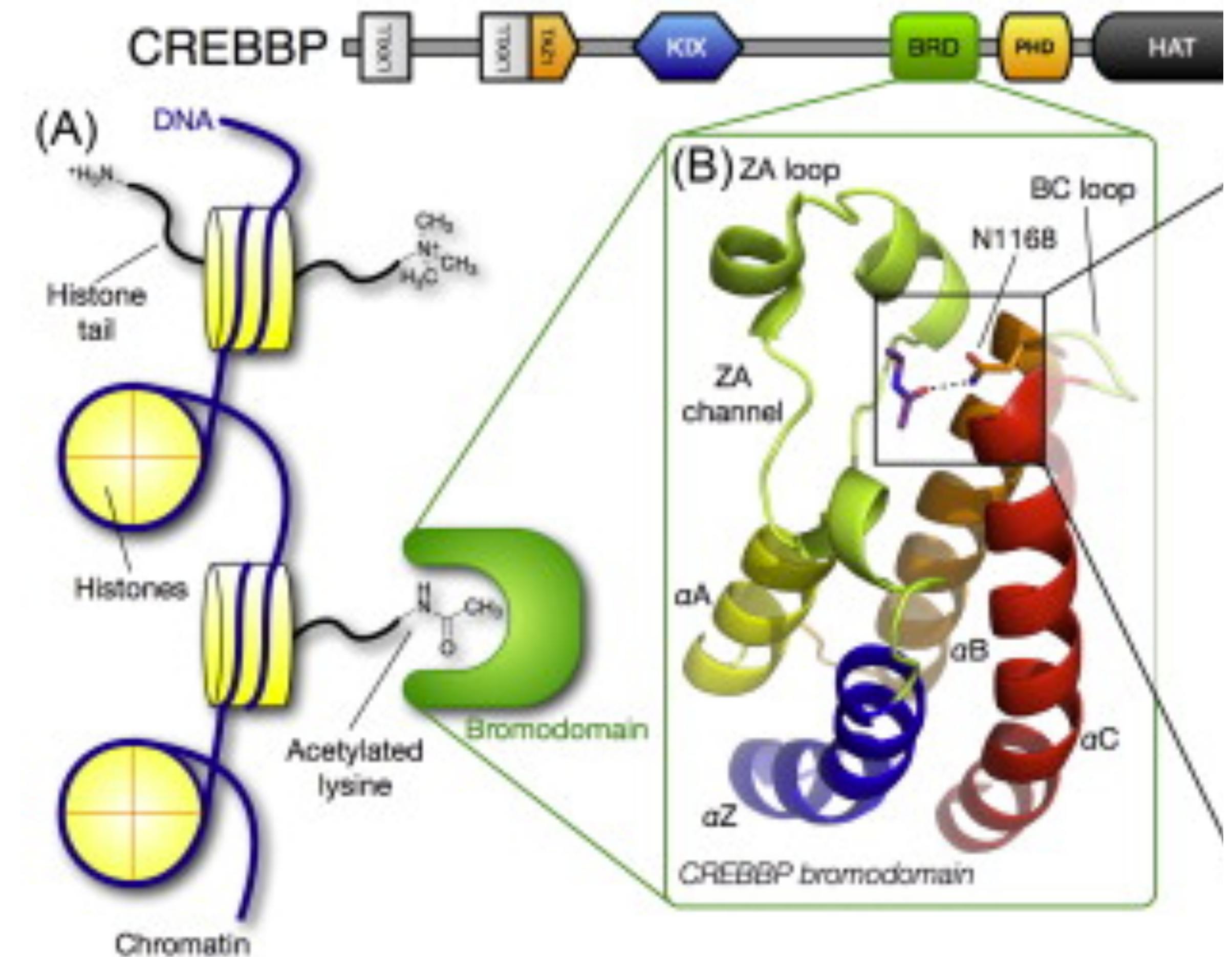


# Projects Introduction

---

# Bromodomain (transcription factors)

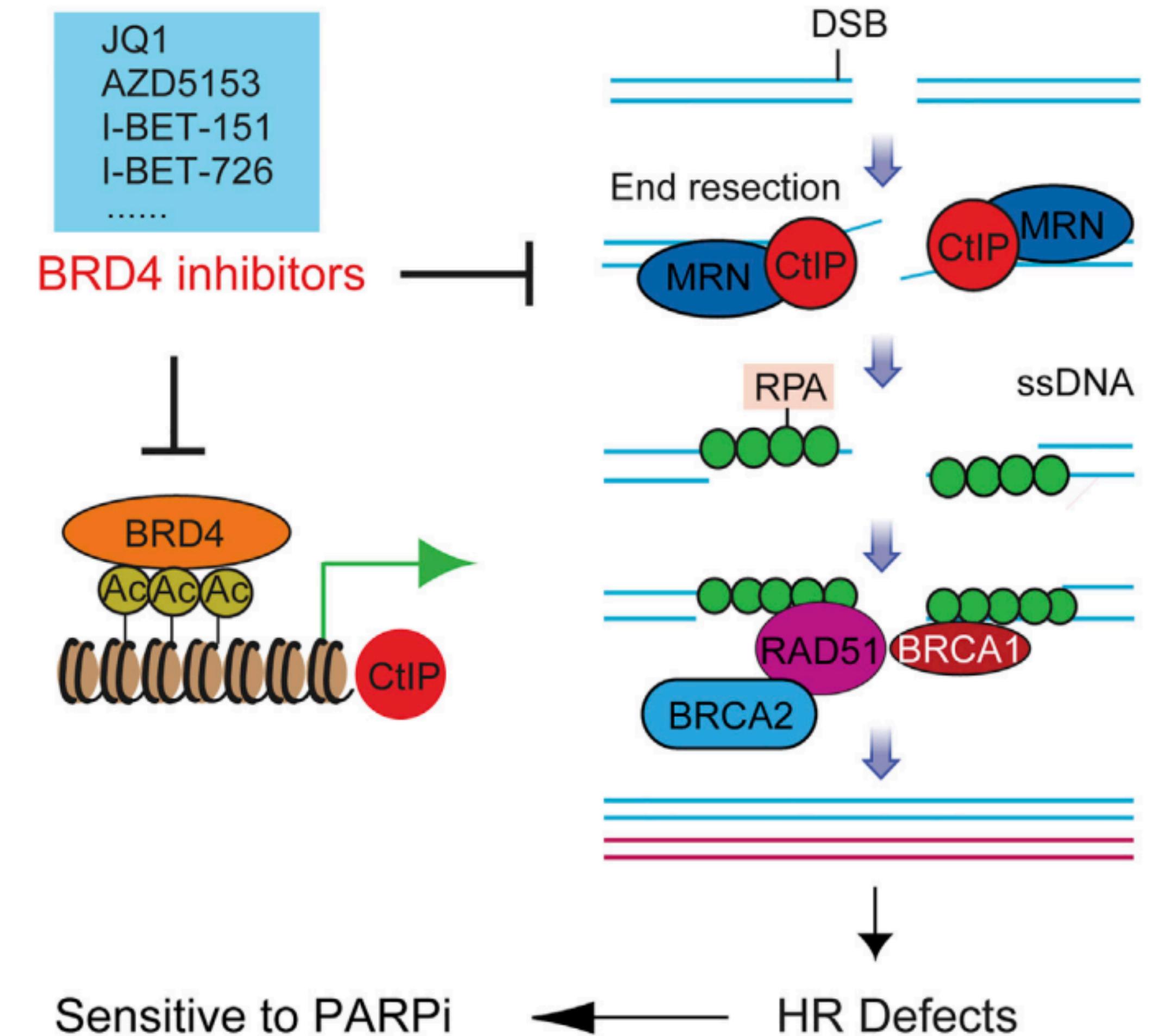
- The bromodomain is a conserved protein motif found in all eukaryotes and is the only known protein module that binds to acetylated lysine residues in histones.
- Bromodomain inhibitors prevent interaction between the bromodomain and the acetyl group, causing downregulation of certain genes.



Structural basis of acetyl-lysine recognition by the bromodomain

# Bromodomain & BRCA

- Bromodomain genes inhibition decreases homologous recombination competency.
- BRCA is a key component in the HR process.



# Identify the Relationship with TCGA data

- Relationship between BRCA mutations and bromodomain genes.

Primary Tumor Site	Tumor Sample Number	Normal Sample Number	BRCA1 CNV Loss	BRCA1 CNV Gain	BRCA2 CNV Loss	BRCA2 CNV Gain	BRCA1 Somatic Mutation	BRCA2 Somatic Mutation
Prostate	495	52	9	36	53	9	2	9
Ovary	374	0	36	18	46	29	15	13
Breast	812	112	87	43	88	70	29	21

Access	File Name	Cases Project	Data Category	Data Format	File Size Annotations
	<a href="#">e3a3f47a-280f-488d-ae87-5b7</a>				
	<a href="#">9c51b9398_gdc_realm_rehead.bam</a>	<a href="#">1 TCGA-KIRC</a>	Sequencing Reads	BAM	8.32 GB
	<a href="#">nationwidechildren.org_clinical.TCGA-BP-4343.xml</a>	<a href="#">1 TCGA-KIRC</a>	Clinical	BCR XML	56.3 KB
	<a href="#">TCGA-B0-4698-01A-01-BS1.6c</a>	<a href="#">1 TCGA-KIRC</a>	Biospecimen	SVS	304.49 MB
	<a href="#">07c221-172c-4244-890c-df21faec69fb.svs</a>				

Key Step: link the mutations with gene expression results.

# Linking Strategy

- Step1: download the mutation files with different variant calling methods from TCGA, file type is MAF.
- Step2: link these mutations to a series of interested genes with expression scores (FPKM).  
Access come from: <https://github.com/yuetang625-UT/BromodomainProteins>
- FPKM: fragments per kilobase of transcript per million fragments mapped

Access	File Name	Cases Project	Data Category	Data Format	File Size	Anno
	TCGA.BRCA.varsca_n.6c93f518-1956-44		varscan			#version gdc-1.0.0 #filedate 20170929 #annotation.spec gdc-1.0.1-public #n.analyzed.samples 986 #tumor aliquots submitter_id TCGA-3C-AAAU-01A-11D-A41F-09,TCGA-3C-AALI-01A-11D-A41F-09,TCGA-3C-AALJ-01A-11D-A41F-09 Hugo_Symbol Entrez_Gene_Id Center NCBI_Build Chromosome Start_Position End_Position CALML6 163688 WUGSC GRCh38 chr1 1916819 1916819 + Missense_Mutation SNP C PRKCZ 5590 WUGSC GRCh38 chr1 2172304 2172304 + Missense_Mutation SNP G CCDC27 148870 WUGSC GRCh38 chr1 3766586 3766586 + Missense_Mutation SNP G KCNAB2 8514 WUGSC GRCh38 chr1 6040634 6040634 + Silent SNP G G C PNRC2 55629 WUGSC GRCh38 chr1 23961791 23961791 + Missense_Mutation 3'UTR SNP C ATPIF1 93974 WUGSC GRCh38 chr1 28236188 28236188 + Missense_Mutation 3'UTR SNP C SMAP2 64744 WUGSC GRCh38 chr1 40422316 40422316 + Missense_Mutation 3'UTR SNP C CCDC30 728621 WUGSC GRCh38 chr1 42577033 42577033 + Missense_Mutation 3'UTR SNP C CCDC17 149483 WUGSC GRCh38 chr1 45621953 45621953 + Missense_Mutation 3'UTR SNP C FAM69A 388650 WUGSC GRCh38 chr1 92843906 92843906 + Missense_Mutation 3'UTR SNP C WDR47 22911 WUGSC GRCh38 chr1 108970228 108970228 + Missense_Mutation 3'UTR SNP A HSD3B1 3283 WUGSC GRCh38 chr1 119514202 119514202 + Missense_Mutation 3'UTR SNP C ADAMTSL4 54507 WUGSC GRCh38 chr1 150555531 150555531 + Missense_Mutation 3'UTR SNP C RFX5 5993 WUGSC GRCh38 chr1 151344268 151344268 + Missense_Mutation 3'UTR SNP C LELP1 149018 WUGSC GRCh38 chr1 153204755 153204755 + Missense_Mutation 3'UTR SNP C ADAM15 8751 WUGSC GRCh38 chr1 155055999 155055999 + Missense_Mutation 3'UTR SNP C IGSF9 57549 WUGSC GRCh38 chr1 159931873 159931873 + Missense_Mutation 3'UTR SNP C HMCN1 83872 WUGSC GRCh38 chr1 186189753 186189753 + Nonsense_Mutation 3'UTR SNP C OBSCN 84033 WUGSC GRCh38 chr1 228246639 228246639 + Missense_Mutation 3'UTR SNP C PCNXL2 80003 WUGSC GRCh38 chr1 233000479 233000479 + Missense_Mutation 3'UTR SNP C RYR2 6262 WUGSC GRCh38 chr1 237456728 237456728 + Missense_Mutation 3'UTR SNP C
	35-9806-37185266d	1,044 TCGA-BRCA	Simple Nucleotide Variation MAF		24.16 MB	
	248.DR-10.0.somatic.maf.gz					
	TCGA.BRCA.muse.b8ca5856-9819-459cf.gz		muse			
	87c5-94e91aca4032.	1,044 TCGA-BRCA	Simple Nucleotide Variation MAF		23.04 MB	
	DR-10.0.somatic.maf.gz					
	TCGA.BRCA.mutect.995c0111-d90b-414		mutect			
	0-bee7-3845436c3b	1,044 TCGA-BRCA	Simple Nucleotide Variation MAF		31.06 MB	
	42.DR-10.0.somatic.maf.gz					
	TCGA.BRCA.somatic.csniper.7dd592e3-59		somaticsniper			
	50-4438-96d5-3c71	1,044 TCGA-BRCA	Simple Nucleotide Variation MAF		16.91 MB	
	8aca3f13.DR-10.0.somatic.maf.gz					

# Linking Strategy

- Step1: download the mutation files.
- Step2: link these mutations to a series of interested genes with FPKM.  
Access come from: <https://github.com/yuetang625-UT/BromodomainProteins>

transfer\_id.py automatically combines the mutation information of a specific gene with metadata and FPKM expression levels of target genes.

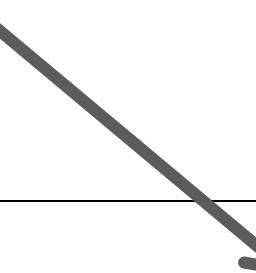
usage transfer\_id.py -h gives you all the parameters

-f MAF, --input\_file=MAF MAF file contains the somatic mutations from TCGA

-i ESMLID, --inquiry\_gene\_file=ESMLID Ensembl id for your research gene

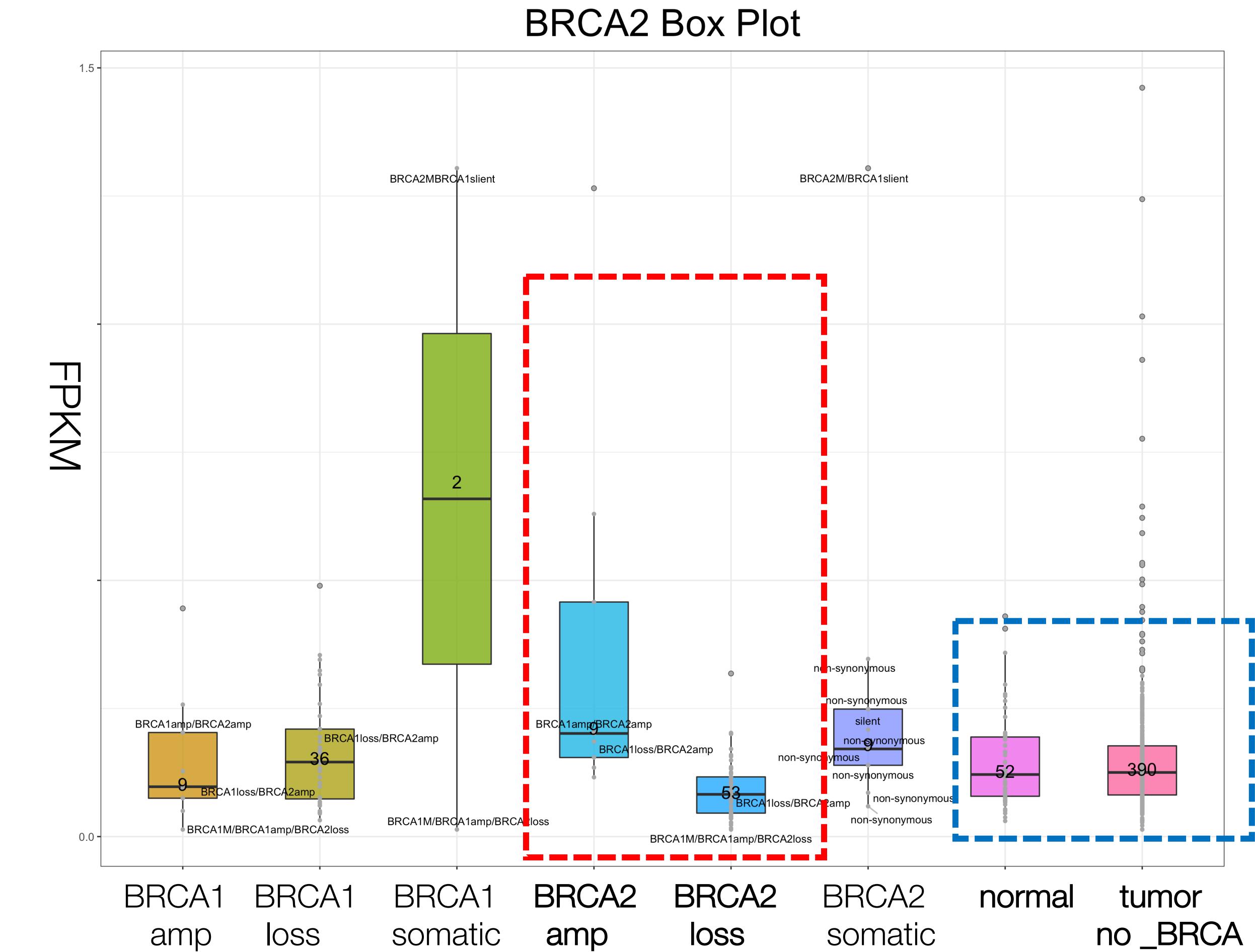
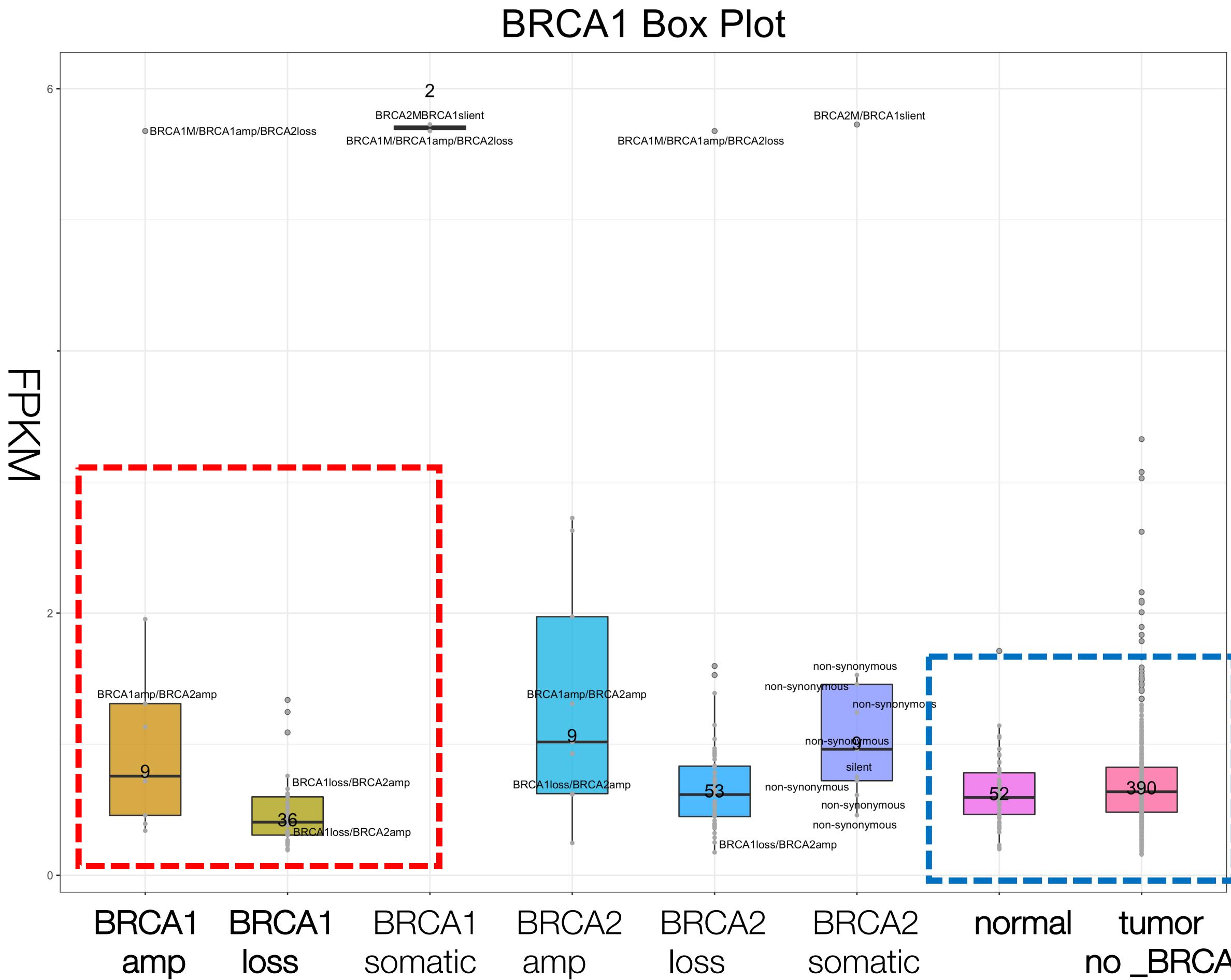
-g GENEID, --gene\_id=GENEID input the gene name you want to inquire

-o OUTDIR, --output-dir=OUTDIR Output directory (default: current)



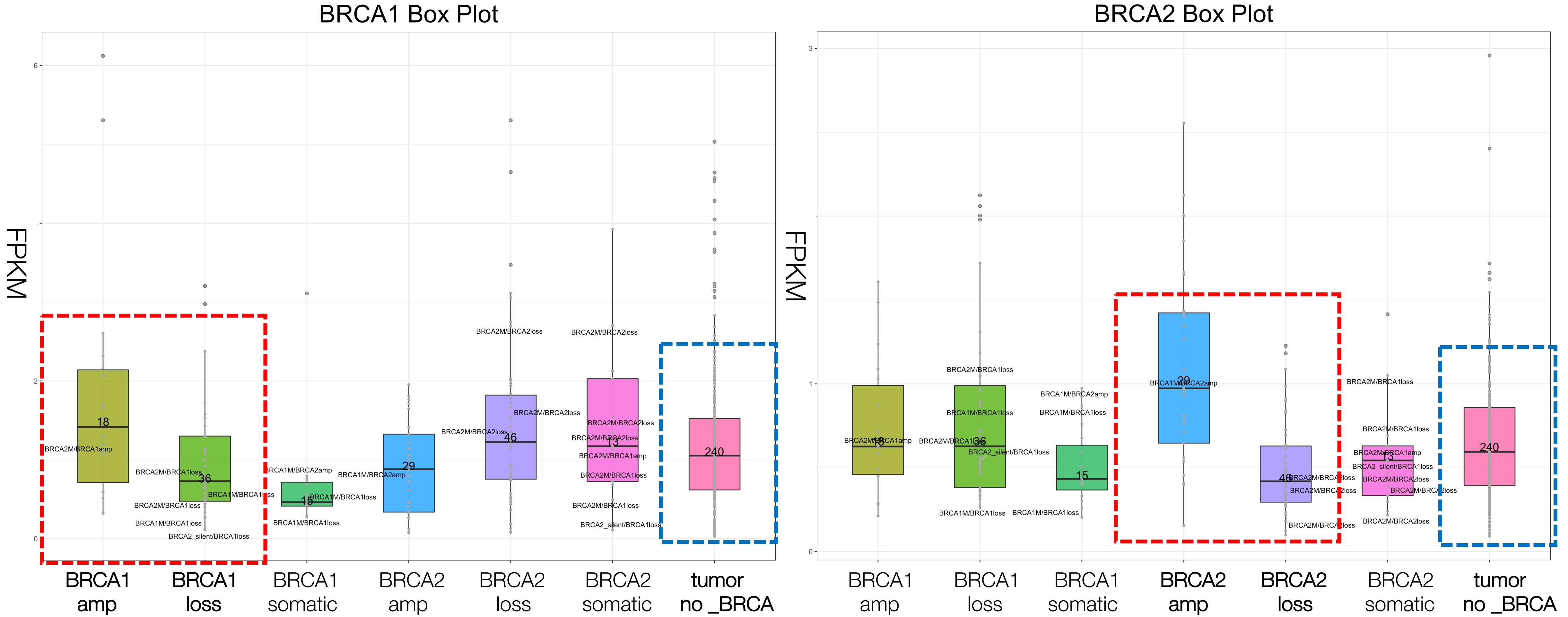
D	E	F	G	H	I	J	K	L	M	N	DW	DX
NCBI_Build	Chromosome	Start_Position	End_Position	Strand	Variant_Class	Variant_Type	Reference_A	Tumor_Seq	Tumor_Seq	dbSNP_RS	ENSG00000204256	ENSG00000169925
GRCh38	chr17	43099779	43099779	+	Missense_M	SNP	T	T	G	novel	51.77293639	9.751364844
GRCh38	chr17	43074377	43074377	+	Silent	SNP	C	C	T	novel	58.46646491	9.158682149
GRCh38	chr17	43091991	43091991	+	Silent	SNP	G	G	A	novel	58.46646491	9.158682149

# Results: BRCA Gene Expression -- Prostate



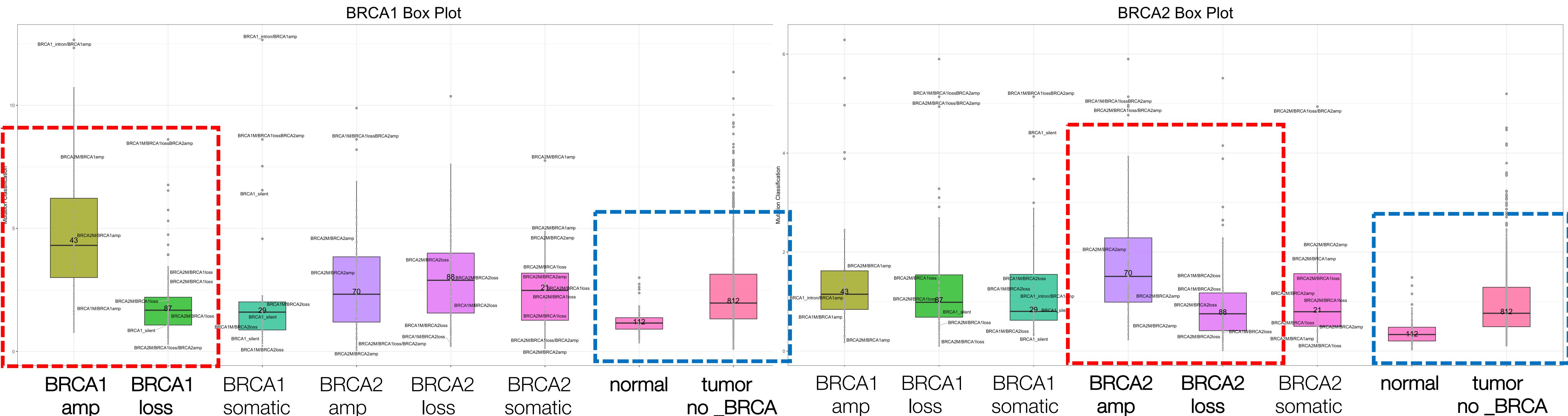
- Amplification of the BRCA gene causes high expression of BRCA

# Results: BRCA Gene Expression -- Ovary



- Amplification of the BRCA gene lead to high expression of BRCA.

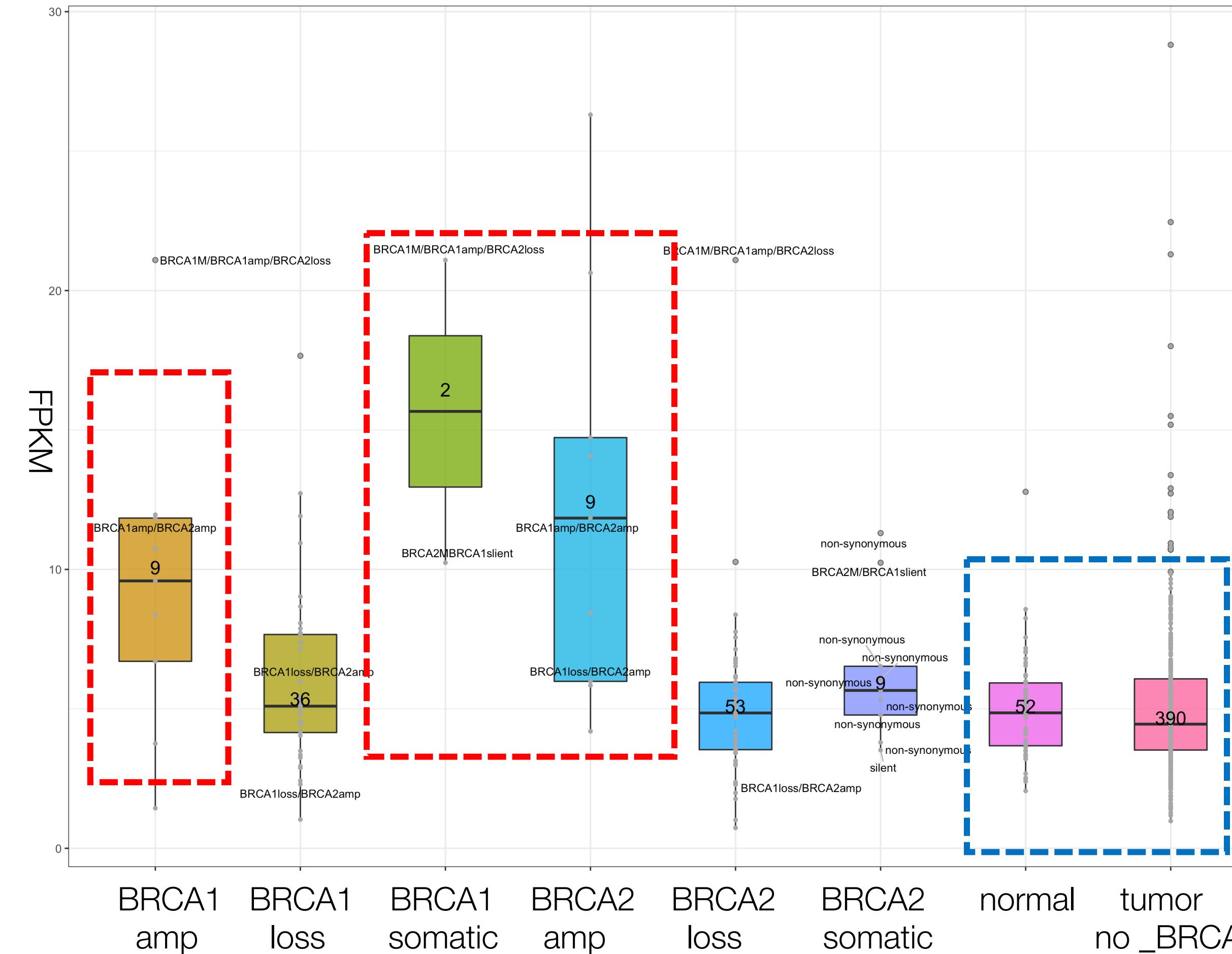
# Results: BRCA Gene Expression -- Breast



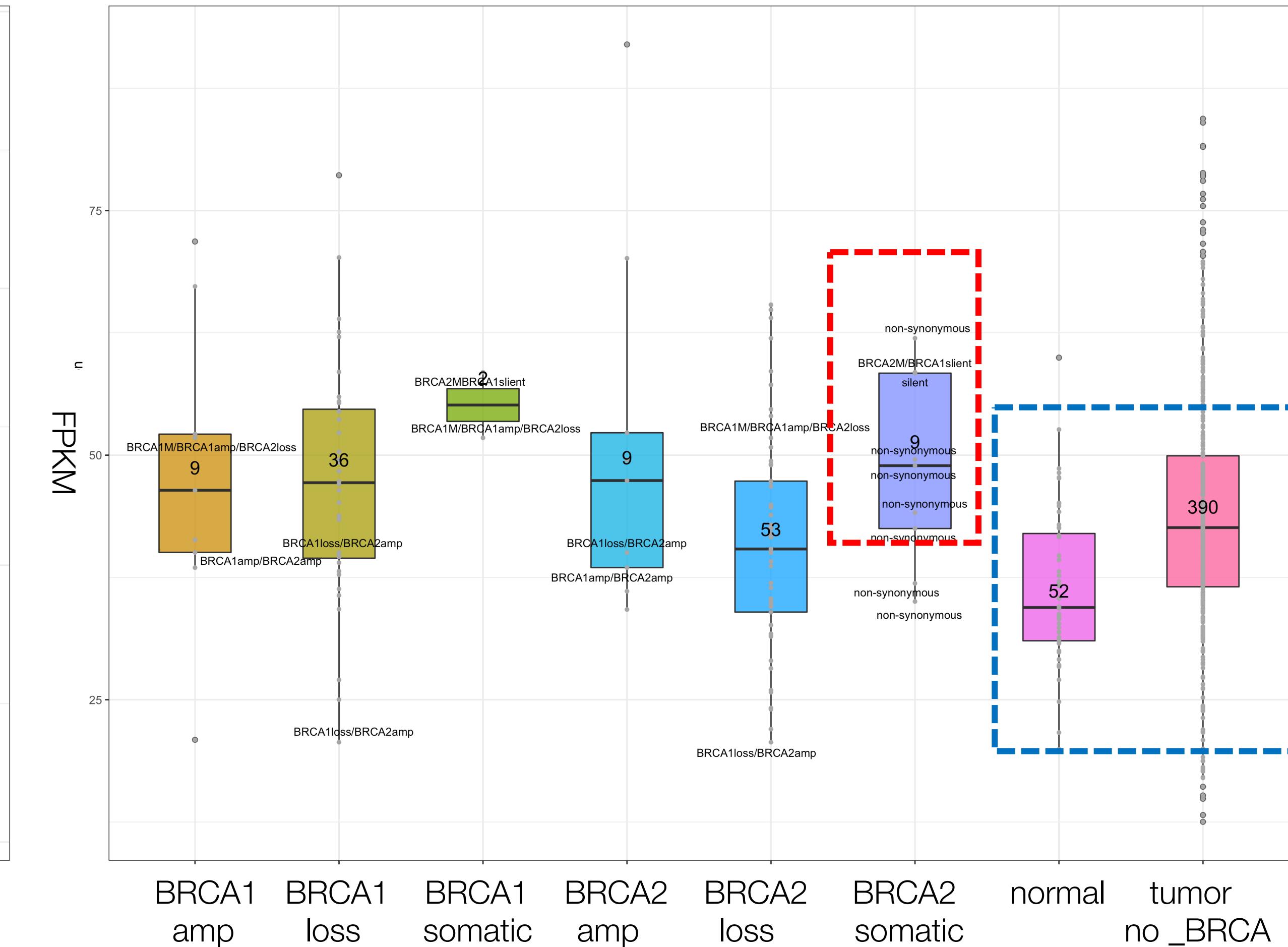
- Amplification of the BRCA gene lead to high expression of BRCA.
- Copy number variation affects gene expression.

# Results: Bromodomain Gene Expression -- Prostate

ATAD2 Box Plot

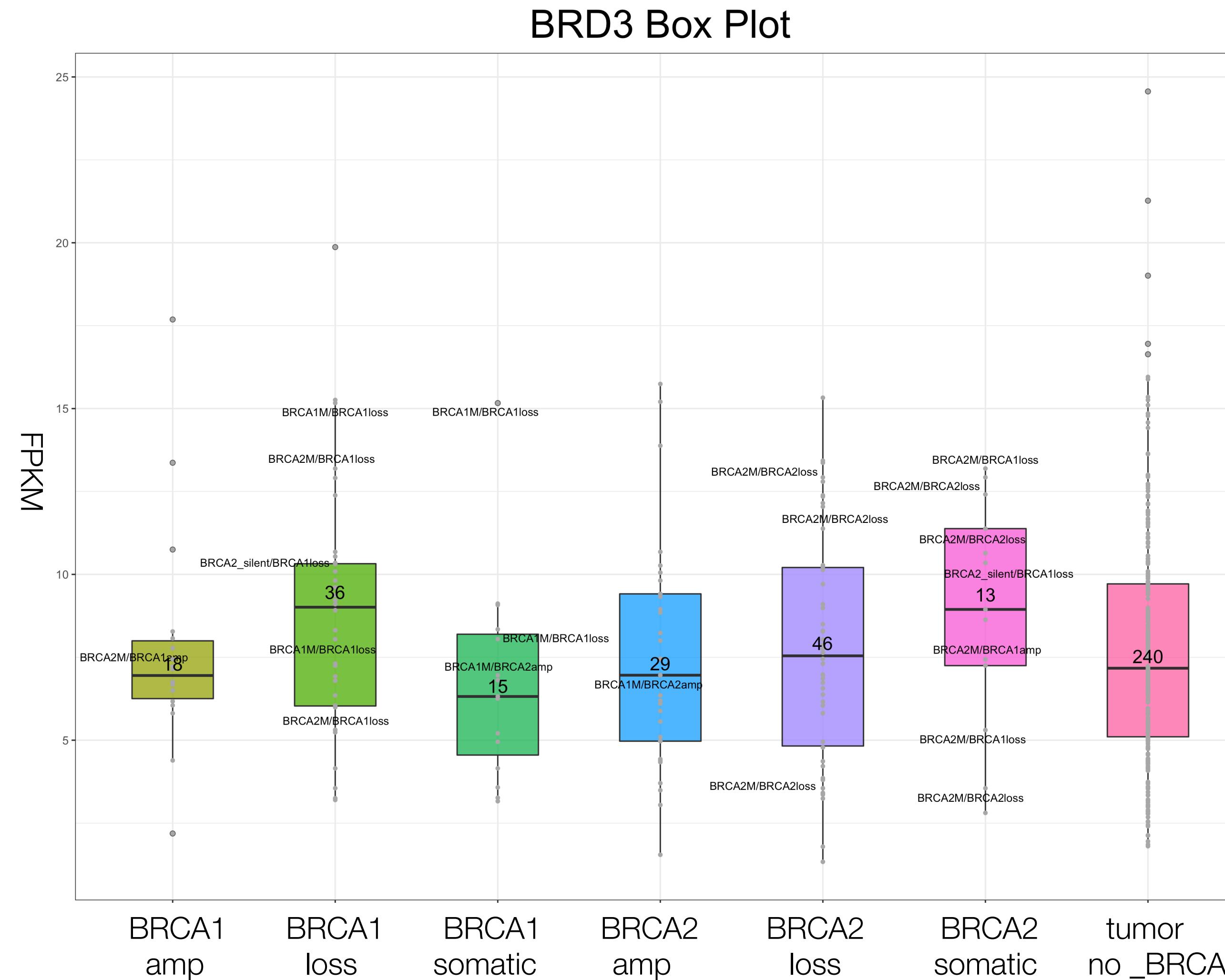


BRD2 Box Plot



- BRCA mutations affect Bromodomain gene expression in the prostate, causing expression abnormal.

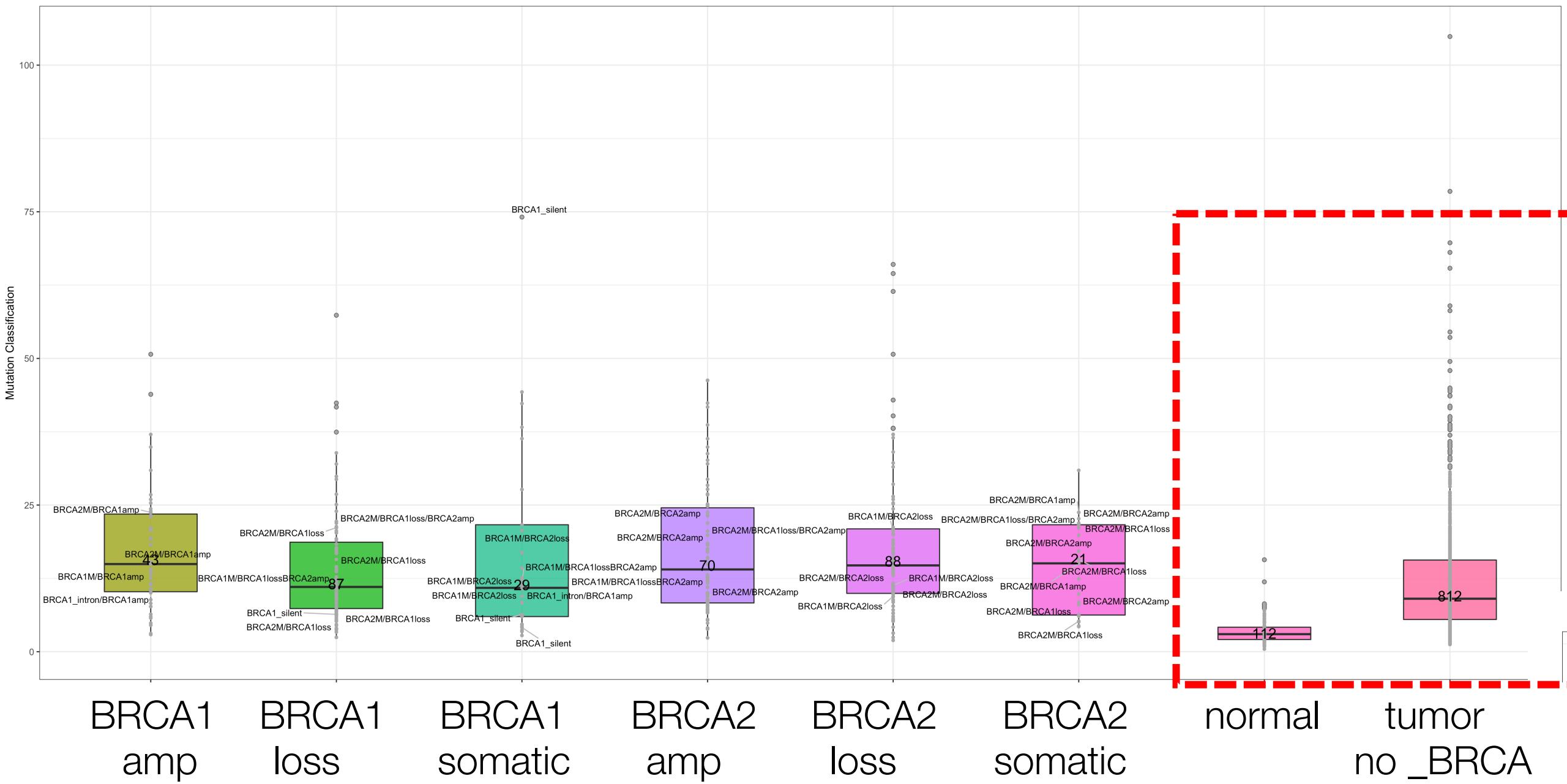
# Results: Bromodomain Gene Expression -- Ovary



- In the ovary, BRCA mutations do not affect Bromodomain gene expression.

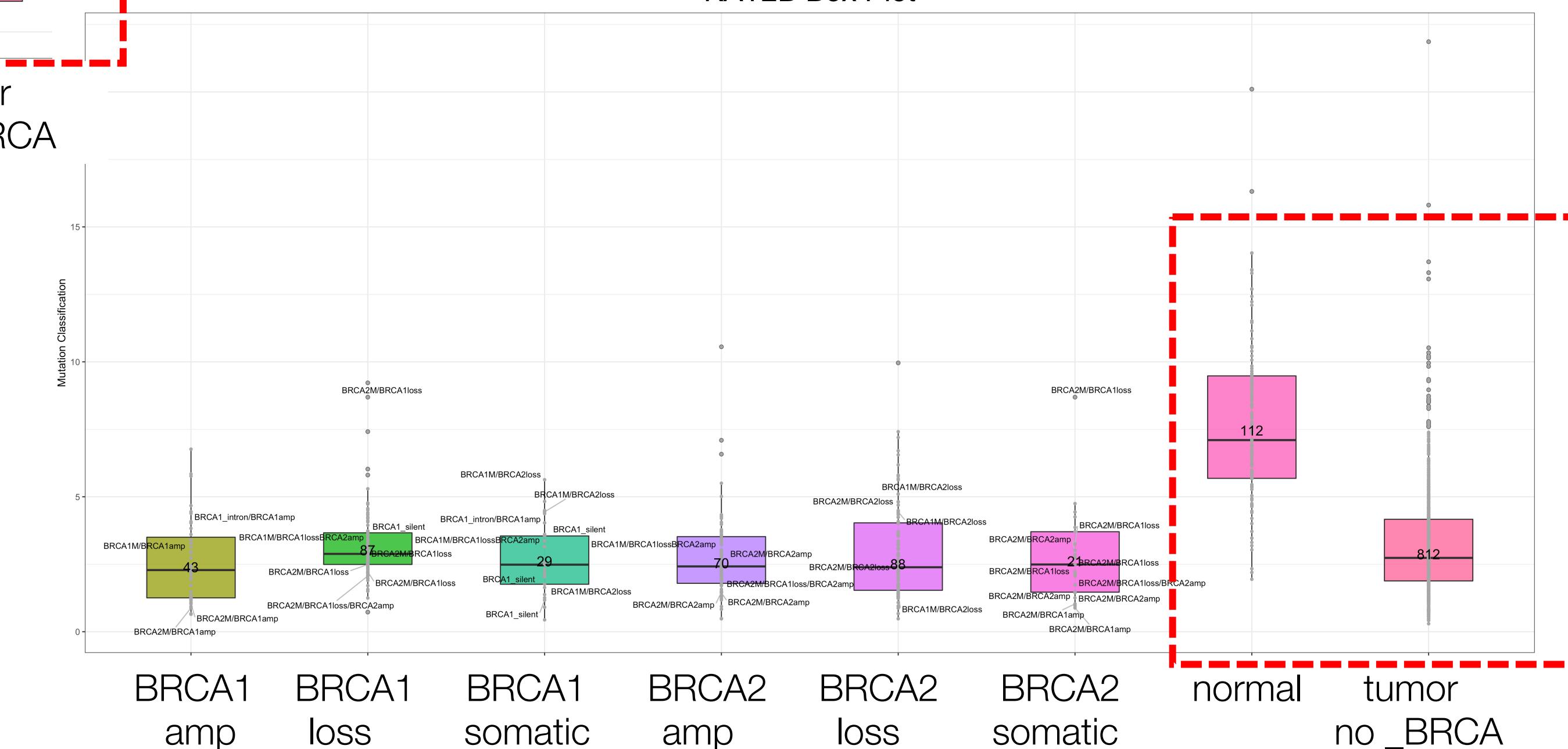
# Results: Bromodomain Gene Expression -- Breast

ATAD2 Box Plot



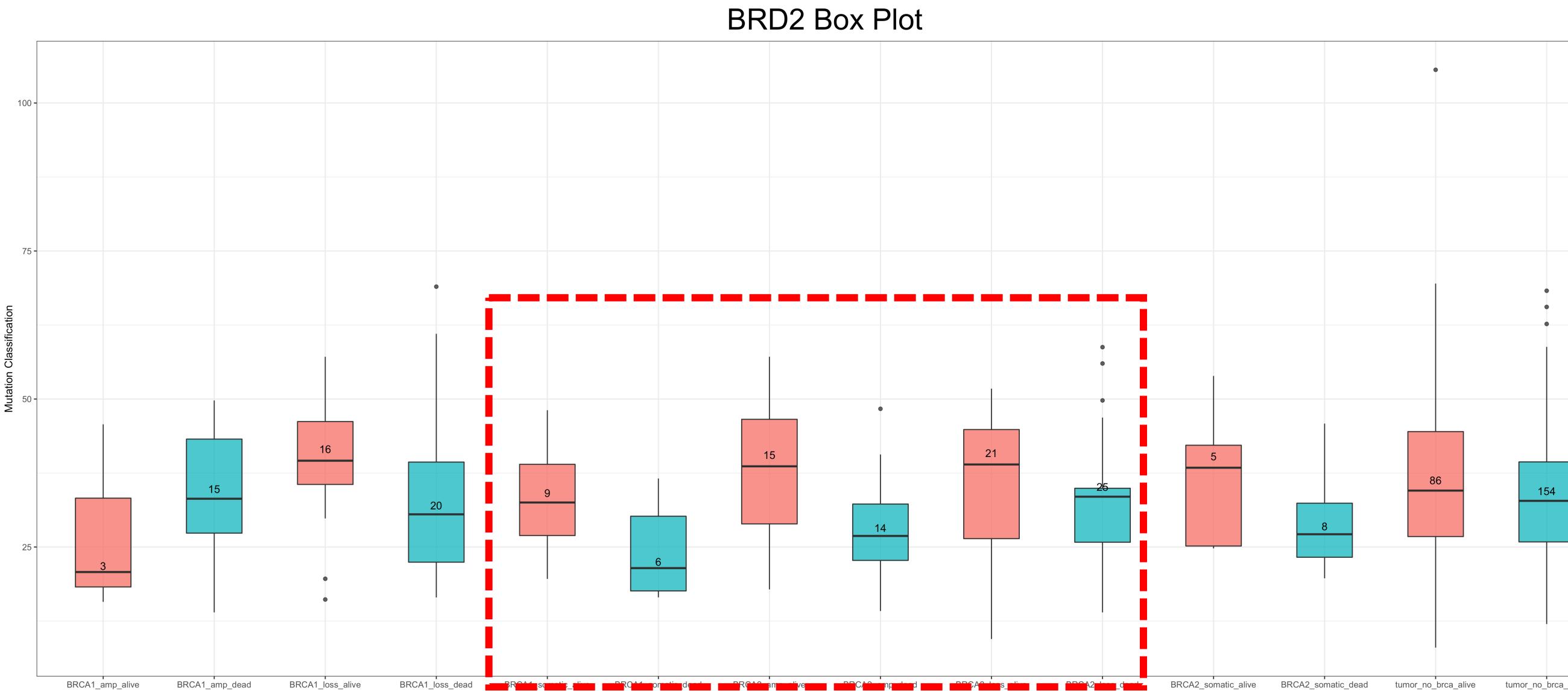
- Expression pattern is different between normal samples and tumor samples.

KAT2B Box Plot



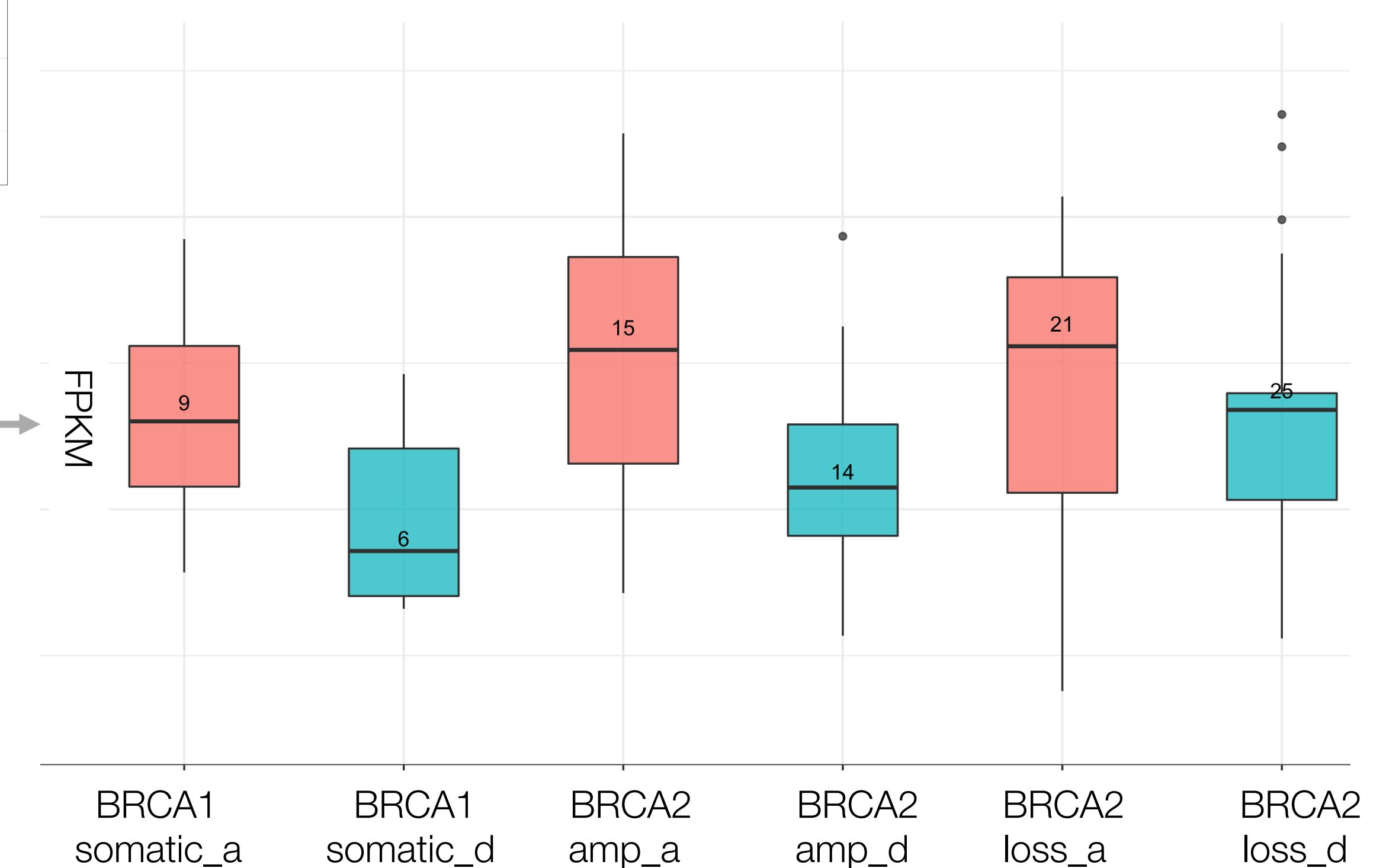
- In the breast, BRCA mutations do not affect Bromodomain gene expression a lot.

# Results: Bromodomain Gene Expression vs. Survival Status -- Ovary



- BRCA mutations might not be the reason for the Bromodomain expression differences.
- Lack samples to confirm this pattern in other cancer types.

- Expression pattern is different between alive samples and dead samples.
- Alive samples express BRD2 higher (~50) than dead samples (~25).



# Conclusion

---

- In the prostate, BRCA mutations have a correlation with the Bromodomain genes.
- However, no significant relationship in the ovary and breast cancer.
- Survival status might relate to expression of Bromodomain genes.

## Next Step:

- *Statistic method could be used to inquire expression difference in Bromodomain genes.*
- *Incorporate BRCA germline mutations into the analysis.*