

Investigate RNA-seq data using public databases

M2D5

Your samples for qPCR

Remember:

- Parental cell line, DLD-1
 - +/- etoposide
- Mutant cell line, BRCA -/-
 - +/- etoposide



Seed cells



Induce DNA damage
(etoposide)

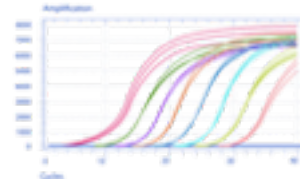


ACAUGGAUUA



TGTACCTAAT

Purify RNA and
prepare cDNA



Measure transcript levels

Your samples for qPCR

Remember:

- Parental cell line, DLD-1
 - +/- etoposide
- Mutant cell line, BRCA -/-
 - +/- etoposide



Seed cells



Induce DNA damage
(etoposide)

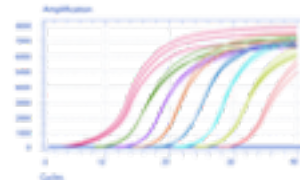


ACAUGGAUUA



TGTACCTAAT

Purify RNA and
prepare cDNA

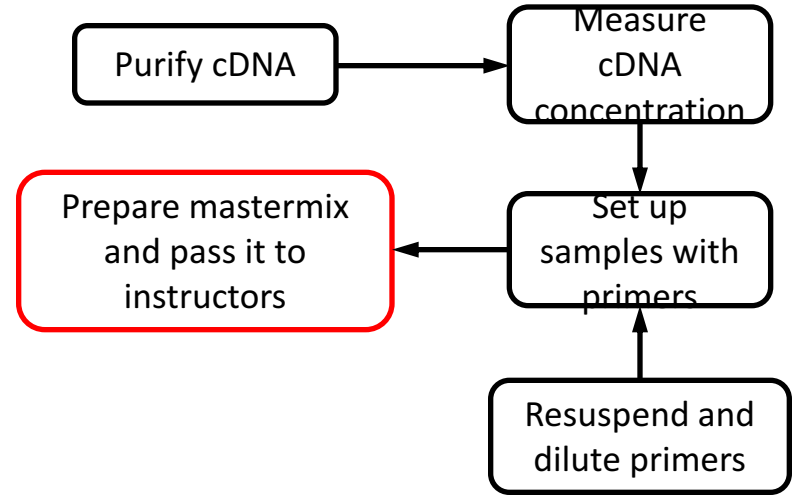


Measure transcript levels

Perform qPCR

Remember:

- p21:
 - Critical gene involved in replication
 - We use it to see if cells replicate normally*
- GAPDH:
 - Reference gene. Our control



| | Primers | | | | | |
|-----------|---------|---|--------------|---|-------------|---|
| | GAPDH | | p21 (instr.) | | p21 (yours) | |
| Etoposide | - | + | - | + | - | + |
| DLD-1 | A | B | E | F | I | J |
| BRCA2 -/- | C | D | G | H | K | L |

Clean up cDNA: remove all enzyme, buffers + dNTPs

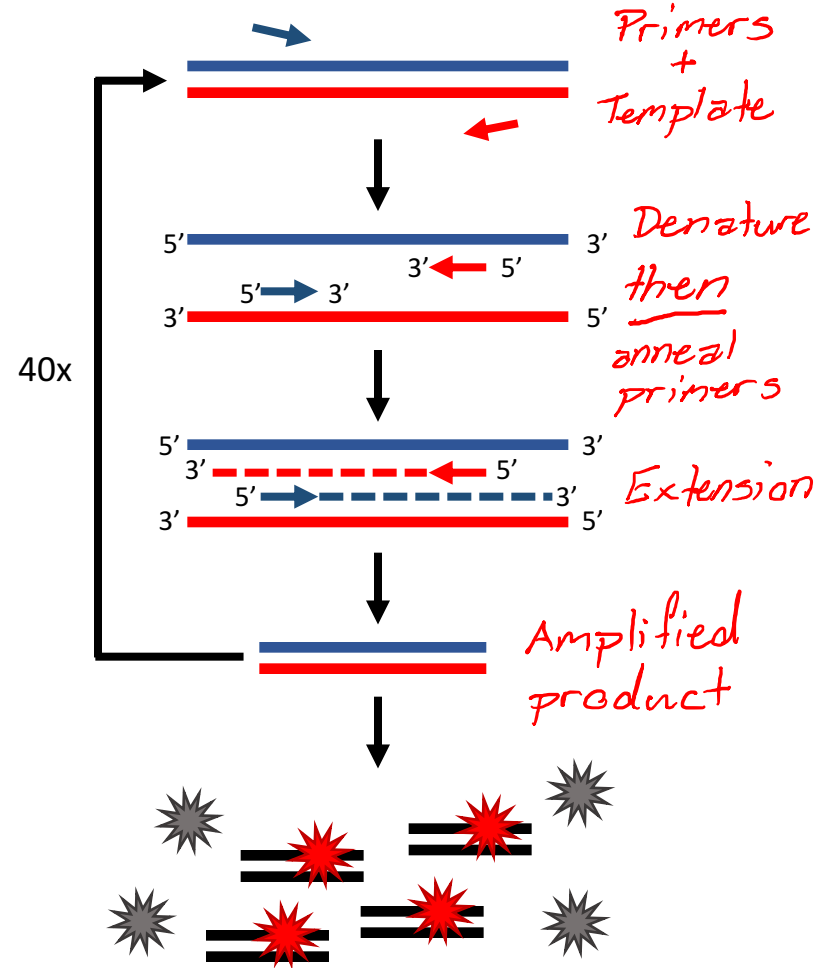
- Qiagen QIAquick PCR purification kit



| steps | contents | purpose |
|---------|---------------------------|--|
| prepare | 5X Buffer PB to cDNA | Contains guanidinium thiocyanate. Ensures DNA can bind to |
| bind | silica membrane in column | silica membrane |
| wash | Buffer PE | Mostly ethanol. Wash away salts/buffers/impurities ** then, get rid of all ethanol |
| elute | water | Compete for binding to membrane. Elutes DNA |

Perform qPCR

- Like a regular PCR:
 - Template DNA (cDNA)
 - Fw and Rv primer
 - dNTPs
 - DNA Polymerase
 - Buffer
- ... *except*:
 - Contains dye that binds dsDNA



Perform qPCR

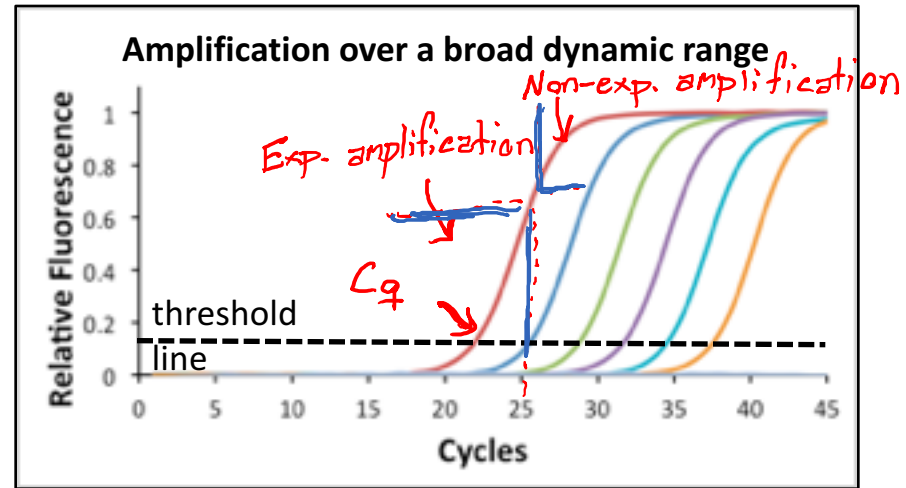
- Dye binds dsDNA
 - Is quenched if unbound
- Fluorescence is a function of [DNA]
 - Remember, DNA is exponentially amplified
- We use controls with a known DNA concentration to calculate [DNA] in our sample
- C_q value = cycle # where the fluorescent signal is significantly above background

↙ Best practices

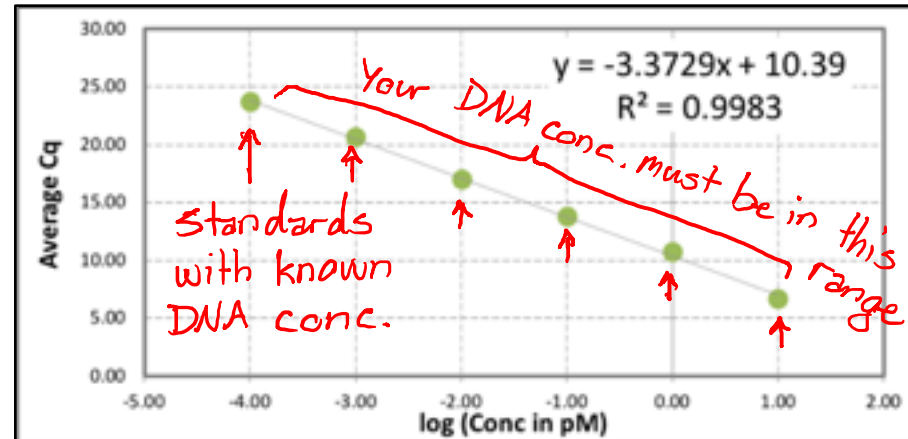
On publishing qPCR data:

<http://clinchem.aaccjnls.org/content/55/4/6>

11



Using controls to calculate [DNA] in sample



Using public databases

The Cancer Genome Atlas – TCGA

- Sequencing data
- Healthy – diseased comparison
- Clinical data

Ensembl

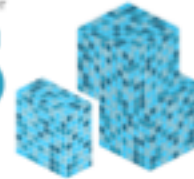
- Genome annotations
- Comparative genomics
- Variation
- Regulatory data

NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data



To put this into perspective, 1 petabyte of data is equal to

212,000
DVDs



TCGA data describes

33
DIFFERENT
TUMOR TYPES



...including

10
RARE
CANCERS

...based on paired tumor and normal tissue sets collected from

11,000
PATIENTS



...using

7
DIFFERENT
DATA TYPES



Exploring the TCGA database

New Feature: Export biospecimen and clinical data in TSV format now available in the GDC Data Portal Dismiss

NIH NATIONAL CANCER INSTITUTE GDC Data Portal Home Projects **Exploration** Analysis Repository Quick Search Manage Sets Login Cart GDC Apps

> Primary Site

> Program

> Project

- TCGA-BLCA
- TCGA-BRCA
- TCGA-AML
- TCGA-WT
- TCGA-GBM
- TCGA-OV
- TCGA-LIAD
- TCGA-UCEC
- TCGA-KIRC
- TCGA-HNSC
- TCGA-LGG
- TCGA-THCA
- TCGA-LUSC
- TCGA-PRAD
- TCGA-SKCM
- TCGA-COAD**

Cases (461) **Genes (19,946)** Mutations (279,629) OncoGrid

Genes

Distribution of Most Frequently Mutated Genes

Overall Survival Plot
442 Cases with Survival Data

Showing 1 - 10 of 19,946 genes

| Symbol Name | Cytoband | Type | # Affected Cases in Cohort | # Affected Cases Across the GDC | # Mutations | Annotations | Survival |
|-------------|--|---------|----------------------------|---------------------------------|----------------|-------------|----------|
| APC | adenomatous polyposis coli | 5q22.2 | protein_coding | 309 / 400 (77.25%) | 893 / 10,202 | 316 | |
| TTN | titin | 2q31.2 | protein_coding | 258 / 400 (64.50%) | 3,858 / 10,202 | 1,031 | |
| TP53 | tumor protein p53 | 17p13.1 | protein_coding | 233 / 400 (58.25%) | 6,008 / 10,202 | 162 | |
| KRAS | Kirsten rat sarcoma viral oncogene homolog | 12p12.1 | protein_coding | 124 / 400 (31.00%) | 839 / 10,202 | 33 | |
| MUC16 | mucin 16, cell surface associated | 19p13.2 | protein_coding | 152 / 400 (38.00%) | 2,486 / 10,202 | 383 | |

KEGG to look at pathways

- Comparison between our sample and sample from TCGA

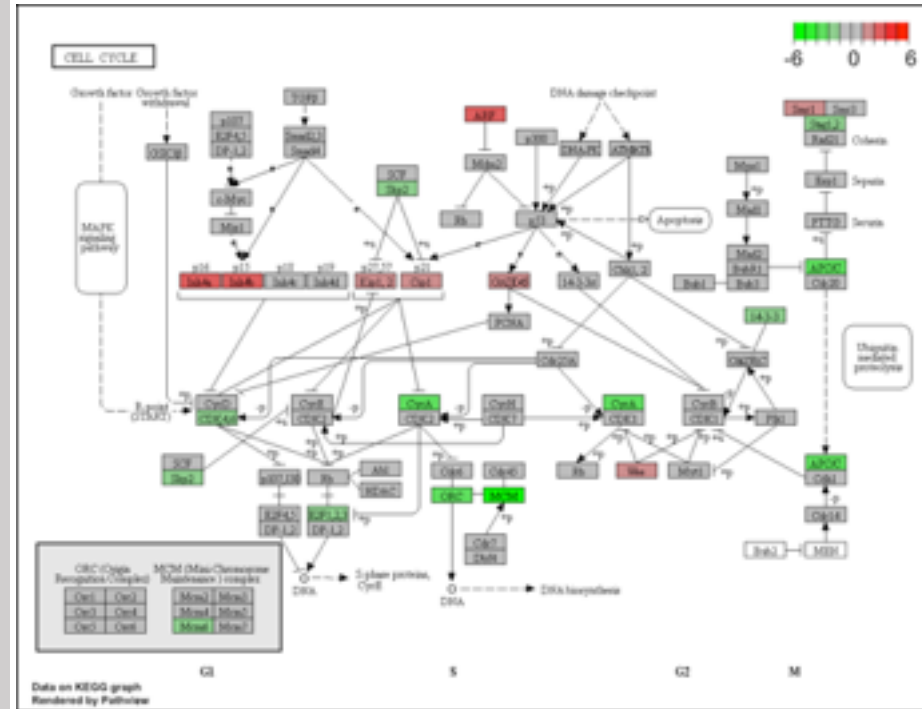
- Are certain pathways different?

- What is a tissue sample?

From patient. Can be healthy or diseased.

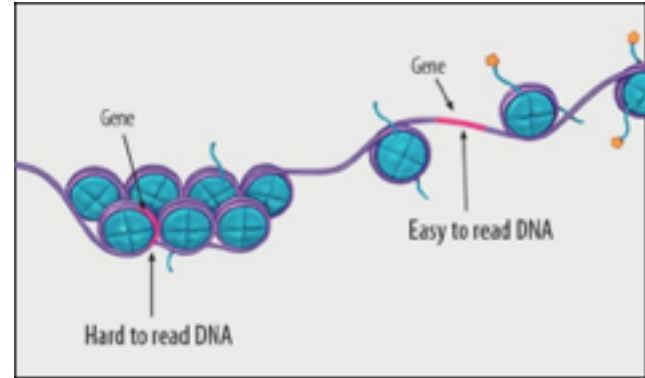
- What is a cell line?

A cell, i.e. a fibroblast (skin cell), that has been immortalized such that it can be passaged indefinitely

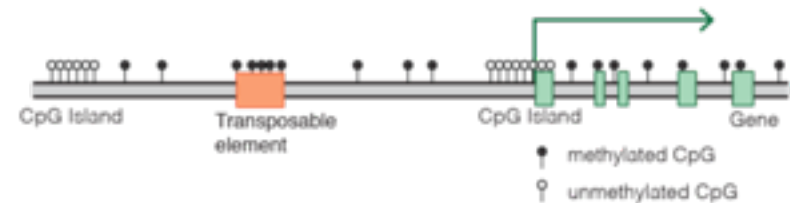


There is more to DNA than just the sequence

- Epigenetic data:
 - **Histones** and post-translational modifications of their tails
 - **DNA methylation**
 - Unmethylated =
 - Methylated =



Typical mammalian DNA methylation landscape



SPRING BREAK 2018!



your  cards
someecards.com



- M1 Data Summary revision due Monday, March 26th 10pm
- Journal Club Presentations, April 3rd and 5th