IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Rationalizing Biological Pathways Responsible In Metastatic Triple Negative Breast Cancer via Identification and Validation of Hub Genes

[1]Architha Sankar,

[1]M. Tech Biotechnology, [1]School of Biosciences and Technology, Department of Biotechnology
[1]Vellore Institute Technology, Vellore, India

Abstract: Triple Negative Breast Cancer (TNBC) is a severe form of breast cancer with an increased incidence of metastasis and relapse. Since, the current conventional method of treatment remains ineffective due to poor diagnosis and lack of therapeutic targets, it remains an unsettling challenge for both researchers and clinicians in the field. The centerpiece of our study was to unwind the Differentially Expressed Genes (DEGs), its pathway enrichment analysis underlying the metastatic condition of the disease. We had chosen multiple datasets from distant metastasis and metastatic TNBC stages for investigation. From these disease-specific expression network analysis, ten hub genes namely: SMC4, BUB1B, CCNA2, KIF2C, KIF15, PBK, CDCA5, CENPE, ZWINT and MELK were identified and validated by survival and correlation analysis. In our attempt to highlight the biological pathway that our hub genes were majorly involved, further analysis revealed that cleavage of centromeric cohesion by ESPL1 (Separase) in mitotic cell cycle were found to be enriched which demonstrates that these hub genes are responsible in triggering tumor invasion and metastasis. Concisely, our report will aid in gaining a better understanding of biological complexities, revealing potential biomarkers and therapeutic targets implicated in metastasis of TNBC.

Keywords **- Hub Genes, Triple-Negative Breast Cancer, In-silico Analysis, DEG's, Metastasis.**

I. INTRODUCTION

Triple Negative Breast Cancer (TNBC) is a subtype of breast cancer that is ER-negative, PR-negative, or HER2 negative. It accounts for 15% to 20% of all breast cancer cases. TNBC is characterized by the deficiency of Estrogen receptor (ER) and Progesterone receptor (PR), and in the absence of overexpression of Human Epidermal Growth factor Receptor 2 (HER2) (Medina et al., 2020). TNBC makes about 13% of all breast cancers and is associated to more aggressive tumor accumulation, a greater recurrence rate, and has the least therapeutic prognosis in comparison with other forms of cancers (Borri & Granaglia, 2021). Previous studies have found that TNBC has a significant chance of metastasis and tends to act aggressively, resulting in poor outcome for patients. Available therapeutics for metastatic TNBC are mainly ineffective due to the lack of gene targets for metastasis. The analysis of metastatic TNBC requires in-depth clarification with respect to understanding its molecular mechanism aiding to the improvement of clinical outcome (Heeke & Tan, 2021). The precise molecular changes associated with the disease state of cancer are highlighted by DNA microarray data, and these signatures can be used to develop innovative hypothesis. Although prior studies on triple-negative breast cancer microarray data have been conducted, the focus was either on a specific stage of TNBC or on a single microarray dataset with a small sample size. With gene expression studies growing by advancements and better bioinformatics approaches on the horizon, a meta-analytical research integrating numerous microarray datasets, each with samples from different phases of the disease, was considered necessary. As a whole, this analysis incorporated many phases of TNBC and provides a thorough understanding of the overlap of critical genes and pathways involved in sequential metastasis of triple-negative breast cancer. We used several statistical and comparative analysis on the acquired microarray data, including meta-analysis and network backbone construction, to identify significant hub genes that might be prospective therapeutic targets as well as possible biomarkers for TNBC with prognostic significance. By further pathway analysis and validation of these hub genes, the significant pathways associated with this disease were identified in order to have a clear picture of the significant role these hub genes possess and its dysregulation that might greatly impact on the triggering metastasis. As a result, we presume that by constructing a protein–protein interaction network and evaluating it in-depth, we will be able to achieve our goal of identifying novel hub genes and its biological pathways that play a key role in the metastatic mutilation and chemo-resistant pattern of TNBC.

II. RESEARCH METHODOLOGY

2. 1 Dataset Collection and Identification of DEGs:

For differential expression analysis, we retrieved microarray datasets 1. GSE5446 (Limame et al., 2013), 2. GSE61724 (Mathe et al., 2015) and 3. GSE95472 (Barton et al., 2017) from GEO database (URL: http://www.ncbi.nlm.nih.gov/geo). The above mentioned datasets were selected with respect to chemo-resistance, metastasis or distant metastasis in TNBC. The following were the inclusion criteria for microarray datasets: (a) metastatic TNBC and non-metastatic TNBC tissues in samples, (b) expression

profiling using an array as the research type, and (c) organism as Homo sapiens. According to the appropriate GEO platform files, the probe names were tagged to Entrez ID. The dataset GSE5446 was constructed upon Illumina Human HT-12 V4.0 expression bead-chip using the platform GPL10558 and was specifically incorporated for the comparison of references between early and late migratory cells to predict early metastatic relapse in TNBC and each of which were represented by 3 samples. GSE61724 was built on [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] via the platform GPL6244 was used as the second dataset which comprises of 69 samples in total. Out of this, 64 samples were Invasive Ductal Carcinoma (IDC) and the rest were Normal Adjacent Tissues (NAT) from human specimen in TNBC type. GSE95472 was taken as the third dataset inclusive of [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] using the platform GPL6244 to compare and contrast BT549 attached with suspended cells since TNBC cells in forced suspension mimics early metastatic stage. The overall process for the identification hub genes is displayed in Figure 1. (Zhong et al., 2020). The differential expression analysis for each of the GEO datasets was performed using the R package Limma (URL: https://www.bioconductor.org/packages/release/bioc/html/limma.html), while the background correction was made using Robust Multi-Array Average (RMA) approach. The R package genefu (URL: https://www.bioconductor.org/packages/release/bioc/html/genefu.html) was used to do the PAM50 molecular subtyping (Cao et al., 2021). Secondly, we then used Robust Rank Aggregation (RRA) analysis using the R package Robust Rank Aggreg package (https://cran.r-project.org/web/ packages/RobustRankAggreg/index.html) to filter out the common DEGs from all of these datasets. This technique is more accurate for identifying DEGs from various datasets since it is sensitive to outliers, noise, and flaws. A criteria of adjusted P-value (Adj.P) 0.05 and |log2 fold-change (FC)| > 1 was employed.

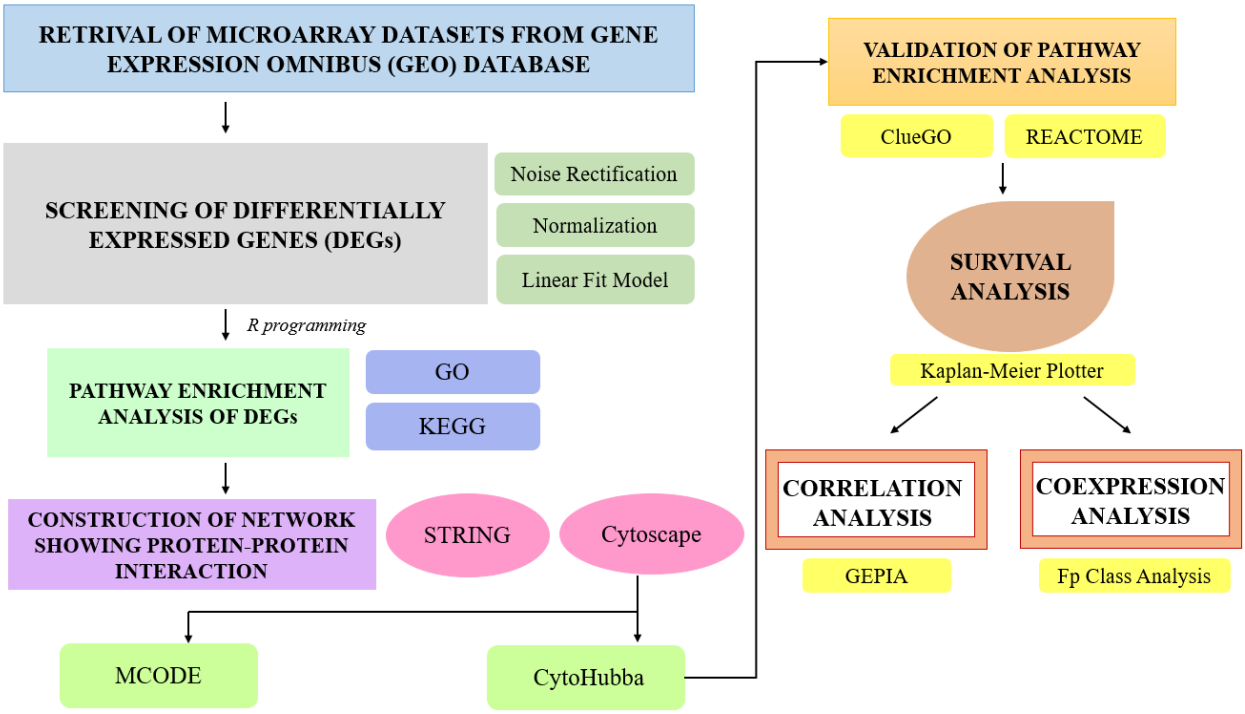

Figure 1. Represents the overall schematic workflow of our study

2. 2 Functional and Pathway Enrichment Analysis:

We analyzed the overall functional and pathway enrichment analysis of the identified DEG's using the most exclusive tools such as: the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) tools. The goal of the GO enrichment analysis was to learn more about how DEGs work biologically.

Gene Ontology is a digital depiction of our current scientific understanding about the activities of genes (or, more precisely, the protein and non-coding RNA molecules encoded by genes) in a variety of creatures, from humans to bacteria (Ashburner et al., 2000). Certain key aspects pertaining to molecular function, biological process, and cellular components were computed as results by the GO resource (https://geneontology.org/). KEGG is a comprehensive platform for deducing high-level functions and efficiencies of biological systems including cells, organisms, and ecosystem from molecular-level data, relatively large molecular information produced by genome sequencing and other high-throughput methodologies (https://www.genome.jp/kegg/). The threshold for KEGG was fixed at $P < 0.05$ (Guo et al., 2017).

2. 3 Analysis of PPI:

We performed the analysis of PPI network my implementing the Cytoscape software. The core programme from Cytoscape allows the user to create and query the network, as well as graphically integrate it with expr ession profiles, phenotypes, and other molecular states, and link it to databases of functional annotations. Analysis of protein-protein interaction and the preliminary identification of hub genes were performed in this step by the implementation of Cytoscape (version 3.9.1) software. Cytoscape is an open source software project that brings together bio molecular interaction networks, high throughput expression data, and other molecular states into a single conceptual framework. The core may be extended using a simple plug-in architecture, allowing for the quick development of new computational analysis and functionality (Shannon et al., 2003). This software was incorporated to build a PPI network based on the data obtained from another database called the Search Tool for the Retrieval of Interacting Genes (STRING; version 11.0, URL: https://string-db.org/). STRING is a database of protein-protein interactions that is both known and anticipated.

The interactions arise from computer prediction, information transmission across species, and interactions gathered from other (pr imary) databases, and they contain both direct (stuctural) and indirect (functional) connections (Szklarczyk et al., 2020).
In this study, it evaluates the interactions of proteins encoded by the observed DEGs which is shown in Fig 2. A total score of 0.7 was used as a cut-off point (Q. Zhai et al., 2019).

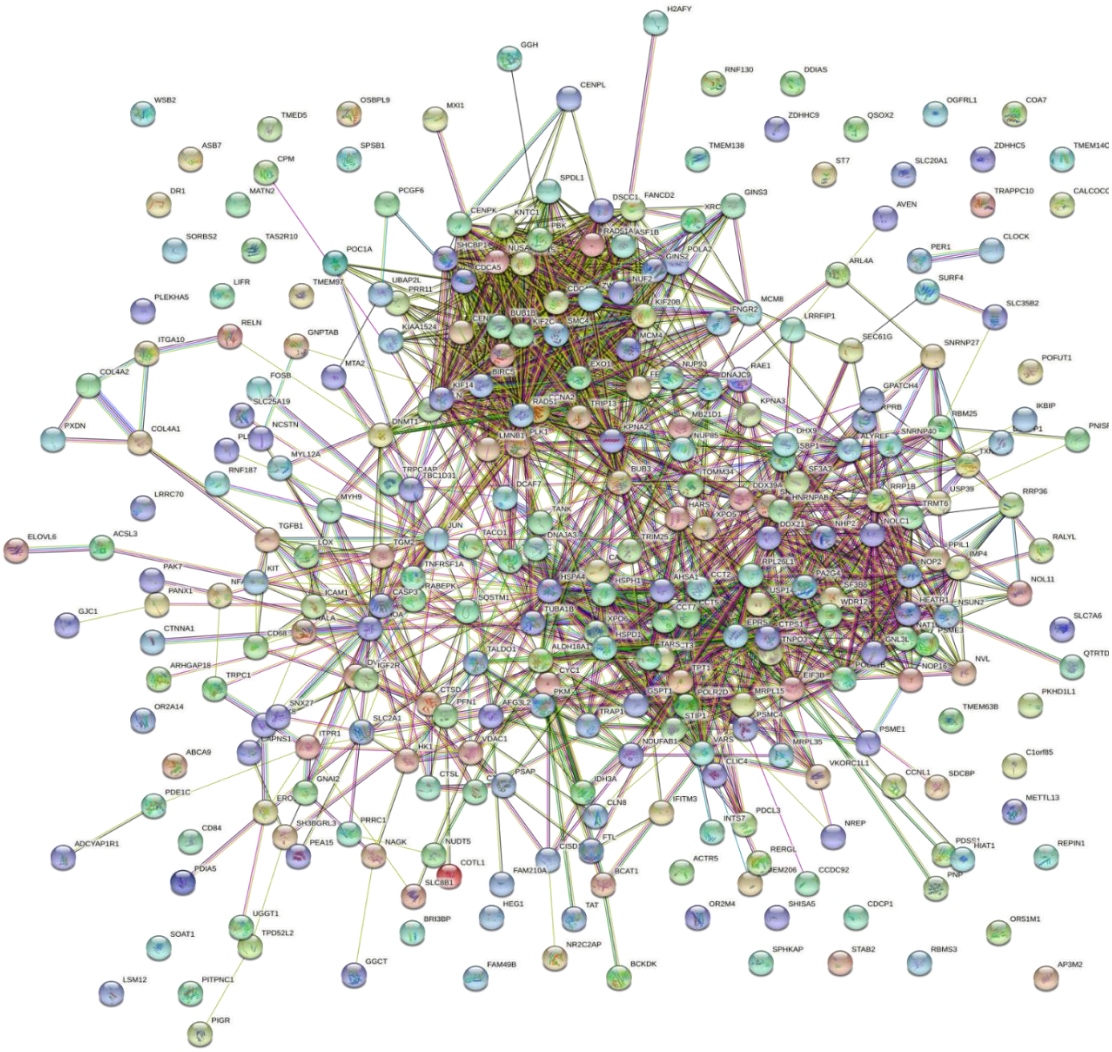

Figure 2. Shows the PPI network analysis using STRING database among the top identified DEGs.

Following that, a Cytoscape plug-in called MCODE (version 1.5, URL: https://apps.cytoscape.org/apps/mcode) was used to screen major functional modules from the PPI network, and another plug-in called cytohubba (version 0.1, URL: apps.cytoscape.org/apps/cytohubba) was used to identify hub genes. Moreover, Maximal Clique Centrality (MCC) technique was used for identifying hub genes through cytohubba plug-in ((Li et al., 2021).
What's more, Cluego plug-in was administered to identify the major biological pathways that our hub genes are majorly involved. Further validation of pathway enrichment was carried using REACTOME (URL: https://reactome.org/) database.

2. 4 Survival and Co-Expression Analysis of the Identified Genes:

The retrieved hub genes were further filtered on the basis of survival analysis. The Kaplan-Meier prediction is among the finest ways to determine the percentage of participants who live for a specific period of time following therapy.
The impact of an intervention is determined in clinical trials or community trials by quantifying the number of individuals who lived or were saved after the intervention over duration. Survival time is defined as thi s time from a defined point to the occurrence of a certain event, such as death, while survival analysis is defined as the study of group data (Goel et al., 2010).
This database (URL: https://kmplot.com/analysis/) has the ability to analyse the impact of 54,675 genes on survival upon 10,293 cancer samples. Using the Kaplan-Meier plotter, we analysed recurrence free survival of important genes in breast cancer patients from the TCGA-BRCA database, with prognosis deemed significant if a log rank p value was less than 0.05 (X. Zhai et al., 2020).
The two most significant genes from survival analysis was further performed using correlation analysis via the Gene Expression Profiling Interactive Analysis (GEPIA; URL: https://gepia.cancer-pku.cn/) online database. To find out which genes were co-expressed with hub genes, we used the FpClass programme (http://dcv.uhnres.utoronto.ca/FPCLASS), an in silico approach for id entifying high confidence protein–protein interactions on the proteome level in order to revalidate our hub genes. The co-expressed genes derived from the combined scores gave us a clear picture of the implicated partner genes and the accompanying molecular networks that might be associated in metastatic TNBC.
While investigating the hub genes in our study, two specific scores were taken into account: gene expression score and network topology score. The network topology score examines whether the genes are present i n the training data as well as the intensity of their interactions (Kotlyar et al., 2015).

III. RESULTS

3. 1 Identification of DEG's:

To identify DEGs using R programming, the three datasets were individually analysed (P value < 0.01, |log FC| = 2). After evaluating GSE54465, a total of 4,877 genes were found, with 1424 genes up-regulated and 3453 genes down-regulated. GSE61724 dataset, revealed 4834 DEGs, 1422 genes that were up-regulated, and 3412 genes that were down-regulated. In addition, GSE95472 yielded 4820 DEGs, with 2211 up-regulated genes and 2609 down-regulated genes. Figure 2. A–C. depicts the distribution of DEGs in each dataset as volcano plots and their expression as heat-maps. According to the Venn diagrams, a total of 504 DEGs overlapped, as seen in Figure 3.

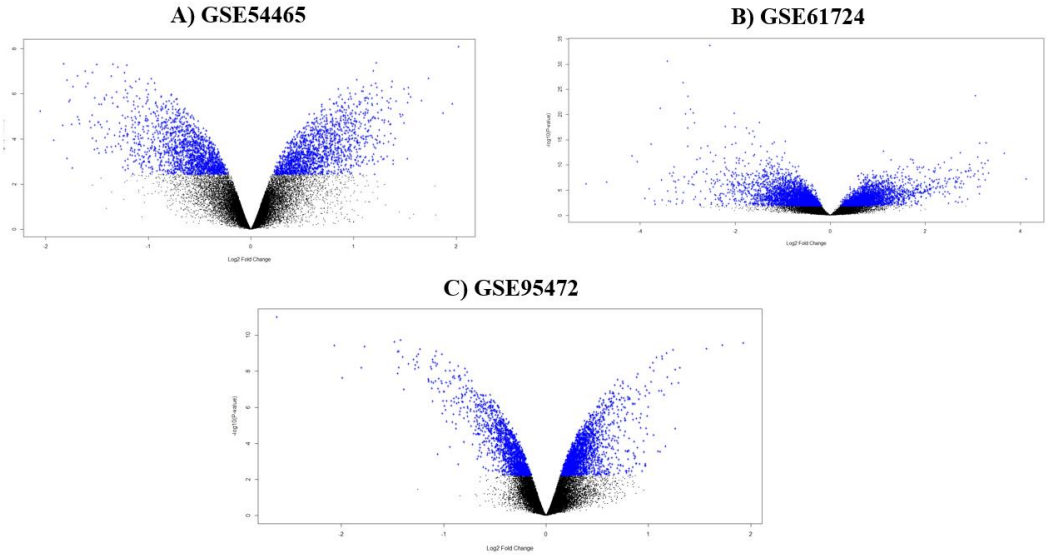

Figure 2A-C. Demonstrates the Volcano Plots for the datasets A) GSE54465, B) GSE61724 and C) GSE95472. The left side of the blue ink mean up-regulated genes screened on the basis |log2 fold change| ≥ 0.585 and P-value < 0.05. The right side of the blue ink mean down-regulated genes screened on the basis |log2 fold change| ≥ 0.585 and P-value < 0.05

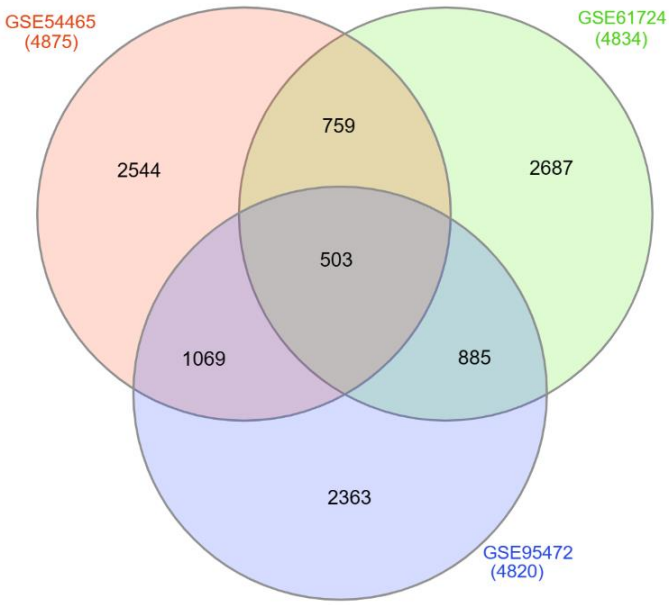

Figure 3. Venn Analysis showing the overlapping 504 DEGs among the three datasets out of which 264 are down-regulated and 236 are up-regulated DEGs.

3. 2 Pathway Enrichment Analysis:

GO functional and KEGG pathway enrichment analysis of 504 significant DEGs were taken to acquire an understanding into the existing Biological Pathways and Molecular Functions involved in TNBC. These findings revealed that those substantial DEGs were considerably abundant in over 200 GO terms and 6 main KEGG pathways (Figure 4A).

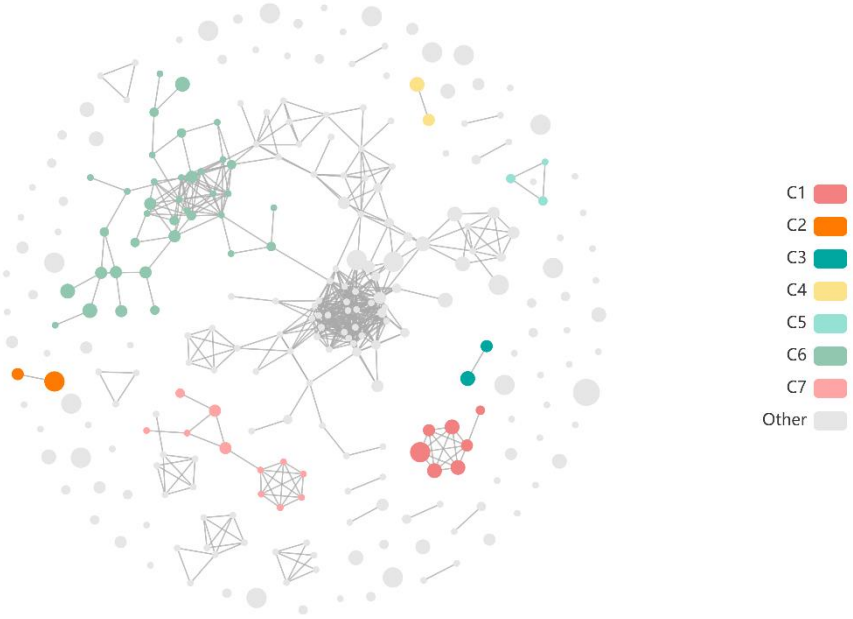

Figure 4A. Each node represents an enriched term, and the node colour represents different clusters; the node size represents 6 levels of enriched p-value, node size from small to large.

The top GO terms for Biological Process such as protein metabolic processes, mitotic cell cycle, DNA replication, protein-containing complex organization, and chaperone-mediated protein folding. Similarly, its molecular functions as well as cellular component have been depicted in the below Table 2.

Significant Terms	Description	P Value	Number of Associated DEGs
Biological Process			
GO:0019538	Protein Metabolic Process	4.28E-06	142
GO:0006807	Nitrogen Compound Metabolic Process	8.83E-21	268
GO:0000278	Mitotic Cell Cycle	4.18E-12	49
GO:0055133	DNA Replication	7.15E-11	219
GO:0043933	Protein-Containing Complex Organization	3.58E-08	71
GO:0006457	Protein Folding	1.46E-11	28
Molecular Function			
GO:0003676	Nucleic Acid Binding	1.46E-11	158
GO:0005515	Protein Binding	1.53E-29	466
GO:0016887	ATP Hydrolysis Activity	3.77E-12	37
GO:0004930	G- Protein Coupled Receptor Activity	9.33E-09	12
GO:0003700	DNA Binding Transcription Factor Activity	2.72E-04	16
GO:0005524	ATP Binding	1.01E-13	89
Cellular Component			
GO:0032991	Protein Containing Complex	1.03E-20	243
GO:0031974	Membrane Enclosed Lumen	6.39E-33	270
GO:0005743	Mitochondrial Inner Membrane	2.31E-04	27
GO:1990904	Ribonucleoprotein Complex	6.44E-11	50
GO:0005737	Cytoplasm	1.56E-28	419

Table 2. GO Enrichment Analysis on the basis of Biological Process, Molecular Function and Cellular Component for the identified DEGs

Besides, KEGG Pathway revealed that DNA replication, 2-oxocarboxylic acid metabolism, one carbon pool by folate, protein export were shown to be strongly related to these robust DEGs (Figure 4B).

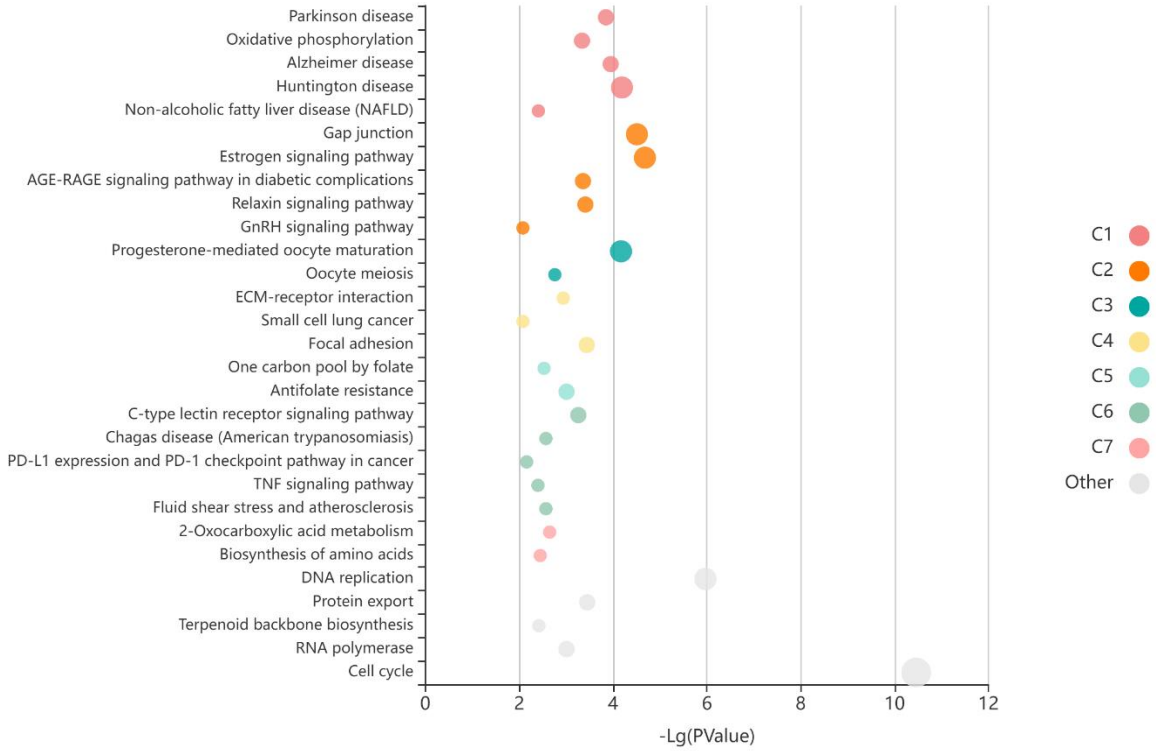

Figure 4B. Each row represents an enriched function, and the length of the bar represents the enrich ratio which is calculated as "input gene number" or "background gene number". The colour of the bar is the same as the colour in the circular network in Figure. 4A, which represents different clusters.

3. 3 Protein-Protein Interaction Network and Cluster Analysis:

A PPI network was built to better investigate the relationships between the 504 DEGs. STRING was used to create a PPI network with 317 nodes and 1077 edges using an interaction score of >0.7 as the cut-off condition. On this basis, Cytoscape software implemented MCODE plug-in for the identification and analysis of hub genes the top 3 clusters were displayed based on number nodes, edges as well as the overall cluster score for each individually. Module 1 (shown in the Figure 5A.) exhibited a maximum score of 21.28 (cluster score ≥ 20 is accepted) while the scores of module 2 and 3 (as shown in the Figure 5B and C) were 15.87 and 5.13 respectively. Furthermore, using the cytoHubba plugin the 10 hub genes (Figure. 6) with high degrees of connection and on the basis of MCC score within the PPI network were identified: SMC4, BUB1B, CCNA2, KIF2C, KIF15, PBK, CDCA5, CENPE, ZWINT and MELK (supplementary Table. 3).

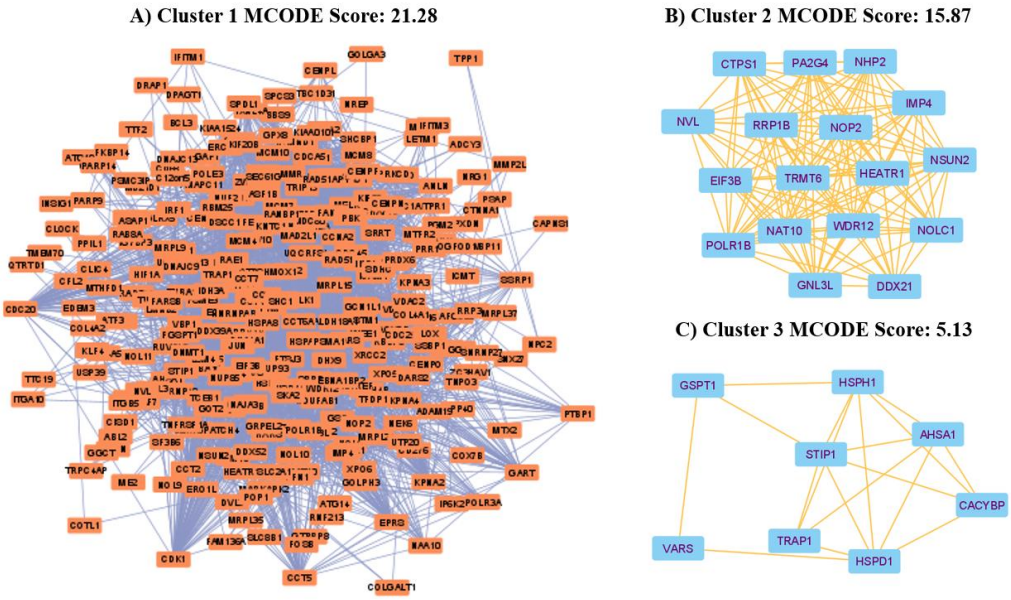

Figure 5 A-C. Captures the top 3 clusters obtained from MCODE plug-in using Cytoscape software

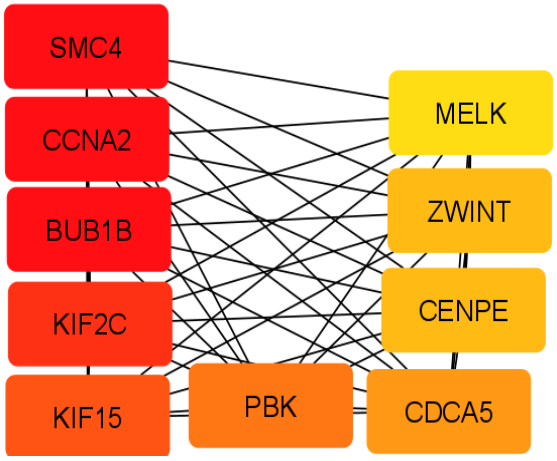

Figure 6. Genes labelled in red boxes shown in the figure states the top ranked genes while the boxes in orange and yellow are to low ranked genes obtained from MCC Algorithm.

Rank	Gene Name	Score
1	SMC4	3.50E+22
1	BUB1B	3.50E+22
1	CCNA2	3.50E+22
4	KIF2C	3.50E+22
5	KIF15	3.50E+22
6	PBK	3.50E+22
7	CDCA5	3.50E+22
8	CENPE	3.50E+22
8	ZWINT	3.50E+22
10	MELK	3.50E+22

Table 3. Represents the ranking of the identified hub genes along with their respective scores.

3. 4 Validation of Pathway Enrichment of the Identified genes:

In accordance with previous evidence of pathway enrichment, the analysis was further filtered for accurate results by evaluating with respect to the identified hub genes. The 10 hub genes were incorporated in Cytoscape software by using ClueGO plugin, a cluster network of all the genes with their respective biological pathways in which they play a maximum role was depicted.

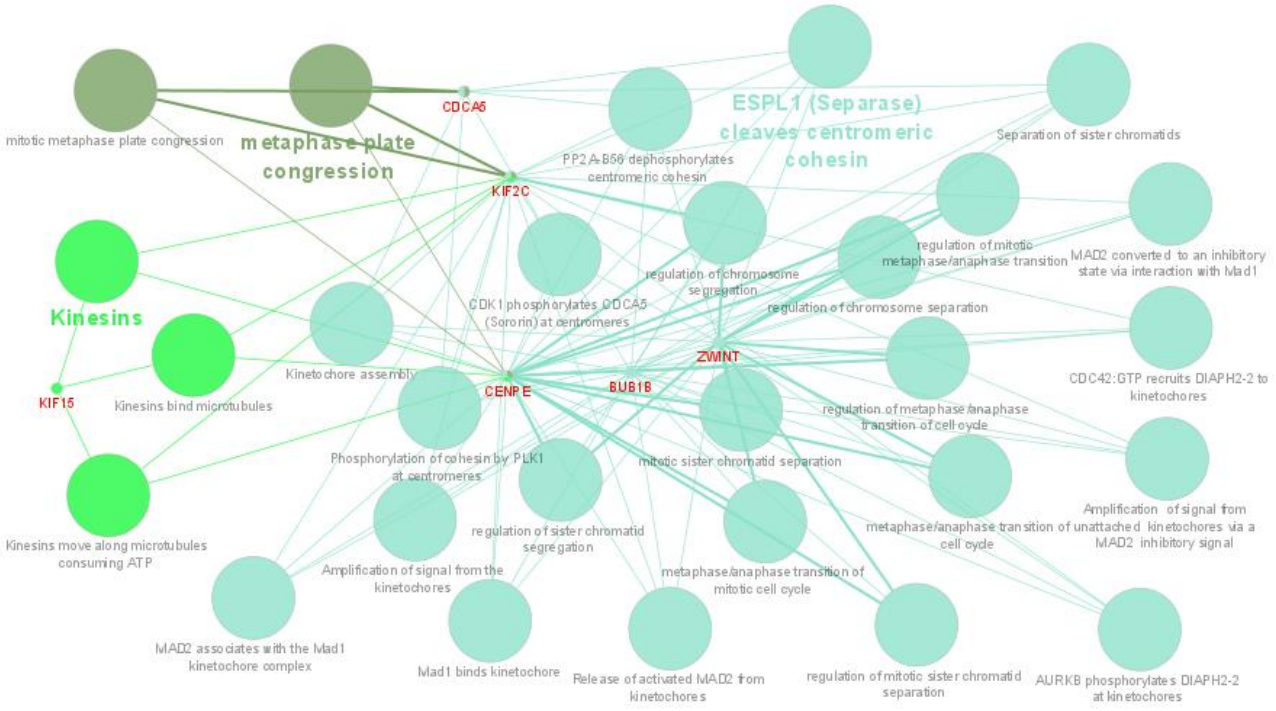

Figure 7. Depicts the validation of biological pathways that the identified hub genes play an important role using Cluepedia tool in Cluego plug-in.

We further refined this using the CluePedia toolbar which displayed a network of clusters shared between pathways which otherwise is described as more than one gene encoding for the same function (Figure. 7).

ID	Biological Pathway	Percentage of Associated Genes	Gene Names
R-HSA:1638803	Phosphorylation of cohesin by PLK1 at centromeres	4.10	BUB1B, CDCA5, CENPE, KIF2C, ZWINT
R-HSA:1638821	PP2A-B56 dephosphorylates centromeric cohesin	4.10	BUB1B, CDCA5, CENPE, KIF2C, ZWINT
R-HSA:2467809	ESPL1 (Separase) cleaves centromeric cohesin	4.07	BUB1B, CDCA5, CENPE, KIF2C, ZWINT
R-HSA:2467811	Separation of sister chromatids	4.10	BUB1B, CDCA5, CENPE, KIF2C, ZWINT
R-HSA:2468287	CDK1 phosphorylates CDCA5 at centromeres	4.00	BUB1B, CDCA5, CENPE, KIF2C, ZWINT
R-HSA:2484822	Kinetochore assembly	4.10	BUB1B, CDCA5, CENPE, KIF2C, ZWINT

Table 4. Various biological pathway names along with the percentage of associated genes and the maximum number of hub genes involved in each pathway has been tabulated.

On the basis of the results featured by ClueGO as well as validation of pathway enrichment was performed by REACTOME which highlighted the major pathway that our genes are involved: "ESPL1 (Separase) cleaves centromeric cohesion in Mitotic Cell Cycle" (Figure. 7).

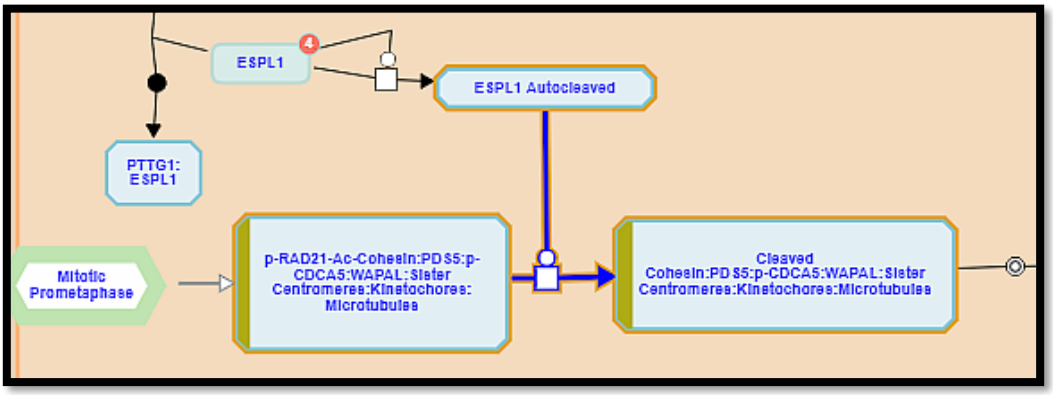

Figure 8. Shows the exact location of ESPL1 as it cleaves centromeric cohesion in mitotic pro-metaphase cell cycle is highlighted in bright blue colour.

3. 5 Survival Analysis:

In 4929 BC patients, we used the Kaplan-Meier plotter tool to assess hub genes for recurrence free survival. Simultaneously, we analysed all 12 hub genes that contributes to the survival in BC patients having high and low expression. KIF2C, MELK, CCNA, and SSK1 were found to be strongly tied to Relapse Free Survival (RFS) in BC patients. Although, there were no significant differences detected in the KIF2C and MELK hub genes.

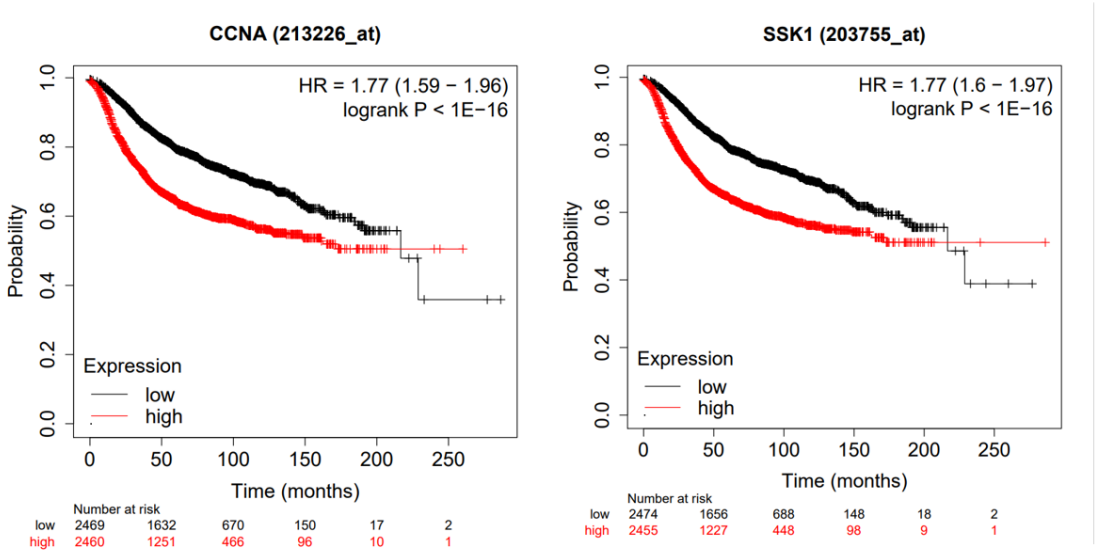

Figure 9. The RFS survival analysis using Kaplan-Meier plotter of CCNA and BUB1B (SSK1) showing HR- Hazard Ratio value of 1.77, red lines represent high expression and black line represent low expression of genes.

However, only the low expression of CCNA and SSK1 were favorable with respect to the RFS survival on the basis of certain criteria like: Hazard Ratio = 1.77 with 95% Confidence Interval (CI) = 0.48 – 0.51, P Val < 1E-16 for both CCNA and SSK1. In TNBC patients, favorable prognostic variables: CCNA and SSK1 were considered to be associated to metastatic-free survival. Figure. 9 depicts the outcomes.

IV. DISCUSSION

Many new frontiers are being explored in different malignancies to identify essential biomolecules and cellular processes with prognostic and diagnostic significance, thanks to recent improvements in gene expression microarray studies and the introduction of innovative methods of bioinformatical analysis. We used a large-scale meta-analysis to combine samples from different stages of TNBC in our study, which allowed us identify significant genes linked to metastatic illness. To begin with, the microarray datasets that matched the inclusion criteria were calibrated before being used in the meta-analysis. Followed by, pathway enrichment analysis was performed by utilizing the identified DEGs. Subsequently, the expression of TNBC tissue specific PPI network were constructed and the hub genes obtained from their conjuncture were subjected to relapse-free survival analysis and partner gene prediction which is otherwise known as co-expression analysis. Further elucidation of biological pathway validation were performed to decipher the original functionalities of the hub genes. In the likewise process, we retrieved 10 hub genes namely: SMC4 (Structural maintenance of chromosomes protein 4), BUB1B (BUB1 Mitotic Checkpoint Serine/Threonine Kinase B), CCNA2 (Cell Cycle Regulator Cyclin-A2), KIF2C (kinesin family member 2C), KIF15 (Kinesin Family Member 15), PBK (PDZ binding kinase), CDCA5 (Cell Division Cycle Associated 5), CENPE (centromere protein E), ZWINT (ZW10 Interacting Kinetochore Protein) and MELK (Maternal Embryonic Leucine Zipper Kinase) through PPI network construction. Subsequently, the RFS survival predicted that CCNA2 and BUB1B to the two most significant hub genes among the ten. This was further validated by correlation analysis to elucidate the relationship between the two robust genes.

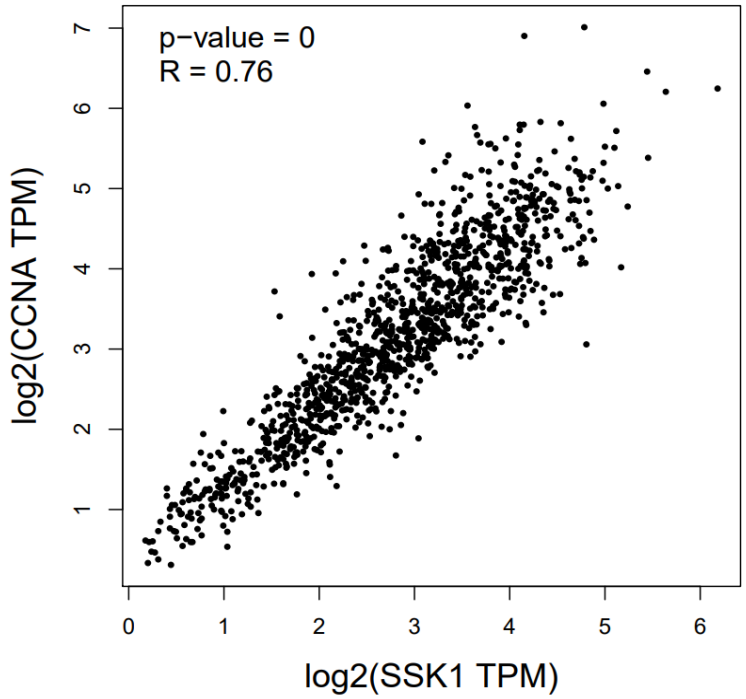

Figure 10. The correlation between CCNA and BUB1B is provided as a graphical representation and any R value in the range of 0.7-1 is considered accepted.

In colorectal, ovarian, and breast cancer, CCNA2 is involved in modulating cell cycle progression as well as increasing proliferation and carcinogenesis. Furthermore, bioinformatics studies has shown that CCNA2 is substantially expressed in TNBC tissues (Lu et al., 2022, p. 1). (Koyuncu et al., 2021) have demonstrated that the loss of function in BUB1B have resulted in anti-apoptotic activity and aggressiveness leading to metastasis in TNBC. Moreover, both CCNA2 and BUB1B has a long history of being linked to the carcinogenic change of breast tissue cells from epithelial to mesenchymal, which leads to metastasis and chemo-resistance (Cai et al., 2019).

Besides this, we further went ahead to estimate the co-expression of SMC4, BUB1B, CCNA2, KIF2C, KIF15, PBK, CDCA5, CENPE, ZWINT and MELK using Fp Class co-expression analysis database (URL: http://dcv.uhnres.utoronto.ca/FPCLASS/) that is shown in Table. 5 to understand and obtain gene-gene expression similarity and its impact on the disease.

Query ID	Predicted Partner Symbol	Total Score	Network Topology Score
SMC4	SMC2 (SMC-2)	0.9429	0.1906
BUB1B	MAD2L1	0.9429	0.206
CCNA2	FEN1	0.9429	0.201
KIF2C	MTUS2	0.8826	0.6791
KIF15	CBX5	0.7915	0.4669
KIF15	MKI67IP	0.7426	0.7184
PBK	RRM2	0.9429	0.111
CDCA5	REC8	0.8577	0.5759
CENPE	NDC80	0.8826	0.2223

ZWINT	NDC80	0.9429	0.2374
MELK	MYBL2	0.8826	0.2984

Table 5. Fp Class tool shows the co-expression analysis of 10 hub genes on the basis of total and network topology scores.

What's more, the two most significant hub genes CCNA2 and BUB1B have already been previously reported as potential biomarkers in TNBC which comes to the fact that our meta-analysis have confirmed the same thereby providing more accuracy towards implementing the identified hub genes as potential biomarkers for disease diagnosis and targeted therapy. Although, our analysis requires in-vitro as well as in-vivo validation, it does provide a basic understanding at the molecular level. Perhaps our study would possibly ignite an in-depth research and its application towards targeting the therapeutic aspects of metastatic TNBC the near future.

V. CONCLUSION

In this study, we strived to elucidate the microarray datasets with identification of hub genes and understanding its biological significance responsible for malignancy through multiple stages leading to metastatic disease. This substantial high-throughput meta-analytical investigation incorporating several datasets to identify DEGs as well as their enriched pathways leading to an observation of total ten highly important hub genes. For this rationale, the PPI network construction was utilized and further validation of pathway enrichment analysis were performed additionally to understand its functional impact. Although, previous studies have reported similar records of the majority of hub genes like: BUB1B, CCNA2, MELK, KIF2C, SMC4 and PBK that have been linked to the development and metastasis of TNBC. Our research has revealed some novel genes like: KIF15, PBK, CDCA5 and CENPE that were solely identified through our study. Moreover, ClueGO analysis combined with REACTOME database for pathway enrichment validation revealed that the identified hub genes were majorly involved in the cleavage of centromeric cohesion by ESPL1 (Separase) in Mitotic Cell Cycle. Therefore, this study provides a foundation for further research to dive deep into the biological and molecular impact of the identified hub genes and its involvement in metastatic TNBC providing new insights in early diagnosis and targeted therapy for triple-negative breast cancer.

VI. ACKNOWLEDGEMENT

The management of Vellore Institute of Technology (VIT), Vellore, is thanked by the author for providing the resources and support necessary to complete this study.

REFERENCE:

[1] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

[2] Barton, V. N., Christenson, J. L., Gordon, M. A., Greene, L. I., Rogers, T. J., Butterfield, K., Babbs, B., Spoelstra, N. S., D'Amato, N. C., Elias, A., & Richer, J. K. (2017a). ANDROGEN RECEPTOR SUPPORTS AN ANCHORAGE-INDEPENDENT, CANCER STEM CELL-LIKE POPULATION IN TRIPLE-NEGATIVE BREAST CANCER. *Cancer Research*, *77*(13), 3455–3466. https://doi.org/10.1158/0008-5472.CAN-16-3240

[3] Barton, V. N., Christenson, J. L., Gordon, M. A., Greene, L. I., Rogers, T. J., Butterfield, K., Babbs, B., Spoelstra, N. S., D'Amato, N. C., Elias, A., & Richer, J. K. (2017b). Androgen Receptor Supports an Anchorage-Independent, Cancer Stem Cell-like Population in Triple-Negative Breast Cancer. *Cancer Research*, *77*(13), 3455–3466. https://doi.org/10.1158/0008-5472.CAN-16-3240

[4] Borri, F., & Granaglia, A. (2021). Pathology of triple negative breast cancer. *Seminars in Cancer Biology*, *72*, 136–145. https://doi.org/10.1016/j.semcancer.2020.06.005

[5] Cai, Y., Mei, J., Xiao, Z., Xu, B., Jiang, X., Zhang, Y., & Zhu, Y. (2019). Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas*, *156*, 20. https://doi.org/10.1186/s41065-019-0096-6

[6] Cao, W., Jiang, Y., Ji, X., Guan, X., Lin, Q., & Ma, L. (2021). Identification of novel prognostic genes of triple-negative breast cancer using meta-analysis and weighted gene co-expressed network analysis. *Annals of Translational Medicine*, *9*(3), 205–205. https://doi.org/10.21037/atm-20-5989

[7] Chuan, T., Li, T., & Yi, C. (2020). Identification of CXCR4 and CXCL10 as Potential Predictive Biomarkers in Triple Negative Breast Cancer (TNBC). *Medical Science Monitor*, *26*. https://doi.org/10.12659/MSM.918281

[8] de Bastos, D. R., Conceição, M. P. F., Michelli, A. P. P., Leite, J. M. R. S., da Silva, R. A., Cintra, R. C., Sanchez, J. J. D., Vilanova-Costa, C. A. S. T., & Silva, A. M. T. C. (2021). An *In Silico* Analysis Identified FZD9 as a Potential Prognostic Biomarker in Triple-Negative Breast Cancer Patients. *European Journal of Breast Health*, *17*(1), 42–52. https://doi.org/10.4274/ejbh.2020.5804

[9] Dong, P., Yu, B., Pan, L., Tian, X., & Liu, F. (2018a). Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis. *BioMed Research International*, *2018*, 1–10. https://doi.org/10.1155/2018/2760918

[10] Dong, P., Yu, B., Pan, L., Tian, X., & Liu, F. (2018b). Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis. *BioMed Research International*, *2018*, 1–10. https://doi.org/10.1155/2018/2760918

[11] Fang, X., Yin, Z., Li, X., Xia, L., Quan, X., Zhao, Y., & Zhou, B. (2017a). Multiple functional SNPs in differentially expressed genes modify risk and survival of non-small cell lung cancer in chinese female non-smokers. *Oncotarget*, *8*(12), 18924–18934. https://doi.org/10.18632/oncotarget.14836

[12] Fang, X., Yin, Z., Li, X., Xia, L., Quan, X., Zhao, Y., & Zhou, B. (2017b). Multiple functional SNPs in differentially expressed genes modify risk and survival of non-small cell lung cancer in chinese female non-smokers. *Oncotarget*, *8*(12), 18924–18934. https://doi.org/10.18632/oncotarget.14836

[13] Gao, Y., Wang, X., Li, S., Zhang, Z., Li, X., & Lin, F. (2021). Identification of a DNA Methylation-Based Prognostic Signature for Patients with Triple-Negative Breast Cancer. *Medical Science Monitor*, *27*. https://doi.org/10.12659/MSM.930025

[14] Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, *1*(4), 274–278. https://doi.org/10.4103/0974-7788.76794

[15] Guo, J., Gong, G., & Zhang, B. (2017). Screening and identification of potential biomarkers in triple-negative breast cancer by integrated analysis. *Oncology Reports*, *38*(4), 2219–2228. https://doi.org/10.3892/or.2017.5911

[16] He, J., Yang, J., Chen, W., Wu, H., Yuan, Z., Wang, K., Li, G., Sun, J., & Yu, L. (2015). Molecular Features of Triple Negative Breast Cancer: Microarray Evidence and Further Integrated Analysis. *PLOS ONE*, *10*(6), e0129842. https://doi.org/10.1371/journal.pone.0129842

[17] He, Y., Cao, Y., Wang, X., Jisiguleng, W., Tao, M., Liu, J., Wang, F., Chao, L., Wang, W., Li, P., Fu, H., Xing, W., Zhu, Z., Huan, Y., & Yuan, H. (2021). Identification of Hub Genes to Regulate Breast Cancer Spinal Metastases by Bioinformatics Analyses. *Computational and Mathematical Methods in Medicine*, *2021*, 5548918. https://doi.org/10.1155/2021/5548918

[18] Heeke, A. L., & Tan, A. R. (2021). Checkpoint inhibitor therapy for metastatic triple-negative breast cancer. *Cancer and Metastasis Reviews*, *40*(2), 537–547. https://doi.org/10.1007/s10555-021-09972-4

[19] Houghton, S. C., & Hankinson, S. E. (2021). Cancer Progress and Priorities: Breast Cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *30*(5), 822–844. https://doi.org/10.1158/1055-9965.EPI-20-1193

[20] *Identification of a prognosis-associated signature associated with energy metabolism in triple-negative breast cancer.pdf*. (n.d.).

[21] *Identification of key genes as potential biomarkers for triple-negative breast cancer using integrating genomics analysis.pdf*. (n.d.).

[22] Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102409/

[23] Kotlyar, M., Pastrello, C., Pivetta, F., Lo Sardo, A., Cumbaa, C., Li, H., Naranian, T., Niu, Y., Ding, Z., Vafaee, F., Broackes-Carter, F., Petschnigg, J., Mills, G. B., Jurisicova, A., Stagljar, I., Maestro, R., & Jurisica, I. (2015). In silico prediction of physical protein interactions and characterization of interactome orphans. *Nature Methods*, *12*(1), Article 1. https://doi.org/10.1038/nmeth.3178

[24] Koyuncu, D., Sharma, U., Goka, E. T., & Lippman, M. E. (2021). Spindle assembly checkpoint gene BUB1B is essential in breast cancer cell survival. *Breast Cancer Research and Treatment*, *185*(2), 331–341. https://doi.org/10.1007/s10549-020-05962-2

[25] Li, H., Guo, J., Cheng, G., Wei, Y., Liu, S., Qi, Y., Wang, G., Xiao, R., Qi, W., & Qiu, W. (2021). Identification and Validation of SNP-Containing Genes With Prognostic Value in Gastric Cancer via Integrated Bioinformatics Analysis. *Frontiers in Oncology*, *11*, 564296. https://doi.org/10.3389/fonc.2021.564296

[26] Li, L., Huang, H., Zhu, M., & Wu, J. (2021). Identification of Hub Genes and Pathways of Triple Negative Breast Cancer by Expression Profiles Analysis. *Cancer Management and Research*, *Volume 13*, 2095–2104. https://doi.org/10.2147/CMAR.S295951

[27] Li, M.-X., Jin, L.-T., Wang, T.-J., Feng, Y.-J., Pan, C.-P., Zhao, D.-M., & Shao, J. (2018). Identification of potential core genes in triple negative breast cancer using bioinformatics analysis. *OncoTargets and Therapy*, *Volume 11*, 4105–4112. https://doi.org/10.2147/OTT.S166567

[28] Limame, R., de Beeck, K. O., Van Laere, S., Croes, L., De Wilde, A., Dirix, L., Van Camp, G., Peeters, M., De Wever, O., Lardon, F., & Pauwels, P. (2013). Expression profiling of migrated and invaded breast cancer cells predicts early metastatic relapse and reveals Krüppel-like factor 9 as a potential suppressor of invasive growth in breast cancer. *Oncoscience*, *1*(1), 69–81. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4295756/

[29] Limame, R., de Beeck, K. O., Van Laere, S., Croes, L., De Wilde, A., Dirix, L., Van Camp, G., Peeters, M., De Wever, O., Lardon, F., & Pauwels, P. (2014). Expression profiling of migrated and invaded breast cancer cells predicts early metastatic relapse and reveals Krüppel-like factor 9 as a potential suppressor of invasive growth in breast cancer. *Oncoscience*, *1*(1), 69–81. https://doi.org/10.18632/oncoscience.10

[30] Limame, R., Op de Beeck, K., Van Laere, S., Croes, L., De Wilde, A., Dirix, L., Van Camp, G., Peeters, M., De Wever, O., Lardon, F., & Pauwels, P. (2014). Expression profiling of migrated and invaded breast cancer cells predicts early metastatic relapse and reveals Krüppel-like factor 9 as a potential suppressor of invasive growth in breast cancer. *Oncoscience*, *1*(1), 69–81. https://doi.org/10.18632/oncoscience.10

[31] Lin, Z., Peng, R., Sun, Y., Zhang, L., & Zhang, Z. (2021). Identification of ribosomal protein family in triple-negative breast cancer by bioinformatics analysis. *Bioscience Reports*, *41*(1), BSR20200869. https://doi.org/10.1042/BSR20200869

[32] Lu, Y., Su, F., Yang, H., Xiao, Y., Zhang, X., Su, H., Zhang, T., Bai, Y., & Ling, X. (2022). E2F1 transcriptionally regulates CCNA2 expression to promote triple negative breast cancer tumorigenicity. *Cancer Biomarkers*, *33*(1), 57–70. https://doi.org/10.3233/CBM-210149

[33] Mathe, A., Wong-Brown, M., Morten, B., Forbes, J. F., Braye, S. G., Avery-Kiejda, K. A., & Scott, R. J. (2015a). Novel genes associated with lymph node metastasis in triple negative breast cancer. *Scientific Reports*, *5*, 15832. https://doi.org/10.1038/srep15832

[34] Mathe, A., Wong-Brown, M., Morten, B., Forbes, J. F., Braye, S. G., Avery-Kiejda, K. A., & Scott, R. J. (2015b). Novel genes associated with lymph node metastasis in triple negative breast cancer. *Scientific Reports*, *5*, 15832. https://doi.org/10.1038/srep15832

[35] McCall, M. N., Bolstad, B. M., & Irizarry, R. A. (2010). Frozen robust multiarray analysis (fRMA). *Biostatistics (Oxford, England)*, *11*(2), 242–253. https://doi.org/10.1093/biostatistics/kxp059

[36] Medina, M. A., Oza, G., Sharma, A., Arriaga, L. G., Hernández Hernández, J. M., Rotello, V. M., & Ramirez, J. T. (2020). Triple-Negative Breast Cancer: A Review of Conventional and Advanced Therapeutic Strategies. *International Journal of Environmental Research and Public Health*, *17*(6), 2078. https://doi.org/10.3390/ijerph17062078

[37] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003a). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303
[38] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003b). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303
[39] Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2020). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074
[40] Wang, S., & Quan, Y. (n.d.). *Identification of invasive key genes in breast cancer by bioinformatics analysis*. 10.
[41] Wang, Y., Zhang, L., Chen, Y., Li, M., Ha, M., & Li, S. (2020). Screening and identification of biomarkers associated with the diagnosis and prognosis of lung adenocarcinoma. *Journal of Clinical Laboratory Analysis*, *34*(10). https://doi.org/10.1002/jcla.23450
[42] Wei, L.-M., Li, X.-Y., Wang, Z.-M., Wang, Y.-K., Yao, G., Fan, J.-H., & Wang, X.-S. (2021). Identification of hub genes in triple-negative breast cancer by integrated bioinformatics analysis. *Gland Surgery*, *10*(2), 799–806. https://doi.org/10.21037/gs-21-17
[43] Wu, T., Wang, X., Li, J., Song, X., Wang, Y., Wang, Y., Zhang, L., Li, Z., & Tian, J. (2015). Identification of Personalized Chemoresistance Genes in Subtypes of Basal-Like Breast Cancer Based on Functional Differences Using Pathway Analysis. *PLOS ONE*, *10*(6), e0131183. https://doi.org/10.1371/journal.pone.0131183
[44] Xie, W., Du, Z., Chen, Y., Liu, N., Zhong, Z., Shen, Y., & Tang, L. (2020). Identification of Metastasis-Associated Genes in Triple-Negative Breast Cancer Using Weighted Gene Co-expression Network Analysis. *Evolutionary Bioinformatics*, *16*, 117693432095486. https://doi.org/10.1177/1176934320954868
[45] Yi, J., Zhong, W., Wu, H., Feng, J., Zouxu, X., Huang, X., Li, S., & Shuang, Z. (2021). Identification of Key Genes Affecting the Tumor Microenvironment and Prognosis of Triple-Negative Breast Cancer. *Frontiers in Oncology*, *11*, 746058. https://doi.org/10.3389/fonc.2021.746058
[46] Yin, J., Lin, C., Jiang, M., Tang, X., Xie, D., Chen, J., & Ke, R. (2021). CENPL, ISG20L2, LSM4, MRPL3 are four novel hub genes and may serve as diagnostic and prognostic markers in breast cancer. *Scientific Reports*, *11*(1), 15610. https://doi.org/10.1038/s41598-021-95068-6
[47] Yin, L., Duan, J.-J., Bian, X.-W., & Yu, S.-C. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research: BCR*, *22*(1), 61. https://doi.org/10.1186/s13058-020-01296-5
[48] Zhai, Q., Li, H., Sun, L., Yuan, Y., & Wang, X. (2019). Identification of differentially expressed genes between triple and non-triple-negative breast cancer using bioinformatics analysis. *Breast Cancer*, *26*(6), 784–791. https://doi.org/10.1007/s12282-019-00988-x
[49] Zhai, X., Yang, Z., Liu, X., Dong, Z., & Zhou, D. (2020). Identification of NUF2 and FAM83D as potential biomarkers in triple-negative breast cancer. *PeerJ*, *8*, e9975. https://doi.org/10.7717/peerj.9975
[50] Zhao, B., Xu, Y., Zhao, Y., Shen, S., & Sun, Q. (2020). Identification of Potential Key Genes Associated With the Pathogenesis, Metastasis, and Prognosis of Triple-Negative Breast Cancer on the Basis of Integrated Bioinformatics Analysis. *Frontiers in Oncology*, *10*, 856. https://doi.org/10.3389/fonc.2020.00856
[51] Zhong, G., Lou, W., Shen, Q., Yu, K., & Zheng, Y. (2020). Identification of key genes as potential biomarkers for triple-negative breast cancer using integrating genomics analysis. *Molecular Medicine Reports*, *21*(2), 557–566. https://doi.org/10.3892/mmr.2019.10867