

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)

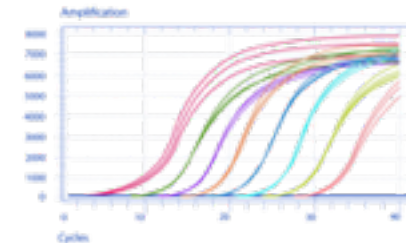


ACAUGGAUUA



TGTACCTAAT

Purify RNA and
prepare cDNA



Measure transcript levels

Your samples for qPCR

Remember:

- Parental cell line, DLD-1
 - +/- etoposide
- Mutant cell line, BRCA²-/-
 - +/- etoposide



Seed cells



Induce DNA damage
(etoposide)

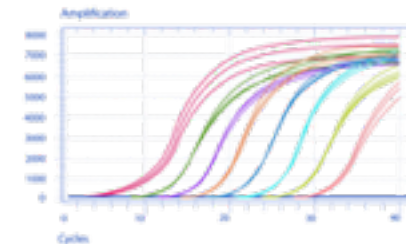


ACAUGGAUUA



TGTACCTAAT

Purify RNA and
prepare cDNA



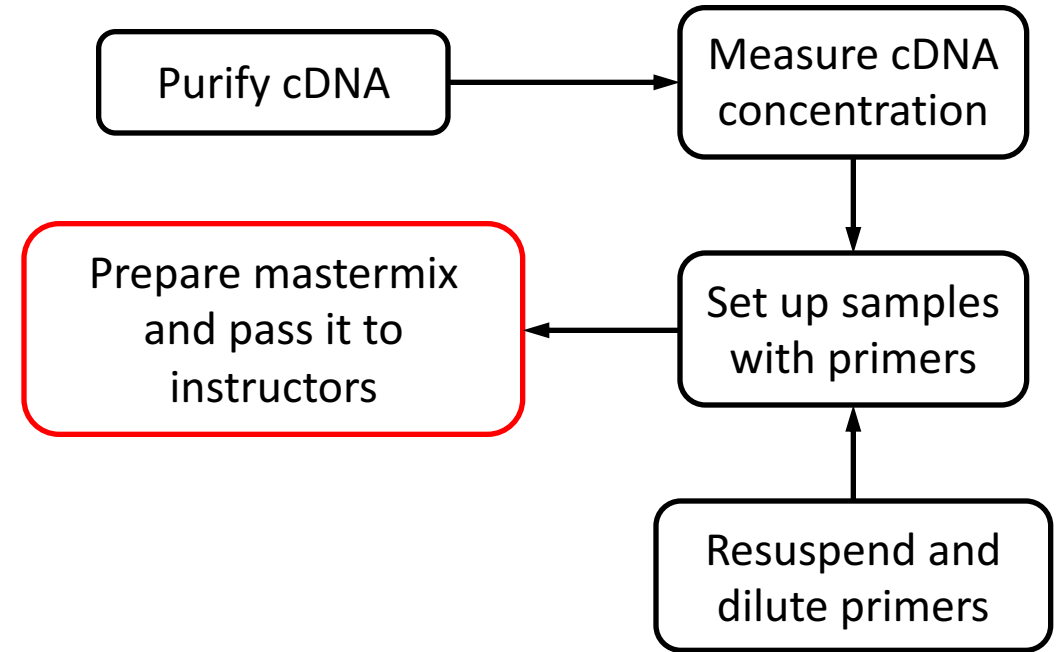
Measure transcript levels

Perform qPCR

Remember:

CDKN1A

- p21:
 - Critical gene involved in replication
- 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

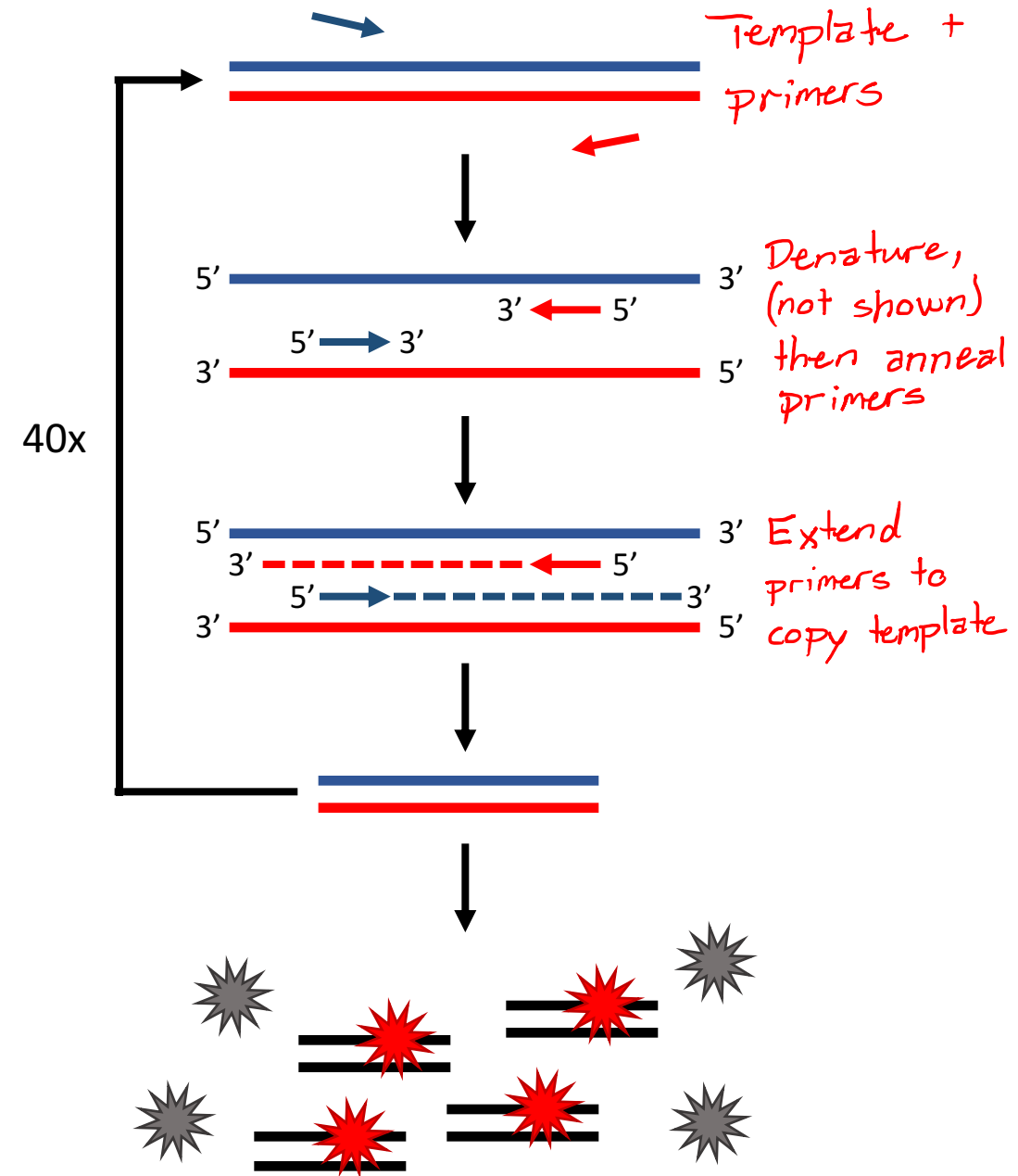
- Qiagen QIAquick PCR purification kit



steps	contents	purpose
prepare	5X Buffer PB to cDNA	Guanidinium thiocyanate Binds DNA to silica membrane
bind	silica membrane in column	
wash	Buffer PE	Ethanol to wash away salts/ buffers ** then, get rid of all ethanol
elute	water	Elute DNA by competing with its binding to the silica membrane

Perform qPCR

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

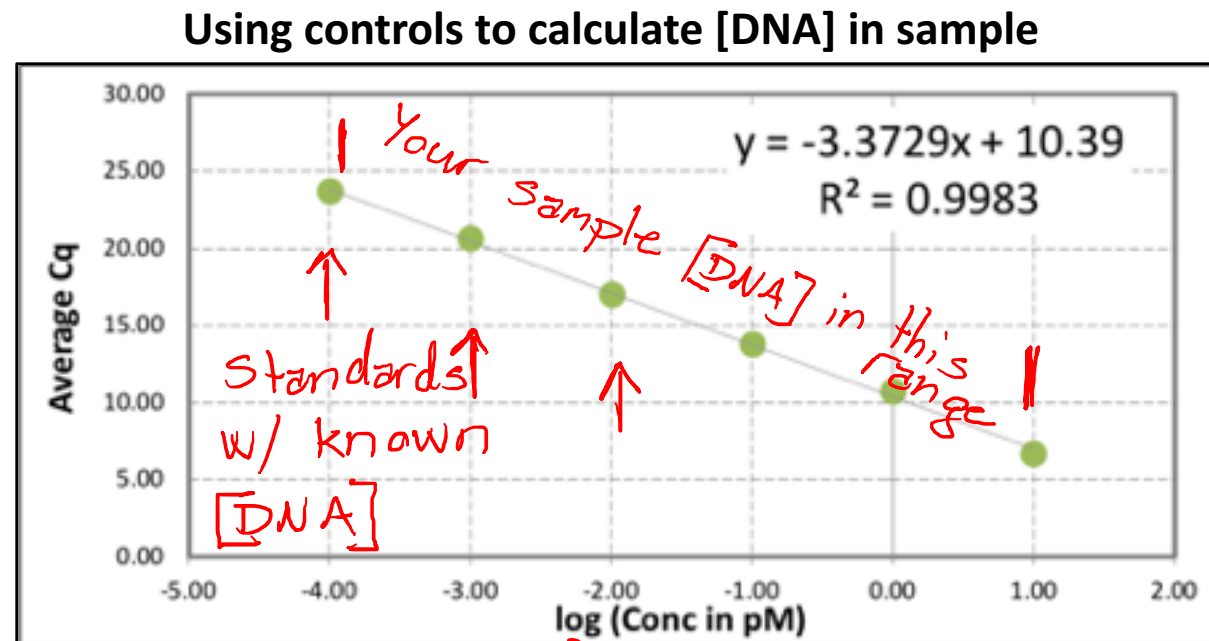
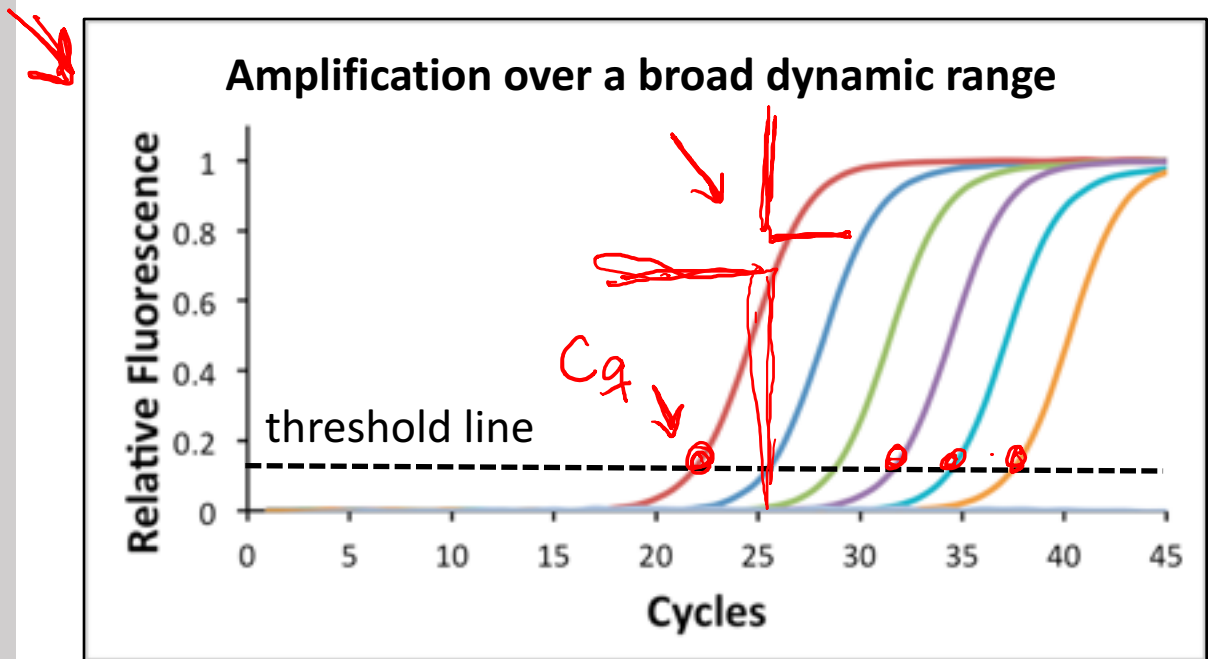


Perform qPCR

- 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

On publishing qPCR data:

<http://clinchem.aaccjnl.org/content/55/4/611>



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



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



TCGA data describes



...including



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



...using



Exploring the TCGA database

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

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects **Exploration** Analysis Repository

Quick Search Manage Sets Login Cart GDC Apps

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

Primary Site

Program

Project

- FM-AD 18,204
- TARGET-NBL 1,187
- TCGA-BRCA 1,206
- TARGET-AML 969
- TARGET-WT 852
- TCGA-GBM 817
- TCGA-OV 808
- TCGA-LUAD 585
- TCGA-UCEC 580
- TCGA-KIRC 537
- TCGA-HNSC 528
- TCGA-LGG 518
- TCGA-THCA 507
- TCGA-LUSC 504
- TCGA-PRAD 500
- TCGA-SKCM 479
- TCGA-COAD 461

Genes

Distribution of Most Frequently Mutated Genes

Gene	% of Cases Affected
APC	77.25%
TTN	64.50%
TP53	58.25%
KRAS	43.50%
MUC16	38.00%
BRCA1	~35%
BRCA2	~32%
OBFC1	~30%
FAT2	~28%
SOX11	~27%
ITIH3	~26%
MUC4	~25%
ZNF134	~24%
FAT3	~23%
PCLD	~22%
CSMD3	~21%
FUT9	~20%
LRP1B	~20%
DNAH5	~20%
NEB	~20%

Overall Survival Plot

442 Cases with Survival Data

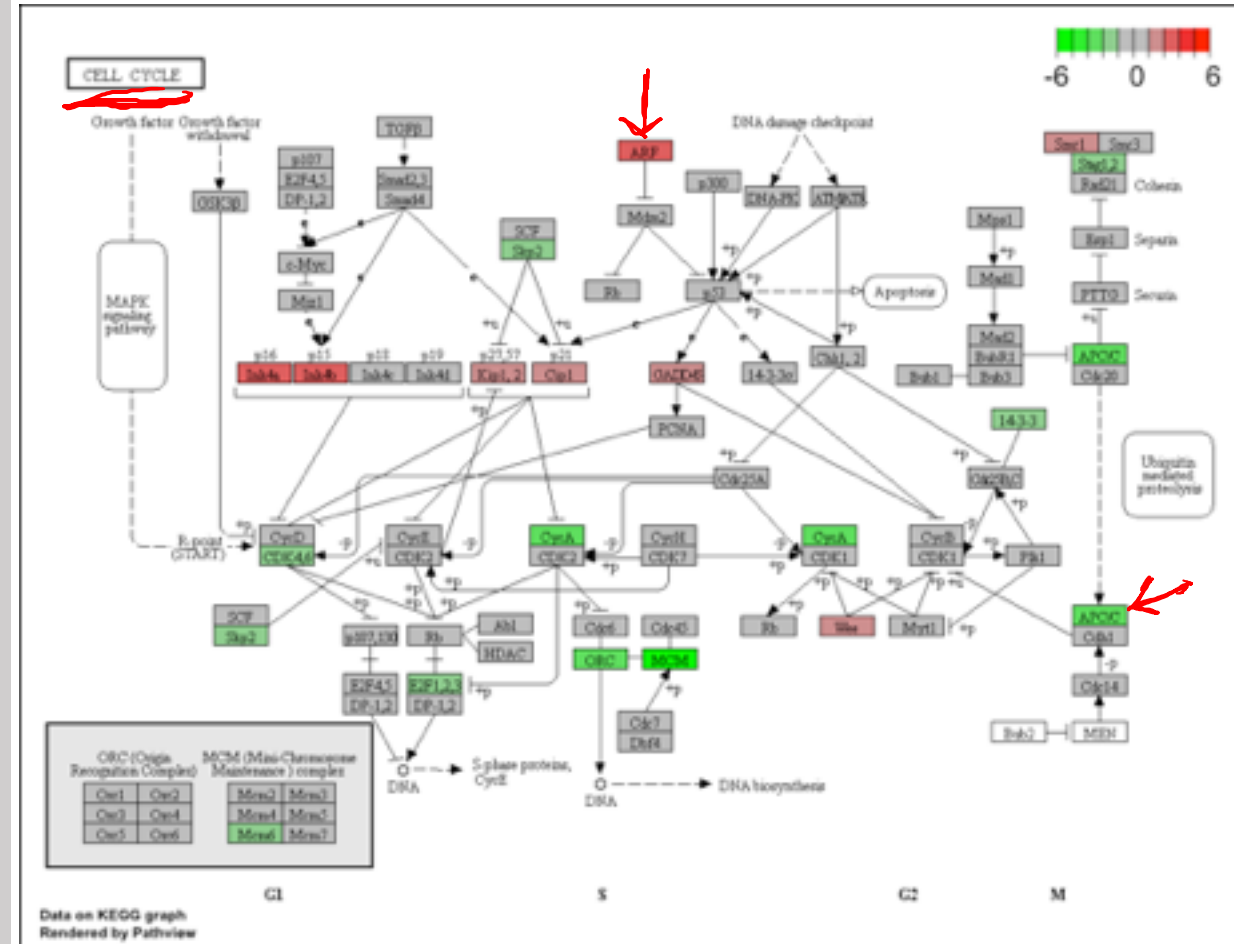
Showing 1 - 10 of 19,946 genes

JSON TSV Save/Edit Gene Set

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%)	4,008 / 10,202	142		
KRAS	Kirsten rat sarcoma viral oncogene homolog	12p12.1	protein_coding	174 / 400 (43.50%)	839 / 10,202	33		
MUC16	mucin 16, cell surface associated	19p13.2	protein_coding	152 / 400 (38.00%)	2,466 / 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?
- What is a cell line?

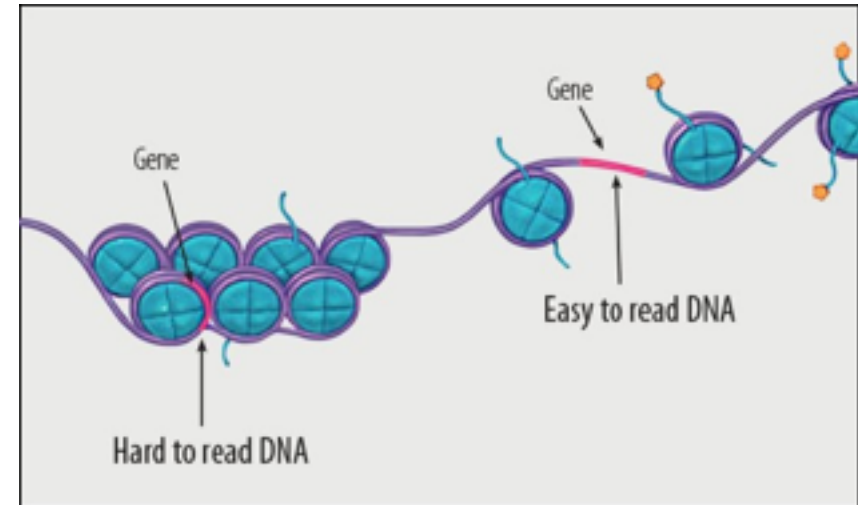


There is more to DNA than just the sequence

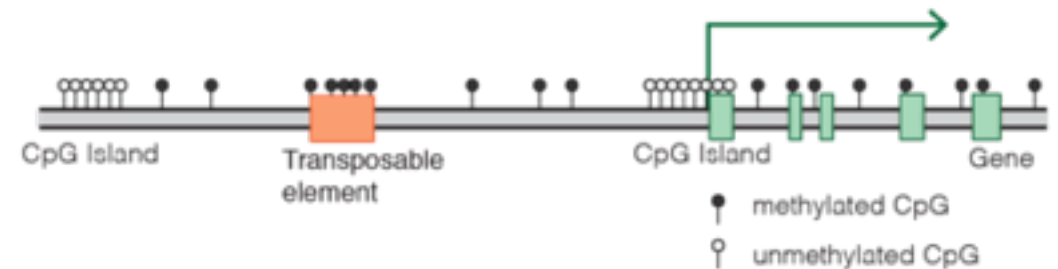
Ignore this slide

☹

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



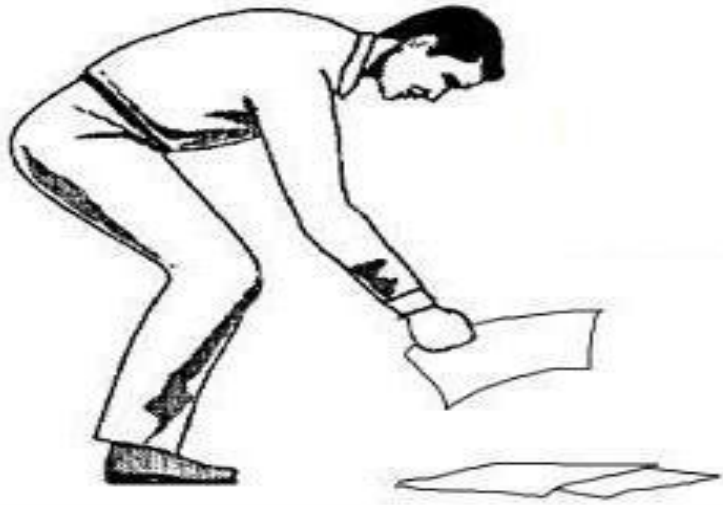
Typical mammalian DNA methylation landscape



SPRING BREAK 2018!



your  ecards
someecards.com



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