CADD ASSIGNMENT – 1

(A)

DISEASE SELCTED: BREAST CANCER

SOURCE: https://www.malacards.org/

REASON FOR SELECTION: Breast cancer is the most common cancer in women globally, making it a highly relevant and impactful area of study. Other reasons that impacted my selection of this disease are the highly extensive research done on breast cancer with availability of abundant literature, data and experimental result available in databases such as PubMed, DRUGBANK, and MalaCards .

(B)

REVIEW PAPER

TITLE: A review of clinical aspects of breast cancer

AUTHORS: Shai Libson , Marc Lippman

DOI: 10.3109/09540261.2013.852971
LINK: https://pubmed.ncbi.nlm.nih.gov/24716497/

SUMMARY:

Breast cancer is the most commonly diagnosed cancer in women and the second leading cause of cancer-related death.
Treatment evolution: Shift from anatomic staging to biologically-based decisions.
Gene array technology: Identifies breast cancer as a heterogeneous disease with different subtypes; genetic profiling predicts chemotherapy response.
Breast conservation: Oncoplastic surgery allows wide excisions without compromising breast shape.
Sentinel lymph node biopsy: Replaces axillary dissection, reducing patient morbidity.
Targeted therapies: Focus on estrogen receptor and HER2 receptor; antibody-drug conjugates improve treatment outcomes.

Survival improvements: Effective new systemic therapies result in longer survival for metastatic breast cancer patients.

RESEARCH PAPER

TITLE: Analysis of a new therapeutic target and construction of a prognostic model for breast cancer based on ferroptosis genes

AUTHORS: Qi Li , Hengchen Liu , Yun Jin , Yuanquan Yu , Yihang Wang , Di Wu , Yinghao Guo , Longfu Xi , Dan Ye , Yanzhi Pan , Xiaoxiao Zhang , Jiangtao Li

DOI: 10.1016/j.compbiomed.2023.107370
LINK: https://pubmed.ncbi.nlm.nih.gov/37643511/

SUMMARY:

Breast cancer is the most common cancer in women and a major cause of death, with current prognostic models being inaccurate due to the disease's resistance to standard treatments.
Ferroptosis, a form of cell death involving iron accumulation and lipid peroxidation, plays a crucial role in the development of breast cancer.
The study analyzed **clinical factors** and gene expression data from breast cancer samples using the **TCGA** and **GEO** databases.
11 prognostic genes were identified (TP63, IFNG, MT3, ANO6, FLT3, PTGS2, SLC1A4, JUN, SLC7A5, CHAC1, and TF) to construct a **survival prediction model**, which showed strong predictive ability.
KEGG pathway analysis revealed that **immune-related pathways** were the primary pathways involved in breast cancer prognosis.
ssGSEA analysis showed significant differences in immune cell distributions (e.g., CD8+ T cells, B cells) and immune gene expressions (e.g., type II IFN response, APC co-inhibition).
10 immune targets related to ferroptosis in breast cancer were identified, including **CD276, CD80, HHLA2, LILRA2, NCR3LG1, NECTIN3, PVR, SLAMF9, TNFSF4, and BTN1A1**.
The study discovered new **ferroptosis-related genes** and developed a **reliable, accurate prognosis model** for breast cancer.
10 potential therapeutic targets, different from traditional treatments, were identified, offering new opportunities to improve outcomes for patients with breast cancer.

TARGET NAME	FUNCTION / ROLE IN BREAST CANCER	SOURCE
BRCA2	plays a critical role in preventing breast cancer by facilitating DNA repair	MalaCards
BRIP1	a tumor suppressor gene that plays a critical role in maintaining genomic stability and regulating HR repair	MalaCards
ERBB2/HER2	promoting tumor progression, aggressiveness, and resistance to endocrine therapy	DRUGBANK
TOP2A	Its expression can be used to predict response to anthracycline-based chemotherapy, identify resistant tumors, and monitor treatment response	DRUGBANK
ESRRG	plays a crucial role in regulating breast cancer cell behavior, and its dysregulation may contribute to tumorigenesis and progression	DRUGBANK

PROTEIN NAME	PDB ID	UNIPROT ID	CHAIN	SOURCE
BRCA2	1MIU	P60896	A	RCSB PDB
BRIP1	1T15	P38398	A	RCSB PDB
ERBB2/HER2	2JWA	P04626	A	RCSB PDB
TOP2A	4FM9	P11388	A	RCSB PDB
ESRRG	6KNR	P62508	A	https://www.rcsb.org/structure/6KNR

(C)FASTA SEQUENCES

BRCA2:

```
>1MIU_1|Chain A[auth B]|Deleted in split hand/split foot protein 1|Homo sapiens (9606)
MSEKKQPVDLGLLEEDDEFEEFPAEDWAGLDEDEDAHVWEDNWDDDNVEDDFSNQLRAELEKHGYKMETS
```

BRIP1:

```
>1T15_1|Chain A|Breast cancer type 1 susceptibility protein|Homo sapiens (9606)
VNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLITEETTHVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGRNHQGPKRARESQDRKIFRGLEICCYGPFT
NMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIP
```

ERBB2:

```
>2JWA_1|Chains A, B|Receptor tyrosine-protein kinase erbB-2|Homo sapiens (9606)
GCPAEQRASPLTSIISAVVGILLVVVLGVVFGILIKRRQQKIRK
```

TOP2A:

```
>4FM9_1|Chain A|DNA topoisomerase 2-alpha|Homo sapiens (9606)
KHNRIKGIPKLDDANDAGGRNSTECTLILTEGDSAKTLAVSGLGVVGRDKYGVFPLRGKILNVREASHKQIMENAEINNIIKIVGLQYKKNYEDEDSLKTLRYGKIMIMTDQDQDGSHIKGLLINFIH
HNWPSLLRHRFLEEFITPIVKVSKNKQEMAFYSLPEFEEWKSSTPNHKKWKVKYYKGLGTSTSKEAKEYFADMKRHRIQFKYSGPEDDAAISLAFSKKQIDDRKEWLTNFMEDRRQRKLLGLPEDYLY
GQTTTYLTYNDFINKELILFSNSDNERSIPSMVDGLKPGQRKVLFTCFKRNDKREVKVAQLAGSVAEMSSYHHGEMSLMMTIINLAQNFVGSNNLNLLQPIGQFGTRLHGGKDSASPRYIFTMLSSLA
RLLFPPKDDHTLKFLYDDNQRVEPEWYIPIIPMVLINGAEGIGTGWSCKIPNFDVREIVNNIRRLMDGEEPLPMLPSYKNFKGTIEELAPNQYVISGEVAILNSTTIEISELPVRTWTQTYKEQVLEP
MLNGTEKTPPLITDYREYHTDTTVKFVVKMTEEKLAEAERVGLHKVFKLQTSLTCNSMVLFDHVGCLKKYDTVLDILRDFFELRLKYYGLRKEWLLGMLGAESAKLNNQARFILEKIDGKIIIENKPK
KELIKVLIQRGYDSDPVKAWKEAQQKVPDEEENEESDNEKETEKSDSVTDSGPTFNYLLDMPLWYLTKEKKDELCRLRNEKEQELDTLKRKSPSDLWKEDLATFIEELEAVEAKEKQDEQVGL
```

ESRRG:

```
>6KNR_1|Chains A, B|Estrogen-related receptor gamma|Homo sapiens (9606)
GPLGSMLNPQLVQPAKKPYNKIVSHLLVAEPEKIYAMPDPTVPDSDIKALTTLCDLADRELVVIIGWAKHIPGFSTLSLADQMSLLQSAWMEILILGVVYRSLSFEDELVYADDYIMDEDQSKLAGLL
DLNNAILQLVKKYKSMKLEKEEFVTLKAIALANSDSMHIEDVEAVQKLQDVLHEALQDYEAGQHMEDPRRAGKMLMTLPLLRQTSTKAVQHFYNIKLEGKVPMHKLFLEMLEAKV
```

(D) PDB FILES

- TARGET: BRCA2

  PDB ID: 1MUI

  RESOLUTION: 3.10 ANGSTROMS

- TARGET: BRIP1

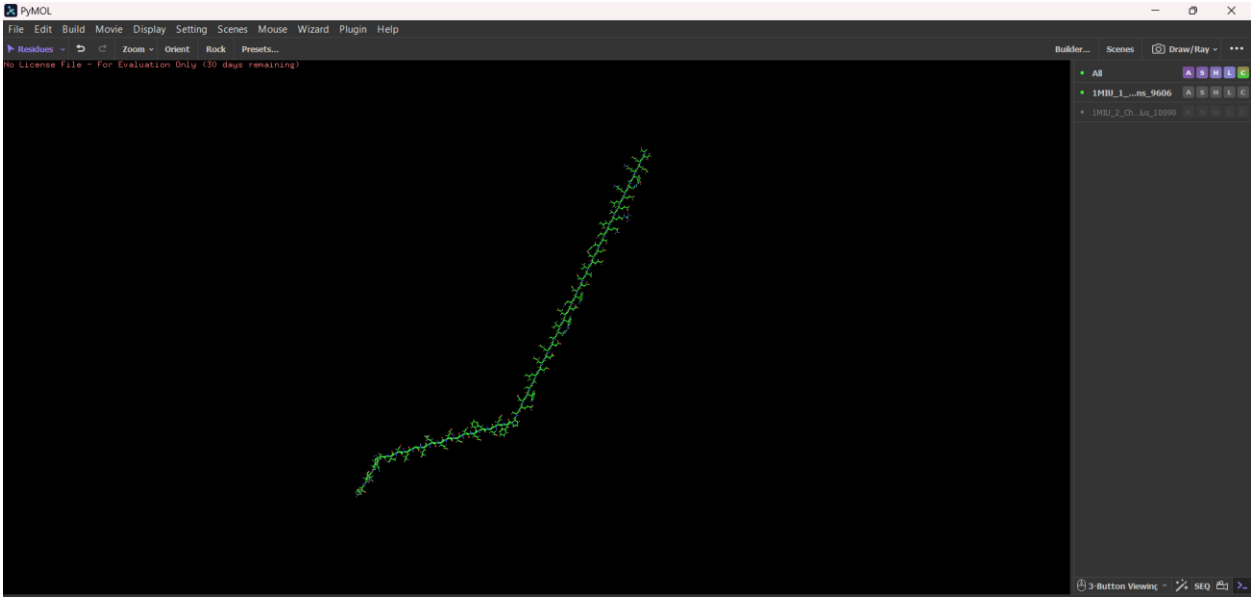
  PDB ID: 1T15

  RESOLUTION: 1.85 ANGSTROMS

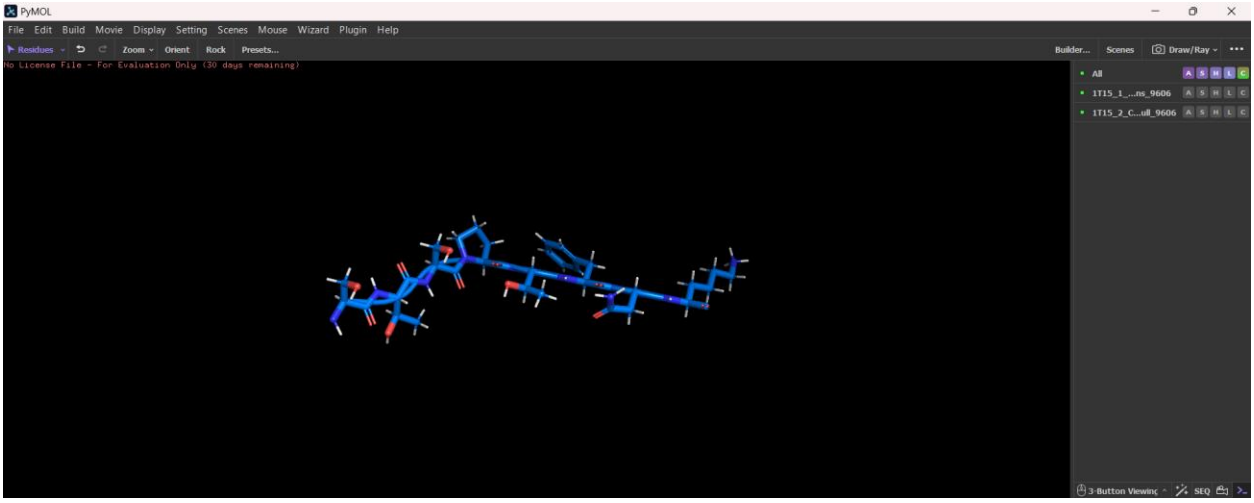
A resolution below 2.0 Angstroms in considered **high-quality**, while 2.0-3.5 Angstroms is considered **moderate-quality** and resolutions above 3.5 Angstroms might lack fine details.

(E) VISUALIZATION OF PROTEINS

1)BRCA2



2)BRIP1



2) DRUG LISTING FOR THE SELECTED DISEASE

DRUG NAME	DETAILS
Tamoxifen	Estrogen receptor antagonist
Trastuzumab	Monoclonal antibody for HER2
Letrozole	Aromatase inhibitor
Paclitaxel	Microtubule Stabilizer
Perutuzumab	HER2 dimerization inhibitor

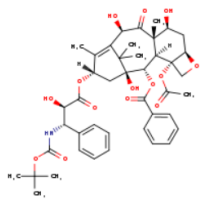
2.a)RELATED DRUGS TO THE ABOVE LISTED DRUGS

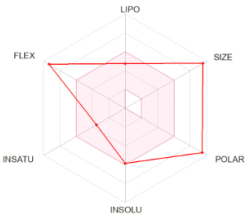
PRIMARY DRUG	RELATED DRUG 1	SOURCE	RELATED DRUG 2	SOURCE
Tamoxifen	Raloxifene	ZINC	Toremifene	ZINC
Trastuzumab	Pertuzumab	PubChem	Lapatinib	PubChem
Letrozole	Anastrozole	ChemDB	Exemestane	ChemDB
Paclitaxel	Docetaxel	DRUGBANK	Cabazitaxel	DRUGBANK
Pertuzumab	Trastuzumab	CHEMBL	Margetuximab	CHEMBL

2.b) ADME for:

DOCETAXEL

Molecule 1





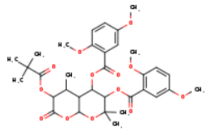
SMILES CC(=O)O[C@@]12CO[C@@H]1C[C@@H]([C@@]1([C@@H]2[C@H](OC(=O)c2ccccc2)[C@]2(O)C[C@H](OC(=O)[C@@H]([C@H](c3ccccc3)NC(=O)OC(C)(C)C)O)C(=C(C2(C)C)C[C@H](C1=O)O)C)C)O

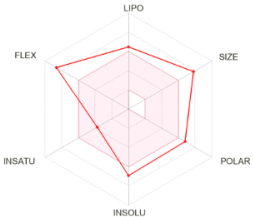
Physicochemical Properties	
Formula	C43H53NO14
Molecular weight	807.88 g/mol
Num. heavy atoms	58
Num. arom. heavy atoms	12
Fraction Csp3	0.56
Num. rotatable bonds	14
Num. H-bond acceptors	14
Num. H-bond donors	5
Molar Refractivity	205.25
TPSA	224.45 Å²
Lipophilicity	
Log $P_{o/w}$ (iLOGP)	4.10
Log $P_{o/w}$ (XLOGP3)	2.81
Log $P_{o/w}$ (WLOGP)	2.94
Log $P_{o/w}$ (MLOGP)	1.06
Log $P_{o/w}$ (SILICOS-IT)	3.51
Consensus Log $P_{o/w}$	2.88

Water Solubility	
Log *S* (ESOL)	-5.85
Solubility	1.15e-03 mg/ml ; 1.42e-06 mol/l
Class	Moderately soluble
Log *S* (Ali)	-7.18
Solubility	5.33e-05 mg/ml ; 6.60e-08 mol/l
Class	Poorly soluble
Log *S* (SILICOS-IT)	-6.66
Solubility	1.76e-04 mg/ml ; 2.18e-07 mol/l
Class	Poorly soluble
Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	Yes
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	Yes
Log K_p (skin permeation)	-9.23 cm/s
Druglikeness	
Lipinski	No; 2 violations: MW>500, NorO>10
Ghose	No; 3 violations: MW>480, MR>130, #atoms>70
Veber	No; 2 violations: Rotors>10, TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 3 violations: MW>600, TPSA>150, H-acc>10
Bioavailability Score	0.17
Medicinal Chemistry	
PAINS	0 alert
Brenk	2 alerts: isolated_alkene, more_than_2_esters
Leadlikeness	No; 2 violations: MW>350, Rotors>7
Synthetic accessibility	8.39

CABAZITAXEL

Molecule 1





SMILES COc1ccc(c(c1)C(=O)OC1C2C(OC(=O)C(C2C)OC(=O)C(C)(C)C)OC(C1OC(=O)c1cc(OC)ccc1OC)(C)C)OC

Physicochemical Properties	
Formula	C34H42O13
Molecular weight	658.69 g/mol
Num. heavy atoms	47
Num. arom. heavy atoms	12
Fraction Csp3	0.53
Num. rotatable bonds	13
Num. H-bond acceptors	13
Num. H-bond donors	0
Molar Refractivity	165.80
TPSA	151.35 Å²
Lipophilicity	
Log $P_{o/w}$ (iLOGP)	4.21
Log $P_{o/w}$ (XLOGP3)	5.82
Log $P_{o/w}$ (WLOGP)	4.37
Log $P_{o/w}$ (MLOGP)	2.46
Log $P_{o/w}$ (SILICOS-IT)	4.25
Consensus Log $P_{o/w}$	4.22

Water Solubility	
Log *S* (ESOL)	-6.92
Solubility	7.89e-05 mg/ml ; 1.20e-07 mol/l
Class	Poorly soluble
Log *S* (Ali)	-8.77
Solubility	1.12e-06 mg/ml ; 1.70e-09 mol/l
Class	Poorly soluble
Log *S* (SILICOS-IT)	-6.67
Solubility	1.39e-04 mg/ml ; 2.11e-07 mol/l
Class	Poorly soluble
Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	Yes
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	Yes
CYP3A4 inhibitor	No
Log K_p (skin permeation)	-6.19 cm/s
Druglikeness	
Lipinski	No; 2 violations: MW>500, NorO>10
Ghose	No; 3 violations: MW>480, MR>130, #atoms>70
Veber	No; 2 violations: Rotors>10, TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 4 violations: MW>600, XLOGP3>5, TPSA>150, H-acc>10
Bioavailability Score	0.17
Medicinal Chemistry	
PAINS	0 alert
Brenk	1 alert: more_than_2_esters
Leadlikeness	No; 3 violations: MW>350, Rotors>7, XLOGP3>3.5
Synthetic accessibility	6.24

<u>TARGETS FOR THE FOLLOWING :</u>

DOCETAXEL

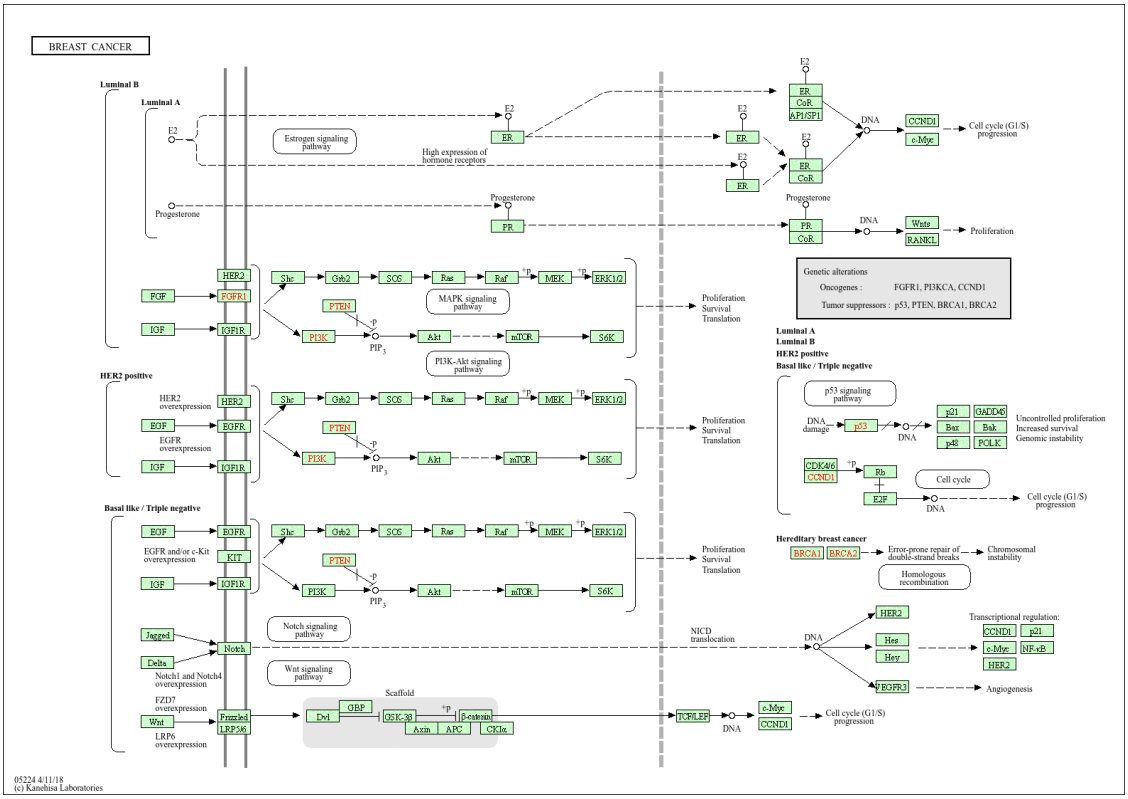
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*
Thrombin and coagulation factor X	F10	P00742	CHEMBL244	Protease	
Protein-tyrosine phosphatase 1B	PTPN1	P18031	CHEMBL335	Phosphatase	
Beta-secretase 1	BACE1	P56817	CHEMBL4822	Protease	
AMY1C	AMY1A	P04745	CHEMBL2478	Enzyme	
Squalene monooxygenase *(by homology)*	SQLE	Q14534	CHEMBL3592	Enzyme	
Tyrosine-protein kinase JAK2	JAK2	O60674	CHEMBL2971	Kinase	
Insulin-like growth factor I receptor	IGF1R	P08069	CHEMBL1957	Kinase	
Matrix metalloproteinase 9	MMP9	P14780	CHEMBL321	Protease	
Platelet activating factor receptor	PTAFR	P25105	CHEMBL250	Family A G protein-coupled receptor	
Cholecystokinin A receptor	CCKAR	P32238	CHEMBL1901	Family A G protein-coupled receptor	

CABAZITAXEL

Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*
Thrombin and coagulation factor X	F10	P00742	CHEMBL244	Protease	
Protein-tyrosine phosphatase 1B	PTPN1	P18031	CHEMBL335	Phosphatase	
Beta-secretase 1	BACE1	P56817	CHEMBL4822	Protease	
AMY1C	AMY1A	P04745	CHEMBL2478	Enzyme	
Squalene monooxygenase *(by homology)*	SQLE	Q14534	CHEMBL3592	Enzyme	
Tyrosine-protein kinase JAK2	JAK2	O60674	CHEMBL2971	Kinase	
Insulin-like growth factor I receptor	IGF1R	P08069	CHEMBL1957	Kinase	
Matrix metalloproteinase 9	MMP9	P14780	CHEMBL321	Protease	
Platelet activating factor receptor	PTAFR	P25105	CHEMBL250	Family A G protein-coupled receptor	
Cholecystokinin A receptor	CCKAR	P32238	CHEMBL1901	Family A G protein-coupled receptor	

3.KEGG PATHWAY AND RELATED GENES

CODE : hsa05224



SOME OF THE ASSOCIATED GENES

CODE DESCRIPTION

2099 ESR1; estrogen receptor 1 [KO:K08550]

2100 ESR2; estrogen receptor 2 [KO:K08551]

8648 NCOA1; nuclear receptor coactivator 1 [KO:K09101] [EC:2.3.1.48]

8202 NCOA3; nuclear receptor coactivator 3 [KO:K11256] [EC:2.3.1.48]

2353 FOS; Fos proto-oncogene, AP-1 transcription factor subunit [KO:K04379]

3725 JUN; Jun proto-oncogene, AP-1 transcription factor subunit [KO:K04448]

6667 SP1; Sp1 transcription factor [KO:K04684]

595 CCND1; cyclin D1 [KO:K04503]

4609 MYC; MYC proto-oncogene, bHLH transcription factor [KO:K04377]

5241 PGR; progesterone receptor [KO:K08556]

7471 WNT1; Wnt family member 1 [KO:K03209]

54361 WNT4; Wnt family member 4 [KO:K00408]

8600 TNFSF11; TNF superfamily member 11 [KO:K05473]

2064 ERBB2; erb-b2 receptor tyrosine kinase 2 [KO:K05083] [EC:2.7.10.1]

2246 FGF1; fibroblast growth factor 1 [KO:K18496]

2247 FGF2; fibroblast growth factor 2 [KO:K18497]

2248 FGF3; fibroblast growth factor 3 [KO:K04358]

2249 FGF4; fibroblast growth factor 4 [KO:K04358]

8822 FGF17; fibroblast growth factor 17 [KO:K04358]

2251 FGF6; fibroblast growth factor 6 [KO:K04358]

2252 FGF7; fibroblast growth factor 7 [KO:K04358]

2253 FGF8; fibroblast growth factor 8 [KO:K04358]

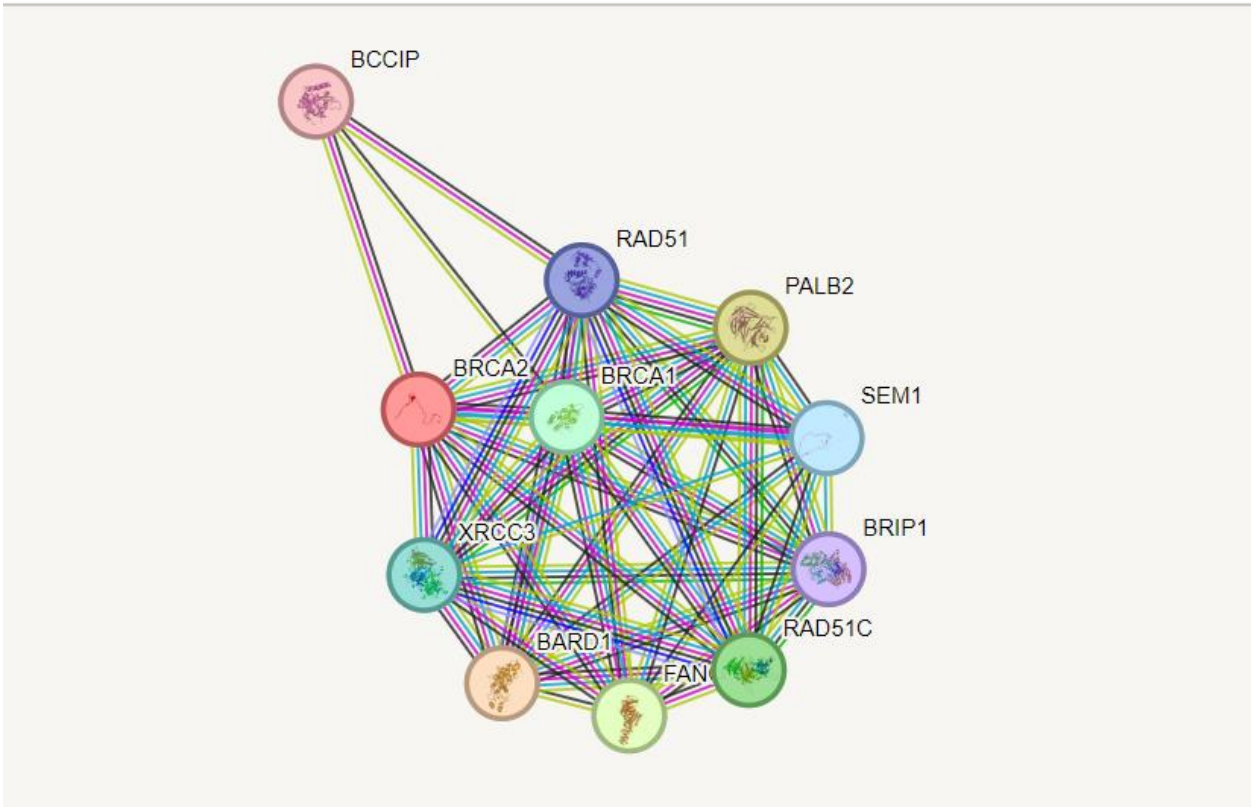
2254 FGF9; fibroblast growth factor 9 [KO:K04358]

2255 FGF10; fibroblast growth factor 10 [KO:K04358]

8823 FGF16; fibroblast growth factor 16 [KO:K04358]

2250 FGF5; fibroblast growth factor 5 [KO:K04358]

4)STRING DIAGRAM FOR BCRA2



a. BRCA2-RAD51

Interaction Type: Direct physical binding.

Role: BRCA2 loads RAD51 onto single-stranded DNA during HR repair, facilitating strand invasion and repair.

Relevance: Dysfunction in this interaction leads to genomic instability, a hallmark of breast cancer.

b. BRCA2-BRCA1

Interaction Type: Functional association.

Role: BRCA1 and BRCA2 coordinate in the DNA damage response pathway, with BRCA1 facilitating the recruitment of BRCA2 to repair sites.

Relevance: Mutations in either genos disrupt this pathway, increasing the risk of breast and ovarian cancers.

c. BRCA2-PALB2

Interaction Type: Physical binding.

Role: PALB2 acts as a bridge between BRCA1 and BRCA2, stabilizing BRCA2 at DNA damage sites.

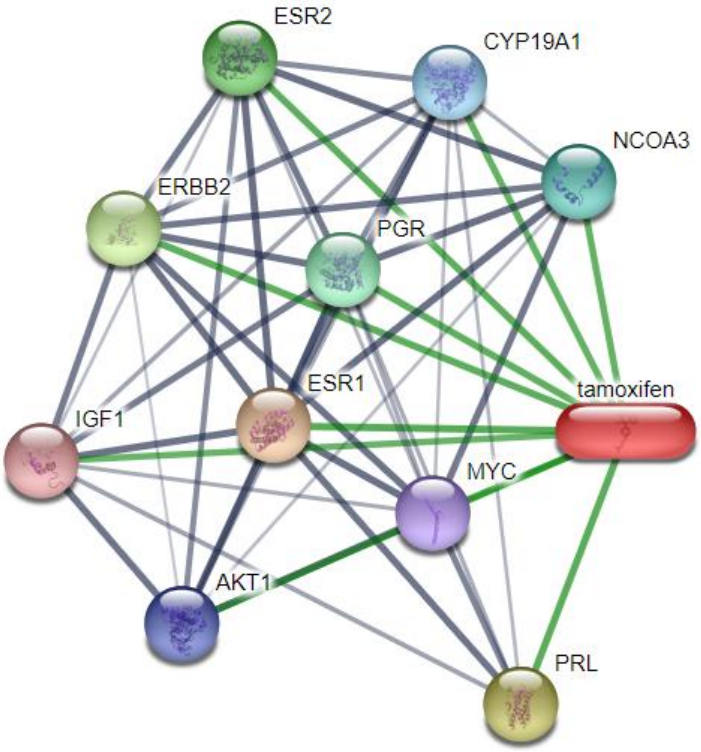
Relevance: This interaction is critical for effective HR repair.

d. BRCA2-ATM

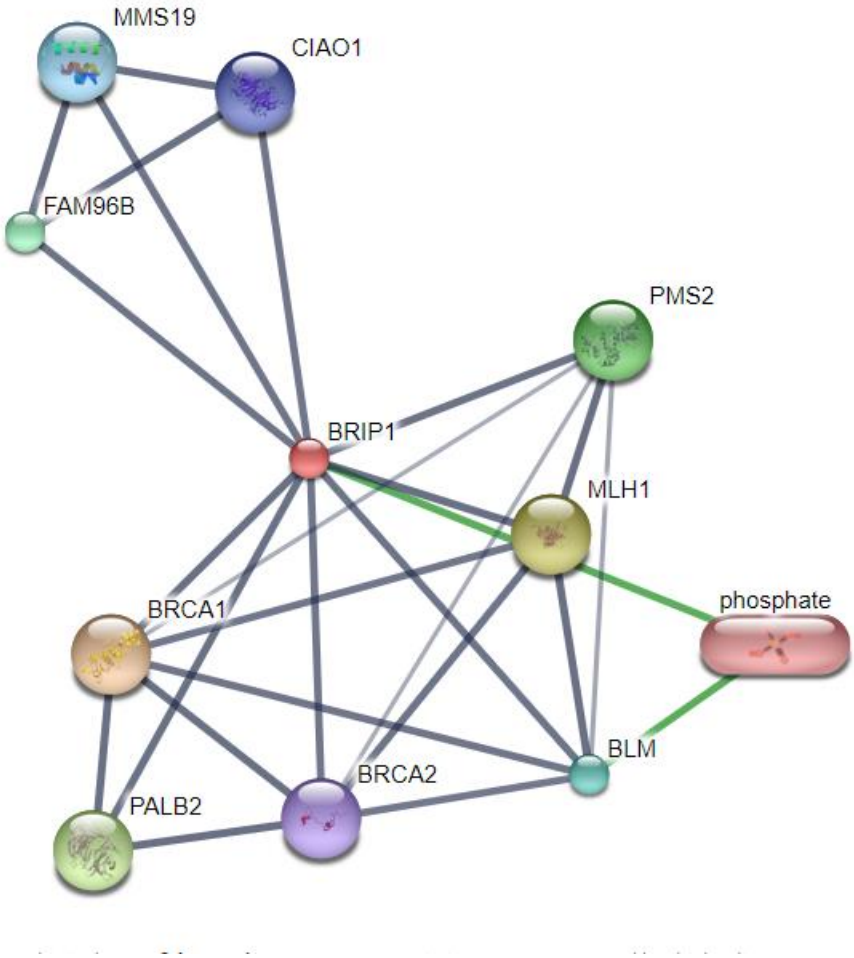
Interaction Type: Functional association.

Role: ATM phosphorylates key proteins in response to DNA damage, indirectly supporting BRCA2-mediated repair.

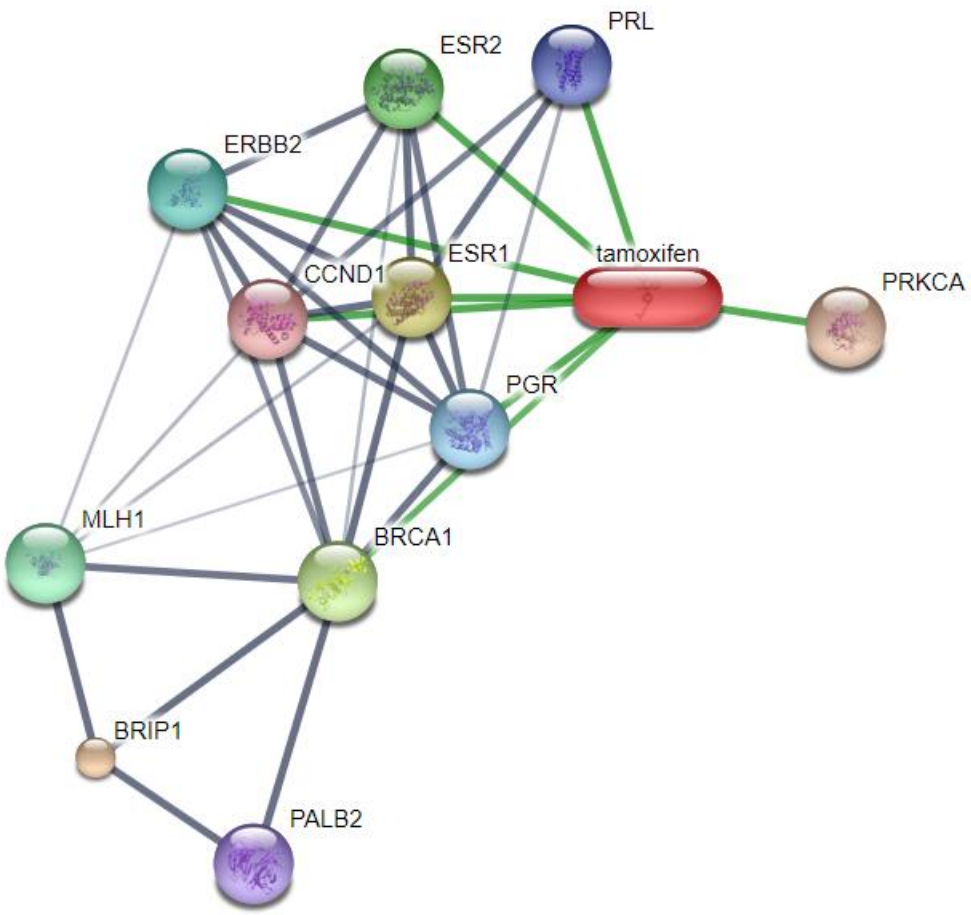
5.a)STITCH NETWORK FOR TAMOXIFEN(DRUG)



5.b)STITCH NETWORK FOR BRIP1(PROTEIN)



5)INTEGRATED STITCH NETWORK OF TAMOXIFEN AND BRIP1



1. Direct Drug-Protein Interactions:

ESR1 (Estrogen Receptor Alpha):

Primary Target of tamoxifen.

Tamoxifen binds to ESR1 and blocks estrogen-mediated signaling, crucial for breast cancer cell growth.

ESR2 (Estrogen Receptor Beta):

Tamoxifen also interacts with ESR2, modulating its activity in estrogen-responsive tissues.

PGR (Progesterone Receptor):

Tamoxifen indirectly affects PGR by modulating estrogen receptor pathways, reducing the downstream effects of estrogen.

PRKCA (Protein Kinase C Alpha):

This protein is involved in signaling pathways influencing cell proliferation and survival, with potential indirect effects from tamoxifen.

2. Indirect Interactions (via Pathways):

CCND1 (Cyclin D1):

A critical regulator of cell cycle progression, influenced by ESR1 activity. Tamoxifen inhibits ESR1, indirectly downregulating CCND1.

BRCA1 (Breast Cancer Type 1 Susceptibility Protein):

Associated with DNA repair and tumor suppression. Tamoxifen's effects on estrogen signaling may indirectly impact BRCA1-related pathways.

ERBB2 (HER2):

Known to interact with ESR1 signaling. While tamoxifen does not directly target ERBB2, it may influence HER2-positive cancer through pathway crosstalk.

PALB2 and BRIP1:

These proteins are involved in DNA damage repair and interact with BRCA1, which is indirectly modulated by tamoxifen's effects on tumor suppression pathways.

6)LIST OF 20 BREAST CANCER ASSOCIATED GENES WITH UNIPROT ID FROM

Gene	Gene Full Name	UniProt
ESR1	estrogen receptor 1	Q9UBT1 | A0A125SXW3 | H0Y4W6 | A8KAF4 | G4XH65 | P03372
BRCA1	BRCA1 DNA repair associated	A0A9Y1QQK3 | P38398
BRCA2	BRCA2 DNA repair associated	P51587
PIK3CA	phosphatidylinositol-4,5-bisphos...	P42336 | Q4LE51
TP53	tumor protein p53	Q53GA5 | A0A087WXZ1 | A0A087X1Q1 | H2EHT1 | P04637 | K7PPA8 | A0A087WT22
CDH1	cadherin 1	B3GN61 | P12830 | A0A0U2ZQU7 | Q9UII7
PTEN	phosphatase and tensin homolog	F6KD01 | P60484
ATM	ATM serine/threonine kinase	Q13315
CHEK2	checkpoint kinase 2	O96017
FGFR2	fibroblast growth factor receptor 2	P21802 | D2CGD1 | A0A141AXF1 | S4R381
RAD51	RAD51 recombinase	Q06609
BRIP1	BRCA1 interacting helicase 1	A0A804HL36 | Q9BX63
BARD1	BRCA1 associated RING domain 1	F6MDI2 | F6MDI0 | A0AVN2 | C9IYG1 | Q99728 | F6MDI1 | A0A087WZ19
AKT1	AKT serine/threonine kinase 1	B3KVH4 | B0LPE5 | P31749
KRAS	KRAS proto-oncogene, GTPase	I1SRC5 | P01116 | L7RSL8
CAV1	caveolin 1	Q03135 | Q59E85 | A9XTE5 | Q2TNI1 | Q7Z4F3
CYP19A1	cytochrome P450 family 19 subfa...	P11511 | A8K6W3 | Q8TCA4 | Q05CU4 | Q8IYG4
MKI67	marker of proliferation Ki-67	P46013
TGFB1	transforming growth factor beta 1	A0A499FJK2 | P01137
MUC1	mucin 1, cell surface associated	A6ZIE6 | A5YRV0 | A0A0C4DGW3 | A5YRU7 | Q7Z538 | P15941 | B6ECB3 | A0A087X0L...

7) Benefits of Bioinformatics in Drug Development:

Target Identification: Identifies disease-associated genes and proteins.

Drug Design: Facilitates structure-based drug design and virtual screening.

Pathway Analysis: Explores disease pathways for precise interventions.

Data Integration: Combines genomics, proteomics, and clinical data.

Cost Efficiency: Reduces costs and time in drug discovery and preclinical testing.

Predictive Models: Anticipates drug efficacy and toxicity through simulations.

Personalized Medicine: Develops drugs tailored to individual genetic profiles.

Repurposing: Identifies new uses for existing drugs.

8)

Drug repurposing is the process of identifying new therapeutic uses for existing drugs that were originally developed for different diseases. This approach reduces time and cost compared to developing drugs from scratch.

Examples of drug repurposing in Breast Cancer

Tamoxifen:

Original Use: Treatment for estrogen receptor-positive (ER+) breast cancer.

Repurposed Use: Being investigated for treating infertility due to its estrogen-modulating effects.

Metformin:

Original Use: Treatment for type 2 diabetes.

Repurposed Use: Shows potential in preventing or treating breast cancer by inhibiting cancer cell growth and reducing insulin levels, which can drive tumor progression.

Bisphosphonates, including **Zoledronic Acid**, are mainly utilized for managing osteoporosis and averting bone-related complications in individuals with bone metastases. Recently, they have been redirected for breast cancer treatment owing to their capacity to inhibit bone resorption and cultivate a bone environment that is less favorable for tumor cell proliferation. By targeting the mevalonate pathway, bisphosphonates obstruct proteins like Farnesyl Diphosphate Synthase (FDPS), which is crucial for the survival and movement of cancer cells. Clinical research has demonstrated that bisphosphonates can diminish cancer recurrence and enhance survival rates, especially among postmenopausal women, hence positioning them as a significant component in breast cancer therapeutic strategies.

ASSOCIATED DISEASE/GENE LIST FOR ZOLEDRONIC ACID

Gene	Gene Full Name	Disease	Disease Class	UniProt
EGFR	epidermal growth factor receptor	Carcinoma,Non Small Cell Lung	Neoplasms|Respiratory T...	F2YGG7 | B7Z2I3 |...
AR	androgen receptor	Neoplasm,Prostatic	Neoplasms|Urogenital Di...	G4VVI6 | P10275 | F...
BRAF	B-Raf proto-oncogene, serine/th...	Melanoma	Neoplasms|Skin and Co...	P15056
PIK3CA	phosphatidylinositol-4,5-bisphos...	Breast Neoplasms	Skin and Connective Tiss...	P42336 | Q4LE51
PIK3CA	phosphatidylinositol-4,5-bisphos...	Cancer,Breast	Skin and Connective Tiss...	P42336 | Q4LE51
AKT1	AKT serine/threonine kinase 1	Breast Neoplasms	Skin and Connective Tiss...	B0LPE5 | P31749 | B...
HIF1A	hypoxia inducible factor 1 subunit...	Breast Neoplasms	Skin and Connective Tiss...	Q16665 | D0VY79
ESR1	estrogen receptor 1	Osteoporoses	Nutritional and Metaboli...	A0A125SXW3 | H0Y4W...
PTGS2	prostaglandin-endoperoxide syn...	Breast Neoplasms	Skin and Connective Tiss...	P35354
PTH	parathyroid hormone	Osteoporoses	Nutritional and Metaboli...	P01270

KEY INTERACTING PROTEINS

FDPS (Farnesyl Diphosphate Synthase):

Target of Zoledronic Acid.

Inhibition of FDPS disrupts the mevalonate pathway, impairing cancer cell survival and bone resorption.

GGPS1 (Geranylgeranyl Diphosphate Synthase 1):

Works in the mevalonate pathway.

A secondary target related to post-translational modification of proteins necessary for cancer progression.

CASP3 (Caspase-3):

Involved in apoptosis (programmed cell death).

Suggests a role of Zoledronic Acid in inducing cancer cell death.

VEGFA (Vascular Endothelial Growth Factor A):

Regulates angiogenesis (blood vessel formation).

Interaction implies Zoledronic Acid may influence the tumor microenvironment by reducing angiogenesis.

AKT1 (Protein Kinase B):

Central in cell survival and proliferation pathways.

Interaction shows potential effects of Zoledronic Acid on inhibiting tumor cell growth.

RAP1A (Ras-related protein RAP-1A):

Linked to cell adhesion and migration.

Zoledronic Acid may reduce tumor metastasis through this pathway.

HRAS (Harvey Rat Sarcoma Viral Oncogene):

A well-known proto-oncogene involved in tumor progression.

Suggests Zoledronic Acid could affect signaling pathways critical for tumor development.

CYP19A1 (Aromatase):

Converts androgens to estrogens.

Relevant for estrogen receptor-positive breast cancer, showing Zoledronic Acid's role in hormone regulation.

TNFRSF11B (Osteoprotegerin):

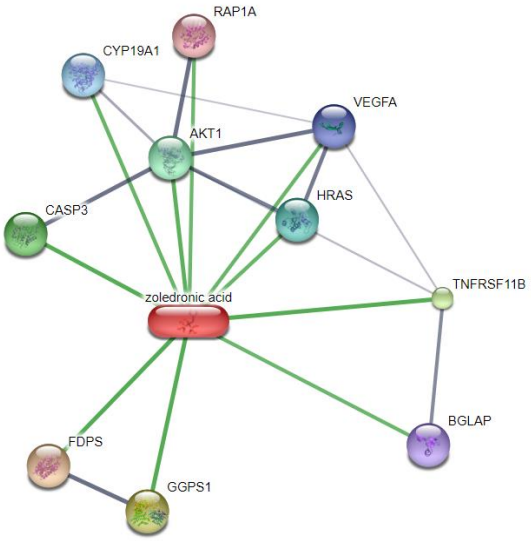
Regulates bone resorption by inhibiting osteoclast activity.

Highlights Zoledronic Acid's effect on the bone microenvironment.

BGLAP (Osteocalcin):

A marker of bone formation.

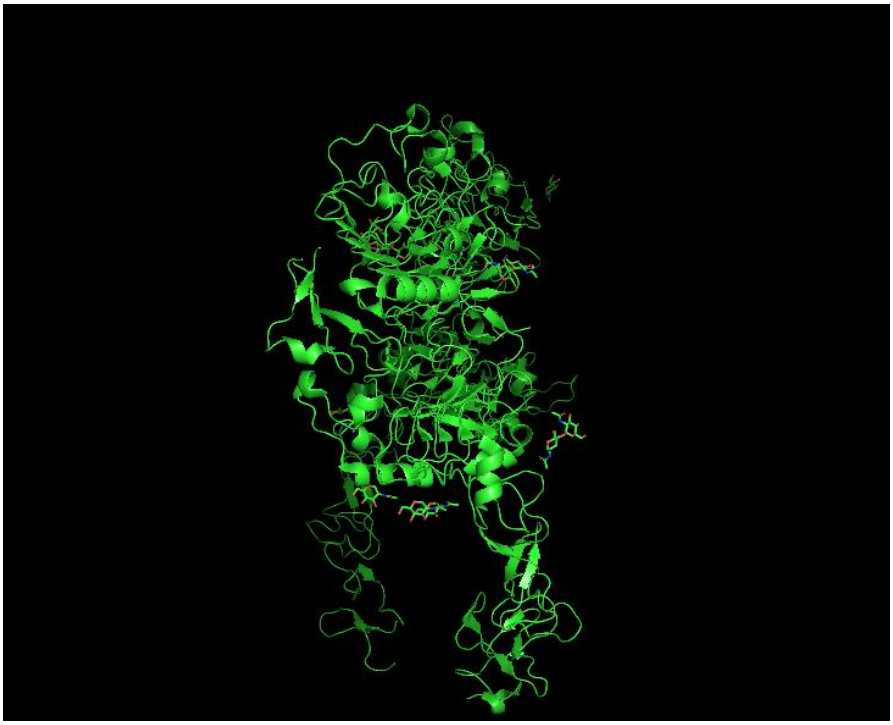
Interaction underscores the dual role of Zoledronic Acid in managing bone health and cancer progression.



INTERACTION MAP

9.

PDB IMAGE OF THE BEST PROTEIN



MSA OF THE 9 SEQUENCES OBTAINED BY PROTEIN BLASTING ERBB2(HER2)

!!AA_MULTIPLE_ALIGNMENT 1.0 squid.msf MSF: 1496 Type: P December 22, 2024 18:07 Check: 5111 ..

Name: 2N2A_A Len: 1496 Check: 726 Weight: -1.00 Name: 8VQD_A Len: 1496 Check: 238 Weight: -1.00 Name: 7MN6_B Len: 1496 Check: 8796 Weight: -1.00 Name: 6BGT_C Len: 1496 Check: 3774 Weight: -1.00 Name: 8HGO_B Len: 1496 Check: 1290 Weight: -1.00 Name: 7MN5_B Len: 1496 Check: 9121 Weight: -1.00 Name: 5KWG_C Len: 1496 Check: 5024 Weight: -1.00 Name: 6ATT_A Len: 1496 Check: 6142 Weight: -1.00

//

1 50

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~TQVCTGTD MKLRLPASPE THLDMLRHLY 7MN6_B MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY 6BGT_C MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY 8HGO_B MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY 7MN5_B MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~TQVCTGTD MKLRLPASPE THLDMLRHLY 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~TQVCTGTD MKLRLPASPE THLDMLRHLY

51 100

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 7MN6_B QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 6BGT_C QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 8HGO_B QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 7MN5_B QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 5KWG_C QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR 6ATT_A QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR

101 150

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 7MN6_B IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 6BGT_C IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 8HGO_B IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 7MN5_B IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 5KWG_C IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK 6ATT_A IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK

151 200

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 7MN6_B GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 6BGT_C GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 8HGO_B GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 7MN5_B GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 5KWG_C GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK 6ATT_A GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK

201 250

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 7MN6_B GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 6BGT_C GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 8HGO_B GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 7MN5_B GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 5KWG_C GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS 6ATT_A GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS

251 300

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A DCLACLHFNH SGICELHCPA LVTFNTDTFE SMPNPEGRYT FGASCVTACP 7MN6_B DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP 6BGT_C DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP 8HGO_B DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP 7MN5_B DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP 5KWG_C DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP 6ATT_A DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP

301 350

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A YNYLSTDVGF CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 7MN6_B YNYLSTDVGF CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 6BGT_C YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 8HGO_B YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 7MN5_B YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 5KWG_C YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL 6ATT_A YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL

351 400

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF 7MN6_B REVRAVTSAN IQEFAGCKKI

FGSLAFLPES FDGDPASNTA PLQPEQLQVF 6BGT_C REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF 8HGO_B REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF 7MN5_B REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF 5KWG_C REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF 6ATT_A REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGDPASNTA PLQPEQLQVF

401 450

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 7MN6_B ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 6BGT_C ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 8HGO_B ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 7MN5_B ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 5KWG_C ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI 6ATT_A ETLEEITGYL YISAWPDSLP DLSVFQNLQV IRGRILHNGA YSLTLQGLGI

451 500

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 7MN6_B SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 6BGT_C SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 8HGO_B SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 7MN5_B SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 5KWG_C SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP 6ATT_A SWLGLRSLRE LGSGLALIHH NTHLCFVHTV PWDQLFRNPH QALLHTANRP

501 550

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 7MN6_B EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 6BGT_C EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 8HGO_B EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 7MN5_B EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 5KWG_C EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL 6ATT_A EDECVGEGLA CHQLCARGHC WGPGPTQCVN CSQFLRGQEC VEECRVLQGL

551 600

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC 7MN6_B PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC 6BGT_C PREYVNARHC LPCHPECQPQ NGSVTCFGPE

ADQCVACAHY KDPPFCVARC 8HGO_B PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC 7MN5_B PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC 5KWG_C PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC 6ATT_A PREYVNARHC LPCHPECQPQ NGSVTCFGPE ADQCVACAHY KDPPFCVARC

601 650

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~AEQRASP 8VQD_A PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 7MN6_B PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 6BGT_C PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 8HGO_B PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 7MN5_B PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 5KWG_C PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP 6ATT_A PSGVKPDLSY MPIWKFPDEE GACQPCPINC THSCVDLDDK GCPAEQRASP

651 700

2N2A_A LTSIISAVVG ILL..... ..VVVLGVVF GILIKR.... .....RQQKI 8VQD_A LTPGSRSPKS CDKTHTCPPC PAPELLGGPS VFLFPPKPKD TLMISRTPEV 7MN6_B LTSIISAVVG ILL..... ..VVVLGVVF GILIKR.... .....RQQKI 6BGT_C LTHHHHHH~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B LTSIISAVVG ILL..... ..VVVLGVVF GILIKR.... .....RQQKI 7MN5_B LTSIISAVVG ILL..... ..VVVLGVVF GILIKR.... .....RQQKI 5KWG_C LTS~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A LT~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

701 750

2N2A_A RKYTMRRLLQ ETELVEPLG~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A TCVVVDVSHE DPEVKFNWYV DGVEVH.NAK TKPREEQYNS TYRVVS.... 7MN6_B RKYTMRRLLQ ETELVEPLTP SGAMPNQ.AQ MRILK...ET ELRKVKVLGS 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B RKYTMRRLLQ EGGSENLYFQ GGGSAQLEKE LQALE..... ..... 7MN5_B RKYTMRRLLQ ETELVEPLTP SGAMPNQ.AQ MRILK...ET ELRKVKVLGS 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

751 800

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ..VLTVLHQD WLNGKEYK.C KVSNKALPAP IEKTISKAKG QPREPQVYTL 7MN6_B GAFGTVYKGI WIPDGENVKI PVAIKVLR.. ..ENTSPKAN KEILDEAYVM 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ....KENAQL EWELQALE.. ..KELAQSNS LEVLFQ~~~~ 7MN5_B GAFGTVYKGI WIPDGENVKI PVAIKVLR.. ..ENTSPKAN KEILDEAYVM 5KWG_C ~~~~~~~~~~

~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

801 850

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A PPS..... .....RE EMTKNQVSLT CLVKG..... .FYPSDIAVE 7MN6_B AGVDSPYVSR LLGICLTSTV QLVTQLMPYG CLLDHVRENR GRLGSQDLLN 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B AGVDSPYVSR LLGICLTSTV QLVTQLMPYG CLLDHVRENR GRLGSQDLLN 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

851 900

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A WESNGQPENN YKTTPPVLDS DGSFFLYSKL TVDKSRWQQG NVFSCSVMHE 7MN6_B WCMQIAKGMS YLEDVRLVH. ....RDLAAR NVLVKSPNHV 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B WCMQIAKGMS YLEDVRLVH. ....RDLAAR NVLVKSPNHV 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

901 950

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ALHNHYTQKS LSLSPGKLEG GGGL..... .....NDIF EAQKIEWHES 7MN6_B KITDFGLARL LDIDETEYHA DGGKVPIKWM ALESILRRRF THQSDVWSYG 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B KITDFGLARL LDIDETEYHA DGGKVPIKWM ALESILRRRF THQSDVWSYG 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

951 1000

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A RHHHHHH~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B VTVWELMTFG AKPYDGIPAR EIPDLLEKGE RLPQPPICTI DVYMIMVKCW 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B VTVWELMTFG AKPYDGIPAR EIPDLLEKGE RLPQPPICTI DVYMIMVKCW 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
~~~~~~~~~~

1001 1050

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B MIDSECRPRF RELVSEFSRM ARDPQRFVVI QNEDLGPASP LDSTFYRSLL 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B MIDSECRPRF RELVSEFSRM ARDPQRFVVI QNEDLGPASP LDSTFYRSLL 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1051 1100

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B EDDDMGDLVD AEEYLVPQQG GGSLEVLFQG PSSPSGSSMK IEEGKLVIWI 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B EDDDMGDLVD AEEYLVPQQG GGSLEVLFQG PSSPSGSSMK IEEGKLVIWI 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1101 1150

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B NGDKGYNGLA EVGKKFEKDT GIKVTVEHPD KLEEKFPQVA ATGDGPDIIF 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B NGDKGYNGLA EVGKKFEKDT GIKVTVEHPD KLEEKFPQVA ATGDGPDIIF 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1151 1200

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B WAHDRFGGYA QSGLLAEITP DKAFQDKLYP FTWDAVRYNG KLIAYPIAVE 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B WAHDRFGGYA QSGLLAEITP DKAFQDKLYP FTWDAVRYNG KLIAYPIAVE 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1201 1250

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B ALSLIYNKDL LPNPPKTWEE IPALDKELKA KGKSALMFNL QEPYFTWPLI 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B ALSLIYNKDL LPNPPKTWEE IPALDKELKA KGKSALMFNL QEPYFTWPLI 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1251 1300

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B AADGGYAFKY ENGKYDIKDV GVDNAGAKAG LTFLVDLIKN KHMNADTDYS 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B AADGGYAFKY ENGKYDIKDV GVDNAGAKAG LTFLVDLIKN KHMNADTDYS 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1301 1350

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B IAEAAFNKGE TAMTINGPWA WSNIDTSKVN YGVTVLPTFK GQPSKPFVGV 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B IAEAAFNKGE TAMTINGPWA WSNIDTSKVN YGVTVLPTFK GQPSKPFVGV 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1351 1400

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B LSAGINAASP NKELAKEFLE NYLLTDEGLE AVNKDKPLGA VALKSYEEEL 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B LSAGINAASP NKELAKEFLE NYLLTDEGLE AVNKDKPLGA VALKSYEEEL 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1401 1450

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN6_B AKDPRIAATM ENAQKGEIMP

NIPQMSAFWY AVRTAVINAA SGRQTVDEAL 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 7MN5_B AKDPRIAATM ENAQKGEIMP NIPQMSAFWY AVRTAVINAA SGRQTVDEAL 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

1451 1496

2N2A_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~ 8VQD_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~ 7MN6_B KDAQTNSSSS GPSSPSAWSH PQFEKGGGSG GGSGGSSAWS HPQFEK 6BGT_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~ 8HGO_B ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~ 7MN5_B KDAQTNSSSS GPSSPSAWSH PQFEKGGGSG GGSGGSSAWS HPQFEK 5KWG_C ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~ 6ATT_A ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~