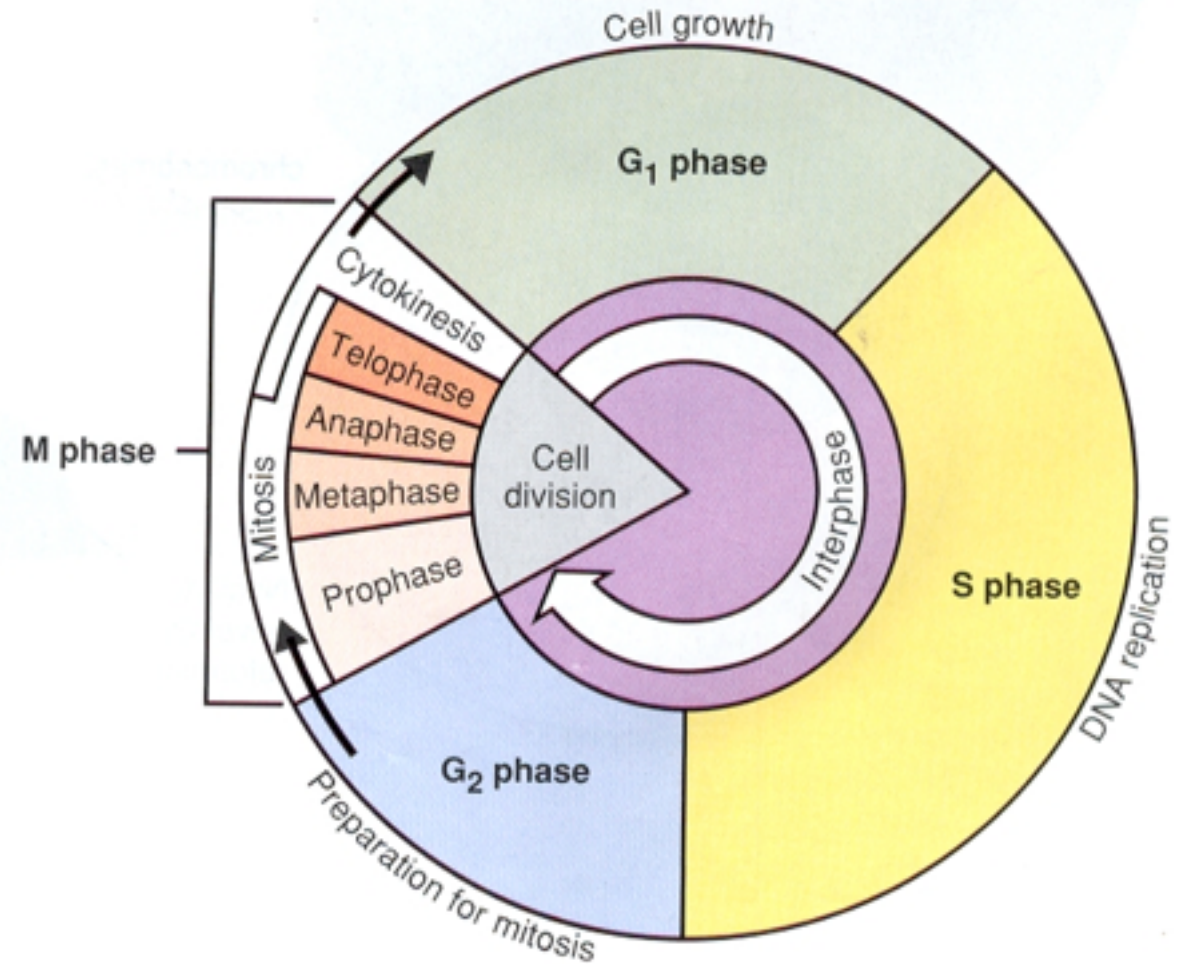


Analyze RNA-seq data and prepare
for quantitative PCR experiment

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 \approx 100 bp
- GC-content = 50-60%
- 3' base preferentially a G or C
- $T_m \approx 60^\circ\text{C}$

- Must span exon-exon junction

We can ignore DNA contaminants

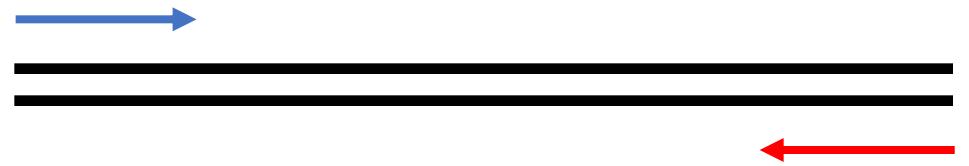
- Does the location of mismatches matter?

- Does the length of alternative products matter?

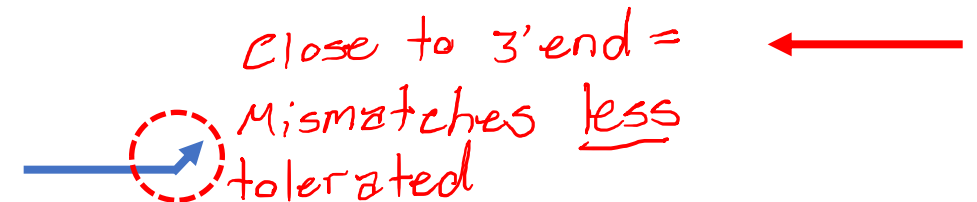
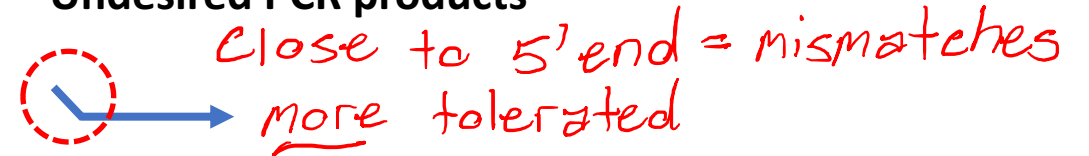
*Short products preferentially amplified.
Cycles are too short for long amplicons*



Desired PCR product



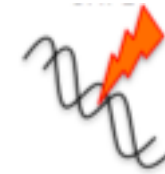
Undesired PCR products



Our RNA-seq data



Seed cells



**Induce DNA damage
(etoposide)**

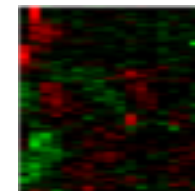


ACAUGGAUUA



TGTACCTAAT

**Purify RNA and
prepare cDNA**



RNA-seq

Our data structure –
preprocessed by Amanda

Rowranges
(Chromosomes,
number of exons,
IDs, etc.)

colData
(DLD-1 or BRCA2-/-,
+/- etoposide)

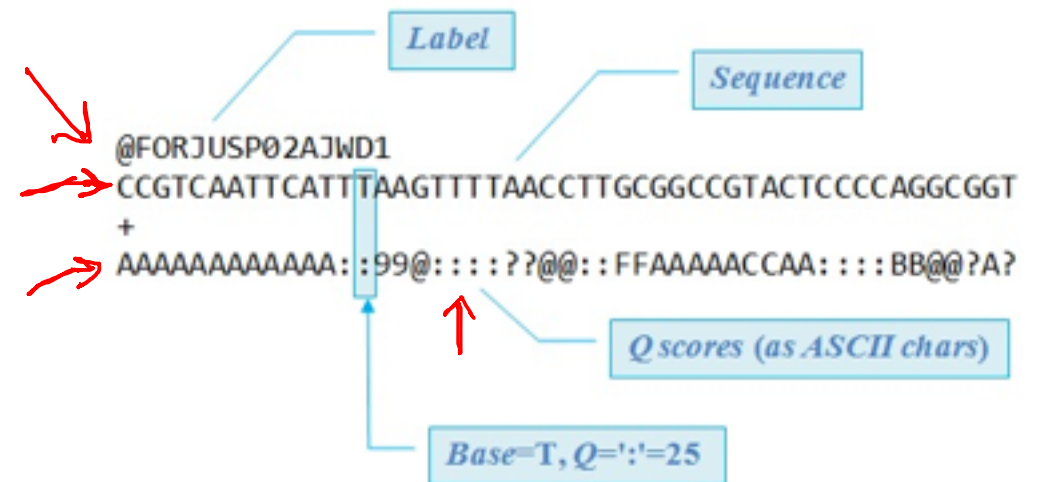
assay
(i.e. “counts”)

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

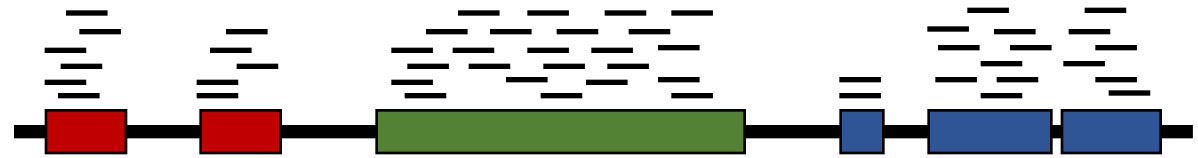


From RNA-seq data to transcriptomics

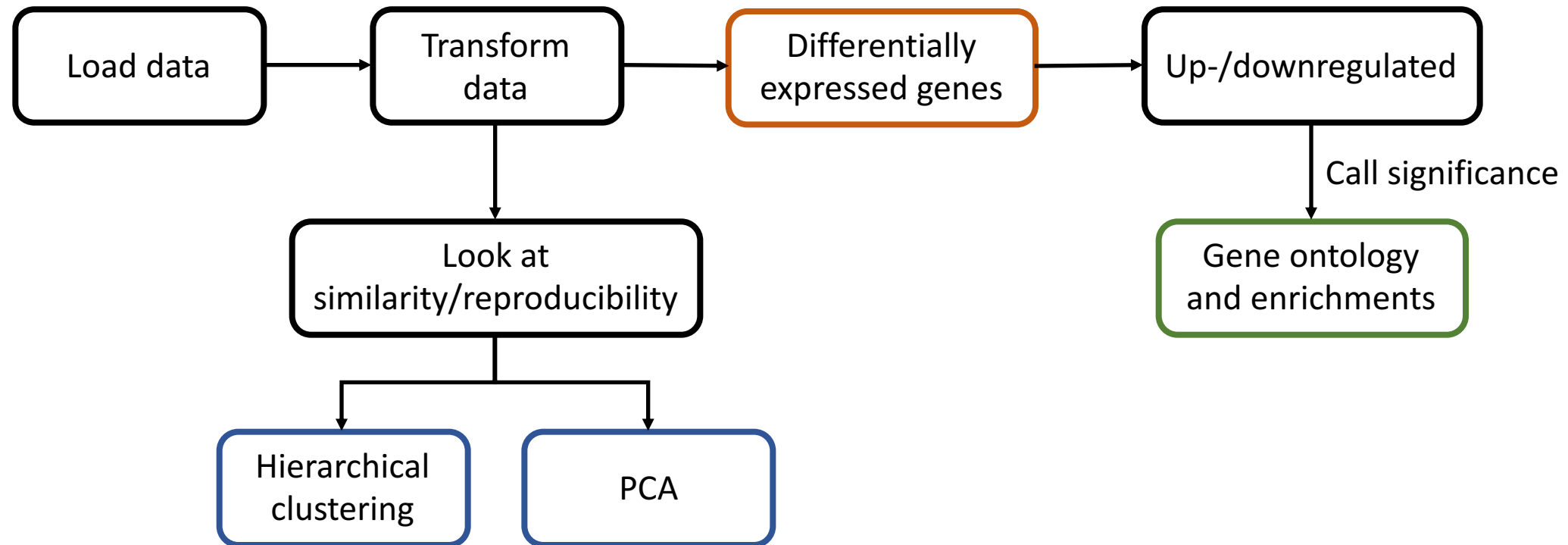
We can have from 8 million to 1.5 billion reads!

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

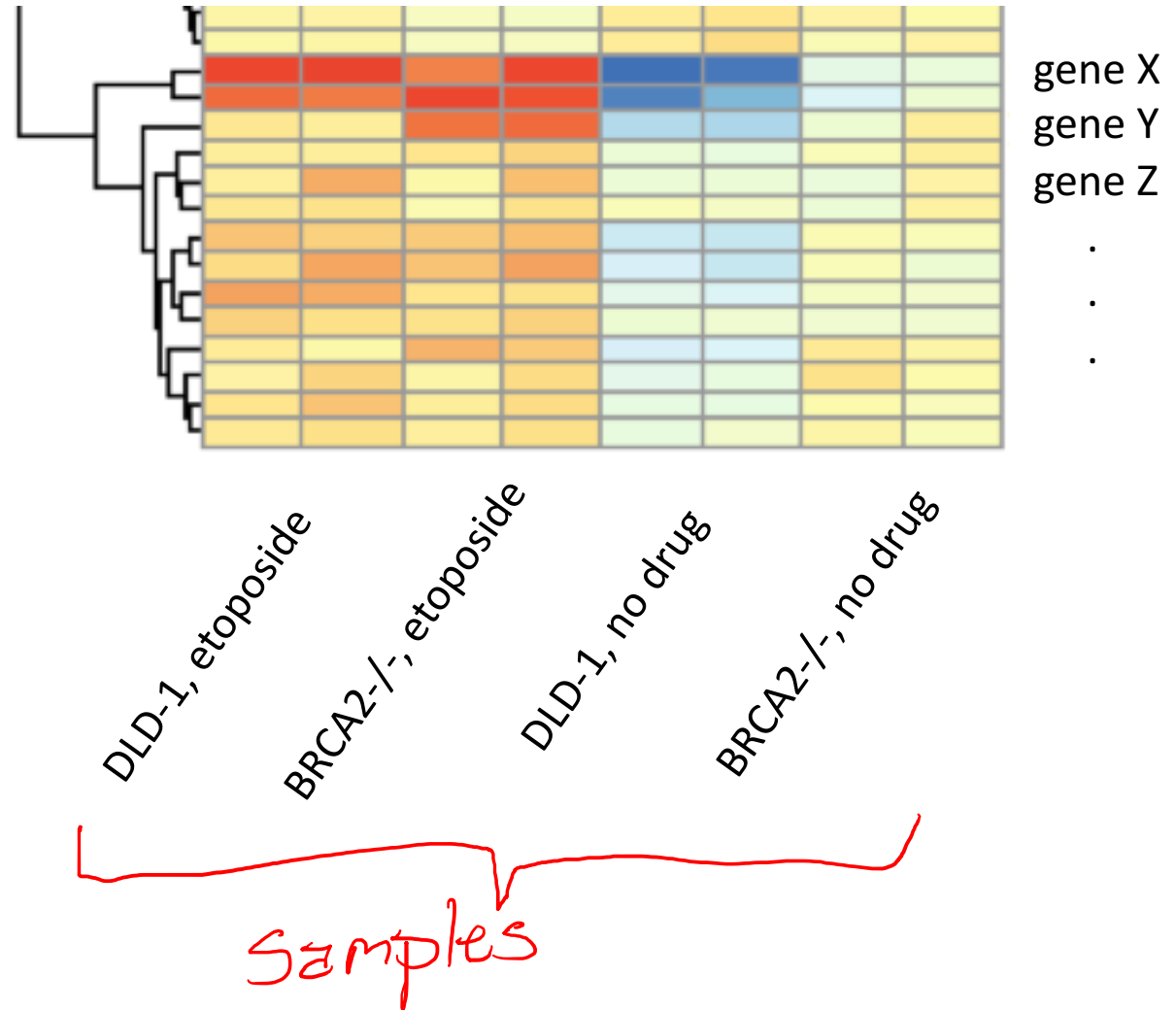


What you will do today



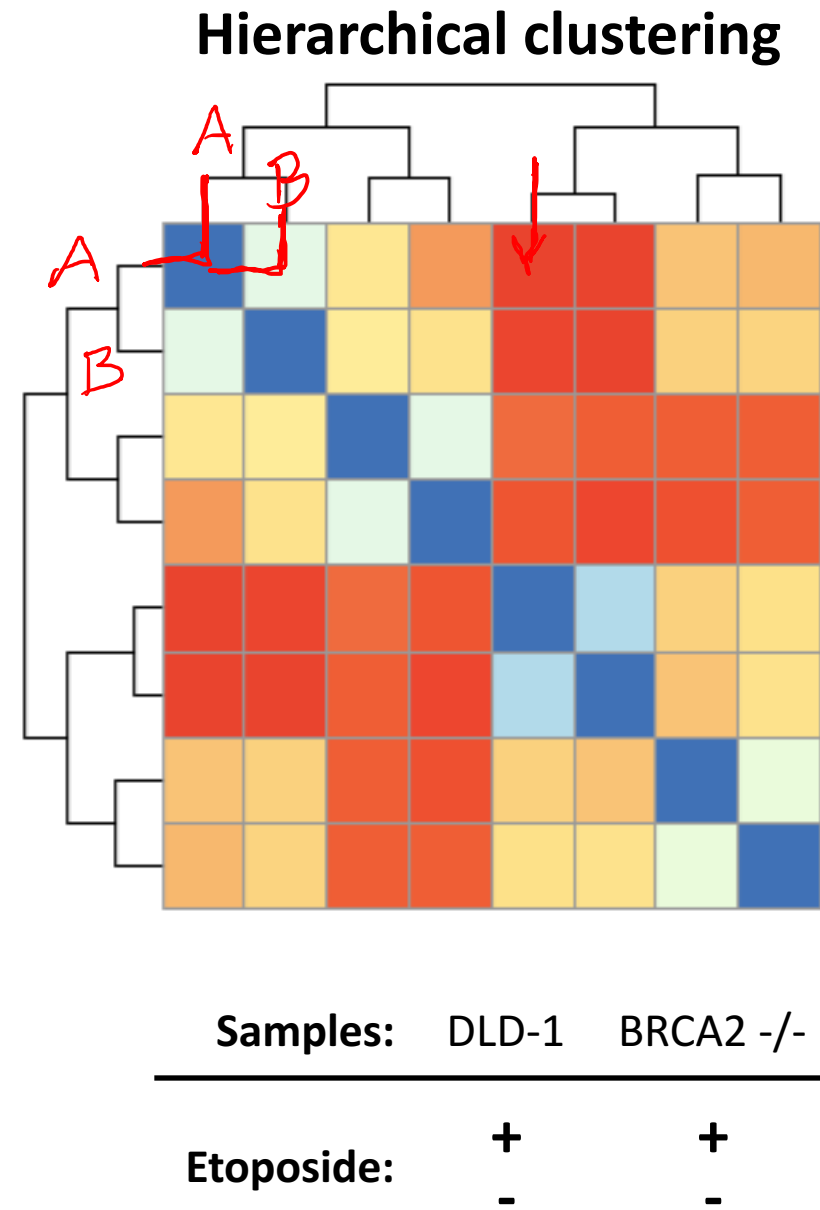
Are specific genes differentially expressed?

genes



Which samples are most similar?

- Distance from one sample to another
- Symmetrical matrix

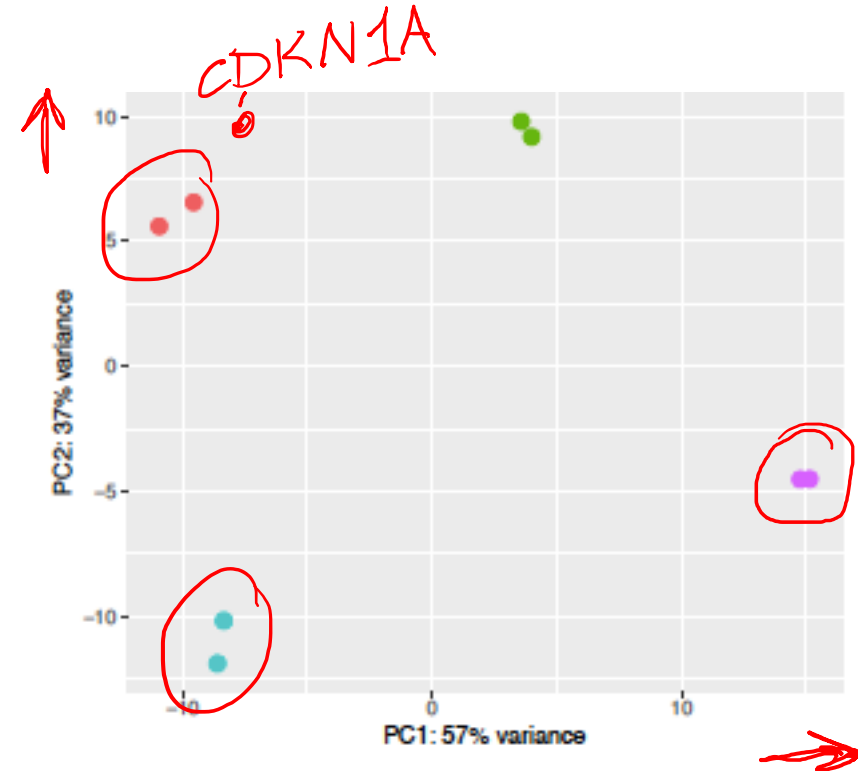


Do samples change together?

- $\approx 20,000$ genes across 8 samples
 - Can we reduce the number of variables?
- PCA reduces dimensionality
 - Each component has the largest possible variance



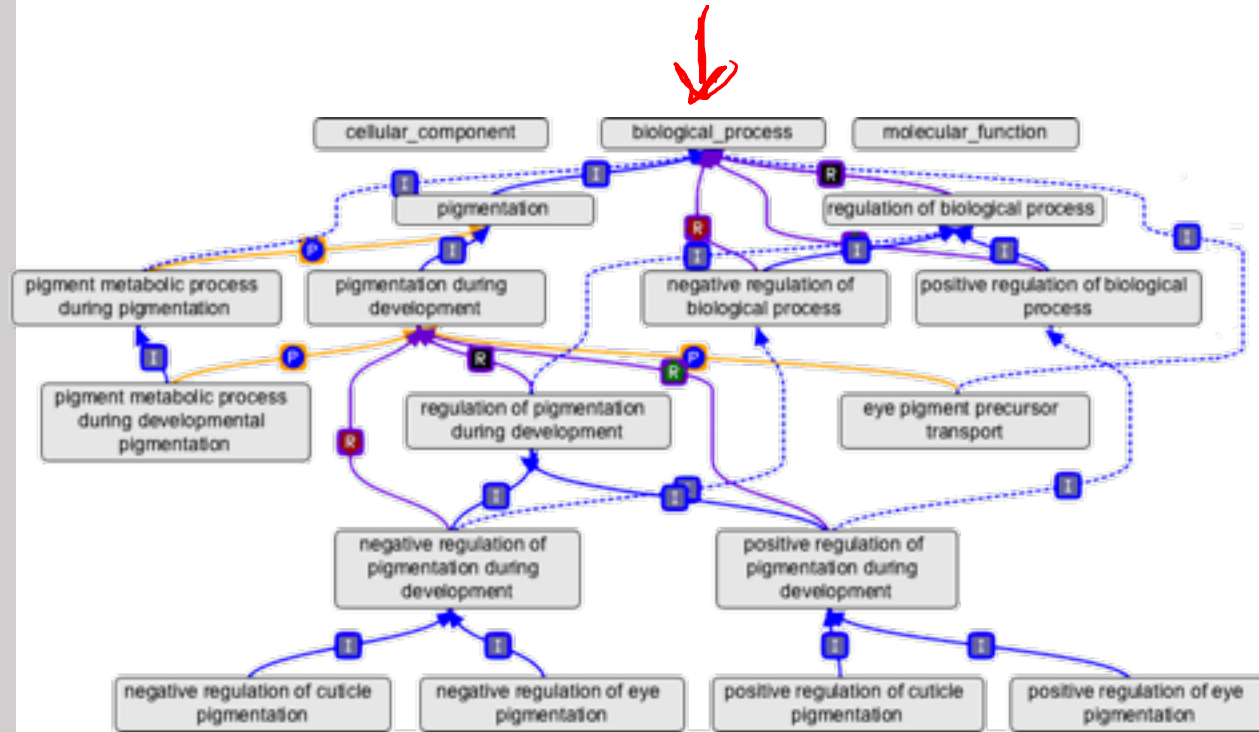
Principal Component Analysis (PCA)



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

What is the function of genes?

- Manual annotation of $\approx 20,000$ genes is not feasible!
- GO (Gene Ontology) provides automatic annotations of biological function



Getting help in R

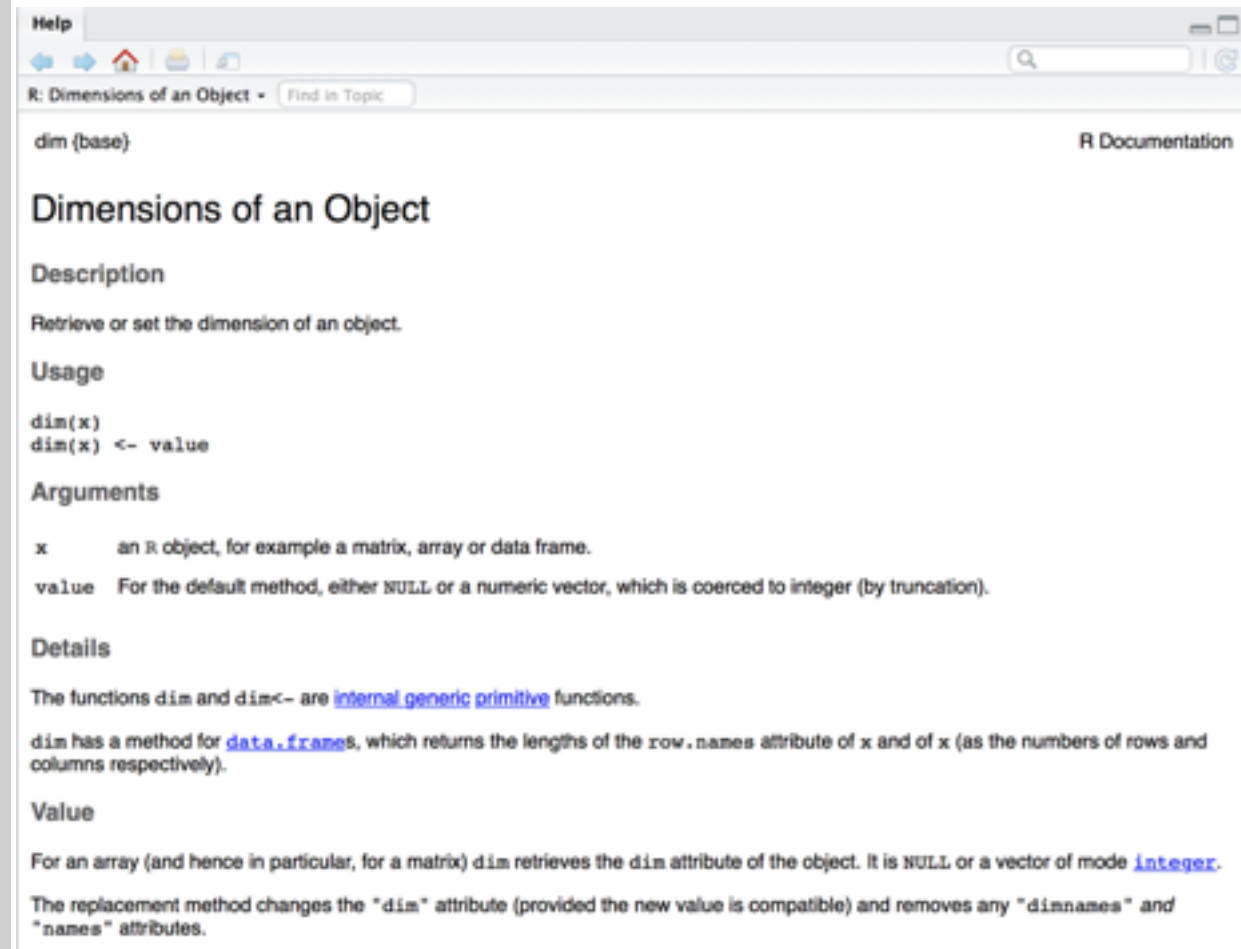
Method 1

`?function`

Method 2

`help(function)`

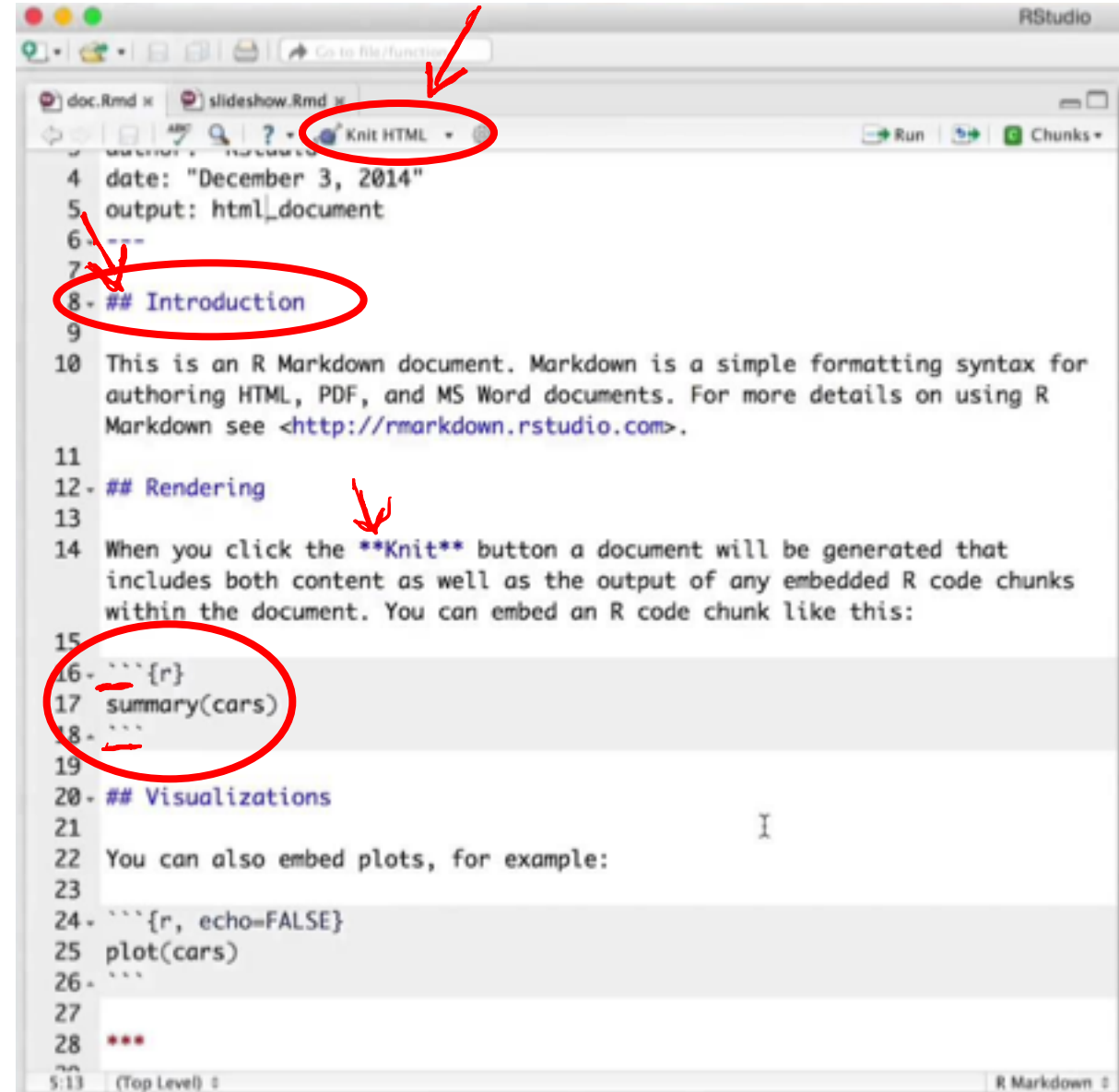
Example: Type in `?dim` or `help(dim)`



The screenshot shows the R Help window for the `dim` function. The title bar reads "Help" and the window content is titled "R: Dimensions of an Object". The main heading is "Dimensions of an Object". Below this, there is a "Description" section stating "Retrieve or set the dimension of an object." and a "Usage" section showing `dim(x)` and `dim(x) <- value`. The "Arguments" section lists `x` as "an R object, for example a matrix, array or data frame." and `value` as "For the default method, either NULL or a numeric vector, which is coerced to integer (by truncation)." The "Details" section notes that `dim` and `dim<=` are internal generic primitive functions and that `dim` has a method for `data.frames`. The "Value" section explains that `dim` retrieves the `dim` attribute of the object, which is NULL or a vector of mode `integer`, and that the replacement method changes the `dim` attribute and removes `dimnames` and `names` attributes.

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



The screenshot shows the RStudio interface with an R Markdown document open. The document content is as follows:

```
4 date: "December 3, 2014"
5 output: html_document
6 ---
7
8 ## Introduction
9
10 This is an R Markdown document. Markdown is a simple formatting syntax for
    authoring HTML, PDF, and MS Word documents. For more details on using R
    Markdown see <http://rmarkdown.rstudio.com>.
11
12 ## Rendering
13
14 When you click the Knit button a document will be generated that
    includes both content as well as the output of any embedded R code chunks
    within the document. You can embed an R code chunk like this:
15
16 ```{r}
17 summary(cars)
18 ```
19
20 ## Visualizations
21
22 You can also embed plots, for example:
23
24 ```{r, echo=FALSE}
25 plot(cars)
26 ```
27
28 ***
```

Annotations in the image include:

- A red circle around the **Knit HTML** button in the top toolbar.
- A red circle around the `## Introduction` header.
- A red circle around the R code chunk `summary(cars)`.
- Red arrows pointing to the `## Introduction` header and the `summary(cars)` code chunk.

R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14



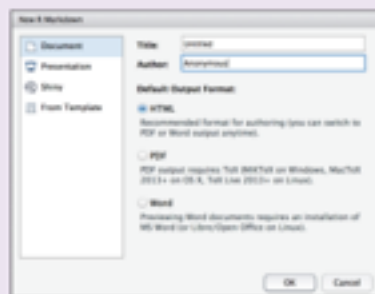
1. Workflow R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

- 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__**
superscript²
~~strikethrough~~
[\[link\]\(www.rstudio.com\)](http://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:

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**
superscript²
~~strikethrough~~
[link](http://www.rstudio.com)

Header 1
Header 2
Header 3
Header 4
Header 5
Header 6

endash: --
emdash: ---
ellipsis: ...
inline equation: $A = \pi r^2$



horizontal rule (or slide break):

| block quote

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

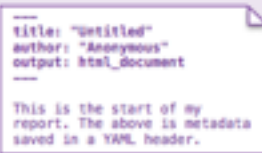
1. ordered list
2. item 2
+ sub-item 1
+ sub-item 2

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 (---)



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