



OPEN

Identification of modules and key genes associated with breast cancer subtypes through network analysis

María Daniela Mares-Quiñones¹, Edgardo Galán-Vásquez², Ernesto Pérez-Rueda³, D. Guillermo Pérez-Ishiwara¹, María Olivia Medel-Flores¹ & María del Consuelo Gómez-García¹✉

Breast cancer is the most common malignancy in women around the world. Intratumor and intertumoral heterogeneity persist in mammary tumors. Therefore, the identification of biomarkers is essential for the treatment of this malignancy. This study analyzed 28,143 genes expressed in 49 breast cancer cell lines using a Weighted Gene Co-expression Network Analysis to determine specific target proteins for Basal A, Basal B, Luminal A, Luminal B, and HER2 ampl breast cancer subtypes. Sixty-five modules were identified, of which five were characterized as having a high correlation with breast cancer subtypes. Genes overexpressed in the tumor were found to participate in the following mechanisms: regulation of the apoptotic process, transcriptional regulation, angiogenesis, signaling, and cellular survival. In particular, we identified the following genes, considered as hubs: *IFIT3*, an inhibitor of viral and cellular processes; *ETS1*, a transcription factor involved in cell death and tumorigenesis; ENSG00000259723 lncRNA, expressed in cancers; AL033519.3, a hypothetical gene; and *TMEM86A*, important for regulating keratinocyte membrane properties, considered as a key in Basal A, Basal B, Luminal A, Luminal B, and HER2 ampl breast cancer subtypes, respectively. The modules and genes identified in this work can be used to identify possible biomarkers or therapeutic targets in different breast cancer subtypes.

Breast cancer (BC) is the most common malignancy in women and the second leading cause of cancer-related deaths in developed and industrialized countries. In 2020, there were 2.3 million new cases and 685,000 deaths worldwide¹. BC is a process characterized by uncontrolled cell proliferation in the cells of the mammary glands, in which mutations are progressively acquired, generating genomic instability².

The two most common types of BC by origin are ductal and lobular³. However, great morphological and biological heterogeneity generates Intrinsic Subtypes of Breast Cancer (ISBC) such as Luminal A and Luminal B, Basal-like, low in claudin, over-expressed with HER2, triple-negative, and normal type⁴. Intrinsic surrogate subtypes are also classified into triple-negative, over-expressed with HER2, Luminal B as HER2+, Luminal B as HER2-, and Luminal A⁴. Intrinsic surrogate subtypes are also classified into triple-negative, over-expressed with HER2, Luminal B as HER2+, Luminal B as HER2-, and Luminal A. This heterogeneity, in turn, generates differences in clinical behavior, pathological characteristics, and treatment responses, among others⁵⁻⁷.

An essential issue for treatment is the selection of the correct therapeutic modality, which largely depends on the Subtype of Breast Cancer (SBC)⁸. This is crucial because when BC is diagnosed early and managed with a multidisciplinary approach, it becomes a potentially curable disease⁹. To date, the estrogen receptor (ER), the progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) are utilized as biomarkers to streamline clinical decision-making, providing prognostic information, and predict responses to targeted therapies¹⁰. Currently, *BRCA1*, *BRCA2*, and *TP53* genes; STAT3, ESR1, and MUC-1 proteins; miRNAs (26b-5p,

¹Laboratorio de Biomedicina Molecular, Programa de Doctorado en Biotecnología, Escuela Nacional de Medicina y Homeopatía, Instituto Politécnico Nacional, Ciudad de México, Mexico. ²Departamento de Ingeniería de Sistemas Computacionales y Automatización, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México, Mexico. ³Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Unidad Académica del Estado de Yucatán, Mérida, Mexico. ✉email: cgomezg@ipn.mx

124-3p, and 201-5p); and lncRNAs (HOTAIR, linc-ROR, TROJAN), have been described as possible targets for treatment or diagnosis for BC^{11–17}.

In this regard, several works aimed at identifying groups of genes, proteins, or consensus regions using different bioinformatics tools in diseases such as asthma, kidney, and colorectal cancer have been described in BC^{18–20}. Therefore, identifying molecular targets can lead to a more accurate diagnosis of the type of BC, precisely targeted treatments, and potentially treatments that are less toxic than the current ones²¹. In this regard, it is possible to highlight some proteins as potential molecular targets, like Transmembrane Protein 170B (TMEM170B)²², Lipocalin 2 (LCN2)²³, Macrophage Migration Inhibitory Factor (MIF)²⁴, Signal Transducer and Activator of Transcription 3 (STAT3)²⁵, Tumor Protein P53²⁶, Vascular Endothelial Growth Factor A (VEGFA)²⁷, Estrogen Receptor 1 (ESR1)²⁸, BCL2 Associated X (BAX)²⁹, Ribonucleotide Reductase Regulatory Subunit M2 (RRM2)³⁰, Matrix Metalloproteinase 1 (MMP1)³¹, and Maternal Embryonic Leucine Zipper Kinase (MELK), among others³². However, some of these molecules are biased to certain types of BC for a specific signal or were identified from a small biological sample³³.

Nevertheless, it is essential to identify biomarkers according to each SBC to detect with high precision intra-tumor and inter-tumor heterogeneity in patients³⁴, prognostic biomarkers^{35,36} that allow for understanding the degree of disease progression, as well as identifying possible specific therapeutic targets for each cancer subtype.

The amount of information on platforms such as The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), or cBioPortal for Cancer Genomics makes information on gene and protein expression from cell lines, as well as tissue samples from different diseases, including breast cancer, increasingly accessible. Therefore, computational approaches are required to analyze the large amount of data³⁷, such as the Weighted Gene Coexpression Network Analysis (WGCNA). This algorithm allows the identification of genes with similar coexpression profiles from different experiments or samples. Indeed, this algorithm has been implemented to identify potential biomarkers for diagnosis, prognosis, and therapeutic targets in diseases such as asthma, coronary artery disease, and different cancer types, like BC^{18,38–41}.

Therefore, understanding the heterogeneity of BC from the gene expression profiles of cell lines and being able to identify genes expressed specifically for each of the main SBC would allow us to identify potential biomarkers that would improve our ability to diagnose this disease with greater precision, to understand its degree of progression, and even to identify specific therapeutic targets for each SBC that would enable us to determine the most appropriate treatments for patients⁴².

In this work, we identified 5 modules of genes with similar co-expression patterns and hub genes significantly associated with each of the Basal A, Basal B, Luminal A, Luminal B, and HER2 ampl SBC. By using the transcriptomic profiles of 49 breast cancer cell lines, through WGCNA and other bioinformatic methods. Our results provide novel possible biomarkers for specific SBC, which need to be explored for their potential.

Materials and methods

Datasets

The workflow chart of data preparation, processing, and analysis is illustrated in Fig. 1. The dataset used in this study (CCLE_expression) was downloaded from the DepMap, database version 21Q1 (<https://depmap.org>). This database contains gene expression data (RNAseq) for 1,376 cell lines (RSEM, gene) of 57,829 genes⁴³, of which 49 cell lines related to BC were selected (Supplementary Table S1). Then, the data were normalized by using $\log_2(\text{TPM} + 1)$ and filtered using the varFilter function in R, to exclude genes that exhibit less than 50% variation among samples, leaving a total of 28,143 expressed genes with greater variability for the subsequent steps.

Cell lines were classified according to the following features: by origin, 23 primary and 26 metastatic; and by subtypes: 15 Basal A, 8 Basal B, 12 Luminal A, 1 Luminal B, and 13 HER2 ampl (Supplementary Table S1)^{4,44}.

Weighted analysis of gene co-expression network

WGCNA⁴⁵ was performed to evaluate the expression profile of the 49 cell lines and 28,143 genes of BC. To this end, a cluster analysis was conducted to identify atypical samples using the flashclust package in the R environment (Fig. 2a). Then, the biological network's scale-free topology property was incorporated by calculating the β parameter using SoftThreshold with a value of 14. An adjacency matrix was calculated using signal correlation networks and the Pearson correlation coefficient among all the genes. This adjacency matrix was transformed into a Topological Overlap Matrix (TOM), where a higher TOM value facilitates the identification of gene modules for each gene pair with strong interconnectivity. Afterwards, gene modules with similar expression patterns were identified using the average linkage hierarchical clustering algorithm, with default parameters (deepSplit set to 3 and a minimum module size set to 30). Additionally, modules with highly correlated eigengenes were merged based on a minimum height of 0.25 (mergeCutHeight function)⁴⁶. Each module was identified with a different color, with the gray color reserved for uncorrelated genes⁴⁷.

Enrichment analysis of all the modules that make up the network was performed to identify and interpret the biological functions based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database^{48–50} and Gene Ontology (GO), using the Annotation Visualization and Integrated Discovery (DAVID). All analyses were performed under the criteria of an adjusted p value of < 0.05 .

To determine the association between each SBC and the modules, the module-trait relationship (MTR) was employed^{46,47}. This approach was used to correlate the gene traits with the sample traits in each SBC, focusing on the most significant correlations.

Identification of hub genes

In this study, we selected the top 50 genes with highly correlated values across different thresholds from each SBC in the modules, as detailed in Supplementary Table S2. This selection was made through the module of high

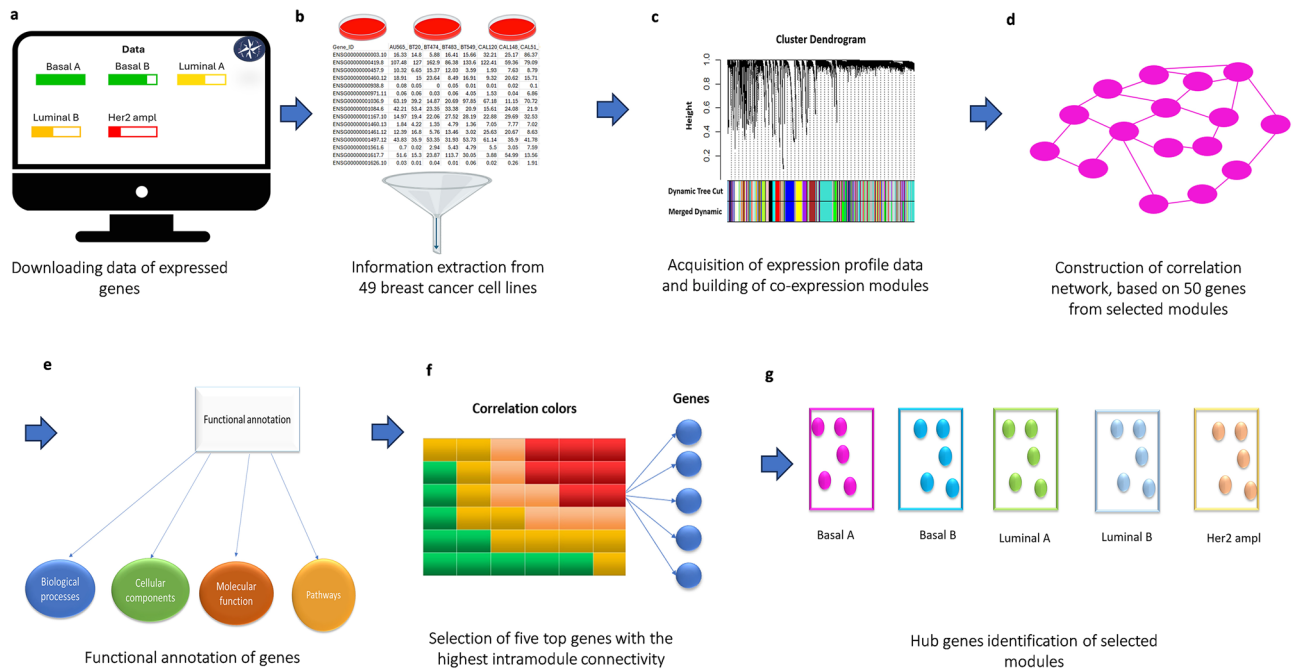


Figure 1. Workflow diagram outlining the proposed method for identifying modules and key genes associated with breast cancer subtypes using gene co-expression networks. **(a)** Initially, gene expression data from breast cancer cell lines were retrieved from the DepMap database version 21Q1. **(b)** A comprehensive analysis was then conducted on 49 breast cancer cell lines to obtain specific details about gene expression in these cell lines. **(c)** Next, a gene co-expression network was reconstructed based on the processed expression profiles using the WGCNA package in the R programming environment, resulting in 65 modules. **(d)** Subsequently, the module with the highest correlation for each breast cancer subtype (pink, turquoise, yellowgreen, skyblue, and navajowhite2) was selected to construct a correlation network consisting of 50 highly correlated genes from each chosen module. The software used for this process was Cytoscape. **(e)** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)^{48–50} were then utilized to perform functional analyses aiming to understand the biological functions and metabolic pathways associated with the genes in the module. **(f)** The top five genes with the highest intramodule connectivity were identified. **(g)** Finally, the hub genes within each module were identified for each subtype of breast cancer.

module membership (MM), high genetic significance (GS), and high intramodular connectivity (IC). Afterward, we identified the five hub genes from each subtype, characterized by the highest intramodule connectivity, which is calculated as the sum of the degree of correlations between that gene and other genes within the module.

Functional annotation

To evaluate the significance of the biological function of genes in each SBC, the Database for Annotation, Visualization, and Integrated Discovery (DAVID; version 6.8; <https://david.ncifcrf.gov>) was used. For this purpose, an enrichment analysis of Gene Ontology terms and KEGG pathways was implemented and displayed using the ggplot2 package in R, including molecular functions (MF), biological processes (BP), and cellular components (CC) with a statistical significance at an adjusted p value of <0.05 . Protein–protein interaction (PPI) networks were built using STRING software (version 10.5; <http://string-db.org>), with the interaction parameter set to confidence maximum of 0.90. Finally, to visualize and analyze the PPI network and the correlation network, Cytoscape software (version 3.6.1; <http://www.cytoscape.org>) was used.

Validation of hub genes in an external dataset

To confirm hub shared genes in each SBC, the mRNA expression dataset of breast cancer was searched using the keywords: “breast cancer samples”, “TNBC”, “Luminal A”, “Luminal B”, “Her2”, “*Homo sapiens*”, “expression profiling by array” against the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>). After a review, the GSE65194 (Affymetrix Human Genome U133 Plus 2.0 Array), GSE96860, GSE52194, and GSE134359 profiles, were selected. We conducted differentially expressed genes (DEG) analysis on the dataset using the R package “limma”. An adjusted p value <0.05 was considered as the cutoff value, where our main hub genes and long non-coding RNAs were found for each SBC.

Results and discussion

Weighted analysis of gene co-expression network

To evaluate the gene expression of BC, RNA seq data from 49 cell lines comprising 28,143 genes were selected (Fig. 2a; Supplementary Table S3). These samples were used to construct a Gene Co-expression Network using

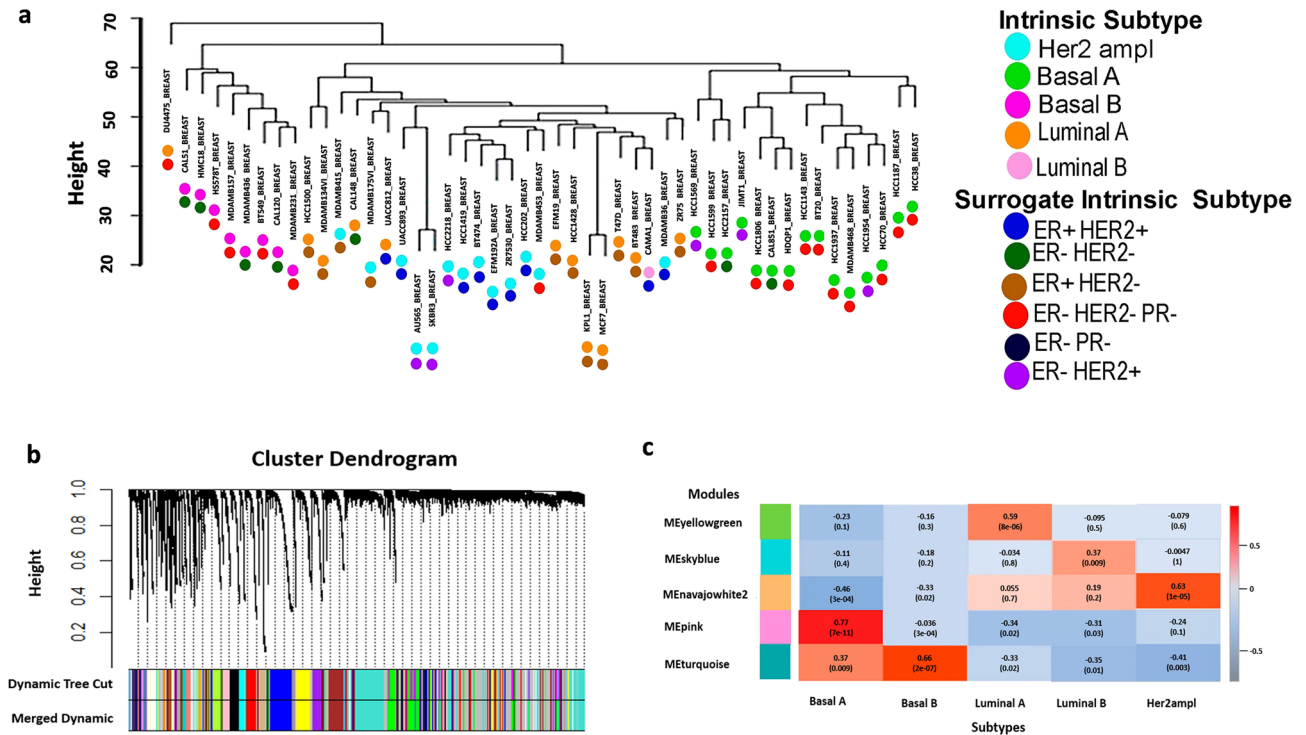


Figure 2. Weighted gene co-expression network analysis. (a) Clustering analysis to remove outliers. Sample dendrogram and trait indicator. Each color circle represents an SBC, five intrinsic subtypes, and six surrogate intrinsic subtype. (b) Network analysis of gene expression in BC identifying distinct modules of co-expression genes. (c) Correlations between module eigengenes and different SBCs with only the most significant module. The most correlated modules are shown for each SBC.

the WGCNA software, with a Soft Threshold power β of 14 as the scale-free topology criterion, a signed network that allows for the identifying modules with more significant enrichment of functional groups, and Pearson correlation (Supplementary Fig. 1a,b)⁴⁵. After, the network was processed by hierarchical clustering, identifying 65 modules (Fig. 2b); that is, a set of genes with similar expression patterns. Then, the eigengene modules were clustered to improve the reliability of the module divisions (Supplementary Fig. S2). The clustering algorithm was used to identify modules of highly correlated genes and was refined to improve the reliability of the module divisions, where each module was assigned a color, with the smallest containing 34 genes (yellow4) and the largest one containing 2297 genes (turquoise).

The modules were found to include genes previously associated with BC, such as *TP53*, which plays a key role in controlling cell division and cell death; *KI67*, related to cellular proliferation; *BRCA1*, *BRCA2*, which are hallmarks for hereditary BC³²; and specific targets that coincide with the different types of BC, such as *BCL2L14*, which is apoptosis facilitator⁵¹, the tumor necrosis factor receptor superfamily member 1A (*TNFRSF1A*)²⁵, and a potassium inwardly-rectifying channel subfamily J member 3 (*KCNJ3*)⁵², among others. Therefore, the role of these genes has already been demonstrated in BC, which supports the representativeness of the genes in our sample.

Significant modules associated to SBCs

A Pearson correlation analysis was conducted to identify the most significant modules associated with each SBC. Below, we describe the most significant modules (Fig. 3, Supplementary Fig. S3). For Basal A, the pink module, with 879 genes ($r=0.77$, p value = $7e-11$) was found to be as significant. For Basal B, the turquoise module was identified with 2297 genes ($r=0.66$, $p=2e-7$); in the Luminal A subtype, the yellowgreen module, consisting of 131 genes ($r=0.59$, $p=8e-6$), was noted. For Luminal B, the skyblue module includes 169 genes with ($r=0.39$ and $p=0.009$). Finally, the navajowhite2 module representing HER2 ampl with 70 genes, was identified, showing $r=0.63$ and $p=1e-5$.

To associate the biological processes with the pink, turquoise, yellowgreen, skyblue, and navajowhite2 modules, which exhibit high correlation values, a Functional Enrichment of Gene Ontology Analysis with DAVID was achieved. In this regard, for Basal A, we identified 171 genes mainly involved in defense response (GO:0006952) with an adjusted p value of $2.60e-33$ (Supplementary Table S3), regulating the body's immune response, and in the body's ability to contend against cancer cells which play a critical role throughout the course of breast carcinogenesis^{53,54}. In breast cancer, specifically in the Basal A subtype, it has been shown that the immune system plays a dual role in tumor initiation and progression. This system can both inhibit and promote tumor expansion. Cytokines TGF β , INF γ , and TNF α have been involved in both functions, by inducing different expression profiles in tumor cells⁵⁵. Additionally, immune-related proteins that are overexpressed in

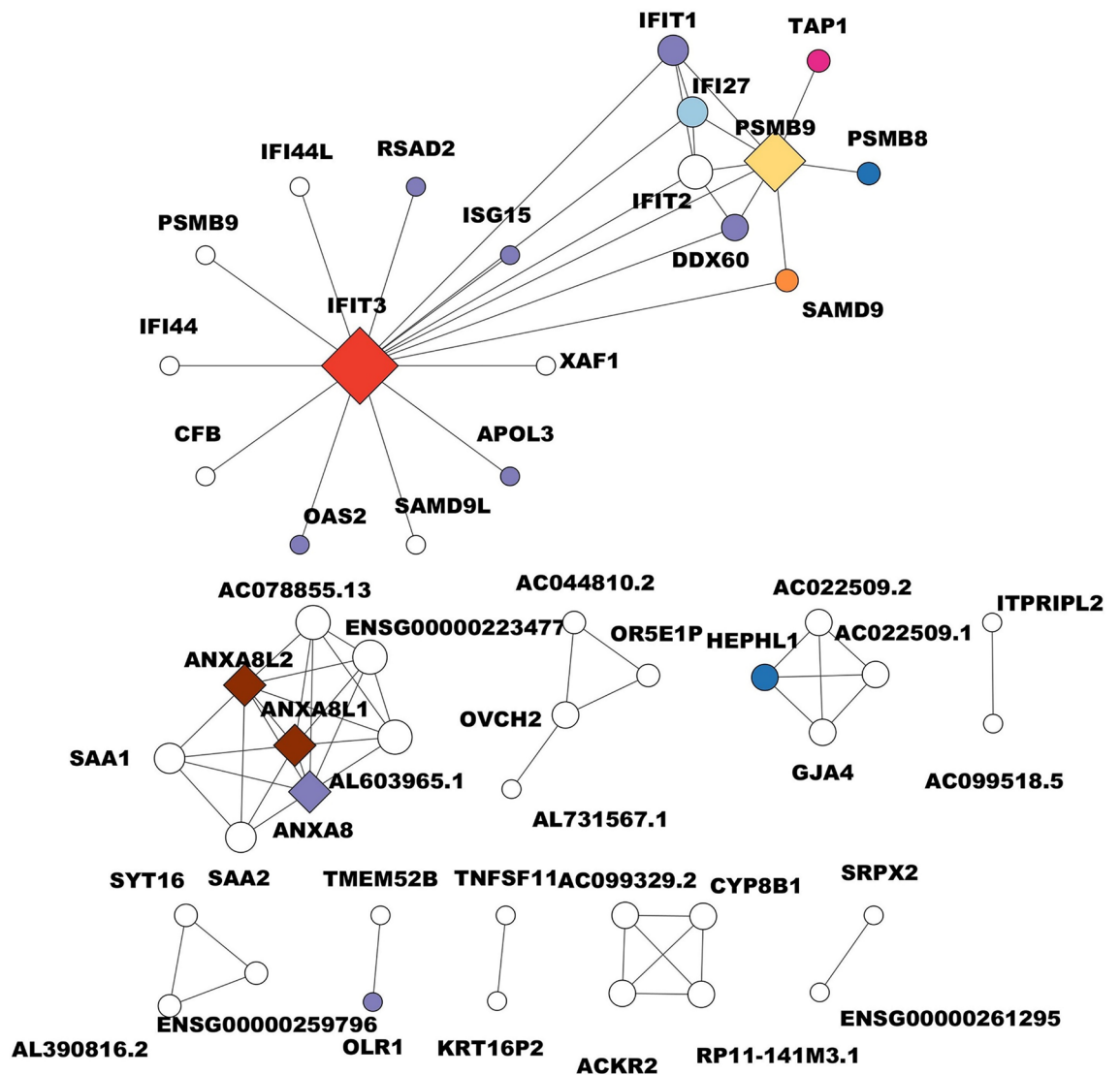


Figure 3. Basal A breast cancer subtype correlation network. The circles in the network represent different biological functions, each color-coded for easy identification. Dark blue circles represent catalytic activity, pink circles represent ATP-dependent activity, and brown circles represent calcium-binding proteins. Purple circles indicate binding functions, while yellow circles are involved in antigen processing to generate class I binding peptides. Red circles represent defense or immunity proteins, light blue circles represent signaling functions, and green circles represent transported activity. Lastly, orange circles represent the regulation of cell proliferation and apoptosis. The diamond shapes in the network represent hub genes, which are genes that have a high degree of connectivity in the network and play a crucial role in the biological processes. The size of each node (circle or diamond) corresponds to the number of connections or the degree of involvement in biological processes. Larger nodes have more connections, indicating a higher degree of involvement in various biological processes.

this subtype of breast cancer have been associated with invasion and metastasis, as is the case with CD58⁵⁶. Furthermore, the differences in the mutation profiles among subtypes underscore the complexity of this disease and the need for specific therapeutic approaches. In the Basal A subtype of breast cancer, a predominance of mutations in the *TP53* gene has been observed in 80% of cases. This high frequency is associated with increased invasion, migration, and cellular resistance, thereby contributing to tumor aggressiveness and a poor prognosis. This pattern suggests that alterations in *TP53* play a fundamental role in the pathogenesis of this subtype⁵⁷. For Basal B, we found 370 genes involved in the movement of cell or subcellular components (GO:0006928) with an adjusted *p* value of $1.70e-24$ (Supplementary Table S3). This movement is essential for many body functions, including tissue growth and repair, migration of immune cells to sites of inflammation, and cancer cells' invasion into surrounding tissues⁵⁸. This is also consistent with work that has documented that patients with the Basal B subtype of breast cancer have a poor prognosis and earlier relapses with metastases. Moreover, metastases occur mainly in the brain, followed by the lungs and distal lymph nodes, suggesting that there are metastatic signatures for each organ⁵⁹. Furthermore, significant overexpression of genes related to migration and invasion has been previously observed, such as *EGFR*, *MMP13*, *SOX4*, and *IGFBP2*^{60,61}. Similarly, in the Basal B subtype,

the prominence of mutations in *TP53* is maintained, followed by alterations in *PIK3CA*, *NF1*, and *FBXW7* in primary tumors. The mutations extend to *ESR1*, *BRCA2*, *RB1*, *ERBB2*, and *AKT1* in metastatic cases, revealing a genetic complexity that affects treatment efficacy. This mutational diversity indicates the intrinsic heterogeneity of breast cancer tumors and the importance of identifying specific biomarkers to improve therapeutic approaches⁶².

For Luminal A, 16 genes were involved in the plasma membrane (GO:0044459) with an adjusted *p* value of 0.19 (Supplementary Table S4), regulating how cells attach to surrounding tissues and other blood vessels. These genes also control the migration and invasion of cancer cells, essential for tumor growth and cancer spread. They regulate how cells respond to cancer treatments, affecting the effectiveness of therapy⁶³. Furthermore, the Luminal A subtype, as reported by Padua et al.⁶⁴ observed that in the Luminal subtype, there was an effect on the expression of genes associated with the plasma membrane and the composition of the extracellular region, as well as the expression of the ER α receptor. This alteration may explain the characteristics of more aggressive cell growth. ER α is a crucial molecule in tumor development and progression due to its ability to activate various signaling pathways, whether cytoplasmic or nuclear, which are present, among other places, in the plasma membrane⁶⁵. Additionally, this subtype shows low expression of genes related to cell proliferation⁶⁶. On the other hand, in the Luminal A subtype, *PIK3CA* emerges as the most frequently mutated gene, followed by *GATA3*, *MAP3K1*, and *TP53*, reflecting a distinctive mutational profile that could influence therapeutic decisions⁶⁷. This pattern suggests that *ESR1*, *GATA3*, *KMT2C*, and *PTEN* also play significant roles in cancer progression and treatment response^{68,69}.

For Luminal B, 19 genes were associated with the extracellular space (GO:0005615) with an adjusted *p* value of 0.14 (Supplementary Table S4), which are important in cancer because they regulate how cancer cells interact with their immediate environment. These genes can control the production of proteins that modulate the adhesion of cancer cells to surrounding tissues and other blood vessels, essential for tumor growth and spread. In addition, these genes can also control the production of proteins that promote angiogenesis, and regulate how cells respond to cancer treatments, affecting the effectiveness of therapy⁷⁰. The composition and three-dimensional structure of the ECM vary significantly, affecting processes such as apoptosis and cell proliferation. Furthermore, changes in the ECM have been linked to resistance to endocrine treatments and cancer recurrence, demonstrating the complexity of the interaction between tumor cells and their microenvironment⁷¹. Additionally, in the Luminal B subtype, a relationship has been established between the heterogeneity of breast cancer tumors and the components of the ECM, whose composition and three-dimensional structure undergo constant remodeling through closely regulated mechanisms during tumorigenesis. It has been evidenced that the ECM of different SBCs, such as Luminal and TNBC, exhibit disparities in the transcriptomic profile, especially in genes that influence cell behavior, such as apoptosis, the cell cycle, DNA replication, and the estrogen signaling pathway, including E-cadherin, vimentin, and ER α . Similarly, a correlation has been established between changes in ECM composition, endocrine resistance, and relapses⁷².

These subtypes also present a high expression rate of genes associated with proliferation, such as *MKI67* and *AURKA*⁷³. The Luminal B subtype is characterized by a high frequency of mutations in *PIK3CA*, *GATA3*, and *TP53*, with a marker difference in the prevalence of P53 mutations compared to Luminal A⁶². This subtype also illustrates how the breast cancer genome can evolve during therapy, with 20% of patients acquiring mutations in *ESR1*, leading to resistance to endocrine therapy. This phenomenon underscores the importance of understanding mutational dynamics to address treatment resistance⁷⁴. Finally, for HER2 ampl, four genes with coenzyme binding function were found (GO: 0050662) with an adjusted *p* value of 0.36 (Supplementary Table S5). These genes can control the production of enzymes involved in metabolic pathways, such as the pentose phosphate pathway, essential for the biosynthesis of molecules needed for cell growth and division⁷⁵. In addition, Moasser et al.⁷⁶ highlights the overexpression of the HER2 gene and its implication in an aggressive tumor phenotype through the activation of signaling pathways such as RAS/MAPK and PI3K/AKT. This subtype shows a clear association between the enzymatic function of involved proteins and a higher propensity for cell proliferation and survival of the HER2-positive subtype.

Approximately 20% of cases exhibit amplification or overexpression of HER2, associated with an aggressive tumor phenotype and poor survival. HER2 encodes for a transmembrane tyrosine kinase receptor, playing a crucial role in cell–cell and cell–stroma communication through signal transduction. Many proteins and intermediaries are involved in these pathways exhibiting enzymatic activity. The main signaling pathways involved, RAS/MAPK, PI3K/AKT, and JAK/STAT, among others, affect cell proliferation, survival, mobility, and adhesion⁷⁷. In this subtype, gene enrichment is primarily related to coenzyme-binding function⁷⁶. Furthermore, this subtype has the second-highest prevalence of P53 mutations (70%), making it a significant biomarker and therapeutic target⁷⁸. Additionally, three mutations (Q429R, Q429H, and T798M) in HER2, identified in the MCF7 cell line, have been shown to diminish the effectiveness of trastuzumab in this cell line, and are also postulated as potential biomarkers for predicting resistance to therapy⁷⁹. These results demonstrated that the SBCs, Basal A, Basal B, Luminal A, Luminal B, and HER2 ampl, exhibit different co-expression patterns, thereby implicating divergent biological processes. This distinction could facilitate the identification of potential biomarkers unique to each subtype.

In Basal A, for metabolic pathways analysis, 21 genes are involved in the phagosome pathway with an adjusted *p* value of $2.4e-5$, and 23 genes are involved in Epstein–Barr virus infection with an adjusted *p* value of $8.0e-5$. A meta-analysis indicates a strong statistical relationship between Epstein–Barr virus infection (EBV) and BC risk, suggesting a potential role for EBV infection in the development of BC⁸⁰. Gupta et al.⁸¹ reported that the presence of EBV in triple-negative breast cancer poses a high risk for patients. Additionally, there is a potential association between serous epithelial ovarian cancer (EOC) and EBV infection, which could be subtype-specific⁸². For the Basal B subtype, 52 genes participate in proteoglycans in cancer with an adjusted *p* value of $3.1e-7$ (Supplementary Table S6). Proteoglycans contribute to the development and progression of the disease and its response to treatment by potentially affecting tumor growth, invasion, and metastasis through interactions

with other extracellular matrix molecules and the regulation of signaling pathways⁸³. These results illustrate that the genes in the pink module, highly correlated with the Basal A subtype, differ from those in the turquoise module, highly correlated with the Basal B subtype, involving them in distinct biological processes and molecular functions.

Correlation network and identification of hub genes

Hub genes are characterized by being genes whose degree value is higher than the average of the threshold network degree value⁸⁴. Many studies have shown that the relationship between connectivity and node significance carries important biological information⁴⁷. In this respect, to identify hub genes for each SBC, we obtained the top 50 highly degree genes for pink, turquoise, yellowgreen, skyblue, and navajowhite2 modules, and their networks were displayed (Supplementary Tables S7, S8). The results show that the Basal A network is composed of 50 nodes and 75 edges with a threshold of 0.05. The gene with the highest degree, corresponding to the main hub gene, is *IFIT3* (Interferon-induced protein with tetratricopeptide repeats 3) with 16 degrees, followed by *PSMB9* (Proteasome 20S Subunit Beta 9) with 12 degrees, *anxa8l2* (annexin A8-like 2), *ANXA8* (annexin A8), and *ANXA8L1* (annexin A8-like 1) with 7 degrees each one (Fig. 3, Supplementary Table S9). In this regard, *IFIT3* encodes for a transcription factor involved in the negative regulation of apoptotic processes and the negative regulation of cell proliferation^{85,86}, indicating its potential role in cancer development and progression.

It inhibits proliferation in human BC cells (treated with T-47D curcumin)⁸⁷, which is consistent with the biological process identified for the pink module. Moreover, *ifit3* is considered a predictive biomarker for chemotherapy and radiation in some human cancers, such as breast, lung, prostate, neck, head cancer, and glioma^{86,87}, indicating its potential utility across multiple cancer types. However, this is the first time that it has been identified as a potential biomarker associated with the Basal A subtype. Recently, Lamsal et al.⁸⁸ demonstrated that *IFIT3* mRNA and its protein are highly expressed in human metastatic breast cancer cell line MDA-MB-231, while its expression is low in non-metastatic MDA-MB-453 cell line. The next hub gene, *PSMB9*, encodes a modified enzyme with peptidase activity involved in the proteasome pathway, which helps cells deal with oxidative and proteotoxic stress. The overexpression of proteasomes is present in a wide variety of cancer types⁸⁹, further highlighting the importance of these genes in cancer progression and treatment response. Proteasomes play a key role in the stress response because they degrade damaged proteins⁹⁰. *PSMB9* is also a potential diagnostic biomarker and tumor suppressor for human uterine leiomyosarcoma⁹¹. *ANXA8L2* Protein 2, *ANXA8* protein (annexin A8), and *ANXA8L1* protein 1 are members of the annexin family of evolutionarily conserved Ca²⁺ and phospholipid-binding proteins; their overexpression has been associated with acute myelocytic leukemia³². In addition, *ANXA8L1* participates in biological regulation, cellular process, localization, metabolic process, calcium-binding protein, response to wounding, and endomembrane system organization^{92,93}. Interestingly, *ANXA8* expression is directly related to a subgroup of basal-type breast cancer patients with poor prognosis⁹⁴, which correlates with the association identified between these three hub genes and the Basal A subtype in this work.

In the case of the Basal B subtype, the network consists of 50 nodes and 57 edges with a threshold of 0.0670. The main hub gene is the transcription factor *ets1* (23 degrees). The remaining four hub genes are *ADAMTS6* with 6 degrees, *MIR100HG* with 6 degrees, *PTGFR* with 6 degrees, and *CAVI* with 5 degrees (Fig. 4, and Supplementary Table S10). *ETS1* belongs to the ETS family of transcription factors and is overexpressed in some types of cancer cells, such as prostate cancer and leukemia⁹⁵. It participates in cancer pathways like angiogenesis, PDGF, VEGF, and RAS signaling, cellular senescence, human T-cell leukemia virus infection, and renal cell carcinoma. The biological processes of *ETS1* include extracellular structure organization, response to wounding, blood vessel development, negative regulation of cell proliferation, regulation of cell death, DNA-binding, and transcription regulator activity^{92,93}.

It was detected as a potential prognostic biomarker in gastric cancer⁹⁶. Recently, *ETS1* was also described as a specific prognostic biomarker for TNBC patients, but not for patients who are not triple-negative or other patients with breast cancer. In addition, a very close association with invasiveness, metastasis, and a poor prognosis was determined in TNBC patients, which is directly related to the overexpression it induces of the genes that code for the metalloproteinases MMP1, MMP-3, and MMP9, as well as VEGF among others⁹⁷. This result agrees with the one obtained in this work where the hub *ETS1* gene is directly associated with the Basal B subtype. *ADAMTS6* has catalytic activity and is involved in cellular processes, protein-modifying enzymes, extracellular structure organization, and blood vessel development. It contains a disintegrin and metalloprotease domains with thrombospondins motifs, members of this family can be secreted by cancer and stromal cells and can contribute to modifying the tumor microenvironment by multiple mechanisms, such as cell adhesion, migration, proliferation, and angiogenesis⁹⁸. Additionally, *ADAMTS6* suppresses tumor progression via the ERK signaling pathway and is a prognostic marker in human BC⁹⁹.

The next hub gene is *MIR100HG*, a long non-coding RNA (lncRNAs). It acts as a regulator of cell proliferation and has been reported as a prognostic biomarker for gastric cancer¹⁰⁰. lncRNAs are endogenous in various cancers, including BC¹⁰¹, and are involved in different biological processes like migration, metastasis, and proliferation¹⁷. *PTGFR* is a tumor suppressor gene involved in a tumor promotion process, perhaps via interaction with its specific ligand (PGF2α). It has been described as a potential biomarker to predict prostate cancer progression, particularly between stage II and subsequent stages of the disease¹⁰². *CAVI* is a membrane protein with a crucial role in the advancement of BC, including cell proliferation, apoptosis, autophagy, invasion, migration, and metastasis¹⁰³. *CAVI* has been identified as a biomarker of colorectal, lung, penile, prostate, and non-small cell lung cancer^{104–108}. Interestingly Al et al. found genes identified from samples of patients with BC

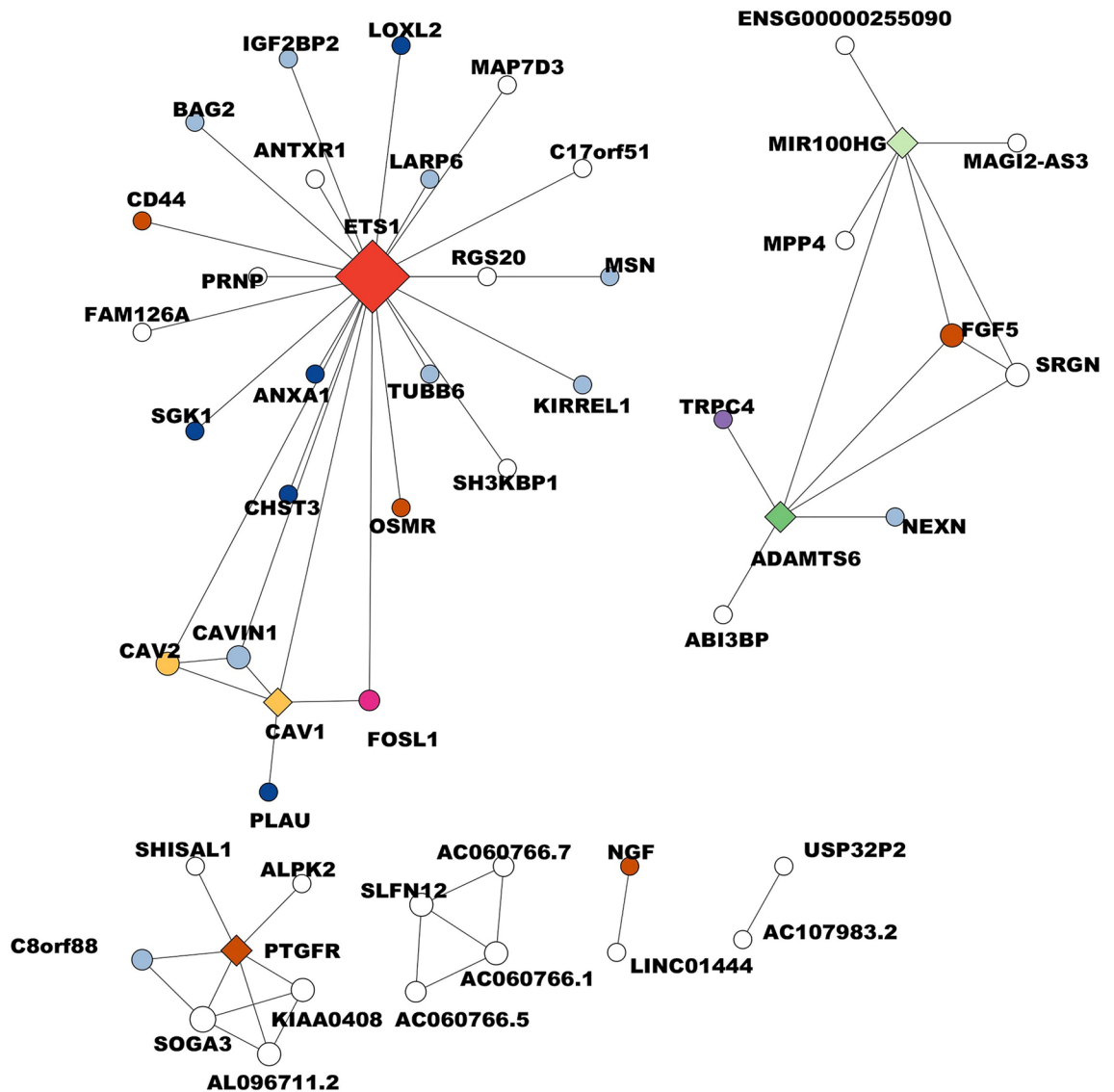


Figure 4. Basal B breast cancer subtype correlation network. The circles in the network symbolize different biological functions, each distinguished by a unique color. Dark blue circles represent catalytic activity, pink circles represent transcriptional regulator activity, and brown circles symbolize molecular adaptor activity. Light blue circles indicate binding functions, while yellow circles are associated with molecular adaptor activity. Red circles represent gene-specific transcriptional regulators, purple circles indicate transporter activity, green circles represent modifying enzymes, and light green circles indicate regulators of cell proliferation. The diamond shapes in the network represent hub genes. These are genes that have a high degree of connectivity within the network and play a significant role in the biological processes. The size of each node, whether a circle or diamond, corresponds to the number of connection degrees involved in biological processes. Larger nodes have more connections, indicating a higher degree of involvement in various biological processes.

of basal subtype to *IFI44L* (Basal A), *PLAU*, and *IL6* (Basal B), which are in the top 50 that we identified of these subtypes¹⁰⁹.

For subtype Luminal A, we identified a network with 50 nodes and 264 edges with a threshold of 0.0445. The gene with the highest degree and the main hub gene is the transcriptional regulator ENSG00000259723 with 32 degrees, followed by ENSG00000266934 with 30 degrees, ENSG00000236055 with 29 degrees, *CA4* with 25 degrees, and AL354813.1 with 24 degrees (Fig. 5, Supplementary Table S11). ENSG00000259723 is a lncRNA expressed in seven cancers, such as bladder cancer, colon adenocarcinoma, head and neck squamous cell carcinoma, lung squamous cell carcinoma, ovarian cancer, prostate adenocarcinoma, and endometrial carcinoma of the uterine body^{110,111}. ENSG00000266934 is present in 31 species, such as *Gorilla gorilla gorilla*, *Carlito syrichta*, *Pteropus vampyrus*, *Dasyurus novemcinctus*¹¹²; and ENSG00000236055, located on chromosome 17, and 1 respectively. *CA4* is a carbonic anhydrase implicated in cell proliferation, and inhibits cell proliferation, invasion, and metastasis³². *CA4* has been reported as a clinicopathological biomarker of immune infiltration and outcomes in kidney carcinoma, lower-grade glioma, pulmonary adenocarcinoma, and uveal melanoma¹¹³, and identified as a biomarker for gastric cancer¹¹⁴. AL354813.1 is a lncRNA located on chromosome 20 and

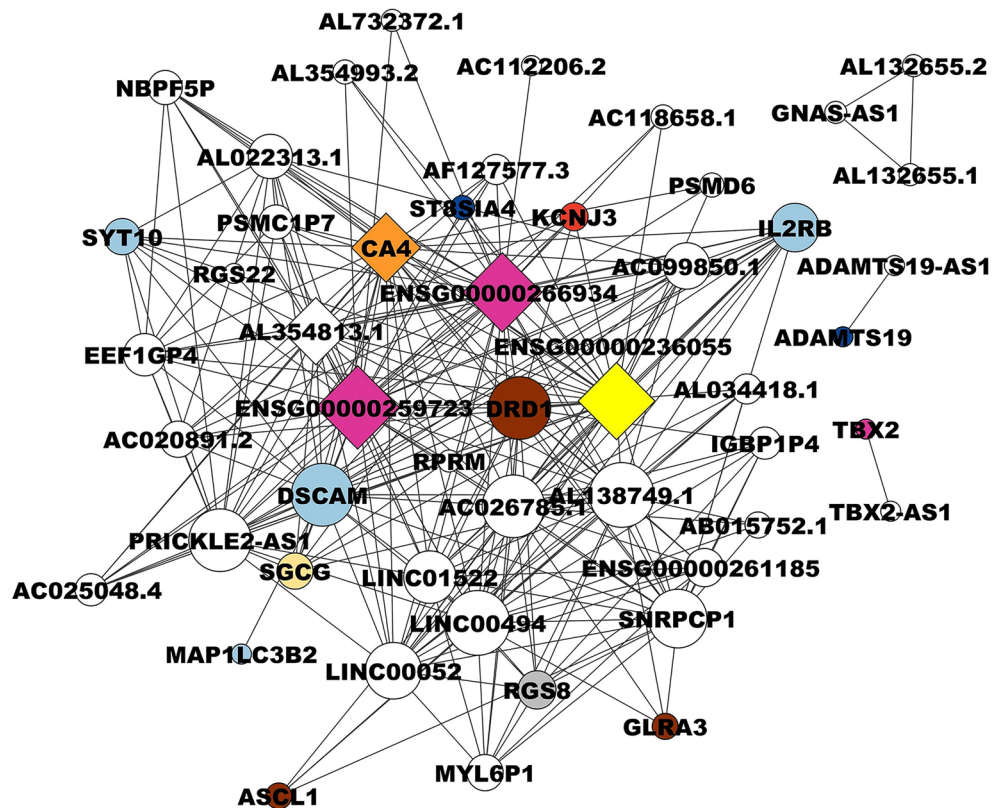


Figure 5. Representation of a correlation network for the Luminal A subtype of breast cancer. Visualising the relationships between different biological functions and hub genes. The circles in the network represent different biological functions, each distinguished by a unique color. Dark blue circles denote catalytic activity, pink circles represent transcriptional regulator activity, and brown circles symbolize molecular transduction activity. Light blue circles indicate binding functions, while orange circles are associated with metabolite interconversion enzyme activity. Red circles represent transporter activity, purple circles indicate transmembrane signal receptor activity, yellow circles indicate cytoskeletal protein functions, and gray circles represent protein-binding activity modulators. The diamond shapes in the network represent hub genes, which are genes with a high degree of connectivity within the network, indicating their important role in the biological processes. The size of each node, whether a circle or diamond, corresponds to the number of connection degrees involved in biological processes. Larger nodes have more connections, indicating a higher degree of involvement in various biological processes.

has one transcript with 2437 bp¹¹⁰. Recent research has shown that the amount of lncRNA in cancer differs in normal tissues, and there are many different prognostic-related expressions. In other types of cancer, lncRNAs have been identified as prognostic markers^{115,116}. It was found that lncRPM plays a crucial role in sustaining and has functioned as a chemo-resistance biomarker in BC cancer stem cells¹¹⁷. Although many lncRNAs have been identified in BC, their functional mechanism remains unknown. Interestingly, we found not only hub genes that encode transcription factors but also lncRNAs, as has also been reported^{118,119}.

The network for the subtype luminal B is composed of 50 nodes and 265 edges, with a threshold of 0.043. The gene with the highest degree is *AL033519.3* with 36 degrees, *MS4A15* with 32 degrees with signal transduction activity, *WASIR1* with 27 degrees, and *PAMR1* with 25 degrees, and *TTR* with 25 degrees (Fig. 6, Supplementary Table S12). *AL033519.3* is a hypothetical gene located on chromosome 6 and comprises 771 nucleotides. Currently, there are no studies linking these genes to tumorigenesis or its progression. *MS4A15* is involved in cellular processes, response to stimulus, and signaling³². It regulates the cellular metabolism involved in the malignant transformation of numerous tumors, including ovarian cancer¹²⁰. Its overexpression correlates with poor survival in patients with ovarian cancer, making it a promising therapeutic target. *MS4A15* has also been suggested as a possible biomarker for colon and gastric cancer¹²¹. *WASIR1* is a lncRNA that appears to interact with *LINC00511*, which is overexpressed in BC and promotes its progression¹²². *PAMR1* is predicted to be involved in proteolysis³². It is inactivated by promoter hypermethylation in BC, therefore, it has been considered a tumor suppressor¹²³ and a potential biomarker that has a negative correlation to the spread of cervical cancer¹²⁴. *TTR* plays a role in stimulating tumor growth mediated by activation of mitogenic and oncogenic molecules, as well as immune and endothelial cells, specifically in lung cancer. It has a significant effect on the range of functions of endothelial cells, managing both tumor and immune cell migration and infiltration. When endothelial cells are treated with *TTR*, there's a suppression of T cell proliferation¹²⁵ linking it to disease pathologies and marking

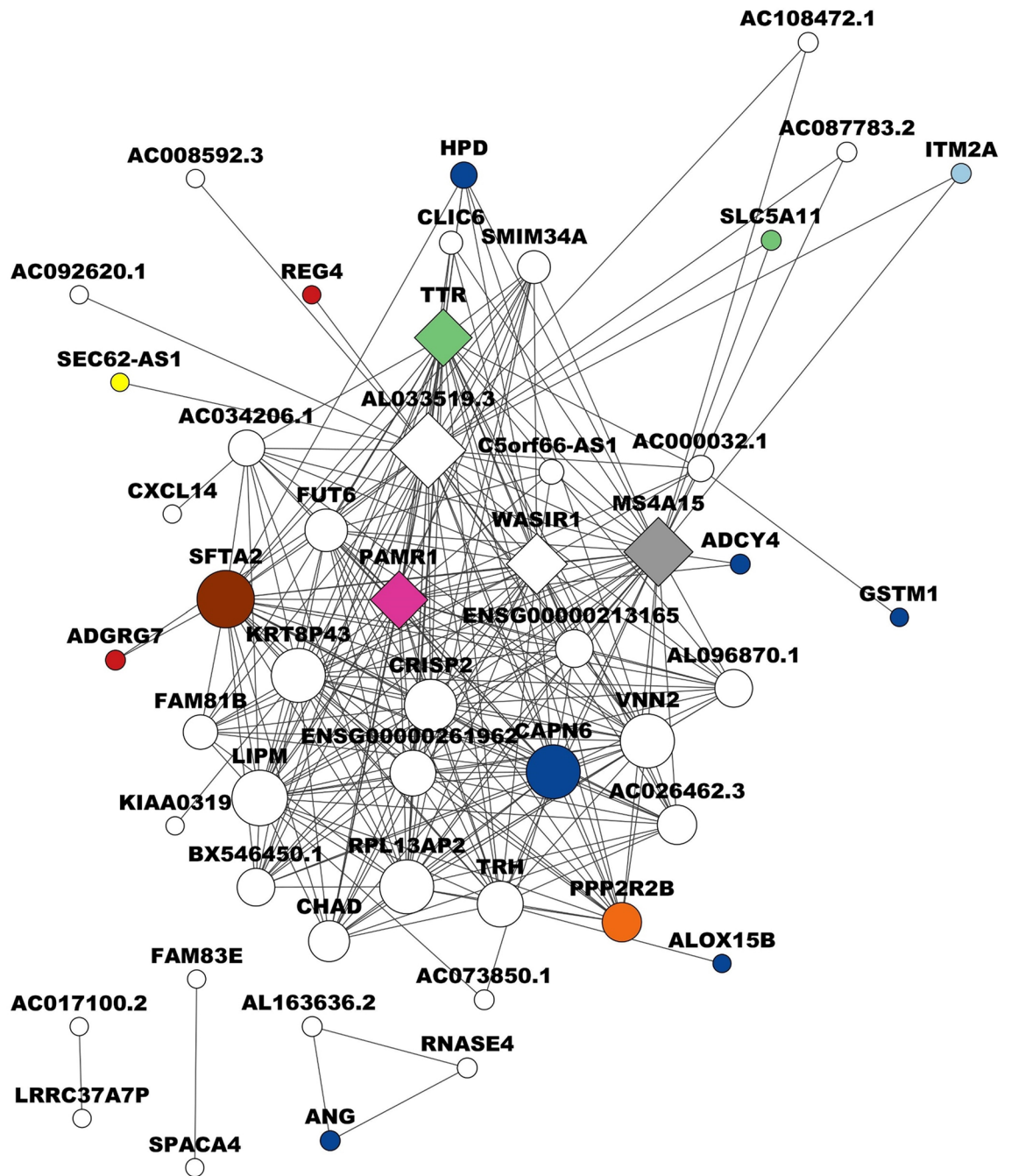


Figure 6. Correlation network for the Luminal B subtype of breast cancer. The circles in the network symbolize different biological functions, each distinguished by a unique color. Dark blue circles denote catalytic activity, pink circles represent proliferation inhibitor activity, and brown circles symbolize disease-susceptibility for panbronchiolitis. Light blue circles indicate binding functions, while orange circles are associated with molecular function regulator activity. Red circles represent molecular transducer activity, green circles denote transporter activity, and gray circles represent signal transduction functions. The diamond shapes in the network represent hub genes. These are genes that have a high degree of connectivity within the network, indicating their significant role in the biological processes. The size of each node, whether a circle or diamond, corresponds to the number of connection degrees involved in biological processes. Larger nodes have more connections, indicating a higher degree of involvement in various biological processes.

it as Oxidative Stress Biomarker¹²⁶. Additionally, the inhibition of *TTR*, associated with conditions such as malnutrition and inflammation, is recognized as a biomarker for various human morbidities¹²⁷.

Finally, the network for the subtype HER2 ampl consists of 50 nodes and 162 edges with a threshold of 0.0211. The gene with the highest grade and considered as the main hub is *TMEM86A* with 32 degrees, linked to ichthyosis disease; followed by four important genes in this network which are *CAPN9* with 25 degrees, *ERVE-1* with 21 degrees, *SCGB2A2* with 14 degrees, and *RIPOR3* with 13 degrees (Fig. 7, Supplementary Table S13). *TMEM86A* is a lysoplasmalogenase that may be important for regulating keratinocyte membrane properties in terminal differentiation cells¹²⁸ proposed as a candidate for tumor surface antigen in HER2 ampl BC, which coincides with our results¹²⁹. *CAPN9*, the calpains are intracellular cysteine proteases that function in various important cellular processes, including signaling, motility, apoptosis, and cellular survival¹³⁰. Down-regulated *CAPN9* is related to poorer patient clinical outcomes following endocrine therapy in BC¹³¹. *ERVE-1* is an

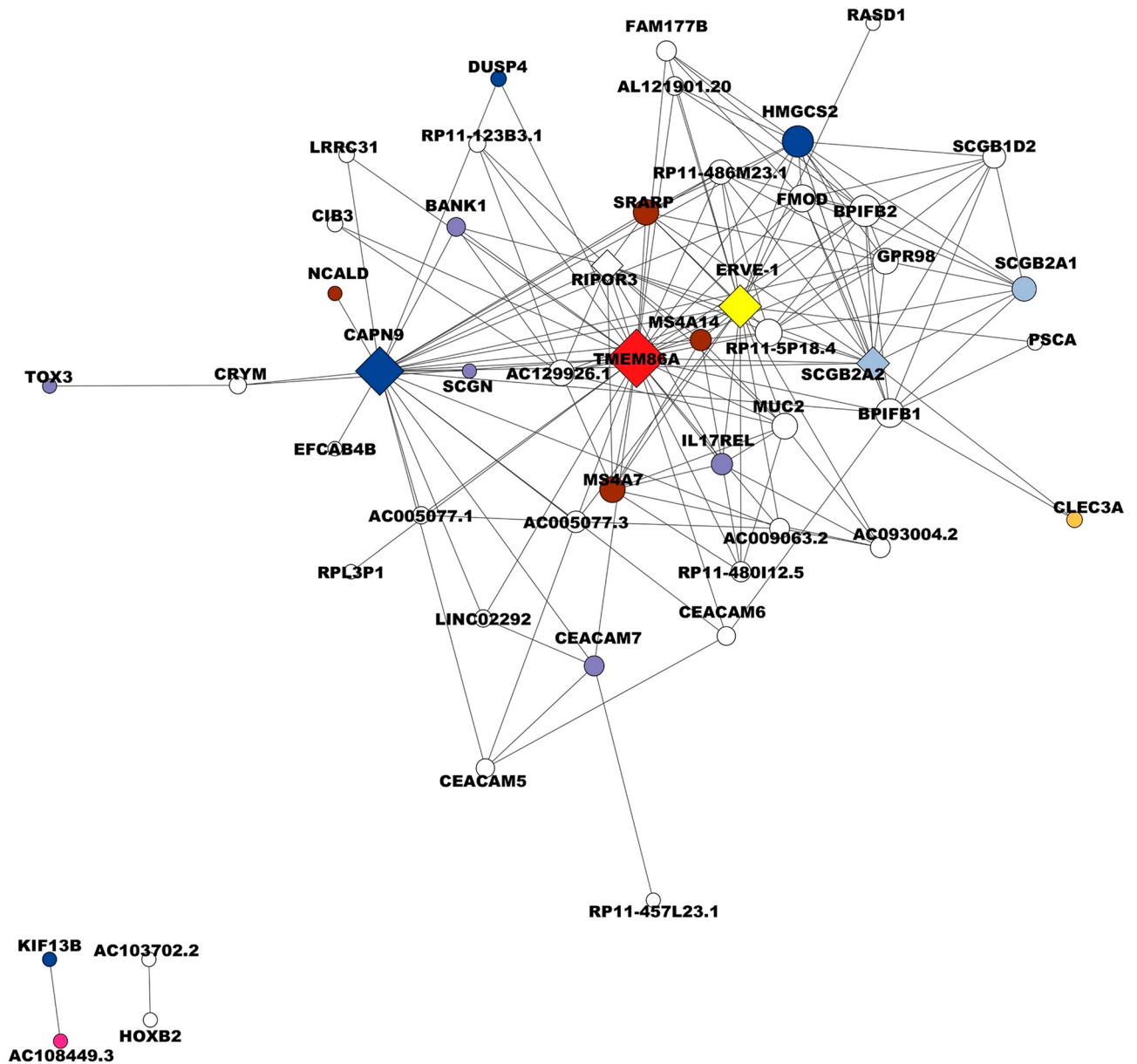


Figure 7. Visualisation of a correlation network for the HER2 ampl subtype of breast cancer. The circles in the network represent different biological functions, each distinguished by a unique color. Dark blue circles denote catalytic activity, pink circles represent ATP-dependent activity, and brown circles symbolize biological regulation. Purple circles indicate binding functions, while yellow circles are associated with multicellular organismal processes. Red circles represent functions linked to autosomal recessive congenital ichthyosis, and light blue circles denote signaling functions. The diamond shapes in the network represent hub genes, which have a high degree of connectivity within the network, indicating their important role in biological processes. The size of each node, whether a circle or diamond, corresponds to the number of connection degrees involved in biological processes. Larger nodes have more connections, indicating a higher degree of involvement in various biological processes. This image is a valuable tool for visualizing complex biological interactions and understanding the role of different functions and genes in HER2 ampl.

endogenous retroviral sequence from the family ERVs (endogenous retroviral sequence) detected in several cancers, while they remain silent in healthy tissues¹³². However, in BC, the expression of *HERV-K* (human endogenous retrovirus type K) is frequent, and *HERV-K* (HML-2) antibodies and mRNA tended to be higher in BC patients with a primary tumor, who later will develop the metastatic disease than in patients who did not develop cancer metastasis¹³³. It is important to note that most of the HER2 ampl cell lines that used for this analysis are metastatic. *SCGB2A2* encodes a glycoprotein indicated as a potential diagnostic and prognostic marker for BC¹³⁴. It is involved in biological regulation, cellular processes, response to stimulus, and signaling. A genetic rearrangement has been detected in breast tumors that overexpress *SCGB2A2*, suggesting changes in transcriptional regulation as a cause of overexpression¹³⁵. Its expression does not seem to be affected by steroid hormones^{135,136}. *SCGB2A2* is a biomarker for the early detection of bone marrow micrometastasis¹³⁴. A downregulation of *RIPOR3* in tongue cancer is associated with a worse prognosis. Its expression is related to the modulation of immune pathway function and hypermethylation, making it a promising prognostic biomarker and is linked to the immune cell infiltration of cell carcinoma of the mobile tongue¹³⁷. The functions and roles of the hub genes described above for each SBC underscore their significance in breast cancer or other cancer types. Our results demonstrate that each SBC evaluated (Basal A, Basal B, Luminal A, Luminal B, and Her2 ampl) presents a different gene expression pattern, identifying highly correlated hub genes for each subtype. Moreover, some of these genes are involved in similar biological and cellular processes including proliferation, invasion, metastasis, and immune response, among others. Therefore, these hub genes may be potential diagnostic biomarkers for each SBC.

Protein–protein interaction (PPI)-networks for each SBC

The interactions between the proteins encoded by the hub genes of each module can provide insights into the pathways or mechanisms activated in the different SBCs. For this reason, we analyzed the PPIs in the STRING database. The results showed that the PPI-network for Basal A consists of 33 nodes and 93 edges (Supplementary Fig. 4a). The protein most closely connected is ISG15 ubiquitin-like protein, which functions as a cytokine to modulate immune responses such as viral infections (14 degrees). Interestingly, its expression has been relevant in the pathogenesis of BC due to its role in motility, invasion, metastasis, and a poor response to chemotherapy and radiotherapy^{138,139}. Additionally, XAF1 counteracts the inhibitory effect of apoptosis (13 degrees)¹⁴⁰, IFIT2 involve positive regulation of the apoptotic process and response to the virus (13 degrees)¹⁴¹, RSAD2 is an interferon-inducible antiviral protein (13 degrees), and OAS2 involved in the innate immune response to viral infection (13 degrees)¹⁴². The Basal B network consists of 40 nodes and 28 edges. The most connected protein is CAV1, recognized as a tumor suppressor (10 degrees) with a dual role in BC progression¹⁰³; CD44, which participates in cell adhesion and tumor metastasis, among others (9 degrees)¹⁴³ PLAU related to apoptotic pathways in synovial fibroblast (4 degrees)¹⁴⁴; PTRE, plays a significant role in caveolae formation, ribosomal transcriptional activity in response to metabolic challenges in the adipocytes, among others (3 degrees); CAV2 involve in signal transduction, lipid metabolism, cellular growth control, and apoptosis (3 degrees)³² (Supplementary Fig. 4b). The PPI network for Luminal A consists of 17 genes and 0 edges (Supplementary Fig. 4c). The Luminal B network includes 29 genes and four edges, with the most connected protein being ANG, which induces angiogenesis (2 degrees)¹⁴⁵; and TRH releasing mature thyrotropin, a hypothalamic regulatory hormone (2 degrees)³²; CAPN6 may play a role in tumor formation by inhibiting apoptosis and promoting angiogenesis (1 degree)¹⁴⁶; TTR transports thyroid hormones and participates in proteolysis, nerve regeneration, autophagy, and glucose homeostasis (1 degree)^{32,147}; ITM2A is a type II membrane protein that may be involved in osteo- and chondrogenic differentiation (1 degree)¹⁴⁸ (Supplementary Fig. 4d). The PPI-network for HER2ampl consists of 34 genes and nine edges. The protein which the highest connection is SCGB2A2, involved in the androgen receptor signaling pathway (3 degrees)¹⁴⁹; SCGB1D2 a chemotherapeutic agent widely used for prostate cancer (2 degrees)³²; BPIFB1 involved in the innate immune response to bacterial exposure in the mouth, nasal cavities, and lungs (2 degrees); SCGB2A1 bind androgens and other steroids, a chemotherapeutic agent used for prostate cancer, under the transcriptional regulation of steroid hormones (2 degrees); CEACAM5 plays a role in cell adhesion, intracellular signaling, and tumor progression(2 degrees)^{32,150} (Supplementary Fig. 4e).

Hub gene expression in BC patients

We conducted a DEG on the GSE65194, GSE96860, and GSE52194 datasets, revealing that our main hub genes, *IFIT3*, *ETS1*, *CA4*, and *TMEM86A*, were overexpressed in each SBC with their respective adjusted *p* values, while *MS4A15* only showed significant overexpression in dataset GSE52194 (Supplementary Table S14). Interestingly, *IFIT3* has been recognized as a possible biomarker for predicting the response to chemotherapy and radiation in specific human cancers (breast, lung, prostate, head and neck cancer, and glioma)^{86,151}. Additionally, the expression of *IFIT3* mRNA could function as a predictive marker for the efficiency of immunostimulant treatments in breast cancer cells¹⁵². Recently, *ETS1* was identified as a predictive biomarker for TNBC, notably enhancing the progression of TNBC by activating the YAP signaling pathway⁹⁷, aligning with our findings. *CA4* has been associated with the prognosis of lung adenocarcinoma¹⁵³. Finally, *TMEM86A* has been suggested as a candidate for a tumor surface antigen in HER2 ampl BC¹²⁹, and another for differential expression in various cancer types, including liver, colon, and gastric¹⁵⁴. A differential gene expression analysis on the GSE134359 dataset, which includes long non-coding RNA (lncRNA) from SBCs, identified ENSG00000259723 and AL033519.3 as the main hub genes for Luminal A and Luminal B, with an adjusted *p* values of 3.43E−01 and 2.46E−05, respectively. In breast cancer, lncRNA functions as both a promoter and inhibitor of metastasis¹¹⁵. It is important to recognize that research into lncRNAs in breast cancer is still evolving, with more comprehensive studies required to fully grasp their functions and mechanisms. Nevertheless, progress in this area has already

demonstrated that lncRNAs significantly impact BC biology. They offer promising potential as therapeutic targets, highlighting the importance of continuing exploration in this domain²¹.

Ultimately, the search for molecules capable of identifying and distinguishing SBCs, as well as serving as therapeutic targets, remains a critical endeavor in cancer treatment. Correct diagnosis enables the administration of the most effective treatments. Moreover, it is crucial to monitor treatment efficacy, prevent cancer cell resurgence, and avert the development of drug resistance, all to eliminate breast cancer cells. Our research underscores the identification of pivotal genes whose overexpression is specifically correlated with certain SBC. While further validation in diverse patient cohorts of various SBCs is necessary, our findings align with those of Charafe-Jauffret et al.¹⁵⁵, who utilized 31 breast cancer cell lines and DNA microarrays to identify, among others, three key genes we also recognized: *ANXA8* (Basal A), *ETS1* and *CAVI* (Basal B), associating them with the basal subtype of breast cancer. Calderon-Gonzalez et al.¹⁵⁶ and Stein et al.¹⁵⁷ have also shown that *ANXA8* is overexpressed in basal subtype cell lines of both human and mouse origin. Regarding *ETS1*, its association with local metastasis in TNBC patient samples was noted⁹⁷. Xie et al.⁹⁹ reported *adamts6* (Basal B) as a potential prognostic marker in breast cancer, supported by a multivariate analysis of patient tumor samples and cell lines (Supplementary Table 15).

In the case of the Luminal A subtype, the *ca4* gene hub identified in our study has also been detected and proposed as a prognostic biomarker from samples of patients with invasive cancer¹⁵⁸. For the Luminal B subtype, none of the gene hubs identified in this work have been proposed as potential biomarkers in other studies. However, Wu et al.¹⁵⁹ suggest the *capn6* gene as a prognostic marker for BC. While it is not the main gene hub correlating with this SBC in our analysis, it is present in the skyblue module. Similarly, the presence of *PPP2R2B*, *HPD*, *ADGRG7*, *ADCY4*, and *ITM2A* supports the correlation of the expression of these genes with the Luminal B subtype (Supplementary Table 15), underscoring the complexity and the nuanced understanding required to navigate the molecular landscape of SBCs. Additionally, these insights highlight the potential of specific genes to serve as biomarkers for BC, emphasizing the need for further research to validate these associations and understand their implications for diagnosis and treatment strategies.

Finally, regarding the gene hubs associated with the HER2 ampl subtype¹³⁴, *SCGB2A2* is predominantly overexpressed in BC patients and shows a high correlation with increased expression and the presence of micrometastases¹³⁴. The similarities and differences in identifying these gene hubs may stem from the origin of the samples, whether they are cell lines or patient tumor samples. While these sources have been shown to possess great similarities¹⁶⁰, they also exhibit certain differences, particularly when compared with samples from advanced-stage tumors^{161,162}. These disparities may also relate to the number of samples used, the types of algorithms applied, and the specific correlation or approach sought in the research. Additionally, this variability underscores the well-documented heterogeneity in breast cancer, highlighting the complexities of diagnosing and treating this multifactorial disease.

Conclusions

Identifying biomarkers that enable the precise diagnosis of specific SBC is crucial for understanding the type of tumor, its prognosis or stage, and selecting the most effective treatment. This study identified a specific module with the highest correlation for each SBC, pinpointing at least one hub gene with the highest correlation for each subtype: Basal A: *IFIT3*; Basal B: *ETS1*; Luminal A: *ENSG00000259723*; Luminal B: *AL033519.3*; HER2amp: *TMEM86A*. These genes are involved in defense response, cell process regulation, and transport for each respective SBC.

Validation analysis of the highly connected hub genes expressed in patients with SBCs, using the external dataset GSE65194, revealed significant differential expression of *IFIT3*, *ETS1*, *CA4*, and *TMEM86A*. Our findings could provide a theoretical basis for enhancing prognosis and diagnosis through potential SBC biomarkers. Nevertheless, the mechanisms underlying these associations need to be explored in future studies.

This research proposes further analysis of these targets as potential biomarkers. It represents the first study to use a large sample of breast cancer cell lines and WGCNA to identify potential biomarkers associated with SBCs. The development of co-expression networks in each SBC may elucidate possible biomarkers for risk assessment, prognostic determination, screening, differential diagnosis, treatment response prediction, and disease progression monitoring. These hubs could be promising targets in drug development for monitoring or treating SBCs. A significant strength of this study is its focus on SBCs rather than BC in general, offering a more nuanced understanding of subtype-specific biomarkers.

Data availability

The dataset analyzed during the current study is available in DepMap database version 21Q1 with the name CCLE_expression.csv. The sample code and data are publicly available at GitHub: https://github.com/DanieMaresQ/WGCNA_BC_Cell.L.

Received: 13 July 2023; Accepted: 10 May 2024

Published online: 29 May 2024

References

1. World Health Organization. *Breast cancer*, accessed 12 April 2021; <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (2021).
2. Henderson Craig, I. *Breast Cancer Fundamentals of Evidence-Based Disease Management* 1–23 (Oxford University Press, 2015).
3. Dunnwald, L. K., Rossing, M. A. & Li, C. I. Hormone receptor status, tumor characteristics, and prognosis: A prospective cohort of breast cancer patients. *Breast Cancer Res.* **9**, R6. <https://doi.org/10.1186/bcr1639> (2007).
4. Harbeck, N. et al. Breast cancer. *Nat. Rev. Dis. Primers* **5**, 66. <https://doi.org/10.1038/s41572-019-0111-2> (2019).

5. Lüönd, F., Tiede, S. & Christofori, G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. *Br. J. Cancer*. **125**, 164–175. <https://doi.org/10.1038/s41416-021-01328-7> (2021).
6. Mei, J., Zhao, J. & Fu, Y. Molecular classification of breast cancer using the mRNA expression profiles of immune-related genes. *Sci. Rep.* **10**, 1–9. <https://doi.org/10.1038/s41598-020-61710-y> (2020).
7. Tomlins, S. A. *et al.* Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur. Urol.* **68**, 555–567. <https://doi.org/10.1016/j.euro.2015.04.033> (2015).
8. Hamam, R. *et al.* Circulating microRNAs in breast cancer: Novel diagnostic and prognostic biomarkers. *Cell Death Dis.* **8**, e3045–e3045. <https://doi.org/10.1038/cddis.2017.440> (2017).
9. Mautner, B. D., Schmidt, K. V. & Brennan, M. B. New diagnostic techniques and treatments for early breast cancer. *Semin. Oncol. Nurs.* **16**, 185–196. <https://doi.org/10.1053/sonc.2000.8112> (2000).
10. Waks, A. G. & Winer, E. P. Breast cancer treatment: A review. *JAMA* **321**, 288–300. <https://doi.org/10.1001/jama.2018.19323> (2019).
11. Peng, L., Xu, T., Long, T. & Zuo, H. Association between BRCA status and P53 status in breast cancer: A meta-analysis. *Med. Sci. Monit.* **22**, 1939–1945. <https://doi.org/10.12659/MSM.896260> (2016).
12. Dees, S., Pontiggia, L., Jasmin, J.-F. & Mercier, I. Phosphorylated STAT3 (Tyr705) as a biomarker of response to pimozone treatment in triple-negative breast cancer. *Cancer Biol. Ther.* **21**, 506–521. <https://doi.org/10.1080/15384047.2020.1726718> (2020).
13. Brett, J. O., Spring, L. M., Bardia, A. & Wander, S. A. ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer. *Breast Cancer Res.* **23**, 85. <https://doi.org/10.1186/s13058-021-01462-3> (2021).
14. Gam, L.-H. Breast cancer and protein biomarkers. *World J. Exp. Med.* **2**, 86–91. <https://doi.org/10.5493/wjem.v2.i5.86> (2012).
15. Adhami, M., MotieGhader, H., Haghdoost, A. A., Afshar, R. M. & Sadeghi, B. Gene co-expression network approach for predicting prognostic microRNA biomarkers in different subtypes of breast cancer. *Genomics* **112**, 135–143. <https://doi.org/10.1016/j.ygeno.2019.01.010> (2020).
16. Liu, Z., Li, M., Hua, Q., Li, Y. & Wang, G. Identification of an eight-lncRNA prognostic model for breast cancer using WGCNA network analysis and a Cox? Proportional hazards model based on L1-penalized estimation. *Int. J. Mol. Med.* **44**, 1333–1343. <https://doi.org/10.3892/ijmm.2019.4303> (2019).
17. Sideris, N., Dama, P., Bayraktar, S., Stiff, T. & Castellano, L. LncRNAs in breast cancer: A link to future approaches. *Cancer Gene Ther.* **29**, 1–12. <https://doi.org/10.1038/s41417-022-00487-w> (2022).
18. Fachal, L. *et al.* Fine-mapping of 150 breast cancer risk regions identifies 178 high confidence target genes. <https://doi.org/10.1101/521054v1> (2019).
19. Zhang, H. *et al.* Integrated analysis of oncogenic networks in colorectal cancer identifies GUC2A as a molecular marker. *Biochem. Res. Int.* **2019**, 1–13. <https://doi.org/10.1155/2019/6469420> (2019).
20. Fachal, L. *et al.* Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat. Genet.* **52**, 56–73. <https://doi.org/10.1038/s41588-019-0537-1> (2020).
21. Zugazagoitia, J. *et al.* Current challenges in cancer treatment. *Clin. Ther.* **38**, 1551–1566. <https://doi.org/10.1016/j.clinthera.2016.03.026> (2016).
22. Li, M. *et al.* Transmembrane protein 170B is a novel breast tumorigenesis suppressor gene that inhibits the Wnt/ β -catenin pathway. *Cell Death Dis.* **9**, 1–15. <https://doi.org/10.1038/s41419-017-0128-y> (2018).
23. Meurer, S. K., Tezcan, O., Lammers, T. & Weiskirchen, R. Differential regulation of Lipocalin 2 (LCN2) in doxorubicin-resistant 4T1 triple negative breast cancer cells. *Cell. Signal* **74**, 109731. <https://doi.org/10.1016/j.cellsig.2020.109731> (2020).
24. Charan, M. *et al.* Macrophage migration inhibitory factor inhibition as a novel therapeutic approach against triple-negative breast cancer. *Cell Death Dis.* **11**, 1–15. <https://doi.org/10.1038/s41419-020-02992-y> (2020).
25. Egusquiaguirre, S. P., Yeh, J. E., Walker, S. R., Liu, S. & Frank, D. A. The STAT3 target gene TNFRSF1A modulates the NF- κ B pathway in breast cancer cells. *Neoplasia* **20**, 489–498. <https://doi.org/10.1016/j.neo.2018.03.004> (2018).
26. Yerlikaya, A., Okur, E. & Ulukaya, E. The p53-independent induction of apoptosis in breast cancer cells in response to proteasome inhibitor bortezomib. *Tumour Biol.* **33**, 1385–1392. <https://doi.org/10.1007/s13277-012-0386-3> (2012).
27. de Amorim, Í. S. S. *et al.* Opposite effects of demethylating treatment on VEGFA and HIF1A expression in MDA-MB-231 breast cancer cell line in hypoxic microenvironment. *Gene Rep.* **16**, 100403. <https://doi.org/10.1016/j.genrep.2019.100403> (2019).
28. Dustin, D. *et al.* RON signalling promotes therapeutic resistance in ESR1 mutant breast cancer. *Br. J. Cancer.* **124**, 191–206. <https://doi.org/10.1038/s41416-020-01174-z> (2021).
29. Amgalan, D. *et al.* A small-molecule allosteric inhibitor of BAX protects against doxorubicin-induced cardiomyopathy. *Nat. Cancer.* **1**, 315–328. <https://doi.org/10.1038/s43018-020-0039-1> (2020).
30. Zhuang, S., Li, L., Zang, Y., Li, G. & Wang, F. RRM2 elicits the metastatic potential of breast cancer cells by regulating cell invasion, migration and VEGF expression via the PI3K/AKT signaling. *Oncol. Lett.* **19**, 3349–3355. <https://doi.org/10.3892/ol.2020.11428> (2020).
31. Wang, K. *et al.* Knockdown of MMP-1 inhibits the progression of colorectal cancer by suppressing the PI3K/Akt/c-myc signaling pathway and EMT. *Oncol. Rep.* **43**, 1103–1112. <https://doi.org/10.3892/or.2020.7490> (2020).
32. Safran, M. *et al.* The gene cards suite. In *Practical Guide to Life Science Databases* (eds Abugessaisa, I. & Kasukawa, T.) 27–56 (Springer, 2021). https://doi.org/10.1007/978-981-16-5812-9_2.
33. Cox, B. & Sneyd, M. Bias in breast cancer research in the screening era. *Breast* **22**, 1041–1045. <https://doi.org/10.1016/j.breast.2013.07.046> (2013).
34. Renfro, L. A., An, M. W. & Mandrekar, S. J. Precision oncology: A new era of cancer clinical trials. *Cancer Lett.* **387**, 121–126. <https://doi.org/10.1016/j.canlet.2016.03.015> (2017).
35. Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: Characterization, validation, and prognostic value. *PLoS Med.* **10**, e1001453. <https://doi.org/10.1371/journal.pmed.1001453> (2013).
36. Liu, Z., Bai, Y., Xie, F., Miao, F. & Du, F. Comprehensive analysis for identifying diagnostic and prognostic biomarkers in colon adenocarcinoma. *DNA Cell Biol.* **39**, 599–614. <https://doi.org/10.1089/dna.2019.5215> (2020).
37. Shi, Z., Derow, C. K. & Zhang, B. Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. *BMC Syst. Biol.* **4**, 74. <https://doi.org/10.1186/1752-0509-4-74> (2010).
38. Abu, N. *et al.* In vivo antitumor and antimetastatic effects of flavokawain B in 4T1 breast cancer cell-challenged mice. *Drug Des. Dev. Ther.* **9**, 1401–1417. <https://doi.org/10.2147/dddt.s67976.2> (2015).
39. Liu, Z. *et al.* Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA. *J. Cell. Biochem.* **120**, 5137–5150. <https://doi.org/10.1002/jcb.27790> (2018).
40. Zheng, P. F., Chen, L. Z., Guan, Y. Z. & Liu, P. Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. *Sci. Rep.* **11**, 6711. <https://doi.org/10.1038/s41598-021-86207-0> (2021).
41. Bao, C. *et al.* Exploring specific prognostic biomarkers in triple-negative breast cancer. *Cell Death Dis.* **10**, 807 (2019).
42. Bettaieb, A. Precision medicine in breast cancer: Reality or utopia?. *J. Transl. Med.* **15**, 139. <https://doi.org/10.1186/s12967-017-1239-z> (2017).
43. Ghandi, M. *et al.* Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508. <https://doi.org/10.1038/s41586-019-1186-3> (2019).
44. Broad DepMap. *DepMap 21Q1 Public* (p. 12817867672 Bytes). figshare. <https://doi.org/10.6084/M9.FIGSHARE.13681534.V2> (2021).

45. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
46. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. *Bioinformatics* **24**, 719–720. <https://doi.org/10.1093/bioinformatics/btm563> (2008).
47. Horvath, S. *Weighted Network Analysis: Applications in Genomics and Systems Biology* 179–206 (Springer, 2011).
48. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
49. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592. <https://doi.org/10.1002/pro.3715> (2023).
50. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1093/nar/gkac963> (2019).
51. Lee, S. *et al.* Landscape analysis of adjacent gene rearrangements reveals BCL2L14-ETV6 gene fusions in more aggressive triple-negative breast cancer. *Proc. Natl. Acad. Sci.* **117**, 9912–9921. <https://doi.org/10.1073/pnas.1921333117> (2020).
52. Kammerer, S. *et al.* KCN3 is a new independent prognostic marker for estrogen receptor positive breast cancer patients. *Oncotarget* **7**, 84705. <https://doi.org/10.18632/oncotarget.13224> (2016).
53. Pavón, L., del Carmen Jiménez, M. & Garcés, M. E. *Inmunología Molecular, Celular y Traslacional* 874–875 (Wolters Kluwer Health, 2020).
54. Goff, S. L. & Danforth, D. N. The role of immune cells in breast tissue and immunotherapy for the treatment of breast cancer. *Clin. Breast Cancer* **21**, e63–e73. <https://doi.org/10.1016/j.clbc.2020.06.011> (2021).
55. Landskron, G., De la Fuente, M., Thuwajit, P., Thuwajit, C. & Hermoso, M. A. Chronic inflammation and cytokines in the tumor microenvironment. *J. Immunol. Res.* **2014**, 149185. <https://doi.org/10.1155/2014/149185> (2014).
56. Xiong, Y. *et al.* High expression of CD58 and ALDH1A3 predicts a poor prognosis in basal-like breast cancer. *Anticancer Res.* **42**, 5223–5232. <https://doi.org/10.21873/anticancer.16029> (2022).
57. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409. <https://doi.org/10.1038/nature11154> (2012).
58. Clark, A. G. & Vignjevic, D. M. Modes of cancer cell invasion and the role of the microenvironment. *Curr. Opin. Cell Biol.* **36**, 13–22. <https://doi.org/10.1016/j.ceb.2015.06.004> (2015).
59. Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *J. Clin. Oncol.* **28**, 3271–3277. <https://doi.org/10.1200/JCO.2009.25.9820> (2010).
60. Du, Y. & Wang, P. Upregulation of MIIP regulates human breast cancer proliferation, invasion and migration by mediated by IGFBP2. *Pathol. Res. Pract.* **215**, 152440. <https://doi.org/10.1016/j.prrp.2019.152440> (2019).
61. Wang, N. *et al.* CXCL1 derived from tumor-associated macrophages promotes breast cancer metastasis via activating NF- κ B/SOX4 signaling. *Cell Death Dis.* **9**, 880. <https://doi.org/10.1038/s41419-018-0876-3> (2018).
62. Kaur, P. *et al.* Identification of putative actionable alterations in clinically relevant genes in breast cancer. *Br. J. Cancer* **125**, 1270–1284. <https://doi.org/10.1038/s41416-021-01522-7> (2021).
63. Choromanska, A. *et al.* Modifications of plasma membrane organization in cancer cells for targeted therapy. *Molecules* **26**, 1850. <https://doi.org/10.3390/molecules26071850> (2021).
64. Padua, M. B. *et al.* Dependence receptor UNC5A restricts luminal to basal breast cancer plasticity and metastasis. *Breast Cancer Res.* **20**, 35. <https://doi.org/10.1186/s13058-018-0964-2> (2018).
65. Arnal, J.-F. *et al.* Membrane and nuclear estrogen receptor alpha actions: From tissue specificity to medical implications. *Physiol. Rev.* **97**, 1045–1087. <https://doi.org/10.1152/physrev.00024.2016> (2017).
66. Eroles, P., Bosch, A., Pérez-Fidalgo, J. A. & Lluch, A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treat. Rev.* **38**, 698–707. <https://doi.org/10.1016/j.ctrv.2011.11.005> (2012).
67. Santarpia, L. *et al.* Deciphering and targeting oncogenic mutations and pathways in breast cancer. *Oncologist* **21**, 1063–1078. <https://doi.org/10.1634/theoncologist.2015-0369> (2016).
68. Gao, J. & Swain, S. Luminal a breast cancer and molecular assays: A review. *Oncologist* **23**, 556–565. <https://doi.org/10.1634/theoncologist.2017-0535> (2018).
69. Mayayo-Peralta, I., Prekovic, S. & Zwart, W. Estrogen receptor on the move: Cistromic plasticity and its implications in breast cancer. *Mol Aspects Med.* **78**, 100939. <https://doi.org/10.1016/j.mam.2020.100939> (2021).
70. Xiong, G. F. & Xu, R. Function of cancer cell-derived extracellular matrix in tumor progression. *JCMT* **2**, 357–364. <https://doi.org/10.20517/2394-4722.2016.08> (2016).
71. Tan, Q. *et al.* Breast cancer cells interact with tumor-derived extracellular matrix in a molecular subtype-specific manner. *Biomater. Adv.* **146**, 213301. <https://doi.org/10.1016/j.bioadv.2023.213301> (2023).
72. Diaz Bessone, M. I., Gattas, M. J., Laporte, T., Tanaka, M. & Simian, M. The tumor microenvironment as a regulator of endocrine resistance in breast cancer. *Front. Endocrinol.* **10**, 547. <https://doi.org/10.3389/fendo.2019.00547> (2019).
73. Cheang, M. *et al.* Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J. Natl. Cancer Inst.* **101**, 736–750. <https://doi.org/10.1093/jnci/djp082> (2009).
74. Toy, W. *et al.* Activating ESR1 mutations differentially affect the efficacy of ER antagonists. *Cancer Discov.* **7**, 277–287. <https://doi.org/10.1158/2159-8290.CD-15-1523> (2017).
75. Thapa, M. & Dallmann, G. Role of coenzymes in cancer metabolism. *Semin. Cell Dev. Biol.* **98**, 44–53. <https://doi.org/10.1186/s12943-022-01668-9> (2020).
76. Moasser, M. M. The oncogene HER2: Its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* **26**, 6469–6487. <https://doi.org/10.1038/sj.onc.1210477> (2007).
77. Razavi, P. *et al.* The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer Cell.* **34**, 427–438.e6. <https://doi.org/10.1016/j.ccell.2018.08.008> (2018).
78. Duffy, M. J., Synnott, N. C. & Crown, J. Mutant p53 in breast cancer: Potential as a therapeutic target and biomarker. *Breast Cancer Res. Treat.* **170**, 213–219 (2018).
79. Kong, X. *et al.* Mechanism of trastuzumab resistance caused by HER-2 mutation in breast carcinomas. *Cancer Manag. Res.* **11**, 5971–5982. <https://doi.org/10.2147/CMAR.S194137> (2019).
80. Farahmand, M. *et al.* Epstein–Barr virus and risk of breast cancer: A systematic review and meta-analysis. *Future Oncol.* **15**, 2873–2885. <https://doi.org/10.2217/fon-2019-0232> (2019).
81. Gupta, I. *et al.* Presence of high-risk HPVs, EBV, and MMTV in human triple-negative breast cancer. *Hum. Vaccines Immunother.* **17**, 4457–4466. <https://doi.org/10.1080/21645515.2021.1975452> (2021).
82. Ingerslev, K. *et al.* The prevalence of EBV and CMV DNA in epithelial ovarian cancer. *Infect. Agents Cancer* **14**, 7. <https://doi.org/10.1186/s13027-019-0223-z> (2019).
83. Ahrens, T. D. *et al.* The role of proteoglycans in cancer metastasis and circulating tumor cell analysis. *Front. Cell Dev. Biol.* **8**, 749. <https://doi.org/10.3389/fcell.2020.00749> (2020).
84. Gwenaëlle Lemoine. GWENA: Pipeline for augmented co-expression analysis. (R package Version 1.0.1); https://rdrr.io/bioc/GWENA/man/get_hub_degree.html (2021).
85. Li, X. *et al.* Upregulation of ADAR promotes breast cancer progression and serves as a potential therapeutic target. *J. Oncol.* **2021**, 1–18. <https://doi.org/10.1155/2021/2012903> (2021).

86. Pidugu, V. K., Pidugu, H. B., Wu, M. M., Liu, C. J. & Lee, T. C. Emerging functions of human IFIT proteins in cancer. *Front. Mol. Biosci.* **6**, 148. <https://doi.org/10.3389/fmolb.2019.00148> (2019).
87. Khazei, K. *et al.* Transcriptome profiling of curcumin-treated T47D human breast cancer cells by a system-based approach. *Gene Rep.* **27**, 101556. <https://doi.org/10.1016/j.gene.2022.101556> (2022).
88. Lamsal, A. *et al.* Opposite and dynamic regulation of the interferon response in metastatic and non-metastatic breast cancer. *J. Cell Commun. Signal* **21**, 1–15. <https://doi.org/10.1186/s12964-023-01062-y> (2023).
89. Rouette, A. *et al.* Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers. *Sci. Rep.* **6**, 34019. <https://doi.org/10.1038/srep34019> (2016).
90. Flick, K. & Kaiser, P. Protein degradation and the stress response. *Semin. Cell Dev. Biol.* **23**, 515–522. <https://doi.org/10.1016/j.semcdb.2012.01.019> (2012).
91. Maia Falcão, R. *et al.* The expression of the immunoproteasome subunit PSMB9 is related to distinct molecular subtypes of uterine leiomyosarcoma. *Cancers* **14**, 5007. <https://doi.org/10.3390/cancers14205007> (2022).
92. Sherman, B. T. *et al.* DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, 216–221. <https://doi.org/10.1093/nar/gkac194> (2022).
93. Thomas, P. D. *et al.* PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **31**, 334–341. <https://doi.org/10.1093/nar/gkg115> (2003).
94. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752. <https://doi.org/10.1038/35021093> (2000).
95. Wu, M. *et al.* Targeting ETS1 with RNAi-based supramolecular nanoassemblies for multidrug-resistant breast cancer therapy. *J. Control Release* **253**, 110–121. <https://doi.org/10.1016/j.jconrel.2017.03.011> (2017).
96. Mei, D. *et al.* Microarray profile analysis identifies ETS1 as potential biomarker regulated by miR-23b and modulates TCF4 in gastric cancer. *World J. Surg. Oncol.* **19**, 1–14. <https://doi.org/10.1186/s12957-021-02417-w> (2021).
97. Li, Y. *et al.* ETS1 is a prognostic biomarker of triple-negative breast cancer and promotes the triple-negative breast cancer progression through the YAP signaling. *Am. J. Cancer Res.* **12**, 5074–5084 (2022).
98. Cal, S. & López-Otín, C. ADAMTS proteases and cancer. *Matrix Biol.* **44**, 77–85. <https://doi.org/10.1016/j.matbio.2015.01.013> (2015).
99. Xie, Y. *et al.* ADAMTS6 suppresses tumor progression via the ERK signaling pathway and serves as a prognostic marker in human breast cancer. *Oncotarget* **7**, 61273. <https://doi.org/10.18632/oncotarget.11341> (2016).
100. Li, J., Xu, Q., Wang, W. & Sun, S. MIR100HG: A credible prognostic biomarker and an oncogenic lncRNA in gastric cancer. *Biosci. Rep.* <https://doi.org/10.1042/BSR20190171> (2019).
101. Crudele, F. *et al.* The network of non-coding RNAs and their molecular targets in breast cancer. *Mol. Cancer* **19**, 1–18. <https://doi.org/10.1186/s12943-020-01181-x> (2020).
102. Alkhateeb, A. *et al.* Transcriptomics signature from next-generation sequencing data reveals new transcriptomic biomarkers related to prostate cancer. *Cancer Inform.* **18**, 1–12. <https://doi.org/10.1177/1176935119835522> (2019).
103. Qian, X.-L. *et al.* Caveolin-1: A multifaceted driver of breast cancer progression and its application in clinical treatment. *Oncotargets Ther.* **12**, 1539–1552. <https://doi.org/10.2147/OTT.191317> (2019).
104. Erdemli, H. K. *et al.* Is serum caveolin-1 a useful biomarker for progression in patients with colorectal cancer. *Clin Lab* **62**, 401–408. <https://doi.org/10.7754/clin.lab.2015.150719> (2016).
105. Gumulec, J. *et al.* Caveolin-1 as a potential high-risk prostate cancer biomarker. *Oncol. Rep.* **27**, 831–841. <https://doi.org/10.3892/or.2011.1587> (2012).
106. Leiser, D. *et al.* Role of caveolin-1 as a biomarker for radiation resistance and tumor aggression in lung cancer. *PLoS ONE* **16**, e0258951. <https://doi.org/10.1371/journal.pone.0258951> (2021).
107. Mahmood, J., Murti, S. C., Zaveri, S. R., Shukla, H. D. & Vujaskovic, Z. Caveolin-1: A novel prognostic biomarker for radioresistance in non-small cell lung carcinoma (NSCLC) and prostate cancer. *IJROBP* **96**, E589–E590. <https://doi.org/10.1016/j.ijrobp.2016.06.2105> (2016).
108. Panic, A. *et al.* The biomarker potential of Caveolin-1 in penile cancer. *Front. Oncol.* **11**, 606122. <https://doi.org/10.3389/fonc.2021.606122> (2021).
109. Al Abo, M. *et al.* Adaptive stress response genes associated with breast cancer subtypes and survival outcomes reveal race-related differences. *NPJ Breast Cancer* **8**, 73. <https://doi.org/10.1038/s41523-022-00431-z> (2022).
110. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688. <https://doi.org/10.1093/nar/gkz966> (2020).
111. Zhang, Y. *et al.* Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nat. Struct. Mol. Biol.* **23**, 522–530. <https://doi.org/10.1038/nsmb.3211> (2016).
112. Xue, Y. *et al.* Database resources of the national genomics data center, China National Center for Bioinformation in 2022. *Nucleic Acids Res.* **50**, 27–38. <https://doi.org/10.1093/nar/gkab951> (2022).
113. Xu, Y. *et al.* Carbonic anhydrase 4 serves as a clinicopathological biomarker for outcomes and immune infiltration in renal cell carcinoma, lower grade glioma, lung adenocarcinoma and uveal melanoma. *J. Cancer* **11**, 6101. <https://doi.org/10.7150/jca.46902> (2020).
114. Wang, B. *et al.* Carbonic anhydrase IV inhibits cell proliferation in gastric cancer by regulating the cell cycle. *Oncol. Lett.* **20**, 1–1. <https://doi.org/10.3892/ol.2020.11865> (2020).
115. Liu, L., Zhang, Y. & Lu, J. The roles of long noncoding RNAs in breast cancer metastasis. *Cell Death Dis.* **11**, 749. <https://doi.org/10.1038/s41419-020-02954-4> (2020).
116. Liu, C., Hu, C., Li, J., Jiang, L. & Zhao, C. Identification of epithelial-mesenchymal transition-related lncRNAs that associated with the prognosis and immune microenvironment in colorectal cancer. *Front. Mol. Biosci.* **8**, 633951. <https://doi.org/10.3389/fmolb.2021.633951> (2021).
117. Liu, S. *et al.* A novel lncRNA ROPM-mediated lipid metabolism governs breast cancer stem cell properties. *J. Hematol. Oncol.* **14**, 1–23. <https://doi.org/10.1186/s13045-021-01194-z> (2021).
118. Liu, Y., Tingart, M., Lecouturier, S., Li, J. & Eschweiler, J. Identification of co-expression network correlated with different periods of adipogenic and osteogenic differentiation of BMSCs by weighted gene co-expression network analysis (WGCNA). *BMC Genom.* **22**, 1–15. <https://doi.org/10.1186/s12864-021-07584-4> (2021).
119. Su, R. *et al.* Construction of a ceRNA network of hub genes affecting immune infiltration in ovarian cancer identified by WGCNA. *BMC Cancer* **21**, 970. <https://doi.org/10.1186/s12885-021-08711-w> (2021).
120. Fang, Y., Yu, H. & Zhou, H. MS4A15 acts as an oncogene in ovarian cancer through reprogramming energy metabolism. *Biochem. Biophys. Res. Commun.* **598**, 47–54. <https://doi.org/10.1016/j.bbrc.2022.01.128> (2022).
121. Sun, L., Zhang, Y. & Zhang, C. Distinct expression and prognostic value of MS4A in gastric cancer. *Open Med.* **13**, 178–188. <https://doi.org/10.1515/med-2018-0028> (2018).
122. Liu, C. *et al.* Upregulation of LINC00511 expression by DNA hypomethylation promotes the progression of breast cancer. *Gland Surg.* **10**, 1418. <https://doi.org/10.21037/gs-21-84> (2021).
123. Lo, P. H. Y., Tanikawa, C., Katagiri, T., Nakamura, Y. & Matsuda, K. Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. *Oncol. Rep.* **33**, 267–273. <https://doi.org/10.3892/or.2014.3581> (2015).
124. Yang, R. *et al.* High expression of PAMR1 predicts favorable prognosis and inhibits proliferation, invasion, and migration in cervical cancer. *Front. Oncol.* **11**, 742017. <https://doi.org/10.3389/fonc.2021.742017> (2021).

125. Lee, C. *et al.* Transthyretin stimulates tumor growth through regulation of tumor. *Immune Endothel. Cells. J. Immunol.* **202**, 991–1002. <https://doi.org/10.4049/jimmunol.1800736> (2019).
126. Sharma, M., Khan, S., Rahman, S. & Singh, L. R. The extracellular protein, transthyretin is an oxidative stress biomarker. *Front. Physiol.* **10**, 1–8. <https://doi.org/10.3389/fphys.2019.00005> (2019).
127. Ingenbleek, Y. Plasma transthyretin is a nutritional biomarker in human morbidities. *Front. Med.* **16**, 540–550. <https://doi.org/10.1007/s11684-022-0940-3> (2022).
128. Zhang, H. *et al.* Exploration of novel candidate genes involved in epidermal keratinocyte differentiation and skin barrier repair in man. *Differentiation* **119**, 19–27. <https://doi.org/10.1016/j.diff.2021.04.001> (2021).
129. Schettini, F. *et al.* Identification of cell surface targets for CAR-T cell therapies and antibody-drug conjugates in breast cancer. *ESMO Open* **6**, 100102. <https://doi.org/10.1016/j.esmoop.2021.100102> (2021).
130. Goll, D. E., Thompson, V. F., Li, H., Wei, W. & Cong, J. The calpain system. *Physiol. Rev.* **83**, 731–801. <https://doi.org/10.1152/physrev.00029.2002> (2003).
131. Davis, J. *et al.* Low calpain-9 is associated with adverse disease-specific survival following endocrine therapy in breast cancer. *BMC Cancer* **14**, 995. <https://doi.org/10.1186/1471-2407-14-995> (2014).
132. Vergara Bermejo, A., Ragonnaud, E., Daradoumis, J. & Holst, P. Cancer associated endogenous retroviruses: Ideal immune targets for adenovirus-based immunotherapy. *Int. J. Mol. Sci.* **21**, 4843. <https://doi.org/10.3390/ijms21144843> (2020).
133. Wang-Johanning, F. *et al.* Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. *Int. J. Cancer* **134**, 587–595. <https://doi.org/10.1002/ijc.28389> (2014).
134. Talaat, I. M. *et al.* Bone marrow mammaglobin-1 (SCGB2A2) immunohistochemistry expression as a breast cancer specific marker for early detection of bone marrow micrometastases. *Sci. Rep.* **10**, 1–12. <https://doi.org/10.1038/s41598-020-70012-2> (2020).
135. Watson, M. A., Darrow, C., Zimonjic, D. B., Popescu, N. C. & Fleming, T. P. Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13. *Oncogene* **16**, 817–824. <https://doi.org/10.1038/sj.onc.1201597> (1998).
136. Span, P. N. *et al.* Mammaglobin is associated with low-grade, steroid receptor-positive breast tumors from postmenopausal patients, and has independent prognostic value for relapse-free survival time. *J. Clin. Oncol.* **22**, 691–698. <https://doi.org/10.1200/JCO.2004.01.07> (2004).
137. Zhang, K. *et al.* Downregulated expression of RIPOR3 correlated with immune infiltrates predicts poor prognosis in oral tongue cancer. *Med. Sci. Monit.* **28**, e935055-1-e935055-17. <https://doi.org/10.12659/MSM.935055> (2022).
138. Albert, M. *et al.* ISG15 is a novel regulator of lipid metabolism during vaccinia virus infection. *Microbiol. Spectr.* **10**, e03893-e3922. <https://doi.org/10.1128/spectrum.03893-22> (2022).
139. Han, H. G., Moon, H. W. & Jeon, Y. J. ISG15 in cancer: Beyond ubiquitin-like protein. *Cancer Lett.* **438**, 52–62. <https://doi.org/10.1016/j.canlet.2018.09.007> (2018).
140. Lee, K. *et al.* XAF1 drives apoptotic switch of endoplasmic reticulum stress response through destabilization of GRP78 and CHIP. *Cell Death Dis.* **13**, 655. <https://doi.org/10.1038/s41419-022-05112-0> (2022).
141. Tran, V. *et al.* Influenza virus repurposes the antiviral protein IFIT2 to promote translation of viral mRNAs. *Nat. Microbiol.* **5**, 1490–1503. <https://doi.org/10.1038/s41564-020-0778-x> (2020).
142. Gu, X. *et al.* Epigenetic regulation of OAS2 shows disease-specific DNA methylation profiles at individual CpG sites. *Sci. Rep.* **6**, 1–10. <https://doi.org/10.1038/srep32579> (2016).
143. Senbanjo, L. T. & Chelliah, M. A. CD44: A multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front. Cell Dev. Biol.* **5**, 18. <https://doi.org/10.3389/fcell.2017.00018> (2017).
144. Fang, L. *et al.* PLAU directs conversion of fibroblasts to inflammatory cancer-associated fibroblasts, promoting esophageal squamous cell carcinoma progression via uPAR/Akt/NF- κ B/IL8 pathway. *Cell Death Discov.* **7**, 32. <https://doi.org/10.1038/s41420-021-00410-6> (2021).
145. Nilsson, U. W., Abrahamsson, A. & Dabrosin, C. Angiogenesis regulation by estradiol in breast tissue: Tamoxifen inhibits angiogenesis nuclear translocation and antiangiogenesis therapy reduces breast cancer growth in vivo. *Clin. Cancer Res.* **16**, 3659–3669. <https://doi.org/10.1158/1078-0432.CCR-10-0501> (2010).
146. Chen, L., Xiao, D., Tang, F., Gao, H. & Li, X. CAPN6 in disease: An emerging therapeutic target (Review). *Int. J. Mol. Med.* **46**, 1644–1652. <https://doi.org/10.3892/ijmm.2020.4734> (2020).
147. Magalhães, J., Eira, J. & Liz, M. A. The role of transthyretin in cell biology: Impact on human pathophysiology. *Cell Mol. Life Sci.* **78**, 6105–6117. <https://doi.org/10.1007/s00018-021-03899-3> (2021).
148. Pittois, K., Wauters, J., Bossuyt, P., Deleersnijder, W. & Merregaert, J. Genomic organization and chromosomal localization of the Itm2a gene. *Mamm. Genome* **10**, 54–56. <https://doi.org/10.1007/s003359900942> (1999).
149. Aurilio, G. *et al.* Androgen receptor signaling pathway in prostate cancer: From genetics to clinical applications. *Cells* **9**, 2653. <https://doi.org/10.3390/cells9122653> (2020).
150. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, eaay5947. <https://doi.org/10.1126/science.aay5947> (2020).
151. Weichselbaum, R. R. *et al.* An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc. Natl. Acad. Sci. USA* **105**, 18490–18495. <https://doi.org/10.1073/pnas.0809242105> (2008).
152. Nushtaeva, A. A. *et al.* Characterization of primary normal and malignant breast cancer cell and their response to chemotherapy and immunostimulatory agents. *BMC Cancer* **18**, 1–11. <https://doi.org/10.1186/s12885-018-4635-8> (2018).
153. Yu, D. H. *et al.* Effects of hub genes on the clinicopathological and prognostic features of lung adenocarcinoma. *Oncol. Lett.* **19**, 1203–1214. <https://doi.org/10.3892/ol.2019.11193> (2020).
154. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507. <https://doi.org/10.1126/science.aan2507> (2017).
155. Charafe-Jauffret, E. *et al.* Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **25**, 2273–2284. <https://doi.org/10.1038/sj.onc.1209254> (2006).
156. Calderón-González, K. *et al.* Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry. *J. Proteom.* **124**, 50–78 (2015).
157. Stein, T. *et al.* Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer. *Clin. Cancer Res.* **11**(19), 6872–6879 (2005).
158. Muthamilselvan, S. & Palaniappan, A. Brcadx: Precise identification of breast cancer from expression data using a minimal set of features. *Front. Bioinform.* **3**, 1103493. <https://doi.org/10.3389/fbinf.2023.1103493> (2023).
159. Wu, J., Liu, X. J., Hu, J. N., Liao, X. H. & Lin, F. F. Transcriptomics and prognosis analysis to identify critical biomarkers in invasive breast carcinoma. *Technol. Cancer Res. Treat.* **19**, 1533033820957011. <https://doi.org/10.1177/1533033820957011> (2020).
160. Jiang, G. *et al.* Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. *BMC Genom.* **17**, 525. <https://doi.org/10.1186/s12864-016-2911-z> (2016).
161. Gambardella, G. *et al.* A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat. Commun.* **13**, 1714. <https://doi.org/10.1038/s41467-022-29358-6> (2022).

162. Liu, K. *et al.* Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nat. Commun.* **10**, 2138. <https://doi.org/10.1038/s41467-019-10148-6> (2019).

Acknowledgements

We acknowledge the contributions of Salvador Pérez Mora, the Cancer Cell line Encyclopedia (CCLE), Gene Expression Omnibus (GEO) database, DepMap Portal and The Cancer Genome Atlas (TCGA) for data sharing.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) Grant 773600, 320012, by the Instituto Politécnico Nacional Grants SIP20211325, SIP20220709 and Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México IA207423, IN-220523.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61908-4>.

Correspondence and requests for materials should be addressed to M.C.G.-G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024