

Protein-Protein Interaction Network Analysis of Subject Z's Coding SNPs

CBB 752 Final project

INVESTIGATING THE HUMAN PROTEIN-PROTEIN INTERACTOME

- Provides an exploration of the functional aspects of the human proteome
- Human Protein-Protein Interaction (PPI) files were downloaded from two databases (after filtering *w.r.t.* corss-DB ID availability):
- Database of Interacting Proteins (DIP) | *4904 distinct proteins with 7,387 interactions*
- Molecule INTeracting Database (MINT)| *4584 distinct proteins with 12,655 interactions*

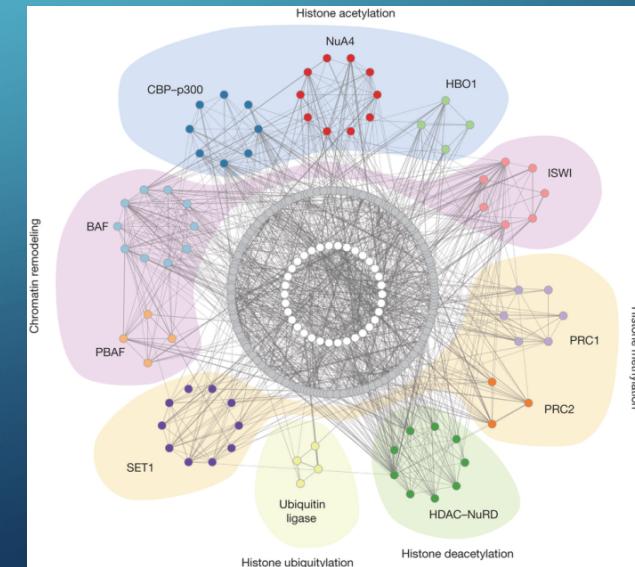
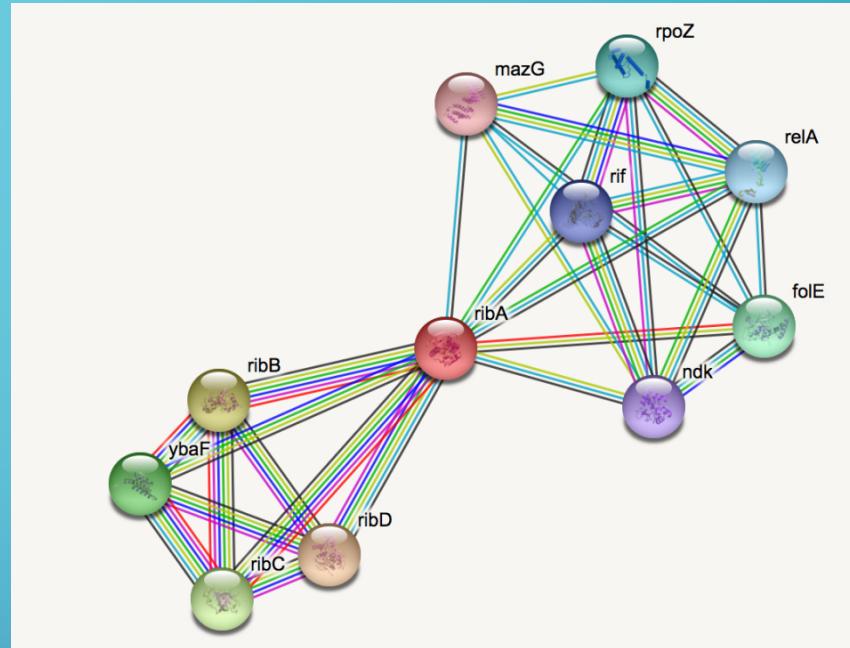
PPI Network and Subject Z's Coding SNPs

- Many ways to determine impacts of SNPs
 - Conservation, statistical modeling....
- Importance of hubs in sparse networks
- Any hubs for genes with SNPs of Carl?
 - A node with many neighbors
 - A node which is centered in communications
- Are they related to disease or any other important traits?

NETWORK THEORY

- Node: individual protein
- Edge: Evidence of interaction
- Degree Centrality: normalized number of edges connected to a node
- Betweenness Centrality: number of times a node is within the shortest path between two other nodes
- Hierarchical Networks: splits a network into layers of subnetworks
 - Scale-free topology
 - Related to high clustering of nodes

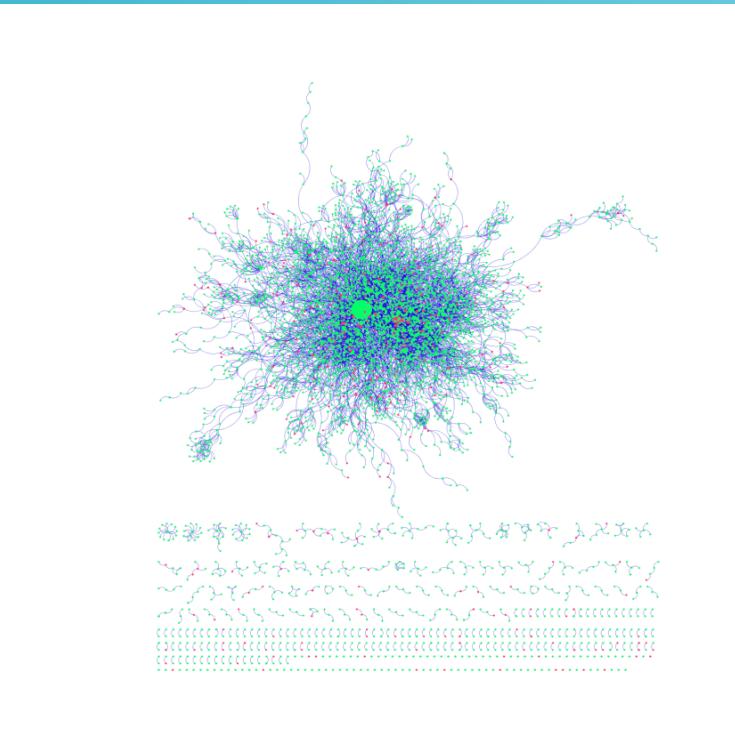
From 3.1 by James, Dingjue, and Zhaolong



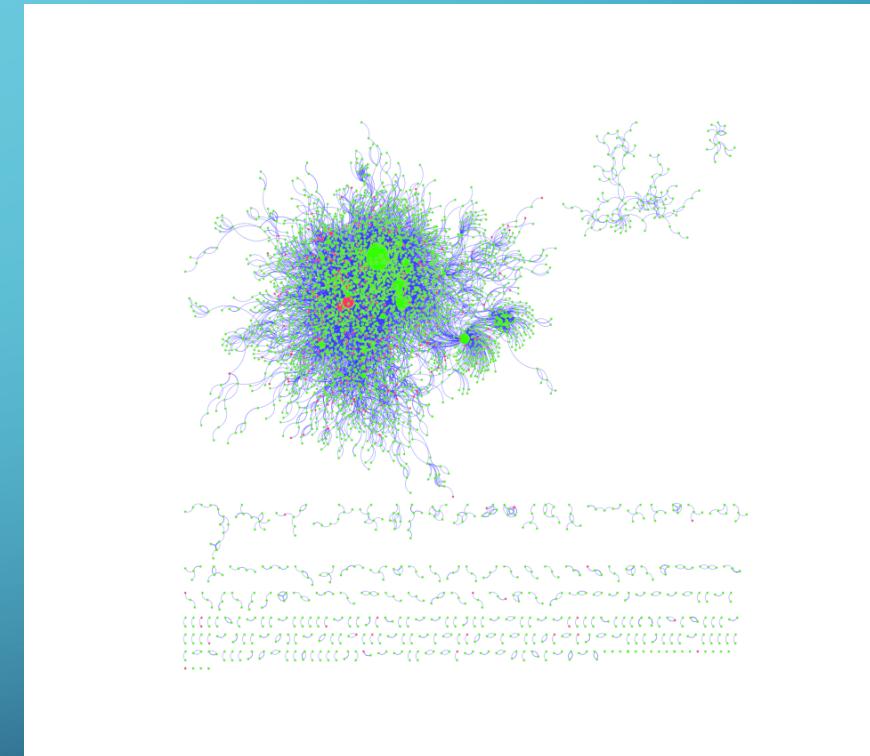
PROJECT AIMS

- Visualize the PPI networks obtained from the two databases
- Propose a tool that calculates measures of centrality and compare to output from Cytoscape
- Determine distribution of degree centrality and betweenness centrality for PPI networks
- Characterize any statistically significant differences between proteins containing and not containing SNPs in Carl's genome
- Perform hierarchical analyses on the PPI networks

Mapping Subject Z's SNPs onto the PPI Network



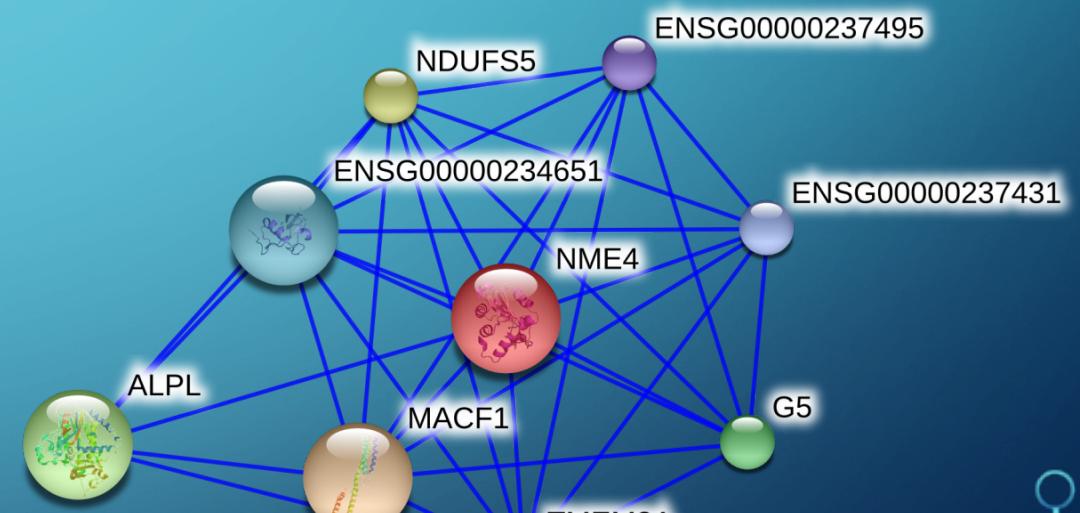
DIP



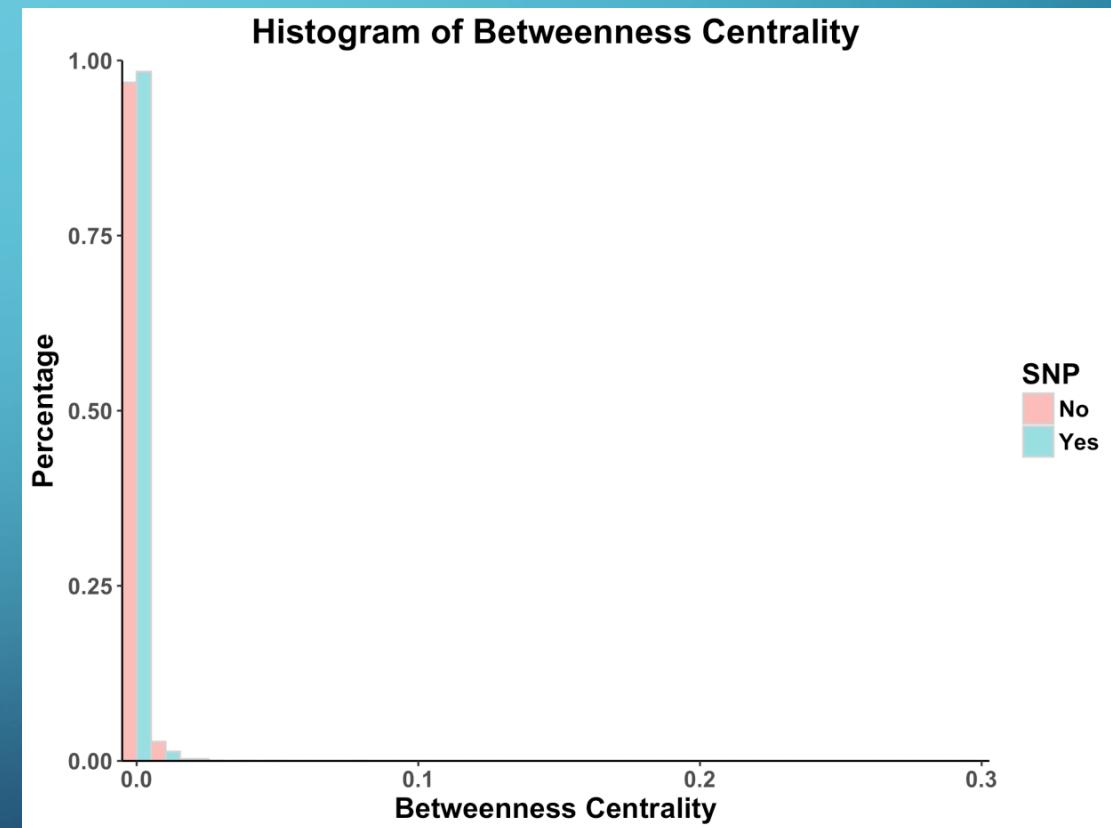
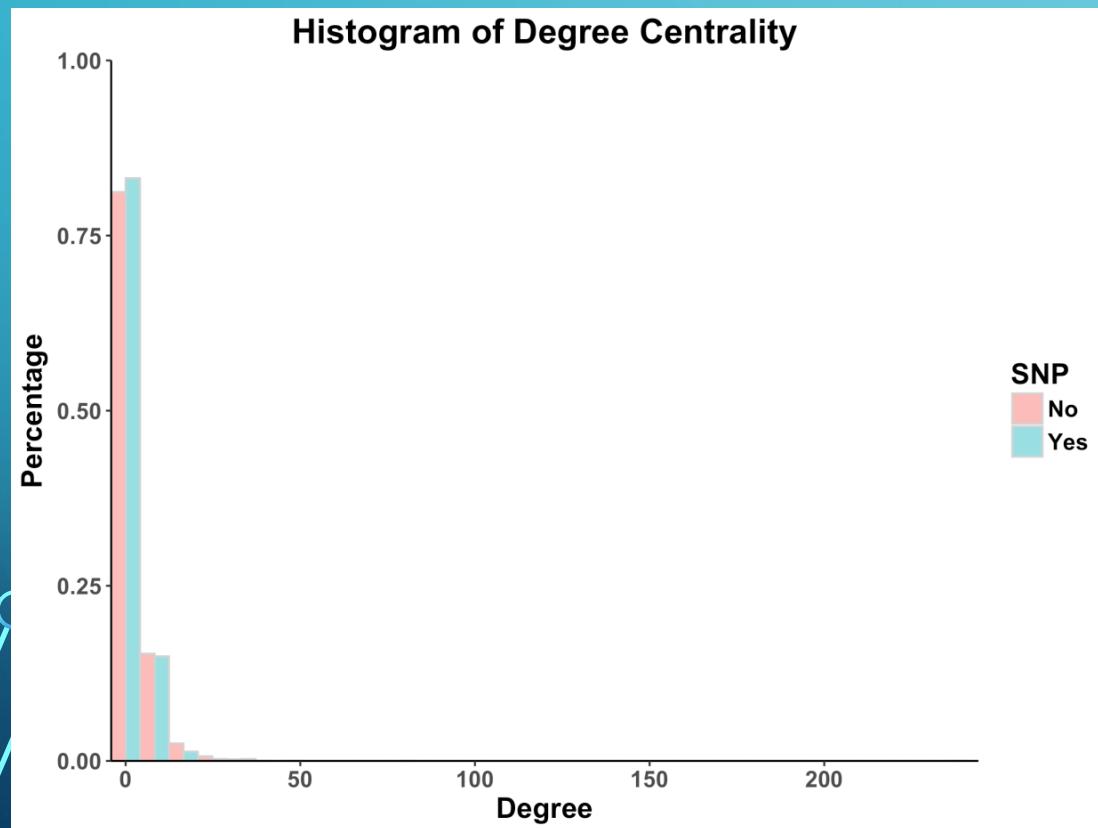
MINT

MEASURES OF CENTRALITY

- Nodes with higher betweenness and degree centralities tend to fulfill important functions and take part in multiple biological processes
- SNPs in genes of these proteins tend to be more deleterious



CYTOSCAPE-GENERATED DISTRIBUTIONS OF CENTRALITY MEASURES

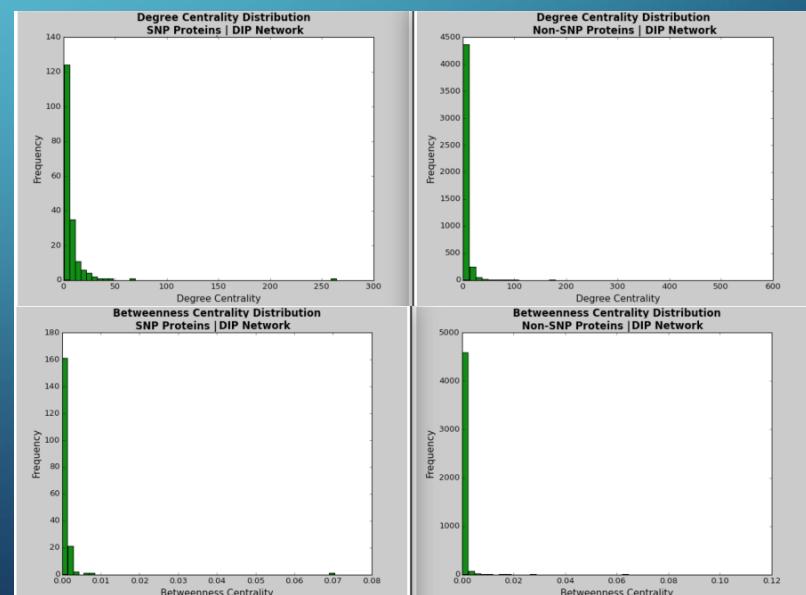


PROPOSED TOOL

- Given a csv file with 3 columns (Protein Interactor A, Protein Interactor B, Weight), calculates the degree centrality and betweenness centrality for each node and outputs distribution of centrality measures

A	B	Weight
Q8WYA6	Q13573	0.51
Q13573	Q7Z5K2	0.35
P24941	Q12974	0.17
P50479	Q12923	0.35
P08151	P40337-1	0.4
Q9BXM7	Q15388	0.43
Q13573	Q13200	0.35
Q8IWY9	Q9Y294	0.35
O00743	O15084	0.74

Input



Output

COMPARISON OF CYTOSCAPE VS PROPOSED TOOL

	Cytoscape	Cy.Degree	Coding	Co.Degree
4146	P62988	241	P62988	238
4880	P04637	86	P04637	85
4325	P10275	47	P10275	43
4583	P35222	41	P35222	41
4897	P06400	34	Q00653	37
4288	P38398	33	P06400	34
4468	P01375	32	P01375	32
3501	P62158	31	P38398	31
4889	P04049	31	P19387	30
2831	P19387	30	P63208	30

Table 1: Top 10 Nodes with Highest Degree Centrality in DIP

	Cytoscape	Cy.Degree	Coding	Co.Degree
2412	Q13573	595	Q13573	595
1111	Q99459	566	Q99459	566
2469	O95758	296	O95758	296
38	P04637	264	P04637	264
1275	Q9UL18	242	Q9UL18	242
4531	Q14197	234	Q14197	234
144	Q9UKV8	203	Q9UKV8	203
196	O35182	191	O35182	191
8	P0CG48	179	P0CG48	179
884	Q8AZK7	172	Q8AZK7	172

Table 2: Top 10 Nodes with Highest Degree Centrality in MINT

	Cytoscape	Cy.Betweenness	Coding	Co.Betweenness
4146	P62988	0.30	P62988	0.19
4880	P04637	0.08	P04637	0.07
4325	P10275	0.05	P10275	0.04
4583	P35222	0.04	P07900	0.03
4487	P21333	0.03	P35222	0.03
4771	P00533	0.03	P00533	0.03
4468	P01375	0.02	P21333	0.03
3669	P31749	0.02	P25963	0.02
4293	Q03135	0.02	P01375	0.02
4666	P24385	0.02	Q9Y4K3	0.02

Table 3: Top 10 Nodes with Highest Betweenness Centrality in DIP

	Cytoscape	Cy.Betweenness	Coding	Co.Betweenness
2412	Q13573	0.15	Q99459	0.12
1111	Q99459	0.14	P04637	0.07
2469	O95758	0.10	P0CG48	0.06
38	P04637	0.09	O95758	0.06
4531	Q14197	0.07	Q13573	0.06
8	P0CG48	0.07	Q14197	0.05
196	O35182	0.04	O35182	0.03
161	P31947	0.04	P31947	0.03
1275	Q9UL18	0.04	Q9UKV8	0.03
150	P63104	0.03	P63104	0.03

Table 4: Top 10 Nodes with Highest Betweenness Centrality in MINT

W.R.T. CARL'S GENOME

	Cytoscape	Cy.Degree	Coding	Co.Degree
4146	P62988	241	P62988	238
4880	P04637	86	P04637	85
4325	P10275	47	P10275	43
4583	P35222	41	P35222	41
4897	P06400	34	Q00653	37
4288	P38398	33	P06400	34
4468	P01375	32	P01375	32
3501	P62158	31	P38398	31
4889	P04049	31	P19387	30
2831	P19387	30	P63208	30

Table 1: Top 10 Nodes with Highest Degree Centrality in DIP

	Cytoscape	Cy.Degree	Coding	Co.Degree
2412	Q13573	595	Q13573	595
1111	Q99459	566	Q99459	566
2469	O95758	296	O95758	296
38	P04637	264	P04637	264
1275	Q9UL18	242	Q9UL18	242
4531	Q14197	234	Q14197	234
144	Q9UKV8	203	Q9UKV8	203
196	O35182	191	O35182	191
8	P0CG48	179	P0CG48	179
884	Q8AZK7	172	Q8AZK7	172

Table 2: Top 10 Nodes with Highest Degree Centrality in MINT

	Cytoscape	Cy.Betweenness	Coding	Co.Betweenness
4146	P62988	0.30	P62988	0.19
4880	P04637	0.08	P04637	0.07
4325	P10275	0.05	P10275	0.04
4583	P35222	0.04	P07900	0.03
4487	P21333	0.03	P35222	0.03
4771	P00533	0.03	P00533	0.03
4468	P01375	0.02	P21333	0.03
3669	P31749	0.02	P25963	0.02
4293	Q03135	0.02	P01375	0.02
4666	P24385	0.02	Q9Y4K3	0.02

Table 3: Top 10 Nodes with Highest Betweenness Centrality in DIP

	Cytoscape	Cy.Betweenness	Coding	Co.Betweenness
2412	Q13573	0.15	Q99459	0.12
1111	Q99459	0.14	P04637	0.07
2469	O95758	0.10	P0CG48	0.06
38	P04637	0.09	O95758	0.06
4531	Q14197	0.07	Q13573	0.06
8	P0CG48	0.07	Q14197	0.05
196	O35182	0.04	O35182	0.03
161	P31947	0.04	P31947	0.03
1275	Q9UL18	0.04	Q9UKV8	0.03
150	P63104	0.03	P63104	0.03

Table 4: Top 10 Nodes with Highest Betweenness Centrality in MINT

P38398 | BRCA1

The screenshot shows the UniProtKB entry for P38398 (BRCA1_HUMAN). The top navigation bar includes links for BLAST, Align, Retrieve/ID mapping, and Peptide search. The main title is "UniProtKB - P38398 (BRCA1_HUMAN)". Below the title, there's a "Display" section with tabs for Entry (selected), Publications, Feature viewer, and Feature table. The "Entry" tab shows the following details:

- Protein: Breast cancer type 1 susceptibility protein
- Gene: BRCA1
- Organism: Homo sapiens (Human)
- Status: Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

The bottom section is titled "Functionⁱ" and contains a detailed description of the protein's function.

E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage. It is unclear whether it also mediates the formation of other types of polyubiquitin chains. The E3 ubiquitin-protein ligase activity is required for its tumor suppressor function. The BRCA1-BARD1 heterodimer coordinates a diverse range of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability. Regulates centrosomal microtubule nucleation. Required for normal cell cycle progression from G2 to mitosis. Required for appropriate cell cycle

- Well-annotated protein
- Mutations increase susceptibility of breast cancer
- Among the proteins with highest degree centrality

[UniProtKB Record]

P38398 | BRCA1

Breast Cancer Risk Elevated In Male BRCA Mutation Carriers

Date: November 29, 2007

Source: Journal of the National Cancer Institute

Summary: Men with mutations in the BRCA1 or BRCA2 genes are at greater risk of breast cancer than the general population. Male breast cancer accounts for less than 1 percent of all breast cancers in the U.S., and it is most common in men with a family history of the disease. Previous studies have shown that men who carry mutations in the BRCA2 gene have a greater risk of developing breast cancer than men in the general population. The association between BRCA1 mutations and breast cancer in men was less clear.

P04637 | P53

UniProtKB - P04637 (P53_HUMAN)

Display

Entry

BLAST Align Format Add to basket History

Protein Cellular tumor antigen p53

Gene TP53

Organism Homo sapiens (Human)

Status Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

- Well-annotated protein
- Mutated and/or incapacitated in 60% of cancers
- Among the proteins with highest betweenness centrality

[UniProtKB Record]

Functionⁱ

Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. In

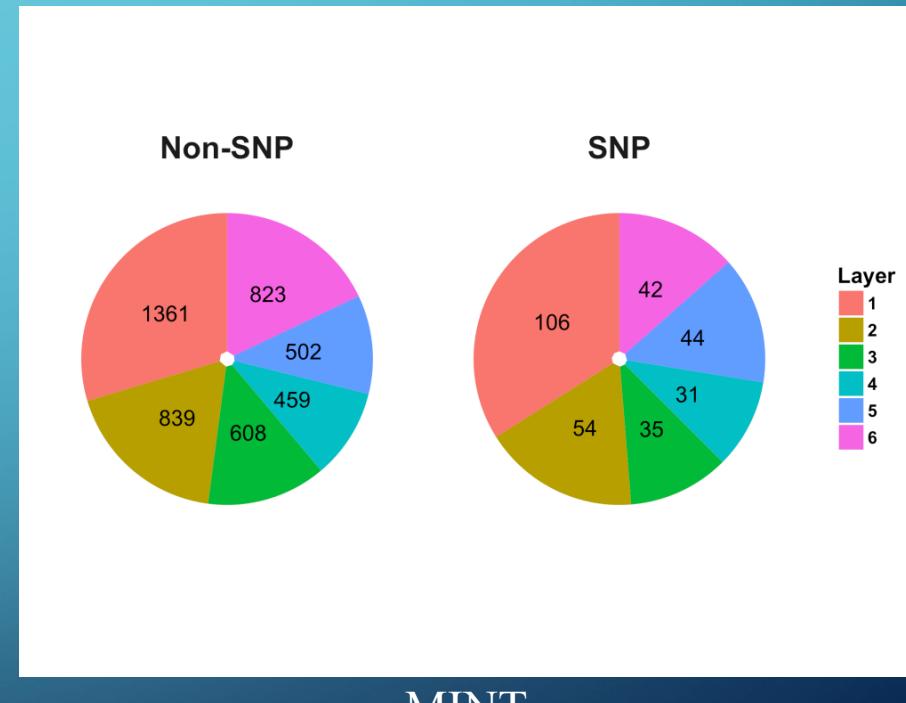
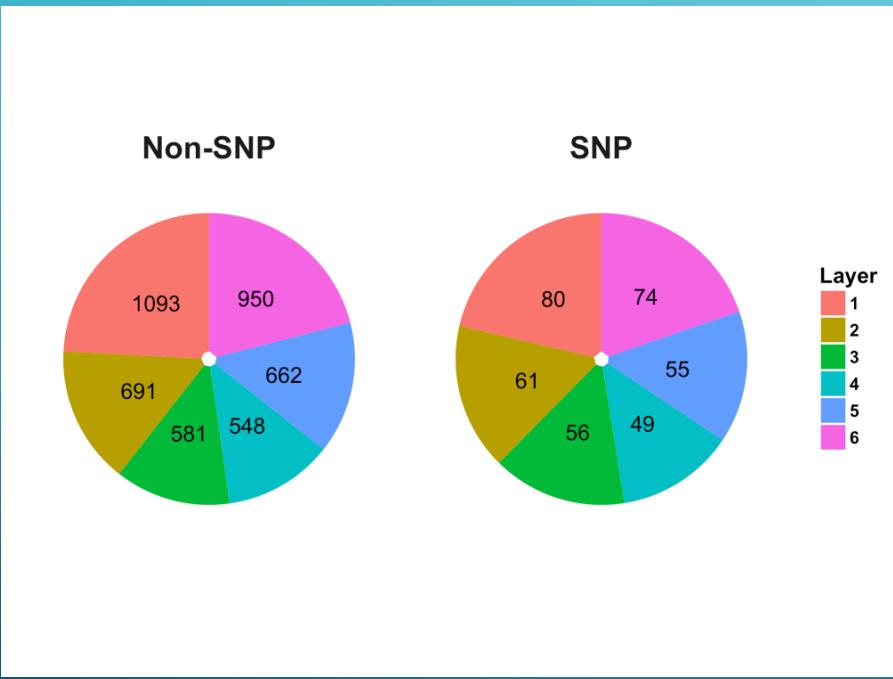
OBSERVATIONS ON THE DISTRIBUTION OF CENTRALITY MEASURES

- Histograms follow power law distributions, not Poisson distributions, suggesting that the PPI networks are scale-free
- Permutation tests show that there is no significant difference between the proteins containing and not containing SNPs
- There are no almost SNP-containing proteins in almost all ten proteins with the highest centrality measures

PPI NETWORK HIERARCHICAL ANALYSIS

- Implement Dr. Chao Cheng's HirNet, which resolves hierarchical network structure and calculates hierarchical score for *directed* networks

DISTRIBUTION OF PROTEINS IN THE HIERARCHICAL LAYERS OF THE PPI NETWORK



Fisher's Exact Test yields no statistically significant difference between layers

CONCLUSIONS

- Successfully proposed a new tool that calculates the degree centrality and the betweenness centrality for all nodes in the PPI network
- Found no statistically significant difference in distributions of proteins containing SNPs and not containing SNPs
- Observed no statistically significant enrichment of proteins/protein types in any hierarchical layers of the PPI network

MEASURES OF CENTRALITY

- Measures of centrality can identify the most important nodes within a network
- We looked at two different measures of centrality. Given a graph $G:=(V,E)$:
 - Degree centrality of node n :
 - $C_D(n) = \deg(n)$
 - Betweenness centrality of node n , summing over distinct nodes l and m :
 - $C_B(n) = \sum_{l \neq m \neq n \in V} \frac{\sigma_{lm}(n)}{\sigma_{lm}}$