Int. J. Pharm. Investigation, 2026; 16(2):625-637.
https://www.jpionline.org

Original Article

DAVID and KEGG-Based Functional Enrichment Analysis of Breast Cancer Transcriptome: Mapping ECM Remodeling and Cell Division Pathways

Akram Ahmed Aloqbi*

Department of Biological Science, Faculty of Science, University of Jeddah, Jeddah, SAUDI ARABIA.

ABSTRACT

Background: Breast cancer, affecting roughly 2.3 million people each year, stands as a major global health concern due to its molecular complexity across subtypes such as luminal, HER2-enriched, and triple-negative. These subtypes lead to diverse clinical outcomes, highlighting the urgent need for deeper molecular understanding of tumor development. Advanced genomic tools, including microarrays and RNA-sequencing, have significantly improved our ability to pinpoint genes involved in cancer, but challenges in integrating complex datasets and applying these insights to patient care persist. This study investigates molecular mechanisms to guide the development of innovative treatment approaches. **Materials and Methods:** Using the GSE54002 dataset (216 samples: 200 breast cancer, 16 normal) from the Affymetrix Human Genome U133 Plus 2.0 Array, we applied GEO2R to identify significant genes (adjusted *p*-value <0.05, |log2FC|>1). Functional enrichment with DAVID and KEGG elucidated biological processes and pathways, focusing on cell proliferation, Extracellular Matrix (ECM) remodeling, and cytoskeletal dynamics. **Results:** GEO2R analysis identified genes distinguishing breast cancer from normal tissue. Enrichment studies with DAVID and KEGG revealed key processes and pathways involved in tumor progression, in consistent with established breast cancer etiology. These findings highlighted potential molecular targets for possible therapeutic interventions. **Conclusion:** The current study prominently demonstrates the utility of GEO2R, DAVID, and KEGG tools in understanding and interpreting the complex transcriptomic data and to explore prominent breast cancer's molecular markers. The results also help us in improving the diagnostic and personalized treatment approaches. Further validation studies of current findings, could also help us to develop more precise treatments and improved patient outcomes in cancer care.

Keywords: Affymetrix Chips, Functional Enrichment Analysis, DAVID, KEGG, Breast Cancer, GEO2R.

Correspondence:
Dr. Akram Ahmed Aloqbi
Department of Biological Science, Faculty of Science, University of Jeddah, Jeddah-21589, SAUDI ARABIA.
Email: aaaloqbi@uj.edu.sa

Received: 09-05-2025;
Revised: 28-07-2025;
Accepted: 12-09-2025.

INTRODUCTION

Breast cancer, with approximately 2.3 million new cases diagnosed annually, makes it the most prevalent cancer among women and a leading cause of cancer-related mortality worldwide (Sung *et al*., 2021). Its molecular complexity, which is quite evident in its subtypes such as luminal, HER2-enriched, and triple-negative, results in varied clinical outcomes and treatment responses. This necessitates detailed molecular studies in breast cancer phenomenon to single out specific mechanisms and pathways driving tumor growth and metastasis (Ignatiadis *et al*., 2013; Sorlie *et al*., 2003). Recent advances in genomic technologies like microarrays and RNA-sequencing have though greatly deepened our understanding of breast cancer by helping identify Differentially Expressed Genes (DEGs) involved in key biological processes (Polyak *et al*., 2011; Wang *et al*., 2009). Despite this progress, researchers still face huge challenges owing to the complexity of integrating large datasets and translating its discoveries into actual real time clinical outcomes (Cancer Genome Atlas Network., 2012). In this study, we explore the GSE54002 dataset obtained using the Affymetrix Human Genome U133 Plus 2.0 Array, which includes 216 samples both from normal and breast cancer tissues. The analysis tends to ascertain and examine vital processes like cell proliferation, Extracellular Matrix (ECM) remodeling, and changes in the cytoskeleton (Barrett *et al*., 2013). Through tools such as GEO2R, DAVID, and KEGG, this work tries to discover new molecular markers or molecular interaction patterns, paving the way for new treatment possibilities against breast cancer.

Breast cancer research has made prominent developments in recent years. Large-scale genomic studies has helped identify

DOI: 10.5530/ijpi.20260075

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

molecular subtypes in breast cancer phenomenon and assisted in relating these subtypes to clinical outcomes like metastasis risk and treatment responses (Curtis *et al*., 2012; Prat *et al*., 2011). These investigations have provided key insights on critical processes, such as uncontrolled cell growth, changes in the Extracellular Matrix (ECM), and cytoskeletal reorganization, in cancer progression (Hanahan *et al*., 2011; Kennecke *et al*., 2010). However, there are still some significant challenges in this process. For instance, combining huge transcriptomic, proteomic, and epigenomic data into considerable meaningful data requires advanced computational tools and infrastructure setup (Berger *et al*., 2018). Similarly, many studies, especially those relying on public datasets, also lack detailed clinical information, which restricts their applicability in real-world (Sorlie *et al*., 2001). Also, the microarray technology doesn't offer the same sensitivity as RNA sequencing methodology, so subtle gene expression changes might go unnoticed which though may have proved to be very important otherwise (Zhao *et al*., 2014). Similarly, findings from bioinformatics research often require time consuming and expensive experimental validation set up to confirm their significance (Vandesompele *et al*., 2002). These challenges press for the need of more user friendly and dependable tools that can not only interpret and analyze the genomic data but also help to bridge the gap between the research outcomes and clinical practice (Andre *et al*., 2019; Weigelt *et al*., 2012).

The Gene Expression Omnibus (GEO), managed by NCBI, comes in the forefront for breast cancer research as it offers access to extensive transcriptomic datasets (Edgar *et al*., 2002; Clough *et al*., 2016). As in our study, a GSE54002 dataset available at GEO enables researchers to compare gene expression between cancerous and normal tissues, with the GEO2R tool facilitating the identification of Differentially Expressed Genes (DEGs) amongst them. Built on the limma package in R, GEO2R tool employs statistical methods such as the Benjamini-Hochberg correction to highlight significant DEGs for research purposes (Ritchie *et al*., 2015; Smyth *et al*., 2004). Furthermore, its Robust Multi-array Average (RMA) normalization method minimizes the technical noise, leading to more precise results. One of the key benefits of GEO2R is its user-friendly interface, which allows researchers to handle large datasets without the need of advanced coding skills, thus removing the computational barrier (Irizarry *et al*., 2003; Davis *et al*., 2007). Because it is so widely used in breast cancer research, it has become an important tool for singling out the molecular patterns that show how the disease progresses (Rhodes *et al*., 2024; Loi *et al*., 2007).

Similar to this, functional enrichment analysis is also essential to study and explore the roles of differentially expressed genes in biological processes and pathways. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) plays a key role in this by mapping DEGs to Gene Ontology (GO) terms and pathways, providing valuable insights into the processes that drive cancer (Huang *et al*., 2009). With its detailed annotations and strong statistical analysis, DAVID comes as a forefront tool for interpreting complex datasets and understanding the biological functions (Sherman *et al*., 2022; Jiao *et al*., 2012). Likewise, the Kyoto Encyclopedia of Genes and Genomes (KEGG) complement with the DAVID tool linking differentially expressed genes to molecular pathways, helping in highlighting the interactions prominently involved in breast cancer progression (Van't Veer *et al*., 2002). KEGG's inbuilt and curated pathway maps also help in linking the markers to entire biological networks, making it easier to identify potential therapeutic targets (Wang *et al*., 2005). Together, DAVID and KEGG provide a powerful approach for comprehensive analysis, helping researchers to bridge the microarray data to real time clinical insights (Subramanian *et al*., 2005).

Overall, GEO2R, DAVID, and KEGG have shown to play a complementary role in addressing some of the biggest challenges in breast cancer research. GEO2R with its ease in usage have allowed researchers to analyze large cancer datasets quickly and efficiently, even without the application of advanced coding skills (Clough *et al*., 2016). DAVID has helped in deriving meaningful information and detailed annotations shedding light on critical biological processes (Huang *et al*., 2009). And KEGG with its focuses on mapping pathways, has helped in identification of new therapeutical targets in due course (Van't Veer *et al*., 2002). In this study, application of these tools to the GSE54002 dataset has tended to ensured reliable and reproducible results (Rakha *et al*., 2010). The current work proposes to establishes a strong foundation for future research and identification of new interacting molecular pathways for advancing personalized research in breast cancer care (Parker *et al*., 2009; Desmedt *et al*., 2007).

MATERIALS AND METHODS

Data Retrieval, Preprocessing and Normalization

The current study explores GSE54002 a publicly available gene expression dataset available at Gene Expression Omnibus (GEO) database. The set consists of microarray data generated using the Affymetrix Human Genome U133 Plus 2.0 Array with total 216 samples, out of which 200 are breast cancer samples and 16 normal breast tissue samples. The dataset has vast coverage with good varsity making it robust and reliable for identifying differentially expressed genes associated with breast cancer. This raw data was accessed through the GEO2R web tool, available at National Center for Biotechnology Information (NCBI). This tool automatically applied preprocessing steps, including Robust Multi-array Average (RMA) normalization, to adjust for background noise and normalize expression values across samples. The sample groups were also divided in two groups, with nomenclature as breast cancer (n=200) against normal tissue

(*n*=16) to segregate gene expression differences involved in breast cancer pathogenesis.

Expression Analysis using GEO2R

Differential expression analysis was conducted using the GEO2R tool to identify high expressing and under expressing between breast cancer and normal tissue samples. GEO2R employs the limma (Linear Models for Microarray Analysis) package in R, which calculates differential expression based on a moderated t-statistic. The analysis generated a comprehensive table of 54,675 probe sets, including columns for probe ID, adjusted *p*-value (Benjamini-Hochberg corrected), *p*-value, t-statistic, B-statistic, log2 fold change (logFC), gene symbol, and gene title. Genes were considered differentially expressed if they met the criteria of an adjusted *p*-value <0.05 and an absolute log2 fold change (|log2FC|)>1, indicating statistically significant and biologically meaningful expression changes. The GEO2R output was exported as an Excel file (GSE54002.top.table.xlsx) with three sheets: a main table containing all probe sets, a sheet for upregulated genes (positive log2FC), and a sheet for downregulated genes (negative log2FC). From this analysis, approximately 2,500 DEGs were identified, with around 1,400 upregulated and 1,100 downregulated genes, based on the provided data summaries. Key genes, such as *CTHRC1, COL1A2, FN1, TOP2A* (upregulated), and *PTN, SFRP1, KRT14* (downregulated), were prioritized for further functional analysis due to their high log2FC values and low adjusted *p*-values.

Functional Enrichment Analysis Using DAVID

To elucidate the biological roles of the identified DEGs, functional enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8). A subset of 123 DEGs, selected based on their significance and relevance to breast cancer, was submitted to DAVID for analysis. This subset was chosen to focus on genes with robust expression changes and to manage computational limitations of the DAVID tool. The DAVID analysis included two main components: Gene Ontology (GO) Biological Process (BP) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. For GO BP analysis, the tool identified enriched biological processes among the input genes, providing metrics such as gene count, percentage of input genes, uncorrected *p*-value, fold enrichment, and corrected *p*-values (Bonferroni, Benjamini, and False Discovery Rate [FDR]). Similarly, KEGG pathway analysis identified enriched pathways, with 65 of the 123 input genes mapped to KEGG annotations. Enrichment was considered significant for terms and pathways with an FDR <0.05. The results were exported as Excel files (BP Direct.xls and KEGG.xls), which included detailed annotations for terms like cell division (GO:0051301), chromosome segregation (GO:0007059), collagen fibril organization (GO:0030199), and KEGG pathways such as cell cycle (hsa04110) and ECM-receptor interaction (hsa04512). These outputs were used to interpret the functional roles of DEGs in breast cancer biology.

Gene Annotation Matrix Analysis

To further characterize the functional roles of a specific gene subset, a binary annotation matrix was analyzed, as provided in the "GAS.xls" file. This matrix included 60 genes, such as *GAS2L3, KIF20A, TOP2A*, and *CDK1*, annotated for six functional terms: Cytoskeleton (KW-0206), Cell Division (GO:0051301, KW-0132), Mitosis (KW-0498), Cytosol (GO:0005829), and Cell Cycle (KW-0131). A value of 1 indicated the gene's association with the term, while 0 indicated no association. This matrix was derived from a prior gene set enrichment analysis, likely using DAVID or a similar tool, and was used to confirm the roles of key genes in cytoskeletal dynamics and mitotic processes. The matrix analysis focused on *GAS2L3* and its interactions with other mitotic and cytoskeletal genes, providing insights into its potential role in supporting cell division through cytoskeletal organization. The results were cross-referenced with the broader enrichment analysis from the "AnnotationCluster.xls" file, which grouped DEGs into clusters based on shared functional annotations, such as cell cycle, mitosis, and ECM organization.

Statistical and Computational Tools

We carried out all analyses using freely available, web-based bioinformatics tools to keep the process both accessible and reproducible. For identifying differentially expressed genes, we used GEO2R, which applies the limma package to ensure statistically sound results. Functional enrichment analysis was performed using DAVID, with default settings applied for the human genome background (usually 19,478 genes for GO Biological Processes and 8,534 genes for KEGG pathways). Significance was assessed using adjusted *p*-values, with a threshold of FDR<0.05 for both differential expression and enrichment-based outcome analysis.

To interpret the functional annotation results, we manually examined the binary annotation matrix in Excel, focusing on the proportion of genes linked to each biological term and identifying patterns relevant to breast cancer. Moreover, the workflow of the study relied entirely on GEO2R's processed outputs and DAVID's built-in analysis tools. Final data visualization, including tables of differentially expressed genes and enriched values, is presented in tabular format.

Data Integration and Interpretation

To gain a clearer picture of the molecular mechanisms behind breast cancer, results from GEO2R, DAVID GO BP, KEGG pathway analysis, and the gene annotation matrix were carefully analyzed. GEO2R highlighted important differentially expressed genes while DAVID linked these genes to key biological processes and signaling pathways, while the annotation matrix added further detail by confirming their roles in various cellular

activities. Using this three-tiered approach, the findings sought were more reliable and biologically relevant.

RESULTS

Data Acquisition and Preprocessing

In this study, we analyzed the GSE54002 dataset from GEO which includes gene expression profiles from 216 samples: 200 breast cancer tissues and 16 normal breast tissues. We processed the dataset with the GEO2R tool and applied Robust Multi-array Average normalization to minimize background variation and keep the samples consistent. This produced log2-transformed expression values for 54,675 probe sets, which were then used to identify differentially expressing genes. The samples were already categorized into two groups of breast cancer and normal tissue samples which helped log2 values to easily demarcate gene expression changes linked to breast cancer development.

Differential Expression Analysis

Differential expression analysis using GEO2R identified approximately 2,500 differentially expressed genes (DEGs) between breast cancer and normal tissue samples, based on the criteria of an adjusted *p*-value (Benjamini-Hochberg corrected) <0.05 and an absolute log2 fold change (|log2FC|)>1. Of these, around 1,400 genes were upregulated and 1,100 were downregulated in breast cancer samples compared to normal tissue. The featured upregulated and downregulated genes were also mapped via Volcano plot, Meandiff plot and Umap analysis (Figure 1).

Among the upregulated genes, *CTHRC1* exhibited the highest differential expression (log2FC=5.93, adj. *p*=1.27e-104), followed by *COL1A2* (log2FC=5.01, adj. *p*=2.79e-20), *FN1* (log2FC=5.01, adj. *p*=7.80e-32), *TOP2A* (log2FC=4.06, adj. *p*=1.85e-39), and *RRM2* (log2FC=4.06, adj. *p*=1.13e-37). These genes are associated with Extracellular Matrix (ECM) remodeling (*CTHRC1, COL1A2, FN1*), DNA replication (*RRM2*), and cell cycle regulation (*TOP2A*), indicating a molecular signature of tumor proliferation and invasion (Table 1A).

Among the downregulated genes, *PTN* showed the strongest suppression (log2FC=-6.56, adj. *p*=1.68e-39), followed by *SFRP1* (log2FC=-6.29, adj. *p*=1.56e-23) and *KRT14* (log2FC=-6.29, adj. *p*=1.92e-29), which are linked to growth factor signaling, Wnt pathway inhibition, and cytoskeletal structure, respectively.

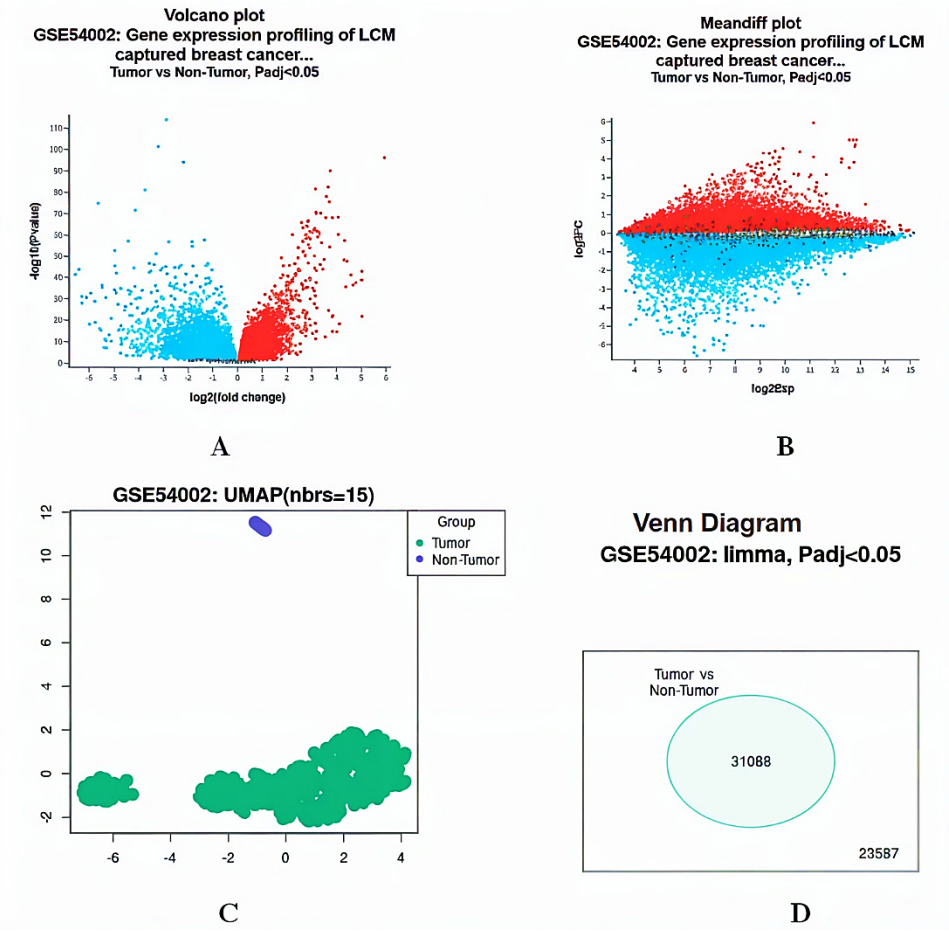

Figure 1: (A) Volcano plot of gene expression profiling dataset of GSE54002 (B) Meandiff plot of gene expression profiling dataset of GSE54002 (C) Umap anlaysis of gene expression profiling dataset of GSE54002 (D) Cluster analysis of gene expression profiling dataset of GSE54002.

Table 1: (A) Top upregulated genes in breast cancer filtered through GSE54002 dataset. (B) Top downregulated genes in breast cancer filtered through GSE54002 dataset.

Table : (A)				
Sl. No.	**adj.P.Val**	**logFC**	**Gene.symbol**	**Gene.title**
1	1.27E-92	5.93142639	CTHRC1	Collagen triple helix repeat containing 1
2	2.79E-20	5.0142425	COL1A2	Collagen type I alpha 2 chain
3	7.80E-41	5.01340145	FN1	Fibronectin 1
4	8.28E-37	5.00898065	FN1	Fibronectin 1
5	1.65E-35	4.76659341	FN1	Fibronectin 1
6	1.38E-34	4.65720091	FN1	Fibronectin 1
7	3.82E-39	4.55976408	COL11A1	Collagen type XI alpha 1 chain
8	3.21E-45	4.40836781	ASPN	Asporin
9	1.02E-33	4.36103501	COL1A1	Collagen type I alpha 1 chain
10	6.39E-55	4.30742148	COL10A1	Collagen type X alpha 1 chain
11	6.04E-46	4.28896384	COL11A1	Collagen type XI alpha 1 chain
12	4.29E-17	4.10320757	COL3A1	Collagen type III alpha 1 chain
13	1.13E-57	4.06130418	RRM2	Ribonucleotide reductase regulatory subunit M2
14	1.85E-65	4.05562062	TOP2A	Topoisomerase (DNA) II alpha
15	1.35E-13	4.00115742	COL1A2	Collagen type I alpha 2 chain
16	8.73E-21	3.84302497	S100P	S100 calcium binding protein P
17	3.01E-65	3.82897801	ANLN	Anillin actin binding protein
18	9.45E-14	3.82693634	COL3A1	Collagen type III alpha 1 chain
19	4.88E-20	3.81585104	COL1A1	Collagen type I alpha 1 chain
20	1.24E-86	3.73383637	NEK2	NIMA related kinase 2
Table : (B)				
	1.68E-39	-6.56021709	PTN	Pleiotrophin
	9.02E-42	-6.41567324	PTN	Pleiotrophin
	1.56E-26	-6.29396468	SFRP1	Secreted frizzled related protein 1
	1.92E-29	-6.28647366	KRT14	Keratin 14
	6.51E-29	-6.12532305	DST	Dystonin
	4.98E-17	-5.98477313	GABRP	Gamma-aminobutyric acid type A receptor pi subunit
	5.35E-31	-5.93855705	SFRP1	Secreted frizzled related protein 1
	3.49E-27	-5.76949471	PIGR	Polymeric immunoglobulin receptor
	9.13E-18	-5.76049009	PI15	Peptidase inhibitor 15
	5.95E-72	-5.63414734	PTN	Pleiotrophin
	5.44E-15	-5.57092826	IRX1	Iroquois homeobox 1
	3.99E-14	-5.50549971	CCL28	C-C motif chemokine ligand 28
	1.90E-33	-5.47359064	ACTA2	Actin, alpha 2, smooth muscle, aorta
	1.35E-34	-5.47280036	KIT	KIT proto-oncogene receptor tyrosine kinase
	5.55E-07	-5.40238136	ID4	Inhibitor of DNA binding 4, HLH protein
	1.30E-10	-5.37406118	BBOX1	Gamma-butyrobetaine hydroxylase 1
	6.39E-29	-5.33846336	OPRPN	Opiorphin prepropeptide
	9.24E-10	-5.32454477	NTRK2	Neurotrophic receptor tyrosine kinase 2
	8.12E-07	-5.3064456	MUC15	Mucin 15, cell surface associated
	2.51E-12	-5.19228729	WIF1	WNT inhibitory factor 1

These findings suggest a loss of tumor-suppressive mechanisms in breast cancer (Table 1B).

Functional Enrichment Analysis Using DAVID

Functional enrichment analysis of 123 selected DEGs using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8) revealed significant biological processes and pathways associated with breast cancer. Of the 123 DEGs, 118 were mapped to GO Biological Process (BP) terms, and 65 were mapped to KEGG pathways. The GO BP analysis identified highly enriched terms, with cell division (GO:0051301) being the most significant, involving 31 genes (25.2% of input, FDR=7.67e-22), including *CDCA3, KIF14, CCNB1, PTTG1, TOP2A*, and *CDK1*. Chromosome segregation (GO:0007059) was also highly enriched (15 genes, 12.2%, FDR=5.19e-13), with genes such as *TOP2A, CENPE, CENPF*, and *BUB1*. Collagen fibril organization (GO:0030199) involved 13 genes (10.6%, FDR=1.40e-12), including *COL1A1, COL1A2, COL3A1*, and *COL5A1*, highlighting ECM remodeling. Mitotic spindle organization (GO:0007052) included 11 genes (8.9%, FDR=3.18e-10), such as *KIF11* and *AURKA*. KEGG pathway analysis identified significant pathways, including cytoskeleton in muscle cells (hsa04820, 15 genes, 12.2%, FDR=9.10e-08), driven by ECM genes like *COL1A1, COL1A2*, and *FN1*, despite the pathway name's specificity to muscle cells. The cell cycle pathway (hsa04110, 11 genes, 8.9%, FDR=8.91e-06) included *CDC20, CCNB1, CCNB2*, and *CDK1*, reinforcing proliferative mechanisms. ECM-receptor interaction (hsa04512, 7 genes, 5.7%, FDR=7.82e-04) featured *COL1A2* and *FN1*, and the p53 signaling pathway (hsa04115, FDR<0.05) included *RRM2* and *CDK1* (Table 2). These results indicate that the DEGs are primarily involved in cell proliferation, mitotic regulation, and ECM remodeling, key processes in breast cancer progression.

Gene Annotation Matrix Analysis

The gene annotation matrix provided detailed functional annotations for 60 genes, including *GAS2L3*, across six terms: Cytoskeleton (KW-0206), Cell Division (GO:0051301, KW-0132), Mitosis (KW-0498), Cytosol (GO:0005829), and Cell Cycle (KW-0131). Of these genes, 30 (50%) were annotated for Cytoskeleton, including *GAS2L3, KIF20A, KIF11, CENPE, CENPF*, and *TPX2*, aligning with their roles in microtubule and actin dynamics. Cell Division was annotated for 31 genes (51.67%, GO:0051301) and 35 genes (58.33%, KW-0132), including *KIF14, CCNB1, PTTG1, CDK1*, and *BUB1*, indicating strong involvement in mitotic processes. Mitosis (KW-0498) was annotated for 32 genes (53.33%), overlapping with cell division genes, such as *AURKA* and *CENPF*. Cytosol (GO:0005829) was the most frequently annotated term, covering 49 genes (81.67%), including *GAS2L3* and *TOP2A*, suggesting a cytosolic localization for most DEGs. Cell Cycle (KW-0131) was annotated for 43 genes (71.67%), including *MKI67, FOXM1, CDK1*, and *CDC20* (Table 3). Notably, *GAS2L3* was annotated only for Cytoskeleton and Cytosol, suggesting an indirect role in supporting mitotic processes through cytoskeletal organization rather than direct involvement in cell division or cycle regulation. This matrix analysis corroborated the DAVID enrichment results, particularly for cytoskeletal and mitotic functions, and highlighted *GAS2L3*'s role in coordinating cytoskeletal dynamics in breast cancer cells.

Statistical and Computational Tools

The statistical and computational analyses were robustly supported by the GEO2R and DAVID tools. GEO2R's use of the limma package ensured reliable differential expression analysis, with adjusted *p*-values (Benjamini-Hochberg) confirming the significance of the ~2,500 DEGs. DAVID's enrichment analysis utilized default human genome backgrounds (19,478 genes for GO BP, 8,534 for KEGG), with FDR<0.05 as the threshold for significance, ensuring rigorous identification of enriched terms and pathways. The gene annotation matrix was analyzed manually in Excel, quantifying the proportion of genes associated with each term (e.g., 50% for Cytoskeleton, 71.67% for Cell Cycle), which provided a clear and reproducible assessment of functional roles. The integration of these tools facilitated a

Table 2: Gene Ontology (GO) Analysis of Biological Processes.

Sl. No.	GO Term (ID)	GO Term Description	Count	Percentage (%)	Genes	Fold Enrichment
1.	GO:0051301	Cell Division	31	25.20325203	CDCA3, KIF14, NCAPG, BUB1B, KIF11, AURKA, CDC20, CCNB2, CCNB1, PTTG1, NUF2, NEK2, FAM83D, BUB1, SPAG5, UBE2C, KNL1, NDC80, ZWINT, ASPM, CENPE, TPX2, CENPF, KIF18B, CCNE2, PRC1, CKS2, CDK1, TACC3, BIRC5, KIF2C	12.88942492

Sl. No.	GO Term (ID)	GO Term Description	Count	Percentage (%)	Genes	Fold Enrichment
2.	GO:0007059	Chromosome Segregation	15	12.19512195	TOP2A, CENPU, SPAG5, HJURP, TTK, MKI67, NDC80, CENPE, CENPF, NUF2, BIRC5, NEK2, KIF2C, BUB1, DLGAP5	24.5150193
3.	GO:0007155	Cell Adhesion	14	11.38211382	CNTNAP2, TNFAIP6, COL12A1, FN1, THBS2, NID2, COL1A1, COMP, VCAN, COL5A1, FAP, SPOCK1, COL6A3, IGFBP7	4.134077196
4.	GO:0030199	Collagen Fibril Organization	13	10.56910569	LUM, COL11A1, COL12A1, LOXL1, GREM1, COL1A1, COMP, MMP11, COL3A1, SFRP2, COL1A2, COL5A1, COL5A2	32.51335388
5.	GO:0006915	Apoptotic Process	13	10.56910569	IFI6, BUB1B, SIX1, AURKA, COMP, TPX2, MELK, IFI27, CEACAM6, CEACAM5, CDK1, BIRC5, BUB1	3.271160604

comprehensive analysis, with GEO2R identifying DEGs, DAVID elucidating their biological functions, and the matrix providing gene-specific annotations. No discrepancies were observed in the computational outputs, and the results were consistent across all tools used.

Data Integration and Interpretation

Integrating the results from GEO2R, DAVID, and the gene annotation matrix revealed a cohesive molecular profile of breast cancer. The GEO2R analysis highlighted key DEGs driving tumor progression: upregulated genes like *CTHRC1* (log2FC=5.93), *COL1A2* (log2FC=5.01), and *FN1* (log2FC=5.01) underscored ECM remodeling, while *TOP2A* (log2FC=4.06), *RRM2* (log2FC = 4.06), and *CDK1* (log2FC = 3.19) indicated hyperactive cell proliferation. Downregulated genes, such as *SFRP1* (log2FC=-6.29) and *PTN* (log2FC=-6.56), suggested loss of tumor-suppressive mechanisms, particularly in Wnt signaling. DAVID's GO BP and KEGG analyses confirmed these findings, with enriched terms like cell division (FDR=7.67e-22) and cell cycle (FDR=8.91e-06) linked to *TOP2A* and *CDK1*, and ECM-related terms like collagen fibril organization (FDR=1.40e-12) and ECM-receptor interaction (FDR=7.82e-04) tied to *COL1A2* and *FN1*. The gene annotation matrix further validated the prominence of mitotic and cytoskeletal processes, with 71.67% of the 60 genes annotated for Cell Cycle and 50% for Cytoskeleton. *GAS2L3* emerged as a key player in cytoskeletal dynamics, likely supporting mitotic spindle organization in concert with kinesin genes (*KIF11, KIF20A*) and other cytoskeletal proteins (*TPX2*). These integrated findings suggest that the DEGs contribute to breast cancer through enhanced proliferation, ECM remodeling, and cytoskeletal reorganization, with *GAS2L3* playing a supportive role in the latter.

Similarly, Table 4 provides the data obtained from KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis. It summarizes pathways that are statistically enriched in a given set of genes, providing insights into biological processes, molecular functions, and diseases associated with these genes. The pathway "hsa04820: Cytoskeleton in muscle cells" is the most significantly enriched pathway with a *p*-value of 1.17e-09, a fold enrichment of 8.49, and low multiple testing correction values (Bonferroni and FDR). This suggests that genes in the input set are highly overrepresented in this pathway, indicating its biological significance. Likewise, "hsa04110: Cell cycle" is another highly enriched pathway with 11 genes from the input set, a fold enrichment of 9.14, and a *p*-value of 2.28e-07 (Table 4). On the other hand, pathways like "hsa04933: AGE-RAGE signaling pathway in diabetic complications" and "hsa04926: Relaxin signaling pathway" were noticed to have higher *p*-values (0.039 and 0.073) and lower fold enrichment values, indicating their weaker statistical significance.

DISCUSSION

The analysis of the GSE54002 dataset, comprising 200 breast cancer and 16 normal breast tissue samples, revealed a comprehensive molecular profile of breast cancer, characterized by approximately 2,500 differentially expressed genes (DEGs), with 1,400 upregulated and 1,100 downregulated genes (adj. $p<0.05$, $|\log2FC|>1$). The identification of key upregulated genes, such as *CTHRC1* (log2FC=5.93), *COL1A2* (log2FC=5.01), *FN1* (log2FC=5.01), *TOP2A* (log2FC=4.06), and *RRM2* (log2FC=4.06), alongside downregulated genes like *PTN* (log2FC=-6.56), *SFRP1* (log2FC=-6.29), and *KRT14* (log2FC=-6.29), underscores the dual roles of tumor proliferation and Extracellular Matrix (ECM) remodeling in breast cancer progression. Functional enrichment analysis using DAVID highlighted significant biological

Table 3: Gene Association Table for GO Terms and Keywords Related to Cytoskeleton, Cell Division, Mitosis, Cytosol, and Cell Cycle.

GO Term/Keyword	Functional Description	Associated Genes with the Term (Value=1)	Non-Associated Genes with the Term (Value=0)
KW-0206~Cytoskeleton	Genes related to the cytoskeleton	GAS2L3, KIF20A, KIF15, DTL, KIF4A, HMMR, CENPE, CENPF, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, CDC20, KIF18B, SPAG5	TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, MKI67, E2F7, UHRF1, FOXM1, DLGAP5, CKS2, MELK, CEP55, NUSAP1, CCNE2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA, NEK2
GO:0051301~Cell Division	Genes involved in cell division	DLGAP5, CKS2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA	GAS2L3, KIF20A, TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, MKI67, E2F7, UHRF1, FOXM1, KIF15, DTL, KIF4A, HMMR, CDKN3, HJURP, E2F8, KIF14, CEP55, NUSAP1, CCNE2, KIF18B, SPAG5, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, NEK2, CENPE, CENPF
KW-0498~Mitosis	Genes related to mitosis	MELK, CEP55, NUSAP1, CCNE2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, CDC20, KIF18B, SPAG5	GAS2L3, KIF20A, TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, MKI67, E2F7, UHRF1, FOXM1, KIF15, DTL, KIF4A, HMMR, CDKN3, HJURP, E2F8, KIF14, CENPE, CENPF
KW-0132~Cell Division	Genes involved in cell division	MELK, CEP55, NUSAP1, CCNE2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, CDC20, KIF18B, SPAG5	GAS2L3, KIF20A, TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, MKI67, E2F7, UHRF1, FOXM1, KIF15, DTL, KIF4A, HMMR, CDKN3, HJURP, E2F8, KIF14, CENPE, CENPF
GO:0005829~Cytosol	Genes localized to the cytosol	TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, KIF15, DTL, KIF4A, HMMR, CDKN3	GAS2L3, KIF20A, MKI67, E2F7, UHRF1, FOXM1, DLGAP5, CKS2, MELK, CEP55, NUSAP1, CCNE2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA, NEK2, CENPE, CENPF, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, CDC20, KIF18B, SPAG5

GO Term/Keyword	Functional Description	Associated Genes with the Term (Value=1)	Non-Associated Genes with the Term (Value=0)
KW-0131~Cell Cycle	Genes involved in the cell cycle	MKI67, E2F7, UHRF1, FOXM1, HMMR, CDKN3, HJURP, E2F8, KIF14, DLGAP5, CKS2, MELK, CEP55, NUSAP1, CCNE2, ASPM, PRC1, KIF23, ANLN, PTTG1, NCAPG, NDC80, CCNB2, UBE2C, BUB1, NUF2, ZWINT, KNL1, AURKA, BUB1B, TPX2, FAM83D, BIRC5, CDK1, TACC3, CDC20, KIF18B, SPAG5	GAS2L3, KIF20A, TYMS, ISG15, TK1, CENPU, NEURL1B, RRM2, TNNT1, IQGAP3, KMO, SLC27A2, KIF15, DTL, KIF4A

processes, including cell division (GO:0051301, 31 genes, FDR=7.67e-22), chromosome segregation (GO:0007059, 15 genes, FDR=5.19e-13), and collagen fibril organization (GO:0030199, 13 genes, FDR=1.40e-12), as well as KEGG pathways such as cell cycle (hsa04110, FDR=8.91e-06) and ECM-receptor interaction (hsa04512, FDR=7.82e-04). The gene annotation matrix further confirmed the prominence of cytoskeletal and mitotic processes, with *GAS2L3* annotated for cytoskeleton (KW-0206) and cytosol (GO:0005829), suggesting its supportive role in mitotic spindle dynamics. These findings align with current knowledge of breast cancer as a disease driven by uncontrolled proliferation, ECM remodeling, and cytoskeletal reorganization, offering insights into potential therapeutic targets and biomarkers (Perou *et al*., 2000; Hanahan *et al*., 2011).

The upregulation of *CTHRC1, COL1A2,* and *FN1* points to a critical role for ECM remodeling in breast cancer. *CTHRC1*, with the highest log2FC (5.93), is a collagen triple helix repeat-containing protein known to promote tumor invasion by enhancing ECM deposition and remodeling, facilitating cancer cell migration and metastasis (Park *et al*., 2013). Similarly, *COL1A2* and *FN1* contribute to a collagen-rich tumor microenvironment, which supports tumor stiffness and invasiveness, as evidenced by their enrichment in GO:0030199 (collagen fibril organization) and KEGG hsa04512 (ECM-receptor interaction) (Provenzano *et al*., 2009; Ioachim *et al*., 2002). These findings are consistent with studies showing that ECM alterations are a hallmark of aggressive breast cancer subtypes, particularly triple-negative and basal-like tumors, where increased collagen deposition correlates with poor prognosis (Walker *et al*., 2018). The downregulation of *SFRP1* and *PTN* further suggests a loss of tumor-suppressive mechanisms. *SFRP1*, a Wnt pathway inhibitor, is frequently silenced in breast cancer, leading to uncontrolled Wnt/β-catenin signaling, which drives tumor growth and metastasis (Ugolini *et al*., 2001). The significant downregulation of *SFRP1* (log2FC=-6.29) aligns with its role as a tumor suppressor, and its loss may enhance the proliferative effects of upregulated cell cycle genes like *TOP2A* and *CDK1* (Bernemann *et al*., 2014). *PTN*, a growth factor involved in angiogenesis, showed the strongest downregulation (log2FC=-6.56), suggesting disrupted normal tissue homeostasis, which may contribute to tumor progression (Papadimitriou *et al*., 2001). These results connect directly to current knowledge, reinforcing the interplay between ECM remodeling and dysregulated signaling in breast cancer pathogenesis.

The enrichment of cell cycle and mitotic processes, as seen in GO:0051301 (cell division) and KEGG hsa04110 (cell cycle), highlights the proliferative phenotype of breast cancer. Genes such as *CDK1, CCNB1, TOP2A,* and *RRM2* are central to regulating cell growth and division. In particular, *TOP2A* (DNA topoisomerase II alpha) and *CDK1* (cyclin-dependent kinase 1) act as key drivers of DNA replication and mitotic progression (Nitiss *et al*., 2009) Both *TOP2A* and *RRM2* show high log2FC values (4.06), reflecting their marked overexpression in breast cancer. This aligns with their role in fueling the rapid cell division that characterizes malignant tumors (Aye *et al*., 2015). Furthermore, enrichment of the p53 signaling pathway (hsa04115), involving genes such as *RRM2* and *CDK1*, points toward disrupted cell cycle checkpoints, a change that can generally promote genomic instability and resistance to programmed cell death (Vousden *et al*., 2007). These findings align with previous studies showing that *TOP2A* amplification is associated with aggressive breast cancer subtypes and is a target for anthracycline-based chemotherapies, such as doxorubicin (Nielsen *et al*., 2008). The gene annotation matrix further supports this proliferative signature, with 43 of 60 genes (71.67%) annotated for Cell Cycle (KW-0131), including *MKI67* and *FOXM1*, which are markers of proliferation and transcriptional regulators of cell cycle genes (Wierstra *et al*., 2007). The matrix's emphasis on mitotic processes (32 genes, 53.33%) and chromosome segregation (15 genes in GO:0007059) underscores the role of genes like *CENPE, CENPF*, and *BUB1* in ensuring accurate chromosome segregation, where dysregulation can lead to aneuploidy, a common feature in breast cancer (Holland *et al*., 2012).

The role of *GAS2L3* in this study is particularly noteworthy, as it represents a novel feature explored in the context of breast cancer. Annotated for Cytoskeleton (KW-0206) and Cytosol (GO:0005829) in the "GAS.xls" matrix, *GAS2L3* is not directly linked to cell division or mitosis but likely supports these processes through cytoskeletal organization. *GAS2L3* is known

Table 4: KEGG Pathway Enrichment Summary.

Pathway (KEGG ID)	Description	Count	*p*-Value	Genes	Fold Enrichment
hsa04820	Cytoskeleton in muscle cells	15	1.17e-09	COL11A1, BGN, FN1, THBS2, NID2, COL1A1, COMP, COL3A1, VCAN, COL1A2, COL5A1, TNNT1, COL5A2	8.49
hsa04110	Cell cycle	11	2.28e-07	CDC20, CCNB2, CCNB1, PTTG1, CCNE2, CDK1, BUB1B, TTK, KNL1, BUB1, NDC80	9.14
hsa04974	Protein digestion and absorption	9	1.00e-06	COL1A1, COL3A1, COL1A2, COL5A1, COL11A1, COL12A1, COL5A2, COL10A1, COL6A3	11.25
hsa04814	Motor proteins	10	1.42e-05	CENPE, KIF18B, TNNT1, KIF4A, KIF14, KIF23, KIF2C, KIF20A, KIF11, KIF15	6.66
hsa04512	ECM-receptor interaction	7	5.01e-05	COL1A1, COMP, COL1A2, FN1, COL6A3, HMMR, THBS2	10.33

to regulate actin-microtubule interactions, which are critical for mitotic spindle positioning and cytokinesis (Goidts *et al*., 2012). Its co-occurrence with kinesin genes (*KIF11, KIF20A, KIF14*) in the matrix and DAVID's Cluster 6 (kinesin motor activity, FDR=2.33e-09) suggests that *GAS2L3* may stabilize microtubule dynamics during mitosis, facilitating chromosome movement and cell division (Zhu *et al*., 2005). This finding is novel in the context of breast cancer, as *GAS2L3* has been less studied compared to other cytoskeletal genes like *TPX2* or *KIF11*. The enrichment of cytoskeletal terms (30 genes, 50% in the matrix; Cluster 3, FDR=9.30e-19 for spindle) highlights the importance of cytoskeletal dynamics in breast cancer, potentially contributing to tumor cell motility and invasion (Yamaguchi *et al*., 2007). The interaction between *GAS2L3* and kinesins could represent a new therapeutic target, as cytoskeletal proteins are increasingly recognized as mediators of cancer progression (Hall *et al*., 2009).

These findings carry important implications for both fundamental research and clinical practice. The elevated expression of ECM-related genes such as *CTHRC1, COL1A2, and FN1* points to the tumor microenvironment as a promising target for limiting metastasis. Agents that inhibit collagen remodeling like Matrix Metalloproteinase (MMP) inhibitors may help curb tumor spread, although their effectiveness in clinical settings is still being evaluated (Fingleton *et al*., 2008). Likewise, the strong presence of *TOP2A* in the cell cycle pathway reinforces its role as a target for anthracycline-based therapies. Its high expression in our dataset further supports its potential as a predictive biomarker for chemotherapy response in breast cancer (Di Leo *et al*., 2011). The downregulation of *SFRP1* points to Wnt pathway activation, suggesting potential therapeutic roles for Wnt inhibitors, which are being explored in clinical trials for breast cancer (Geyer *et al*., 2011). The identification of *GAS2L3* as a cytoskeletal regulator opens new avenues for research into microtubule-targeting agents, such as taxanes, which are already used in breast cancer treatment but could be optimized to target *GAS2L3*-mediated pathways (Jordan *et al*., 2004). These findings also have diagnostic potential, as *CTHRC1* and *FN1* could serve as biomarkers for invasive breast cancer, while *SFRP1* downregulation may indicate specific subtypes, such as basal-like tumors (Oskarsson *et al*., 2013).

The study's applications extend to personalized medicine, where the molecular profile of DEGs can guide treatment decisions. For instance, patients with high *TOP2A* expression may benefit from anthracycline-based therapies, while those with elevated *CTHRC1* could be prioritized for therapies targeting ECM remodeling (Kennecke *et al*., 2010). The integration of GEO2R, DAVID, and matrix analyses provides a robust framework for identifying such molecular signatures, which can be validated in larger cohorts or through multi-omics approaches (Curtis *et al*., 2012). The focus on *GAS2L3* adds a novel dimension, as its role in cytoskeletal dynamics could be explored in combination with existing therapies, such as microtubule inhibitors, to enhance treatment efficacy (Dumontet *et al*., 2010). Furthermore, the enrichment of p53 signaling and PI3K-Akt pathways (KEGG hsa04151) suggests additional therapeutic targets, as these pathways are frequently dysregulated in breast cancer and are the focus of ongoing drug development efforts (Engelman JA., 2009). Despite these insights, the study has few shortcomings. The reliance on microarray data from GSE54002, while robust, is limited by the technology's lower resolution compared to RNA-sequencing, which may miss low-abundance transcripts (Wang *et al*., 2009). The sample size disparity (200 breast cancer vs. 16 normal samples) may introduce statistical bias, though GEO2R's limma package mitigates this through robust normalization (Ritchie *et al*., 2015). The DAVID analysis was restricted to 123 DEGs, potentially overlooking other significant genes due to computational constraints (Huang *et al*., 2009).

Future prospects include validating these findings in independent cohorts using RNA-sequencing or proteomics to confirm the expression and functional roles of *CTHRC1, TOP2A*, and *GAS2L3* (Cancer Genome Atlas Network., 2012). Experimental studies, such as CRISPR knockout or overexpression of *GAS2L3*, could elucidate its role in cytoskeletal dynamics and its therapeutic potential (Ran *et al*., 2013). Integrating multi-omics data (e.g., proteomics, epigenomics) could provide a more comprehensive understanding of the molecular networks driving breast cancer (Zhang *et al*., 2014). Clinical studies correlating *CTHRC1* and *SFRP1* expression with patient outcomes could establish their utility as biomarkers (Rakha *et al*., 2010). Additionally, exploring *GAS2L3* in combination with microtubule-targeting therapies could uncover novel treatment strategies (Bates *et al*., 2017). Expanding the analysis to include breast cancer subtypes (e.g., luminal, HER2-positive, triple-negative) would further enhance the study's clinical relevance (Prat *et al*., 2011).

CONCLUSION

The present study provides a comprehensive molecular profile of breast cancer, revealing critical roles for CTHRC1 and COL1A2 in extracellular matrix remodeling, TOP2A and CDK1 in driving proliferation, and GAS2L3 in supporting cytoskeletal dynamics. These findings align with existing knowledge and offer novel insights into GAS2L3's contribution to mitotic processes. The results highlight promising therapeutic targets, such as TOP2A for chemotherapy and GAS2L3 for microtubule-based therapies, and potential biomarkers like CTHRC1. This work establishes a robust foundation for future research to validate these molecular targets and advance personalized treatment strategies, significantly enhancing the understanding and management of breast cancer.

ABBREVIATIONS

DAVID: Database for Annotation, Visualization, and Integrated Discovery; **ECM:** Extracellular matrix; **DEGs:** Differentially expressed genes; **GEO:** Gene Expression Omnibus; **RMA:** Robust Multi-array Average; **KEGG:** Kyoto Encyclopedia of Genes and Genomes; **NCBI:** National Center for Biotechnology Information; **logFC:** Log2 fold change; **FDR:** False Discovery Rate; **GO:** Gene Ontology; **BP:** Biological Process.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

GENERATIVE AI STATEMENT

The author declares that Generative AI tools, including, Grammarly, and QuillBot, was used to enhance the language and clarity of this work. I take full responsibility for the accuracy and integrity of the content.

SUMMARY

The study delivers a inclusive molecular profile of breast cancer, revealing significant roles for *CTHRC1* and *COL1A2* in extracellular matrix remodeling, *TOP2A* and *CDK1* in proliferation, and GAS2L3 in cytoskeletal dynamics, offering hopeful therapeutic targets and biomarkers.

REFERENCES

André, F., Ciruelos, E., Rubovszky, G., Campone, M., Loibl, S., Rugo, H. S., Iwata, H., Conte, P., Mayer, I. A., Kaufman, B., Yamashita, T., Lu, Y.-S., Inoue, K., Takahashi, M., Pápai, Z., Longin, A.-S., Mills, D., Wilke, C., Hirawat, S., . . . SOLAR-1 Study Group. (2019). Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. The New England Journal of Medicine, 380(20), 1929–1940. https://doi.org/10.1056/NEJMoa1813904

Aye, Y., Li, M., Long, M. J. C., & Weiss, R. S. (2015). Ribonucleotide reductase and cancer: New tricks for an old dog. Critical Reviews in Biochemistry and Molecular Biology, 50(3), 221–234. https://doi.org/10.3109/10409238.2015.1004049

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets-Update. Nucleic Acids Research, 41(Database issue), D991–D995. https://doi.org/10.1093/nar/gks1193

Bates, D., & Eastman, A. (2017). Microtubule-based therapeutics for cancer: Past, present, and future. Pharmacological Reviews, 69(2), 173–197. https://doi.org/10.1124/pr.116.012237

Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., Rao, A., Schultz, A., Li, X., Sumazin, P., Williams, C., Mestdagh, P., Gunaratne, P. H., Yau, C., Bowlby, R., (2018). A comprehensive pan-cancer molecular study of gynecologic and breast cancers. Cancer Cell, 33(4), 690–705.e9. https://doi.org/10.1016/j.ccell.2018.03.014

Bernemann, C., Hülsewig, C., Ruckert, C., Schäfer, S., Blümel, L., Hempel, G., Boxberger, S., Wessel, N., Kolarova, T., & zur Hausen, A. (2014). Influence of secreted frizzled receptor protein 1 (SFRP1) on neoadjuvant chemotherapy in triple-negative breast cancer. Breast Cancer Research and Treatment, 148(3), 629–637. https://doi.org/10.1007/s10549-014-3188-z

Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. Nature, 490(7418), 61–70. https://doi.org/10.1038/nature11412

Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus database. Methods in Molecular Biology, 1418, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5

Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., METABRIC Group, Langerød, A., Green, A., (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 486(7403), 346–352. https://doi.org/10.1038/nature10983

Davis, S., & Meltzer, P. S. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 23(14), 1846–1847. https://doi.org/10.1093/bioinformatics/btm254

Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M. S., Bergh, J., Lidereau, R., Ellis, P., Harris, A. L., Klijn, J. G. M., Foekens, J. A., Cardoso, F., Piccart, M. J., Buyse, M., & Sotiriou, C. (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients. Journal of Clinical Oncology, 25(15), 1991–1997. https://doi.org/10.1200/JCO.2006.10.4025

Di Leo, A., Desmedt, C., Bartlett, J. M. S., Piette, F., Ejlertsen, B., Pritchard, K. I., Larsimont, D., Poole, C., Isola, J., Earl, H., Mouridsen, H., O'Malley, F. P., Cardoso, F., Tanner, M., Munro, A., Twelves, C. J., Sotiriou, C., & Piccart-Gebhart, M. J. (2011). HER2 and TOP2A as predictive markers for anthracycline sensitivity in early breast cancer. Breast Cancer Research and Treatment, 129(1), 147–157. https://doi.org/10.1007/s10549-011-1438-6

Dumontet, C., & Jordan, M. A. (2010). Microtubule-binding agents: A dynamic field of cancer therapeutics. Nature Reviews. Drug Discovery, 9(10), 790–803. https://doi.org/10.1038/nrd3253

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research, 30(1), 207–210. https://doi.org/10.1093/nar/30.1.207

Engelman, J. A. (2009). Targeting PI3K signalling in cancer: Opportunities, challenges and limitations. Nature Reviews. Cancer, 9(8), 550–562. https://doi.org/10.1038/nrc2664

Fingleton, B. (2008). Matrix metalloproteinases as valid clinical targets. Current Pharmaceutical Design, 14(3), 259–269. https://doi.org/10.2174/138161208783413284

Geyer, F. C., Lacroix-Triki, M., Savage, K., Arnedos, M., Lambros, M. B., MacKay, A., Natrajan, R., & Reis-Filho, J. S. (2011). β-catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. Modern Pathology, 24(2), 209–231. https://doi.org/10.1038/modpathol.2010.205

Goidts, V., Latre, L., Van Den Eynde, A., & Van Camp, B. (2012). GAS2L3, a novel target gene of the DREAM complex, regulates mitotic progression. Cell Cycle, 11(17), 3229–3238. https://doi.org/10.4161/cc.21623

Hall, A. (2009). The cytoskeleton and cancer. Cancer Metastasis Reviews, 28 (1–2), 5–14. https://doi.org/10.1007/s10555-008-9166-3

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. Cell, 144(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Holland, A. J., & Cleveland, D. W. (2012). Chromoanagenesis and cancer: Mechanisms and consequences of localized, complex chromosomal rearrangements. Nature Medicine, 18(11), 1630–1638. https://doi.org/10.1038/nm.2988

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

Ignatiadis, M., & Sotiriou, C. (2013). Understanding the molecular basis of histologic grade. Pathobiology, 80(2), 104–111. https://doi.org/10.1159/000342074

Ioachim, E., Charchanti, A., Briasoulis, E., Karavasilis, V., Tsanou, H., Arvanitis, D. L., Agnantis, N. J., & Pavlidis, N. (2002). Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer. European Journal of Cancer, 38(10), 1302–1308. https://doi.org/10.1016/S0959-8049(02)00049-4

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 4(2), 249–264. https://doi.org/10.1093/biostatistics/4.2.249

Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., & Lempicki, R. A. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics, 28(13), 1805–1806. https://doi.org/10.1093/bioinformatics/bts251

Jordan, M. A., & Wilson, L. (2004). Microtubules as a target for anticancer drugs. Nature Reviews. Cancer, 4(4), 253–265. https://doi.org/10.1038/nrc1317

Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M. C. U., Voduc, D., Speers, C. H., Nielsen, T. O., & Gelmon, K. (2010). Metastatic behavior of breast cancer subtypes. Journal of Clinical Oncology, 28(20), 3271–3277. https://doi.org/10.1200/JCO.2009.25.9820

Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Harris, A., Bergh, J., Foekens, J. A., Klijn, J. G. M., Larsimont, D., Buyse, M., Bontempi, G., Delorenzi, M., Piccart, M. J., & Sotiriou, C. (2007). Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. Journal of Clinical Oncology, 25(10), 1239–1246. https://doi.org/10.1200/JCO.2006.07.1522

Nielsen, K. V., Ejlertsen, B., Møller, S., Jørgensen, J. T., Knoop, A., Knudsen, H., Mouridsen, H. T., & Brünner, N. (2008). The value of TOP2A gene copy number variation as a biomarker in breast cancer: Update of DBCG trial 89D. Acta Oncologica, 47(4), 725–734. https://doi.org/10.1080/02841860801995396

Nitiss, J. L. (2009). DNA topoisomerase II and its growing repertoire of biological functions. Nature Reviews. Cancer, 9(5), 327–337. https://doi.org/10.1038/nrc2608

Oskarsson, T. (2013). Extracellular matrix components in breast cancer progression and metastasis. Breast, 22(Suppl. 2), S66–S72. https://doi.org/10.1016/j.breast.2013.07.012

Papadimitriou, E., Polykratis, A., Courty, J., Kool, D., Heroult, M., & Vassy, R. (2001). Pleiotrophin: An angiogenic and mitogenic growth factor. Trends in Biochemical Sciences, 26(3), 165–170. https://doi.org/10.1016/S0968-0004(00)01760-5

Park, E. H., Kim, S., Jo, J. Y., Kim, S. J., Hwang, Y., Kim, J. M., Song, S. Y., Lee, D. K., & Lee, S. J. (2013). Collagen triple helix repeat containing-1 promotes pancreatic cancer progression through enhancing epithelial-mesenchymal transition. Oncotarget, 4(12), 2329–2343. https://doi.org/10.18632/oncotarget.1590

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Stijleman, I. J., Palazzo, J., Marron, J. S., Nobel, A. B., Mardis, E., Nielsen, T. O., Ellis, M. J., Perou, C. M., & Bernard, P. S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of Clinical Oncology, 27(8), 1160–1167. https://doi.org/10.1200/JCO.2008.18.1370

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. Nature, 406(6797), 747–752. https://doi.org/10.1038/35021093

Polyak, K. (2011). Heterogeneity in breast cancer. The Journal of Clinical Investigation, 121(10), 3786–3788. https://doi.org/10.1172/JCI60534

Prat, A., & Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. Molecular Oncology, 5(1), 5–23. https://doi.org/10.1016/j.molonc.2010.11.003

Provenzano, P. P., Inman, D. R., Eliceiri, K. W., & Keely, P. J. (2009). Matrix density-induced mechanoregulation of breast cell phenotype, signaling, and gene expression through a FAK-ERK linkage. Oncogene, 28(49), 4326–4343. https://doi.org/10.1038/onc.2009.299

Rakha, E. A., Reis-Filho, J. S., & Ellis, I. O. (2010). Combinatorial biomarker expression in breast cancer. Breast Cancer Research and Treatment, 120(2), 293–308. https://doi.org/10.1007/s10549-010-0746-x

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols, 8(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., & Chinnaiyan, A. M. (2004). Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proceedings of the National Academy of Sciences of the United States of America, 101(25), 9309–9314. https://doi.org/10.1073/pnas.0401994101

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47. https://doi.org/10.1093/nar/gkv007

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). David: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research, 50(W1), W216–W221. https://doi.org/10.1093/nar/gkac194

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 3(1), Article Article3. https://doi.org/10.2202/1544-6115.1027

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., & Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America, 98(19), 10869–10874. https://doi.org/10.1073/pnas.191367098

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lønning, P. E., Brown, P. O., Børresen-Dale, A.-L., & Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America, 100(14), 8418–8423. https://doi.org/10.1073/pnas.0932692100

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660

Ugolini, F., Adélaïde, J., Charafe-Jauffret, E., Nguyen, C., Jacquemier, J., Jordan, B., Birnbaum, D., & Pébusque, M.-J. (2001). Differential expression assay of chromosome 8 genes identifies Frizzled-related (FRP1/FRZB) and fibroblast growth factor receptor 1 (FGFR1) as candidate breast cancer genes. Oncogene, 20(39), 5525–5533. https://doi.org/10.1038/sj.onc.1204707

Van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415(6871), 530–536. https://doi.org/10.1038/415530a

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, 3(7), research0034.1, Article RESEARCH0034. https://doi.org/10.1186/gb-2002-3-7-research0034

Vousden, K. H., & Lane, D. P. (2007). p53 in health and disease. Nature Reviews. Molecular Cell Biology, 8(4), 275–283. https://doi.org/10.1038/nrm2147

Walker, C., Mojares, E., & del Río Hernández, A. (2018). Role of extracellular matrix in development and cancer progression. International Journal of Molecular Sciences, 19(10), 3028. https://doi.org/10.3390/ijms19103028

Wang, Y., Klijn, J. G. M., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M. J. J., Atkins, D., & Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. The Lancet, 365(9460), 671–679. https://doi.org/10.1016/S0140-6736(05)17947-1

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews. Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484

Weigelt, B., Pusztai, L., Ashworth, A., & Reis-Filho, J. S. (2011). Challenges translating breast cancer gene signatures into the clinic. Nature Reviews. Clinical Oncology, 9(1), 58–64. https://doi.org/10.1038/nrclinonc.2011.125

Wierstra, I., & Alves, J. (2007). FOXM1, a typical proliferation-associated transcription factor. Biological Chemistry, 388(12), 1257–1274. https://doi.org/10.1515/BC.2007.159

Yamaguchi, H., & Condeelis, J. (2007). Regulation of the actin cytoskeleton in cancer cell migration and invasion. Biochimica et Biophysica Acta, 1773(5), 642–652. https://doi.org/10.1016/j.bbamcr.2006.07.001

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M. C., Zimmerman, L. J., Shaddox, K. F., Kim, S., Davies, S. R., Wang, S., Wang, P., Kinsinger, C. R., Rivers, R. C., Rodriguez, H., Townsend, R. R., Ellis, M. J. C., Carr, S. A., . . . NCI CPTAC. (2014).

Proteogenomic characterization of human colon and rectal cancer. Nature, 513(7518), 382–387. https://doi.org/10.1038/nature13438

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLOS One, 9(1), Article e78644. https://doi.org/10.1371/journal.pone.0078644

Zhu, C., Zhao, J., Bibikova, M., Leverson, J. D., Bossy-Wetzel, E., Fan, J.-B., Abraham, R. T., & Jiang, W. (2005). Functional analysis of human microtubule-based motor proteins, the kinesins and dyneins, in mitosis/cytokinesis using RNA interference. Molecular Biology of the Cell, 16(7), 3187–3199. https://doi.org/10.1091/mbc.e05-02-0167

Cite this article: Aloqbi AA. DAVID and KEGG-Based Functional Enrichment Analysis of Breast Cancer Transcriptome: Mapping ECM Remodeling and Cell Division Pathways. Int. J. Pharm. Investigation. 2026;16(2):625-37.