Architecture and topologies of gene regulatory networks associated with breast cancer, adjacent normal, and normal tissues

Swapnil Kumar, Vaibhav Vindal *

Department of Biotechnology & Bioinformatics, School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad, 500046, India.

* Correspondence to: Dr. Vaibhav Vindal

Department of Biotechnology & Bioinformatics, School of Life Sciences, University of Hyderabad, Prof. CR Rao Road, Gachibowli, Hyderabad - 500046, India.

Email ID: vaibhav@uohyd.ac.in

Phone: +91-40-23134589

Abstract

Breast cancer is the most commonly diagnosed cancer and a leading cause of death in women worldwide. It is heterogeneous in nature, comprising multiple subtypes with varying clinical features, treatment responses, and insufficient treatment regimens. The present study aims to identify early-stage candidate genes for efficiently managing this disease. For this, first, we identified differentially expressed genes (DEGs) which followed the construction and analysis of seven gene regulatory networks (GRNs), including five subtypes (*viz.* Basal, Her2, LuminalA, LuminalB, and Normal-Like), one adjacent normal (ANT), and one normal tissue using the gene expression and regulatory information. Further, the DEGs of each subtype and ANT were analyzed using differential correlation, disease annotation, and functional enrichment analysis. The topologies of these GRNs enabled us to identify tumor-specific features of different subtypes and ANT networks. Further, GRN analysis using escape velocity centrality (EVC+) identified 24 common functionally significant genes, including well-known genes such as *E2F1*, *FOXA1*, *JUN*, *BRCA1*, *GATA3*, *ERBB2*, and *ERBB3* across subtypes and ANT responsible for the cancer development and progression. Similarly, the EVC+ also helped us to identify subtype-specific key genes (Basal: 18, Her2: 6, LuminalA: 5, LuminalB: 5, Normal-Like: 2, and ANT: 7). Moreover, the differential correlation and functional enrichment analysis also highlighted the cancer-associated role of these genes. Thus, it can be concluded that these common and ANT-specific genes have therapeutic potential, which can be explored further using *in vitro* and *in vivo* experiments to develop more accurate and effective early-stage diagnosis and treatment strategies for better disease management.

Keywords: Breast cancer subtypes, Adjacent normal tissues, Gene regulatory networks, Escape velocity centrality, Influential genes, Differential correlations, Switching genes.

1. Introduction

Breast cancer is the leading cause of death (15.5%) and also the most commonly diagnosed cancer (24.5%) in women worldwide, as per GLOBOCAN 2020 [1]. Despite the availability of various treatments, including hormone therapy, chemotherapy, radiotherapy, targeted therapy, and surgery for this disease, it is still the leading cause of concern. For example, patients diagnosed with early-stage tumors in about 30% of cases, even after removing the tumors, are prone to recurrence in distant organs [2]. Additionally, those diagnosed with late-stage and distant metastases are incurable [3].

Breast cancer is a heterogeneous disease with multiple subtypes. It can be stratified into five different molecular subtypes based on expression profiles of hormone receptors such as estrogen/progesterone receptor (ER/PR) and human epidermal growth factor receptor 2 (HER2). The five subtypes include Basal-like (ER-/PR- and HER2-), HER2-positive (ER+/PR+ and HER2+), Luminal A (ER+/PR+, HER2- and Ki67 low), Luminal B (ER+/PR+, HER2- or HER2+, and Ki67 high), and Normal-like (ER+/PR+, HER2-, and Ki67 low) [4].

These molecular subtypes of breast cancer exhibit varying biological features and hence respond differently to treatment processes leading to disparate clinical outcomes [5]. Further, these also occur at different rates and are more or less aggressive in nature, with variations in long-term survival rates [5]. A group of patients with different subtypes may have completely different molecular profiles and hence regulatory patterns of molecules. With the knowledge of the regulatory organization of molecules and behavior, it is possible to infer the causal mechanisms leading to the development and progression of different subtypes of the same cancer.

Despite the substantial advancement in the treatment process of breast cancer, to date, a proper treatment at the level of its different subtypes is still limited in terms of the lack of precise molecular targets for each one of them [6]. In this regard, some progress has been made toward identifying subtype-specific molecular targets or biomarkers [7-9]. However, these are inadequate to tackle subtype-specificity-based treatments' limitations efficiently. The heterogeneity in breast tumors makes treating them challenging, particularly for metastatic triple-negative patients, where available treatment options are limited [10, 11]. Therefore, exploring the mechanisms underlying the development and progression of different molecular subtypes of breast cancer is important. This may further lead to identifying more specific and novel

candidate genes at the subtype level to improve the treatment decisions for an early and more effective disease diagnosis.

The network theory-based approach allows us to investigate the vast amount of high-throughput data in a novel way [12, 13]. Genes and their products work together by interacting with each other to perform various essential cellular functions and processes. These genes, by interacting with each other, form a complex network. A diseased cell, including a cancerous cell, can be represented as a gene regulatory network (GRN) or a protein-protein interactions network (PPIN). The GRNs are very important for understanding the regulatory mechanism of genes and thus gene expression heterogeneity across normal and diseased cells. A comprehensive analysis of GRN representing cancerous cells can provide valuable insight into its behavior and may lead to the discovery of new and efficient biomarkers.

The study by Slaughter et al. on adjacent normal tissues (ANT) to tumors suggests the presence of an acquired step-by-step cumulative carcinogenic process comprising various genetic alterations, thus portraying the ANT as a pre-neoplastic, intermediate state which is composed of morphologically normal but altered cells at the molecular level [14]. Another study on ANT has also demonstrated the differentiating characteristics of ANT from healthy tissues [15]. Hence, the detailed exploration of the GRNs of ANT and different breast cancer subtypes can help to identify key candidate genes that can be considered early-stage specific genes. These early-stage-specific candidate genes can be utilized further to develop more efficient novel biomarkers and therapeutics for the disease.

In the present study, first, we identified differentially expressed genes (DEGs) across different subtypes of breast cancer and adjacent normal tissues using gene expression profiles of the tumor, adjacent normal, and normal tissues of the human breast. It followed the construction and subsequent analysis of GRNs of each subtype, adjacent normal (ANT), and normal tissues (NBT) using the regulatory information of transcription factors (TFs) and target genes available in various public databases. Furthermore, the TF-gene regulatory information for these tissue types was also predicted using expression profiles of DEGs along with prior data such as PPI and motif. Besides, all DEGs of each subtype were also analyzed using differential correlation, disease annotation, and functional enrichment analysis. This comprehensive study helped us identify breast cancer's condition and subtype-specific features. The structural properties of GRNs also enabled us to identify some functionally significant genes responsible for the

development and progression of the disease. These important genes can be explored further using *in vitro* and *in vivo* experiments to develop a more accurate and effective treatment strategy for better disease management.

2. Methods

2.1 Gene expression data

Gene expression profiles of breast tumor tissues and adjacent normal tissues (ANT) from The Cancer Genome Atlas (TCGA) project [16], along with gene expression profiles of normal breast tissues (NBT) from the Genotype-Tissue Expression (GTEx) database [17] were accessed through the recount2 project [18]. The recount2 project provides uniformly processed and quantified RNA-seq data derived from the GTEx, Sequence Read Archive (SRA), and TCGA projects for various downstream analyses, including differential expression analysis. All these three datasets were pre-processed and normalized for further use using the TCGAbiolinks R/Bioconductor package [19].

The gene expression data of tumor tissues were further subdivided or stratified into expression profiles for five different subtypes of breast cancer, *viz.*, Basal, Her2, Luminal A (LumA), Luminal B (LumB), and Normal-like (NormL), based on the sample's information as available from the Genomics Data Commons (GDC) portal of TCGA.

2.2 Identification of differentially expressed genes

To identify differentially expressed genes (DEGs) in each subtype, gene expression data of tumor tissues of all five subtypes were compared with gene expression profiles of normal breast tissues using the DESeq2 [20], edgeR [21], and limma R package [22]. Similarly, DEGs across both adjacent normal and normal tissues compared to tumor tissues, including all five subtypes, were also identified. These resulted in a list of significantly up and down-regulated genes across all conditions. Only significantly differentially expressed genes (up and down) as per any two tools out of three were considered for further analysis. The expression profiles of the DEGs of different subtypes were extracted from the corresponding tumor tissue data for subsequent analysis. Further, the expression data of these DEGs from normal tissues were also extracted for the downstream analysis.

2.3 Construction of subtype and normal networks

The DEGs of each of five subtypes (Basal, Her2, LumA, LumB, and NormL), along with DEGs of ANT and a list of breast tissue genes, were used to construct gene regulatory networks using

seven different databases *viz.*, GRAND [23], GRNdb [24], hTFtarget [25], ICON [26], ORTI [27], RegNetwork [28], and TRRUSTv2 [29]. The GRAND database hosts several gene regulatory networks derived from cell lines, tumor tissues, normal tissues, and small molecules. Out of these GRAND gene regulatory networks, only TF-gene regulations predicted using the gene expression data of breast tumor tissues from TCGA and normal breast tissues from GTEx were considered in the current study. These GRAND regulatory data have more than one million TF-gene regulations; therefore, only the top 10% of TF-gene regulations based on the edge score > 0 were used to identify regulations of DEGs for subtypes and of breast tissue genes for NBT network constructions from the TF-gene regulations of the breast tumor tissues (TCGA) and normal breast tissues (GTEx). The list of breast tissue genes was retrieved from The Human Protein Atlas [30] to construct the NBT network. Further, there was no TF-gene regulatory information on ANT in the GRAND database; therefore, the gene expression profile data of 19,061 protein-coding genes across 111 samples from ANT was used to predict TF-gene regulations with the help of the PANDA R package [31]. From the predicted TF-gene regulations of ANT, only the top 10% of TF-gene regulations based on the edge score > 0 were considered to identify regulations of DEGs for ANT network construction.

2.4 Regulatory network analysis

All five subtype networks (Basal, Her2, LumA, LumB, and NormL) along with two normal breast networks (ANT and NBT) were analyzed using the igraph R package [32] by assessing various topological properties of these networks. The topological properties considered for the present analysis were degree, clustering coefficient, betweenness centrality, average path length, diameter, eigen centrality, and heterogeneity. Further, corresponding random networks were constructed for all real networks (five subtypes, one ANT, and one NBT) with the same degree sequences present in the original networks. These random networks were used to analyze the structure and organization of corresponding real networks by comparative analysis of their different topological properties.

Degree: The total number of nodes directly connected with a node of interest is called the Degree of that node. It is the most basic property of a node participating in any network, and the fraction of nodes having degree value k is called Degree distribution, i.e., $P(k)$. Highly connected genes (hubs) in biological networks are thought to play key roles in the behavioral organization of the networks [33-36].

Betweenness centrality: The fraction of shortest paths between all the nodes' pairs passing through a node of interest is defined as the Betweenness centrality of that node [37].

Clustering coefficient: The ratio of the number of triangles a node of interest makes with its neighboring nodes and the total number of possible triangles that node can have is defined as the Clustering coefficient [38, 39].

Average path length: It is the average number of edges along the shortest paths for all possible node pairs in a network. It is also known as the Characteristic path length or the Average shortest path length [38].

Diameter: It is the length of the longest "shortest path" connecting any two nodes present in the network. The diameter of a network describes its interconnectedness and the ability of nodes to communicate with other nodes in the network [33, 40].

Eigen centrality: It is the value of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix of a network, also known as eigenvector centrality [41, 42]. It measures the influence of nodes in a network. If a node is connected to the high-scoring nodes, have more contributions to the score of the node of interest compared to the equal number of connections with low-scoring nodes. A high eigenvector score of any node means that the node is connected with other nodes having high scores.

Heterogeneity: It is defined as the coefficient of variation of the connectivity distribution [43]. Biological networks are very heterogeneous in nature; because some nodes are highly connected while the majority of nodes are less or very less connected.

2.5 Identification of influential genes

All subtype and ANT networks were analyzed to identify influential genes with the help of extended escape velocity centrality (EVC+). As proposed by Ullah et al., it is an extended version of the escape velocity centrality (EVC) which considers the number of connections a node can have along with its position in the network to identify key players (nodes) of the network. Thus, it provides a better ranking than any other, i.e., baseline network properties, including the degree, betweenness, and clustering coefficient of nodes based on the connections and positions of nodes participating in a network under study [44].

The EVC+ of node 'i' can be defined as follows:

$$\text{EVC+ (i)} = \sum_{j \epsilon \psi} \sqrt{\frac{2\alpha(dv_i(ks_{vi} + ks_{vj}))}{d_{ij}}}$$

Where, α is a tunable factor which controls the degree effect, and its values range between 0.1 - 1, dv_i is the degree of node 'i', ks_{vi} and ks_{vj} are k-shell values of node 'i' and 'j' respectively, d_{ij} is the shortest path between two nodes 'i' and 'j' using Dijkstra algorithm, and $j\square_{\psi}$ is the set of neighbors of node 'i'.

2.6 Differential correlation analysis

Network edges coupled with expression profiles of DEGs in tumor conditions (Basal, Her2, LumA, LumB, and NormL) were analyzed by comparing with normal conditions (NBT) for each subtype of breast cancer using the DiffCorr R package [45]. This resulted in a list of gene pairs (edges) significantly differentially correlated for each subtype in the tumor condition compared to the normal condition. From the list of differentially correlated gene pairs of each subtype, gene pairs showing switching behavior between the tumor and normal conditions were filtered out. Gene pairs with positive correlations in tumor tissues, while negative correlations in normal tissues and vice-versa are considered gene pairs with switching behavior.

2.7 Functional enrichment analysis

All DEGs in the case of subtypes and ANT were further subjected to functional and pathway enrichment or over-representation analysis to decipher significantly over-represented terms *viz.* biological processes (BPs), molecular functions (MFs), and pathways (KEGG) associated with these genes. The functional and pathway enrichment analysis of DEGs was carried out using the clusterProfiler v4.4.4 [46] and visualized with the enrichplot R package [47].

2.8 Disease ontology and gene-disease associations

The DEGs of subtypes and ANT were also checked for their associations with various cancers, including breast cancer. The gene-disease associations using the DisGeNet database [48] and disease enrichments using Disease Ontologies (DO) were inferred with the help of the clusterProfiler R package. The DisGeNet database hosts information on gene-disease associations as reported in previously published literature. All the enrichment terms with an adjusted p-value less than 0.05 were considered significant.

3. Results

3.1 Expression profiles of subtypes and normal tissues

The gene expression profiles of breast tumor tissues contained expression values for 56,002 genes across 1,210 samples. The expression profiles of normal breast tissues (NBT) contained expressions of 56,002 genes across only 92 samples. Among tumor tissue samples, there were 200 samples of Basal, 82 samples of Her2, 570 samples of LumA, 207 samples of LumB, 40 samples of NormL, and 111 samples of adjacent normal tissues (ANT). We considered only the expression profiles of 19090 protein-coding genes reported by the Human Gene Nomenclature Committee (HGNC) for the current study.

Before further analysis, these expression profiles were pre-processed and normalized, as discussed in the Methods section. After normalization, the data contained expression profiles of only 19079 genes in the case of tumors and normal tissues and 19060 genes in adjacent normal.

3.2 Differentially expressed genes

The differentially expressed genes (DEGs) were identified with |FoldChange| > 1.5 and adj P-value < 0.05 across different subtypes, adjacent normal, and normal tissues using three commonly used R packages *viz.* DESeq2, edgeR, and limma. All subtypes and adjacent normal tissues were compared with normal breast tissues to identify DEGs. However, normal breast tissues were compared with all tumor tissues combined (including all subtypes) to identify DEGs across normal breast tissues. Thus, significantly up and down-regulated genes were selected for further analysis as per any two out of the three tools across all seven tissue types, including adjacent normal and normal tissues. The distributions of significantly up and down-regulated genes across all tumor and normal tissue samples are available in Table 1.

Out of all DEGs in the Basal subtype, there were 792 genes unique to Basal. Similarly, Her2, LumA, LumB, and NormL have 337, 150, 233, and 169 unique genes, respectively. Furthermore, the ANT also has 359 unique genes. These unique genes were present only in respective tissues but not in others.

Table 1. Distribution of significantly differentially expressed genes across all tissue samples with |FoldChange| > 1.5 and adj. P-value < 0.05.

Tissues	**DESeq2**		**edgeR**		**limma**		**Common in Any Two methods**		**Total DEGs (Common in Any Two methods)**
	Up	**Down**	**Up**	**Down**	**Up**	**Down**	**Up**	**Down**	
Basal	4883	3028	4700	3169	3783	4032	4684	3162	7846

HER2	4630	3699	4639	3744	3949	4309	4603	3741	8344
LumA	3987	3231	3896	3314	3343	3724	3861	3314	7175
LumB	4600	3347	4579	3411	3850	4027	4484	3433	7917
NormL	3289	2778	3383	2808	2884	3122	3285	2805	6090
ANT	2192	2090	2263	2132	2115	2187	2196	2117	4313

3.3 Gene regulatory networks of subtypes, adjacent normal, and normal tissues

Gene regulatory networks (GRN) for all five subtypes, adjacent normal and normal tissues were constructed by retrieving interactions data from five publicly available databases as discussed in the Methods section. The distributions of extracted edges across all five subtypes and normal tissues are available in Table 2. The numbers of nodes and interactions (edges) present among them and in the largest connected components across all seven networks are described in Table 3.

Table 2. Distribution of edges collected from nine different sources for all seven tissue types

Source	**Basal**	**Her2**	**LumA**	**LumB**	**NormL**	**ANT**	**NBT**
GRAND	282347	328704	254804	293235	168693	30457	183875
GRNdb	47045	46191	39753	44378	26279	5547	81421
hTFtarget	44031	51656	41270	50848	29235	18115	157885
ICON	1146	1468	1117	1279	826	402	3376
ORTI	16713	16960	18253	20067	10827	6589	43897
RegNetwork	34367	40219	29956	38608	18736	9944	123603
TRRUST	2419	2735	2207	2610	1577	813	6898

Table 3. Distribution of edges and vertices participating in the networks and largest connected components of the networks across all five subtypes and normal tissues

Network	**N_V**	**N_E**	**N_V (LCC)**	**N_E (LCC)**	**alpha**
Basal	6100	32885	6099	32884	2.581
Her2	6558	36240	6556	36238	2.564
LumA	5641	32941	5639	32939	2.708

LumB	6307	37652	6305	37650	2.628
NormL	4536	20455	4534	20453	2.598
ANT	2514	5826	2506	5820	2.887
NBT	11463	45952	11462	45951	1.769

3.4 Topological properties of networks

All networks, including ANT and NBT, were assessed based on different topological properties like degree, betweenness centrality, clustering coefficient, average path length, diameter and heterogeneity (Table 4). By analyzing these properties, we were able to draw some interesting inferences from these networks. All five subtypes and one ANT network follow the power-law of degree distribution with alpha values between 2 to 3 (Table 3); however, the NBT network exhibit less than two, which may be due to the effect of the finite size observed in the case of other biological networks [49]. This proves that all the networks constructed in the present study are significant and represent true biological networks in terms of the structural organization of these networks, similar to other biological networks. The average degree values of all subtype networks are higher than the average degree values of normal network NBT. However, the average degree value of ANT is lower than that of all subtypes and NBT networks. Further, the average betweenness values of all five subtypes and one ANT network, *viz.*, Basal, Her2, LumA, LumB, NormL, and ANT, are lower than normal network NBT. However, the average betweenness value of the ANT network is the lowest among all networks. Next, all subtypes and ANT networks exhibited higher average values of clustering coefficient (<CC>) and eigenvector centrality (<EC>) than the normal network NBT. Moreover, the number of nodes with clustering coefficients equal to one and zero (CC = 1 and CC = 0) present in all five subtype networks was less than that of the normal network NBT. The number of nodes with CC = 1 and CC = 0 present in ANT was also less than that of the normal network NBT. The average path lengths of all five subtype networks and one ANT network showed lower values than that of the normal network (NBT). Furthermore, the values of diameters of all five subtype networks and one ANT network were also less than that of the normal network NBT. Similarly, the heterogeneity values of all five subtype networks and one ANT network were lower than that of the normal network NBT.

When these properties of random networks were compared with those of the corresponding real networks, they were found to deviate significantly. Hence, the comparative analysis of real networks with corresponding random networks indicated that they indeed possess some intrinsic properties leading to their diseased and healthy behavior.

Table 4. Various topological properties of all subtype and normal networks

Network	**<k>**	**<BC>**	**<EC>**	**<CC>**	**CC=1**	**CC=0**	**APL**	**D**	**HC**
Basal	10.783	364.855	0.128	0.541	781	335	3.056	7	7.855
Her2	11.055	416.099	0.115	0.562	925	362	3.128	7	8.098
LumA	11.683	323.677	0.144	0.550	692	284	3.003	7	7.722
LumB	11.943	348.718	0.134	0.568	838	292	2.930	7	7.742
NormL	9.022	248.568	0.145	0.545	668	289	2.995	8	7.831
ANT	4.645	148.833	0.138	0.556	486	282	3.355	8	6.936
NBT	8.018	759.965	0.065	0.426	1162	1645	3.552	10	10.032

Average degree, <k>; Average betweenness centrality, <BC>; Average eigenvector centrality, <EC>; Average clustering coefficient, <CC>; Number of nodes with CC = 1, N (CC=1); Number of nodes with CC = 0, N (CC=0); Average path length, APL; Diameter, D; Heterogeneity centrality, HC.

The degree distribution plots of all seven networks showed that all, including tumor, adjacent normal, and normal networks, follow the power law of degree distribution which is an intrinsic feature of true biological networks (Figure 1A - 1G). Next, the degree-betweenness and degree-correlation coefficient plots of all seven networks (Figure 2A - 2G and Figure 3A - 3G) revealed that these networks follow positive and negative correlations, respectively. Several other biological networks have also exhibited the same properties of degree-betweenness and degree-clustering coefficient correlations.

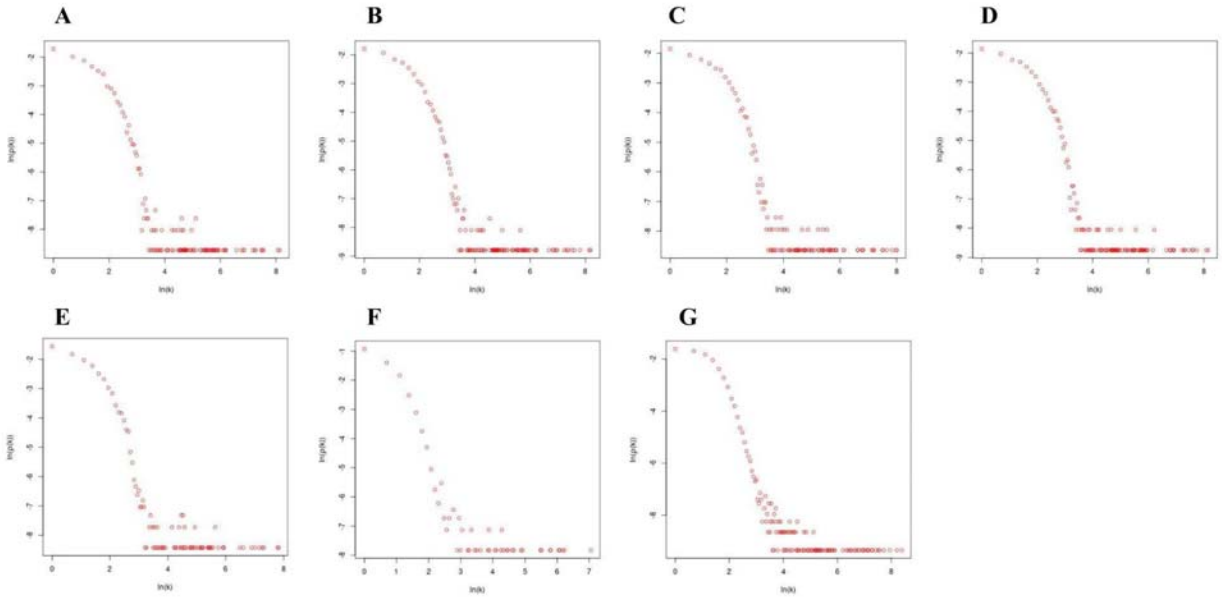

Figure 1. Degree distribution plot of (A) Basal, (B) Her2, (C) LumA, (D) LumB, (E) NormL, (F) ANT, and (G) NBT network.

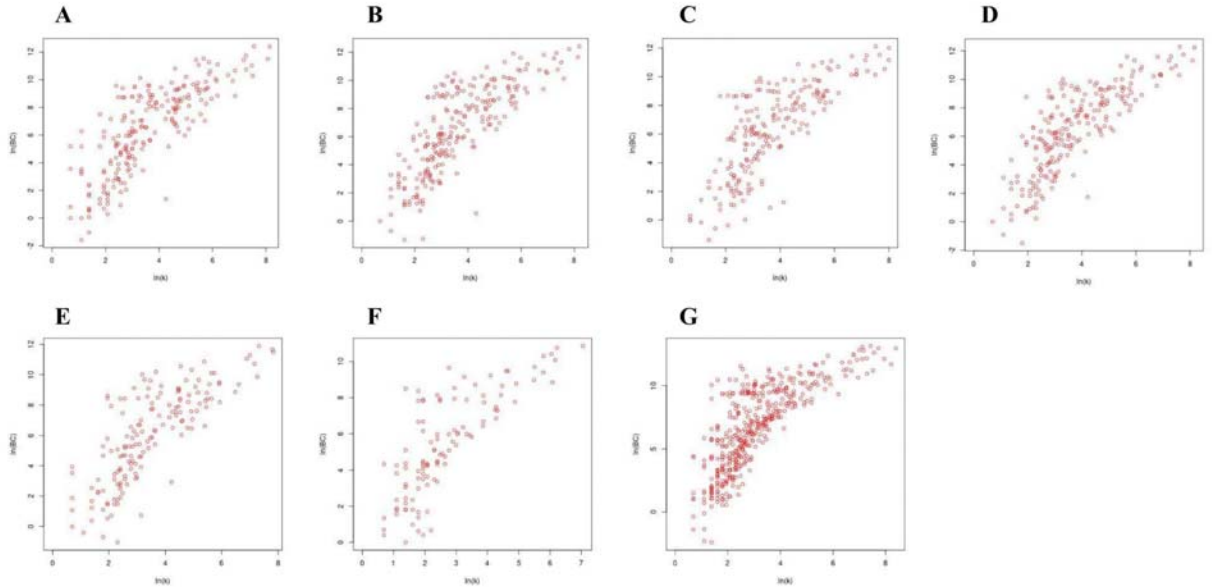

Figure 2. Degree-Betweenness centrality plot of (A) Basal, (B) Her2, (C) LumA, (D) LumB, (E) NormL, (F) ANT, and (G) NBT network.

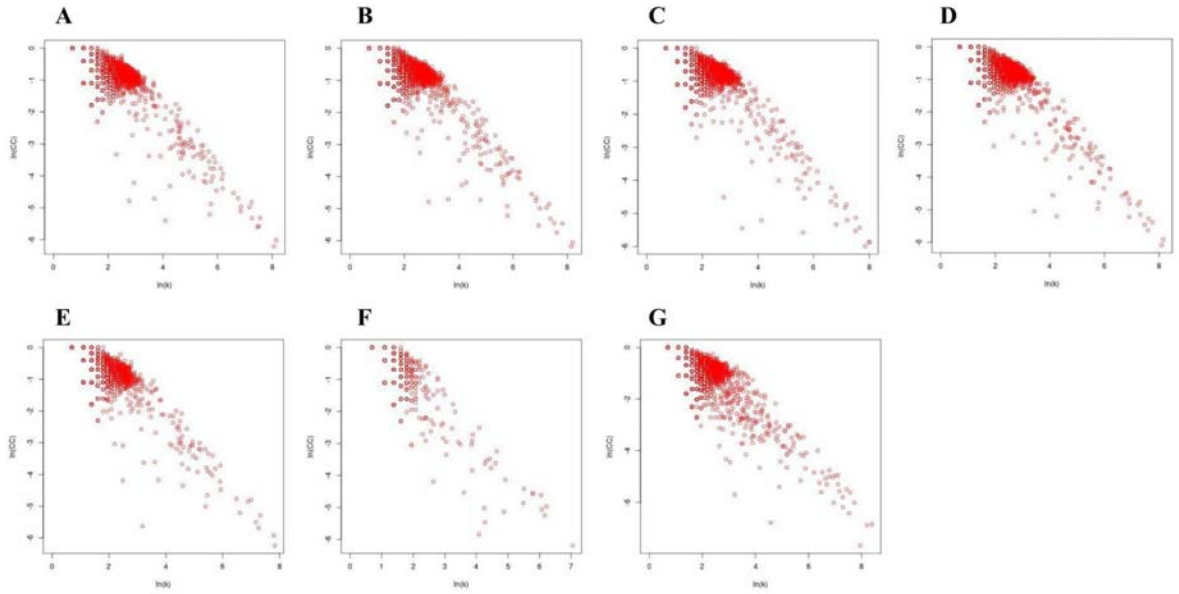

Figure 3. Degree-Clustering coefficient plot of (A) Basal, (B) Her2, (C) LumA, (D) LumB, (E) NormL, (F) ANT, and (G) NBT network.

3.5 Influential genes across tissue types

All subtype and ANT networks were analyzed using EVC+, and key influential genes were identified for each one of them. Based on the 80:20 principle, mean values of EVC+ of the top twenty percent genes were considered as cutoff values to determine influential genes across different tissue types. The distribution of the total number and tissue-specific number of influential genes are available in Tables 5 and S1. There were 24 influential genes that were commonly present in all tissue types, *viz.*, Basal, Her2, LumA, LumB, NormL, and ANT. Among these common influential genes, genes such as *E2F1*, *FOXA1*, *JUN*, *AR*, *GATA3*, *BRCA1*, and *ERBB2* were influential among the top key genes based on the value of EVC+ across all subtype networks (Table S1). Whereas *ESR1* was present only in Basal, LumA, LumB, NormL, and ANT, but not in Her2. Similarly, *TFAP2A* was present in four subtypes, *viz.* Basal, Her2, LumA, and LumB, but not in NormL.

Table 5. Distribution of influential genes across different subtypes and ANT

Network	**Total Number of Influential Genes**	**Specific Influential Genes**	**Common Influential Genes**

Basal	237	18 (*ATP6V0A1, BCAR3, CAMTA1, DAPK1, ECE1, ELF1, ELF2, FOXP1, HNRNPL, IGF2BP3, MAD1L1, MEF2D, NFKB2, NUMA1, PHF12, PRMT1, RABGAP1L, SPIB*)	24 (*E2F1, FOXA1, JUN, AR, KLF4, GATA3, STAT1, STAT2, NR2C2, KLF16, BRCA1, MAZ, MZF1, TCF7L2, CEBPA, SREBF2, GRHL2, TAL1, BCL6, DDIT3, ERBB3, CDH1, CCND1, ERBB2*)
Her2	254	6 (*ARNT, CALM1, MED24, MYO1B, NFIA, SMAD7*)	
LumA	206	5 (*AKAP1, MYO18A, PBX3, SSBP3, TFAP2C*)	
LumB	229	5 (*AHR, COL4A2, PTK2B, REL, TP73*)	
NormL	145	2 (*IRF8, RUNX1*)	
ANT	78	7 (*APC, BMPR2, NFIC, TNS3, USF1, ZBTB7A, ZNF436*)	

3.6 Differential correlation analysis and switching gene networks

After differential correlation analysis of each subtype and adjacent normal tissues with normal tissues at FDR 0.05 and correlation-coefficient (r) 0.4, we identified the significant number of switching gene pairs present in subtype networks and corresponding normal networks. We identified a total of 525 switching gene pairs involving 380 genes in the basal subtype and 460 gene pairs involving 362 genes in the normal counterpart of the basal subtype (Figure 4 and Table S2). In the case of Her2, a total of 428 switching gene pairs involving 465 genes were found. On the other hand, 444 switching gene pairs involving 480 genes were present in the normal counterpart of Her2 (Figure 5 and Table S3). Next, the LumA subtype had a total of 176 switching gene pairs comprising 175 genes, whereas the normal counterpart of the LumA had 185 switching gene pairs with 187 genes (Figure 6 and Table S4). Further, the LumB contained 186 switching gene pairs involving 206 genes and its normal counterpart 181 switching gene pairs with 193 genes (Figure 7 and Table S5). Similarly, the NormL subtype had 813 switching gene pairs involving 677 genes, and the normal network had 824 switching gene pairs among 678 genes (Figure 8 and Table S6). The ANT also exhibited 105 switching genes which involved

120 genes, whereas 108 switching gene pairs were present in the normal counterpart of the ANT involving 115 genes (Figure 9 and Table S7).

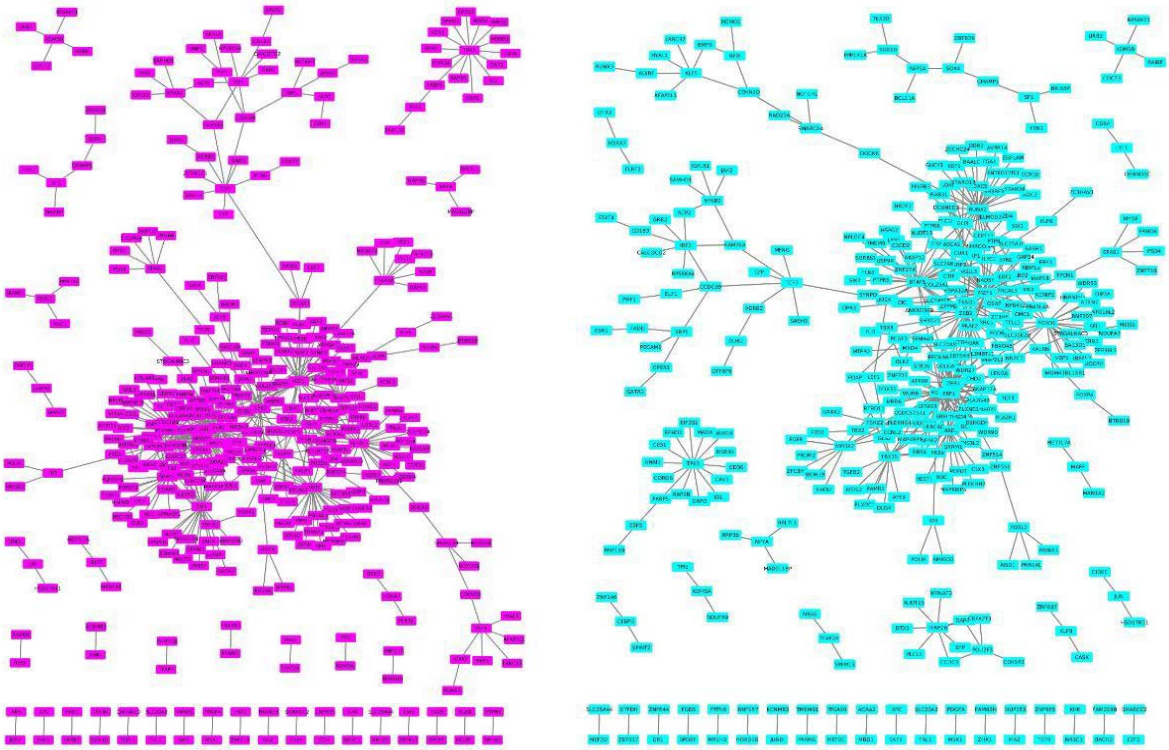

Figure 4. Switching gene regulatory networks of Basal tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

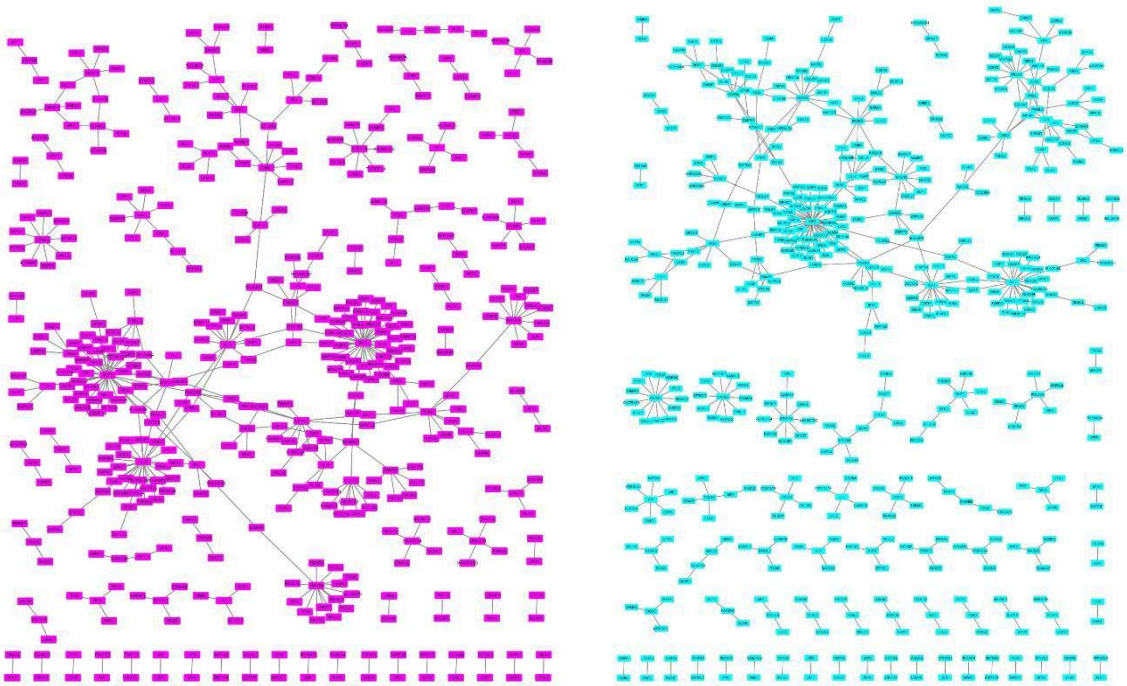

Figure 5. Switching gene regulatory networks of Her2 tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

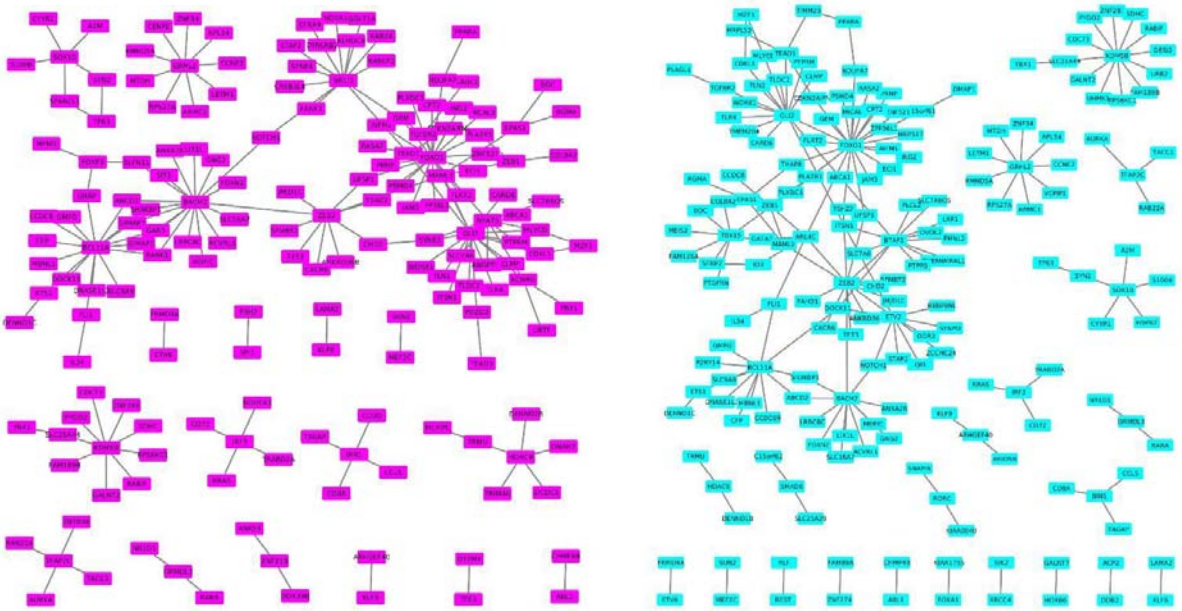

Figure 6. Switching gene regulatory networks of LumA tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

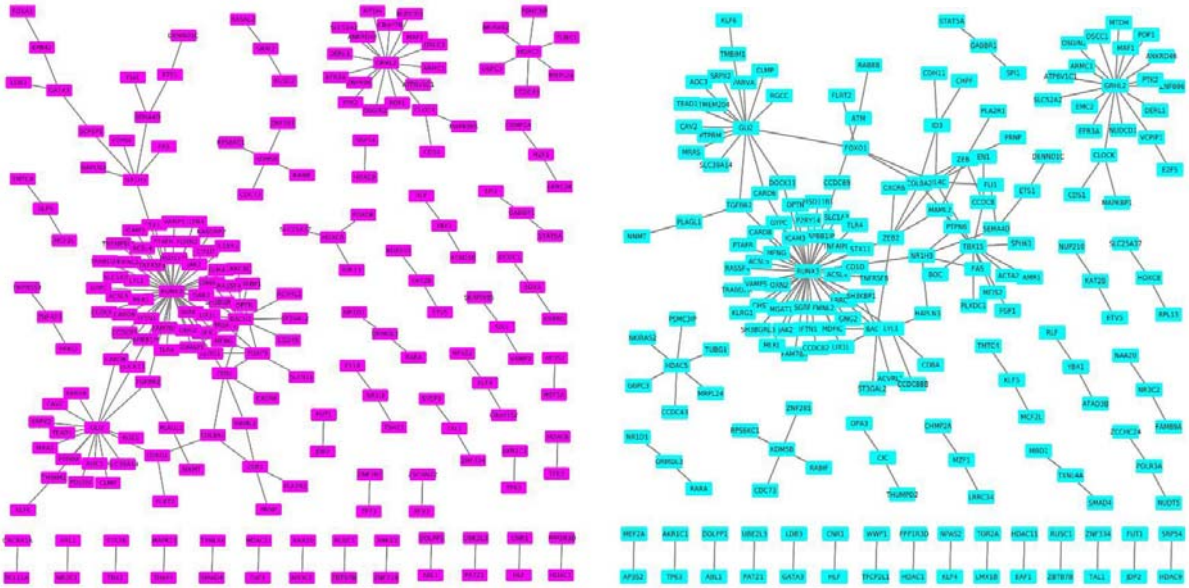

Figure 7. Switching gene regulatory networks of LumB tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

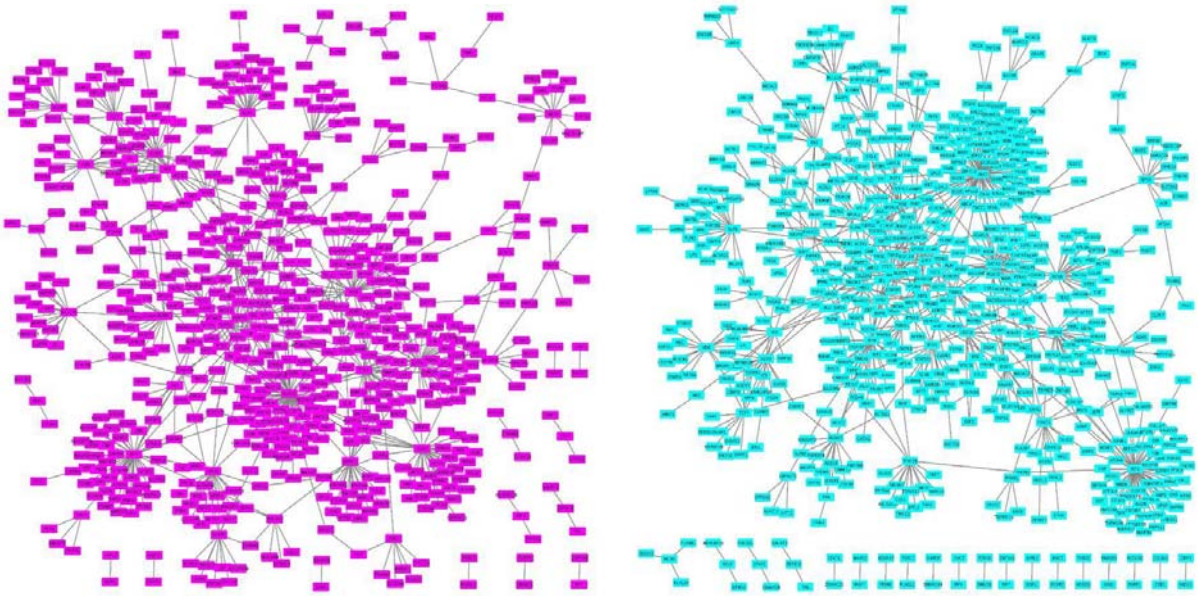

Figure 8. Switching gene regulatory networks of NormL tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

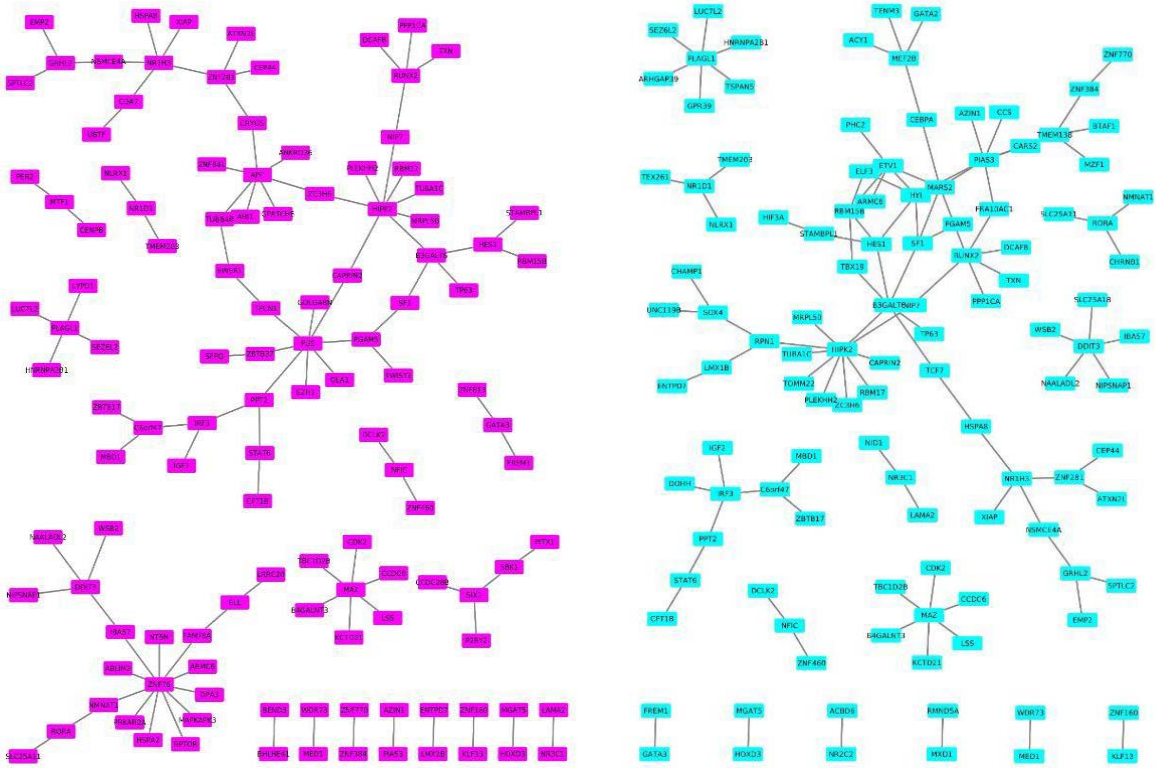

Figure 9. Switching gene regulatory networks of adjacent normal tissues (purple color nodes, left panel) and corresponding normal counterparts (cyan color nodes, right panel).

3.7 Functional annotation of DEGs

The DEGs identified across all five subtypes were subjected to functional and pathway enrichment analysis to infer the biological processes (BPs) these genes are participating in, the molecular functions (MFs) these genes perform, and the pathways they are involved. The details of the number of significant GO terms (*viz.* BP and MF) and pathway terms (*viz.* KEGG, Reactome, and Wiki pathway) are available in Table 5. The dot plots in figures 6 to 9 represent the top significantly over-represented GO (BP and MF) terms and pathway (KEGG and Wiki pathway) terms.

Biological processes: There were a total of 2542 significant BP terms associated with Basal, Her2, LumA, LumB, NormL, and ANT (Table 6). Among these enriched terms, BP terms like skeletal system development, external encapsulating structure organization, extracellular matrix organization, extracellular structure organization, positive regulation of cell adhesion, cell-substrate adhesion, leukocyte cell-cell adhesion, urogenital system development, and renal system development were the most significantly enriched terms across all five subtypes and ANT (Figure 10 (A), Table S8).

Molecular functions: There were a total of 573 molecular function terms associated significantly with DEGs across all five subtypes and ANT (Table 6). Of these enriched terms, MF terms like actin binding, glycosaminoglycan binding, extracellular matrix structural constituents, heparin binding, integrin binding, growth factor binding, and transmembrane receptor protein kinase activity were among the significantly enriched terms with a high number of genes as compared to others (Figure 10 (B), Table S9).

KEGG pathways: A total of 404 KEGG pathway terms were found significantly distributed among all five subtypes (Basal, Her2, LumA, LumB, and NormL) and ANT. Among all enriched pathway terms, Her2 had the highest number of terms. While ANT had the least number of enriched pathway terms (Tables 6 and S10). Among all significantly enriched KEGG pathways terms, Rap1 signaling pathway, Focal adhesion, Proteoglycans in cancer, Cell adhesion molecules, Phagosome, Cell cycle, and ECM-receptor interaction related terms were found associated with all subtypes and ANT. Whereas the PI3K-Akt signaling pathway was associated with all except Basal subtypes, and the Calcium signaling pathway was associated with only ANT (Figure 10 (C), Table S10).

Table 6. Distribution of functional and pathway enrichment terms over-represented among DEGs across all five breast cancer subtypes.

Subtype	GO-BP	GO-MF	KEGG
Basal	462	116	71
Her2	492	123	102
LumA	457	99	77
LumB	453	120	66
NormL	475	83	85
ANT	203	32	3

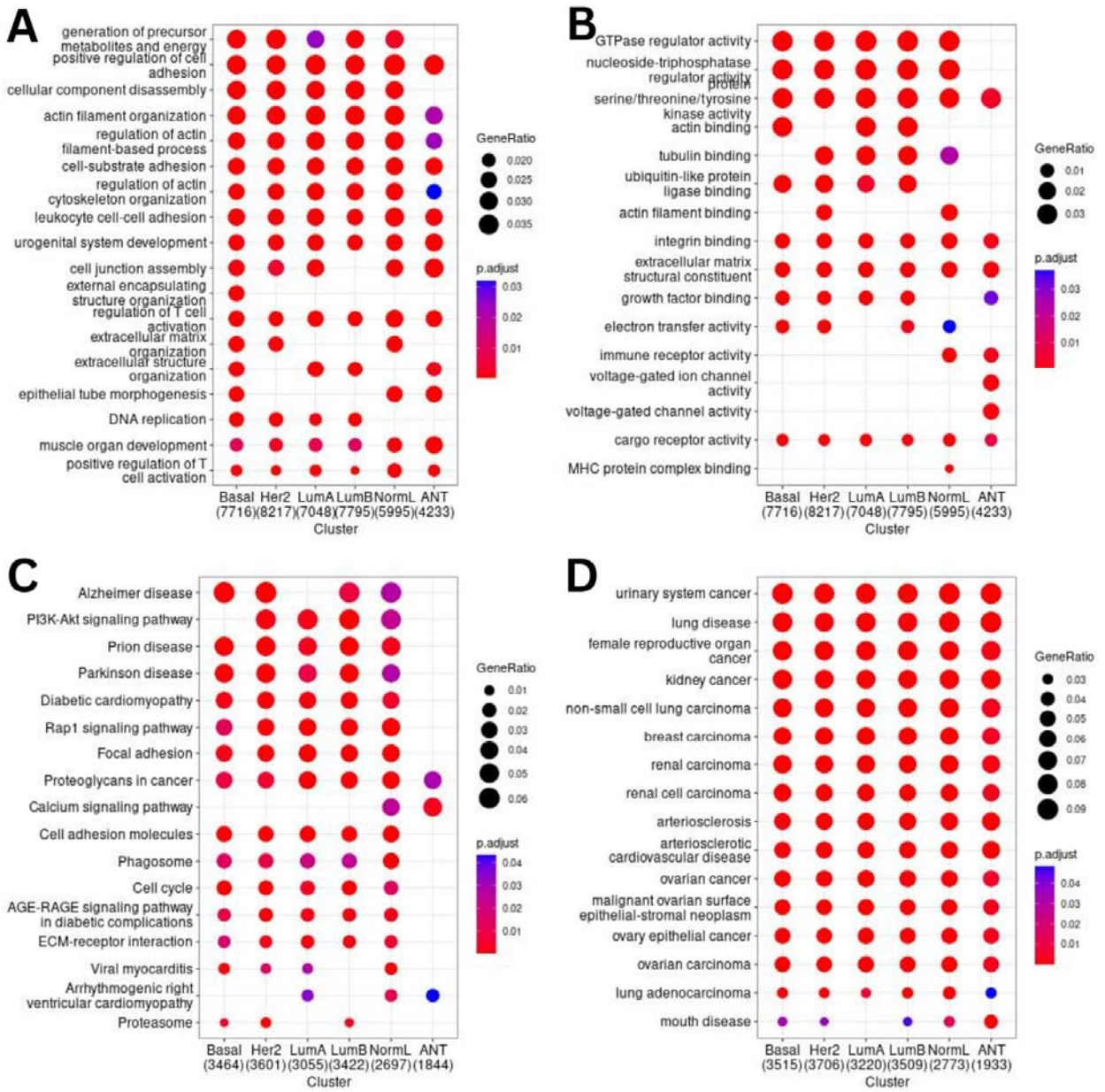

Figure 10. Distribution of top biological process (A), molecular function (B), KEGG pathway (C), and disease ontology (D) terms based on gene counts associated with DEGs across all five subtypes and ANT.

3.8 Disease ontology and disease-gene associations

The DEGs across all five different subtypes and ANT of breast cancer were subjected to the enrichment of Disease Ontology (DO) terms using the clusterProfiler. Further, these DEGs were also looked for their associations with various diseases using the disease-gene network (DGN)

association functionality of clusterProfiler based on the DisGENET database. There were a total of 1029 DO terms and 17429 DGN terms significantly associated with all these DEGs across different subtypes and ANT (Table 7).

Disease ontology enrichment of DEGs showed significant associations of most of them with several cancers like urinary system cancer, female reproductive organ cancer, kidney cancer, lung cancer, breast cancer, and ovarian cancer (Figure 10 (D) and Table S11). Similarly, the Disease-Gene association analysis also showed significant associations of these DEGs with cancers like Non-infiltrating intraductal carcinoma, Invasive carcinoma of breast, Metastatic malignant neoplasm to brain, Basal-like breast carcinoma, and Anaplastic carcinoma. Some other diseases, e.g., Infection, Hepatitis C, and Adult meningioma, were also found to be significantly associated with DEGs of different subtypes and ANT (Table S12).

Table 7. Distribution of disease ontology and disease-gene terms across five subtypes and ANT.

Enrichment	Basal	Her2	LumA	LumB	NormL	ANT
DO	182	184	204	153	209	97
DGN	3210	3386	2909	2770	3204	1950

4. Discussion

The present study explores comprehensive gene regulatory networks (GRN) of tumor and normal tissues of the female breast to uncover tumorogenic features of adjacent normal tissues and identify key novel genes responsible for the pathogenesis of breast cancer. While the tumor tissues involve five different subtypes of breast cancer, the normal tissues involved morphologically normal adjacent tissues (ANT) and healthy normal tissues (NBT). Each subtype and ANT exhibited different sets of differentially expressed genes (DEGs) and indicated gene expression profile differences among them. The *TP53* was among DEGs across all subtypes but not in the case of ANT. The DEG *MYC* was present in Her2, LumA, and LumB but not in Basal, NormL, and ANT. Some well-known genes of breast cancer, *viz.*, *BRCA1*, *ESR1*, and *ERBB2* were too differentially expressed in ANT in addition to all subtypes.

The topological analyses (diameter, betweenness centrality, clustering coefficient, heterogeneity, and average path length) of GRNs of five subtypes, one ANT and one NBT, clearly highlighted the key differences between tumor and normal tissues. The less number of nodes having clustering coefficients' values zero and one in tumor and adjacent normal tissue compared to

normal tissue network indicated that the tumor and adjacent normal networks have similar patterns of wiring among TF and target genes with a decrease in the number of complete sub-graphs and terminal nodes. Similar findings have been reported in cervical cancer [50] and breast cancer [51], involving protein-protein interaction network-based studies of tumor and normal conditions. Nodes with CC equal to one form complete sub-graphs, and these sub-graphs are building blocks of a network. Thus, from the above-mentioned result, it can be inferred that the tumor networks are disrupted compared to the normal ones. Similarly, the network of adjacent tissues also has fewer building blocks and hence, is disrupted. Recent studies have indicated that adjacent normal tissues of tumors are not entirely normal. Though it is morphologically normal tissues, these tissues have some tumor-similar features in terms of expression profiles and wiring among TF-target genes. The current findings, such as DEGs of ANT, and topological properties of ANT network, indicate the same.

Escape velocity centrality-based influential genes prediction led to the identification of 24 key genes, *viz.*, *E2F1*, *FOXA1*, *JUN*, *AR*, *KLF4*, *GATA3*, *STAT1*, *STAT2*, *NR2C2*, *KLF16*, *BRCA1*, *MAZ*, *MZF1*, *TCF7L2*, *CEBPA*, *SREBF2*, *GRHL2*, *TAL1*, *BCL6*, *DDIT3*, *ERBB3*, *CDH1*, *CCND1*, and *ERBB2*, which were present in all subtypes as well as ANT. These common influential genes are well-known for their role in the development and progression of various cancers, including breast carcinoma. The *E2F1*, a member of the E2F family, is a transcription factor involved in the G1/S transition; it has been reported to control the expression of numerous pro-metastatic genes and act as a master regulator of metastasis in breast cancer. For example, it drives breast cancer metastasis via regulating the gene *FGF13* and altering cell migration [52]. Next, the expression of *FOXA1* correlates with LumA and is a significant cancer-specific survival predictor in ER-positive patients [53]. However, a study by Zhang et al. has confirmed that the overexpression of *JUN* leads to a more invasive phenotype in breast cancer cells and hence, is involved in breast cancer metastasis [54]. Further, it has been demonstrated that androgen-induced anti-proliferative effects on *AR*-positive breast cancer cells are *AR*-mediated [55]. The *GATA3* (GATA binding protein 3) gene is a transcriptional activator, and a study by Usary et al. suggests that loss of *GATA3* may lead to tumorigenesis in ER+ breast cancers [56]. However, the *BRCA1* gene, a tumor suppressor with binding activity associated with DNA damage response, has been found among mutated candidate genes in breast cancer [57]. Similarly, the *ERBB2* (Human epidermal growth factor receptor 2) gene has involvement in

breast cancer progression and is evident from the studies reporting that the overexpression of *ERBB2* leads to increased metastasis of breast cancer [58-60].

The differential correlation analysis and uncovering of switching gene regulatory networks (SGRNs) across tissue types revealed the essential differences in regulatory patterns or wiring among TFs and target genes acting as switching entities between tumor and normal tissues. Many well-known breast cancer genes were significantly differentially correlated with switching behavior across subtypes and ANT. The TF-gene pairs *ESR1* - *FADD*, *GATA3* - *GPER1*, and *NFKB2* - *JAK2* were found to be positively differentially correlated in the Basal subtype while negatively differentially correlated in the case of normal tissues. The TF-gene pairs *ESR2* - *ALDH1A1* and *GATA3* - *HSPA12A* were positively differentially correlated in the Her2 and negatively differentially correlated in the normal tissues. However, the TF-gene pairs *GATA3* - *TMEM51* and *BCL11A* - *SYN2* were negatively and positively differentially correlated in the Her2 and the normal tissues, respectively. Further, the TF-gene pairs *TP63* - *SYN2*, *BACH2* - *NOTCH1*, and *BCL11A* - *DOCK11* were positively and negatively differentially correlated in the LumA and normal tissues, respectively. The TF-gene pairs *BCL11A* - *CACNA1A*, *TP63* - *AKR1C1*, *GATA3* - *EPB42*, and *RUNX3* - *JAK2* were positively and negatively differentially correlated in the LumB and normal tissues, respectively. Whereas the TF-gene pair *GATA3* - *SCPEP1* was negatively and positively differentially correlated in the LumB and normal tissues, respectively. Next, the TF-gene pairs *ESR1* - *ADCY5*, *RUNX1* - *NOTCH4*, *TP53* - *PANK1*, *TP53* - *EN1*, and *STAT2* - *NRAS* were positively and negatively differentially correlated in the NormL and normal tissues, respectively. While the TF-gene pairs *ESR2* - *SLC27A4*, *ESR1* - *PTK7*, *STAT2* - *ZNF544*, *BRCA1* - *FAM219B*, and *BCL11A* - *TMEM87B* were negatively and positively differentially correlated in the NormL and normal tissues, respectively. The TF-gene pairs *GATA3* - *ZNF813*, *TP63* - *B3GALT6*, and *RUNX2* - *TXN* were positively and negatively differentially correlated in the ANT and normal tissues, respectively. While the TF-gene pairs *STAT6* - *CPT1B*, *STAT6* - *PPT2*, and *GATA3* - *FREM1* were negatively and positively differentially correlated in the ANT and normal tissues, respectively.

The functional enrichment and pathway analysis indicated that the DEGs across subtypes and adjacent normal tissues were significantly associated with cancer-related terms such as cell-cell adhesion, DNA repair, calcium metabolism, and cell cycle. The disease ontology across these DEGs indicated the maximum number of genes associated with various types of cancers,

including breast, renal, liver, and lungs. The comparative analysis of these terms across subtypes and adjacent normal tissue indicated their similarity and differences. For example, the KEGG pathway term “Proteoglycans in cancer” was present across all tissue types, including different subtypes and adjacent normal, while the term “Calcium signaling pathway” was present only in NormL and ANT.

5. Conclusions

The present framework of disease exploration using network biology provides a time and cost-effective approach for understanding the disease complexity. The gene regulatory networks of different subtypes and ANT coupled with differential correlation analysis detected structural and expressional patterns of genes responsible for the disease's onset and progression. Further, the network analysis using EVC+ helped to identify 24 common and functionally significant genes, including well-known cancer genes such as *E2F1*, *FOXA1*, *AR*, *JUN*, *BRCA1*, *GATA3*, and *ERBB2* across subtypes and ANT. Similarly, the EVC+ also helped us to identify subtype-specific genes (Basal: 18, Her2: 6, LuminalA: 5, LuminalB: 5, Normal-Like: 2, and ANT: 7). Moreover, the differential correlation and functional and disease annotation also highlighted the cancer-associated role of these genes. Thus, these 24 common genes and subtype-specific key genes especially ANT-specific seven genes *viz. APC*, *BMPR2*, *NFIC*, *TNS3*, *USF1*, *ZBTB7A*, and *ZNF436* could be utilized for the development of early-stage-specific efficient diagnostic and therapeutic candidates. Further, this analysis framework could also be helpful for other cancer types and diseases, e.g., lung cancer, cervical cancer, head and neck cancer, neuronal disorders, and cardiovascular diseases, towards an understanding of the disease by the disease state's structural and functional aspects.

Acknowledgments

The authors would like to thank the Center for Modeling, Simulation & Design (CMSD), University of Hyderabad, for providing computational facilities. Further, VV thanks the Indian Council of Medical Research (ICMR) New Delhi, India (ISRM/12(72)/2020, ID: 2020-2951) and the Institution of Eminence (IoE), University of Hyderabad (UoH/IoE/RC3-21-052) for funding supports. SK would also like to acknowledge the ICMR for the funding support as Senior Research Fellowship (Grant No.: 3/2/2/113/2019/NCD-III, ID: 2019-6723).

Author contributions

Conceived and designed the experiments: Swapnil Kumar, Vaibhav Vindal; Performed the experiments: Swapnil Kumar; Analyzed the data: Swapnil Kumar; Contributed materials and analysis tools: Vaibhav Vindal; Wrote and edited the paper: Swapnil Kumar, Vaibhav Vindal; Supervision: Vaibhav Vindal.

All Authors have read and approved the final manuscript.

Conflicts of interest

There are no conflicts of interest to declare.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2021, 71, 209-249.
2. McAllister SS, Gifford AM, Greiner AL, Kelleher SP, Saelzler MP, Ince TA, et al. Systemic endocrine instigation of indolent tumor growth requires osteopontin. Cell, 2008, 133, 994–1005.
3. Redig AJ, McAllister SS. Breast cancer as a systemic disease: a view of metastasis. J Intern Med, 2013, 274, 113–26.
4. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, Shi B. Breast cancer intrinsic subtype classification, clinical use and future trends. American journal of cancer research, 2015, 5(10), 2929–2943.
5. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World journal of clinical oncology, 2014, 5(3), 412.
6. Meng L, Xu Y, Xu C, Zhang W. Biomarker discovery to improve prediction of breast cancer survival: using gene expression profiling, meta-analysis, and tissue validation. Oncotargets Ther, 2016, 9, 6177–85.
7. Calza S, Hall P, Auer G, Bjöhle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Research, 2006, 8, 1-9.

8. Walker LC, Harris GC, Holloway AJ, Mckenzie GW, Wells JE, Robinson BA, et al. Cytokeratin KRT8/18 expression differentiates distinct subtypes of grade 3 invasive ductal carcinoma of the breast. Cancer genetics and cytogenetics, 2007, 178, 94-103.
9. Shan L, Li X, Liu L, Ding X, Wang Q, Zheng Y, et al. GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation. Oncogene, 2014, 33(24), 3205-16.
10. Brady-West DC and Mcgrowder DA. Triple negative breast cancer: therapeutic and prognostic implications. Asian Pac J Cancer Prev, 2011, 12, 2139-2143.
11. Podo F, Buydens L, Degani H, Hilhorst R, Klipp E, Gribbestad IS, et al. Triple-negative breast cancer: present challenges and new perspectives. Molecular oncology, 2010, 4, 209-229.
12. Shi H, Zhang L, Qu Y, Hou L, Wang L, Zheng M. Prognostic genes of breast cancer revealed by gene co-expression network analysis. Oncol Lett, 2017, 14, 4535–42.
13. Srivastava A, Kumar S, Ramaswamy R. Two-layer modular analysis of gene and protein networks in breast cancer. BMC systems biology, 2014, 8(1), 81.
14. Slaughter, D. P., Southwick, H. W. & Smejkal, W. 'Field cancerization' in oral stratified squamous epithelium: clinical implications of multicentric origin. Clinical implications of multicentric origin. Cancer, 1953, 6, 963–968.
15. Aran, D., Camarda, R., Odegaard, J. et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun, 2017, 8, 1077.
16. Weinstein J.N. Collisson E.A. Mills G.B. Shaw K.R. Ozenberger B.A. Ellrott K. Shmulevich I. Sander C. Stuart J.M. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013, 45, 1113-1120.
17. Lonsdale, J., Thomas, J., Salvatore, M. et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013, 45, 580–585. https://doi.org/10.1038/ng.2653.
18. Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B. and Leek, J.T. Reproducible RNA-seq analysis using recount2. Nature biotechnology, 2017, 35(4), 319-321.
19. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot T, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks:

An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research. doi: 10.1093/nar/gkv1507.

20. Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biology, 2014, 15(12):550.
21. Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010, 26, 139-140.
22. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 2015, 43(7), e47.
23. Ben Guebila, Lopes-Ramos et al. GRAND: A database of gene regulatory network models across human conditions, bioRxiv, 2021. doi:10.1101/2021.06.18.448997.
24. Fang et al. GRNdb: decoding the gene regulatory networks in diverse human and mouse conditions, Nucleic Acids Research, 2020.
25. Zhang, Q. et al. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. Genomics, Proteomics & Bioinformatics, doi:10.1016/j.gpb.2019.09.006.
26. Aaron Clauset, Ellen Tucker and Matthias Sainz, The Colorado Index of Complex Networks. 2016, https://icon.colorado.edu/.
27. Vafaee F, Krycer JR, Ma X, Burykin T, James DE, Kuncic Z. ORTI: An Open-Access Repository of Transcriptional Interactions for Interrogating Mammalian Gene Expression Data. PloS one. 2016, 11(10):e0164535.
28. Zhi-Ping Liu, Canglin Wu, Hongyu Miao and Hulin Wu. RegNetwork: an integrated database of transcriptional and posttranscriptional regulatory networks in human and mouse. Database 2015. doi: 10.1093/database/bav095
29. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon HN, Jung H, Nam S, Chung M, Kim JH, Lee I. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. 2018, 46(D1):D380-D386.

30. Daniel Schlauch, Joseph N Paulson, Albert Young, Kimberly Glass, John Quackenbush, Estimating gene regulatory networks with pandaR, Bioinformatics, 2017, Volume 33, Issue 14, 2232–2234, https://doi.org/10.1093/bioinformatics/btx139.
31. Uhlén M et al., Tissue-based map of the human proteome. Science, 2015, DOI: 10.1126/science.1260419.
32. Csardi G, Nepusz T. The igraph software package for complex network research, InterJournal, Complex Systems, 2006, 1695.
33. Albert, R., Jeong, H. & Barabási, AL. Error and attack tolerance of complex networks. Nature, 2000, 406, 378–382..
34. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: Lethality and centrality in protein networks. Nature. 2001, 411: 41- 10.1038/35075138.
35. Han JD, Bertin N, Hao T, Goldberg D, Berriz G, Zhang L, Dupuy D, Walhout A, Cusick M, Roth F, Vidal M: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004, 430 (6995): 88-93.
36. Carlson M, Zhang B, Fang Z, Mischel P, Horvath S, Nelson SF: Gene connectivity, function, and sequence conservation: Predictions from modular yeast co-expression networks. BMC Genomics. 2006, 7 (40).
37. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977 Mar 1:35-41.
38. Watts DJ, Strogatz SH: Collective dynamics of 'small-world' networks. Nature. 1998, 393 (6684): 440-2.
39. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: Hierarchical organization of modularity in metabolic networks. Science. 2002, 297 (5586): 1551-5.
40. Wasserman, S. and Faust, K. Social network analysis: Methods and applications. 1994, Cambridge University Press, Cambridge.
41. Bonacich, P., Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology, 1972, 2, 113-120.
42. Ruhnau B. Eigenvector-centrality—a node-centrality?. Social networks. 2000, 22(4): 357-65.
43. Dong, J., Horvath, S. Understanding network concepts in modules. BMC Syst Biol 2007, 1, 24. https://doi.org/10.1186/1752-0509-1-24.

44. Ullah, A., Wang, B., Sheng, J. and Khan, N. Escape velocity centrality: escape influence-based key nodes identification in complex networks. Applied Intelligence, 2022, pp.1-19.
45. Fukushima A. DiffCorr: an R package to analyze and visualize differential correlations in biological networks. Gene. 2013, 518(1):209-14. doi: 10.1016/j.gene.2012.11.028.
46. T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021, 2(3):100141.
47. Yu G. enrichplot: Visualization of Functional Enrichment Result. 2022, R package version 1.16.1, <https://yulab-smu.top/biomedical-knowledge-mining-book/>.
48. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford), 2015:bav028.
49. Lotem EY, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter YR, et al. Network motifs in integrated cellular networks of transcription regulation and protein protein interaction. PNAS. 2004; 101(16): 5934–5939. doi: 10.1073/pnas.0306752101.
50. Jalan S, Kanhaiya K, Rai A, Bandapalli OR, Yadav A. Network Topologies Decoding Cervical Cancer. PLoS ONE, 2015, 10(8): e0135183.
51. Rai, A., Menon, A.V. & Jalan, S. Randomness and preserved patterns in cancer network. Sci. Rep. 2014, 4, 6368.
52. Hollern, D.P., Swiatnicki, M.R., Rennhack, J.P. et al. E2F1 Drives Breast Cancer Metastasis by Regulating the Target Gene FGF13 and Altering Cell Migration. Sci Rep, 2019, 9, 10718.
53. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, Dunn S, Huntsman DG, Nakshatri H. FOXA1 expression in breast cancer--correlation with luminal subtype A and survival. Clin Cancer Res. 2007, 13(15 Pt 1):4415-21.
54. Zhang, Y., Pu, X., Shi, M. et al. Critical role of c-Jun overexpression in liver metastasis of human breast cancer xenograft model. BMC Cancer, 2007, 7, 145.
55. Szelei J, Jimenez J, Soto AM, Luizzi MF, Sonnenschein C. Androgen-induced inhibition of proliferation in human breast cancer MCF7 cells transfected with androgen receptor. Endocrinology. 1997, 138(4):1406-12. doi: 10.1210/endo.138.4.5047.

56. Usary J, Llaca V, Karaca G, et al. Mutation of GATA3 in human breast tumors. Oncogene, 2004, 23:7669–78.
57. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science, 2006, 314, 268–74.
58. Tan M, Yao J, Yu D. Overexpression of the c-erbB-2 gene enhanced intrinsic metastatic potential in human breast cancer cells without increasing their transformation abilities. Cancer Res. 1997;57:1199–1205.
59. Moody SE, Sarkisian CJ, Hahn KT. et al. Conditional activation of Neu in the mammary epithelium of transgenic mice results in reversible pulmonary metastasis. Cancer Cell. 2002;2(6):451–461.
60. Holbro T, Civenni G, Hynes NE. The ErbB receptors and their role in cancer progression. Exp Cell Res. 2003; 284(1):99–110.

Appendix A: Supplementary Tables

Supplementary Table S1. Topological properties of random networks.

Supplementary Table S2. Switching gene pairs of Basal subtypes with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S3. Switching gene pairs of Her2 subtypes with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S4. Switching gene pairs of LumA subtypes with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S5. Switching gene pairs of LumB subtypes with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S6. Switching gene pairs of NormL subtypes with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S7. Switching gene pairs of ANT with adjP-value < 0.05 and |Correlation coefficients| > 0.4.

Supplementary Table S8. GO-BP terms significantly associated with DEGs at adjP-value < 0.05.

Supplementary Table S9. GO-MF terms significantly associated with DEGs at adjP-value < 0.05.

Supplementary Table S10. KEGG pathway terms significantly associated with DEGs at adjP-value < 0.05.

Supplementary Table S11. Disease ontology terms significantly associated with DEGs at adjP-value < 0.05.

Supplementary Table S12. Disease-gene network terms significantly associated with DEGs at adjP-value < 0.05.