# RESEARCH ARTICLE

# Identification of Potential Breast Cancer Stem Cell Biomarkers in the Secretome Using a Network Interaction Approach Analysis

## Ay Ly Margaret[1], Septelia Inawati Wanandi[2,3]* Fadilah Fadilah[4], Rafika Indah Paramita[4]

## Abstract

**Background:** Breast cancer stem cells (BCSCs) play a role in the high rates of resistance, recurrence, and metastasis. The precise biomarkers of BCSCs can assist effectively in identifying cancer, assessing prognosis, diagnosing, and monitoring therapy. The aim of this study was to give a complete analysis for predicting specific biomarkers of BCSCs. **Methods:** We aggregated profile datasets in this work to shed light on the underlying critical genes and pathways of BCSCs. We obtained two expression profiling by array datasets (GSE7513 and GSE7515) from the Gene Expression Omnibus (GEO) database to identify biomarkers in BCSCs. Enrichr was used to do functional analysis, including gene ontology (GO) and reactome pathway. Furthermore, the protein-protein interaction (PPI) of these differential expression genes (DEGs) was visualized using Cytoscape with the search tool for the retrieval of interacting genes (STRING). The hub genes in the PPI network were chosen for further investigation. **Results:** We identified 65 up-regulated and 190 down- regulated DEGs and the GO enrichment analysis revealed that these DEGs were enriched in biological process related to tumorigenesis and stemness, including alter the extracellular matrix's physicochemical properties, cytoskeletal reorganisation, adhesion, motility, migration, growth, and survival. The Reactome analysis indicated that these DEGs were also involved in modulating function of ECM, regulation cancer metabolism and angiogenesis, tumor growth, proliferation, and metastasis. **Conclusion:** Our bioinformatic study revealed that FYN, INADL, OCLN, F11R, and TOP2A were potential biomarker panel of BCSCs from secretome.

**Keywords:** Breast cancer stem cells- bioinformatics- biomarker- network interaction- secretome

## Introduction

Breast cancer is the second most common cancer in women and is the leading cause of death worldwide [1]. In recent years, several studies have demostrated that the population of breast cancer stem cells (BCSCs) play a pivotal role in influencing therapy resistance, recurrence, and metastasis of breast cancer [2]. BCSCs are a subpopulation of cancer cells that exhibit characteristics similar to normal stem cells but possess a high tumorigenicity [3]. Tumor heterogeneity and interpatient variability make it challenging to identify accurate and precise biomarkers for BCSCs [4]. The presence of BCSCs is associated with a poor prognosis, which serves as a strong motivator for researchers to conduct investigations into BCSC in order to eradicate them [5]. CD24-/CD44+ and aldehyde dehydrogenase (ALDH) positivity are among the most commonly used biomarkers for BCSC, although their specificity remains a subject of ongoing research [3].

The gold standard for identifying cancer biomarkers is tissue biopsy, despite its invasive nature. However, invasive techniques are gradually being supplanted by non-invasive methods, such as liquid biopsies [5]. Liquid biopsy offers several advantages, including being non-invasive or less invasive, cost-effectiveness, real-time monitoring of tumor states, and addressing the challenge of tumor heterogeneity [5]. Liquid biopsies contain both soluble and insoluble factors, with the secretome being of particular interest. The secretome derived from BCSCs is influenced by the tumor microenvironment (TME), which helps maintaining their phenotype [6]. Through

*[1]Doctoral Program in Biomedical Sciences, Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No. 6, Jakarta, 10430, Indonesia. [2]Molecular Biology and Proteomics Core Facilities, Indonesian Medical Education and Research Institute (IMERI), Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No. 6, Jakarta, 10430, Indonesia. [3]Department of Biochemistry and Molecular Biology, Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No. 6, Jakarta, 10430, Indonesia. [4]Bioinformatics Core Facilities, Indonesian Medical Education and Research Institute (IMERI), Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No. 6, Jakarta, 10430, Indonesia. *For Correspondence: septelia.inawati@ui.ac.id*

their secretome, BCSCs exert control over various tumor characteristics, including angiogenesis, tumor development and progressiveness, such as metastasis, therapy resistance, and immunological dysregulation [6]. The BCSC secretomes can be obtained from liquid biopsies (in vivo) or cell-conditioned medium (in vitro). Proteomic profiling of cell-conditioned media has emerged as an alternative strategy for identifying secreted tumor markers [7].

Nowadays, the availability of ''omic'' data that are high quality, reproducible, and comprehensive is crucial for accelerating the discovery of biomarkers in support of personalized medicine. Network biomarkers are increasingly recognized as being more effective in characterizing cancer than individual molecules [1]. In this study, we applied a network interaction approach to predict a panel of biomarkers specific to BCSCs from their secretome.

## Materials and Methods

The gene expression profiles of GSE7513 and GSE7515 were obtained from GEO database (https://www.ncbi.nlm.nih.gov/geo/browse/) [8]. Based on the platform GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), GSE7513 comprises of 14 CD24-/CD44+ and 15 non CD24-/CD44+, while GSE7515 comprises 15 cancer mammospheres and 11 non mammospheres primary breast cancer, respectively.

### Identification of differential expression genes (DEGs)

The GSE7513 and GSE7515 datasets were used to identify the DEGs between BCSCs and non BCSCs based on the interactive web tool GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE7513;https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE7515) [9]. With the define groups command in GEO2R, topics may be categorized into many categories. The panel GEO2R's settings enable simple modification of the analysis. The option panel enables us to select the statistical corrector, the data normalization method, and the cut-off value to filter out the genes not holding the defined cut-off. GEO2R exploits the limma package to perform inter- and intra-sample normalization. We used the log data transformation approach to standardize the findings and the false discovery rate (FDR) p value correction for multiple testing in order to conduct DEG analysis. Ultimately, we chose and downloaded the subsequent outcomes: p value, logFC, gene symbol, and adjusted p value. The differences were considered statistically significant when |log2fold change (FC)| was greater than 1.0 and p value was less than 0.05. Each DEG profile was downloaded and overlapped using Venn Diagram software (https://www.biotools.fr/misc/venny) [10]. Finally, the common DEGs among the two datasets were selected for further analysis.

### Functional enrichment analysis

Enrichr software was used for enrichment analysis with the following ontology sources: pathway (Reactome), gene ontology (GO) biological processes (BPs), molecular functions (MFs), and cellular components (CCs) [11]. A p

value < 0.05 were set as the critical standard for significant enrichment.

### Protein-protein interaction (PPI) network construction and visualization

A number of 258 possible proteins correlated to BCSCs were retrived from the protein query of STRING database (http://string-db.org/) (version 12.0) [12]. This database is one of the Cytoscape software 3.9.1 applications. The species was set as Homo sapiens. The required score was set as a medium confidence score cutoff (0.40). The PPI network was exported for further analysis with Cytoscape software 3.9.1 [13]. The interaction with a combined score between 0.90-0.99 were considered statistically significant. Degree (D), betweenness centrality (BC), closeness centrality (CC), stress (S), and average shortest path length (ASPL) distributions were provided for further screening of the hub nodes. The degree of centrality is the number of links to a particular node. Betweenness centrality measures the number of shortest paths that pass through a node in a network [14]. Closeness centrality sums the distance from one node to every other node. The stress of a node indicates the relevance of a protein as functionally capable of holding together communicating nodes [15]. The average shortest path length is the average number of steps along the shortest path for all possible pairs of network nodes [14].

## Results

### Identification of DEGs

A total of 1422 (681 up-regulated and 741 down-regulated genes) were identified in GSE7513 dataset, while 4142 (1294 up-regulated and 2848 down-regulated genes) were found in GSE7515 dataset (Figure 1A). The selection of DEGs was based on specific criteria, including a fold change [Log2FC] greater than 1.0 and p-value less than 0.05, across the different expression profile of GSE7513 and GSE7515 datasets. In the intersection of these two data sets, a total of 258 DEGs (65 upregulated and 193 downregulated genes) were identified and were visually represented in a Venn diagram (Figure 1B and 1C).

### Functional enrichment analysis

All terms in BP, MF, and CC were chosen for visualization and displayed in bar graphs (Figure 2, Figure 3, and Table S1). The more significant the terms, the longer and lighter the colour of the bar graph. The presence of grey bars indicates that the term is not statistically significant. The BP terms of 65 up-regulated proteins were predominantly related to glycosaminoglycan, sulfur compound, and chondroitin sulfate proteoglycan biosynthetic process, microvillus organization and assembly, negative regulation of cell differentiation, plasma membrane bounded cell projection morphogenesis (Figure 2A). The MF terms were mainly related to sulfotransferase activity, guanyl nucleotide exchange factor activity, and cAMP binding (Figure 2B). The CC terms were relevant to glial cell projection, neuron projection, golgi membrane, bounding membrane, apical dendrite, axon, and neuron to neuron synapse (Figure 2C).

Identification of potential breast cancer stem cell biomarkers in the secretome using a network interaction approach analysis
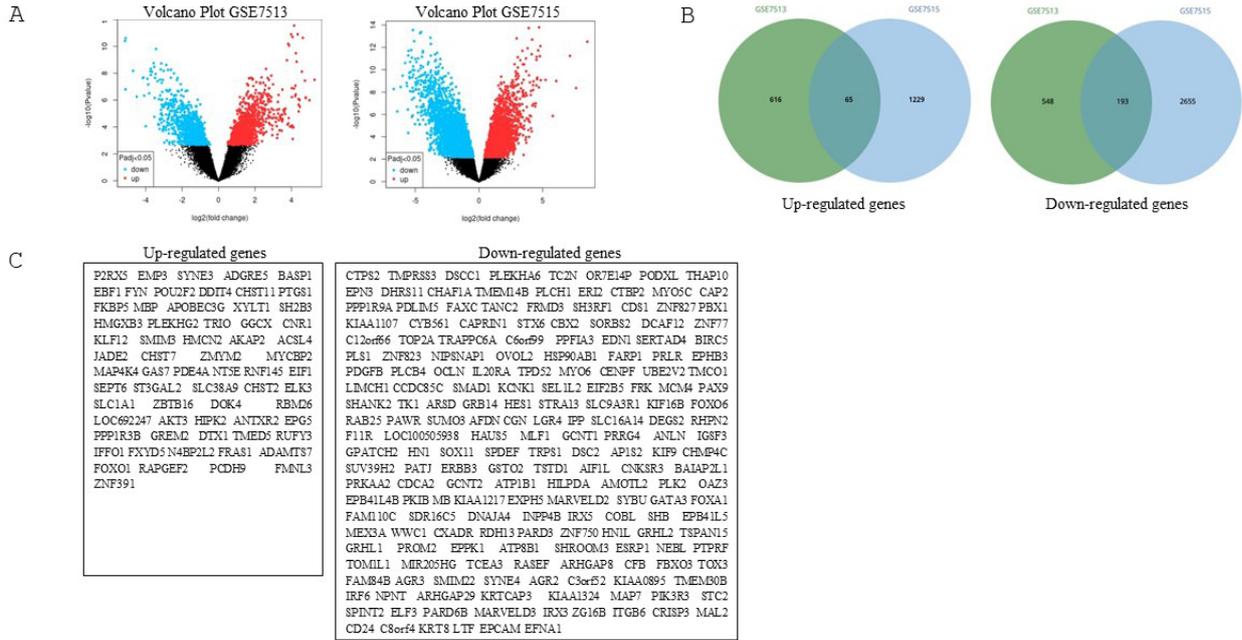


Figure 1. (A) Volcano plot identified the DEGs in two datasets. Red dots stand for up regulated genes and blue dots stand for down regulated genes. (B) In two datasets, a Venn diagram depicted the common up-regulated (left) and down-regulated DEGs (right). (C) List of 65 up-regulated (left) and 193 down-regulated (right) DEGs.

Reactome analysis revealed that the DEGs were mainly related to glycosaminoglycan metabolism, FLT3 signaling, AKT phosphorylates target in nucleus, chondroitin sulfate / dermatan sulfate metabolism, regulation of localization of FOXO (Forkhead box protein O1) transcription factor, DCC (deleted in colorectal cancer) mediated attrative signaling, and regulation of KIT (receptor tyrosine kinase) signaling (Figure 2D).

The BP terms of 193 down-regulated proteins were related to regulation of epithelial to mesenchymal transition, tight junction, cell junction, epithelial cell apical polarity, cellular component, and actin filament (Figure 3A). The MF terms were mainly related to cadherin binding, protein binding, cell to cell adhesion, and actin binding (Figure 3B). The CC terms pertinent to apical junction complex, tight junction, actin cytoskeleton, cell-cell junction, actin filament, intercalated disc, and adherens junction (Figure 3C). Reactome analysis revealed that the DEGs were mainly related to tight junction interaction, cell to cell junction, TGF beta receptor signaling in EMT,
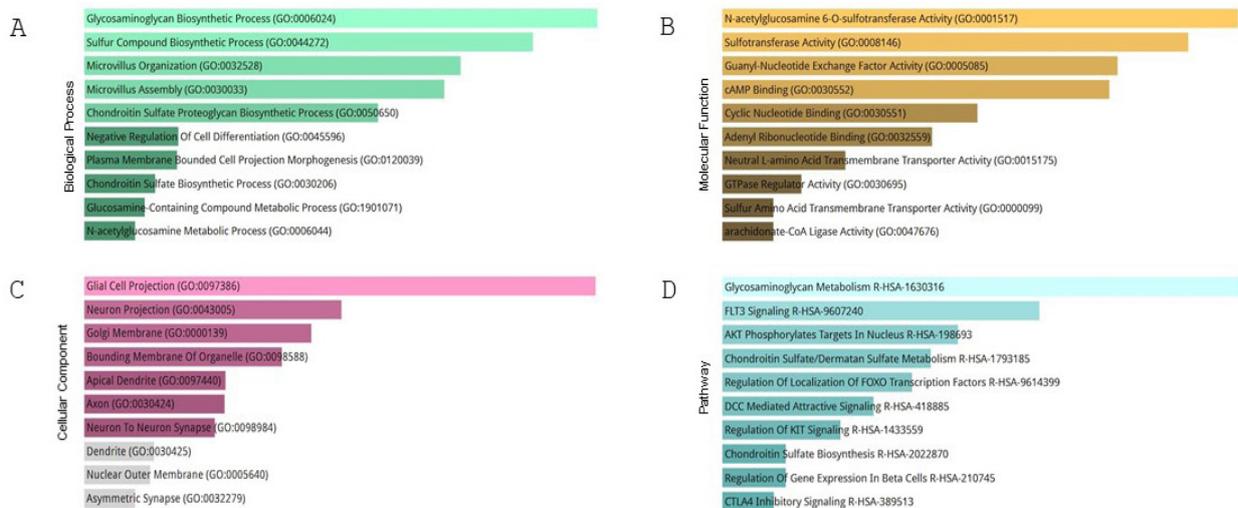


Figure 2. Enrichment Analysis of DEGs of 65 Upregulated Genes Common to All Two Datasets using the Enrichr ($p < 0.05$) and the results are top 10 terms of (A) Biological Processes (BP), (B) Molecular Function (MF), (C) Cellular Component (CC), and (D) Pathway. The longer and the lighter colored of Enrichr rank terms, the more the term is significant.
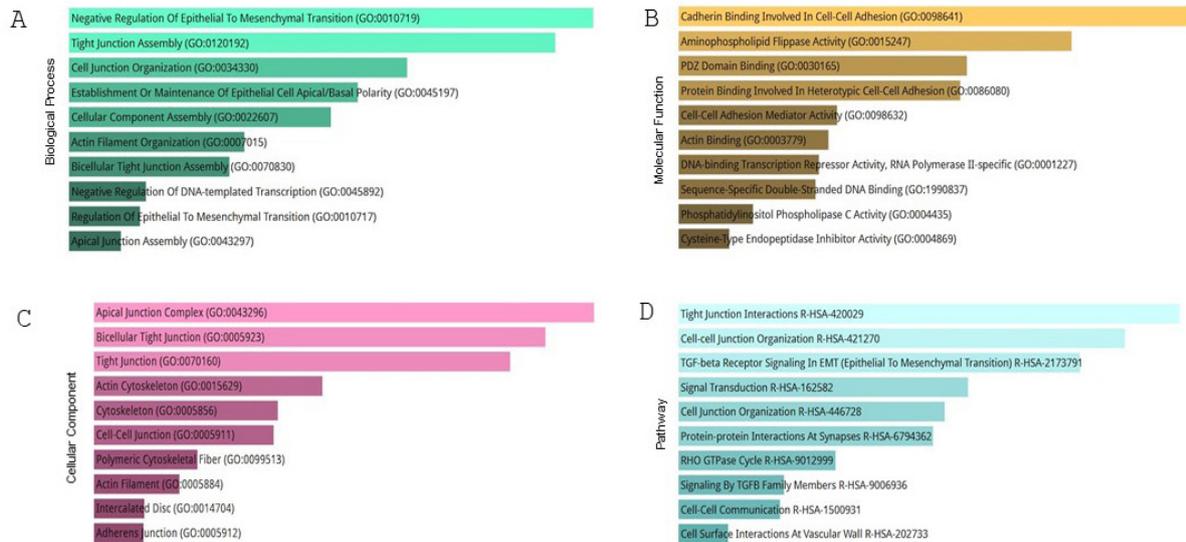
Figure 3. Enrichment Analysis of DEGs of 193 Downregulated Genes Common to All Two Datasets Using the Enrichr (p < 0.05) and the results are top 10 terms of (A) Biological Processes (BP), (B) Molecular Function (MF), (C) Cellular Component (CC), and (D) Pathway. The longer and the lighter colored of Enrichr rank terms (Enrichment score), the more the term is significant.

Table 1. Functional Protein Association Network Score of up-Regulated and down-Regulated

| Association | Protein Name | Stringdb score |
| --- | --- | --- |
| Up-regulated genes | AKT3 (pp) FOXO1 | 0.955 |
| | FYN (pp) MBP | 0.950 |
| Down-regulated genes | OCLN (pp) F11R | 0.986 |
| | CENPF (pp)TOP2A | 0.986 |
| | FOXA1 (pp) GATA3 | 0.985 |
| | INADL (pp) WWC1 | 0.982 |

signal transduction, cell-cell communication, and cell surface interaction (Figure 3D).

*PPI network analysis*

A PPI network of 65 up-regulated (Figure 4A) and 193 down-regulated (Figure 5A) DEGs were generated using STRING, and an interaction score >0.40 was considered a medium confidence interaction relationship. The nodes with the most interactions with neighboring nodes were considered as the key node [16]. The upregulated
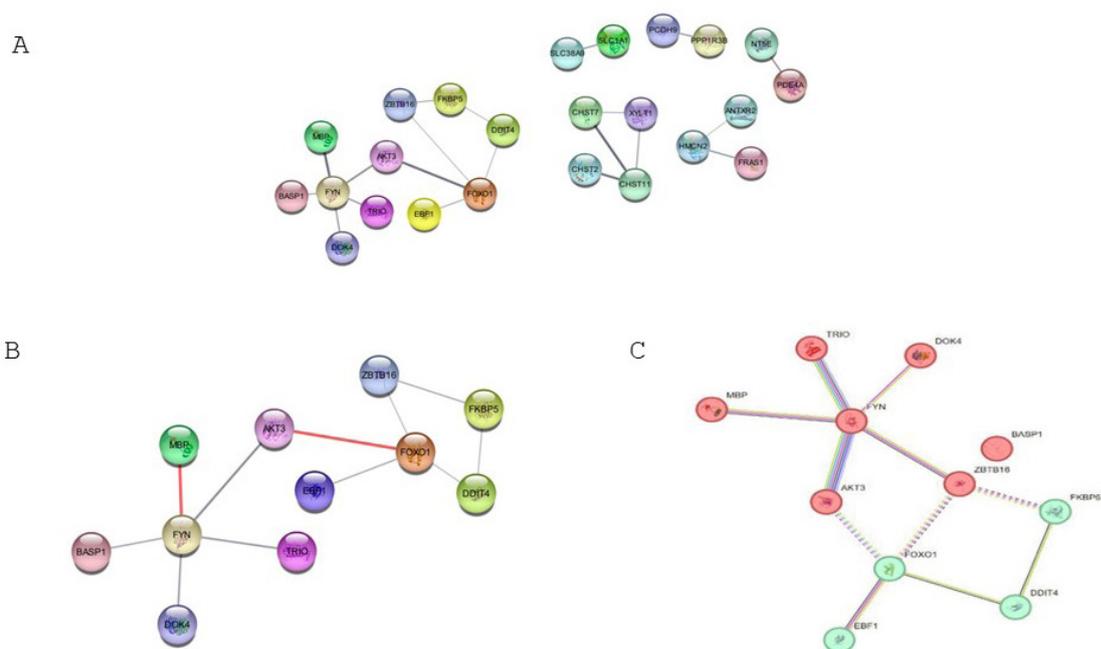


Figure 4. Protein-Protein Interaction (PPI) Network and Modular Analysis of DEGs. (A) The PPI network of 65 upregulated genes was structured by Cytoscape software. (B) The modules extracted from the PPI network comprises 11 nodes with 11 edges (stringdb score between 0.95-0.99) (red line). (C) Two cluster groups of 11 nodes using STRING online database.
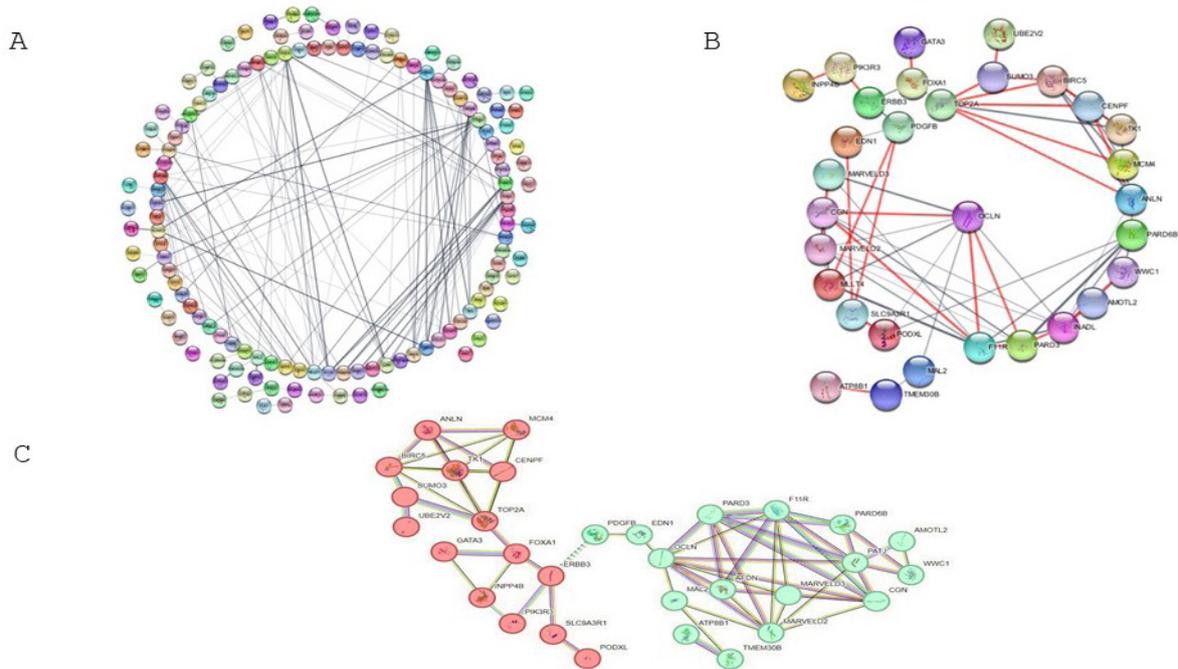
Figure 5. PPI Network and Modular Analysis of DEGs. (A) The PPI network of 193 down-regulated genes was structured by Cytoscape software. (B) The modules extracted from the PPI network comprises 31 nodes with 65 edges (Stringdb score between 0.95-0.99) (red line). (C) Two cluster groups of 31 nodes using STRING online database.

protein had 11 nodes and 11 edges (Figure 4B), while the downregulated protein had 31 nodes and 65 edges (Figure 5B). Protein-protein interactions were then filtered using a Stringdb score of 0.95-0.99 (Table S2 and Table S4). The functional protein association network score of up-regulated (AKT3 pp FOXO1, FYN pp MBP) and down-regulated (OCLN pp F11R, CENPF pp TOP2A, FOXA1 pp GATA3, INADL pp WWC1) modules is shown in Table

1 (Stringdb score 0.95-0.99).

In order to prioritize genes for experimental validation, the K-means clustering method in STRING (http://string-db.org/) (version 12.0) was used to conduct a cluster analysis on the predicted 11 up-regulated and 31 down-regulated proteins. Clustering is grouping proteins that have greater similarities in the same cluster. The analysis resulted in a network with high clustering
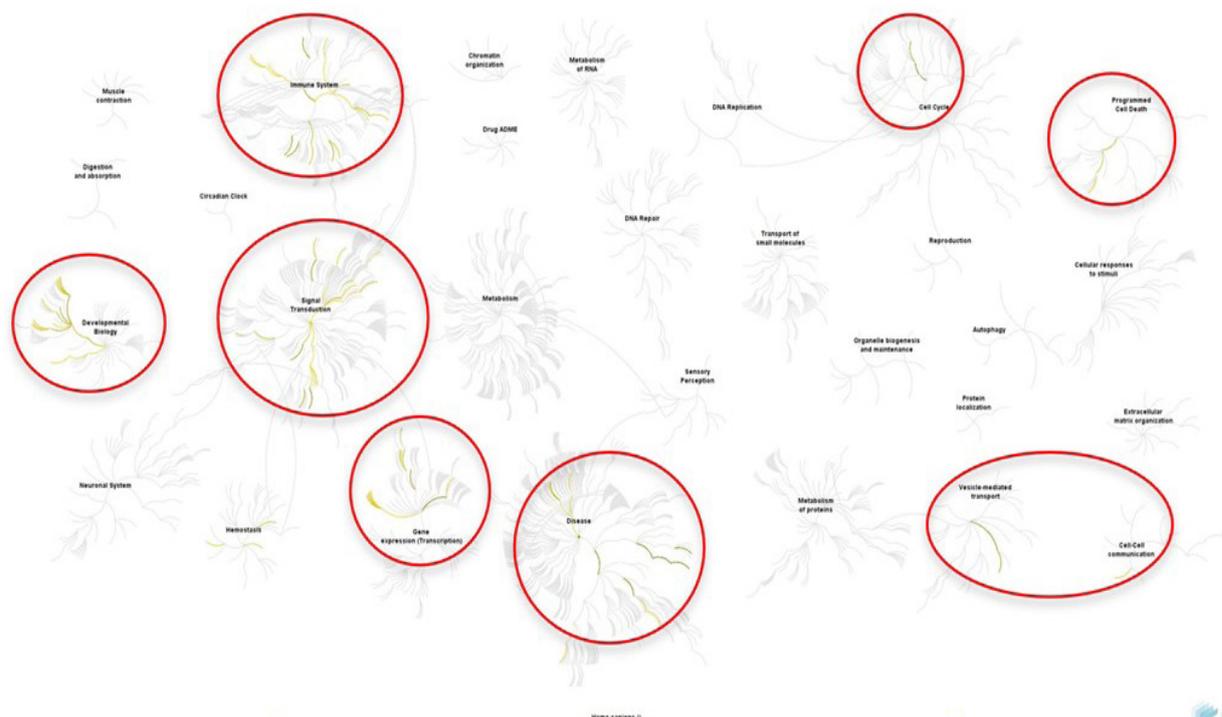


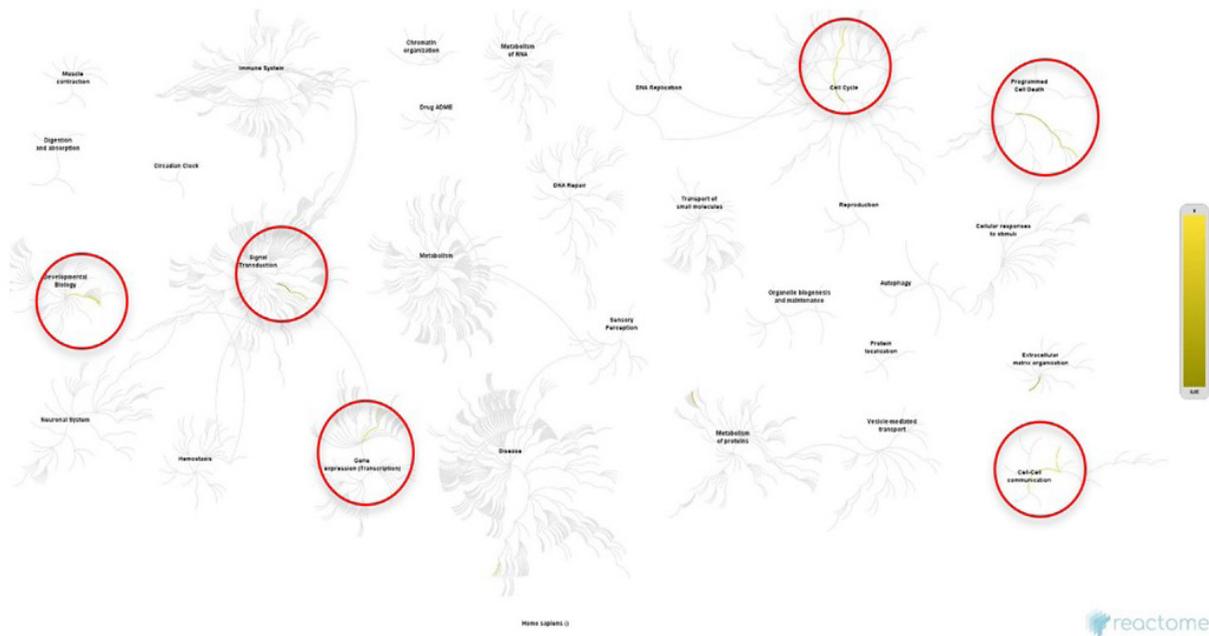Figure 6. Up-regulated Protein Pathways (17)

Figure 7. Down-regulated Protein Pathways (18)

coefficient of 0.714 and the average degree of 1.43. The identified targets were clustered into 2 groups.

The cluster 1 of up-regulated proteins has 7 members (AKT3, BASP1, DOK4, FYN, MBP, TRIO, ZBTB16) and was found to be enriched in genes controlling the organization of actin, PI3K-AKT and MAPK signaling, and transcriptional regulation of pluripotent stem cells (Figure 4C). The cluster 2 of up-regulated proteins has 4 members (EBF1, FOXO1, DDIT4, FKNP5) that were found to be related with transcription factor involved in white fat cell differentiation (Figure 4C).

The cluster 1 of down-regulated proteins has 15 members (ANLN, MCM4, BIRC5, TK1, CENPF, SUMO3, UBE2V2, TOP2A, GATA3, FOXA1, INPP4B, ERBB3, PIK3R3, SLC9A3R1, PODXL) and was found to be enriched in cell cycle, cell cycle checkpoint, cancer network, microvillus, DNA synthesis (Figure 5C). The cluster 2 of down-regulated proteins has 16 members (PDGFB, EDN1, OCLN, MAL2, ATP8B1, TMEM30B, PARD3, AFDN, MARVELD2, F11R, MARVELD3, PARD6B, PATJ, CGN, AMOTL2, WWC1) that were found to be related with tight junction interaction, TGF
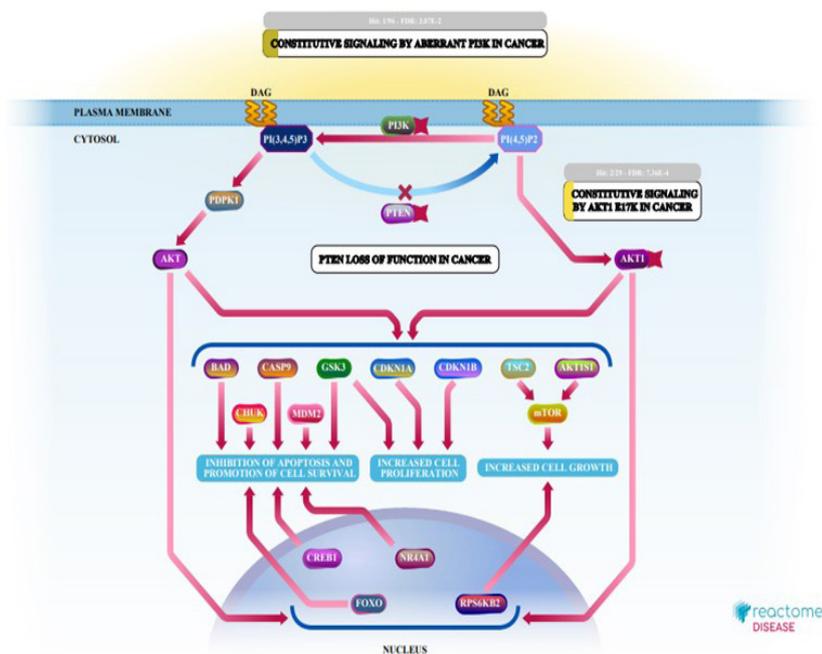


Figure 8. Role of up-Regulated Proteins in PI3K/AKT Signaling Cancer Pathway
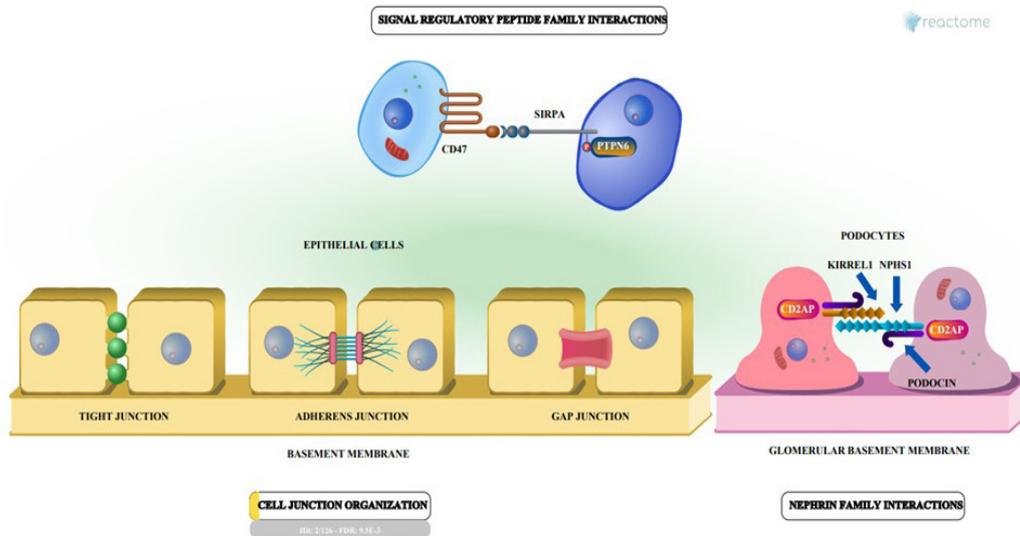
Figure 9. Role of Down-Regulated Proteins in Cell to Cell Communication

Table 2. Centrality Measurement of the Top Three up-Regulated and Eight down-Regulated Proteins were Introduced as Potential Spesific Biomarkers to BCSCs.

| | Protein | Description | D | BC | CC | S | ASPL |
|---|---|---|---|---|---|---|---|
| Up-regulated | FOXO1 | Forkhead box O1 | 5 | 0.67 | 0.5 | 68 | 2 |
| | FYN | Src family tyrosine kinases | 4 | 0.67 | 0.5 | 70 | 2 |
| | AKT3 | AKT Serine/Threonine Kinase 3 | 2 | 0.56 | 0.53 | 60 | 1.9 |
| Down-regulated | INADL/PATJ | InaD-like protein / Pals1-associated tight junction | 9 | 0.28 | 0.41 | 706 | 2.5 |
| | OCLN | Occludin | 9 | 0.23 | 0.4 | 394 | 2.5 |
| | F11R/JAM-A | F 11 receptor /Junctional adhesion molecule-A | 8 | 0.05 | 0.38 | 222 | 2.6 |
| | TOP2A | Transcription associated topoisomerase 2$\alpha$ | 7 | 0.21 | 0.37 | 402 | 2.7 |
| | FOXA1 | Forkhead box A1 | 3 | 0.16 | 0.32 | 220 | 3.2 |
| | GATA3 | GATA binding protein 3 | 1 | 0 | 0.24 | 0 | 4.1 |
| | CENPF | Centromere protein F | 5 | 0.01 | 0.33 | 56 | 3 |
| | WWC1 | WW domain containing C1 | 3 | 0 | 0.31 | 8 | 3.3 |

BC, betweenness centrality; CC, closeness centrality; D, degree; S, stress; ASPL, average shortest path length
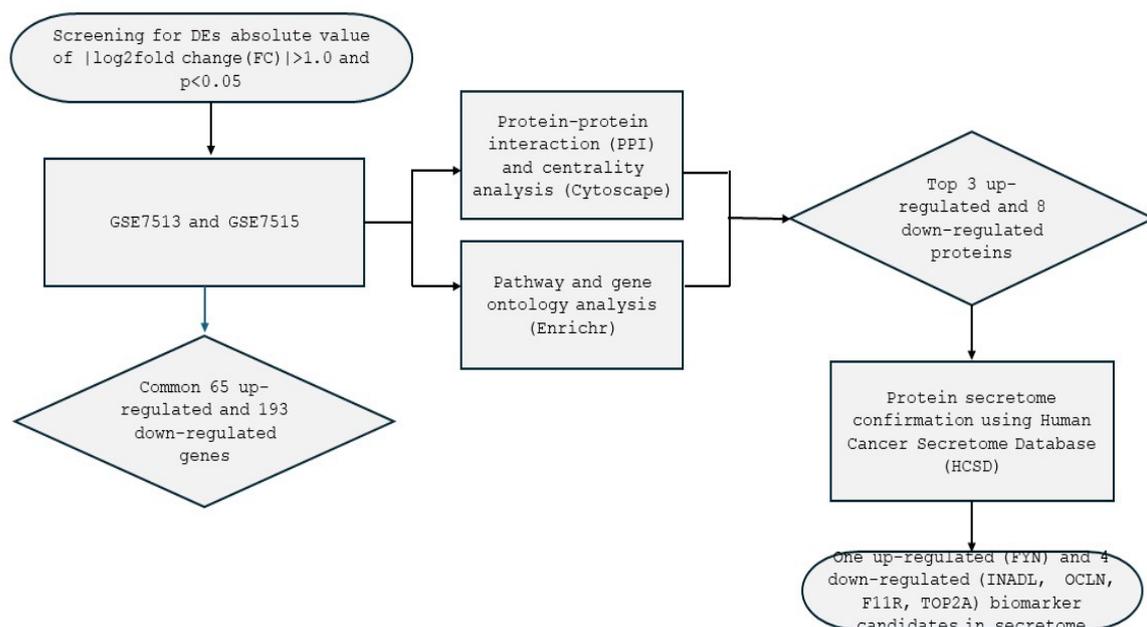


Figure 10. Schematic Summary of *in silico* Analysis

beta receptor signaling in EMT, signaling by Hippo, cell-cell junction organization (Figure 5C).

The nodes with a higher betweenness centrality, higher closeness centrality, higher node degree, higher stress, and lower average shortest path length were considered as hub proteins (FOXO1, FYN, AKT3, INADL, OCLN, F11R,TOP2A, dan FOXA1) (Table 2, Table S3, and Table S5) [13, 14]. They were likely to play a more vital role in maintaining the stability of the entire network.

Reactome pathways are arranged in a hierarchy. The center of each of the circular bursts is the root of one toplevel pathway, for example DNA Repair. Each step away from the center represents the next level lower in the pathway hierarchy. The color code denotes over-representation of that pathway in your input dataset. Light grey signifies pathways which are not significantly over-represented. FYN, FOXO1, and AKT3 have roles in immune system pathways, signal transduction, gene expression, disease, cell cycle, vesicle-mediated transport, developmental biology, programmed cell death, and cell-cell communication (Figure 6). INADL, OCLN, F11R, TOP2A, and FOXA1 have roles in signal transduction pathways, gene expression, the cell cycle, developmental biology, programmed cell death, and cell-cell communication (Figure 7).

## Discussion

Recent research has shown that BCSCs play a significant role in the progression of breast cancer. Consequently, there is a need for a reliable, non-invasive, and specific diagnostic biomarker for BCSCs. The exploration of gene expression profiles through bioinformatics can assist in the development of effective diagnostic and treatment strategies.

In this study, we identified 681 up-regulated genes in GSE 7513 and 1294 in GSE 7515 datasets, along with 741 down-regulated genes in GSE 7513 and 2848 in GSE 7515 datasets. The primary connected component was constructed, comprising 11 nodes with 11 edges for up-regulated genes and 31 nodes with 65 edges for down-regulated genes (Figure 2B). This construction nodes employed a medium confidence cut-off (0.40).

Typically, three types of confidence scores are used for PPIs: a) low confidence with a score < 0.4, b) medium confidence with a score between 0.4 and 0.7, and c) high confidence with a score > 0.7 [19]. In our analysis, we found high confidence values close to 1 (0.95-0.99) for protein interactions of the up-regulated genes, AKT3-FOXO1 and FYN-MBP (Table 1). Additionally, for down-regulated genes, OCLN-F11R, CENPF-TOP2A, FOXA1-GATA3, and INADL-WWC1 protein interaction also displayed high confidence values (Table 1). Hub genes were determined based on the highest degree scores of the PPI results, including AKT3, FOXO1, FYN for up-regulated, and INADL, OCLN, F11R, TOP2A, FOXA1, CENPF, GATA3, and WWC1 for down-regulated proteins (Table 1). However, based on the centrality analysis results [Table 2], CENPF, GATA3, and WWC1 were not included due to their BC values being close to 0, with relatively small D, CC, and S values, while ASPL had

relatively higher values compared to other candidate proteins. Hence, we describe three up-regulated proteins (FOXO1, FYN, AKT3) and five down-regulated proteins (INADL, OCLN, F11R, TOP2A, FOXA1) as potential biomarkers for BCSCs.

The bioinformatics analysis carried out in this study is similar to studies conducted by Tu H, et al. [20] and Liao Y, et al. [21], used bioinformatics to search for potentially useful biomarkers for the early detection of non-small cell lung cancer (NSCLC). This study used three microarray mRNA datasets (GSE19188, GSE33532, and GSE44077) from early NSCLC patient samples and non-cancer samples. GEO2R was used to identify differentially expressed genes (DEGs) with |log2FC|>1.5 and p value < 0.05 (20). The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to conduct functional and pathway enrichment studies for the DEGs. The analysis's results are derived from the expression of the upregulated gene in NSCLC related to mitotic nuclear division, cell division, and proliferation. In NSCLC, downregulated genes are related to angiogenesis, signal transduction, and cell adhesion. The STRING was used to plot the protein-protein interaction (PPI) network, and Cytoscape was used to visualize it [20]. A crucial module including 27 genes was then found, with a cut-off value of degree ≥10 [20].

Although there are some differences in the methods, Liao Y, et al. [21] also used bioinformatic search techniques to find biomarkers for cancer stem cell lung squamous cell carcinoma. The TCGA website provided RNA-seq data downloads. Using the Ensembl database, Ensembl IDs were translated to gene names, and log2 processing was carried out on the data. Multiple expression values for a given gene were averaged. DEGs between lung squamous cell carcinoma (LUSC) and normal tissues were found using the limma program. For DEGs, log2-fold change (FC) > 1.0 and adjusted p < 0.05 were the inclusion criteria [21]. The next step is to identify the significant modules using the module eigengene (ME), module significance (MS), and genetic significance (GS) calculations. The degree of connection between gene expression and the mRNAsi and EREG-mRNAsi was determined as GS. The average significance of every gene in the module was referred to as MS. ME was defined as the first principal component that resulted from each module's gene expression matrix's principal component analysis. The module that had the highest MS out of all of the modules was chosen for more study since it was thought to be connected to the mRNAsi and EREG-mRNAsi. The chosen modules were functionally enhanced for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) investigations through the use of functional enrichment analysis. Three terms are used in GO analysis such as molecular function (MF), cellular component (CC), and biological process (BP). The cutoff point was an adjusted P < 0.05 (21). Protein interactions and relationships using the STRING web database [21].

FOXO proteins govern a wide range of catabolic and anabolic processes across multiple cellular levels. Dysregulated modulation of FOXOs can contribute to tumor proliferation. FOXO1 is present in both the nucleus

and cytoplasm. This subfamily of FOXO proteins plays a vital role in stimulating antioxidant enzymes in response to cellular stress, safeguarding cells against oxidative damage [22, 23]. FOXOs are distributed throughout the human body but exhibit unique enrichment in specific tissue types. FOXO1 is implicated in diverse functions, which are pivotal in the context of cancer stem cells [23]. Despite overwhelming evidence that FOXO family genes act as tumor suppressors, a high level of FOXO1 expression has been linked to cancer metastasis and MMP overexpression, leading to the podocyte EMT in a high glucose condition [22]. Notably, the IGF-1/Akt signaling pathway increases FOXO1 phosphorylation (pFOXO1), which is associated with enhanced proliferation and decreased apoptosis of tumor cells [24].

FYN is a non-receptor tyrosine kinase belonging to the protein tyrosine kinase oncogene family. It is predominantly located in the cytoplasm [25]. FYN plays a pro-oncogenic role in cancer development [25]. It has been associated with cell motility and proliferation and has been observed to be overexpressed in the MDAMB231 cell line [23]. FYN mediates the Akt/PKB-induced anti-apoptotic effects of growth factors and exerts a significant influence on the proliferation of various cell types, including breast cancer cells [26]. Elevated FYN expression activates downstream signaling pathways, including PI3K/Akt, MAPK, and STAT, resulting in increased cell proliferation, migration, invasion, apoptosis inhibition, and EMT [26]. FYN has been found to be more abundant in highly invasive breast cancer cell lines such as MDAMB231 compared to those that are low-invasive such as MCF7 and normal mammary cell line such as MCF10A [22, 25]. The upregulation of FYN has been linked to the increased expression of Snail, a transcription factor associated with EMT, in breast cancer cells [18]. FOXO1 controls FYN transcription and mediates FGF2-induced EMT through the PI3K/AKT and ERK/MAPK pathways [22]. Furthermore, FYN is notably expressed in various drug-resistant cancer cells and is implicated in the development of resistance to cancer treatments [25].

AKT is a downstream effector of PI3K, which plays a critical role in cancer cell proliferation, resistance to treatment, and the maintenance of cancer stem cell populations [27]. Analysis by the The Cancer Genome Atlas Project (TCGA) revealed that AKT3 expression is upregulated in 28% of triple-negative breast cancers (TNBCs). AKT3 is significantly involved in mammosphere formation and the enhancement of ALDH+ or CD24-/CD44+ populations in MDAMB231 cells [19]. Depletion of AKT3 in MDAMB231 cells effectively suppresses spheroid formation, resulting in a remarkable 57% reduction in spheroid size [27]. In contrast, the deficiency of AKT1 and AKT2 has limited impact on spheroid grwoth. AKT3 is overexpressed in 21% of TNBCs and 2% of luminal breast cancers [27]. Through the YB1 (Y box binding protein 1)-Snail/Slug axis, AKT3 exerts its regulatory influence on stemness in TNBC [27]. In addition, AKT3 overexpression leads to the downregulation of p53, p21, and p27, and the upregulation of Cyclin D1, Bcl2, and XIAP [28]. In breast cancers, AKT3 expression is closely associated

with epithelial-mesenchymal transition (EMT) activators such as ZEB1 [29].

INADL plays a crucial role in regulating apico-basal polarity and directional migration in epithelial cells by modulating the positioning of atypical protein kinase C (aPKC) and PAR3 [30]. To date, there is limited data available that establishes a link between INADL protein or PATJ and breast cancer stemness. A study by Li P in 2020 reported that EMT markers such as SANIL and ZEB1 can transcriptionally decrease INADL expression in MDAMB231 cells [30]. Occludin (OCLN), an integral membrane protein found in tight junctions, plays a pivotal role in tumor development and metastasis. Tight junction proteins also contribute to the aquisition of stem cell phenotypes in cancer cells [31]. Knocking down OCLN in breast cancer cell lines results in reduced cell-cell adhesion, decreased sensitivity to apoptotic signals, induction of EMT, and pertubation of Ca2+ homeostasis, leading to enhanced invasiveness [32]. Another tight junction protein, F11 Receptor (F11R), also known as Junctional Adhesion Molecule-A (JAM-A) can serve as a cell surface marker for characterizing TNBC cancer stem cells [33]. It has been reported that F11R deficiency in the MDAMB231 cell line promotes breast cancer cell invasion [34]. The F11R expression can be inhibited by TGF-1, affecting both transcriptional and post-translational regulations of F11R, thereby promoting cell invasion [30].

TOP2A, a DNA topoisomerase, plays a role in various malignancies, including breast cancer. Downregulation of TOP2A is associated with cancer stem cell properties such as quiescence, resistance to apoptosis, and chemotherapy chemotherapy [35]. In breast cancer, TOP2A low-expressing cells are less sensitive to TOP2A-inhibiting drugs compared to TOP2A high-expressing cells [36]. Similar to TOP2A, under-expression of FOXA1 enhances malignancy and cancer stemness [37]. FOXA1 expression is found to be significantly lower in triple negative compared to non-triple negative tumors [37]. Low FOXA1 expression attenuates apoptosis by increasing SOD2 transcription and maintains cancer stem cell-like properties through enhanced IL6 transcription [35, 38].

Further investigation in this study utilizing the Human Cancer Secretome Database (HCSD) have identified FYN, INADL, OCLN, F11R, and TOP2A as secretome proteins across various human cancers, whereas AKT3, FOXO1, and FOXA1 were not found in this context [39]. Specifically, FYN was detected in the secretome of colorectal carcinoma, while INADL appeared in colorectal carcinoma, ovarian cancer, and glioblastoma secretomes [39]. OCLN was observed in prostate cancer and colorectal carcinoma secretomes, and F11R was found in colorectal carcinoma, liver cancer, oral cancer, bladder cancer, breast cancer, lung cancer, pancreatic cancer, epidermoid cancer, ovarian cancer and bladder cancer secretomes [39]. Additionally, TOP2A was detected in nasopharyngeal carcinoma, colorectal carcinoma, liver cancer, oral cancer, bladder cancer, ovarian cancer, lung cancer, pancreatic cancer, epidermoid cancer, lymph cancer secretomes [39].

Based on Reactome pathway tracing the three up-

regulated proteins have the same role. Their role such as PI3K/AKT pathway in cancer (Figure 8), PIP3 activates AKT signaling, cytokine signaling in the immune system, intracellular signaling by second messenger and diseases of signal transduction via growth factor receptors and second messengers. AKT (AKT1, AKT2, or AKT3) and PDPK1 (PDK1) are recruited to the plasma membrane by PIP3, which functions as a messenger. Due to PDPK1's poor affinity for PIP2, minute quantities of the protein are constantly present at the membrane. AKT's conformational shift upon binding to PIP3 allows the TORC2 complex to phosphorylate AKT at a conserved serine residue. AKT can attach to PDPK1 when the serine residue is phosphorylated, exposing a conserved threonine residue that is phosphorylated by PDPK1. After being phosphorylated at both serine and threonine residues, AKT separates from the plasma membrane and functions as a serine/threonine kinase, phosphorylating several nuclear and cytosolic substrates that are important in controlling gene expression, survival, and cellular metabolism. Signaling by PI3K/AKT is frequently constitutively activated in cancer. This activation can be via gain-of-function mutations in PI3KCA, PIK3R1, and AKT1. Tumor suppressor genes like PTEN that have loss-of-function mutations can also constitutively activate the PI3K/AKT pathway.

Based on Reactome pathway tracing the five down-regulated proteins have the same role. Intercellular communication is essential because it enables the coordination of cell activity (Figure 9). Certain types of intercellular communication necessitate direct interactions between cells, which are facilitated by cell surface receptors. Some of the cell surface receptors involved in cell-cell recognition, communication, and other facets of axon guidance and synapse formation are members of the immunoglobulin proteins. These functions are vital during development of the embryo.The creation and maintenance of adherens junctions, tight junctions, and gap junctions, as well as features of cellular contacts with extracellular matrix and hemidesmosome assembly, are mediated by processes identified here as aspects of cell junction organization. Hence, further research is essential to determine the potential of these identified secretome proteins as biomarker candidates in BCSCs.

In conclusion, the expression of FOXO1, FYN, and AKT3 exhibited up-regulation, whereas INADL, OCLN, F11R, TOP2A, and FOXA1 expressions showed down-regulation in BCSCs. FYN, INADL, OCLN, F11R, and TOP2A emerge as potential biomarker candidates for BCSCs due to their presence in the secretome (Figure 10).

## Author Contribution Statement

Septelia Inawati Wanandi contributed to study design, data collection, interpretation, manuscript preparation, supervision, and funding acquisition. Fadilah Fadilah participated in data collection, study design, statistical analysis, and supervision. Rafika Indah Paramita participated in data collection and data interpretation. Ay, Ly Margaret contributed to study design, data collection, statistical analysis, data interpretation, manuscript preparation, literature search, and funding acquisition.

## References

1. Afzal S, Hassan M, Ullah S, Abbas H, Tawakkal F, Khan MA. Breast cancer; discovery of novel diagnostic biomarkers, drug resistance, and therapeutic implications. Front Mol Biosci. 2022;9:783450. https://doi.org/10.3389/fmolb.2022.783450.
2. Wu ZH, Zhang YJ, Jia CL. Cancer stem cell characteristics by network analysis of transcriptome data stemness indices in breast carcinoma. J Oncol. 2020;2020:8841622. https://doi.org/10.1155/2020/8841622.
3. Yoshioka Y, Katsuda T, Ochiya T. Extracellular vesicles and encapusulated mirnas as emerging cancer biomarkers for novel liquid biopsy. Jpn J Clin Oncol. 2018;48(10):869-76. https://doi.org/10.1093/jjco/hyy120.
4. Crichton DJ, Alphan A, Amos CI, Anton K, Cinquini L, Colbert M, et al. Cancer biomarkers and big data: A planetary science approach. Cancer Cell. 2020;38(6):757-60. https://doi.org/10.1016/j.ccell.2020.09.006.
5. Lone SN, Nisar S, Masoodi T, Singh M, Rizwan A, Hashem S, et al. Liquid biopsy: A step closer to transform diagnosis, prognosis and future of cancer treatments. Mol Cancer. 2022;21(1):79. https://doi.org/10.1186/s12943-022-01543-7.
6. López de Andrés J, Griñán-Lisón C, Jiménez G, Marchal JA. Cancer stem cell secretome in the tumor microenvironment: A key point for an effective personalized cancer treatment. J Hematol Oncol. 2020;13(1):136. https://doi.org/10.1186/s13045-020-00966-3.
7. Méndez O, Villanueva J. Challenges and opportunities for cell line secretomes in cancer proteomics. Proteomics Clin Appl. 2015;9(3-4):348-57. https://doi.org/10.1002/prca.201400131.
8. NCBI. Gene Expression Omnibus [Internet]. USA: Bethesda; c2024 [updated 2024 Feb; cited 2023 Jan 25]. Available from: https://www.ncbi.nlm.nih.gov/geo/browse/
9. NCBI Geo2R. Geo2R [Internet]. USA: Bethesda; c2024 [updated 2024; cited 2023 Jan 25]. Available from:https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE7513; https://www.ncbi.nlm.nih.gov/geo/

geo2r/?acc=GSE7515.

10. BioTools.Fr. Venn Diagram [Internet]. Frances: Maarseille; c2024 [updated 2024 Feb; cited 2023 Jan 25]. Available from: https://bioinfogp.cnb.csic.es/tools/venny/

11. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. Curr Protoc. 2021;1(3):e90. https://doi.org/10.1002/cpz1.90.

12. String db. String Database [Internet]. Swiss: String Consortium; c2023 [updated 2023 July; cited 2023 August 25]. Available from: http://string-db.org/

13. Cytoscape. Cytoscape [internet]. Us: Cytoscape consortium; c2023 [cited 2023 august 25]. 2020. Available from: https://cytoscape.org/download.html.

14. Soofi A, Taghizadeh M, Tabatabaei SM, Rezaei Tavirani M, Shakib H, Namaki S, et al. Centrality analysis of protein-protein interaction networks and molecular docking prioritize potential drug-targets in type 1 diabetes. Iran J Pharm Res. 2020;19(4):121-34. https://doi.org/10.22037/ijpr.2020.113342.14242.

15. Centiserver. Stress Centrality [Internet]. US: Cytoscape Consortium; c2023 [cited 2023 August 25]. 2020. Available from: https://www.centiserver.org/centrality/Stress_Centrality/

16. Cai XY, Wang ZF, Ge SW, Xu G. Identification of hub genes and immune-related pathways for membranous nephropathy by bioinformatics analysis. Front Physiol. 2022;13:914382. https://doi.org/10.3389/fphys.2022.914382.

17. Reactome. Up-regulated protein pathway [Internet]. 2024. Available from: https://reactome.org/PathwayBrowser/#/R-HSA-2219528&DTAB=AN&ANALYSIS=MjAyNDAzMDgwNDM1MzNfMTI1MjE%253D%0A%0A%0A18. Reactome. Down-regulated protein pathway [Internet]. 2024. Available from: https://reactome.org/PathwayBrowser/#/R-HSA-9758919&DTAB=AN&ANALYSIS=MjAyNDAzMDgwNTExMjZfMTI1MjY%253D%0A

19. Tian Y, Li J, Cheung TC, Tam V, Lau CG, Wang X, et al. Akt3 promotes cancer stemness in triple-negative breast cancer through yb1-snail/slug signaling axis. Genes Dis. 2022;10(2):301-6. https://doi.org/10.1016/j.gendis.2022.08.006.

20. Tu H, Wu M, Huang W, Wang L. Screening of potential biomarkers and their predictive value in early stage non-small cell lung cancer: A bioinformatics analysis. Transl Lung Cancer Res. 2019;8(6):797-807. https://doi.org/10.21037/tlcr.2019.10.13.

21. Liao Y, Xiao H, Cheng M, Fan X. Bioinformatics analysis reveals biomarkers with cancer stem cell characteristics in lung squamous cell carcinoma. Front Genet. 2020;11:427. https://doi.org/10.3389/fgene.2020.00427.

22. Sadaf, Hazazi A, Alkhalil SS, Alsaiari AA, Gharib AF, Alhuthali HM, et al. Role of fork-head box genes in breast cancer: From drug resistance to therapeutic targets. Biomedicines. 2023;11(8):2159. https://doi.org/10.3390/biomedicines11082159.

23. Xie J, Zhang J, Tian W, Zou Y, Tang Y, Zheng S, et al. The pan-cancer multi-omics landscape of foxo family relevant to clinical outcome and drug resistance. Int J Mol Sci. 2022;23(24):15647. https://doi.org/10.3390/ijms232415647.

24. Jiang S, Li T, Yang Z, Hu W, Yang Y. Deciphering the roles of foxo1 in human neoplasms. Int J Cancer. 2018;143(7):1560-8. https://doi.org/10.1002/ijc.31338.

25. Peng S, Fu Y. Fyn: Emerging biological roles and potential therapeutic targets in cancer. J Transl Med. 2023;21(1):84. https://doi.org/10.1186/s12967-023-03930-0.

26. Elias D, Ditzel HJ. Fyn is an important molecule in cancer pathogenesis and drug resistance. Pharmacol Res. 2015;100:250-4. https://doi.org/10.1016/j.phrs.2015.08.010.

27. Chin YR, Yoshida T, Marusyk A, Beck AH, Polyak K, Toker A. Targeting akt3 signaling in triple-negative breast cancer. Cancer Res. 2014;74(3):964-73. https://doi.org/10.1158/0008-5472.Can-13-2175.

28. Hinz N, Jücker M. Distinct functions of akt isoforms in breast cancer: A comprehensive review. Cell Commun Signal. 2019;17(1):154. https://doi.org/10.1186/s12964-019-0450-3.

29. Stottrup C, Tsang T, Chin YR. Upregulation of akt3 confers resistance to the akt inhibitor mk2206 in breast cancer. Mol Cancer Ther. 2016;15(8):1964-74. https://doi.org/10.1158/1535-7163.Mct-15-0748.

30. Li P, Lan P, Liu S, Wang Y, Liu P. Cell polarity protein pals1-associated tight junction expression is a favorable prognostic marker in clear cell renal cell carcinoma. Front Genet. 2020;11:931. https://doi.org/10.3389/fgene.2020.00931.

31. Nehme Z, Roehlen N, Dhawan P, Baumert TF. Tight junction protein signaling and cancer biology. Cells. 2023;12(2). https://doi.org/10.3390/cells12020243.

32. Rachow S, Zorn-Kruppa M, Ohnemus U, Kirschner N, Vidal-y-Sy S, von den Driesch P, et al. Occludin is involved in adhesion, apoptosis, differentiation and ca2+-homeostasis of human keratinocytes: Implications for tumorigenesis. PLoS One. 2013;8(2):e55116. https://doi.org/10.1371/journal.pone.0055116.

33. Bednarek R, Wojkowska DW, Braun M, Watala C, Salifu MO, Swiatkowska M, et al. Triple negative breast cancer metastasis is hindered by a peptide antagonist of f11r/jam-a protein. Cancer Cell Int. 2023;23(1):160. https://doi.org/10.1186/s12935-023-03023-4.

34. Bednarek R, Selmi A, Wojkowska D, Karolczak K, Popielarski M, Stasiak M, et al. Functional inhibition of f11 receptor (f11r/junctional adhesion molecule-a/jam-a) activity by a f11r-derived peptide in breast cancer and its microenvironment. Breast Cancer Res Treat. 2020;179(2):325-35. https://doi.org/10.1007/s10549-019-05471-x.

35. Li X, Liu Y, Chen W, Fang Y, Xu H, Zhu HH, et al. Top2ahigh is the phenotype of recurrence and metastasis whereas top2aneg cells represent cancer stem cells in prostate cancer. Oncotarget. 2014;5(19):9498-513. https://doi.org/10.18632/oncotarget.2411.

36. Nandeesh BN, Naskar S, Shashtri AH, Arivazhagan A, Santosh V. Recurrent glioblastomas exhibit higher expression of biomarkers with stem-like properties. J Neurosci Rural Pract. 2018;9(1):86-91. https://doi.org/10.4103/jnrp.jnrp_417_17.

37. Dai X, Cheng H, Chen X, Li T, Zhang J, Jin G, et al. Foxa1 is prognostic of triple negative breast cancers by transcriptionally suppressing sod2 and il6. Int J Biol Sci. 2019;15(5):1030-41. https://doi.org/10.7150/ijbs.31009.

38. Yamaguchi N, Nakayama Y, Yamaguchi N. Down-regulation of forkhead box protein a1 (foxa1) leads to cancer stem cell-like properties in tamoxifen-resistant breast cancer cells through induction of interleukin-6. J Biol Chem. 2017;292(20):8136-48. https://doi.org/10.1074/jbc.M116.763276.

39. Human cancer secretome database. Human cancer secretome [internet]. 2023 [cited 2023 oct 5]. Available from: Http://176.58.113.186/search.