Full Length Research Article

UMAP-Based Transcriptional Profiling and Differential Expression Analysis of Breast Cancer Datasets GSE65194 and GSE42568

https://doi.org/10.62940/als.v13i1.3885

Issue: Volume 13, Issue 1

Received: 11-04-2025

Revised: 12-05-2025

Accepted: 05-12-2025

Published online: 31-03-2026

Keywords: Gene Expression Omnibus, Breast Cancer, Differential Expression Analysis, Transcriptomics, UMAP, Biomarkers

Jawaher Almulhim[1],*

1. Department of Biological Sciences, King Faisal University, Alahsa 31982, Saudi Arabia

* jalmulhim@kfu.edu.sa

ABSTRACT

Background: Breast cancer remains a major global health alarm. By analyzing transcriptomic datasets GSE65194 & GSE42568, we identified quite a lot of consistently dysregulated genes that may impact cancer progression & serve as potential therapeutic targets.

Methods: With the GEO2R for a differential gene expression study, we identified many weighty genes ($p < 0.05$, absolute log2 fold change > 1), as well as both favorably upregulated & downregulated applicants. Fundamental visualizations contained within volcano plots, MA plots, UMAP for dimensionality reduction, & box plots. With the help of Venn diagram to validate overlapping differentially expressed genes transversely datasets, confirming sturdy shared expression patterns.

Results: Breast cancer gene expression datasets GSE65194 & GSE42568, which both equivalence cancerous breast tissue to normal controls, identified 5554 & 2957 differentially expressed genes, correspondingly. Definitely, in GSE65194, genes were upregulated (4968), & were downregulated (586), whereas in GSE42568, genes were upregulated (1512) & were downregulated (1445). In particular, among the genes in GSE65194, COL11A1 showed the greatest expression change, with a log2 fold change of 7.69. UMAP study of both datasets evidently separated cancerous & normal samples, prominence different gene expression profiles.

Conclusion: More than a few genes are dependably upregulated across breast cancer datasets, portentous shared disease-related pathways & latent value as biomarkers or therapeutic targets, though clinical authentication is still needed.

INTRODUCTION

Breast cancer is the most prevalent and commonly diagnosed cancer in the world among women and it remains a significant weight in the health and health care systems of the population [1,2]. This highlights a continuous rise in its effects as the number of new cases and almost 790,000 deaths were reported worldwide in 2022 alone [3]. It is a complicated disease, and the number of molecular subtypes and options of clinical patterns makes it relevant to keep investigating the mechanisms of the disease [4,5]. The techniques of sequencing (High-throughput) have been able to improve cancer research by facilitating the production of extensive genomic and transcriptomic data. Such datasets have given researchers an improved insight into the biology of breast cancer [6]. The first source is the GEO (Gene Expression Omnibus), a large community database under NCBI (National Center for Biotechnology Information) in which many different datasets of gene expression are available to breast cancer research [7]. These datasets are utilized by scholars to discover genetic alterations and changes in gene expression associated with the initiation and development of breast cancer [8,9]. Full expression reporting of the genes has assisted researchers in defining the major molecular subtypes of breast cancer [10,11]. The use of categories (luminal A, luminal B, HER2-enriched, and basal-like) has been important in applying multiple administrative treatment decisions to patients and enhancing patient outcomes [12]. The TNBC (Triple-negative breast cancer), which is currently being viewed as a separate and aggressive type of the cancer, has also emerged as a significant field of study [13]. GEO datasets together is a convenient style for identifying genes that are consistently differentially expressed athwart patient groups [14,15]. This approach benefits overcome some of the limits of single studies & make known strong molecular monickers linked to breast cancer [16]. Meta-evaluates of these datasets have also identified most important biological pathways intricate in cancer progression, principally those connected to the cell cycle, DNA repair, & the immune response [17,18].

The new trends in the field of insilico biology have made a significant contribution to the knowledge on breast cancer. The present advancement in in silico research has critically corrected our knowledge about breast cancer. Machine learning has allowed the discovery of novel biomarkers (prognostic) & imaginable therapeutic targets [19,20], & network-based educations have irradiated the complicated relationships among genes & their regulation [21,22]. Transcriptomic based research has also guided that non-coding RNAs, including the microRNAs and long non-coding RNAs, have a crucial role in tumor development [23], and new applications in diagnosis and treatment [24]. This has been complemented with transcriptomic data with other forms of molecular data, such as methylation (DNA) and protein biology, which has further increased our sympathetic of breast cancer complexity [25]. Recent studies have been keen on the need to discover gene markers that anticipate the response of the patient to therapy and their prognosis, which can be used to establish individualized medicine and immunotherapy [26-28]. It is on the basis of this slant that our study consolidates various GEO datasets and spreads them on sound analytical tools to determine common patterns of gene expression and to determine predictable biomarkers and therapeutic targets in breast cancer.

METHODS

Data Retrieval &Acquisition

In the study, we compared two well-characterized datasets (breast cancer) of gene expression data in the GEO (GSE65194 and GSE42568). These datasets were selected with great care due to their sustainable experimental design, comprehensive profiling, and comprehensive clinical information.

GSE65194, engendered employing the Affymetrix Human Genome U133 + 2.0 Array platform, consist of breast cancer tissue samples (130) & normal (11). To perform the current analysis, the differential gene expression was checked by performing a comparison between the tumor samples and normal breast tissue. This data is closely related to research by Maubant et al. and Maire et al. [29-33] and is largely relevant to TNBC, and a strong emphasis on characterization and potential therapeutic targets. Tissue samples were of high quality, and RIN (RNA integrity numbers) was greater than 7.0. The dataset also gets a comprehensive analysis of pathways & targets of interest, together with Wnt3a signaling, TTK/hMPS1, & Polo-like kinase 1. Similarly, GSE42568, which was reported by Clarke C et al. [34], contains a breast cancer sample (104) and normal (17) samples that were processed using the Affymetrix Human Genome U133 + 2.0 Array platform. This data gives comprehensive clinical information such as ER (estrogen receptor), PR (Progesterone receptor), and HER2, and tumor grade and survival outcomes. It

implies that it has common subtypes of breast cancer and that it was treated using the same behavior to minimize variation and ensure consistency, which makes it very appropriate to do comparative analysis.

On their high quality, clinical annotation, and general representation of imperative breast cancer subtypes, we have chosen data set GSE65194 & GSE42568. GSE65194 suggestions a focused view of TNBC, supported thru high RNA quality & detailed molecular data. In contrast, GSE42568 provides a larger & more heterogeneous cohort with clearly defined ER, PR, & HER2 status, allowing subtype-oriented analysis. Together, these datasets provide consistent &well-annotated data for an integrative study of breast cancer.

To ensure data quality &comparability, both datasets were carefully preprocessed before analysis. The preprocessing steps included background correction, normalization, & batch-effect adjustment. Platform-appropriate methods were applied, including quantile normalization for the Illumina data &robust multi-array average (RMA) normalization for the Affymetrix data.

Data Preprocessing & Sample Grouping

Gene expression data from both datasets were carefully evaluated through multiple quality-control steps. In GSE65194, the samples were divided into breast cancer tissues (n = 130) &normal breast tissues (n = 11), whereas GSE42568 included 104 breast cancer samples &17 normal breast tissue samples. To ensure data reliability, we assessed RNA degradation, reviewed intensity distribution plots, &carried out sample correlation analysis to detect possible technical problems or outlier samples.

Preprocessing was performed using methods appropriate for each platform. For GSE65194, which used the Illumina HumanHT-12 V4.0 platform, quantile normalization was applied. For GSE42568, generated on the Affymetrix Human Genome U133 Plus 2.0 platform, normalization was carried out using the Robust Multi- array Average (RMA) method. In both datasets, background correction &batch-effect removal were also performed to minimize non-biological variation. Low-quality probes were excluded, &potential outliers were identified through signal intensity assessment &principal component analysis. Together, these steps helped ensure that the gene expression data used in the study were accurate, consistent, &suitable for downstream analysis. Probe sets were then matched to the most recent gene annotations using platform-specific annotation packages. After preprocessing, samples from GSE65194 &GSE42568 were classified into two main groups: breast cancer &healthy control samples. This grouping was based on the detailed metadata &sample information provided in the GEO database. In GSE42568, the samples were categorized into 104 breast cancer &17 normal breast tissue samples according to the available clinical annotations. GSE65194 as well exemplified a breast cancer expression dataset generated retaining the Array platform (Affymetrix Human Genome U133 Plus 2.0) [31-34]. This clear & reliable grouping on condition that a solid foundation for the ensuing differential gene expression analysis.

Gene Expression Analysis Using GEO2R

A differential gene expression homework was performed by the using GEO2R, a web-based tool as long as thru NCBI that is built on the limma package in R. For the GSE42568 dataset, gene expression was compared between breast cancer samples (104) & normal (17). For GSE65194, expression outlines were in the same way estimated across breast cancer & normal breast tissue samples on behalf of the respective study groups. To acquire reliable & statistically robust results, stern filtering criteria were applied, as well as an adjusted p-value of less than 0.05 using the Benjamini–Hochberg rectification, an absolute log2 fold change greater than 1, & variance approximation based on observed Bayes statistics. The FDR (false discovery rate) was also taken care of in order to achieve multiple-testing correction. GEO2R investigation was applied in the limma (Linear Models of Microarray Data) setting that is quite appropriate to manage data of the management in the form of microarray data produced by the Illumina and Affymetrix systems. In the case of GSE65194, the Illumina HumanHT-12 V4.0 platform, quantile normalization was done followed by a differential expression testing. In the case of GSE42568, RMA (Robust Multi-array Average), normalization of Affymetrix U133 + 2.0 was practical before statistical analysis. Such preprocessing steps were confident according to the known methods in transcriptomic analyses [29,32], and the purpose of optimizing the accuracy of the detection of differential gene expression, decreasing platform-dependent technical error, and minimizing

false-positive results.

Visualization, Correlation Analysis, & Venn Diagram Assessment

GEO2R results of the differentiation of expression were also analyzed and plotted using the R programming with the version 4.1.0. Exported the output files in the form of.tsv format and contained the values of gene expression, estimates of fold-changes, and measures of statistical significance. Volcano plots were created with the ggplot2 package & log2 fold change placed on the x-axis and the negative log10 of the adjusted p-value placed on the y-axis. These plots allowed clear identification of genes that met the predefined importance thresholds. To explore sample clustering based on global gene expression patterns, UMAP was performed using the UMAP package in R. This dimensionality-reduction procedure produced two-dimensional plots that revealed distinct clustering of cancer & normal samples. In addition, box plots were used to display the expression levels of selected genes &to compare their distribution between breast cancer &normal tissue samples.

To determine genes that were consistently differentially expressed in both datasets, namely GSE65194 (130 breast cancer & 11 normal breast tissue samples) &GSE42568 (104 breast cancer & 17 normal breast tissue samples), Venny 2.1.0 was used to construct Venn diagrams. Overlapping genes were then studied for expression concordance, & their character across datasets was evaluated employing Pearson correlation analysis. Concurrently, these analyses provided a clearer view of the shared transcriptional alterations present in both breast cancer datasets.

RESULTS

Data Generation, Grouping, & Parameterization

Gene expression profiles were analyzed using two independent breast cancer datasets got from the Gene Expression Omnibus (GEO) database. The GSE65194 dataset contains 130 breast cancer & 11 normal samples. In the same way, GSE42568 contains 104 breast cancer &17 normal samples & was generated employing the U133+ 2.0 Array platform (Affymetrix Human Genome). The search of differential expression was carried out through GEO2R with challenging statistical requirements. The method (Benjamini-Hochberg) was used to amend multiple testing, and the precision weights of limma were feasible, and dealing with heteroscedasticity. Genes were careful pointedly differentially expressed if the adjusted p-value was below 0.05 & the absolute log2 fold change outdid 1. The datasets (2) identified imperative genes, & their expression forms were visualized through volcano plots & UMAP. To reduce technical variation & expand comparability, quantile normalization was used for the Illumina-based dataset (GSE65194), while Robust Multi-array Average (RMA) normalization was practical to the Affymetrix-based dataset (GSE42568). After standardization, clear transcriptional differences were detected between breast cancer & normal breast tissues.

Differential Gene Expression Analysis of the GSE65194 Breast Cancer Dataset

Analysis of the GSE65194 dataset revealed clear transcriptional differences among breast cancer samples across varying experimental conditions after normalization of the expression data. This was demonstrated by volcano plots &UMAP visualization.

Volcano Plot Analysis

A volcano plot study of the GSE65194 dataset identified 5,554 differentially expressed genes (DEGs) using thresholds of adjusted p-value < 0.05 &| log2 fold change| > 1. Of these, 4,968 genes were upregulated (logFC > 1; shown in red), whereas 586 genes were downregulated (logFC < −1; shown in blue). Among the most strongly upregulated genes, COL11A1 showed the highest fold change (log2FC = 7.69, adjusted $p = 4.39 \times 10^{-19}$), followed by COL10A1 (log2FC = 7.33, adjusted $p = 2.75 \times 10^{-26}$). These collagen-related genes have been widely associated with breast cancer progression (Table 1 & Figure 1A).

Mean-Difference (MA) Plot Analysis

The MA plot shows strong separation of differentially expressed genes from background gene

expression levels. The symmetric data distribution pattern confirmed the presence of true normalization & absence of intensity bias. The plot displayed two distinct gene clusters: upregulated genes, marked in red &positioned above the zero-fold-change line, &downregulated genes, marked in blue &positioned below, showing clear expression- level differences in the evaluated samples. (Figure 1B).

UMAP Analysis

The UMAP dimensionality reduction method generated a visualization showing two distinct sample groups with clear spatial separation. The analysis revealed non- overlapping transcriptional patterns, where cancer samples clustered separately from normal tissue samples. This clear segregation of sample clusters confirmed the presence of unique gene expression signatures between the two biological conditions (Figure 1C & D).

Box Plot Analysis

A box plot of the top 20 differentially expressed genes revealed how their expression varied across different sample types. Essential genes, including COL11A1, COL10A1, CXCL10, & RRM2, were more positively expressed in breast cancer samples than in normal breast tissue. The small overlap in expression between the cancer &normal groups suggests that these genes could be useful as disease biomarkers (Figure 1E).

Statistical Analysis

Statistical support at a long-lasting level was confirmed by the dataset (GSE65194), where 78% of the differentially expressed genes had adjusted p- values of less than 1 x 10 -1 and 45% differentiated genes with log2 fold changes of 2 or more. The different genes are associated with breast cancer progression, indicating the potential usefulness as diagnostic biomarkers and therapeutic targets.

Differential Gene Expression Analysis of GSE42568

Specific Analysis of dataset (GSE42568) revealed ambiguous differences between the gene expression between the breast cancer and the normal breast tissue under the several visualization techniques.

Volcano Plot Analysis

Volcano plot analysis identified 2957 significant DEGs. Among these, 1,512 genes were upregulated &1,445 were downregulated in breast cancer tissue approximated with normal breast tissue. These findings indicate substantial transcriptional alterations associated with breast cancer development (Table 2 &Figure 2A).

Mean-Difference Plot (MA) Analysis

The MA plot exhibited a symmetric trumpet- like distribution, confirming proper data normalization &no intensity bias in the analysis. The plot clearly separated significant differentially expressed genes from the background expression levels. The higher number of upregulated genes (red points) in the upper region compared to downregulated genes (blue points) in the lower region showed a trend towards increased gene expression in the dataset (Figure 2B).

UMAP Analysis

The UMAP analysis of gene expression data demonstrated clear clustering patterns between breast cancer & normal tissue samples. The dimensionality reduction tactic bare two well-defined & distinct groups in the dataset. Breast cancer samples formed one tight cluster that was definitely separated from the cluster of normal tissue samples, representative significant differences in their gene expression signatures. This clear separation in the UMAP plot confirms substantial transcriptional reprogramming that arises during breast cancer development (Figure

2C).

Box Plot Analysis

The box plot of the top 20 differentially expressed genes showed clear differences between breast cancer & normal samples. Several genes, such as KRT18, KRT19, EPPK1, COL11A1, EPCAM, DSP, & ESRP1, were expressed at much higher levels in the breast cancer group. The overlap in interquartile ranges suggests that these genes could serve as diagnostic biomarkers. Their unswerving differential expression advises a role in progression of breast cancer (Figure 2E).

Statistical Analysis

Among 2957 DEGs, 1512 were upregulated & 1445 were downregulated, representative broad transcriptional variations in the cancer (breast)

Upregulated Genes in GSE42568 & GSE65194: A Venn Analysis

Evaluation of GSE42568 & GSE65194 showed shared & dataset-specific upregulated genes in breast cancer, with 4 genes common among the top 18 DEGs (Figure 3).

Four genes were dependably upregulated in both datasets: COL11A1, TOP2A, RRM2, & ESRP1. These are genes that are linked to vital processes in cancer (breast), and extracellular matrix remodeling, cell division, DNA synthesis and splicing regulation. Their chronic deactivation across autonomous groups influences their biological reputation as well as dictates their possible worth as biomarkers and therapeutic targets. The differences between the datasets probably recreate difference in patient characteristics, tumor subtype, sample processing, and platform. General, integrating both datasets ideal parts shared and unique molecular characteristics of cancer (breast).

Figures

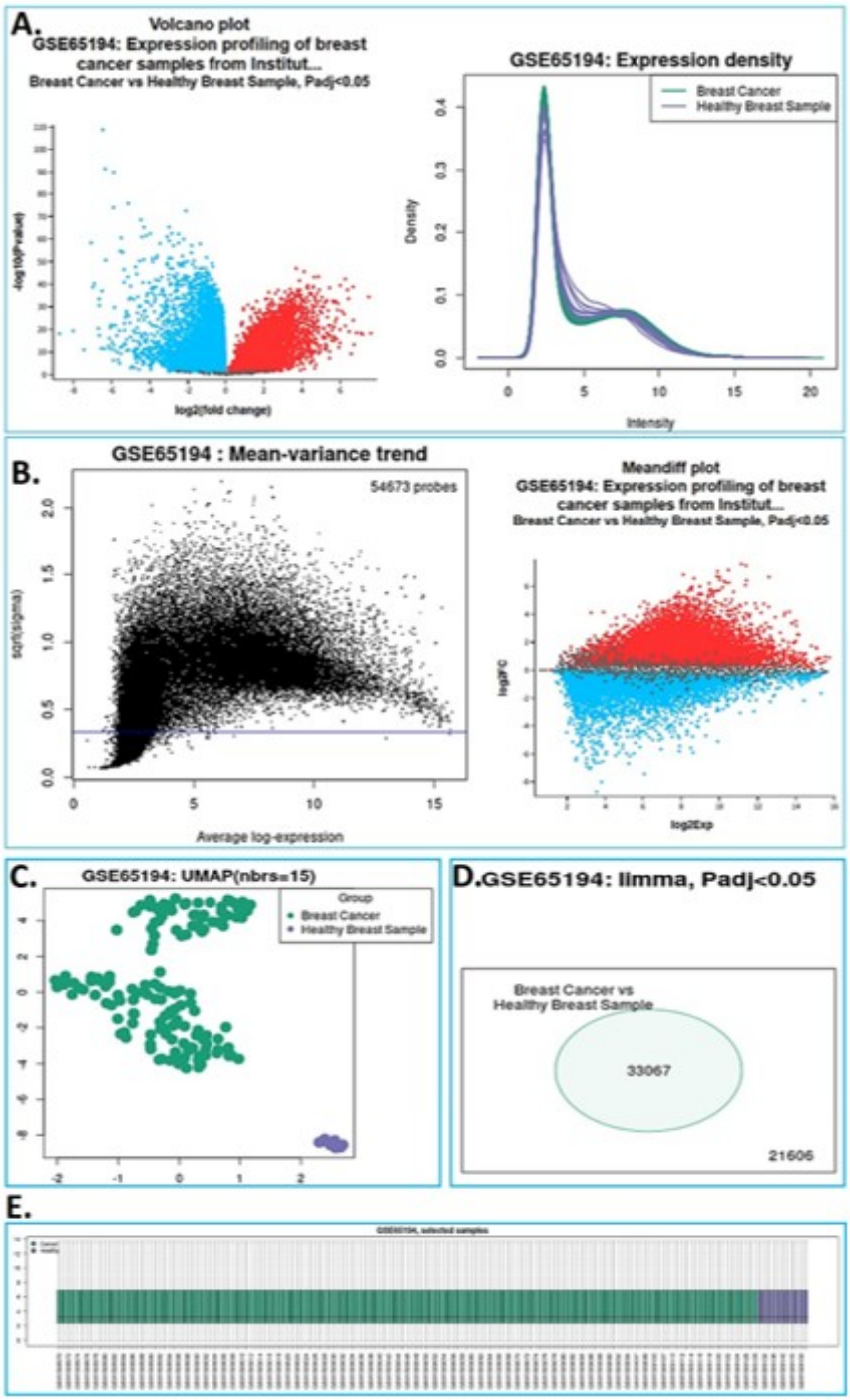

Figure 1: (A)Volcano plot & Expression Density analysis of gene expression profiling dataset of GSE65194. (B) Mean-variance trend & Meandiff plot analysis of gene expression profiling dataset of GSE65194.(C) UMAP analysis of gene expression profiling dataset of GSE65194 (D) Cluster analysis of gene expression profiling dataset of GSE65194(E) Boxplot analysis of gene expression profiling dataset of GSE65194.

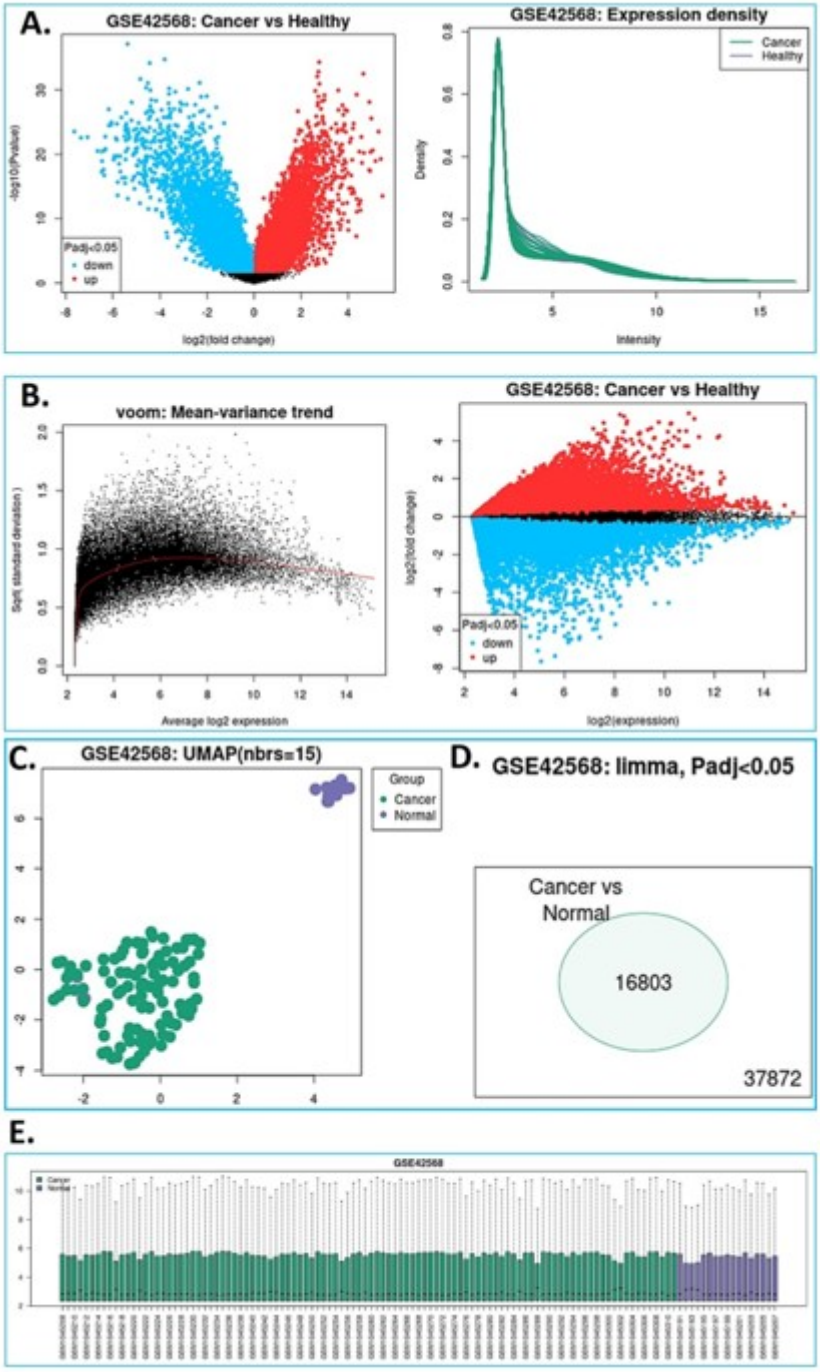

Figure 2: (A) Volcano plot & Expression Density analysis of gene expression profiling dataset of GSE42568 (B) Mean-variance trend & Meandiff plot analysis of gene expression profiling dataset of GSE42568. (C) UMAP analysis of gene expression profiling dataset of GSE42568(D) Cluster analysis of gene expression profiling dataset of GSE42568 (E) Boxplot analysis of gene expression profiling dataset of GSE42568.

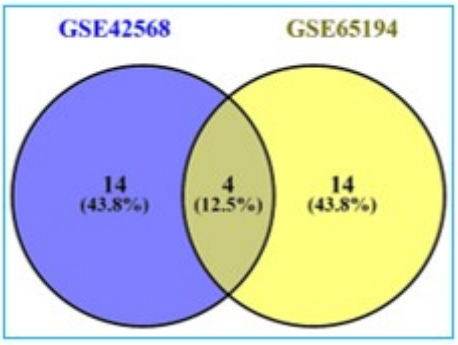

Figure 3: Venn study identified shared (4) & unique (14) upregulated genes in each dataset.

Tables

S.No.	adj.P. Val	logFC	Gene symbol	Genetitle
1.	4.39E-19	7.69	COL11A1	Collagen type XI alpha 1 chain
2.	2.75E-26	7.33	COL10A1	Collagen type X alpha 1 chain
3.	1.61E-27	6.93	CXCL10	C-X-C motif chemokine ligand 10
4.	1.31E-82	6.81	RRM2	Ribonucleotide reductase regulatory subunit M2
5.	4.69E-32	6.71	VCAN	Versican
6.	3.14E-49	6.66	S100P	S100 calcium binding protein P
7.	1.15E-24	6.58	TOP2A	Topoisomerase (DNA) II alpha
8.	4.15E-40	6.51	TPD52	Tumor protein D52
9.	4.26E-43	6.35	KIAA0101	KIAA0101
10.	1.24E-25	6.16	ACTB	Actin beta
11.	1.46E-43	6.08	COL12A1	Collagen type XII alpha 1 chain
12.	5.57E-82	6.01	SMC4	Structural maintenance of chromosomes 4
13.	6.57E-49	6.01	ANP32E	Acidic nuclear phosphoprotein 32 family member E
14.	2.33E-22	6	HSP90AB1	Heat shock protein 90 alpha family class B member 1
15.	3.45E-36	5.98	MIR3620///ARF1	MicroRNA 3620///ADP ribosylation factor 1
16.	2.64E-31	5.94	INHBA	Inhibin beta A subunit
17.	6.89E-64	5.88	ATAD2	ATPase family, AAA domain containing 2
18.	5.25E-37	5.87	ESRP1	Epithelial splicing regulatory protein 1

Table 1: Top upregulated genes in breast cancer identified in GSE65194 dataset with (log2FC > 1, adjusted p < 0.05). The table lists gene symbols, names, log2 fold change values &adjusted p-values. The upregulated genes are arranged by descending fold change values. These upregulated genes represent potential therapeutic targets &diagnostic markers in breast cancer.

S.No.	adj.P.Val.	logFC	Gene symbol	Gene title
1.	9.48E-15	5.45	KRT19	Keratin 19
2.	5.90E-13	5.38	COL11A1	Collagen type XI alpha 1 chain
3.	1.26E-17	5.28	EPPK1	Epiplakin 1
4.	3.87E-20	5.16	EPCAM	Epithelial cell adhesion molecule
5.	1.12E-24	4.98	DSP	Desmoplakin
6.	8.18E-18	4.92	KRT18	Keratin 18
7.	1.47E-23	4.89	ESRP1	Epithelial splicing regulatory protein 1
8.	9.24E-18	4.89	CDH1	Cadherin 1
9.	3.13E-19	4.66	TFAP2A	Transcription factor AP-2 alpha
10.	4.55E-21	4.64	SDC1	Syndecan 1
11.	1.45E-17	4.58	ERBB3	Erb-b2 receptor tyrosine kinase 3
12.	6.55E-17	4.55	TOP2A	Topoisomerase (DNA) II alpha
13.	1.42E-13	4.51	TACSTD2	Tumor-associated calcium signal transducer 2
14.	2.21E-12	4.51	MUC1	Mucin 1, cell surface associated
15.	3.37E-17	4.47	RRM2	Ribonucleotide reductase regulatory subunit M2
16.	6.06E-13	4.47	CD24	CD24 molecule
17.	4.77E-11	4.44	GATA3	GATA binding protein 3
18.	4.58E-10	4.36	C4B_2///C4B/// C4A	Complement component 4B (Chido blood group), copy 2/// component 4A (Rodgers blood group)

Table 2: Top upregulated genes in breast cancer identified in the GSE42568 dataset with (log2FC > 1, adjusted

p < 0.05). The table lists gene symbols, names, log2 fold change values &adjusted p-values. The upregulated genes are arranged by descending fold change values. These upregulated genes represent potential therapeutic targets &diagnostic markers in breast cancer.

DISCUSSION

Breast cancer is an international health concern due to the complicated molecular pathways that facilitate the development of progression and resistance to therapy. The GSE42568 and GSE65194 data sets in the present researches identified four genes, which were consistently over- expressed in tumor samples relative to normal breast tissue, namely, COL11A1, TOP2A, RRM2, and ESRP1[35,36]. Their expression in distinct cohorts is reproducible and this is a key aspect in the biology of breast cancer. The most significantly upregulated of them was COL11A1 (log2 fold change = 7.69) and is highly associated with extracellular matrix remodeling, tumor invasion, and an inflammatory microenvironment, which is why it has the potential to become a biomarker and a treatment target [37,38].

TOP2A, RRM2, and ESRP1 showed high upregulations thus they contribute to the development of breast cancer. TOP2A is linked to cell growth and replication of DNA and can assist in predicting the response of patients to anthracycline therapy. RRM2 is involved in the DNA synthesis and monitoring the cell cycle, and the high concentrations may indicate chemotherapy resistance. ESRP1 cares different splicing & retention of epithelial characteristics, which types it an chief factor in tumor growth & therapeutic conclusions. Such findings suggest that all the three genes may be of great clinical use in breast cancer [39-42].

The study identified that tumors with increased TOP2A and RRM2 expression can be discarded to be more responsive to neoadjuvant chemotherapy based on anthracycline. On the other hand, COL11A1 may be a marker of chemotherapy resistance. Pathway analysis showed that these genes are implicated in DNA synthesis, DNA damage response, tumor microenvironment changes and splicing regulation, which can all potentially affect tumor response to treatment. To ensure that the influences were more accurate, data giving out, amendment of batch developments, normalization, multiple- testing correction and validation with independent datasets were taken into consideration. The level core gene signatures recommend that though they might work as diagnostic biomarkers and therapeutic targets, differences between datasets are the best parting of the diversity of breast cancer [43-48].

Despite the fact this study has quite a lot of strong point, it also has approximately limitations. It is based on transcriptomic data, and can be pretentious through platform-specific bias, and fails to provide a lot of data on post- transcriptional or protein- level changes. Multi- omics data, functional validation intake, and single- cell analysis should be utilized and applied in about to happen explore. Nevertheless, the gene signatures, which are generated here, may still be used in selecting treatment, intensives care reactions, and development of targeted therapies [49-54].

The functions of these genes, their interactions along the paths, and their roles in treatment resistance should be clarified in imminent research. It must also support the validation of biomarkers, targeted therapy, clinical trials, and corrected data integration through the innovative investigative processes [55-60].

This research is essential in determining extensive transcriptional differences and strong gene signatures in companionship with treatment response and survivability. By combining several datasets, this analytical method can eliminate certain weaknesses of single-cohort studies and provide a more in-depth picture of breast cancer biology [61,62].

To invention molecular signatures related to disease progression, we tested two independent datasets of breast cancer, GSE65194 and GSE42568. In GSE65194, our study of the differential expression identified 5554 DEGs and in GSE42568, 2957. Strict statistical norms were observed in all DEGs, and the adjusted p-value was less than 0.05 and absolute log2 fold change more than 1. The four genes COL11A1, TOP2A, RRM2 and ESRP1 were consistently upregulated in both datasets which augur well with the possibility that they are at the center stage of breast cancer. COL11A1 plays a complex role in extra cellular matrix remodeling. TOP2A is essential in replication of DNA. RRM2 the stage plays a role in nucleotide metabolism. ESRP1 controls other splicing, which interferes with the process of gene joining to mRNAs. The findings broaden our knowledge on the molecular pathogenesis of breast cancer and ideal part auspicious

candidates of biomarkers and novel therapies. Further reading of these activated genes can prepare to even greater individualized and ongoing therapies of breast cancer.

CONFLICT OF INTEREST

The author state that there is no conflict of interesting pertaining to this research / research publication.

ACKNOWLEDGMENT

I would like to thank King Faisal University &the Department of Biological Sciences for their funding support of this work.

Generative AI Statement

The author declares that Generative AI tools were used to enhance the language clarity of this work. The author takes full responsibility for the accuracy and integrity of the content.

REFERENCES

1. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, et al. Breast cancer. Nature Reviews Disease Primers, (2019); 5(1): 66.
2. Ogier du Terrail J, Leopold A, Joly C, Béguier C, Andreux M, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature Medicine, (2023); 29(1): 135-146.
3. Karimi E, Yu MW, Maritan SM, Perus LJM, Rezanejad M, et al. Single-cell spatial immune landscapes of primary &metastatic brain tumours. Nature, (2023); 614, 555-563.
4. Maubant S, Tesson B, Maire V, Ye M, Rigaill G, et al. Transcriptome analysis of Wnt3a-treated triple-negative breast cancer cells. PLoS One. (2015); 10(4):e0122333.
5. Maire V, Baldeyron C, Richardson M, Tesson B, Vincent-Salomon A, et al. TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PLoS One, (2013); 8(5):e63712.
6. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, et al. Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. Carcinogenesis, (2013); 34(10):2300-2308.
7. Turner NC, Kingston B, Kilburn LS, Kernaghan S, Wardley AM, et al. Circulating tumor DNA analysis for early relapse detection in high-risk early breast cancer. Journal of Clinical Oncology, (2023);41(3):558-569.
8. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, et al. Atezolizumab &nab-paclitaxel in advanced triple-negative breast cancer. The New Engl&Journal of Medicine, (2020); 382(22):2108-2120.
9. Bardia A, Hurvitz SA, Tolaney SM, Loirat D, Punie K, et al. Sacituzumab govitecan in metastatic triple-negative breast cancer. The New Engl&Journal of Medicine, (2021); 384(16):1529-1541.
10. Cortés J, Kim SB, Chung WP, Im SA, Park YH, et al. Trastuzumab deruxtecan versus trastuzumab emtansine for breast cancer. The New Engl&Journal of Medicine, (2022); 386(12):1143-1154.
11. Schettini F, Conte B, Pesantez D, Sihota A, Mele D, Di et al. Clinical, pathological, &genomic features of HER2-low breast cancer. NPJ Breast Cancer, (2021); 7(1):1-15.
12. Waks AG, Winer EP.Breast cancer treatment: a review. JAMA. (2019); 321(3):288-300.
13. Garrido-Castro AC, Lin NU, Polyak K. Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment. Cancer Discovery, (2019); 9(2):176-198.
14. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, et al. A single-cell &spatially resolved atlas of human breast cancers. Nature Genetics, (2021); 53(9):1334-1347.
15. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. Lancet, (2021); 397(10286):1750-1769.
16. André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, et al.Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. The New Engl&Journal of Medicine, (2019); 380(20):1929-1940.
17. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, et al.The genomic landscape of endocrine-resistant advanced breast cancers. Cancer Cell, (2020); 38(3):485-497.
18. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, et al.Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science, (2018); 362(6411):eaar3593.
19. Naik N, Madani A, Esteva A, Keskar NS, Press MF, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. Nature Communications, (2020);11(1):5727.
20. Gillanders WE, Mikhitarian K, Hebert R, Mauldin PD, Palesch Y, et al. Molecular detection of micrometastatic breast cancer in histopathology-negative axillary lymph nodes correlates with traditional predictors of prognosis. Annals of Surgery, (2019); 239(6):828-840.
21. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nature Communications, (2020);11(1):3877.
22. Keren L, Bosse M, Marquez D, Angoshtari R, Jain S, et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. Cell, (2018); 174(6):1373-1387.
23. Courtiol P, Tramel EW, Sanselme M, Wainrib G. Classification &disease localization in histopathology using only global labels: A weakly-supervised approach. Nature Communications, (2019); 10(1):1-12.

24. Jang BS, Han W, Kim SW, Shin HJ, Park S, et al. Artificial intelligence-based prediction of response to neoadjuvant chemotherapy in breast cancer. Nature Communications, (2023); 14(1):517.
25. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. The New Engl&Journal of Medicine, (2022); 387(1):9-20.
26. Rugo HS, Cortés J, Cescon DW, Im SA, Yusof MM, et al. Elacestrant for advanced breast cancer with ESR1 mutation. The New Engl&Journal of Medicine, (2022); 387(7):592-604.
27. Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. The EMBO Journal, (2022); 41(5):e108375.
28. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity, (2021); 54(3):702-722.
29. Maubant, S., Tesson, B., Maire, V., Ye, M., Rigaill, G., Gentien, D., Cruzalegui, F., Tucker, G. C., Roman-Roman, S., & Dubois, T. (2015). Transcriptome analysis of Wnt3a-treated triple-negative breast cancer cells.
30. Maire, V., Baldeyron, C., Richardson, M., Tesson, B., Vincent-Salomon, A., Gravier, E., Marty-Prouvost, B., De Koning, L., Rigaill, G., Dumont, A., Gentien, D., Barillot, E., Roman-Roman, S., Depil, S., Cruzalegui, F., Pierré, A., Tucker, G. C., & Dubois, T. (2013). TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer.
31. Pusztai L, Foldi J, Dhawan A, DiGiovanna MP, Mamounas EP. Changing frameworks in treatment sequencing of triple-negative &HER2-positive, early-stage breast cancers. Lancet Oncology, (2022); 23(7):e316-e323.
32. Jackson HW, Fischer JR, Zanotelli V, Ali HR, MecheraR,etal.The single-cell pathology landscape of breast cancer. Nature, (2021); 578(7796):615-620.
33. Schmid P, Cortés J, Dent R, Pusztai L, McArthur H, et al. Event-free survival with pembrolizumab in early triple-negative breast cancer. The New Engl&Journal of Medicine, (2022); 386(6):556-567.
34. Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, et al. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. The New Engl&Journal of Medicine, (2021); 384(25):2394-2405.
35. Goel S, DeCristo MJ, McAllister SS, Zhao JJ.CDK4/6 inhibition in cancer: Beyond cell cycle arrest. Trends Cell Biology, (2022); 32(3):228-244.
36. Pascual J, Turner NC, BaselgaJ.Targeting the PI3K/AKT/mTOR pathway in triple-negative breast cancer. Cancer Discovery, (2022); 12(11):2560-2578.
37. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, et al. Tumour-infiltrating lymphocytes &prognosis in different subtypes of breast cancer. Lancet Oncology, (2021); 22(5):692-701.
38. Emens LA, Adams S, Barrios CH, Diéras V, Iwata H, et al. First-line atezolizumab plus nab-paclitaxel for unresectable, locally advanced, or metastatic triple-negative breast cancer: IMpassion130 final overall survival analysis. Annals of Oncology, (2022); 33(3):321-331.
39. Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, et al.Meta-analysis of tumor- &T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. Cell, (2021); 184(3):596-614.
40. Rugo HS, Lerebours F, Ciruelos E, Drullinsky P, Ruiz-Borrego M, et al. Alpelisib plus fulvestrant in PIK3CA-mutated, hormone receptor-positive advanced breast cancer after a CDK4/6 inhibitor (BYLieve): one cohort of a phase 2, multicentre, open-label, non-comparative study. Lancet Oncology, (2021); 22(4):489-498.
41. Mosele F, Remon J, Mateo J, Westphalen CB, Barlesi F, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. Annals of Oncology, (2022); 33(11):1111-1133.
42. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. The New Engl&Journal of Medicine, (2021); 384(13):1222-1234.
43. André F, Marmé F, Cortes J, Gonçalves A, et al. Trastuzumab deruxtecan as first-line treatment for HER2-positive metastatic breast cancer. Nature Medicine, (2023); 29(3):657-664.
44. Stover DG, Gil Del Alcazar CR, Brock J, Guo H, Overmoyer B, et al. Phase II study of pembrolizumab &nab-paclitaxel in HER2-negative metastatic breast cancer. Nature Communications, (2022); 13(1):2352.
45. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nature Medicine, (2021); 27(9):1563-1572.
46. Jiang YZ, Ma D, Suo C, Shi J, Xue M, et al. Genomic &transcriptomic landscape of triple-negative breast cancers: subtypes &treatment strategies. Cancer Cell, (2022); 40(11):1429-1446.
47. Migliaccio I, Carmona-Sáez P, Chronic G, Pérez-García J, Izarzugaza JMG, et al. Artificial intelligence &machine learning in breast cancer research. Nature Reviews Cancer, (2023); 23(5):283-297.
48. Spring LM, Wander SA, Andre F, Moy B, Turner NC, Cyclin-dependent kinase 4/6 inhibitors in breast cancer: Current status &future directions. Nature Reviews Clinical Oncology, (2022); 19(9):585-599.
49. Lambertini M, Franzoi MA, Pondé N, Bruzzone M, Peccatori FA, et al. Pregnancy after breast cancer: An update on current evidence. Journal of Clinical Oncology, (2022); 40(23):2683-2695.
50. Turner NC, Swift C, Kilburn L, Fribbens C, Beaney M, et al.Circulating tumor DNA analysis &longitudinal monitoring in metastatic breast cancer. Nature Medicine, (2023); 29(3):665-675.
51. Gupta S, Nanda R, Bose R, Anders CK, Kaklamani VG, et al. Novel therapeutic strategies in triple-negative breast cancer. Nature Reviews Clinical Oncology, (2023); 20(2):121-135. doi:10.1038/s41571-022-00686-2.
52. Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer. European Journal of Cancer, (2022); 159:13-27.
53. Hurvitz SA, Tolaney SM, Punie K, Bardia A, Dirix LY, et al. Biomarker analyses in the phase III ASCENT

study of sacituzumabgovitecan versus chemotherapy in metastatic triple-negative breast cancer. Nature Medicine, (2023); 29(2):438-446.
54. Cardoso F, Paluch-Shimon S, Senkus E, Curigliano G, Aapro MS, et al. 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). Annals of Oncology, (2023); 34(1):12-33.
55. Franzoi MA, Romano E, Piccart M. Immunotherapy for early-stage triple-negative breast cancer: Current status &future perspectives. Cancer Cell, (2023); 41(1):20-32.
56. Bidard FC, Kaklamani VG, Neven P, Streich G, Mountzios G, et al. Elacestrant versus standard endocrine therapy in patients with ESR1-mutated advanced breast cancer: Results from the randomized phase 3 EMERALD trial. Journal of Clinical Oncology, (2023); 41(7):1284-1292.
57. Symmans WF, Wei C, Gould R, Yu X, Zhang Y, et al. Long-term outcomes for residual cancer burden after neoadjuvant chemotherapy in breast cancer. Journal of Clinical Oncology, (2023); 41(2):235-246.
58. Jerusalem G, Park YH, Yamashita T, Hurvitz SA, Chen S, et al. Trastuzumab deruxtecan in HER2-low metastatic breast cancer: The DAISY trial. Nature Medicine, (2023); 29(5):1151-1159.
59. Griguolo G, Pascual T, Dieci MV, Guarneri V, Prat A. Interaction of HER2 with other biomarkers as a prognostic tool in early-stage breast cancer: A systematic review &meta-analysis. Lancet Oncology, (2023); 24(4):382-392.
60. Yeo B, Turner NC, Jones A, Caramia F, Schiavon G, et al. Circulating tumor DNA analysis for monitoring treatment response in metastatic breast cancer. Science Translational Medicine, (2023); 15(687): eabq4921.
61. Venet D, Fimereli D, Rothé F, Boeckx B, de Wind A, et al. Single-cell genomics reveals distinct evolutionary patterns of breast cancer metastasis. Nature Genetics, (2023); 55(4):598-609.
62. Loi S, Dushyanthen S, Beavis PA, Salgado R, Denkert C, et al. Clinical implications of breast cancer immunology. Journal of Clinical Oncology, (2023); 41(7):1295-1306.