

Machine Learning Characterization and Prediction of Intrinsically Disordered Proteins Interactions:

A Focus on BRCA1

ASHAY ATHALYE

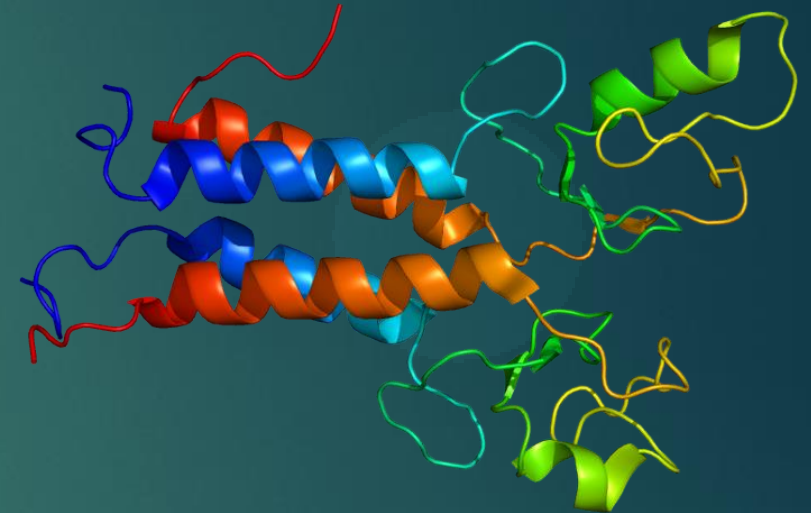
MENTOR: DR. GIL ALTEROVITZ

5TH ANNUAL PRIMES CONFERENCE

MAY 17TH, 2015

BRCA1 (Breast Cancer 1, early onset)

- ▶ Tumor suppressor gene (Caretaker gene)
- ▶ BRCA1 protein: DNA repair and transcriptional regulation
- ▶ Mutation → malfunctioning protein
 - ▶ Cancer Development
 - ▶ Responsible for > 80% breast and ovarian cancers



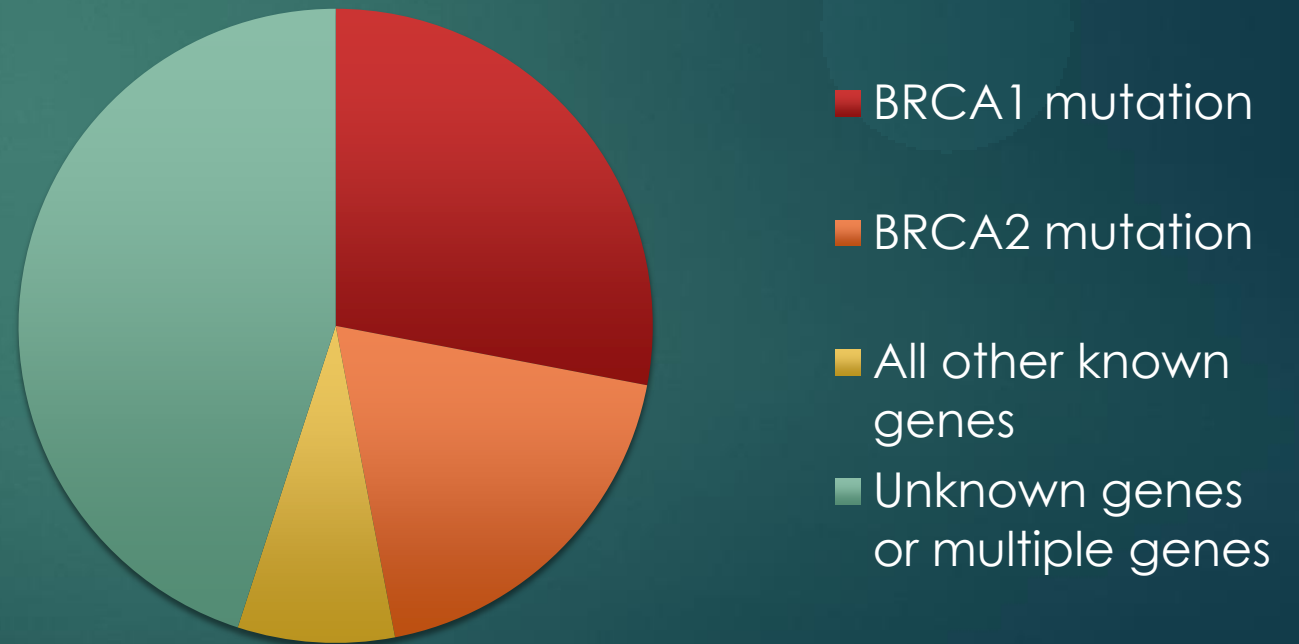
www.rscb.org

Figure 1: PDB rendering
based of model 1jm7

Clinical Significance: BRCA Mutations and Cancer

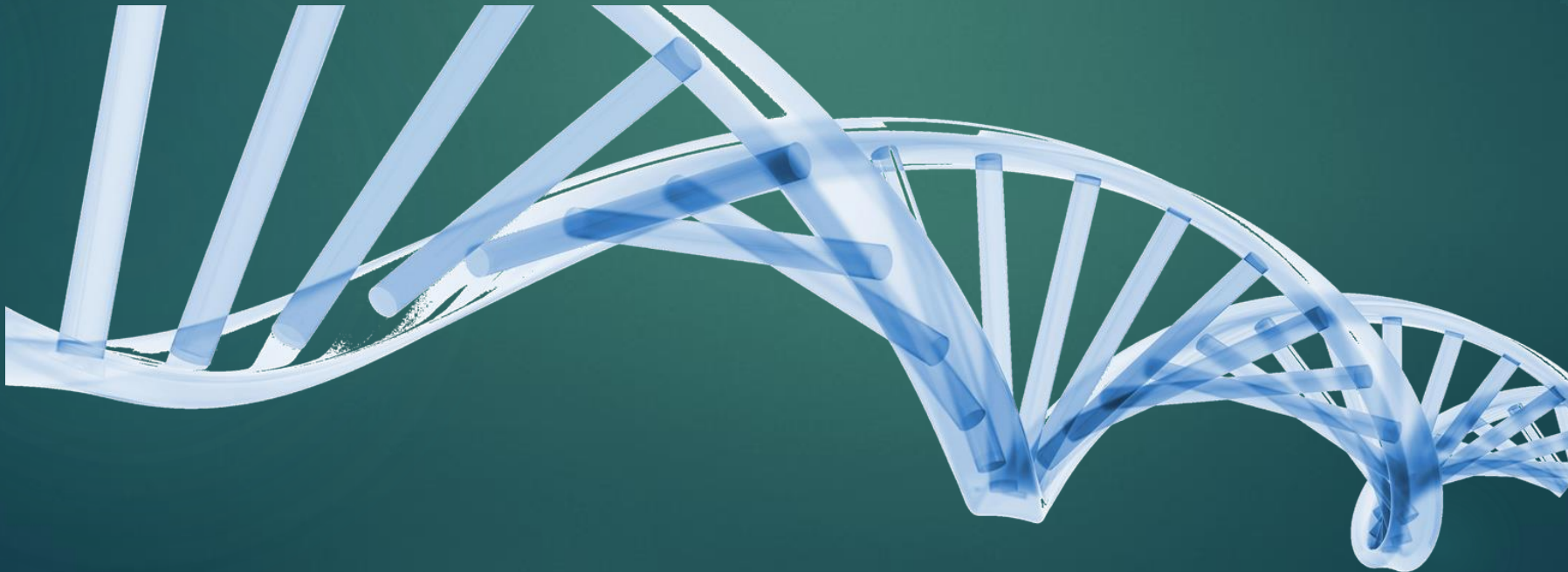
- ▶ BRCA mutation in 90% of familial breast and ovarian cancers
- ▶ BRCA mutation → **5x** breast cancer risk, **30x** ovarian cancer risk

Figure 2: Proportion of hereditary breast cancer cases caused by each gene



Problem

- ▶ Mutations in the BRCA1 lead to cancer
- ▶ Need to create a drug to replicate the behavior of the BRCA1 protein to ensure damaged DNA repair



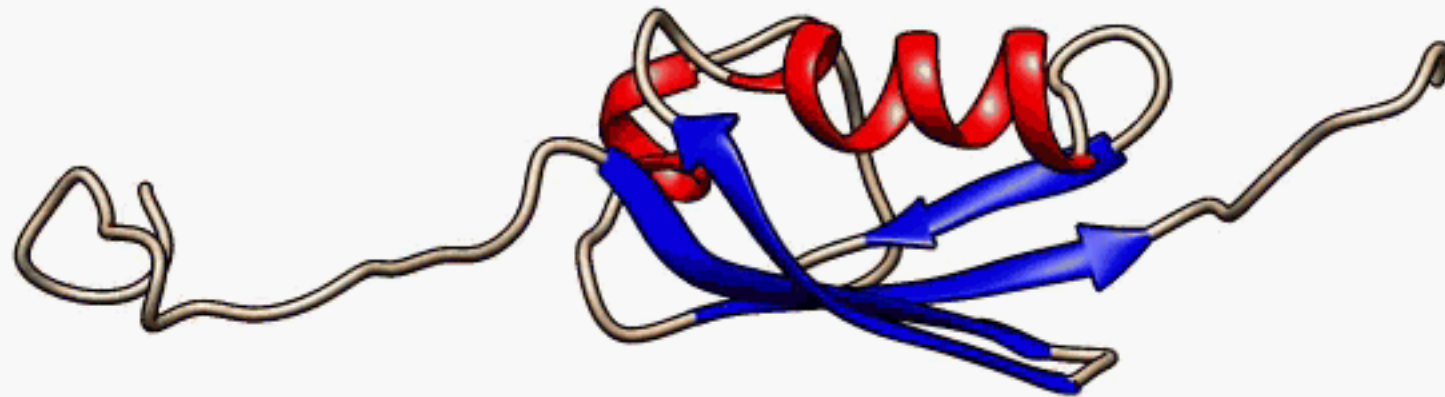


Figure 3: Conformational flexibility in PDB model 1a5r

(Majorek K, Kozlowski L, Jakalski M, Bujnicki, JM (December 18, 2008). *Prediction of Protein Structures, Functions, and Interactions*. John Wiley & Sons, Ltd.)

BRCA1 and Intrinsic Disorder

- ▶ 1863 amino acid protein
 - ▶ N-terminal RING domain (1 – 103; ~5%)
 - ▶ Central disordered region (104 – 1645; ~79%)
 - ▶ C-terminal BRCT domains (1646 – 1863; ~11%)

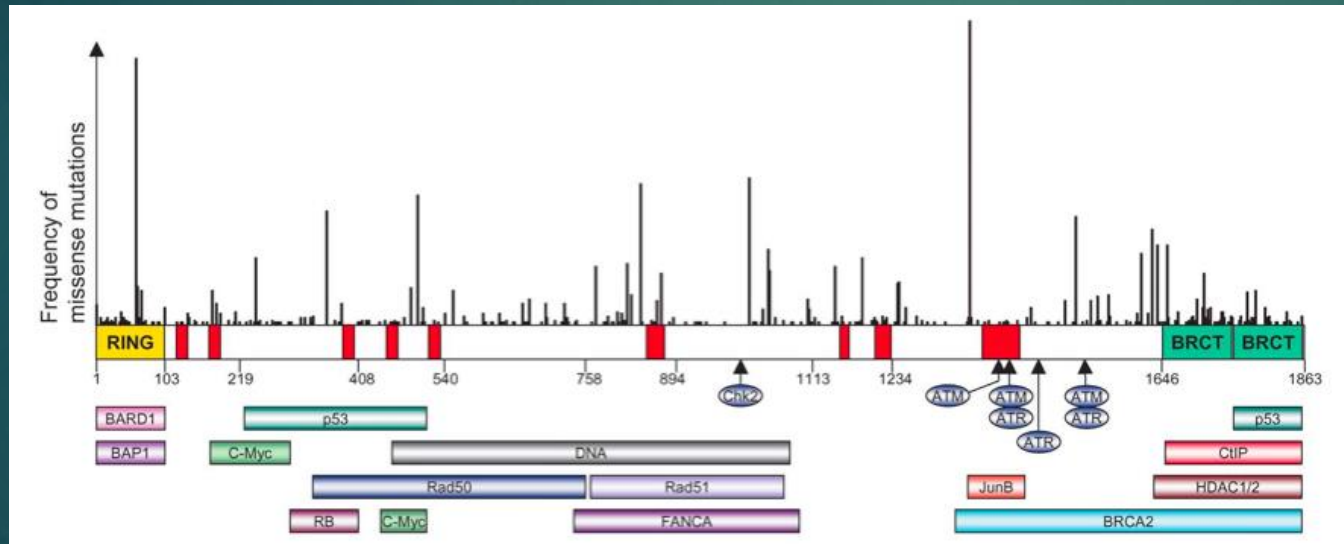


Figure 4: Representation of BRCA1 protein with interactors

(Mark, W., et al (October 14, 2004). Characterization of Segments from the Central Region of BRCA1)

Goal

- ▶ Compile data on BRCA1 interactors & interactions, identify binding-site residues
- ▶ Run, adapt, and improve current framework for IDP binding prediction
- ▶ Develop new algorithm that characterizes and predicts IDP binding interactions

- ▶ Potential for disruption or blocking of protein-protein interactions
- ▶ Potential for synthesis of mimic regions
- ▶ Develop novel drugs that inhibit or replicate BRCA1 behavior

Methods: Gene Metacore and String-DB

- ▶ Databases of protein-protein interactions
- ▶ Rich ontologies for diseases and processes with hierarchical or graphic output

Results: Pathway Mapping

- ▶ BRCA1 functions independently and jointly in a complex
 - ▶ Complexes include BASC, BARD
- ▶ Proteins that interact with BRCA1 function independently and jointly in a complex

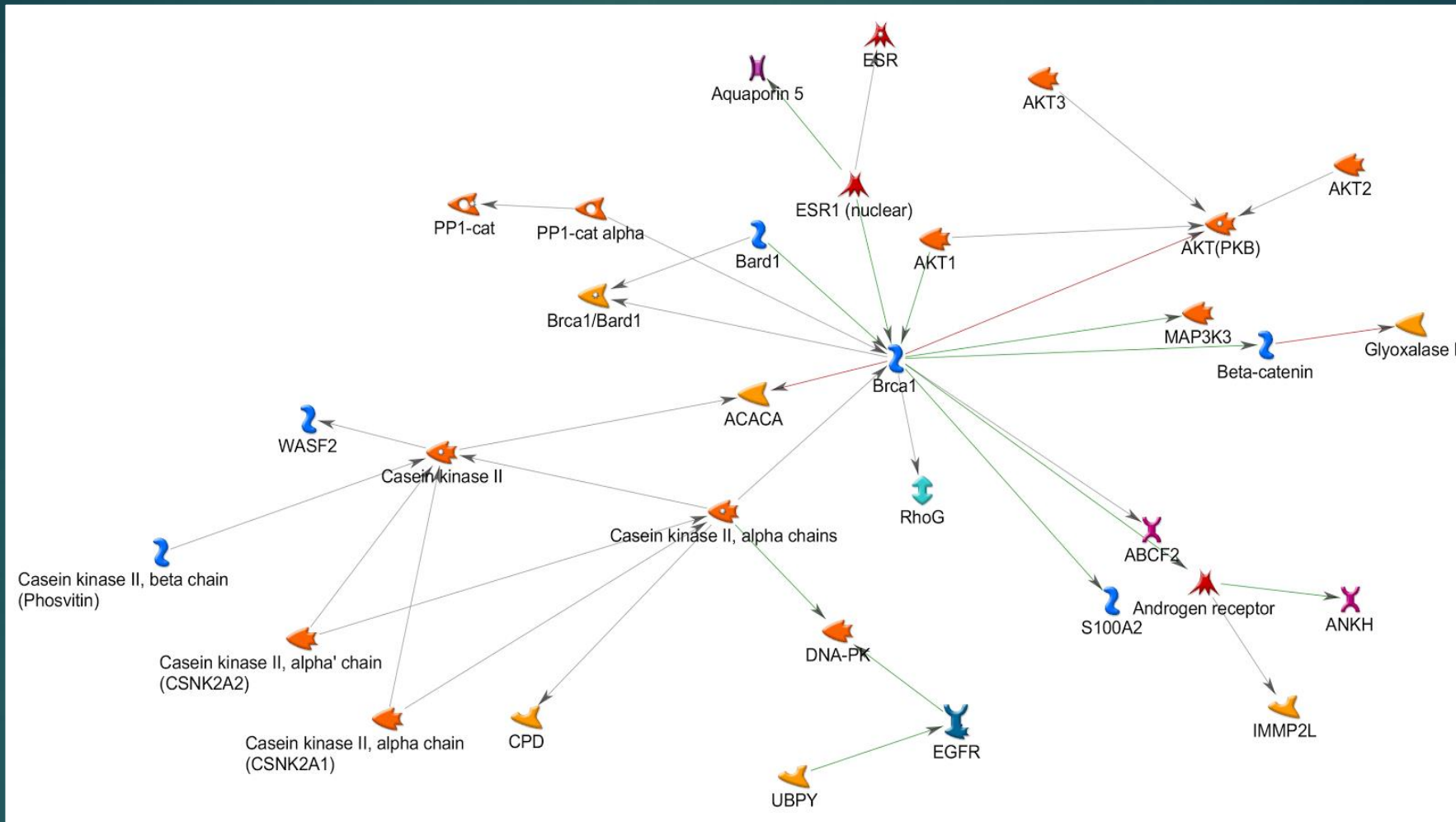


Figure 5: Example of BRCA1 interaction map. Image generated by Gene Metacore (shown is one map among many other pathway network maps & lists)

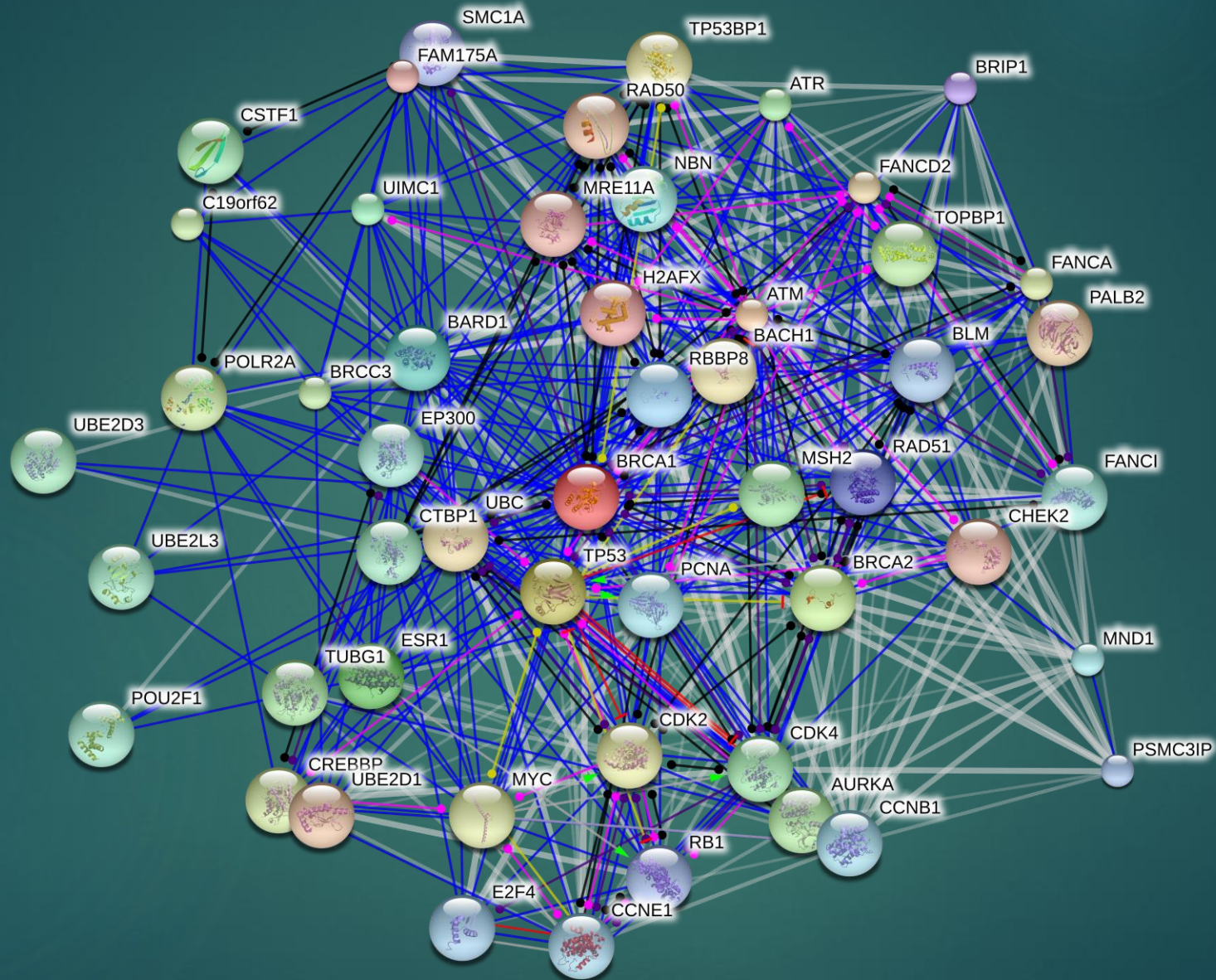


Figure 6: Example of BRCA1 interaction map. Image generated by String-DB (shown is one map among many other pathway network maps & lists)

Interactions Report									
	From		To						
#	Network Object "FROM"	Object Type	Network Object "TO"	Object Type	Effect	Mechanism	Homo sapiens	Link Info	References
1	Brca1	Generic binding protein	MEK3(MAP2K3)	Protein kinase	Unspecified	Binding	x	Brca1 probably binds to MEK3(MAP2K3).	15205325;17922014
2	miR-584-5p	RNA	Brca1	Generic binding protein	Inhibition	miRNA binding	x	BRCA1 is a predicted target of miR584.	22536353
3	miR-623	RNA	Brca1	Generic binding protein	Inhibition	miRNA binding	x	Brca1 is a predicted target of miR-623.	22166496
4	Brca1/Bard1	Generic enzyme	Histone H2B	Generic binding protein	Unspecified	Ubiquitination	x	BRCA1/BARD1 complex can ubiquitylate both free H2A and H2B histones and histones in the context of nucleosomal particles.	19916563
5	Brca1	Generic binding protein	DRIP130	Generic binding protein	Unspecified	Binding	x	Brca1 physically interacts with DRIP130.	15208681
6	Brca1	Generic binding protein	COMMD1 (MURR1)	Transporter	Unspecified	Binding	x	BRCA1-COMMD1 interaction generated in a Y2H screen was defined as true positive in affinity purification experiments.	22990118
7	Brca1	Generic binding protein	XIAP	Generic binding protein	Activation	Transcription regulation	x	Brca1 physically interacts with XIAP and increases its activity.	16322207;23435429
8	LDB2	Generic binding protein	Brca1	Generic binding protein	Unspecified	Binding	x	LDB2 physically interacts with Brca1.	11751867
9	DNA polymerase beta	Generic enzyme	Brca1	Generic binding protein	Activation	Binding	x	DNA polymerase beta binds to and activates Brca1.	23826138
10	ASH2	Transcription factor	c-Myc	Transcription factor	Unspecified	Transcription regulation	x	SET1A and ASH2 bind to the Myc and BRCA1 promoters and the H19 DMR andshRNA knockdown of SET1A decreases both H3K4 dimethylation and gene expression.	18765639;22384170
11	E2F5	Transcription factor	Brca1	Generic binding protein	Unspecified	Transcription regulation	x	E2F5 probably regulates transcription of Brca1 in human B-cell lymphoma.	18277385
12	Brca1	Generic binding protein	RPA1	Generic binding protein	Activation	Binding	x	Brca1 interacts with RPA1.	21240188;23901102
13	Brca1	Generic binding protein	Filamin A	Generic binding protein	Unspecified	Binding	x	Brca1 physically interacts with Filamin A.	11602572;20936779
14	Brca1	Generic binding protein	microRNA 145	RNA	Activation	Binding	x	Brca1 binds to and activates microRNA 145.	22492723

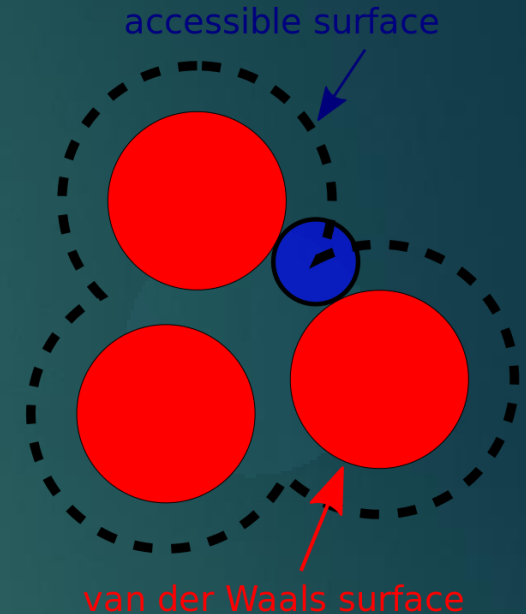
Figure 7: Example of BRCA1 interaction table. Image generated by Gene Metacore (shown is one table among many other network data tables & lists)

Enrichment analysis report								
Enrichment by Process Networks					brca1 (INTERACTIONS)			
#	Networks	Total	pValue	Min FDR	p-value	FDR	In Data	Network Objects from Active Data
1	Immune response Phagosome in antigen presentation	243	1.076E-02	1.184E-01	1.076E-02	1.184E-01	2	SEC15L, PSMB5
2	Immune response Innate immune response to RNA viral infection	85	5.809E-02	1.305E-01	5.809E-02	1.305E-01	1	IRF9
3	Inflammation IFN-gamma signaling	110	7.465E-02	1.305E-01	7.465E-02	1.305E-01	1	IRF9
4	Inflammation Interferon signaling	110	7.465E-02	1.305E-01	7.465E-02	1.305E-01	1	IRF9
5	Development Skeletal muscle development	144	9.680E-02	1.305E-01	9.680E-02	1.305E-01	1	Utrophin
6	Development Neuromuscular junction	147	9.874E-02	1.305E-01	9.874E-02	1.305E-01	1	Utrophin
7	Proteolysis Ubiquitin-proteasomal proteolysis	166	1.109E-01	1.305E-01	1.109E-01	1.305E-01	1	PSMB5
8	Transport Synaptic vesicle exocytosis	175	1.166E-01	1.305E-01	1.166E-01	1.305E-01	1	SEC15L
9	Cytoskeleton Actin filaments	176	1.173E-01	1.305E-01	1.173E-01	1.305E-01	1	Utrophin
10	Transport Potassium transport	194	1.286E-01	1.305E-01	1.286E-01	1.305E-01	1	KCNK10

Figure 8: Enrichment analysis on BRCA1 interactors. Data from Gene Metacore

Methods: Δ ASA and DSSP

- ▶ Accessible Surface Area
 - ▶ => infer binding-site residues
- ▶ Define Secondary Structure of Protein
 - ▶ => infer secondary structure from 3D coordinates of atoms
- ▶ *Problem*: no crystallized structures of BRCA1's disordered region in Protein Data Bank



Results: Δ ASA and DSSP

index	insertCode	res	buried
1649	ARG	0	
1650	MET	0	
1651	SER	0	
1652	MET	0	
1653	VAL	0	
1654	VAL	1.40794	
1655	SER	8.27484	
1656	GLY	27.5648	
1657	LEU	19.211	
1658	THR	8.9448	
1659	PRO	7.01606	
1660	GLU	0	
1661	GLU	0	
1662	PHE	0	
1663	MET	0	
1664	LEU	0	
1665	VAL	0	
1666	TYR	0	
1667	LYS	0	
1668	PHE	0	
1669	ALA	0	
1670	ARG	0	
1671	LYS	0	
1672	HIS	0	
1673	HIS	0	
1674	ILE	0	
1675	THR	0	
1676	LEU	0	
1677	THR	0	

Figure 9: Δ ASA data on 1t15 model on PDB. Calculations conducted on every model of BRCA1 in PDB.

Methods: Machine Learning

- ▶ MoRF: Molecular Recognition Feature
- ▶ Given MoRF, categorize potential binding sites in terms of class, fold, family, and superfamily and provide a yes/no binding prediction
- ▶ Database of 482 MoRFs
- ▶ Bayesian network learning

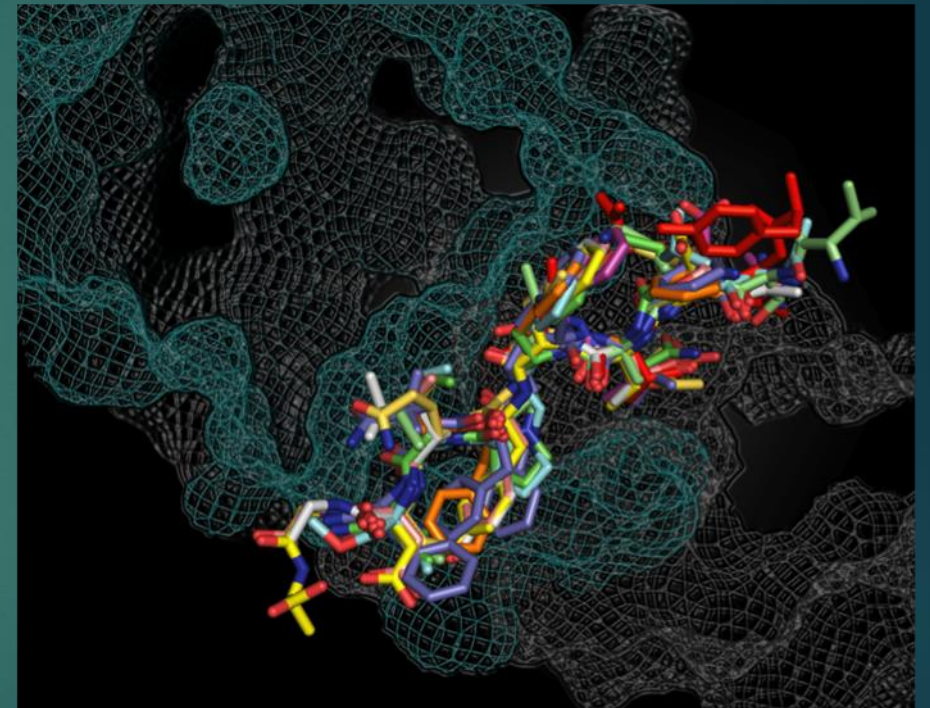
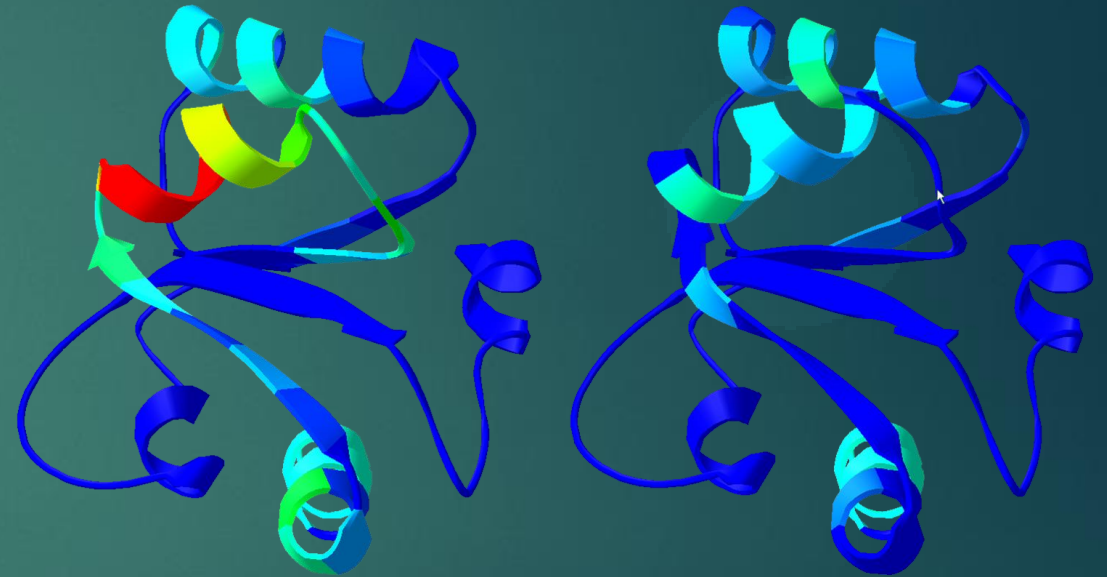


Figure 10: HIV MoRF (Jonah Kallenbach, personal communication)

Methods: Homology

- ▶ Correlate IDPs with other proteins
- ▶ Infer function of same residues on BRCA1
- ▶ BLAST: Basic Local Alignment Search Tool



www.rscb.org

Figure 13: Example of proteins with similar sequences

Ongoing Work and Future

- ▶ Draw subnetworks of BRCA1 functional network based on protein function and results of enrichment analysis
- ▶ Developing a more accurate prediction algorithm based on
 - ▶ Binding-site residue data and MoRF data
 - ▶ Interaction data from databases and enrichment analyses
 - ▶ Tailor to BRCA1

Future

- ▶ Validation by experts at Dana-Farber Cancer Institute
- ▶ Development of drugs that replicate BRCA1 behavior by imitating MoRFs and binding residues

Acknowledgements

- ▶ MIT PRIMES for this opportunity
- ▶ Dr. Gil Alterovitz for his guidance
- ▶ Mr. Mingjie Chen for his guidance
- ▶ Ms. Wei-lun Hsu for her guidance
- ▶ Mentor group
- ▶ Parents

Questions



Extra: Bayesian Networks

- ▶ Directed acyclic graph, probabilistic graphical model
- ▶ Nodes: random variables
- ▶ Edges: conditional probabilities

