

Assignment: Identification of Cancer-Associated Genes/Proteins and PDB Structures

Cancer Type: Breast invasive carcinoma

No	Gene Symbol	Protein (Uniprot)	Cancer type	PDB structure available*	PDB id
1.	PDSS2	Q86YH6 – All trans-polyprenyl-diphosphate synthase PDSS2	breast invasive carcinoma	No	
2.	ADK	P55263 – Adenosine kinase	breast invasive carcinoma	Yes	1BX4
3.	PSMD10	O75832 – 26S proteasome non-ATPase regulatory subunit 10	breast invasive carcinoma	Yes	1UOH
4.	PCMT1	P22061 – Protein-L-isoaspartate(D-aspartate) O-methyltransferase	breast invasive carcinoma	Yes	1I1N
5.	MS4A1	P11836 – B-lymphocyte antigen CD20	breast invasive carcinoma	Yes	3PP4
6.	CXCL2	P19875 – C-X-C motif chemokine 2	breast invasive carcinoma	Yes	5OB5
7.	DEF6	Q9H4E7 – Differentially expressed in FDCP 6 homolog	breast invasive carcinoma	No	
8.	GPR18	Q14330 – N-arachidonyl glycine receptor	breast invasive carcinoma	No	
9.	ZNF853	P0CG23 – Zinc finger protein 853	breast invasive carcinoma	No	
10.	LALBA	P00709 – Alpha-lactalbumin	breast invasive carcinoma	Yes	1B9O
11.	LACRT	Q9GZZ8 – Extracellular glycoprotein lacritin	breast invasive carcinoma	No	
12.	CSN2	P05814 – Beta-casein	breast invasive carcinoma	No	
13.	CST9	Q5W186 – Cystatin-9	breast invasive carcinoma	No	
14.	ANKRD30A	Q9BXX3 – Ankyrin repeat domain-containing protein 30A (NY-BR-1)	breast invasive carcinoma	Yes	6R2L
15.	GLRA3	O75311 – Glycine receptor subunit alpha-3	breast invasive carcinoma	Yes	9BWC
16.	FLT3	P36888 – Receptor-type tyrosine-protein kinase FLT3	breast invasive carcinoma	Yes	1RJB
17.	BRCA1	P38398 – Breast cancer type 1 susceptibility protein	breast invasive carcinoma	Yes	4IGK
18.	BRCA2	P51587 – Breast cancer type 2 susceptibility protein	breast invasive carcinoma	Yes	1MJE
19.	PTGES3	Q15185 – Prostaglandin E synthase 3	breast invasive carcinoma	Yes	1EJF
20.	PHAX	Q9H814 –Phosphorylated adaptor for RNA export protein	breast invasive carcinoma	Yes	8PNT

Notes:

*The selection of 20 associated genes/proteins was based on a combination of statistical significance (p-value), clinical survival validation, and RNA tissue-specificity scores. Consequently, the proteins were categorized into four functional groups: **validated** prognostic markers with unfavorable (Positions 1–4) and favorable (Positions 5–9) impacts on survival, diagnostic markers with high organ-specificity alongside therapeutic targets (Positions 10–16), and a group of fundamental associated genes responsible for DNA repair and hormonal regulation (Positions 17–20).

** For proteins where no experimental PDB structure is available, high-confidence predicted 3D models are accessible via the AlphaFold Protein Structure Database or SWISS-MODEL, providing structural insights through computational modeling

References:

Databases and Web Resources

1. RCSB Protein Data Bank. (n.d.). Protein structure data. Retrieved from <https://www.rcsb.org>
2. The Human Protein Atlas. (n.d.). Breast invasive carcinoma data. Retrieved from <https://www.proteinatlas.org>
3. The UniProt Consortium. (n.d.). UniProt Knowledgebase. Retrieved from <https://www.uniprot.org>

Datasets

1. The Human Protein Atlas. (2024). cancer_prognostic_data.tsv [Data set]
2. The Human Protein Atlas. (2024). cancer_category_rna_breast.tsv [Data set]

Software

1. R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>