BIOINFORMATICS ASSIGNMENT 1 (Day 1 - 5)

Rasitha Arafa R

rasithaarafarafiudeen@gmail.com

1. Gene Name: Cytochrome b5 type B (CYB5B) [*Homo sapiens*]

2. Function of the Gene
   - Enables heme binding activity.
   - Contributes to nitrite reductase activity.
   - Involved in nitric oxide biosynthetic process.
   - Located in membrane of mitochondria.
   - Part of nitric-oxide synthase complex.

3. NCBI accession number: NC_000016.10

4. Forward Primer: GTCAGCGGTTTGTCTTGTGT
5. Reverse primer: ACAGGGCATAGCTTTGGAGT

6. Features of primers
   - Left primer (Forward):   Length – 20 bases, GC% - 50%, Tm – 58.99 degree Celsius
   - Right primer (Reverse): Length – 20 bases, GC% - 50%, Tm – 59 degrees Celsius

7. Amplicon length: 250 bp
8. Amplicon sequence:

   TTTATATTTGTGTTTACTCAAGTAGGACTGCTTTTTGAAACATTTTTTCTTAACAAGAGAAGTTAC
   AAAGTATTTACTTTTTCCCCAAGCAAAAATCCTATTTTTCTGGAATTTGGACTCAGTATCATCTCA
   GGAATAAAAGAATAGCTGAGTCTTGAACAGTAGGAAACATTTTGCTAATGCCTTTATACGCTTTT
   TTTTTTTAACTGAAACTCCAAAGCTATGCCCTGT

qPCR Data analysis (DAY 5)

The following data are results of qPCR from cancer cell lines. HER2 stands for human epidermal growth factor. It's healthy in normal amounts, but too much may be a sign of a certain type of breast cancer. Calculate the 2 Delta Ct values for the following data.

Housekeeping genes(GAPDH)	Ct values		Gene of interest (HER2)	Ct values	
	Ct 1	Ct 2		Ct 1	Ct 1
Untreated (control)	18.5	18.5	Untreated(control)	23.3	22.5
Untreated (control)	17.8	17.8	Untreated(control)	22.5	22.2
Untreated (control)	17.5	17.5	Untreated(control)	21.2	21.9
Treated	18.3	18.3	Treated	25.3	25.3
Treated	18.5	18.5	Treated	26.5	26.5
Treated	18.2	18.2	Treated	27.5	27.5

SOLUTION:

	House Keeping Genes(GAPHD)		Gene of Interest (HER2)		Average ct value of HG	Average ct value of GOI	Δct value	ΔΔct value	Fold Change
	ct1	ct2	ct1	ct2					
Untreated (control) 1	18.5	18.5	23.3	22.5	18.5	22.9	4.4	0	1
Untreated (control) 2	17.8	17.8	22.5	22.2	17.8	22.35	4.55	0	1
Untreated (control) 3	17.5	17.5	21.2	21.9	17.5	21.55	4.05	0	1
Treated 1	18.3	18.3	25.3	25.3	18.3	25.3	7	2.6	0.16493849
Treated 2	18.5	18.5	26.5	26.5	18.5	26.5	8	3.45	0.09150536
Treated 3	18.2	18.2	27.5	27.5	18.2	27.5	9.3	5.25	0.02627801

NCBI – Primer BLAST Results

Primer-BLAST » JOB ID:QEqf0sJaz_LozNXJ2Knx-6Ky4MmPofvUjg

Primer-BLAST Results ❓

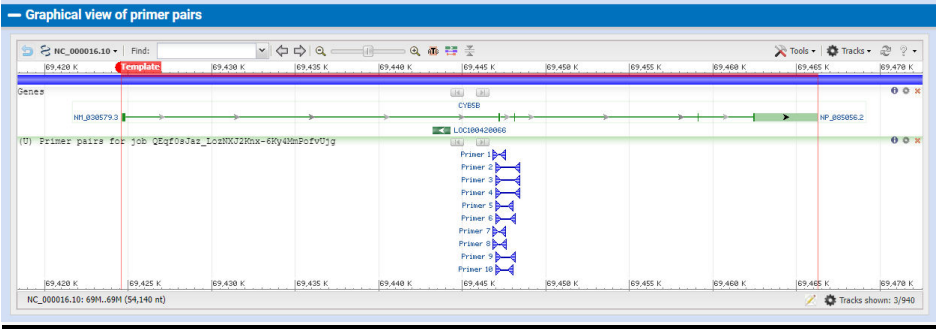| Input PCR template	NC_000016.10 Homo sapiens chromosome 16, GRCh38.p14 Primary Assembly
Range	69424619 - 69466264
Specificity of primers	Primer pairs are specific to input template as no other targets were found in selected database: Refseq mRNA (Organism limited to Homo sapiens)
Other reports	▷Search Summary

— Graphical view of primer pairs

Detailed primer reports

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTGTAGCACCCTGGAGGAGA	Plus	20	69447144	69447163	59.88	55.00	4.00	0.00
Reverse primer	TCTGCATAAGGGCTCCCATC	Minus	20	69447313	69447294	59.23	55.00	4.00	2.00
Product length	170								

Primer pair 2

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGATGGGAGCCCTTATGCAG	Plus	20	69447293	69447312	59.23	55.00	4.00	2.00
Reverse primer	CCGCTGACTTCTTACCTTGC	Minus	20	69448159	69448140	58.92	55.00	2.00	2.00
Product length	867								

Primer pair 3

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGAGATGGGAGCCCTTATGC	Plus	20	69447291	69447310	58.94	55.00	4.00	2.00
Reverse primer	CCGCTGACTTCTTACCTTGCT	Minus	21	69448159	69448139	60.34	52.38	2.00	0.00
Product length	869								

Primer pair 4

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GATGGGAGCCCTTATGCAGAG	Plus	21	69447294	69447314	60.27	57.14	4.00	2.00
Reverse primer	ACACAAGACAAACCGCTGACT	Minus	21	69448171	69448151	60.41	47.62	3.00	1.00
Product length	878								

Primer pair 5

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CAGAGATGGGAGCCCTTATGC	Plus	21	69447290	69447310	60.27	57.14	4.00	2.00
Reverse primer	GCTACCTGGCAAAGAAGAAGC	Minus	21	69447761	69447741	59.53	52.38	4.00	2.00
Product length	472								

Primer pair 6

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ACTATCAGAGATGGGAGCCCT	Plus	21	69447285	69447305	59.49	52.38	4.00	2.00
Reverse primer	AGGAGCTTGCAAAAACAAGCA	Minus	21	69447888	69447868	59.52	42.86	6.00	2.00
Product length	604								

Primer pair 7

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCTTGTAGCACCCTGGAGGA	Plus	20	69447142	69447161	60.91	60.00	4.00	2.00
Reverse primer	AAGGGCTCCCATCTCTGATAGT	Minus	22	69447306	69447285	60.09	50.00	4.00	1.00
Product length	165								

Primer pair 8

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTCCTTGTAGCACCCTGGAG	Plus	20	69447140	69447159	59.02	55.00	3.00	2.00
Reverse primer	TCCTGTTCTGCAGCTGTTAAGT	Minus	22	69447342	69447321	59.90	45.45	6.00	1.00
Product length	203								

Primer pair 9

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ATGGGAGCCCTTATGCAGAGA	Plus	21	69447295	69447315	60.70	52.38	4.00	2.00
Reverse primer	AGCAAGGAGCTTGCAAAAACA	Minus	21	69447892	69447872	59.52	42.86	7.00	0.00
Product length	598								

Primer pair 10

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGGGAGCCCTTATGCAGAGAA	Plus	21	69447296	69447316	60.91	52.38	4.00	1.00
Reverse primer	ACATGTGCTACCTGGCAAAG	Minus	20	69447767	69447748	58.46	50.00	6.00	2.00
Product length	472								

Primer 3 Results

```
PRIMER PICKING RESULTS FOR NC_000016.10:69424619-69466264 Homo sapiens chromosome 16, GRCh38.p14 Primary Assembly

Template masking not selected
No mispriming library specified
Using 1-based sequence positions
OLIGO           start _len _____ tm ____gc% _any_th _3'_th hairpin seq
LEFT PRIMER     23534  20   58.99  50.00    0.00   0.00    0.00 GTCAGCGGTTTGTCTTGTGT
RIGHT PRIMER    23783  20   59.00  50.00    6.29   0.00    0.00 ACAGGGCATAGCTTTGGAGT
SEQUENCE SIZE: 41646
INCLUDED REGION SIZE: 41646

PRODUCT SIZE: 250, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00
```

```
ADDITIONAL OLIGOS
                start _len _____ tm ____gc% _any_th _3'_th hairpin seq

 1 LEFT PRIMER   37842  20   59.02  50.00    7.64   0.00    0.00 AGGCGCTGTTCTCTTAGGTT
   RIGHT PRIMER  38049  20   59.01  50.00    7.01   0.00    0.00 TGAGGTGATGTCTGGCCAAT
   PRODUCT SIZE: 208, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00

 2 LEFT PRIMER    7324  20   59.02  55.00    0.00   0.00    0.00 CTGGGAAAGCCTCTGTGAGA
   RIGHT PRIMER   7504  20   59.01  50.00    5.51   0.00    0.00 AGCTCTCTGGCCATTCTTGT
   PRODUCT SIZE: 181, PAIR ANY_TH COMPL: 11.70, PAIR 3'_TH COMPL: 6.38

 3 LEFT PRIMER   38120  20   59.04  50.00    5.05   0.00    0.00 ACAATTGCCGGTGTTTCCTC
   RIGHT PRIMER  38358  20   59.01  55.00    0.00   0.00    0.00 CCACCACACCTAGCCCTAAA
   PRODUCT SIZE: 239, PAIR ANY_TH COMPL: 2.92, PAIR 3'_TH COMPL: 0.00

 4 LEFT PRIMER   22374  20   59.04  50.00    0.00   0.00    0.00 GCTGTGTTCTTCCCATTGCA
   RIGHT PRIMER  22541  20   59.02  55.00    0.00   0.00    0.00 CTCCAGGGTGCTACAAGGAA
   PRODUCT SIZE: 168, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00
```

BIOINFORMATICS ASSIGNMENT 1 (Day 6 - 10)

Rasitha Arafa R

rasithaarafarafiudeen@gmail.com

NGS DATA QUALITY CHECK (DAY 6)

1. SRA accession number: **SRR24518778**
2. NGS platform: **ILLUMINA sequencing**
3. Layout: **Pair-end sequencing**

4. Basic statistics:

Measure	Value
Filename	SRR24518778_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3516206
Sequences flagged as poor quality	0
Sequence length	35-151
%GC	38

•**This basic statistic report gives the overall outline of the sequencing strategy and inferences from the input.**
•**The given input data was sequenced through illumine sequencing which has a total sequence of 3516206 but the average sequence length is referred as 35-151 with a GC content of 38%**

5. Per Base sequence quality:
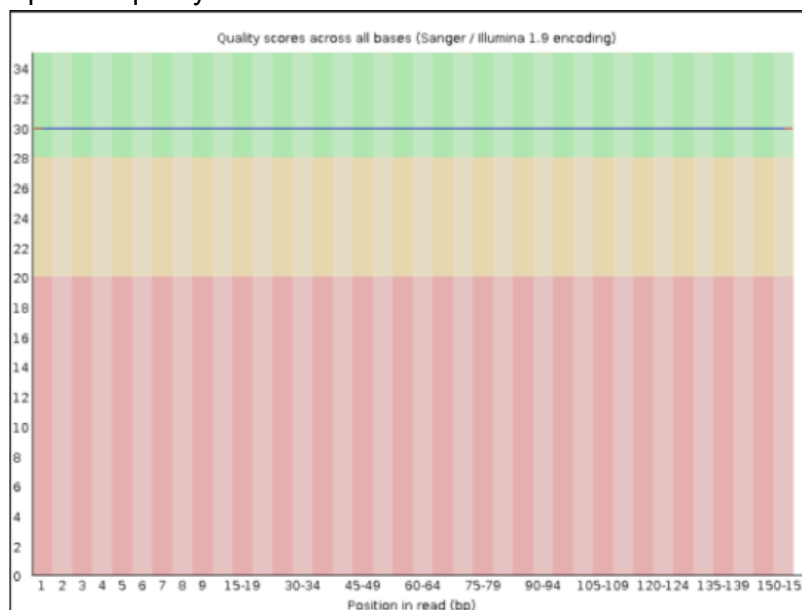
- **This defines the quality of each base read.**
- **The mean quality, represented by the blue line**
- **The y-axis shows the quality scores.**

- **The x-axis is the base position.**
- **This sample contains reads up to 151 bp long.**
- **The higher the score, the better the base call.**
- **The background of the graph divides the y-axis into very good quality scores (green), scores of reasonable qualities (orange), and reads of poor quality (red).**
- **In this all the reads are come out in the green region showing greater quality.**

6. Per sequence quality score:
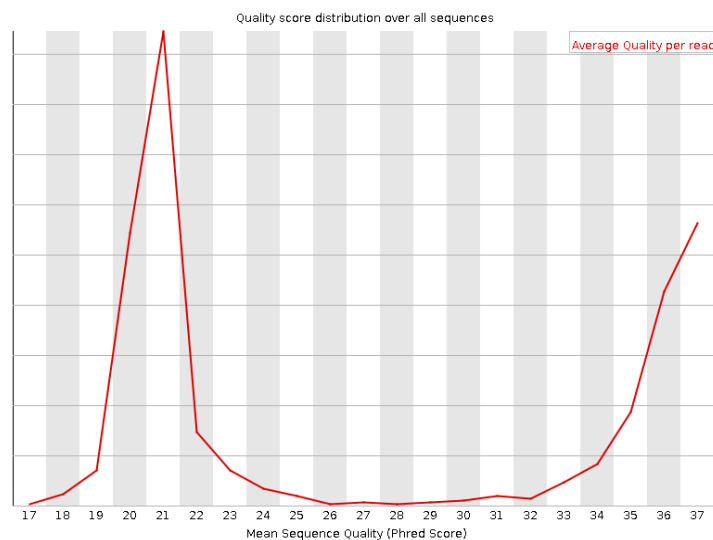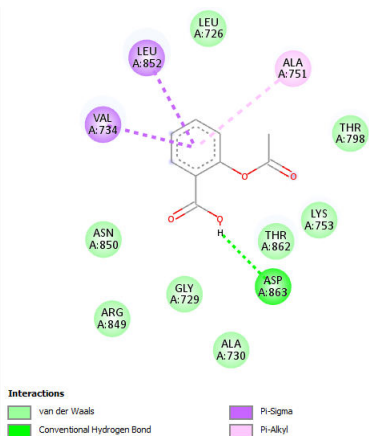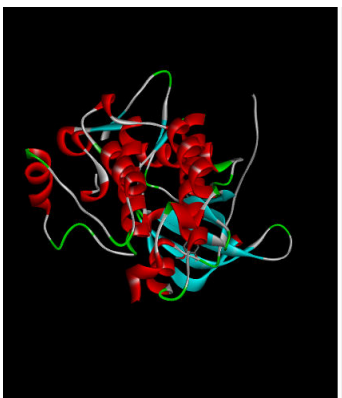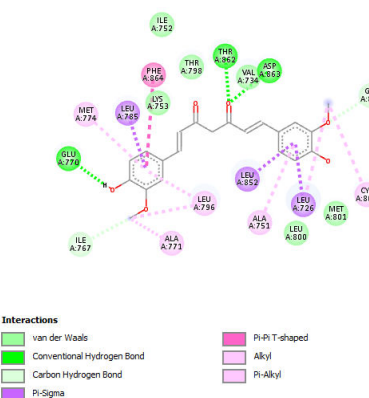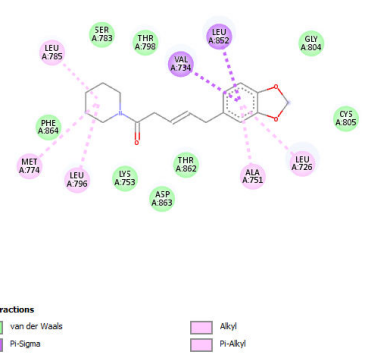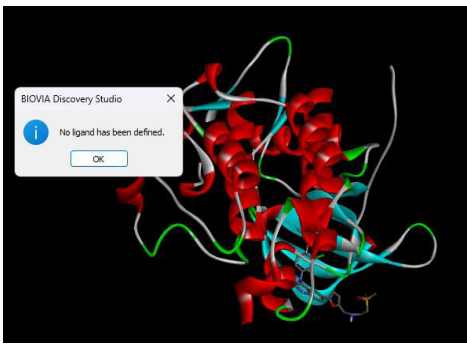
- **This defines the quality of the sequence reads.**
- **The phred score is calculated on the basis of errors per 1000 sequence reads.**
- **It is fine to have 1 error with 1000 base reads to be considered as good quality.**
- **The phred score should be less than 30 always.**
- **The x-axis is the average quality score over the full length of all reads.**
- **The y-axis gives the total number of reads with this score.**
- **The read quality is peak in the upper range of the plot.**
- **Here the graph shows higher peak at 151 range on y axis around phred score of 21, which signifies good overall quality but in the range of 70 – 100 there is a phred score increase but that is not significant.**
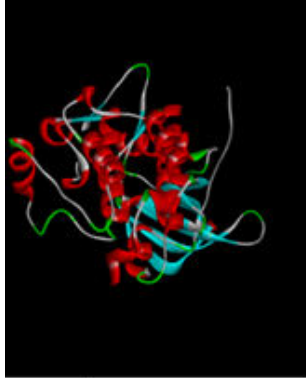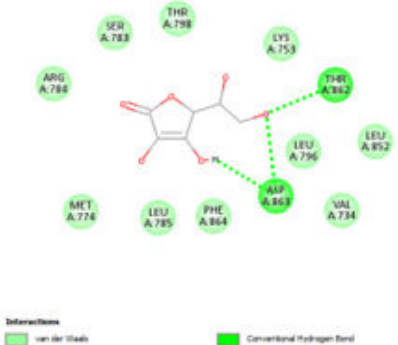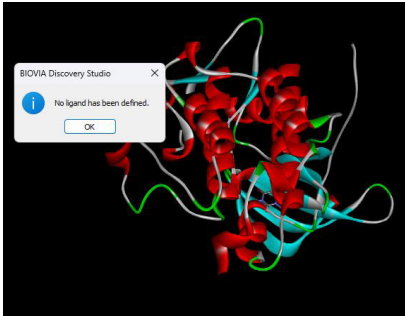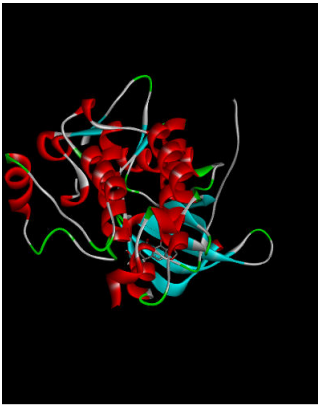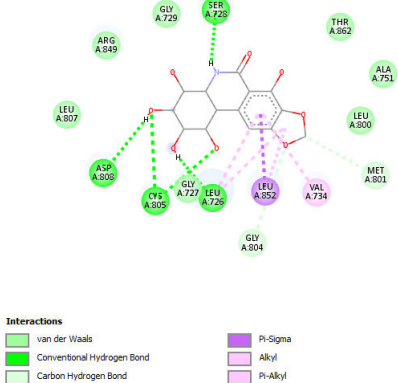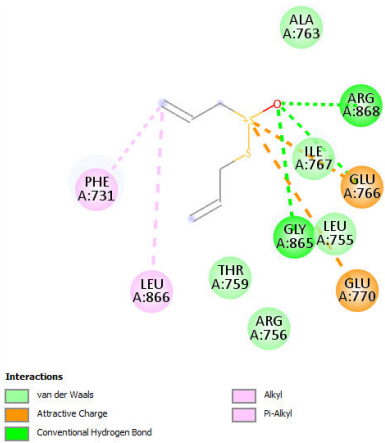
GitHub (DAY 7)

7. GitHub account link - **https://github.com/RasithaArafaRafiudeen**

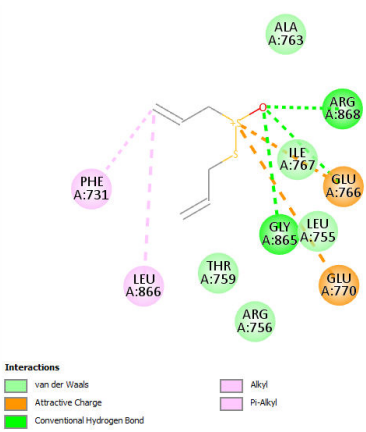## Molecular Docking (DAY 8 and 9)

8. Protein Name: **Crystal structure of the Kinase domain of Human HER2 (erbB2)**
9. Protein ID – **3PP0**
10. PDB – DOI: **https://doi.org/10.2210/pdb3PP0/pdb**

Ligand Name	Ligand ID	Energy value	Dock Image - 2D
Aspirin	2244	-6.7	
Curcumin	638024	-9.6	
Piperine	969516	-10	
Lapatinib	969516	-8.2	**No ligand found**

Ascorbic acid	54670067	-5.8	
Caffeine	2519	-6.1	**No ligand found**
Pancratistatin	441597	-7.5	
Allicin	65036	-4.1	

Ajoene	5386591	-4.1	

Cancer therapy (DAY 10)

Cancer type	Hallmarks	Drug	Mechanism of drug
Mesothelioma	Cancer affecting the lining of the lungs, abdomen, or heart	Pemetrexed and Cisplatin	Pemetrexed inhibits multiple enzymes involved in DNA and RNA synthesis, disrupting cancer cell growth. Cisplatin cross-links DNA and induces apoptosis
Cholangiocarcinoma (Bile Duct Cancer)	Uncontrolled cell growth in the bile ducts	Gemcitabine and Cisplatin	Gemcitabine inhibits DNA synthesis, preventing cell division, while cisplatin cross-links DNA, leading to DNA damage and cell death
Breast Cancer	Uncontrolled growth of breast cells	Tamoxifen	It is a selective estrogen receptor modulator (SERM) that blocks the estrogen receptors in breast cells which inhibits the growth-promoting effects of estrogen and helps prevent estrogen-driven breast cancer growth

Colorectal Cancer	Uncontrolled growth of cells in the colon or rectum	Oxaliplatin, Fluorouracil (5-FU), and Leucovorin	Oxaliplatin is a platinum-based chemotherapy drug that induces DNA damage, preventing cell division. Fluorouracil (5-FU) is an antimetabolite that interferes with DNA and RNA synthesis, while Leucovorin enhances the efficacy of 5-FU
GIST (Gastrointestinal Stromal Tumors)	Abnormal growth of specialized cells in the gastrointestinal tract	Imatinib	It is a tyrosine kinase inhibitor that targets mutated forms of the KIT protein, which is often present in GIST. By inhibiting KIT, imatinib prevents cell growth and promotes apoptosis
Merkel Cell Carcinoma	Aggressive skin cancer originating from Merkel cells	Avelumab.	It is a monoclonal antibody that targets and inhibits the programmed cell death-ligand 1 (PD-L1) on cancer cells. By blocking PD-L1, avelumab enhances the immune system's ability to recognize and eliminate cancer cells
Prostate Cancer	Cancer originating from the prostate gland	Enzalutamide	Enzalutamide is an androgen receptor antagonist. It blocks the binding of androgens (such as testosterone) to the androgen receptor, inhibiting the signalling pathways that promote prostate cancer cell growth