

# Data

Non-Small Cell Lung Cancer

Squamous Cell Lung Cancer

Adenocarcinoma

# Mutation

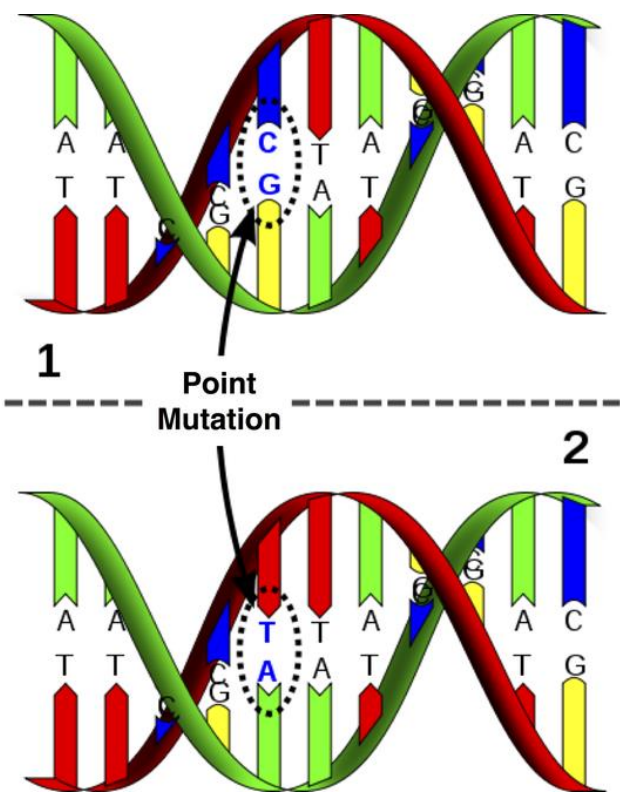
Mutation | 389296 obs. of 9 variables

- Mutation data has:
  - SampleID
  - Chromosomal Location
    - Chr
    - Start, Stop
  - Base pair change
    - Ref: Reference
    - Alt: Alternative

- Gene

- Effect

Sample_ID	Gene	Chr	Start	Stop	Ref	Alt	Amino_Acid_Change	Effect
TCGA-49-4490-01	UBR4	chr1	19155452	19155452	C	T	p.E2097K	missense_variant
TCGA-49-4490-01	PLA2G2D	chr1	20115562	20115562	G	A	p.C79C	synonymous_variant
TCGA-49-4490-01	MTF1	chr1	37822366	37822366	A	G	p.S508P	missense_variant
TCGA-49-4490-01	CLCA2	chr1	86443922	86443922	G	A	p.D542N	missense_variant
TCGA-49-4490-01	WNT2B	chr1	112525957	112525957	G	C		downstream_gene_variant
TCGA-49-4490-01	LINGO4	chr1	151801168	151801168	C	T	p.D513N	missense_variant

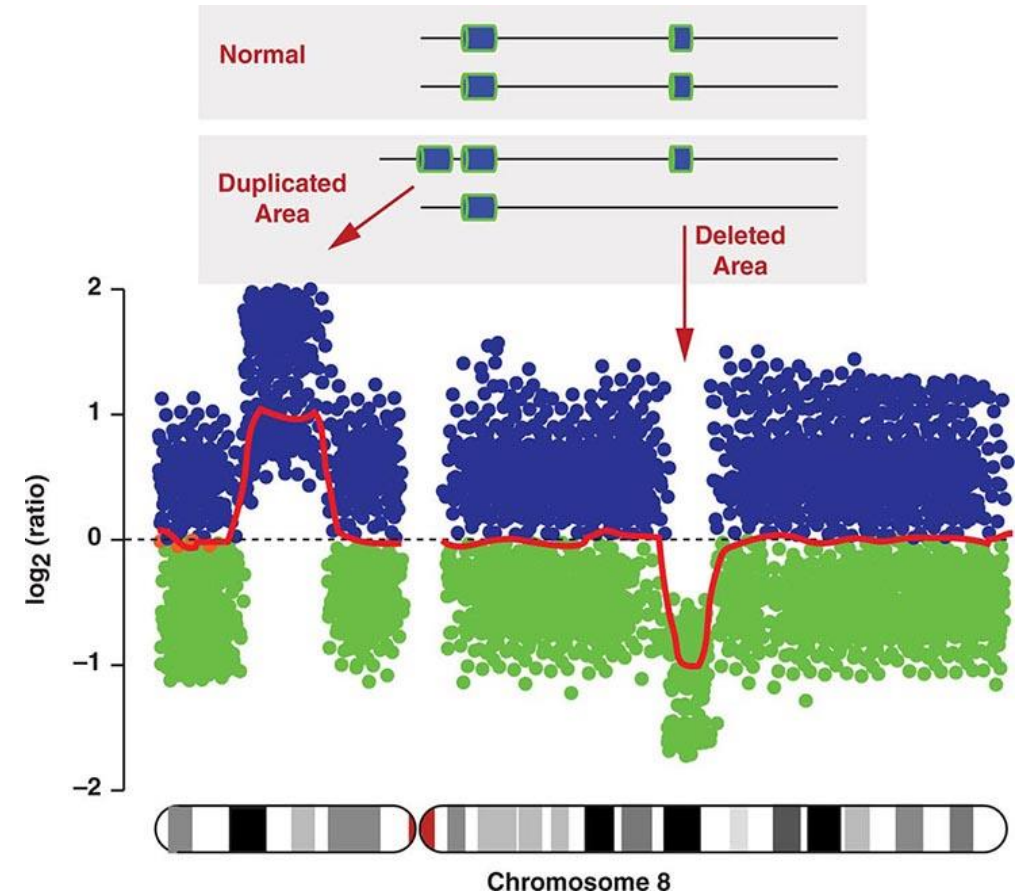


# CNV

• CNV

24791 obs. of 1022 variables

- CNV data has:
  - Gene
  - Chromosomal location of the gene
  - Amount of DNA for a gene per sample
    - -2: 0 copies of the gene (double deletion)
    - -1: 1 copy of the gene (deletion)
    - 0: 2 copies of the gene (normal)
    - 1: 3 copies of the gene (duplication)
    - 2: 4 copies of the gene (double duplication)

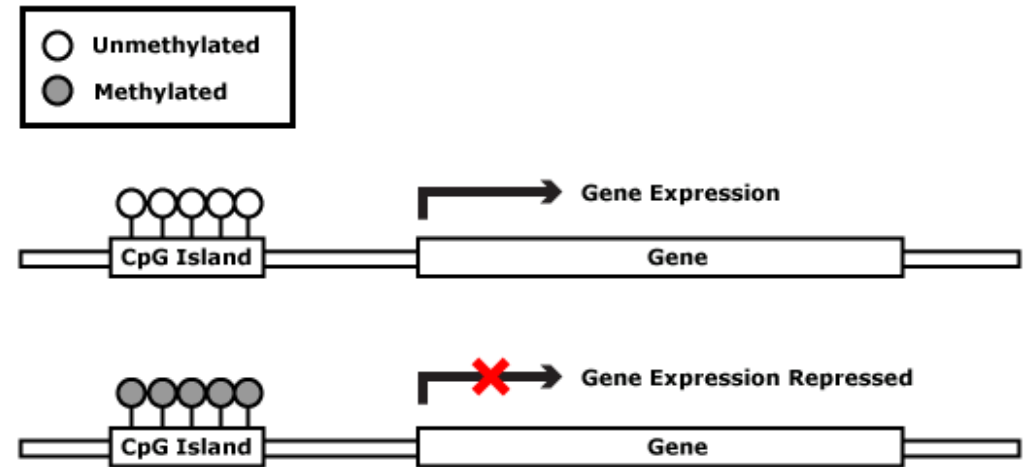


Gene	Chr	Start	Stop	Strand	TCGA-18-3406-01	TCGA-18-3407-01	TCGA-18-3408-01	TCGA-18-3409-01	TCGA-18-3410-01
A1BG	chr19	58345178	58353499	-1	2	1	0	-1	1
A1CF	chr10	50799409	50885675	-1	-1	1	0	0	0
A2M	chr12	9067664	9116229	-1	0	-1	0	0	1
A2ML1	chr12	8822472	8887001	1	0	-1	0	0	1
A2MP1	chr12	9228533	9275817	-1	0	-1	0	0	1
A3GALT2	chr1	33306766	33321098	-1	-1	0	-1	0	0

# Methylation

► Methylation 485577 obs. of 914 variables

- Methylation data has:
  - probeID
  - Chromosomal location of the probe
  - Gene
  - $\beta$ -value for a probe per sample
    - Ranges from 0 to 1
    - Roughly speaking 0: Unmethylated --> 1: Methylated



- Methylation studies use M-values for statistical validity

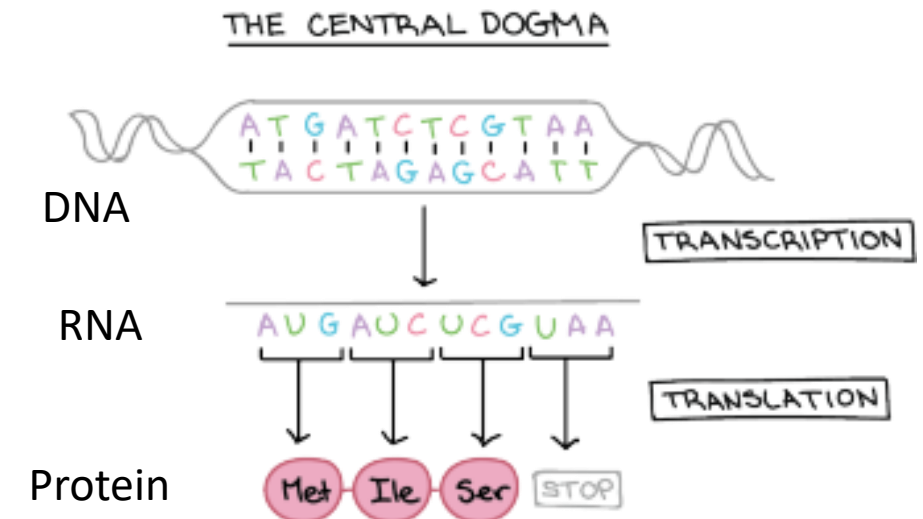
$$M_i = \log_2\left(\frac{\beta_i}{1-\beta_i}\right)$$

probeID	chr	Start	Stop	Strand	Gene	Relation_CpG_Island	TCGA-44-4112-01	TCGA-NJ-A4YP-01	TCGA-22-4613-01
cg00000029	chr16	53434199	53434201	-	RBL2	N_Shore	0.34335	0.3992	0.3073
cg00000108	chr3	37417714	37417716	-	C3orf35		NA	NA	NA
cg00000109	chr3	172198246	172198248	-	FNDC3B		NA	NA	NA
cg00000165	chr1	90729116	90729118	+	BARHL2	S_Shore	0.48315	0.5491	0.6697
cg00000236	chr8	42405775	42405777	+	VDAC3		0.85435	0.7632	0.9144
cq00000289	chr14	68874421	68874423	-	ACTN1		0.66545	0.5226	0.7565

# Gene Expression

GeneExpression | 60483 obs. of 1140 variables

- Gene Expression data has:
  - Gene
  - Chromosomal location of the gene
  - Gene Expression value for a gene per sample
    - $\log_2$  transformed



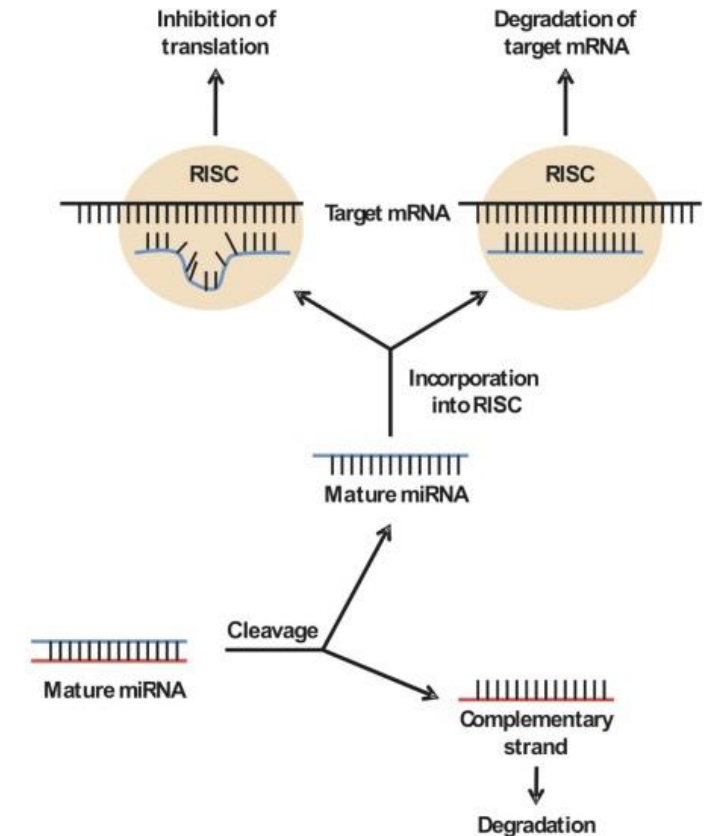
Gene	Chr	Start	Stop	Strand	TCGA-05-4244-01	TCGA-05-4249-01	TCGA-05-4250-01	TCGA-05-4382-01	TCGA-05-4384-01
TSPAN6	chrX	100627109	100639991	-	4.159317	3.985334	4.24140648	3.384259	3.276697
TNMD	chrX	100584802	100599885	+	0.000000	0.000000	0.06642033	0.000000	0.000000
DPM1	chr20	50934867	50958555	-	4.278499	4.727652	5.22596122	4.742701	4.527401
SCYL3	chr1	169849631	169894267	-	2.109782	2.375002	1.38837836	1.537189	2.376222
C1orf112	chr1	169662007	169854080	+	1.683045	1.217652	1.58297357	1.298543	1.017860
FGR	chr1	27612064	27635277	-	3.272660	2.667326	2.90317370	3.983659	3.302965

# miRNA expression

miRNA

2281 obs. of 881 variables

- miRNA expression data has:
  - MIMATID
  - Name
  - Chromosomal location of the miRNA
  - miRNA expression value for a miRNA per sample
    - $\log_2$  transformed

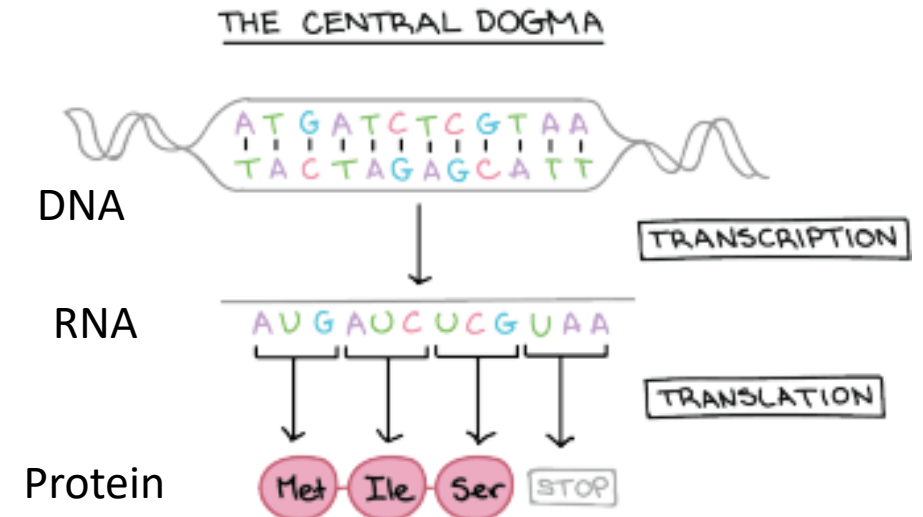


MIMATID	Name	Chr	Start	Stop	Strand	TCGA-NJ-A4YP-01	TCGA-86-8278-01	TCGA-62-A470-01	TCGA-44-6778-01
MIMAT0000062	hsa-let-7a-5p	chr11	122146568	122146589	-	14.939725	14.771870	14.332978	14.454362
MIMAT0000063	hsa-let-7b-5p	chr22	46113691	46113712	+	14.327167	14.010682	13.384376	13.834904
MIMAT0000064	hsa-let-7c-5p	chr21	16539838	16539859	+	9.595746	9.645857	9.848584	10.302141
MIMAT0000065	hsa-let-7d-5p	chr9	94178841	94178862	+	7.405218	6.728663	7.684090	7.625472
MIMAT0000066	hsa-let-7e-5p	chr19	51692793	51692814	+	9.040771	9.723797	9.878375	9.523750
MIMAT0000067	hsa-let-7f-5p	chr9	94176353	94176374	+	12.597921	12.543699	12.880784	12.667485

# Proteome

Proteome | 276 obs. of 694 variables

- Proteome data has:
  - ProteinID
  - Protein expression value
    - Very limited (only 355 proteins measured)



sample	TCGA-44-7670-01	TCGA-44-5644-01	TCGA-55-6979-01	TCGA-73-7498-01	TCGA-78-7148-01
14-3-3_beta-R-V	0.2057878	-0.063604647	-0.1206294	0.02659158	0.085083210
14-3-3_epsilon-M-C	0.1950392	0.003267256	-0.1022047	0.25745218	0.094750030
14-3-3_zeta-R-V	0.7223250	0.412202881	-0.6572866	0.31065868	-0.216132474
4E-BP1-R-V	0.8593884	0.466939612	0.1727313	-0.71101007	-0.152257826
4E-BP1_pS65-R-V	0.4204227	0.620014099	0.1968878	0.13785554	0.006086408
4E-BP1_pT37_T46-R-V	0.7252506	0.794508908	0.3934152	0.01051546	-0.107475736

# Sample Overlap

