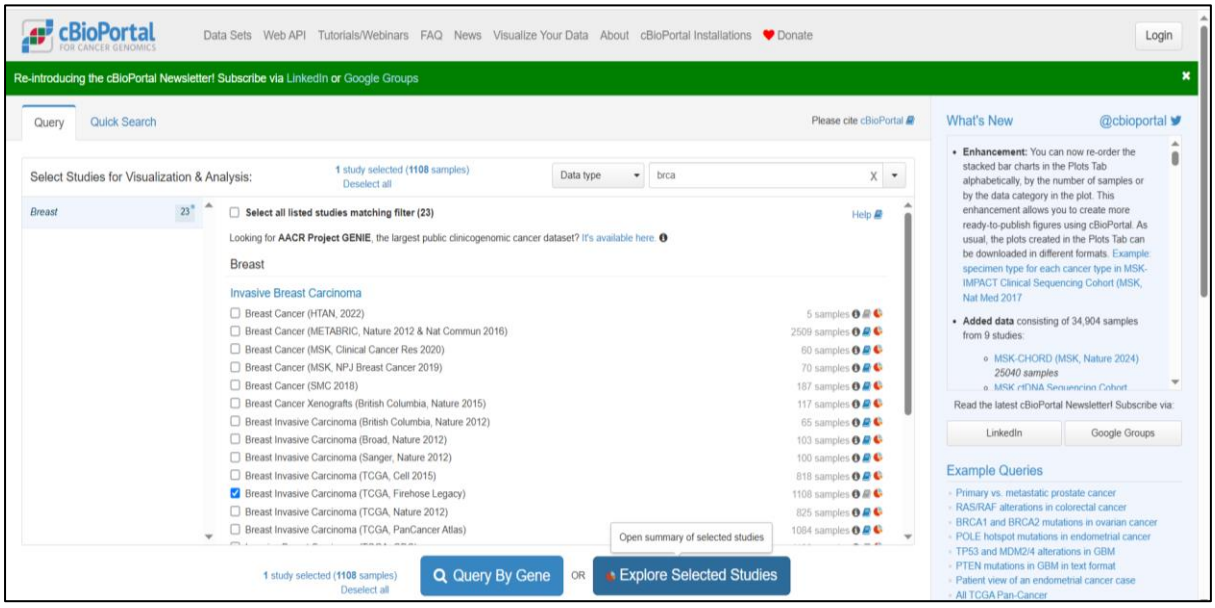
Assignment: Identify key mutations in a cancer dataset and classify them as oncogenes or tumour suppressors.

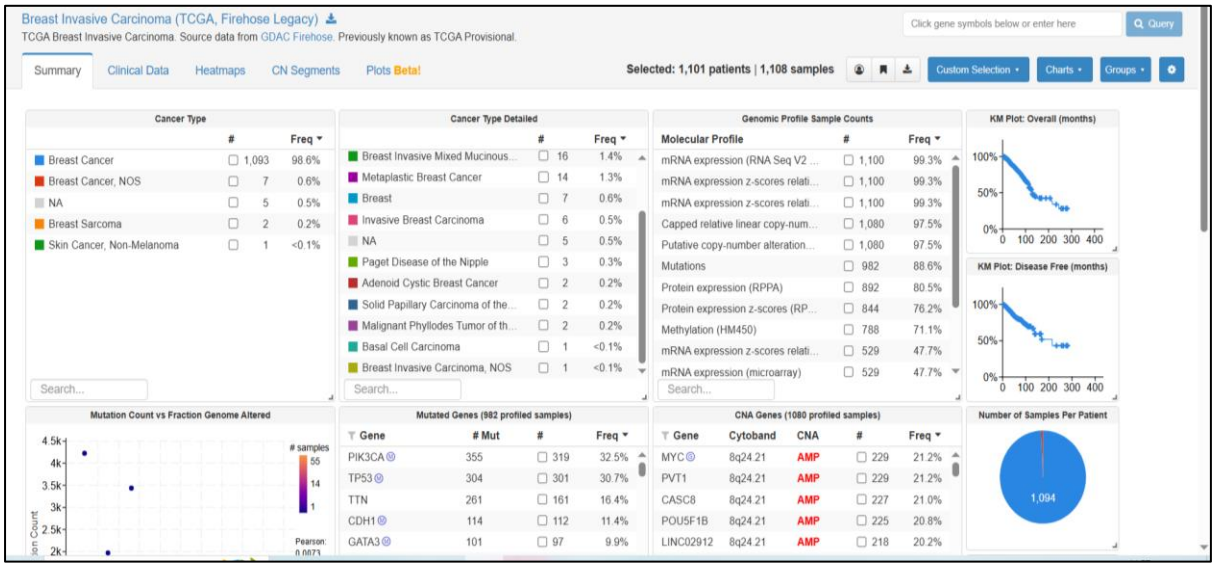
Cancer dataset chosen for assignment-**BRCA**

1. Retrieve mutation data from cBioPortal or COSMIC for a cancer type.

Database used for the assignment – cBioPortal

- BRCA was entered in search bar
- Breast Cancer information from TCGA, Firehouse Legacy was selected
- "Explore Selected Studies" option was selected





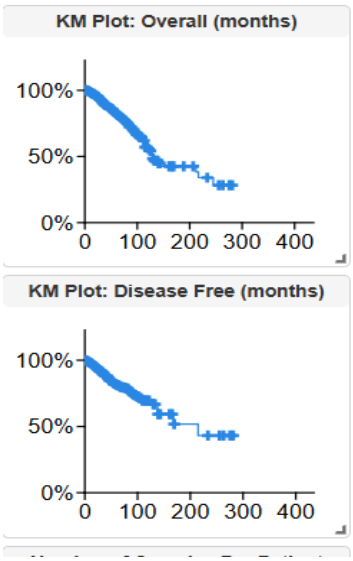
Analysing Data

1. Km Plot

*KM_Plot_Overall_(months) - Notepad

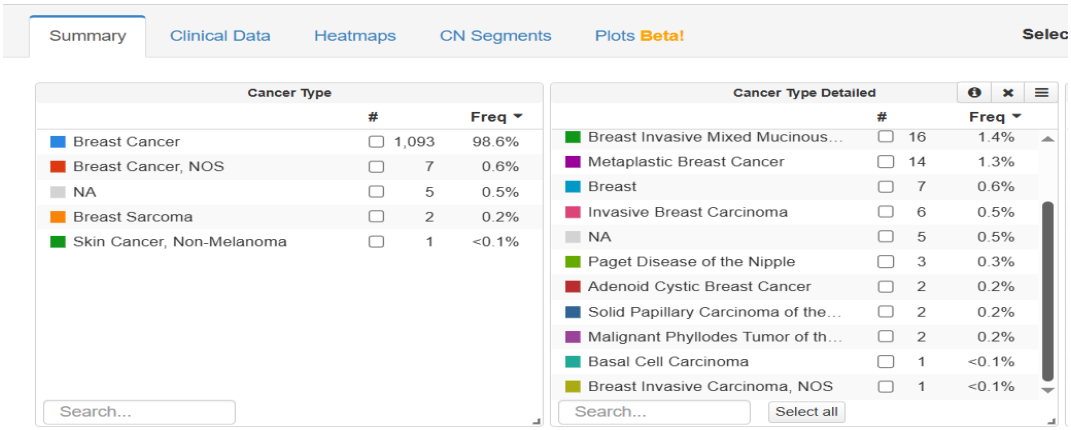
File Edit Format View Help

Study ID	Patient ID	OS_STATUS	OS_MONTHS
brca_tcga	TCGA-3C-AAAU	0:LIVING	132.95
brca_tcga	TCGA-3C-AALI	0:LIVING	131.57
brca_tcga	TCGA-3C-AALJ	0:LIVING	48.42
brca_tcga	TCGA-3C-AALK	0:LIVING	47.57
brca_tcga	TCGA-4H-AAAK	0:LIVING	11.43
brca_tcga	TCGA-5L-AAT0	0:LIVING	48.52
brca_tcga	TCGA-5L-AAT1	0:LIVING	48.32
brca_tcga	TCGA-5T-A9QA	0:LIVING	9.95
brca_tcga	TCGA-A1-A0SB	0:LIVING	8.51
brca_tcga	TCGA-A1-A0SD	0:LIVING	14.36
brca_tcga	TCGA-A1-A0SE	0:LIVING	43.4
brca_tcga	TCGA-A1-A0SF	0:LIVING	48.06
brca_tcga	TCGA-A1-A0SG	0:LIVING	14.26
brca_tcga	TCGA-A1-A0SH	0:LIVING	47.21
brca_tcga	TCGA-A1-A0SI	0:LIVING	20.86
brca_tcga	TCGA-A1-A0SJ	0:LIVING	13.67
brca_tcga	TCGA-A1-A0SM	0:LIVING	7.95
brca_tcga	TCGA-A1-A0SN	0:LIVING	39.29
brca_tcga	TCGA-A1-A0SO	0:LIVING	27.99
brca_tcga	TCGA-A1-A0SP	0:LIVING	19.19
brca_tcga	TCGA-A1-A0SQ	0:LIVING	18.2
brca_tcga	TCGA-A2-A04N	0:LIVING	143.04
brca_tcga	TCGA-A2-A04P	1:DECEASED	18
brca_tcga	TCGA-A2-A04Q	0:LIVING	78.35
brca_tcga	TCGA-A2-A04R	0:LIVING	121.85



- **100% survival rate at time = 0** which means everyone was alive at the start, then **steep drop meant** Patients started dying quickly. Only **30%** of that group **survived** to the end of the observed period.
- Patients significantly worse disease-free survival, with only 40% remaining disease-free by the end of the study period suggesting that this alteration may be associated with increased recurrence risk.

2. Cancer typed associated with BRCA



- BRCA causes breast cancer in high frequencies and specifically Breast Invasive Ductal Carcinoma (73.5% frequency) followed by Breast Invasive Lobular Carcinoma (18.7%)

3. Mutated genes

The most commonly mutated genes were PIK3CA (32%), TP53 (30%), TTN (16.4%), CDH1(11%) and GATA3 (9%),

Mutated Genes (982 profiled samples)

Gene	# Mut	#	Freq
PIK3CA	355	319	32.5%
TP53	304	301	30.7%
TTN	261	161	16.4%
CDH1	114	112	11.4%
GATA3	101	97	9.9%
MUC16	105	74	7.5%
MAP3K1	98	71	7.2%
KMT2C	83	70	7.1%
MUC12	80	54	5.5%
MUC4	62	53	5.4%
FLG	51	45	4.6%

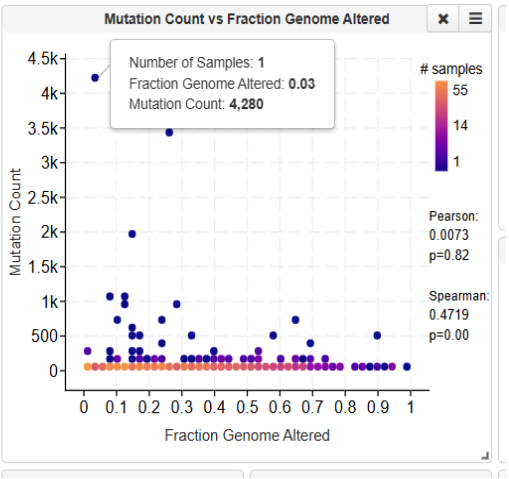
Search...

4. CNA genes and Mutation count vs Fraction genome altered

CNA Genes (1080 profiled samples)

Gene	Cytoband	CNA	#	Freq
LRATD2	8q24.21	AMP	209	19.4%
RN7SL826P	8q24.11	AMP	208	19.3%
TRIB1	8q24.13	AMP	206	19.1%
LINC00964	8q24.13	AMP	206	19.1%
WASHC5	8q24.13	AMP	206	19.1%
UTP23	8q24.11	AMP	206	19.1%
MAL2	8q24.12	AMP	205	19.0%
MTSS1	8q24.13	AMP	205	19.0%
RN7SL228P	8q24.11	AMP	205	19.0%
SAMD12	8q24.11-q...	AMP	205	19.0%
COLEC10	8q24.12	AMP	205	19.0%

Search...



In Mutation count vs Fraction genome altered

- High Mutation count and low fraction genome altered was seen in 1 sample. These tumours have a **high mutation load** but **low chromosomal instability**. Likely driven by **point mutations** rather than large-scale genomic alterations.
- Low Mutation count and increase in fraction genome altered was seen in about 55 samples indicated by pink and yellow dots. This pattern is suggestive of **chromosomally unstable tumors**, potentially driven by **copy number alterations** rather than point mutations
- Low Mutation count and high fraction genome altered was seen in 1 sample These tumours have **low mutation load** but **high chromosomal instability**. Characterized by **large-scale copy number alterations**, such as **amplifications** or **deletions**. This indicates that the tumour might be driven by **copy number-driven oncogenes** or **loss of tumour suppressors**, rather than by **point mutations**

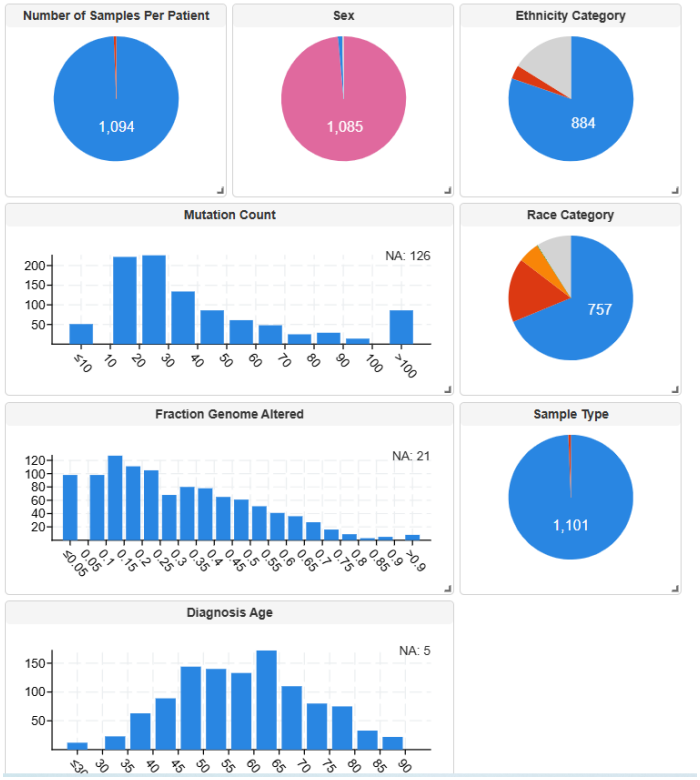
In CNA genes
AMP -Amplification **and HOMDEL-** Homologous Deletion

CNA Type	Typical Role	Example
AMP	Oncogene	**MYC, PVT1, CASC8, POU5F1B**
HOMDEL	Tumour suppressor	**LINC01193, CHST6, RN7SL515P**

AMP- Often seen in **oncogenes**. Amplification can **drive cancer** by overactivating growth-promoting pathways.

HOMODEL- Typically affects **tumour suppressor genes**. Loss of these genes removes the cell's natural defence against uncontrolled growth or DNA damage.

5. Other information

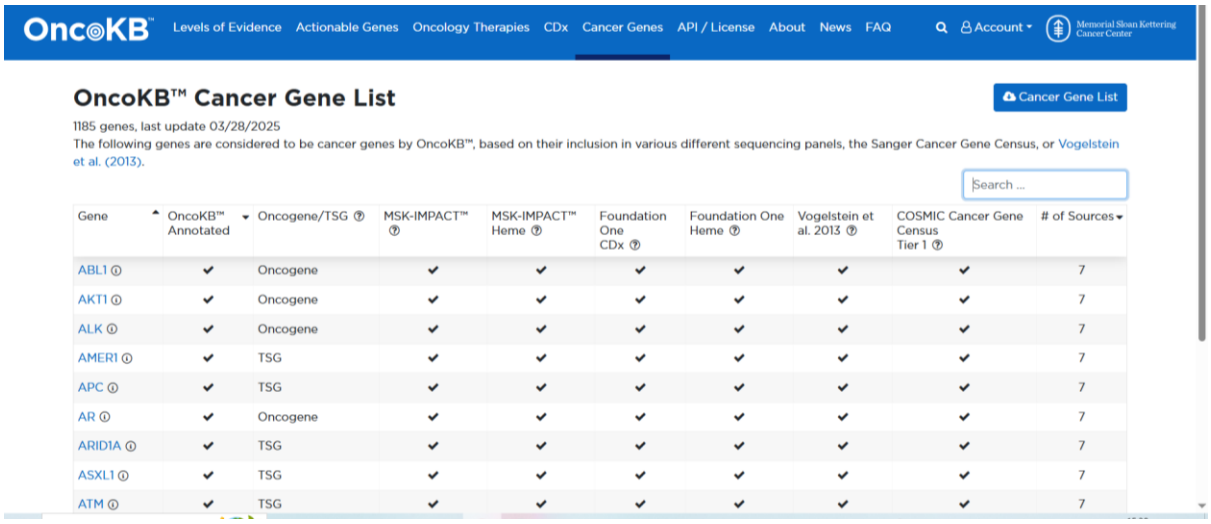


Number of Samples Per Patient- 99.4% patients gave one sample
Sex- 98.5% seen in females, 1.1% in males
Ethnicity- maximum % seen in NOT HISPANIC OR LATINO (80%)
Sample type- 99.4% have sample taken from original tumour
Race- White (68.8%), Africans (16%)
Diagnosis Age- Maximum cases in age range of 60 to 65

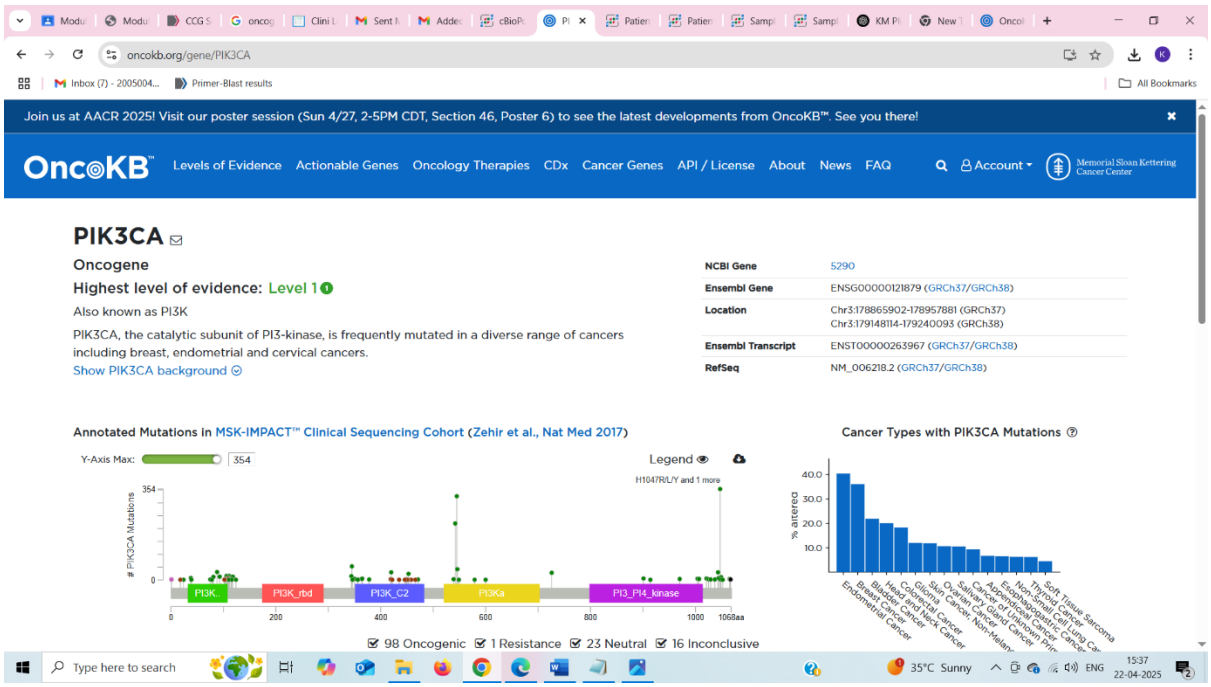
2. Use OncoKB to classify genes as oncogenes or tumour suppressors.

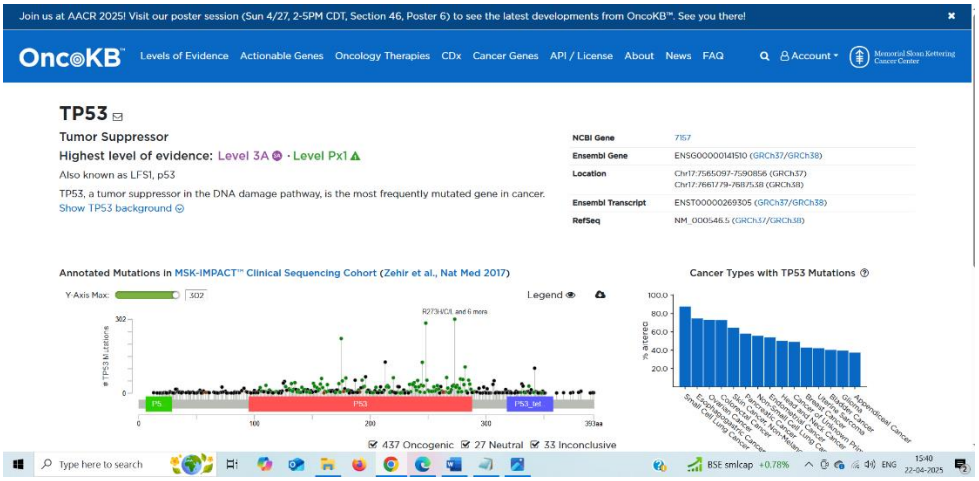
Genes with High frequency of mutation from Cbioportal was obtained-

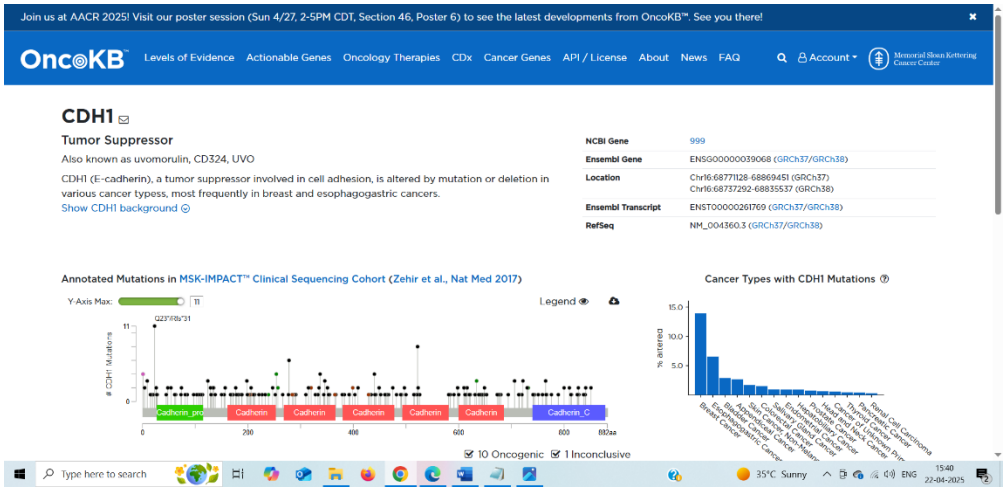
PIK3CA, TP53, CDH1, GATA3, MAP3K1



From OncoKB gene list the genes can be searched and whether they are tumour suppressing or oncogenic can be found out. First PIK3CA was searched and OncoKB classified it as an Oncogene, Similarly information about other genes were also obtained







Sr.NO	Gene	Oncogene/Tumour Suppressor
1	PIK3CA	Oncogene
2	TP53	Tumour Suppressor
3	CDH1	Tumour Suppressor
4	GATA3	Tumour Suppressor
5	MAP3K1	Tumour Suppressor
6	NCOR1	Tumour Suppressor
7	PTEN	Tumour Suppressor

3. Analyse mutational hotspots in IGV (Integrative Genomics Viewer).

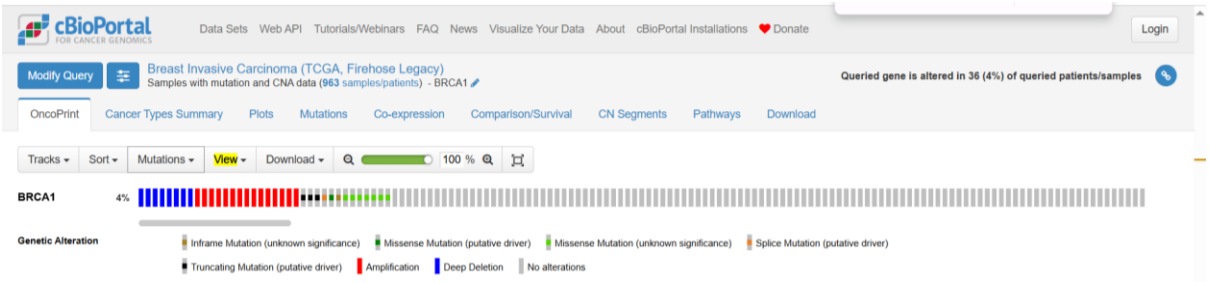
Mutational hotspots are **regions in a gene** (often specific codons or exons) that are **frequently mutated across many samples**, usually because they affect the gene's function in cancer.

Utilized IGV (Integrative Genomics Viewer) of cBioportal to analyze copy number variation (CNV) and identify mutation hotspots using segmented CNV data obtained from cBioPortal.

Genes	Protein Change	Colour represented in IGV	Type of mutation
GATA3	***X308_splice***	Yellow	**Splice**
CDH1	***A575Gfs*13***	Black	**FS ins**
AR	***L881Q***	Green	**Missense**
CHEK2	***D77H***	Green	**Missense**
DPYD	***S260R***	Green	**Missense**
USP8	***R763W***	Green	**Missense**
USP8	***N764K***	Green	**Missense**
DBH	***R329C***	Green	**Missense**
CACNA1C	***A2055T***	Green	**Missense**

The data was loaded into IGV, where blue and red color tracks represent deletions and amplifications respectively. IGV's human ref seq and CNV tracks were used to locate and interpret mutation hotspots based on their log2 ratio values:

- **Log2 ratio > 0**: Copy number gain/amplification (**red**)
- **Log2 ratio < 0**: Copy number loss/deletion (**blue**)
- Normal copy number regions appeared near the baseline (log2 ≈ 0)



The presence of **dense vertical red lines** and consistent red CNV segments in IGV across multiple samples highlights **potential amplification hotspots**. Conversely, **clusters of blue segments** indicate **deletion hotspots**.

These CNV-driven mutations, particularly in genes like **MLL3, CDKN2A, and PHKG1**, may represent **driver events** in tumorigenesis. For instance:

- **MLL3** amplifications could dysregulate chromatin and transcription.
- **CDKN2A**, while typically deleted in cancers, showed copy gain here, possibly reflecting complex alterations.
- **PHKG1** deletion may disrupt glycogen metabolism in tumor cells.

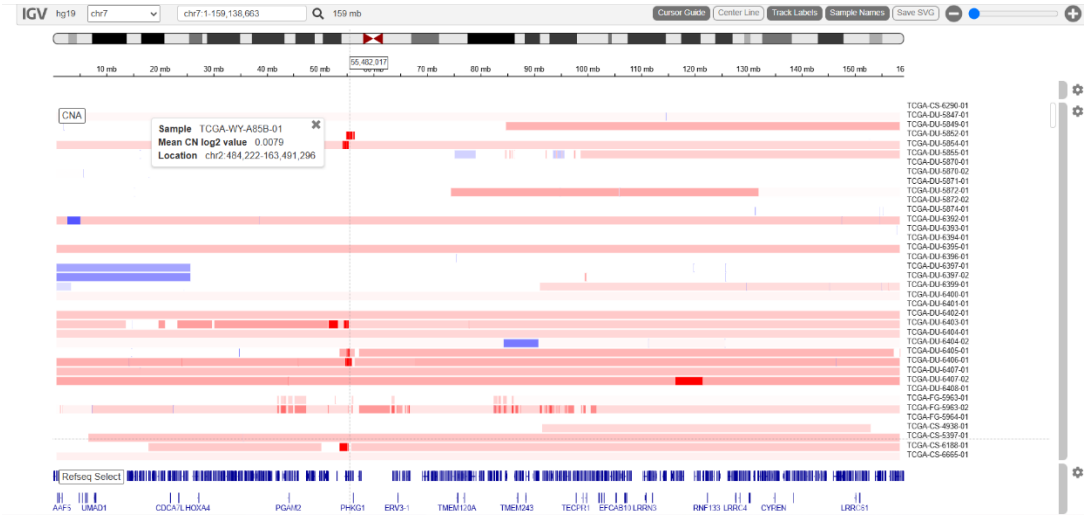


Figure showing Amplification Hotspots

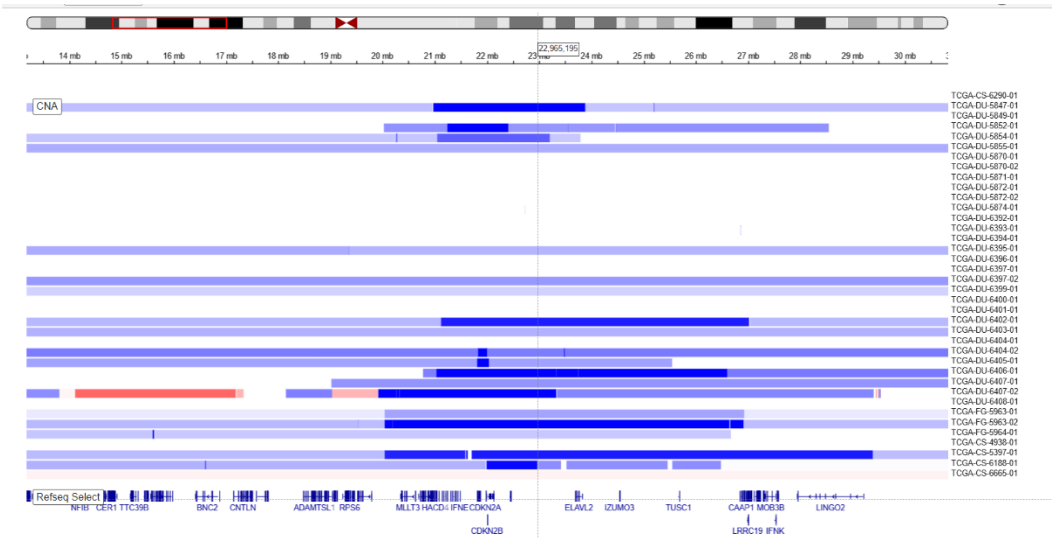


Figure showing Deletion Hotspots

Red Regions

Region	Mutation Hotspots	Mean CN Log2value	Interpretation
chr7:53,604,469–55,109,914	PHKG1	2.54	High Amplification in PHKG1
chr7:86,456,237–87,008,509	TMEM243	0.3198	Slight Amplification in TMEM243
chr7:92,185,806–92,463,876	TECPR1	0.77	Moderate Amplification in TECPR1
chr4:53,918,177–54,948,410	LNX1	1.6355	High Amplification in LNX1

Blue Regions

Mutation Hotspots	Location	Log2 values	Interpretation
MLLT3	**chr9:19,529,532-20,039,441 to chr9:21,702,086-29,383,583**	-3.28	Strong homozygous deletion
HACD4		-1.5569	Deep deletion
IFNE		-2.02	Homozygous deletion
CDKN2B		-3.28	Strong homozygous deletion
ELAVL2		-4.4531	Very Deep Deletion
IZUMO3		-3.37	Strong homozygous deletion
TUSC1		-0.827	Shallow deletion
CAAP1		-0.8028	Shallow deletion
MOB2B		-4.606	Very Deep Deletion
LINGO2		-0.76	Shallow deletion

Other Blue Regions-

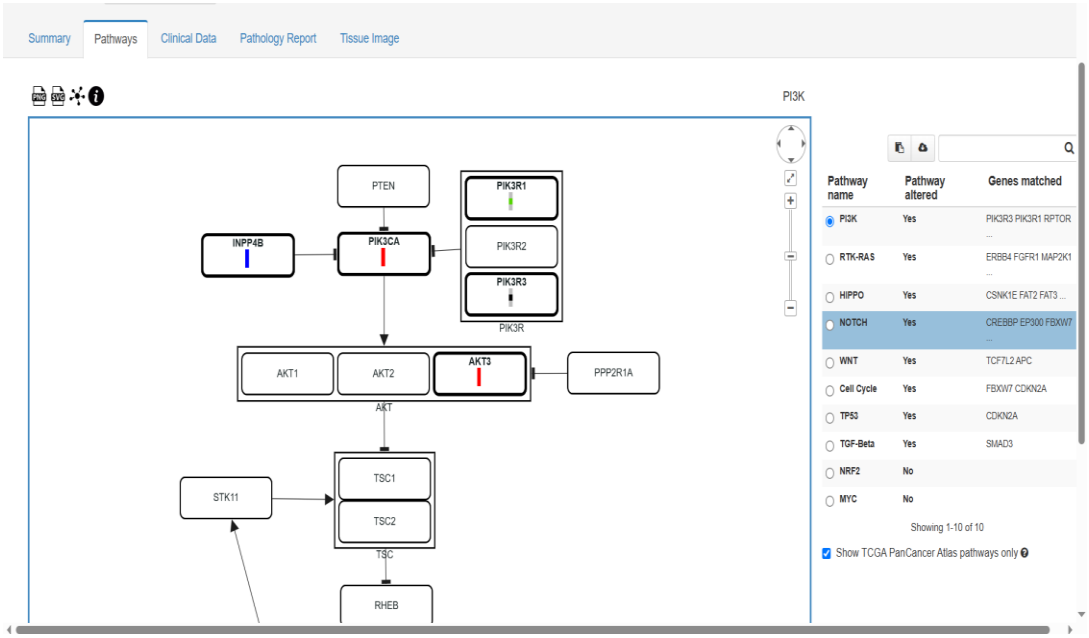
chr1:81,048,805-81,338,857 - GNG5 Amplification- Moderate Deletion

chr10:22,973,855-26,115,303 – AB1 Amplification – Moderate Deletion

4. Discuss how mutations in DNA repair genes contribute to tumour progression.

- DNA repair genes (like **BRCA1**, **BRCA2**, **ATM**, **ATR**, **MLH1**, **MSH2**, **RAD51**, etc.) are responsible for fixing errors or damage in the DNA. When these genes are mutated, cells accumulate more genetic errors over time, which can lead to **genomic instability** and cancer.

  There is pathway alteration in the following - PI3K,WNT,NOTCH,HIPPO,PI3K,MYC CELL CYCLE



Pathway	Function	What Alterations Imply in Cancer
PI3K/AKT/mTOR	Controls cell survival, growth, and metabolism	Alterations often lead to enhanced cell proliferation and resistance to cell death
WNT	Regulates cell fate, stem cell renewal, and differentiation	Dysregulation promotes uncontrolled growth and maintenance of cancer stem cells
Notch	Involved in cell communication and differentiation	Abnormal activity can result in unregulated cell proliferation or impaired differentiation
Hippo	Controls organ size, cell contact inhibition, and apoptosis	Alterations lead to tissue overgrowth and contribute to tumor formation
MYC	Regulates cell cycle, metabolism, and DNA replication	Overactivation supports aggressive growth and metabolic reprogramming in tumors
Cell Cycle	Governs cell division and checkpoints	Alterations cause loss of control over cell division and genomic instability