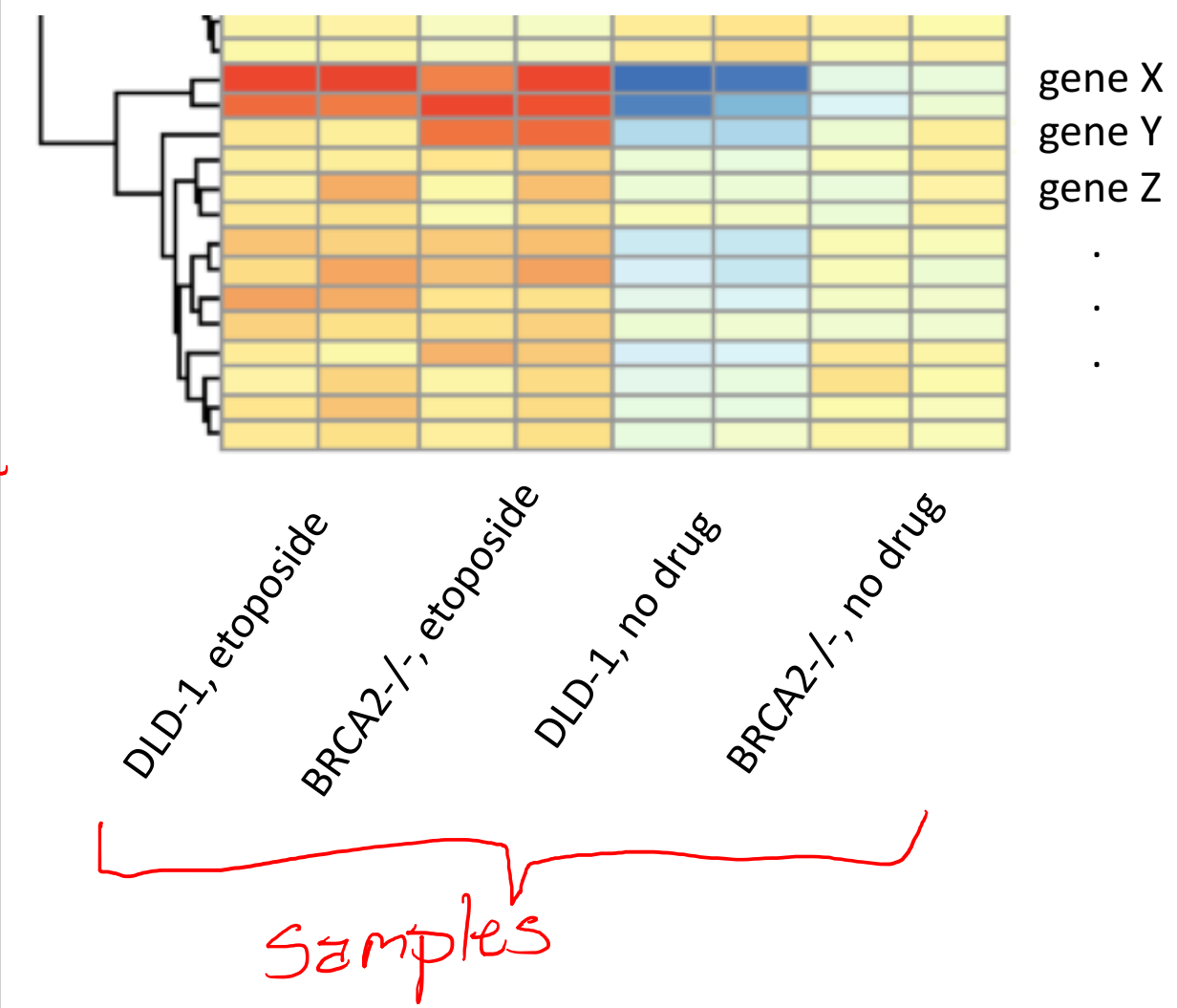
Processing the data:

- We count the number of times a gene is expressed
- We adjust for the total number of reads
- We adjust for the length of the gene

What you will do today



Are specific genes differentially expressed?

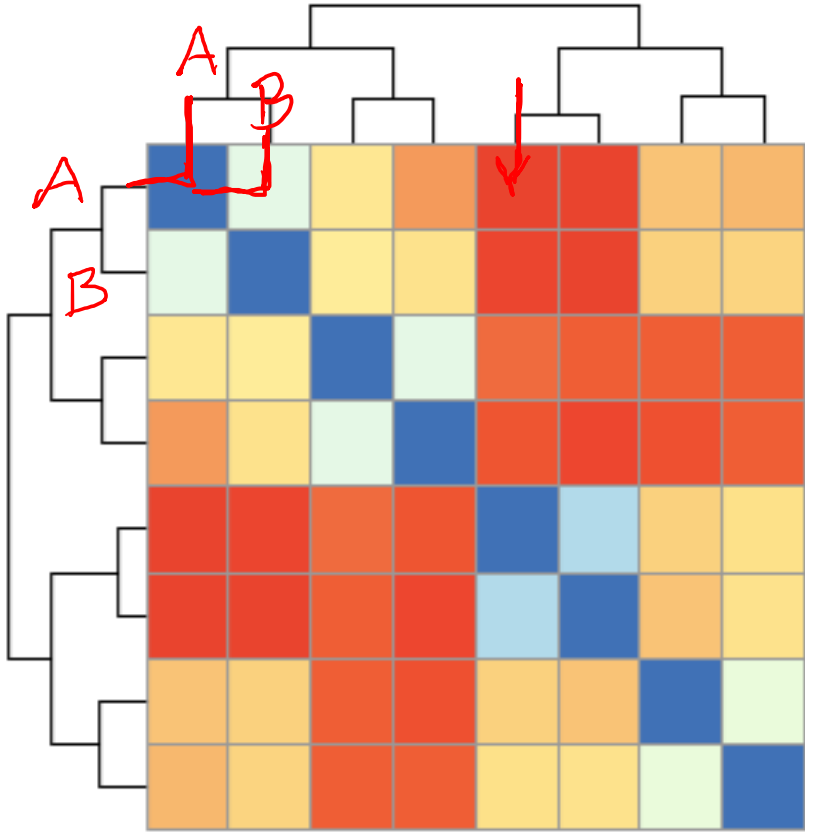


Which samples are most similar?

- Distance from one sample to another

- Symmetrical matrix

Hierarchical clustering

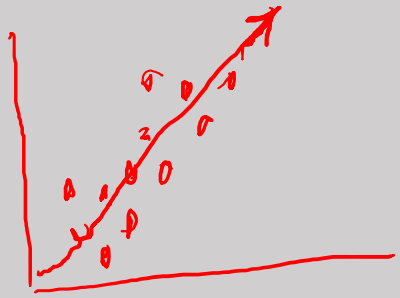


Samples:	DLD-1	BRCA2 -/-
Etoposide:	+	+
	-	-

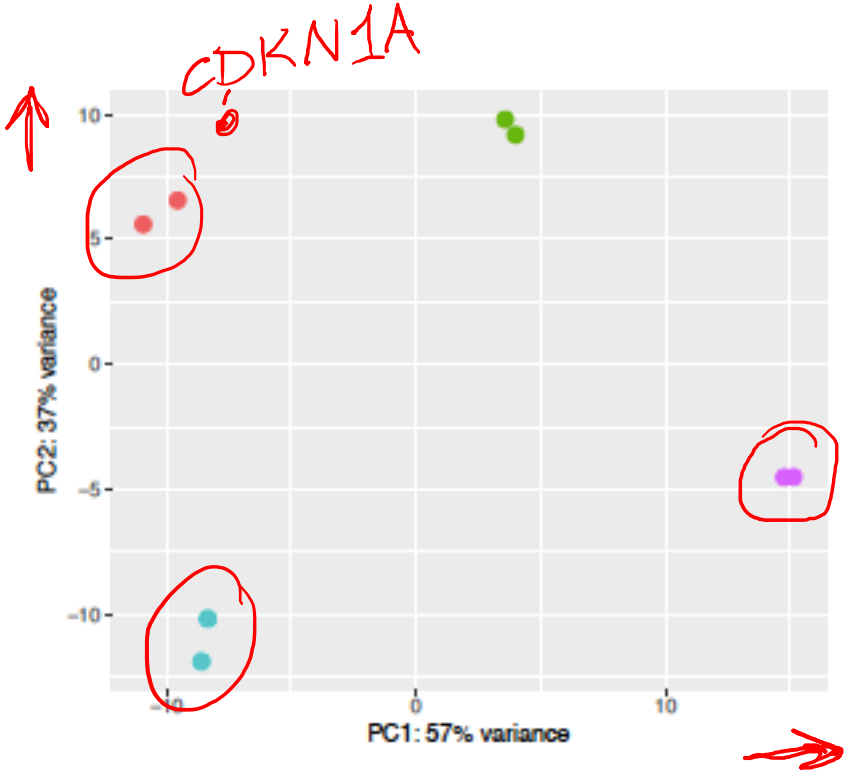
Do samples change together?

- ≈20,000 genes across 8 samples
  - Can we reduce the number of variables?

- PCA reduces dimensionality
  - Each component has the largest possible variance



Principal Component Analysis (PCA)

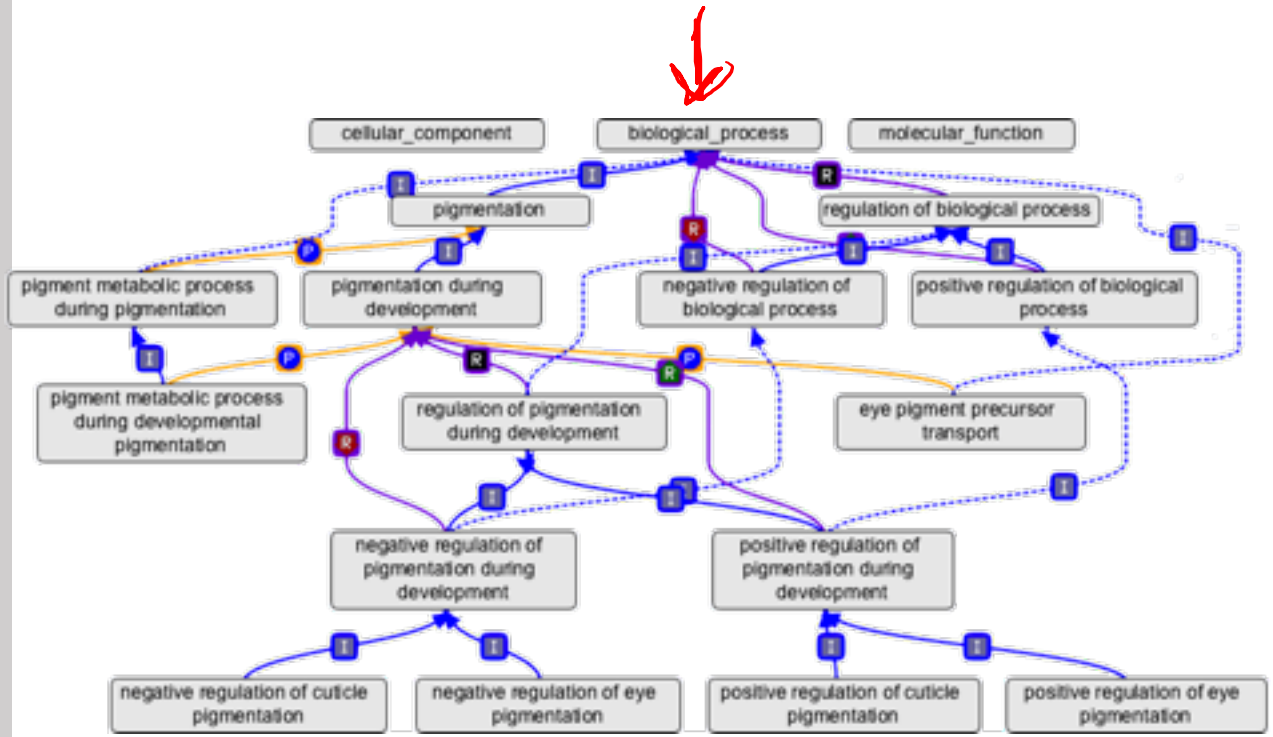


Samples:	DLD-1	BRCA2 -/-
Etoposide:	+	+
	-	-

What is the function of genes?

- Manual annotation of ≈20,000 genes is not feasible!

- GO (Gene Ontology) provides automatic annotations of biological function

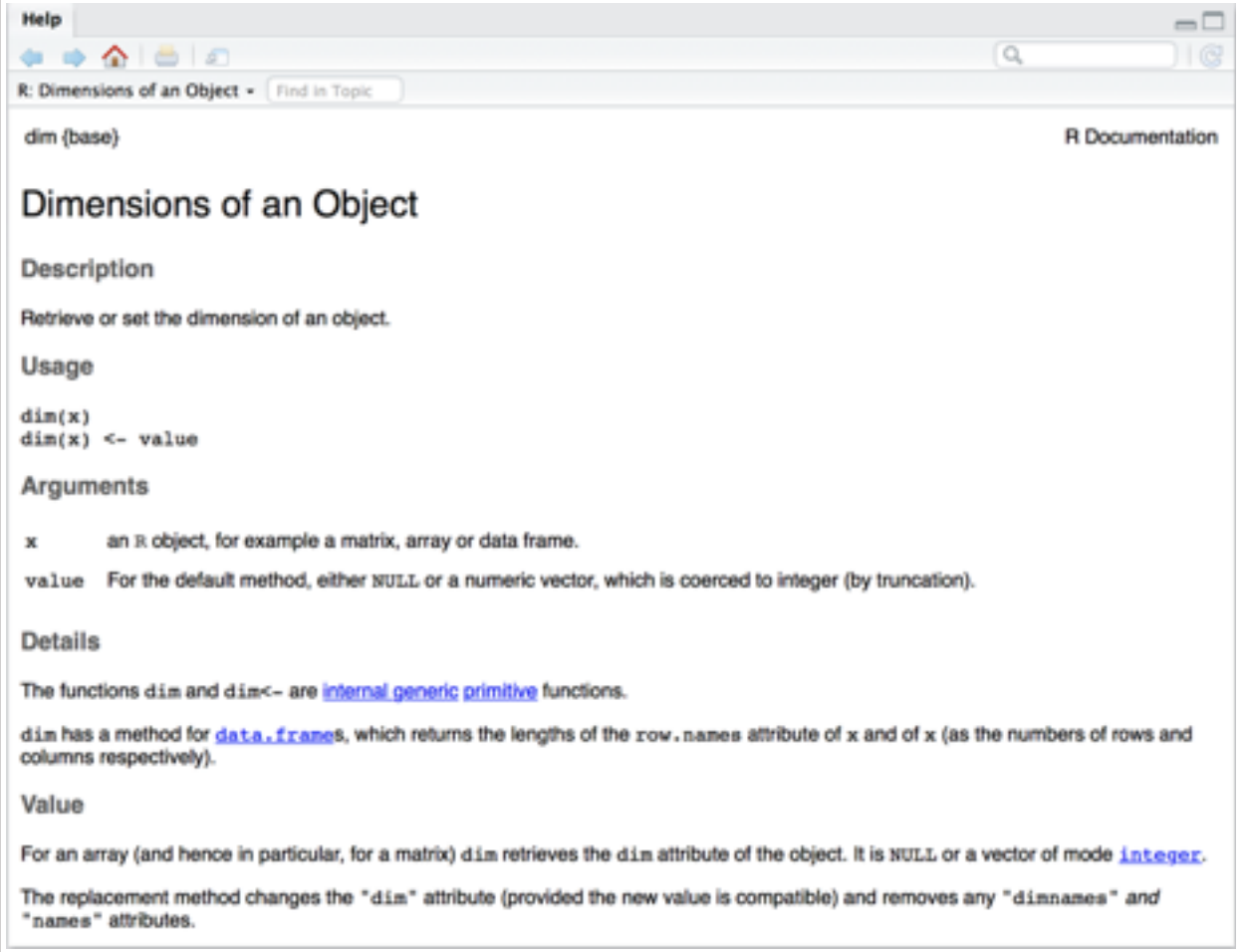


Getting help in R

Method 1
`?function`

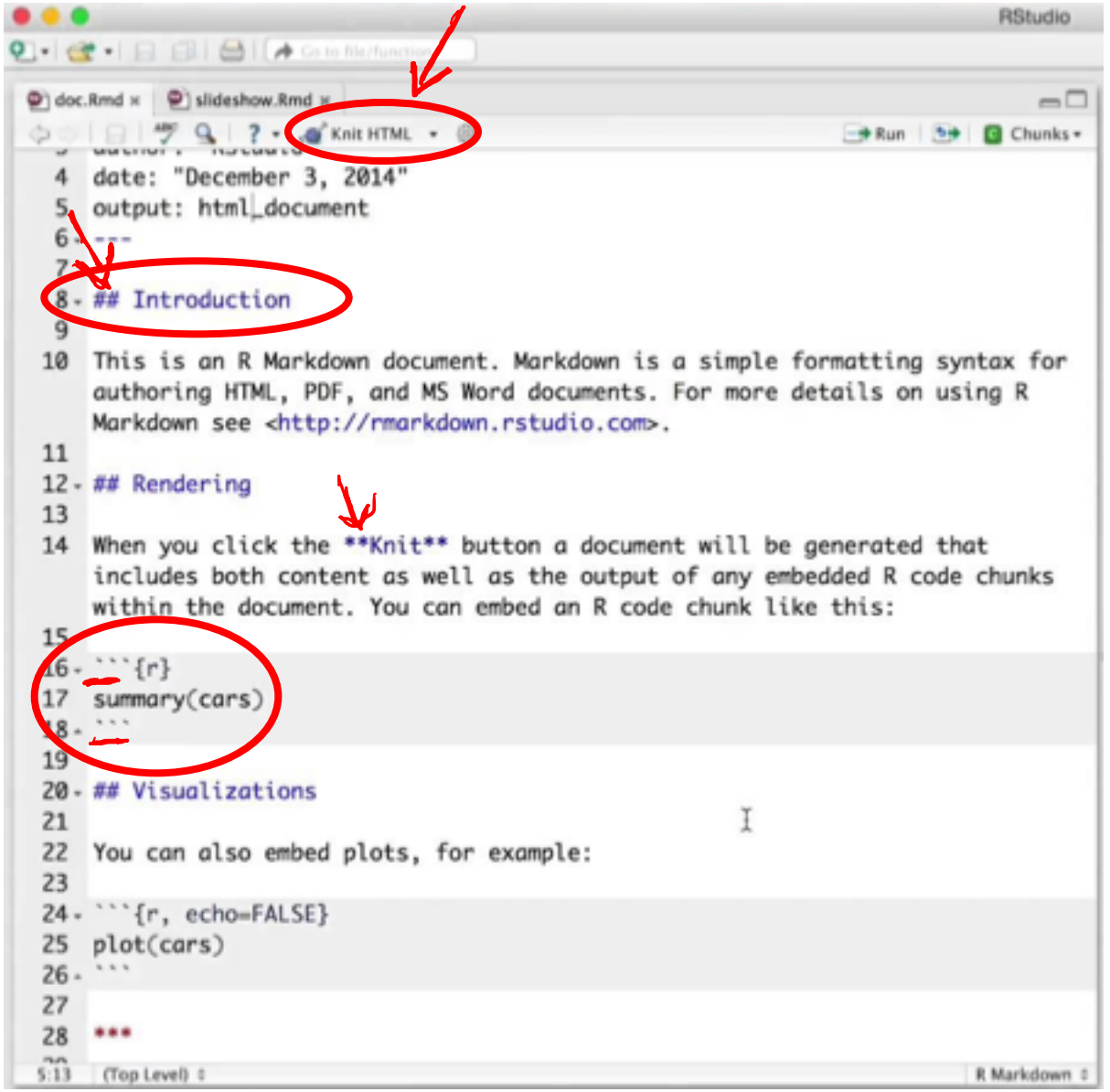
Method 2
`help(function)`

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



Writing code in R Markdown

- `File → New File → R Markdown…`
- Run selected line:
  - `Cmd + Enter`
- Run current chunk:
  - `Cmd + shift + Enter`
- Insert chunk:
  - `Alt + Cmd + I`



R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14

RStudio

1. Workflow R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

i. Open - Open a file that uses the .Rmd extension.

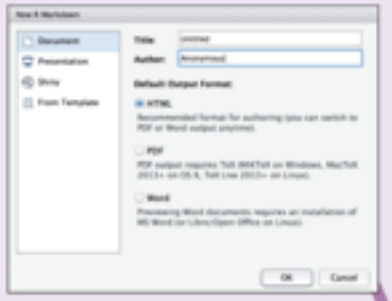
ii. Write - Write content with the easy to use R Markdown syntax

iii. Embed - Embed R code that creates output to include in the report

iv. Render - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.

2. Open File Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ▸ New File ▸ R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



3. Markdown Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax

```
Plain text
End a line with two spaces to start a new paragraph.
*italics* and _italics_
**bold** and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6

endash: --
emdash: ---
ellipsis: ...
inline equation: $A = \pi*r^{2}$
image: ![](path/to/smallorb.png)

horizontal rule (or slide break):

***

> block quote

* unordered list
* item 2
    + sub-item 1
    + sub-item 2

1. ordered list
2. item 2
    + sub-item 1
    + sub-item 2

Table Header  | Second Header
------------- | -------------
Table Cell    | Cell 2
Cell 3        | Cell 4
```

becomes

Plain text
End a line with two spaces to start a new paragraph.
italics and *italics*
bold and **bold**
superscript^2
~~strikethrough~~
link

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

endash: –
emdash: —
ellipsis: …
inline equation: $A = \pi * r^2$
image:

horizontal rule (or slide break):

> block quote

- unordered list
- item 2
  - sub-item 1
  - sub-item 2

1. ordered list
2. item 2
   - sub-item 1
   - sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

4. Choose Output Write a YAML header that explains what type of document to build from your R Markdown file.

YAML

A YAML header is a set of key: value pairs at the start of your file. Begin and end the header with a line of three dashes (---)

```
---
title: "Untitled"
author: "Anonymous"
output: html_document
---

This is the start of my
report. The above is metadata
saved in a YAML header.
```

The RStudio template writes the YAML header for you

The output value determines which type of file R will build from your .Rmd file (in Step 6)

output: html_document	html file (web page)
output: pdf_document	pdf document
output: word_document	Microsoft Word .docx
output: beamer_presentation	beamer slideshow (pdf)
output: ioslides_presentation	ioslides slideshow (html)

Homework Due M2D5

- Peer review of methods

  -before you leave today I will give you a copy of a classmate's methods homework to review

  -review for clarity, completeness, and organization (see prompts on the wiki for full description)

  -you must submit typed comments as a separate document using the "numbering method"

  -turn your comments into the instructors M2D5, not to the peer you reviewed