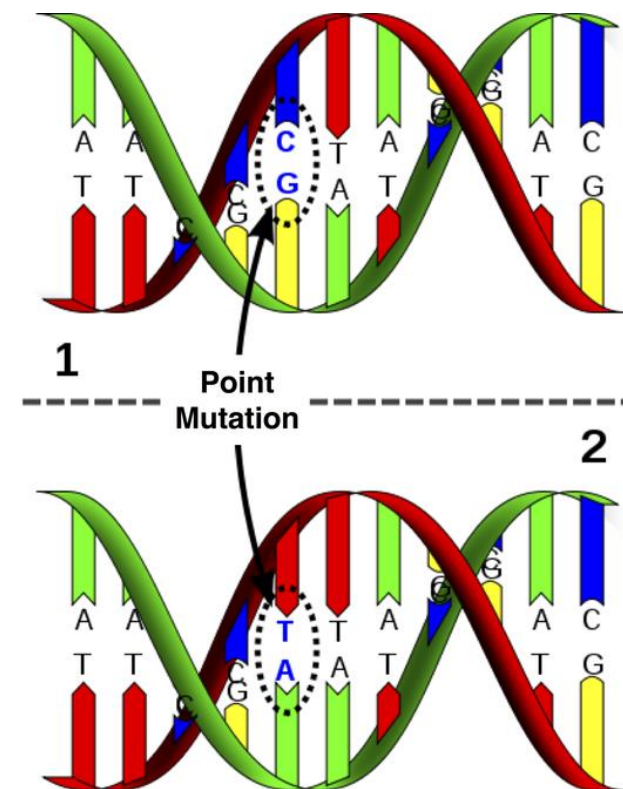# Hackathon Data

Melanoma

Same for Non-Small Cell Lung Cancer

Mutation

Mutation   422553 obs. of 9 variables

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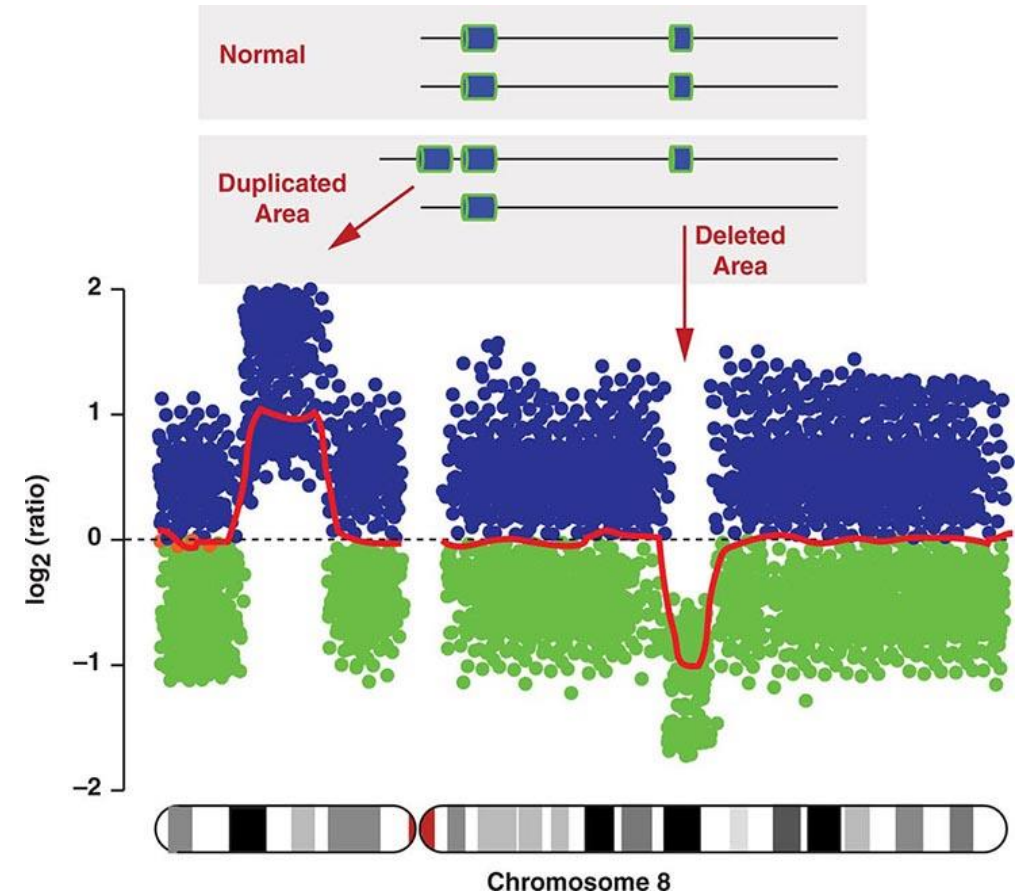   - Ref: Reference
    - Alt: Alternative
  - Gene
  - Effect

Sample	Chr	Start	Stop	Ref	Alt	Gene	Effect	Amino_Acid_Change
TCGA-D3-A3ML-06	chr5	140182973	140182973	G	A	PCDHA3	Missense_Mutation	p.D731N
TCGA-D3-A3ML-06	chr2	133541884	133541884	C	T	NCKAP5	Missense_Mutation	p.E834K
TCGA-D3-A3ML-06	chr19	51217544	51217544	C	T	SHANK1	Missense_Mutation	p.G179R
TCGA-D3-A3ML-06	chr8	2820757	2820757	G	A	CSMD1	Silent	p.I3148I
TCGA-D3-A3ML-06	chr19	43708353	43708353	C	T	PSG4	Missense_Mutation	p.E39K
TCGA-D3-A3ML-06	chr5	125919676	125919676	A	G	ALDH7A1	Missense_Mutation	p.V114A

CNV

Normal

Duplicated Area

Deleted Area

Chromosome 8

⊙ CNV   `24802 obs. of 372 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-3N-A9WB-06 TCGA-3N-A9WC-06 TCGA-3N-A9WD-06 TCGA-BF-AAP0-06 TCGA-D3-A1Q1-06
  A1BG chr19 58856544 58864865      -               0               0               0               0               1
  A1CF chr10 52559169 52645435      -               0               0              -1              -1               0
   A2M chr12  9220260  9268825      -               0               0               0               0               1
 A2ML1 chr12  8975068  9039597      +               0               0               0               0               1
 A2MP1 chr12  9381129  9428413      -               0               0               0               0               1
A3GALT2  chr1 33772367 33786699     -               0               1               1               1               0
```
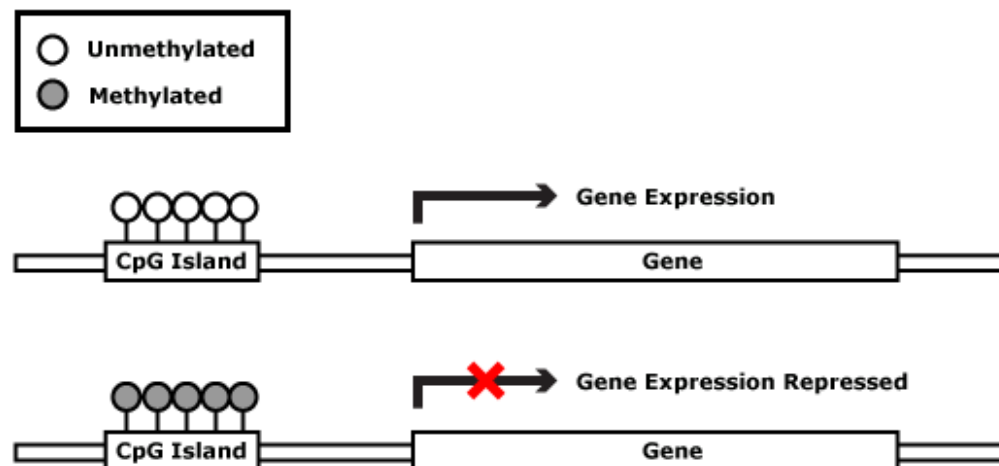
Methylation

Methylation   485577 obs. of 483 variables

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

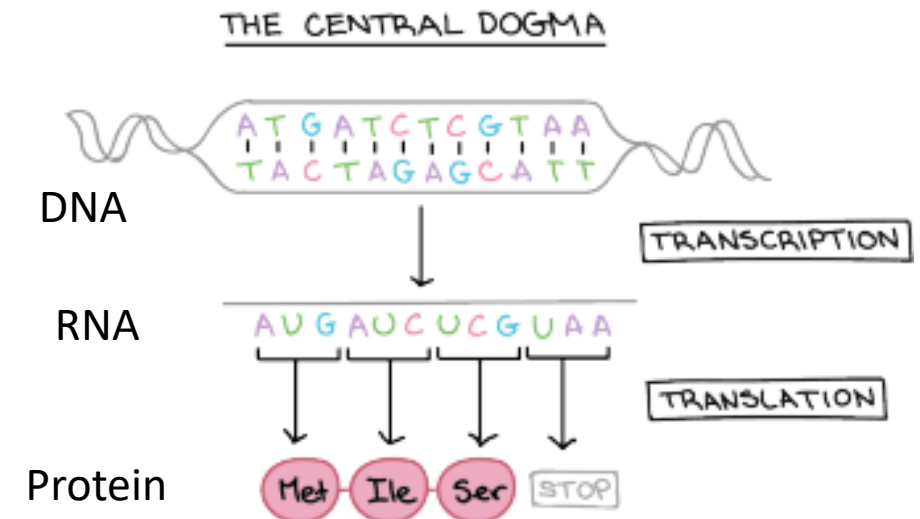- Methylation studies use M-values for statistical validity
  - $M_i = \log_2(\frac{\beta_i}{1-\beta_i})$

```
   probeID    Chr       Start        Stop Strand      Gene Relation_CpG_Island TCGA-FR-A728-01 TCGA-GN-A4U8-11 TCGA-GN-A8LL-06
cg00000029  chr16    53468111    53468113      -      RBL2             N_Shore          0.6603          0.5481          0.0729
cg00000108   chr3    37459205    37459207      -   C3orf35                                  NA              NA              NA
cg00000109   chr3   171916036   171916038      -    FNDC3B                                  NA              NA              NA
cg00000165   chr1    91194673    91194675      +    BARHL2             S_Shore          0.4969          0.3550          0.8921
cg00000236   chr8    42263293    42263295      +     VDAC3                              0.8443          0.9040          0.7284
cg00000289  chr14    69341138    69341140      -     ACTN1                              0.5590          0.6232          0.5661
```

Gene Expression

THE CENTRAL DOGMA

| GeneExpress... | 60531 obs. of 477 variables |

DNA

TRANSCRIPTION

RNA

TRANSLATION

Protein

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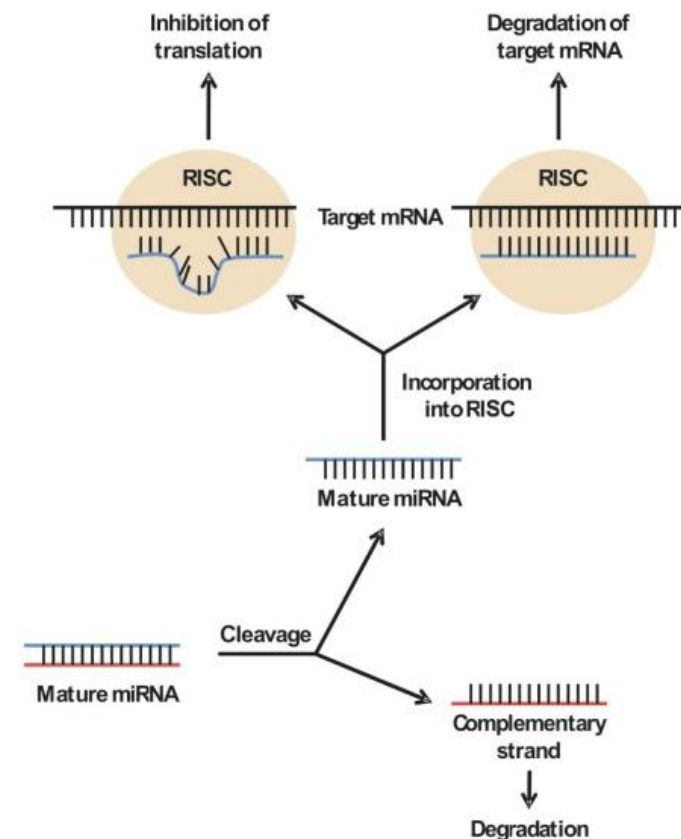 - Gene Expression value for a gene per sample
    - $\log_2$ transformed

Gene	Chr	Start	Stop	Strand	TCGA-3N-A9WB-06	TCGA-3N-A9WC-06	TCGA-3N-A9WD-06	TCGA-BF-A1PU-01	TCGA-BF-A1PV-01
A1BG	chr19	58856544	58864865	–	0.3971954	0.204836583	0.299306617	0.2128563	0.36825465
A1BG-AS1	chr19	58859117	58866549	+	1.5353568	1.473962214	1.284750861	1.1630289	1.39528537
A1CF	chr10	52559169	52645435	–	0.0000000	0.005726971	0.004329595	0.0000000	0.00000000
A2M	chr12	9220260	9268825	–	4.1967727	10.368824437	6.973470574	5.6173997	4.05991218
A2M-AS1	chr12	9217773	9220651	+	0.5874336	1.595183474	0.729753892	1.3490659	0.24652655
A2ML1	chr12	8975068	9039597	+	0.1202575	0.015012785	0.011358577	0.1725125	0.01486633

miRNA expression

▶ miRNA	2220 obs. of 458 variables

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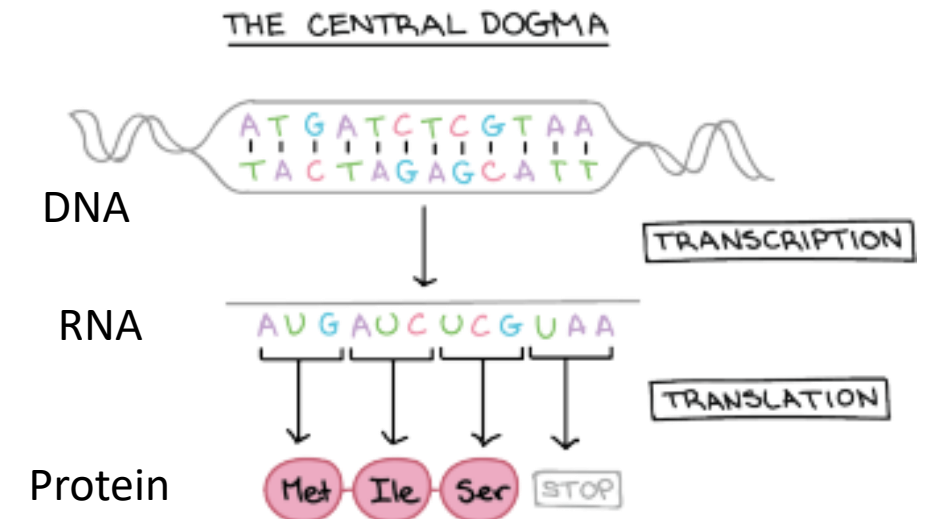   - $\log_2$ transformed

MIMATID	Name	Chr	Start	Stop	Strand	TCGA-EB-A4P0-01	TCGA-D3-A1Q5-06	TCGA-EE-A17X-06	TCGA-EE-A2MI-06
MIMAT0000062	hsa-let-7a-5p	chr11	122017276	122017297	–	15.996174	14.498864	16.695377	16.412696
MIMAT0000063	hsa-let-7b-5p	chr22	46509571	46509592	+	15.312683	13.393522	16.140099	15.675306
MIMAT0000064	hsa-let-7c-5p	chr21	17912158	17912179	+	8.542391	11.666346	8.265981	11.648024
MIMAT0000065	hsa-let-7d-5p	chr9	96941123	96941144	+	10.328891	8.195814	10.212315	8.369484
MIMAT0000066	hsa-let-7e-5p	chr19	52196046	52196067	+	11.930089	11.860895	11.890573	11.150990
MIMAT0000067	hsa-let-7f-5p	chr9	96938635	96938656	+	13.441136	11.335334	15.154012	12.983452

Proteome

THE CENTRAL DOGMA

DNA

RNA

Protein

TRANSCRIPTION

TRANSLATION

▶ Proteome   282 obs. of 355 variables

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

ProteinID	TCGA-EE-A2MP-06	TCGA-EE-A29B-06	TCGA-ER-A19C-06	TCGA-ER-A19T-01	TCGA-ER-A19K-01	TCGA-DA-A3F3-06
14-3-3_epsilon-M-C	-0.01888623	0.11638697	-0.1727033	-0.15173205	0.33662444	-0.063201033
4E-BP1-R-V	-0.30092073	-0.16322719	0.1460885	-0.58830481	0.32431452	-0.814006254
4E-BP1_pS65-R-V	0.26081385	-0.18179514	0.4417653	0.00145398	0.39180912	-0.402395626
4E-BP1_pT37_T46-R-V	0.83365814	0.09730264	1.0165106	0.21570561	1.20961019	-0.708621877
53BP1-R-E	0.15024322	-0.10594928	-0.7467942	-0.03148151	0.01517961	0.000736119
ACC_pS79-R-V	-0.20049804	-0.78342625	0.8651674	-0.21404717	0.33962337	0.012636966

Sample Overlap