European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 11, 2020

Identification of Potential Biomarkers and Their Pathways for Breast Cancer using Integrated Bioinformatics Analysis

Swaminathan Venkataramanan[1], Sarmilah Mathavan[2]

[1,2] *Faculty of Health and Life Sciences, Management and Science University, Shah Alam, Selangor, Malaysia*

Abstract**: Breast cancer (BC) is a serious risk to women's health and became the nation's top cause of death in women. To date, the potential to treat advanced cancer is constrained by a lack of understanding of the precise hub genes. The objective of this study would be to classify hub genes and identify novel biomarkers for designing therapeutic drugs as well as establish prognostically important DEG within upregulated and downregulated expression. There is a total of 433 genes breast cancer which retrieved from GEO database. A protein-protein interaction (PPI) network was constructed for network analysis using Cytoscape. CytoHubba app in Cytoscape discovered the 6 hub genes. Moreover, GO and KEGG pathway analysis mainly involved cellular component, biological process, biological pathway and molecular function using DAVID database. MYLK, MYH11, PK5, ACTA2, TAGLN and CNN1 were recognized as the hub genes in the network based on MCC scoring method. A comprehensive perspective was provided by the bioinformatics analysis such as Cytoscape, GEO, DAVID to understand the mechanism underlying breast cancer development.

Keywords: breast cancer, hub genes, upregulated, downregulated, bioinformatics

1. INTRODUCTION:

Breast cancer (BC) has been a major threat to women's health and has become the leading cause of death among women worldwide (Zhang et al., 2016). BC has been recognized as the malignant breast cell tumor. It was identified as the most significant occurrence of cancer types, particularly in developed countries (Wu, Zhao, Zhou, & Qin, 2020). Usually cases of BC have often occurred in health care institutions with metastatic or advanced cancer due to poor prognosis and impoverished wealth to be diagnosed. BC can prompt to a very horrific and abandoned mortality if it is not treated earlier (Panieri, 2013).

Studies have proposed discovering multiple risk factors associated with breast cancer including obesity, lack of exercises, unhealthy lifestyles, high-fat food consumption (Zhang et al., 2016), consuming alcohol beverages or even radioactive materials (Fang & Zhang, 2017). In addition, early menarche, delayed menopause, contraceptive pills, hormonal therapy, delayed first full pregnancy (Schwab et al., 2015), continuous lactation and family medical history are also among the factors in breast cancer research findings (Adami, Signorello, & Trichopoulos, 1998).

These clinical symptoms may include a mass in the breast, pain and swelling, breast discharge, nipple retraction, and breast irritation and few breast skin changes such as redness, scaly or thickening (Barlow, 2002). Survival of BC closely associated with the stages of the cancer. Ultimately, breast cancer 's expected five - year survival is 90%, however if the tumor has remote metastases, the survival rate would decline to 26%.

Women commonly tested for BC with few clinical diagnoses, such as mammography and conventional tumor markers (Uehara et al., 2008). Diagnostic mammography is appropriate for the assessment of clinical findings through image analysis (Barlow, 2002). A really great potential to enhance life period if diagnosed and treated at an initial stage (Fang & Zhang, 2017). Common treatment of breast cancer including radiotherapy or chemotherapy, surgery and hormonal treatment (Y. Wang, Zhang, Huang, & Li, 2018).

In the detection and verification of genetic markers, cancer bioinformatics is anticipated to play a crucial role. The incorporation of information on protein annotations, interaction, and signaling pathways has analyzed network biomarkers as a new form of biomarkers through protein-protein interactions. Bioinformatics interplay with the field of genomics in cancer research. To gain a more comprehensive view of cancer and interpret intricate data, bioinformatics tools are critical to accelerating up a research process, including the detection of biomarkers(Wu, Rice, and Wang 2012).

To date, the capability to treat advanced cancer is constrained by a lack of comprehension about the precise hub genes with corresponding research underlying this cancer growth. Numerous gene alterations expressing hub genes have been shown to be associated with the development and progression of this cancer with innovations in bioinformatics analysis software. However, molecular mechanism of BC occurrence and prognosis of BC are not fully known (Fang & Zhang, 2017). Hence, in the current study, hub genes discovery influence breast tumour medical diagnosis and treatment.

Due to its highly predominant and metastasis characteristics, designing novel diagnostic biomarkers and drug product to improve prognosis is in such an urgent demand (Y. Wang et al., 2018). The objective of this study would be to classify hub genes and therapeutic drugs. Besides, to establish prognostically important DEG within upregulated and downregulated and perhaps to recognize biological process and pathways in breast cancer using bioinformatics analysis, and then to differentiate novel possible diagnostic biomarkers of breast cancer.

2. MATERIAL AND METHODOLOGY

Microarray data

The gene expression profile of GSE54002 based on the platform GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) was obtained from Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/), which was deposited by Tuan Zea Tan (Tan et al., 2014). Gene Expression Omnibus (GEO) serves a publicly available repository in storing microarray raw data and the next generation sequencing where profiles of expression of genes were extracted from (Liu et al., 2019; Shen et al., 2019).

Identification of differential expressed genes

In the GEO2R database, a number of genes were analyzed for the detection of differentially expressed between tumor samples and non-tumor samples (https:/www.ncbi.nlm.nih.gov / geo / geo2r/) with a threshold of positive and negative and a value of $P<0.05$ which was rendered statistical significance (Shen et al., 2019). Result showed in a form of table of 250 genes which has been classified into downregulated and upregulated. It serves as an online resource web resource designed to compare minimum of two specimens in a GEO array for experimental state recognition of DEGs (Liu et al., 2019).

GO term and KEGG pathway enrichment analysis

Analysis of differentially expressed genes (DEGs) at functional level performed by using Gene Ontology (GO) datasets (Ashburner et al., 2002) and Kyoto Encyclopaedia of Genes and Genome (KEGG) pathway enrichment (Shen et al., 2019) which available in Database

for Annotation, Visualization and Integrated Discovery (DAVID) (Huang, Sherman, & Lempicki, 2009). GO analysis used to identified features of biological alterations in given dataset from high-throughput genomes and transcriptomes (Ashburner et al., 2002). KEGG is an integrated database providing information regarding gene products, systems biology, drug and medications and chemical compounds (Du et al., 2014). DEGs were uploaded in DAVID database to execute biological process, molecular function and cellular component as well as pathway enrichment analysis (Dong et al., 2018). A cut off of $P<0.05$ was fixed.

PPI network construction

Differentially expressed genes from GSE54002 were uploaded in STRING protein query in Cytoscape 3.7.2 (Paul Shannon et al., 1971). Search Tool for the Retrieval of Interactive Genes (STRING) known as a precomputed global resource designed to critically evaluate protein-protein interaction (PPI) information (Szklarczyk et al., 2015). The constructed PPI network visualized in Cytoscape software (Paul Shannon et al., 1971). Confidence score (cut off) was set to 0.4 and maximum additional interactors remains default parameter (Liu et al., 2019) in order to obtain more closely related genes to the targeted protein.

Retrieval of hub genes

The plug-in CytoHubba of Cytoscape is a convenient application for extracting the subnetwork which contain hub genes from the entire large PPI collection (Chin et al., 2014). A number of ID nodes can be obtained from the existing network. Scoring method Maximum Clique Centrality (MCC) is one of the topological analysis methods which has been selected to discover featured nodes (Shen et al., 2019). Top 5 ranked nodes, scored by MCC in CytoHubba, were chosen. MCC's insight is that vital proteins appear to be grouped.

3. RESULTS

Table 1: 6 of which were downregulated genes and 245 which were upregulated genes in breast cancer tissues were identified out of top 250 genes. DEGs with $P>0.05$ and $|logFC|$ classified into two values including positive and negative obtained from GEO2R tool.

Expression	Gene symbol
Downregulated	MYH11, PAK5, TSHZ2, MYLK, TSHZ2
Upregulated	KCNJ16, SDPR, DST, PDLIM3, CARMN, ARHGAP20, TSHZ2, ADAMTS9-AS2, NPY2R, ANGPTL7, NRG1, LOC401433///NUB1, GAS1RR, TSHZ2, PDLIM3, CNN1, DKK3, OPRPN, LOC101927841, DSC1, SNCA, TSHZ2, COL17A1, HAS3, CLCA4, ACTA2, TCEAL7, TLE4, NR4A3, SCN3A, PTN, OXTR, ZNF638, FN1, SMYD1, PAK3, MYLK, PAMR1, TMEM255A, MYH11, KLHL29, TSLP, KCNE1, CCDC8, MME, MYH11, PTN, PDE1C, EFCAB1, CARMN///MIR145, CAPZA3, SPOCK3, NRG1, FN1, FN1, TCEAL2, LOC100128727///DGKI, FN1, TSHZ2, CARMN, IFRD1, BCLAF1, SCN2B, CNTNAP3P2, SLC7A3, PDE1C, PTN, SAMD5, CNTNAP3B///CNTNAP3, LMOD1, C2orf40, ADAMTS9-AS2, WIF1, TP63, KY, C4BPA, ADAMTS9-AS2, ITIH5, EMP1, GRIA4, GIPC2, EFCAB1, TLE4, IL33, DCAF1, CGB7///NTF4, ADAMTS5, CACUL1, CD200, TUSC3, MAGI2-AS3, PIGR, ABCB1, PER2, PAK3, DKK3, GRIA4, PAK3, CTHRC1, DST, NSG1, LOC202025, CCDC8, MGARP, NDRG2, ANXA1, SCN4B, TAGLN, CCL28, RCSD1, MT1M, KLHL29, CXCL2, RAPGEF2, MME, NUB1, HOXA4, TP63, GNAL, LOC105371352, RBMS3, LOC101929497, MOCS1,

	EPB41L4A- AS1, WDR86, KCNH8, CREB5, PDE1C, ADAMTS5, SAMD5, EGR2, PALMD, C8orf88, PPP2R2B, SLC27A6, P2RY12, RNF186, CLDN11, ADAMTS5, PPP1R12B, ADAMTS5, NGFR, SART1, HLF, MYH11, RPP30, PAK3, HS3ST4, SCARA5, OR5E1P, EMP1, ISM1, SLC25A27, ZBTB24, EIF1, RNF180, CCL28, UACA, PPP2CA, CHL1, PPP1R14A, FGF1, TPT1-AS1, PIK3C2G, ABCF2, P2RY12, GIPC2, CA3, TPT1-AS1, RUFY2, ITIH5, COL1A1, PLA2R1, DCT, ACTR3, AK5, PELI2, DGKI, NRG1, LOC101928307, MICU3, ABCB1, CCL28, TP53AIP1, NDRG2, TPT1-AS1 , PRDM5, GGTA1P, MAMDC2, EDN3, GPM6B, EOGT, DCAF1, SMIM10L2A

Table 2: Top 5 Gene Ontology (GO) terms in functional annotation for downregulated genes in DAVID with P<0.05

Term	P value	Genes	Fold enrichment
Aortic aneurysm	0.00160264	MYH11, MYLK	935.5
GO:0006939~smooth muscle contraction	0.00321256	MYH11, MYLK	466.4444444
nucleotide phosphate-binding region:ATP	0.00711424	PAK5, MYH11, MYLK	15.13807847
ATP-binding	0.01307844	PAK5, MYH11, MYLK	11.09687275
GO:0006936~muscle contraction	0.01899582	MYH11, MYLK	78.46728972

Table 3: KEGG pathway enrichment analysis for downregulated genes for selected top 5- fold enrichment

Genes	KEGG pathway
MYH11	Tight junction, Myosin S1 fragment, SH3 like, N-terminal, Loop containing nucleoside triphosphate hydrolase, Like IQ motif containing domain
MYLK	Calcium signalling pathway, cGMP PKG signalling pathway, vascular smooth muscle contraction, focal adhesion, platelet activation, Regulation of action cytoskeleton
PAK5	ErbB signalling pathway, Ras signalling pathway, Axon guidance, Focal adhesion, T-cell receptor signalling pathway, regulation of actin cytoskeleton

Table 4: Top 5 Gene Ontology (GO) terms in functional annotation for upregulated genes in DAVID with P<0.05

Term	P value	Genes	Fold enrichment
Secreted	1.14E-05	EDN3, CTHRC1, SPOCK3, MAMDC2, PAMR1, CXCL2, SNCA, IL33, CCL28, ISM1, ANGPTL7, COL17A1, PTN, ITIH5, NRG1, FGF1, PLA2R1, FN1, CLCA4, ANXA1, C4BPA, PIGR, OPRPN, C2ORF40, TSLP, DKK3, WIF1, COL1A1, ADAMTS5, CHL1	2.398578171
GO:005615 ~extracellul	0.001019711	CTHRC1, EDN3, SPOCK3, ACTA2, CXCL2, SNCA, ANXA1, C4BPA, PIGR, IL33,	2.219654788

ar space		OPRPN, CCL28, C2ORF40, DKK3, TSLP, PTN, COL1A1, FGF1, NRG1, ADAMTS5, FN1	
GO:0050699~WW domain binding	0.00126603	TP63, TCEAL7, TCEAL2, RAPGEF2	18.45926736
Extracellular matrix	0.00128063	CTHRC1,COL17A1,SPOCK3,MAMDC2,COL1A1,ADAMTS5, CHL1, FN1	4.87153086
Glycosylation site:N-linked (GlcNAc..)	0.001581897	CTHRC1, HS3ST4, MAMDC2, SCN3A, NPY2R, PAMR1, MME, OXTR, GPM6B, CCL28, ISM1, DCT, ANGPTL7, COL17A1, EOGT, KCNE1, ITIH5, NRG1, PLA2R1, FN1, CLCA4, SCN2B, ABCB1, C4BPA, GRIA4, PIGR, P2RY12, TSLP, DKK3, SLC7A3, SCN4B, KCNH8, WIF1, DSC1, COL1A1, HAS3, NGFR, DST, SCARA5, CD200, ADAMTS5, CHL1, EMP1	1.567364921

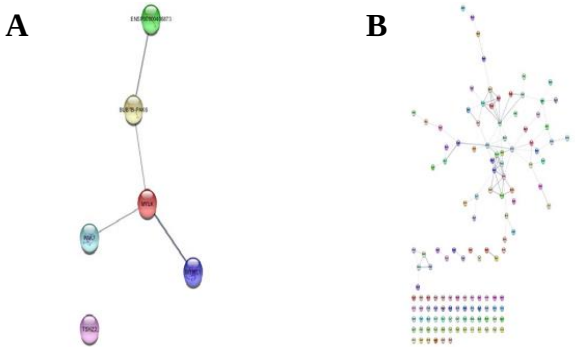

Figure 1: The PPI network construction of differentially expressed genes in breast cancer tissues built using STRING through the Cytoscape database. The network consists of 129 edges (interaction) between 104 nodes for upregulated (B) and 4 edges between 6 nodes for downregulated (A) based on a confidence score of 0.4 and the maximum additional interactors default parameter. Nodes represent protein, edges represent interaction between two nodes (proteins).

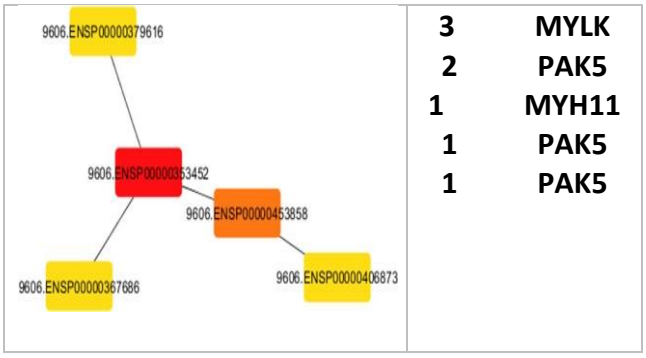

Figure 2: The PPI sub-network consisting of 5 highly essed genes downregulated based on MCC scoring method ranking was constructed. The 5 hub genes network are indicated by yellow and red nodes based on ranking score.

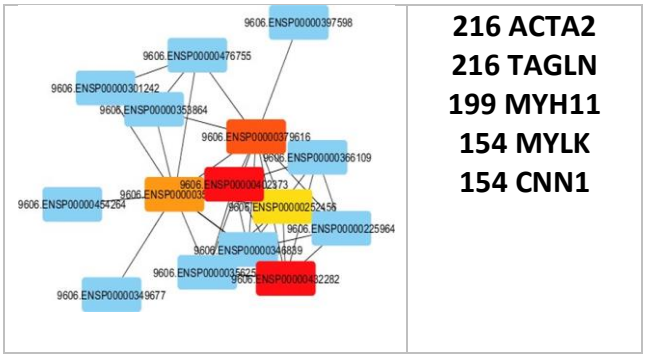

Figure 3: The PPI sub-network consisting of 5 highly expressed genes upregulated based on MCC scoring method ranking was constructed. The 5 hub genes network are indicated by yellow and red nodes based on ranking score.

Table 5: Top 5 hub genes for downregulated and upregulated genes of breast cancer obtained from MCC scoring method using CytoHubba plug-in

Downregulated hub genes	Upregulated hub genes
MYLK	ACTA2
PAK5	TAGLN
MYH11	**MYH11**
PAK5	**MYLK**
PAK5	CNN1

4. DISCUSSION

An ontology includes a set of definite terms with well-defined relationships. Processes also require a chemical or physical transition, in the sense that anything goes through a process then something else emerges out from that process. Molecular structure represents the potential a gene expression (or complex gene product) carries. Cell component indicates the position in which a target gene is involved in the cell (Ashburner et al., 2002). In the present work, downregulated genes involved in several GO terms including aortic aneurysm, smooth muscle contraction, nucleotide phosphate-binding region: ATP, ATP-binding and muscle contraction. MYH11 and MYLK involved in aortic aneurysm which resembles in enlargement of the aorta to a greater size than usual which later leads to rupture. Abnormal bulge can be observed on aorta wall. It also influences the development of malignancies of breast cancer (J. C. Wang et al., 2018).

Besides, smooth muscle contraction is one of the biological process involving the same genes to act on regulation of intracellular calcium levels in vascular smooth muscle is conducted mainly by plasmalemma including sarcoplasmic reticulum (SR) Ca2 + channels and ATP-dependent Ca2 + pumps in breast. Highest number of genes functioned in nucleotide phosphate binding region in breast tissue causes Her2 degradation and the growth arrest and differentiation of breast cancer cells (Chiosis et al., 2001). ATP binding represent molecular function GO terms whereby PAK5, MYH11, and MYLK genes involved to support transporters to be resistance to drugs (Leonessa & Clarke, 2003).

Moreover, MYH11 gene is mainly involving in a pathway of tight junction. Tight junctions be active specifically in paracellular sealing and also in distinction of membrane domains. Tight junctions are indeed an environment in which the plasma membrane of neighbouring

cells develops a sequence of connections which tend to completely obscure the extracellular space. It establishes an intercellular shield and a diffusion barrier within the intramembrane. Consequently, tight junctions are also the first barrier cancerous cells need to defeat for metastasis. Failure of tight joining plaque molecules in tissues of breast cancer is associated with poor prognosis in breast cancer patients (Martin, Watkins, Mansel, & Jiang, 2004).

MYLK downregulated gene primarily involved in calcium signalling pathway whereas it able to regulate cell proliferation and cause cell death mainly through Ca2+. Alterations in all these processes may lead to cell cycle arrest and apoptosis. Cancer cells can be selectively killed and/or arrested by targeting Ca2+ channels (Humeau et al., 2018). Expression of the ErbB family of tyrosine kinase receptors mostly through genitive Epidermal Growth Factor (EGF)- like peptide ligands is a large family of similar signalling pathways which regulate breast cancer proliferation, survival, cell growth and proliferation triggered by PAK5 gene. Antibodies blocking receptor ErbB and specific tyrosine kinase inhibitors have been proven. This pathway can discretely stage of mammary gland development (Hardy, Booth, Hendrix, Salomon, & Strizzi, 2010).

In the present study, upregulated genes involved in several GO terms including secreted, extracellular space, WW domain binding, extracellular matrix, and glycosylation site: N-linked. A large proportion of human breast cancers is characterized by its ability to exhibit metastasis and to be regulated by oestrogens. The proteins that are secreted, such as growth factors and proteases (Harley et al., 1988). Genes in extracellular space secreted by stromal fibroblasts in primary breast tumors increases the proliferation of tumor cells, as well as angiogenesis (Luker, Steele, Mihalko, Ray, & Luker, 2010). WW domain containing proteins, the WW motifs should play a role in protein-protein interactions commonly affected by abnormalities in breast and other cancers (Bednarek et al., 2000). The extracellular matrix (ECM) in breast cancer is being increasingly recognized as an essential regulator. ECM proteins play a significant functional role in the development and metastasis of breast cancer (Insua-Rodríguez & Oskarsson, 2016). Glycosylation site proteins from breast cancer lines. Analysing it from breast cancer lines could help generating a set of potential biomarkers (Whelan et al., 2009).

Protein associations can be concluded from the corresponding gene-coded interactions. The STRING database is often used to assess and pre-calculate global view protein associations comprising 89 full genome sequencing datasets including 261033 ontological genes. The visual representation of the predicted, ranked protein interactions network in STRING database offers a high-level overview of the functional association, enabling extensibility analysis in biological systems (Szklarczyk et al., 2015).

Nodes and edges are most important because they able to associated with gene expression data and protein structure information. To explore the hub genes in breast cancer network, we applied CytoHubba app, one of Cytoscape plugin to analyze the GSE54002 dataset from constructed STRING network. First, scores across all 11 methods are given to each node throughout a pre-loaded protein- protein interaction network by selecting CytoHubba options in the network panel. Maximum Clique Centrality (MCC) scoring method which belongs to local-based method was chosen for this study.

A local rank approach will only recognize the relationship between both the node and its significant neighbors in order to measure the node scoring within a network. MCC used to explore featured nodes in order to boost efficacy. The nodes and interactions of the steppingstone comprise the shortest path (Chin et al., 2014).

The hub genes involved in development of breast cancer including MYLK [Myosin light chain kinase], MYH11 [Myosin heavy chain 11], PAK5 [p21 (RAC1) activated kinase 5], ACTA2 [Actin, alpha 2, smooth muscle, aorta], TAGLN [transgelin], and CNN1 [calponin 1]. MYLKP1 promoter is triggered preferentially in neoplastic processes mirrored in silent

MYLKP1 promoter activities in healthy cells but also in significantly active cancer cell promoter activity. Novel function of MYLKP1 pseudogene expressed specifically in cancerous conditions, stimulating cell proliferation (Han, Ma, Yourek, Park, & Garcia, 2011). MYH11 (also known as SMMHC) encodes the heavy chain of smooth muscle myosin which plays a key role in the smooth contraction of muscles. Myosins have recently been involved in a variety of other intracellular functions which include cellular functions, migration, adhesion, cellular form regulation, and membrane traffic. Myosins has a very well-characterized biological function in use of the ATP hydrolysis energy to migrate actin filaments and generate muscle strength, but are also involved in a wide range of other cellular functions, some of which are pertinent for cancer development (Alhopuro et al., 2008).
P21-activated kinases (PAKs) activate various extracellular stimuli and activate other kinases by phosphorylating them at specific serine / threonine residues or by interacting with protein-proteins. Suppression of PAK5 in the lines of human breast cancer cells inhibited cell growth due to arrest in step G1. Knockdown of PAK5 could significantly inhibit the proliferation of breast cancer cells (X. X. Wang et al., 2013). Transgelin (TAGLN, SM22a) is an actin-binding pro tein that expresses itself abundantly in smooth muscle cells. It has transformed fibroblasts, and several lines and tissues of cancer cells. It is clear in expression and recurrent hypermethylation in the cell lines of the breast carcinoma (Sayar et al., 2015).
The most potent inducer of fibroblast transformation into 'activated' fibroblasts with elevated levels of smooth actin-alpha (SMAa) muscle (encoded by ACTA2) is transforming growth factor-beta (TGFß) (Busch et al., 2017). CNN1 was identified as the key gene in breast cancer by the bioinformatics analysis. Upregulating CNN1 inhibited cell survival, migration, invasion, and adhesion, but enhanced cell apoptosis. CNN1 can inhibit actin-activated myosin ATPase and Ca2+ dependent mobility of actin. Abilities of cell proliferation and invasion in BRCA cells were impaired via upregulating the CNN1 expression (Z. Wang, Li, Chen, Pan, & Shen, 2020).

5. CONCLUSION

In conclusion, methods of system biology, especially PPI networks beneficial in studying complex related diseases. A comprehensive perspective was provided by the bioinformatics analysis to understand the mechanism underlying breast cancer development. In this study, the hub genes involved including MYLK, MYH11, PAK5, ACTA2, TAGLN and CNN1 in breast cancer through network analysis. A total of 433 gene disease association and 6 hub genes have been identified and recognized as target biomarkers for breast cancer. A detailed study of the biological mechanism of this gene and its pathways may provide a potential target for the therapeutic drug monitoring of breast cancer.

6. REFERENCES

[1] Adami, H. O., Signorello, L. B., & Trichopoulos, D. (1998). Towards an understanding of breast cancer etiology. *Seminars in Cancer Biology*, *8*(4), 255–262. https://doi.org/10.1006/scbi.1998.0077

[2] Alhopuro, P., Karhu, A., Winqvist, R., Waltering, K., Visakorpi, T., & Aaltonen, L. A. (2008). Somatic mutation analysis of MYH11 in breast and prostate cancer. *BMC Cancer*, *8*, 1– 5. https://doi.org/10.1186/1471-2407-8-263

[3] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2002). Gene Ontology: tool for the unification of biology. *Nature America*, *25*(1), 25–30. https://doi.org/https://doi.org/10.1038/75556

[4] Barlow, W. E. (2002). Performance of Diagnostic Mammography for Women With Signs or Symptoms of Breast Cancer. *CancerSpectrum Knowledge Environment*, *94*(15), 1151– 1159. https://doi.org/10.1093/jnci/94.15.1151

[5] Bednarek, A. K., Laflin, K. J., Daniel, R. L., Liao, Q., Hawkins, K. A., & Aldaz, C. M. (2000). WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Research*, *60*(8), 2140–2145.

[6] Busch, S., Andersson, D., Bom, E., Walsh, C., Ståhlberg, A., & Landberg, G. (2017). Cellular organization and molecular differentiation model of breast cancer-associated fibroblasts. *Molecular Cancer*, *16*(1), 1–12. https://doi.org/10.1186/s12943-017-0642-7

[7] Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(4), 1–7. https://doi.org/10.1186/1752-0509-8-S4-S11

[8] Chiosis, G., Timaul, M. N., Lucas, B., Munster, P. N., Zheng, F. F., Sepp-Lorenzino, L., & Rosen, N. (2001). A small molecule designed to bind to the adenine nucleotide pocket of Hsp90 causes Her2 degradation and the growth arrest and differentiation of breast cancer cells. *Chemistry and Biology*, *8*(3), 289–299. https://doi.org/10.1016/S1074- 5521(01)00015-1

[9] Dong, H., Zhang, S., Wei, Y., Liu, C., Wang, N., Zhang, P., … Huang, J. (2018). Bioinformatic analysis of differential expression and core GENEs in breast cancer. *Int J Clin Exp Pathol*, *11*(3), 1146–1156. Retrieved from www.ijcep.com/

[10] Du, J., Yuan, Z., Ma, Z., Song, J., Xie, X., & Chen, Y. (2014). KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Molecular BioSystems*, *10*(9), 2441–2447. https://doi.org/10.1039/c4mb00287c

[11] Fang, E., & Zhang, X. (2017). Identification of breast cancer hub genes and analysis of prognostic values using integrated bioinformatics analysis. *Cancer Biomarkers*, *21*(1), 169–177. https://doi.org/10.3233/CBM-170550

[12] Han, Y. J., Ma, S. F., Yourek, G., Park, Y., & Garcia, J. G. N. (2011). A transcribed pseudogene of MYLK promotes cell proliferation . *The FASEB Journal*, *25*(7), 2305–2312. https://doi.org/10.1096/fj.10-177808

[13] Hardy, K. M., Booth, B. W., Hendrix, M. J. C., Salomon, D. S., & Strizzi, L. (2010). ErbB/EGF signaling and emt in mammary development and breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, *15*(2), 191–199. https://doi.org/10.1007/s10911-010-9172- 2

[14] Harley, C. B., Lawrie, J., Betlach, M., Crea, R., Boyer, H. W., & Hedgpeth, J. (1988). *Nucleic Acids Research*. *16*(15), 7269–7285.

[15] Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

[16] Humeau, J., Bravo-San Pedro, J. M., Vitale, I., Nuñez, L., Villalobos, C., Kroemer, G., & Senovilla, L. (2018). Calcium signaling and cell cycle: Progression or death. *Cell Calcium*, *70*, 3–15. https://doi.org/10.1016/j.ceca.2017.07.006

[17] Insua-Rodríguez, J., & Oskarsson, T. (2016). The extracellular matrix in breast cancer.

[18] *Advanced Drug Delivery Reviews*, *97*, 41–55. https://doi.org/10.1016/j.addr.2015.12.017

[19] Leonessa, F., & Clarke, R. (2003). ATP binding cassette transporters and drug resistance in breast cancer. *Endocrine-Related Cancer*, *10*(1), 43–73. https://doi.org/10.1677/erc.0.0100043

[20] Liu, F., Wu, Y., Mi, Y., Gu, L., Sang, M., & Geng, C. (2019). Identification of core genes and potential molecular mechanisms in breast cancer using bioinformatics

analysis. *Pathology Research and Practice*,*215*(7), 152436. https://doi.org/10.1016/j.prp.2019.152436

[21] Luker, K. E., Steele, J. M., Mihalko, L. A., Ray, P., & Luker, G. D. (2010). Constitutive and chemokine-dependent internalization and recycling of CXCR7 in breast cancer cells to degrade chemokine ligands. *Oncogene*, *29*(32), 4599–4610. https://doi.org/10.1038/onc.2010.212

[22] Martin, T. A., Watkins, G., Mansel, R. E., & Jiang, W. G. (2004). Loss of tight junction plaque molecules in breast cancer tissues is associated with a poor prognosis in patients with breast cancer. *European Journal of Cancer*, *40*(18), 2717–2725. https://doi.org/10.1016/j.ejca.2004.08.008

[23] Panieri, E. (2013). Breast-cancer awareness in low-income countries. *The Lancet Oncology*, *14*(4), 274–275. https://doi.org/10.1016/S1470-2045(13)70020-2

[24] Paul Shannon, 1, Andrew Markiel, 1, Owen Ozier, 2 Nitin S. Baliga, 1 Jonathan T. Wang, 2 Daniel Ramage, 2, Nada Amin, 2, Benno Schwikowski, 1, 5 and Trey Ideker2, 3, 4, 5, 山本隆久, … 大森敏行. (1971). Cytoscape: A Software Environment for Integrated Models. *Genome Research*, *13*(22), 426. https://doi.org/10.1101/gr.1239303.metabolite

[25] Sayar, N., Karahan, G., Konu, O., Bozkurt, B., Bozdogan, O., & Yulug, I. G. (2015). Transgelin gene is frequently downregulated by promoter DNA hypermethylation in breast cancer. *Clinical Epigenetics*, *7*(1), 1–16. https://doi.org/10.1186/s13148-015-0138-5

[26] Schwab, F. D., Kilic, N., Huang, D. J., Schmid, S. M., Vetter, M., Schötzau, A., & Güth, U. (2015). Personal or first-degree family breast cancer history: which has higher impact on tumor detection and tumor size in breast cancer. *Archives of Gynecology and Obstetrics*, *291*(6), 1387–1394. https://doi.org/10.1007/s00404-014-3592-y

[27] Shen, J., Yu, S., Sun, X., Yin, M., Fei, J., & Zhou, J. (2019). Identification of key biomarkers associated with development and prognosis in patients with ovarian carcinoma: Evidence from bioinformatic analysis. *Journal of Ovarian Research*, *12*(1), 1–13. https://doi.org/10.1186/s13048-019-0578-1

[28] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., … Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

[29] Tan, T. Z., Miow, Q. H., Miki, Y., Noda, T., Mori, S., Huang, R. Y., & Thiery, J. P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Molecular Medicine*, *6*(10), 1279–1293. https://doi.org/10.15252/emmm.201404208

[30] Uehara, M., Kinoshita, T., Hojo, T., Akashi-Tanaka, S., Iwamoto, E., & Fukutomi, T. (2008). Long-term prognostic study of carcinoembryonic antigen (CEA) and carbohydrate antigen 15-3 (CA 15-3) in breast cancer. *International Journal of Clinical Oncology*, *13*(5), 447–451. https://doi.org/10.1007/s10147-008-0773-3

[31] Wang, J. C., Chien, W. C., Chung, C. H., Liao, W. I., Tsao, C. H., Wu, Y. F., & Tsai, S. H. (2018). Increased risk of malignancy in patients with an aortic aneurysm: A nationwide population-based retrospective study. *Oncotarget*, *9*(2), 2829–2837. https://doi.org/10.18632/oncotarget.20181

[32] Wang, X. X., Cheng, Q., Zhang, S. N., Qian, H. Y., Wu, J. X., Tian, H., … Zheng, J. N. (2013). PAK5-Egr1-MMP2 signaling controls the migration and invasion in breast cancer cell. *Tumor Biology*, *34*(5), 2721–2729. https://doi.org/10.1007/s13277-013-0824-x

[33] Wang, Y., Zhang, Y., Huang, Q., & Li, C. (2018). Integrated bioinformatics analysis reveals key candidate genes and pathways in Breast cancer. *Molecular Medicine Reports*, *17*(6), 8091–8100. https://doi.org/10.3892/mmr.2018.8895

[34] Wang, Z., Li, T. E., Chen, M., Pan, J. J., & Shen, K. W. (2020). miR-106b-5p contributes to the lung metastasis of breast cancer via targeting CNN1 and regulating Rho/ROCK1 pathway. *Aging*, *12*(2), 1867–1887. https://doi.org/10.18632/aging.102719

[35] Whelan, S. A., Lu, M., He, J., Yan, W., Saxton, R. E., Faull, K. F., … Chang, H. R. (2009). Mass spectrometry (LC-MS/MS) site-mapping of N-glycosylated membrane proteins for breast cancer biomarkers. *Journal of Proteome Research*, *8*(8), 4151–4160. https://doi.org/10.1021/pr900322g

[36] Wu, J. R., Zhao, Y., Zhou, X. P., & Qin, X. (2020). Estrogen receptor 1 and progesterone receptor are distinct biomarkers and prognostic factors in estrogen receptor-positive breast cancer: Evidence from a bioinformatic analysis. *Biomedicine and Pharmacotherapy*,*121*(June2019),109647. https://doi.org/10.1016/j.biopha.2019.109647

[37] Zhang, M., Liu, D., Li, W., Wu, X., Gao, C., & Li, X. (2016). Identification of featured biomarkers in breast cancer with microRNA microarray. *Archives of Gynecology and Obstetrics*, *294*(5), 1047–1053. https://doi.org/10.1007/s00404-016-4141-7

[38] Wu, Duojiao, Catherine M. Rice, and Xiangdong Wang. 2012. "Cancer Bioinformatics: A New Approach to Systems Clinical Medicine." *BMC Bioinformatics* 13(1):13–16.