AI & Biotechnology Bioinformatics

Computational Approaches in Drug Discovery Course

Name : Abeera Iftikhar

Assignment 1 : Comprehensive Analysis of 20 Breast Cancer-Associated Genes and Their Structural Characterization

Introduction:

Breast cancer is a malignant tumor originating in breast tissue, commonly in the milk ducts (ductal carcinoma) or lobules (lobular carcinoma). It is the most common cancer in women worldwide and a **leading cause of cancer-related death**. Risk factors include female gender, age, family history, genetic mutations **(BRCA1/2)**, hormonal factors, and lifestyle choices such as obesity, alcohol, and smoking (WHO, 2023).

Early detection through screening methods like mammography is crucial, as early-stage breast cancer may be asymptomatic. Common symptoms include lumps, nipple discharge, or changes in breast shape or skin. Treatment typically involves a combination of surgery, radiation, chemotherapy, hormonal therapy, and targeted therapies, depending on tumor type and stage (Britannica, 2023).

Globally, while high-income countries report higher incidence due to lifestyle and screening, mortality remains higher in low-resource areas due to limited access to diagnosis and treatment (WHO, 2023).

Structural Profiling of Breast Cancer-Associated Proteins:

For this study, 20 genes and their corresponding proteins associated with breast cancer were selected from **UniProt** based on their documented roles in tumor development, progression, or as biomarkers. Each protein was then checked for experimentally resolved 3D structures in the **Protein Data Bank (PDB)** using gene symbols, protein names, or UniProt IDs. The availability of structures was recorded, including the PDB ID **and the method of determination (X-ray, NMR, or Cryo-EM)**. Proteins without experimentally resolved structures were noted as “Not Available”. This approach ensures a clear mapping between breast cancer-related genes, their protein products, and structural information crucial for downstream structural and functional analyses.

Breast Cancer–Associated Genes/Proteins and PDB Structure Availability

No.	Gene Symbol	Protein Name	UniProt ID	PDB Structure	PDB ID	Structure Method
1	BRCA1	Breast cancer type 1 susceptibility protein	P38398	Yes	1JM7	NMR
2	BRCA2	Breast cancer type 2 susceptibility protein	P51587	Yes	1N0W	X-RAY
3	TP53	Cellular tumor antigen p53	P04637	Yes	1A1U	NMR
4	HER2 (ERBB2)	Receptor tyrosine-protein kinase erbB-2	P04626	Yes	1MFG	X-RAY
5	EGFR	Epidermal growth factor receptor	P00533	Not Available	No	No
6	PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	P42336	Yes	2ENQ	NMR
7	PTEN	Phosphatase and tensin homolog	P60484	Yes	1D5R	X-RAY
8	AKT1	RAC-alpha serine/threonine-protein kinase	P31749	Yes	1H10	X-RAY
9	ESR1	Estrogen receptor alpha	P03372	Yes	1A52	X-RAY
10	ESR2	Estrogen receptor beta	Q92731	Yes	1L2J	X-RAY
11	CDH1	E-cadherin	P12830	Yes	1o6s	X-RAY
12	MYC	Myc proto-oncogene protein	P01106	Yes	1A93	NMR
13	CCND1	Cyclin D1	P24385	Yes	2W96	X-RAY
14	RB1	Retinoblastoma-associated protein	P06400	Yes	1AD6	X-RAY
15	MAPK1	Mitogen-activated protein kinase 1 (ERK2)	P28482	Yes	1PME	X-RAY
16	MAPK3	Mitogen-activated protein kinase 3 (ERK1)	P27361	Yes	2ZOQ	X-RAY
17	BCL2	Apoptosis regulator Bcl-2	P10415	Yes	1G5M	NMR
18	MMP9	Matrix metalloproteinase-9	P14780	Yes	1GKC	X-RAY
19	VEGFA	Vascular endothelial growth factor A	P15692	Yes	1BJ1	X-RAY
20	FOXA1	Forkhead box protein A1	P55317	Yes	7VOX	X-RAY

Refernce links

https://www.who.int/news-room/fact-sheets/detail/breast-cancer
https://www.britannica.com/science/breast-cancer
https://www.uniprot.org/
https://www.rcsb.org/