ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 22(8): 4781-4786 (2026)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 22(8)

Research Article

Received August 1, 2026; Revised August 31, 2026; Accepted August 31, 2026, Published August 31, 2026

DOI: 10.6026/973206300224781

SJIF 2026 (Scientific Journal Impact Factor for 2026) = 8.478
2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Edited by P Kangueane

Citation: Kumar *et al.* Bioinformation 22(8): 4781-4786 (2026)

Elucidation of putative key genes involved in triple-negative breast cancer progression

Alok Kumar[1], Gauri Shankar Upadhyay[1], Mohammad Kashif[2], Md. Zubbair Malik[2], Naidu Subbarao[2] & Maitreyi S. Rajala[1,*]

[1]School of Biotechnology, Jawaharlal Nehru University, New Delhi, India; [2]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India; *Corresponding author

Affiliation URL:
https://www.jnu.ac.in/Faculty/msrajala/

Author contact:
Alok Kumar - E-mail: iamaloklalchand@gmail.com
Gauri Shankar Upadhyay - E-mail: gsupadhyay060801@gmail.com
Mohammad Kashif - E-mail: kashifjmi.bioinfo@gmail.com

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(8): 4781-4786 (2026)

Md. Zubbair Malik - E-mail: zubair.bioinfo@gmail.com
Naidu Subbarao - E-mail: nsrao.jnu@gmail.com
Maitreyi S. Rajala - E-mail: msrajala@mail.jnu.ac.in

Abstract:
The molecular basis of TNBC development is poorly understood. Therefore, it is of interest to identify key genes and pathways involved in TNBC development and progression using a systems biology approach, followed by experimental validation. Gene ontology and KEGG enrichment analyses were performed to identify DEG-regulated biological functions and pathways; a PPI network was constructed using the STRING online database. The expression and prognostic value of the hub genes were validated by TCGA survival analysis, resulting in the identification of 727 DEGs (473 downregulated and 254 upregulated) in TNBC samples. The GO and KEGG analyses indicated that the DEGs were mainly related to cell adhesion and tumorigenesis and the PPI network showed six hub genes: CCND1, CDH1, ESR1, FN1, IL6 and PPARG as the top key genes. These were validated by real-time PCR in the TNBC cell line using a non-TNBC cell line as a calibrator and the obtained results were in accordance with the bioinformatics data.

Keywords: Triple negative breast cancer (TNBC), non-TNBC, systems biology, Kyoto Encyclopaedia of genes and genomes (KEGG) analysis, expression profiling, QPCR

Background:
Triple negative breast cancer (TNBC) is a subtype of breast cancer characterized by the absence of human epidermal growth factor receptor 2 (HER2) amplification and lack of progesterone receptor (PR) and estrogen receptor (ER) expression **[1]**. Due to these reasons, no specific therapy exists for TNBC tumors besides chemotherapy, immunotherapy and surgery **[1].** High-throughput technologies, including microarray and RNA sequencing, have provided a novel understanding of the molecular complexity of TNBC **[2]**. Several studies have reported the expression of genes such as EGFR, KRT17, KRT14 and KRT5 in TNBC cells that are usually characteristic of normal myoepithelial/basal cells **[3]**. Other molecular features of TNBC include deregulated cell cycle, inactivation of BRCA1 and BRCA2, aberrant activation of proliferative signal pathways and a high frequency of mutations in tumor suppressor genes **[4]**. Therefore, it is of interest to identify novel therapeutic targets and to enhance our understanding of TNBC development and progression in order to develop effective treatment options.

Materials and methods:
Data retrieval:
TNBC datasets were retrieved from GEO database (https://www.ncbi.nlm.nih.gov/geo/, accessed on 20 December 2022). Two human transcriptomic datasets, GSE27447 (n=12) and GSE39004 (n=73) were taken for the study. GSE27447 dataset consisted of 5 TNBC and 7 non-TNBC human tissue samples, while GSE39004 dataset consisted of 30 TNBC and 47 non-TNBC human tissue samples.

Data pre-processing and identification of DEGs:
Patient datasets were analysed using GEO-2R (www.ncbi.nlm.nih.gov/geo/geo2r/). Two groups were defined as non-TNBC and TNBC; p-value was adjusted to Benjamini and Hochberg false discovery rate method with a significance level of 0.05 and auto-detect log transformation was performed group wise. The DEGs were selected by a Bioconductor software package called LIMMA in R program language using linear models from the normalized data of TNBC and non-TNBC tissues. The DEGs were visualised in a volcano plot using the R package ggplot2. The BRCW computing website (http://jura.wi.mit.edu/bioc/tools/compare.php_was used to choose unique DEGs that were shared by at least two gene expression profile datasets.

Gene ontology analysis:
The DEGs functional interpretation was evaluated and visualized in Database for Annotation, Visualization and Integration Discovery (DAVID) **[5]** . For the metabolic pathway enrichment study, the KEGG **[6]** tool was used.

Protein-protein (PPI) network construction and analysis:
The STRING database was used to extract interconnected genes to create a network of PPI **[7]**. To visualize PPI, the tool Cytoscape version 3.8.2 was used **[8]**. Cytohubba **[9]** and the Network Analyzer plugin of Cytoscape was used to calculate the topological properties of PPI network, namely Degree (k), Betweenness centrality ($C_{B)\ and}$ Bottleneck (B_N).

Genetic alteration and validation of key genes expression paradigm:
The GEPIA data tool was used to see whether the expression of key genes was linked to the survival of breast cancer patients from the TCGA data. The breast cancer patients were divided into two classes based on median gene expression values. Their overall survivals (OSs) were evaluated using the Kaplan-Meier approach with a log-rank test for obtaining the survival plot of the key gene **[10]**.

Validation of key gene expression in TNBC patients using UALCAN database:
The expression of key genes was further validated using the UALCAN web portal (https://ualcan.path.uab.edu/index.html), which derives transcriptome data from TCGA and allows analysis of gene expression across various cancer subtypes **[11]**. The gene

expression values were reported as transcripts per million (TPM).

Quantitative real time PCR:

To experimentally validate the expression patterns of the identified hub genes, quantitative real-time PCR (qRT-PCR) analysis was carried out utilizing two breast cancer cell lines that represent distinct molecular subtypes: MCF7 and MDA-MB-231. Total RNA was isolated from MDA-MB-231 and MCF-7 cells cultured in DMEM with 10% FBS by Trizol (Thermo Fisher) method. A fraction of RNA was reverse transcribed and 100ng of cDNA was subjected to real time PCR using specific primers (**CCND1**-FP: 5′TCTACACCGACAACTCCATCCG 3′, RP: 5′ TCTGGCATTTTGGAGAG GAAGTG 3′; **ESR1**- FP: 5′AGACAGCCACTCACC TCTTCAG 3′, RP: 5′TTCTGCCAGTGCCTCTTTGCTG 3′; **CDH1**- FP: 5'GCCTCC TGA AAAAGAGAGTGGAAG 3′, RP: 5′TGG CAGTGTCTCTCCAAATCCG 3'; **FN1**- FP: 5′ACAACACCGA GGTGACTGAGAC 3′, RP: 5′GGACACAACGATGCTTCCTGAG 3′; **PPARG**- FP: 5′AGCCTGCGAAAGCCTTTTGGTG 3′, RP: 5′GGCTTCACATTCAGCA AACCTGG 3′; **IL-6-** FP: 5′AGA CAG CCA CTCACCTCTTCAG 3′, RP: 5′TTCTGCCAGTGCCTCTTTGCTG 3′; **GAPDH**- FP: 5′TGC ACCA CCA ACTG CTT AGC 3′, RP: 5′GGC ATGGACTGTGGTCATGAG 3′)) to the identified key genes by the SYBR green method. Each sample was run in technical triplicates and the mean Ct value was taken for each set of reactions. The relative expression levels of hub genes were quantified using the $2^{-\Delta\Delta Ct}$ method, with the expression in MCF7 cells serving as the reference baseline.

Results:

A total number of 22193 and 11219 annotated transcripts obtained from the datasets GSE27447 and GSE39004_GPL6244 respectively was taken for evaluation. A total of 368 downregulated and 343 upregulated genes were selected in the GSE27447 dataset on the basis of selection criteria (p-value < 0.05 and [log2FC] > +1, [log2FC] < −1) in TNBC and compared to non-TNBC samples. Subsequently, 473 downregulated and 254 upregulated genes were identified in the GSE39004_GPL6244 dataset. A comparison of the complete genes and top 100 genes expression profile of TNBC and non-TNBC patient samples was shown in a graded manner by heatmap construction. The distribution of the DEGs in each dataset with the heatmap. In the GO analysis, the biological process category showed the genes that are involved in immune response were significantly downregulated. Under the cellular component, membrane and cytoskeletal elements were significantly downregulated. While under the molecular function category, downregulated genes were identified in association with cellular and immunological signalling. According to KEGG pathway enrichment analysis, the down regulated genes were significantly enriched in pathways associated with cancer, infections and diseases. Genes that were identified to be upregulated under biological process category were associated mainly with signal transduction, cell adhesion, migration and cellular response to external factors. Under the cellular component category, the up regulated genes were in association with non-nuclear components of the cell. For the molecular functions category, up regulated genes in TNBC were enriched in various metabolic pathways. A total of 1810 DEGs were submitted to STRING for the PPI construction and analysis of the network which had shown that it has 1608 interacting nodes and 15524 edges **(Figure 1A)**. Further analysis that was executed using Cytoscape had shown 20 top hub genes. The top 20 higher-grade bottleneck genes in TNBC PPI regulatory network were also identified (**Figure 1B**). To identify key genes in the PPI network, top 20 genes from each of the three topological properties were taken to compare with each other. After analysing the topological properties (degree, bottleneck and Betweenness centrality) and tracing them to top 20-degree (hub genes) CDH1, PPARG, FN1, CCND1, IL6 and ESR1 were found to be common genes for all topological properties suggesting the significance of these 6 genes in TNBC. The Venn diagram shows common genes of topological properties (**Figure 1C to 1F**). The expression pattern of identified key genes in general breast cancer tissue samples was predicted from mRNA expression using GEPIA (http://gepia.cancer-pku.cn/) online server. The expression pattern of CDH1, FN1, CCND1 and ESR1 was found to be significantly up regulated in breast cancer tissue samples as compared to that of non-cancerous tissue samples while IL6 and PPRAG were down regulated. The mRNA expression of CCND1, ESR1, IL6 and PPARG genes was found to be related to overall survival. The Kaplan Meier (KM) plot also showed that these genes were significantly associated with breast cancer. Thus, these genes can be correlated with increasing or decreasing risk of TNBC progression. The expression levels of CCND1, CDH1, ESR1, FN1, IL6 and PPARG were assessed across various molecular subtypes of breast cancer (using the UALCAN web tool), which include luminal, HER2-positive and triple-negative breast cancer (TNBC), in addition to normal breast tissues. The analysis uncovered subtype-specific variations in gene expression patterns. Notably, ESR1 expression was significantly diminished in TNBC samples when compared to luminal breast cancer. Likewise, the expression levels of CDH1 and PPARG were found to be lower in TNBC samples in comparison to luminal tumors. Conversely, FN1 and IL6 exhibited relatively elevated expression in TNBC compared to luminal breast cancer subtypes. CCND1 expression was detected across several breast cancer subtypes, showing moderate variation among the groups **(Figure 2A)**. The results from the qRT-PCR analysis revealed significant subtype-associated differences in gene expression between the two cell lines (**Figure 2B**). Notably, IL6 expression was higher in MDA-MB-231 cells compared to MCF7 cells, indicating an increase in inflammatory signalling within the TNBC model. Likewise, FN1 expression was elevated in MDA-MB-231 cells, suggesting enhanced interaction with the extracellular matrix. Conversely, the expression levels of CDH1 and ESR1 were markedly diminished in MDA-MB-231 cells in comparison to MCF7 cells. CCND1 expression exhibited a moderate increase in MDA-MB-231 cells,

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(8): 4781-4786 (2026)

indicating possible changes in cell cycle regulation. Interestingly, PPARG expression was found to be elevated in MDA-MB-231 cells relative to MCF7 cells. Although this expression pattern diverges from the subtype distribution noted in TCGA datasets, such discrepancies may reflect variations between patient tumor samples and *in vitro* cell line models, including tumor heterogeneity and the metabolic adaptations that occur in cultured cancer cells.

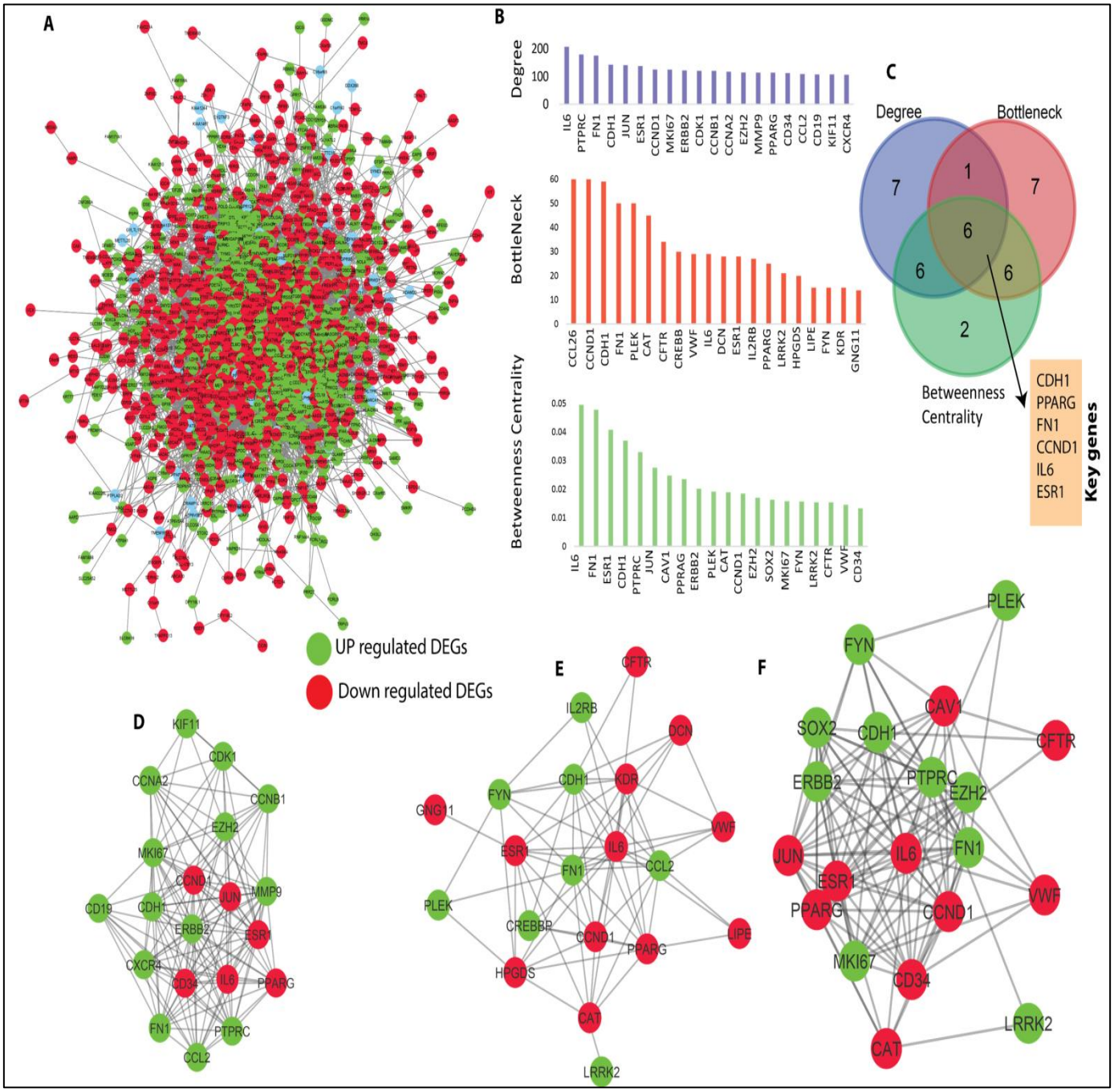

Figure 1: PPI regulatory network of TNBC DEGs (A) PPI network of DEGs. Red and green colour nodes show down-regulated and up-regulated genes of TNBC versus non-TNBC, respectively. (B) The bar plot shows topological properties (degree, bottleneck and betweenness). (C) Venn diagram shows the number of common genes among various centralities of the top 20 genes' topological parameters, i.e., degree, bottleneck and betweenness centrality. (D-F) Top hubs, bottleneck and betweenness centrality genes of the regulatory network retrieved from the PPI network of TNBC.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(8): 4781-4786 (2026)

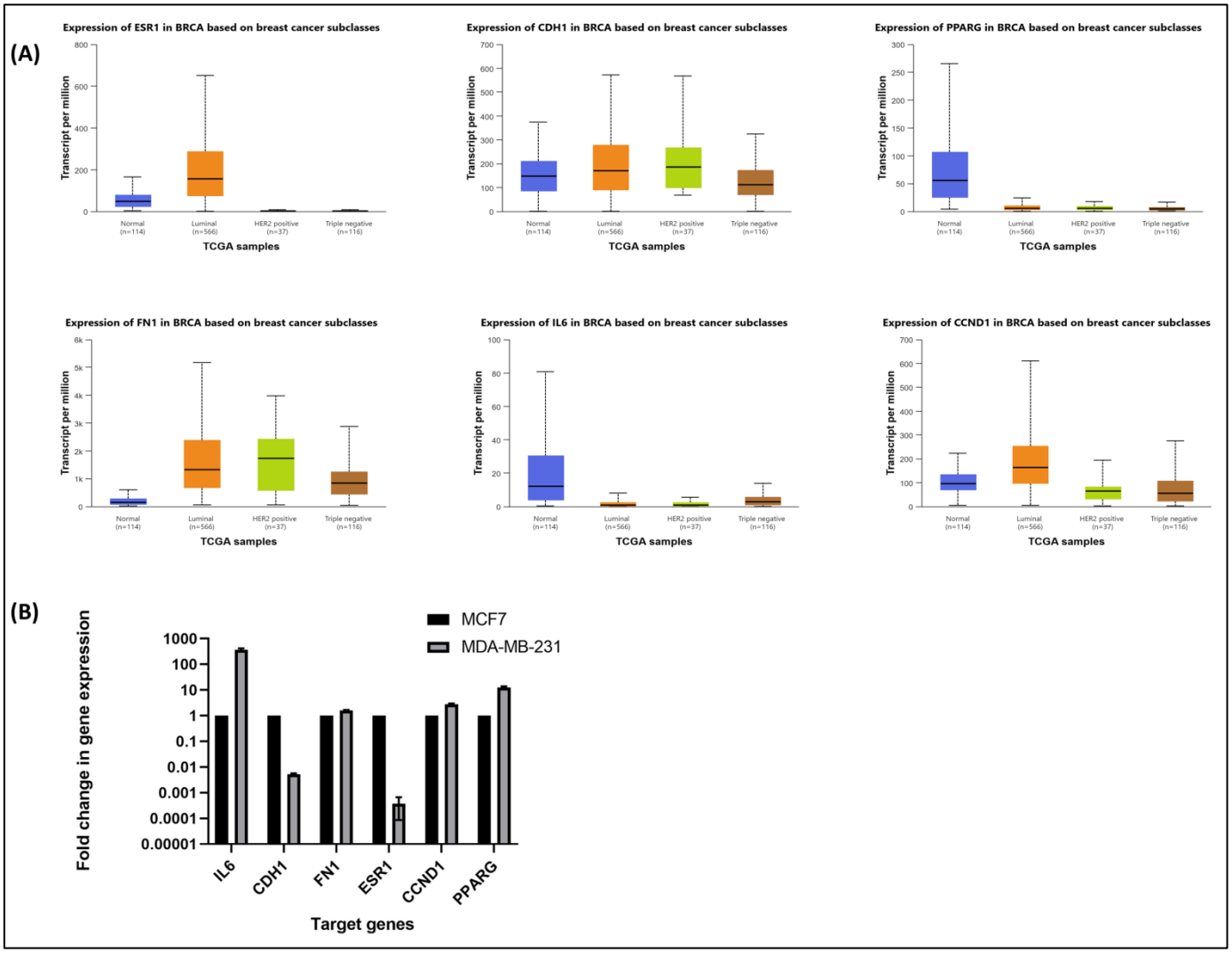

Figure 2: Validation of identified hub genes **(A)** Subtype-specific expression patterns of hub genes in breast cancer (**B)** Validation of hub gene expression at transcript level.

Discussion:

Network analysis and functional enrichment are well-known techniques to find novel potential key genes for any biological network **[12]**. In our study, DEGs between TNBC and non-TNBC tissues of patients in two datasets from GEO were analysed. This study highlights the genes associated with changes in cellular structure and disease progression. Predicting good candidate genes before experimental analysis will save time and effort. Using the network analysis, CCND1, CDH1, ESR1, FN1, IL6 and PPARG were identified to be key players involved in TNBC pathogenesis. CCND1 (Cyclin D1) plays a role in cell cycle regulation and has been reported to be overexpressed in various tumors, including TNBC tumors. Association of higher expression of CCND1 with the aggressive nature of TNBC and poor prognosis was also reported **[13]**. CDH1 (E-cadherin) is another gene identified to be a key regulator in the maintenance of epithelial tissue integrity. It is involved in the regulation of cell-cell adhesion, cell-matrix interactions and is responsible for mediating calcium-dependent interactions between adjacent cells. Loss of CDH1 expression is associated with various cancers, including breast, gastric and ovarian cancer **[14]**. It is used as a biomarker to classify the subtype and also determines the patient outcome. ESR1 is a well-established therapeutic target in hormone receptor-positive breast cancer, where it is targeted with hormone therapy such as tamoxifen or aromatase inhibitors **[15]**. FN1 is being investigated as a therapeutic target in breast cancer, as inhibition of its expression or function may reduce cancer cell migration and invasion **[16]**. IL-6 has been shown to promote breast cancer cell proliferation and survival; targeting the IL-6 pathway with drugs such as tocilizumab or siltuximab is under investigation as a potential therapy to treat breast carcinoma **[17]**. PPARG is also being investigated as a therapeutic target in breast cancer; the inhibition of its

expression or function may reduce cancer cell proliferation and promote survival **[18]**.

Conclusion:
The current study identified multiple important genes, differentially expressed in TNBC patients versus controls through DEG analysis and network building, suggesting their significance as diagnostic and prognostic markers. Considering the importance of the identified genes in TNBC, this study proposes the use of these markers in combination with other clinical, pathological and imaging data for a patient-specific treatment regimen. These biomarkers are not only being investigated in breast cancer but also in other cancers, as they may have different implications depending on the type and the stage of the disease.

Funding:
Financial support was received from the Anusandhan National Research Foundation (ANRF/PAIR/2025/000029/PAIR-A), New Delhi.

Authors Contributions:
Alok Kumar: Performed the bioinformatic studies and analysis, wrote the first draft. Mohammad Kashif, Md. Zubbair Malik: Performed the analysis; Gauri Shankar Upadhyay: Performed gene validation studies *in vitro*; Naidu Subbrao: Conceptualization; Maitreyi S. Rajala: Conceptualization, writing, review and editing.

Ethics approval: Not applicable

Declaration of competing interest: The authors declare no competing interest

Pre-print declaration:
This manuscript has been previously posted as a preprint on *bioRxiv* and has not undergone peer review. The preprint is available at: bioRxiv 2026.04.15.718835; doi: https://doi.org/10.64898/2026.04.15.718835

References:
[1] Zagami P & Carey LA. *NPJ Breast Cancer.* 2022 **8:**95. [PMID: 35987766]
[2] Banerji S *et al. Nature.* 2012 **486:**405. [PMID: 22722202]
[3] Parker JS *et al. Journal of the American Society of Clinical Oncology.* 2009 **27:**1160. [PMID: 19204204]
[4] Anders CK *et al. American Society of Clinical Oncology Educational book.* 2016 **35:**34. [PMID: 27249684]
[5] Huang da W *et al. Nat Protoc.* 2009 **4:**44. [PMID: 19131956]
[6] Kanehisa M & Goto S *Nucleic Acids Research.* 2000 **28:**27. [PMID: 10592173]
[7] Szklarczyk D *et al. Nucleic Acids Res.* 2015 **43**:D447. [PMID: 25352553]
[8] Shannon P *et al. Genome Research.* 2003 **13:**2498. [PMID: 14597658]
[9] Chin CH *et al. BMC Systems Biology.* 2014 **8:**S11. [PMID: 25521941]
[10] Lánczky A & Gyorffy B. *J Med Internet Res.* 2021 **23:** e27633. [PMID: 34309564]
[11] Chandrasekhar DS *et al. Neoplasia* 2022 **25:**18. [PMID: 35078134]
[12] Malik MZ *et al. BMC Cancer.* 2019 **19:**1129. [PMID: 31752757]
[13] Umekita Y *et al. Int J Cancer.* 2002 **98:**415. [PMID: 11920593]
[14] Mahler-Araujo B *et al. J Clin Pathol.* 2008 **61:**615. [PMID: 18156431]
[15] Brett JO *et al. Breast Cancer Res.* 2021 **23:**85. [PMID: 34392831]
[16] Lin YC *et al. Ecotoxicol Environ Saf.* 2025 **306:**119303. [PMID: 41166853]
[17] Heo TH *et al. Oncotarget.* 2016 **7:**15460. [PMID: 26840088]
[18] Panigrahy D *et al. Cancer Biol Ther.* 2005 **4:**687. [PMID: 16082179]