

Gene Expression Profiling of Breast Cancer

Network Biology Mini Project

1. Mohammad Atif Quamar - 2020523
2. Aditya Choudhary - 2020489
3. Shreya Bhatia - 2020542

Introduction

- Gene expression profiling : Genetic microarray analysis of gene-expression dataset.
- It details the expression levels of thousands of genes in breast cancer & draws molecular portraits of breast cancer.

Problem Statement

Thorough Analysis and Biological Interpretation of Molecular Interactomes Linked with Breast Cancer through Network Analysis.

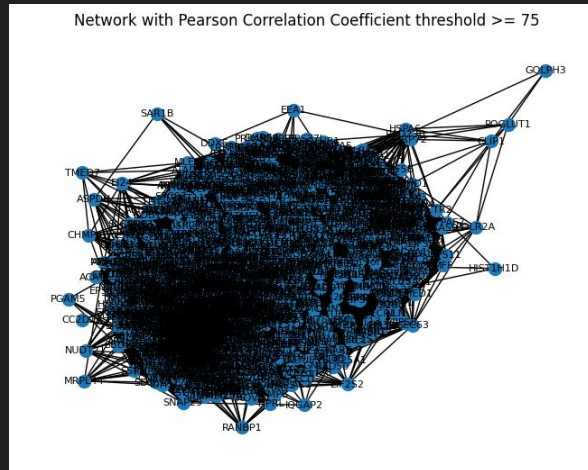
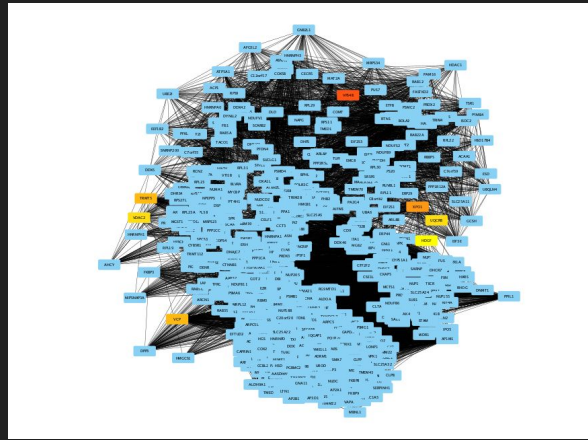
Deliverables

Microarray Dataset Preparation → Network Analysis → Visualization → Biological Interpretation → Result

Methods

Step 1: Data Preparation

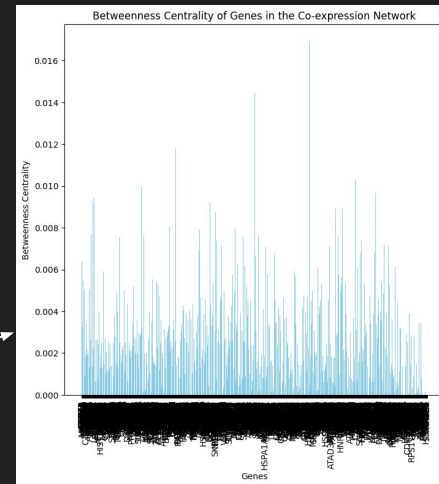
- Used breast cancer microarray dataset to generate the Expression - Correlation Matrix as it gave the similarity between different genes as the dataset gathered comprised of values for different cell lines.
- Used the metric of Pearson product-moment correlation coefficients between any genes in the microarray dataset to determine whether there is possible correlation between them.
- If the value of correlation coefficient was above a certain threshold (0.75), we considered it as an edge in our network.



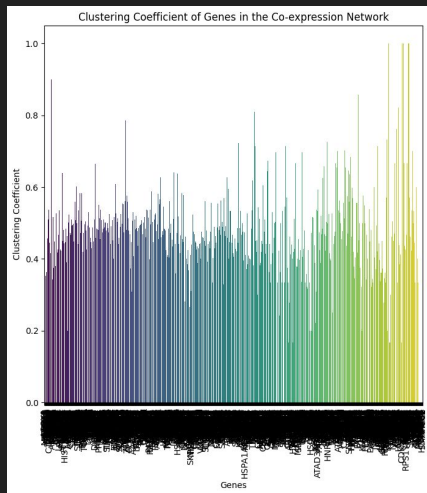
Step 2: Network Analysis

- With our network being prepared, we used the NetworkX library in python to calculate the following parameters and study the network based upon it:
 - Degree Distribution
 - Clustering coefficient of each node
 - Centrality between each node
 - Closeness centrality
 - Average clustering coefficient
 - Degree correlation coefficient
 - Closeness centrality
 - Identified hubs
 - Clustering Modularity of graph
- Visualized the graph on Cytoscape and Cytohub to study the data further

Degree Centrality
(peaks suggesting some genes having higher numbers of connections)



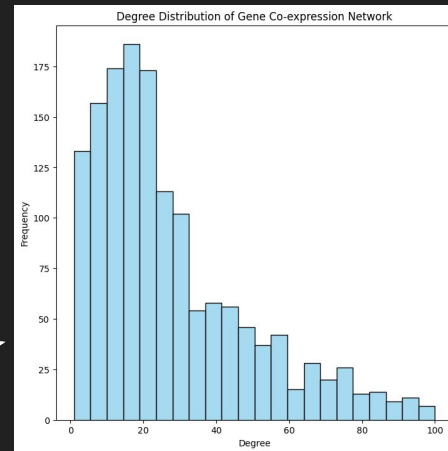
We created 4 distributions for our dataset. Each describing how the Gene nodes can be ranked.



Clustering Coefficient

(Peaks indicating important hubs)

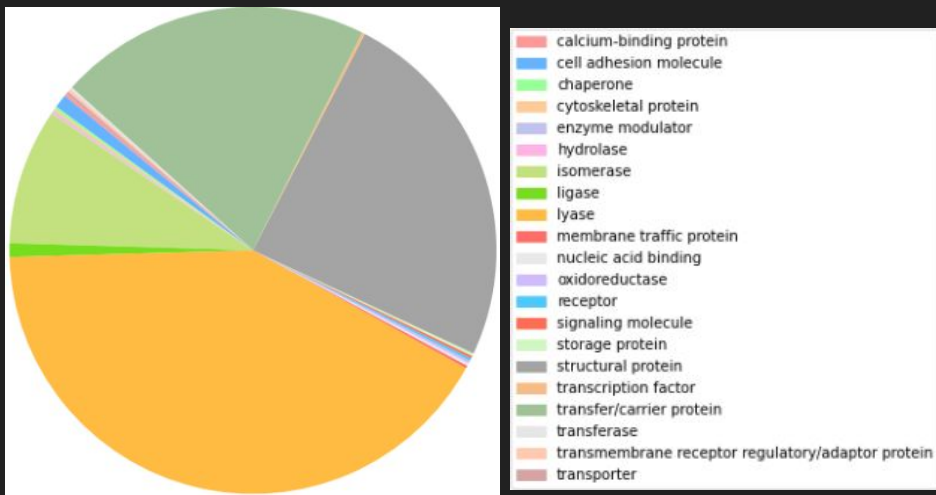
Degree Distribution (Power Law)



Observations

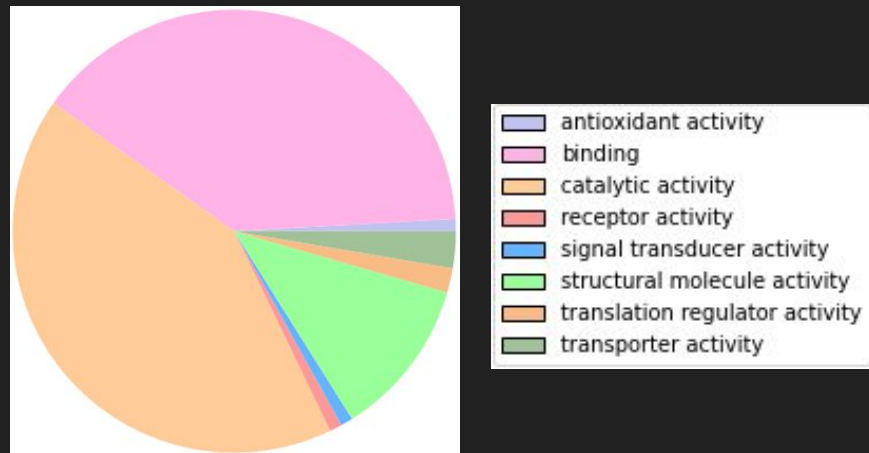
Cellular Analysis

- Majority of breast cancer genes possess “lyase” which is an enzyme which catalyzes the joining of specific molecule.
- Also, there exists transfer/carrier protein cellular function in most of genes.



Catalytic activity

- Most of the genes possess similar catalytic activity like PPIB, CTSB, GSPT1, TXNDC17 to name a few.



Network Observations

- Nodes follow a proper power law distribution.
- Most of the nodes have low clustering coefficient.
- Some nodes have extreme degrees and there is peak downfall for others.

Results

Some of the important hubs observed on the basis of connectivity are:

Gene Name	Function
TLN1	Plays a significant role in the assembly of actin filaments and in spreading and migration of various cell types, including fibroblasts and osteoclasts.
TUBB	Provides instructions for making of a protein called beta-tubulin, which is part of the tubulin family of proteins that form and organize cell structures called microtubules.
RANGAP1	Encodes a protein that associates with the nuclear pore complex and participates in the regulation of nuclear transport

Top 10 Genes in Degree Centrality

	Gene	Degree Centrality
696	TLN1	0.201743
482	NCLN	0.201072
617	CORO1C	0.201072
403	ANXA5	0.200402
509	PSMA7	0.199732
682	RANGAP1	0.199732
506	PRKACA	0.199062
396	ACTG1	0.197721
521	RQCD1	0.197721
703	TUBB	0.197721
598	AP1M1	0.196381
704	TUBB3	0.195710
659	OXSR1	0.195040
475	MSN	0.194370
518	RHOA	0.194370

Top 10 Genes in Betweenness Centrality

	Gene	Betweenness Centrality
1307	PHB2	0.004747
1306	C7orf55	0.003689
1025	NARS	0.003641
1269	MCM7	0.003488
717	C22orf28	0.003272
987	GNL1	0.003152
1037	ARL8B	0.003111
1213	PRDX3	0.002933
1226	PRPF38B	0.002931
1464	KHDRBS1	0.002925
143	ARL3	0.002903
1287	NRAS	0.002886
1046	STX12	0.002872
1045	NAA10	0.002806
219	TMX1	0.002747

Top 15 Genes in Closeness Centrality

	Gene	Closeness Centrality
703	TUBB	0.406761
843	NCAPG	0.406429
677	PSMD14	0.406097
682	RANGAP1	0.405325
617	CORO1C	0.405215
751	CAPZA1	0.404995
766	GSTO1	0.404885
627	DEK	0.404775
753	CNP	0.404665
854	STX7	0.403570
598	AP1M1	0.403461
696	TLN1	0.403461
785	PSME1	0.403461
838	KIF4A	0.403461
804	THOP1	0.403352

Top 15 Genes in Clustering Coefficient

	Gene	Clustering Coefficient
1492	GOLPH3	1.000000
1350	RPS8	0.867692
381	ASPDH	0.830882
1395	ATPAF2	0.813793
1461	PGP	0.812312
1182	ACAT1	0.789855
885	KPNA6	0.757595
990	MBLAC2	0.734921
1454	RPL29	0.731765
1068	MRPL44	0.730769
989	DDX1	0.719920
952	STMN2	0.711020
788	RAB11B	0.704856
1020	RANBP1	0.704762
1032	TPP2	0.702359

Top 15 Genes ranked in every category