Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deciphering the recurrence-associated coding and non-coding genes in triple-negative breast cancer using an integrated omics approach

Ganesh S. Bose
Krishna Vishwa Vidyapeeth (Deemed to be University)
Richa Kumari
Krishna Vishwa Vidyapeeth (Deemed to be University)
Prajakta Reur
Krishna Vishwa Vidyapeeth (Deemed to be University)
Smriti PK. Mittal
Savitribai Phule Pune University
Abhijeet P. Kulkarni
Savitribai Phule Pune University
Suresh J. Bhosale
Krishna Vishwa Vidyapeeth (Deemed to be University)
Sumit Aggarwal
Indian Council of Medical Research
Pragati Agarwal
Indian Council of Medical Research
Kailas D. Datkhile
hodgeneticslab@kimskarad.in

Krishna Vishwa Vidyepeeth (Deemed to be University)

Research Article

Keywords: Recurrent TNBC, DEGs, GSEA, hub genes, miRNA, biomarker

Posted Date: July 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10095883/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Breast cancer exhibits significant molecular heterogeneity, with triple-negative breast cancer (TNBC) being the most biologically complex subtype, having a higher probability for recurrence and presenting greater clinical challenges. The increased mortality rate of BC patients in India might be linked to a higher prevalence of TNBC. In this study, transcriptome analysis of 24 breast tumor tissues and their corresponding adjacent tissues was conducted to understand the differential expression profile of the study cohort, which revealed 9,049 significantly differentially expressed genes (DEGs). To identify TNBC recurrence-specific molecular alterations, whole-exome sequencing and transcriptomic analyses of representative 4 recurrent and 4 non-recurrent TNBC patients were performed. The expression data identified a total of 913 DEGs in recurrent TNBC patients. Network analysis of protein-coding DEGs identified KNL1, KIF14, CKAP2L, SPC25, MCM10, and GTSE1 as key recurrence-associated hub genes. Furthermore, the identified hub genes, CKAP2L, KIF1A, KIF4A, KIF14, and MYL4 were analyzed for their prognostic significance using Kaplan-Meier Plotter based on TCGA-TNBC dataset, which revealed poor patient survival due to their elevated expression in recurrent-TNBC patients. In contrast, lower expression levels of CNTNAP2, MYL7, SNAP25, LTF, and PDCD1 showed the worst survival. Additionally, two non-coding RNAs, miR-635 and miR-647, with a significant level of their increased expression, were identified in recurrent TNBC patients. In summary, the combined analysis of exome and transcriptome data revealed pathways and central hub genes associated with recurrence, which may have prognostic and diagnostic value in patients with recurrent TNBC. These genes may be used as biomarkers to evaluate the risk of recurrence and possible targets for treatment strategies.

Introduction

Breast cancer (BC) continues to remain one of the most frequently diagnosed cancers in India, with a global estimate of 1,92,020 incidences and 98,337 reported mortalities (1). Given the increased incidence of this disease, it is crucial to identify the complex nature of BC to develop effective therapeutic strategies that significantly rely on the presence or absence of hormone receptors, estrogen (ER), progesterone (PR), and human epidermal growth factor receptor 2 (HER2) (2). Among the molecular subtypes, triple-negative breast cancer (TNBC) accounts for 15–20% of all invasive BC (3, 4). It is highly aggressive and clinically challenging, with a higher prevalence, which may contribute to the higher mortality rate in India. An Indian meta-analysis reported a TNBC prevalence of 27%–35%, with higher histological grade (grade III) disease, larger tumors, greater lymph node involvement, premenopausal cases, and often diagnosed at a younger age (5, 6). It is reasonable to assume that there is a substantial degree of intrinsic variance at the state level that is not sufficiently reflected by India's regional divisions. Hence, a more stratified approach is required to thoroughly understand the higher prevalence of TNBC in the Indian population (6).

The clinical complexity of TNBCs, such as a lack of hormone receptors and appropriate treatment strategies, leaves patients unresponsive to treatments such as chemotherapy, hormonal therapy, and immunotherapy. This promotes the growth of therapy-resistant cell populations that can lead to the re-

emergence of tumors, a phenomenon called tumor recurrence (7, 8). According to global research, approximately 30% of women diagnosed with BC experience recurrence following primary treatment. Among BC subtypes, TNBC patients have a threefold higher tendency to develop disease recurrence at distant sites, most commonly in the lymph nodes, lungs, and brain (9, 10). A study conducted in an Asian cohort demonstrated that TNBC patients exhibited a significantly elevated risk of locoregional recurrence compared to other subtypes (11). Therefore, a comprehensive characterization of the TNBC molecular landscape is essential for elucidating recurrence-specific driver mechanisms.

Currently, next-generation sequencing (NGS) has evolved as a promising tool for elucidating the molecular profile of tumors, enabling the identification of drivers of progression, resistance, and recurrence (12). Given the elevated risk of metastasis and recurrence in TNBC, molecular signatures and genomic footprints associated with chemoresistant responses can be evaluated systematically (13). Focusing on the Indian scenario, recurrent mutations in some BC-specific crucial genes, such as ABCC3, BRCA1, BRCA2, BIRC6, CDH7, CSMD3, MUC12, MUC16, NT5C1B, PIK3CA, POLESTK31, TTN, and ZFHX4, highlighted the role of mutations in Indian TNBC patients (14). Further, in context with transcriptome analysis, clinically actionable frequent alterations in TP53 and PIK3CA have been identified in BC patients, thereby implicating dysregulated pathways of cell cycle regulation and DNA repair (15). Although NGS has significantly progressed in BC research globally, transcriptomic studies involving Indian patients remain limited. Additionally, there is a lack of significant information on the availability of molecular profiles and recurrence-specific sequencing studies in rural Indian populations, one of them being the rural Maharashtrian population (16). Therefore, developing a detailed profile of sequenced data from rural Indian recurrent BC patients will help to understand population-specific differences. In this context, a comprehensive study incorporating genomic and transcriptomic analyses was conducted on a rural population of Maharashtra. This study identified candidate coding and non-coding molecular biomarkers for the early, economically efficient diagnosis and prognostic evaluation of recurrent TNBC.

Materials and methods

Patient enrolment, sample collection, and clinical data collection

Primary tumors and their paired adjacent tissues from BC patients (N = 24) were retrieved from the Krishna Breast Tumor Bank (KBTB), Department of Molecular Biology and Genetics, Krishna Institute of Medical Sciences, Krishna Vishwa Vidyapeeth, Karad, India. The clinicopathological and demographic information of all BC patients, along with informed consent, was collected. Patient confidentiality was ensured by assigning KBTB identifiers to each patient.

RNA and DNA extraction, library preparation, and sequencing

The HiPurA Multi-Sample RNA Purification Kit (HiMedia Laboratories, Cat. No. MB619) was used to extract RNA from cancer tissue (CT) and adjacent tissues (AT). Qubit-4-Fluorometer (Invitrogen, Thermo Fisher Scientific, Model No. Q33226) was used for the quantification of the extracted RNA, and its quality was checked on QSep1 Plus Bio-Fragment Analyzer (Bioptic, Model No. C100001-P) and only RNA with RNA Quality Number ≥ 7 was selected for library preparation. Illumina Stranded mRNA Prep Ligation Kit (Illumina, Cat. No 20040898) was used to prepare the RNA libraries. Similarly, DNA of BC patients were extracted from both CT and AT using the HiPurA Mammalian Genomic DNA Purification Kit (HiMedia Laboratories, Cat No. MB506) following manufacturing guidelines. The quantity and quality of DNA were evaluated, and high-quality DNA was selected for library preparation using the Twist Library Preparation EF Kit 1.2.0 (Twist Bioscience, Cat No. 104207) and Twist Universal Adaptor System (Twist Bioscience, Cat No. 101311) according to the guidelines provided by the manufacturer. Sequencing libraries were loaded on an Illumina NextSeq 1000 (Illumina, Inc., Model no. V200987) platform and subjected to 150 bp paired-end reads for exome sequencing and 100 bp paired-end reads for RNA sequencing The FASTQ files were generated after sequencing and were used for *in-silico* analysis.

Exome sequence analysis

Following a FastQC quality assessment, raw paired-end FASTQ (v0.12.1) data were aligned with the human reference genome (GRCh38) using the BWA-MEM (v0.7.17) (17). Duplicate reads were obtained using Picard, and base quality scores were recalculated using the Genome Analysis Toolkit (GATK) (18). GATK mutect2 (v4.6.0.0) tool was used for the Somatic variant identification, and the called somatic variants were annotated through the Ensembl Variant Effect Predictor (VEP) software (v112) (19). Mutation annotation format (MAF) files were generated from the Variant Call Format (VCF) files for downstream analysis, such as somatic mutation profiling, mutation burden estimation, and determination of recurrently mutated genes, using Maftools (R Bioconductor package) (20).

RNA sequence analysis

FastQC (v0.12.1) tool was used to evaluate the quality of raw paired-end FASTQ files obtained after sequencing. STAR Aligner (v2.7.11b) was used to align the reads with the human reference genome, and aligned BAM files were obtained which were further used to generate gene-level read counts using featureCounts (v2.0.6). Gene-level annotation was carried out using GENCODE (v45) assembly. Genes with lower grade levels were filtered out before conducting the analysis. DESeq2 (Bioconductor package) was used for analyzing the differential gene expression levels between the CT and AT (21). Gene expression changes for recurrent and non-recurrent samples were calculated for each gene as Δ-expression (CT – AT) using normalized expression values. Differential analysis of Δ-expression values between recurrent and non-recurrent patients was performed using the limma Bioconductor package with empirical Bayes moderation (22).

Pathway enrichment of DEGs using Gene Set Enrichment Analysis (GSEA)

GSEA was carried out through MSigDB Hallmark pathway gene sets on a pre-ranked gene list derived from differential expression data. The significantly enriched pathways were identified on the basis of normalized enrichment scores (NES) (23).

Network analysis of protein-coding DEGs using Cytoscape

To evaluate the functional interactions among the identified genes, a protein-protein (PPI) network of recurrence-associated candidate genes was built using the STRING database and imported to Cytoscape for downstream visualization and topological analysis. CytoHubba plugin of Cytoscape was used for determining the hub genes. Network topology parameters, such as degree centrality, were used to find highly ranked genes. Hub genes with strong connections to recurrence-related biological pathways were determined. Network visualization and prioritization of hub genes were performed using the default parameters unless otherwise specified.

Pathway enrichment analysis of miRNA target genes

Target genes of miRNA were retrieved using miRDB and subjected to pathway enrichment analysis using the Enrichr web server (https://maayanlab.cloud/Enrichr/). The KEGG and MSigDB Hallmark gene set library was used for the gene pathway enrichment analysis. Enriched pathways were ranked on the basis of Enrichr combined score, and statistical significance was set to p-value (FDR) < 0.05 (24).

Survival analysis using Kaplan-Meier plotter

Kaplan-Meier plotter was used to evaluate the prognostic significance of the hub genes, using the TCGA dataset. Survival analysis was restricted to TNBC samples by selecting ER^-, PR^-, and $HER2^-$ patients, and statistical significance was calculated using the log-rank test, considering $p < 0.05$ as statistically significant (25).

Statistical analysis

The statistical test of the data was conducted using R (4.3.3) packages. The relationship between gene-level mutation frequency and expression was assessed using Spearman's rank correlation test, which is robust compared to non-linear connections and non-normally distributed data. The correlation coefficient (ρ) and associated p-value are displayed, with a $p < 0.05$ considering statistically significant.

Results

Clinicopathological and demographic data of BC patients

Table 1 provides a concise overview of the 24 BC patients, with a median age of 59 years, ranging from 36 to 80 years. The majority of the study population were above the age of 50 years, with a body mass index (BMI) of < 25, and 66.67% of postmenopausal cases. Histopathological analysis revealed that half of the population were diagnosed with stage III tumors, and 15 of 24 patients (62.50%) had lymph node metastasis. The addiction history of the patients showed that 58.33% of the study population were addicted to tobacco usage. The data highlighted the predominant occurrence of the TNBC subtype (45.83%) compared with other BC subtypes. Among the documented cases of tumor recurrence, 4 patients experienced recurrence of the TNBC subtype.

Table 1
Clinicopathological and demographic data of BC patients. The table summarizes patient-related parameters associated with tumor progression and disease pathology.

Demographic and Clinicopathological Parameter		Number of Patients (N = 24)	Percentage
Age in years	Median age = 59		
	≤ 50	8	33.33
	> 50	16	66.67
BMI (kg/m2)	< 25	15	62.50
	> 25	9	37.50
Menopausal Status	Premenopausal	8	33.33
	Postmenopausal	16	66.67
Tobacco Habit	Yes	14	58.33
	No	10	41.67
Tumor Laterality	Right	13	54.16
	Left	11	45.83
Pathological Lymph Node Status	N0	9	37.50
	N+	15	62.50
Tumor Stage	IA	1	4.17
	IIA	5	20.83
	IIB	6	25.00
	IIIA	5	20.83
	IIIB	3	12.50
	IIIC	4	16.67
Tumor Grade	I	1	4.17
	II	9	37.50
	III	14	58.33
Molecular Subtypes	ER+/PR+/HER2-	7	29.17
	ER+/PR+/HER2+	3	12.50
	ER-/PR-/HER2+	1	4.17

Demographic and Clinicopathological Parameter		Number of Patients (N = 24)	Percentage
	ER-/PR-/HER2-	11	45.83
	ER+/PR-/HER2+	1	4.17
	ER+/PR-/HER2-	1	4.17
Recurrence Status	Recurrent	4	16.67
	Non-Recurrent	20	83.33

Transcriptome of BC patients and KEGG pathway analysis of DEGs

Transcriptomic analysis of 24 BC tumors and their paired adjacent tissues was carried out. The volcano plot revealed a clear separation of 9,049 significantly differentially expressed genes (DEGs), with 4,289 downregulated genes on the left side and 4,760 upregulated genes on the right. Among these, 7,204 genes were protein-coding DEGs, of which 3,388 were downregulated, and 3,816 were upregulated. A total of 1,431 non-coding RNAs were identified, of which 763 were downregulated, and 668 were upregulated (Fig. 1A). Heatmap analysis based on hormone status revealed a distinct pattern of DEGs, highlighting the substantial differences in expression between cancer and adjacent tissues. Among these, PLIN4, PLIN1, GPD1, and HSPB7, which play roles in tumor suppression, were found to be downregulated, whereas SOX4, MMP11, MMP13, and CCN4, involved in cancer progression, exhibited considerably higher expression patterns (Fig. 1B). Moreover, KEGG analysis of DEGs was performed to identify their involvement in biological pathways. The analysis identified significant enrichment of key pathways of cell cycle, microRNAs (miRNAs) in cancer, focal adhesion, cellular senescence, and the MAPK signaling pathway (Fig. 1C). Overall, transcriptomic and pathway enrichment analysis highlighted the pattern of gene expression and key biological pathways connected with BC.

Transcriptomic and genomic investigation of recurrent and non-recurrent TNBC patients

To further investigate the recurrence-specific molecular patterns associated with TNBC, 4 recurrent and 4 non-recurrent TNBC patients were selected as representative samples for the subsequent studies. Whole-exome sequencing and transcriptomic analyses were performed on the tumor and paired adjacent tissues. Somatic variant analysis revealed a total of 51,814 variants in recurrent patients and 52,837 variants in non-recurrent patients, with 2,197 variants identical between recurrent and non-recurrent groups (Fig. 2A). Single nucleotide polymorphisms (SNPs) were the commonly observed variant types in both groups, although in the non-recurrent group a slightly greater number of SNPs were identified (Fig. 2B). Among these patients, missense variants were predominantly present, with a high

frequency in non-recurrent patients, followed by nonsense mutations, which were slightly higher in number within the recurrent group, and splice-site variants, which were fairly elevated in non-recurrent patients (Fig. 2C). Additionally, SNV class analysis revealed a higher proportion of C > A transversions in the recurrent group, followed by greater C > T and T > C transitions in non-recurrent patients compared to recurrent patients (Fig. 2D and E). Transcriptomic sequencing was conducted on identical samples, which identified 913 significant DEGs through differential expression analysis. Among these, 536 genes were upregulated, while 377 were downregulated in recurrent TNBC patients. The volcano plot displayed a clear distribution of these genes, with downregulated genes lying on the left side of the plot and upregulated genes on the right side of the plot (Fig. 2F). Moreover, Spearman's correlation test was executed to examine the relationship between mutations and gene expression; however, result suggested no significant correlation observed between the mutation frequency and expression level (Fig. 2G).

Gene expression profiling and GSEA of protein-coding genes in recurrent TNBC patients

A total of 516 significant protein-coding DEGs were identified, of which 269 were upregulated, and 247 were downregulated in patients with recurrence. A few of the upregulated genes were KNL1, KIF4A, KIF14, and HOXD3, while a few of the downregulated genes included PTPRT, DIRAS3, RET, and LTF (Fig. 3A). Furthermore, GSEA of DEGs was conducted to identify the biological pathways enriched in recurrent patients. The data revealed that Hallmark pathways, including Wnt/β-catenin signaling, Hedgehog signaling, Reactive Oxygen Species pathway, and E2F Targets, were positively enriched (Fig. 3B-E). In contrast, pathways including Estrogen Response Late, Estrogen Response Early, Apical Junction, and KRAS Signaling DN were negatively enriched in recurrent TNBC patients (Fig. 3F-I). The findings suggest that the enriched pathways identified could aid in tumor recurrence by enhancing tumor growth, proliferation, stemness, and cancer progression.

Regulatory network analysis of DEGs for hub gene identification in recurrent TNBC patients

A PPI network was generated using the STRING database and the Cytoscape application with an input of 516 significant protein-coding DEGs to determine the functional relationships between DEGs. The Maximal Clique Centrality (MCC) algorithm of CytoHubba was used to determine hub genes linked with tumor recurrence. Genes were assigned MCC scores based on their participation and interconnections within the network. KNL1, KIF14, CKAP2L, SPC25, MCM10, and GTSE1 were identified as top hub genes, which may suggest their potency in TNBC recurrence (Fig. 4).

TCGA-based survival analysis of hub genes from recurrent TNBC patients

Survival probability of the TNBC patients was evaluated using the Kaplan-Meier Plotter with the TCGA-TNBC database. The data showed that higher expression of CKAP2L, KIF1A, KIF4A, KIF14, and MYL4 was linked to poor patient survival (Fig. 5A-E). In contrast, lower expression levels of CNTNAP2, MYL7, SNAP25, LTF, and PDCD1 were marked with the worst survival in the TCGA-TNBC dataset (Fig. 5F-J). Among these genes, CKAP2L, KIF1A, CNTNAP2, MYL7, and LTF showed significant associations between expression and overall survival. The pattern of the expression profile observed was consistent with the recurrence-specific DEGs identified in this study. In particular, DEGs such as CKAP2L, KIF1A, KIF4A, KIF14, and MYL4 were upregulated, while DEGs including CNTNAP2, MYL7, SNAP25, LTF, and PDCD1 were downregulated in recurrent TNBC patients. These genes were also linked with worst survival in the TCGA cohort. (Fig. 5K). Collectively, the findings imply that these genes may act as potential diagnostic and prognostic biomarkers in recurrent TNBC patients.

Pathway enrichment and regulatory network analysis of recurrent specific miRNAs and their predicted target genes

Differential expression analysis also revealed the presence of non-coding DEGs, in which two miRNAs, miR-635 and miR-647, were identified with significantly higher expression in recurrent than in non-recurrent TNBC patients (Fig. 6A and B). The potential regulatory roles of these miRNAs were determined by identifying the predicted target genes using the miRDB target prediction database (20). Thereafter, MSigDB Hallmark pathway enrichment analysis of miR-635 target genes revealed that pathways such as the inflammatory response, androgen response, and mitotic spindle were enriched, while the predicted target genes of miR-647 showed enriched pathways, including KRAS signaling DN, G2-M checkpoint, and KRAS signaling UP (Fig. 6C). Additionally, to determine the interactions among these target genes, a PPI network was constructed using the MCC algorithm in CytoHubba. Network analysis of miR-635 target genes identified hub genes, including PCDHGB1-5, PCDHGA2, PCDHGA4, PCDHGA9, and PCDHGA12 (Fig. 6D). Likewise, network analysis of miR-647 target genes characterized hub genes such as SMAD2, RUNX2, EGF, ENG, SOX9, and FN1 (Fig. 6E). Collectively the data suggest the probable involvement of both the miRNAs in regulating genes that function in cell cycle regulation, cellular growth, and survival.

Discussion

TNBC is particularly known as a highly aggressive and clinically challenging subtype, characterized by the absence of ER, PR, and HER2 (4). Although chemotherapy remains the standard treatment option owing to the restricted number of targeted therapies, most patients carry a greater risk of recurrence (3). Unlike other BC subtypes, TNBC is distinguished by a unique pattern of recurrence, with most relapses occurring within the initial 5 years following diagnosis. This early peak in TNBC recurrence is accompanied by distant metastasis and poor survival outcomes (26). Recently, sequencing technologies have advanced and improved our understanding of disease progression; however, further research is

required to identify the early molecular determinants of recurrence. This will help in the development of more effective therapeutic strategies aimed at enhancing overall survival (27).

The clinicopathological summary of 24 BC patients suggested that 45.8% of the study population had TNBC, among which 4 were of the recurrent type. Therefore, WES and transcriptomic analyses of the 4 recurrent and 4 non-recurrent TNBC patients were performed in the present study. WES showed a higher proportion of nonsense mutations and C > A transversions in the recurrent group. Differential expression analysis identified 913 significant coding and non-coding genes, of which 536 were upregulated, and 377 were downregulated. Nevertheless, somatic mutational events and gene expression did not significantly correlate.

Previous reports on the association between somatic mutations and gene expression have faced numerous challenges, such as differentiating between causal driver effects and neutral-passenger mutations. Additionally, mutational analysis in WES focuses on coding genes, which overlooks non-coding regulatory changes and confounding factors such as copy number variations, methylation, and tumor purity. Most somatic mutations do not exhibit a clear impact on expression, highlighting the need for experimental validation and enhanced multi-omics integration (28–30).

The subsequent study was segregated according to coding and non-coding DEGs. The first section involved identification of differentially expressed protein-coding DEGs. Genes including HOXD3, KIF4A, and KIF14 were upregulated, whereas PTPRT and DIRAS3 were found to be downregulated in recurrent TNBC patients. Earlier studies showed that few of the genes were linked to BC while, none of them have been linked to recurrence (31–34). Furthermore, GSEA of these protein-coding DEGs showed a strong negative enrichment of the estrogen early and late response, apical junction, and KRAS signaling pathways, while a positive enrichment of the Wnt/β-catenin signaling and ROS pathways, suggesting a stronger suppression and activation of the downstream targets of these pathways, respectively in recurrent patients at the transcriptional level. Reported evidence of significant negative enrichment of the estrogen signaling pathway has been documented in early-stage recurrent TNBC patients (35). Previous studies have reported an association between high ROS levels and intratumor heterogeneity, homologous recombination defects, higher mutation rates, clinical aggressiveness, and poorer overall survival in the GSE96058 and METABRIC datasets in BC (36). Similar results have also been evidenced in this study. In addition, Wnt/β-catenin signaling was correlated with reduced overall survival rates and was positively enriched in recurrent TNBC patients, highlighting its potential as a pharmacological target for treating particularly aggressive BC subtypes.

Further, these coding DEGs were evaluated at the translational level by constructing a PPI network of significant protein-coding DEGs, which identified the top 30 hub genes. The validation of the prognostic significance of the hub genes was performed using Kaplan-Meier plotter, among which CKAP2L, KIF1A, CNTNAP2, MYL7, and LTF showed a significant association with patient survival. Notably, the data revealed a consistent expression pattern of the recurrent-specific DEGs with that of the TCGA dataset

survival plot expression pattern, which reinforces the biological relevance of the identified hub genes and their contribution to clinical outcomes in recurrent patients.

Among these, kinesin superfamily (KIFs) are microtubule-based motor proteins that function as ATP-dependent intracellular transporters and play roles in cellular morphogenic pathways and cell division. The increased expression levels of KIFs, including KIF1A, KIF4A, and KIF14, in the recurrent TNBC patients were observed. Prior studies related to the overexpression of these proteins, specifically, KIF4A and KIF14, have been linked to worse overall survival, with distant metastasis-free survival, and relapse-free survival rates. They are known to serve as potential biomarkers in BC (34). Moreover, previous report related to increased KIF1A expression showed docetaxel resistance. (37), highlighting its role in BC progression.

CKAP2L (cytoskeleton-associated protein 2-like), a mitotic spindle protein, is an important component of the centrosome that regulates cell cycle, microtubule dynamics, and chromosome segregation with higher expression levels in various cancers, particularly in BC, with poor clinical outcomes and patient survival. The expression of levels of CKAP2L in recurrent-TNBC patients was also high compared with non-recurrent TNBC patients, indicating its potential role in disease progression. Its inhibition by FOXP3 (forkhead box P3) has been reported in BC. Notably, the recurrent group exhibited low FOXP3 expression in the present study, suggesting its contribution to CKAP2L regulation (38, 39).

LTF (Lactotransferrin) is an iron-binding protein that is usually found in various body fluids, including mammary secretions, neutrophil granules, and epithelial cell secretions. It is involved in several biological functions, such as antimicrobial, anticancer, anti-inflammatory, and immune regulatory mechanisms. Studies related to nasopharyngeal carcinoma have reported the tumor suppressive effect of LTF by suppressing the AKT signaling pathway. In addition, its poor prognosis and enhanced epithelial-to-mesenchymal transition have been reported in clear cell renal cell carcinoma with increased metastatic potential and decreased expression levels. Consistent with the results of this study, LTF downregulation in recurrent tumor samples suggests its function in cancer progression (40, 41).

Cytoskeleton assembly proteins, such as Myosin light chains, are normally dysregulated in cancer pathways and are notably linked with tumor metastasis (42). In this study, MYL4 and MYL7 were differentially expressed in recurrent patients. However, further experimental validation is required to elucidate their roles in TNBC.

The transmembrane protein CNTNAP2 (Contactin Associated Protein 2) functions in cell adhesion. Its expression was marked to be downregulated in recurrent TNBC patients. Recently, it has been reported that its overexpression caused cell cycle arrest, while its downregulation induced proliferative properties (43).

The second section of this study focused on non-coding DEGs that may act as potential contributors to disease pathogenesis and recurrence. They can be considered promising candidates for the development of recurrence-specific biomarkers, given that non-coding genes are notably involved in

controlling various biological pathways. Advancements in NGS have revealed many regulatory miRNAs that are involved in post-transcriptional gene regulation. In the present study, two miRNAs, miR-635 and miR-647, were overexpressed in recurrent TNBC patients compared to non-recurrent TNBC patients. Previous reports of BC indicate that miR-635 expression is increased in breast tumor tissue compared to adjacent tissue, suggesting its role in tumor progression (44). Additionally, involvement of miR-647 in regulating cell proliferation and apoptosis in colon and gastric cancer has been found in previous studies (45, 46). Liu et al. experimentally confirmed NFIX as a target of miR-647, with reduced luciferase activity observed only in the wild-type NFIX 3'-UTR construct, using dual Luciferase assays. Downregulation of NFIX may trigger proliferation, invasion, and migration in colon cancer cells, due to its tumor suppressive activity (46). Moreover, the enriched pathway of predicted target genes demonstrated the participation of both miRNAs in regulating cell cycle, growth, and survival. While comprehensive research on these miRNAs in the context of recurrent TNBC is still lacking, the collective evidence indicates that the identified miRNAs could serve as potential biomarkers, supporting the ideology of effective diagnosis and treatment of TNBC. Although this is a preliminary study, the data obtained hold the potential to deliver crucial insights into TNBC recurrence in the rural Indian population.

Conclusion

In conclusion, we identified coding and non-coding DEGs in recurrent TNBC patients. A PPI network of coding DEGs was constructed to investigate their functional relationships and identify hub genes in recurrent patients. The potency and prognostic significance of these hub genes as candidate biomarkers for recurrence were further confirmed using Kaplan-Meier survival analysis. CKAP2L, KIF1A, CNTNAP2, MYL7, and LTF, along with two miRNAs (miR-635 and miR-647), were identified in recurrent patients. Collectively, these findings suggest a possibility of these genes being potential candidate biomarkers, and with a future perspective of detecting tumor recurrence in TNBC patients. However, further validation with a larger sample size is required.

Declarations

Competing interests:

None of the authors declare conflict of interest.

Funding:

Authors are thankful to the Indian Council of Medical Research (ICMR) for financial assistance to the research project (Grant No. ID: IIRPIG-2024/01595 dated 10th January 2025).

Author Contribution

GB, RK, PR: Primary data curation, Methodology, Visualization, Formal Analysis, Writing Original Draft. KD, SM, AK, SB: Resources and Funding acquisition, Supervision, Original draft preparation, Review & Editing. SA, PA: Review & Editing All authors reviewed the manuscript.

Acknowledgement

We gratefully acknowledge the Indian Council of Medical Research (ICMR) for providing the financial support for this study. We also extend our sincere thanks to Krishna Vishwa Vidyapeeth (Deemed to be University), Karad and Savitribai Phule Pune University (SPPU), Pune, for providing the essential facilities and a conducive environment for the smooth conduct of this research. Authors are thankful to the Indian Council of Medical Research (ICMR) for financial assistance to the research project (Grant No. ID: IIRPIG-2024/01595 dated 10th January 2025).Competing interests: None of the authors declare conflict of interest.

Data Availability

The datasets generated and analysed during the current study are available in the repository National Center of Biotechnology Information-Sequence Read Archive with BioProject accession number for RNA sequence data PRJNA1481169 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA1481169?reviewer=nj49o62mqlvri075qrt376gpra) and BioProject accession number for Exome sequence data PRJNA1481988 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA1481988?reviewer=6efe03n9d5v0amgrgnmq3grpu7).The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Declaration

The study protocol was approved by the Institutional Ethics Committee of Krishna Vishwa Vidyapeeth (Deemed to be University), Karad, and Maharashtra, India (Ethics Committee Approval Certificate: KVV /IEC/Protocol No. 110/2024-2025). Informed consent was obtained from all participants.

References

1. Srivastava TP, Goel I, Gogia A et al. Comprehensive Overview of Breast Cancer in India. JCO Glob Oncol: e2500083, 2025.
2. Xiong X, Zheng L-W, Ding Y, et al. Breast cancer: pathogenesis and treatments. Signal Transduct Target Ther. 2025;10:49.
3. Chen F, Li Y, Qin N, et al. RNA-seq analysis identified hormone-related genes associated with prognosis of triple negative breast cancer. J Biomed Res. 2020;34:129–38.

4. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R. and Stanisławek A: Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. Cancers. 2021;13:4287.
5. Kulkarni A, Kelkar DA, Parikh N, Shashidhara LS, Koppiker CB, Kulkarni M. Meta-Analysis of Prevalence of Triple-Negative Breast Cancer and Its Clinical Features at Incidence in Indian Patients With Breast Cancer. JCO Glob Oncol: 1052–62, 2020.
6. Sandhu GS, Erqou S, Patterson H, Mathew A. Prevalence of Triple-Negative Breast Cancer in India: Systematic Review and Meta-Analysis. J Glob Oncol. 2016;2:412–21.
7. Ahmad A. Pathways to breast cancer recurrence. ISRN Oncol 2013: 290568, 2013.
8. Wu P, Gao W, Su M, Nice EC, Zhang W, Lin J, Xie N. Adaptive Mechanisms of Tumor Therapy Resistance Driven by Tumor Microenvironment. Front Cell Dev Biol. 2021;9:641469.
9. Noman SM, Fadel YM, Henedak MT, et al. Leveraging survival analysis and machine learning for accurate prediction of breast cancer recurrence and metastasis. Sci Rep. 2025;15:3728.
10. Zagami P, Carey LA. Triple negative breast cancer: Pitfalls and progress. Npj Breast Cancer. 2022;8:95.
11. Shim HJ, Kim SH, Kang BJ, Choi BG, Kim HS, Cha ES, Song BJ. Breast Cancer Recurrence According to Molecular Subtype. Asian Pac J Cancer Prev. 2014;15:5539–44.
12. Lips EH, Michaut M, Hoogstraat M, et al. Next generation sequencing of triple negative breast cancer to find predictors for chemotherapy response. Breast Cancer Res BCR. 2015;17:134.
13. Balko JM, Giltnane JM, Wang K, et al. Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. Cancer Discov. 2014;4:232–45.
14. Preeti P, Das AM, Kumar P, et al. Whole-exome sequencing for identification of specific gene mutations in an Indian cohort of triple-negative breast cancer patients. Discov Oncol. 2025;17:20.
15. Gardi N, Chaubal R, Gandhi KA, et al. Clinically actionable alterations in Indian breast cancer patients derived through whole transcriptome sequencing. Indian J Med Res. 2026;163:95–103.
16. Jadhav N, Dravid NV, Kaderi A, Newadkar D, Gadre A, Agrawal S. High prevalance of triple negative markers in breast cancer patients in North Maharashtra. Indian J Pathol Oncol. 2025;10:29–33.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
18. Van der Auwera GA, Carneiro MO, Hartl C et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al 11: 11.10.1–11.10.33, 2013.
19. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.
20. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.

21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
22. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3, 2004.
23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
24. Xie Z, Bailey A, Kuleshov MV, et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021;1:e90.
25. Posta M, Győrffy B. Pathway-level mutational signatures predict breast cancer outcomes and reveal therapeutic targets. Br J Pharmacol. 2025;182:5734–47.
26. Radosevic-Robin N, Selenica P, Zhu Y, et al. Recurrence biomarkers of triple negative breast cancer treated with neoadjuvant chemotherapy and anti-EGFR antibodies. NPJ Breast Cancer. 2021;7:124.
27. Golestan A, Tahmasebi A, Maghsoodi N, Faraji SN, Irajie C, Ramezani A. Unveiling promising breast cancer biomarkers: an integrative approach combining bioinformatics analysis and experimental verification. BMC Cancer. 2024;24:155.
28. Boström M, Larsson E. Somatic mutation distribution across tumour cohorts provides a signal for positive selection in cancer. Nat Commun. 2022;13:7023.
29. Ding J, McConechy MK, Horlings HM, et al. Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. Nat Commun. 2015;6:8554.
30. Jia P, Zhao Z. Impacts of somatic mutations on gene expression: an association perspective. Brief Bioinform. 2017;18:413–25.
31. Dai J, Chen Q, Li G, Chen M, Sun H, Yan M. DIRAS3, GPR171 and RAC2 were identified as the key molecular patterns associated with brain metastasis of breast cancer. Front Oncol. 2022;12:965136.
32. Li L, Xu F, Xie P, Yuan L, Zhou M. PTPRT Could Be a Treatment Predictive and Prognostic Biomarker for Breast Cancer. BioMed Res Int 2021: 3301402, 2021.
33. Zhang Y, Zhang Q, Cao Z, Huang Y, Cheng S, Pang D. HOXD3 Plays a Critical Role in Breast Cancer Stemness and Drug Resistance. Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol. 2018;46:1737–47.
34. Li T-F, Zeng H-J, Shan Z, et al. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. Cancer Cell Int. 2020;20:123.
35. TSAI C-H, CHIU J-H, YANG C-W, et al. Molecular characteristics of recurrent triple-negative breast cancer. Mol Med Rep. 2015;12:7326–34.
36. Oshi M, Gandhi S, Yan L, et al. Abundance of reactive oxygen species (ROS) is associated with tumor aggressiveness, immune response, and worse survival in breast cancer. Breast Cancer Res Treat. 2022;194:231–41.

37. De S, Cipriano R, Jackson MW, Stark GR. Overexpression of Kinesins Mediates Docetaxel Resistance in Breast Cancer Cells. Cancer Res. 2009;69:8035–42.
38. Chi F, Chen L, Jin X, He G, Liu Z, Han S. CKAP2L, transcriptionally inhibited by FOXP3, promotes breast carcinogenesis through the AKT/mTOR pathway. Exp Cell Res. 2022;412:113035.
39. Liu Z, Zhang J, Shen D, Hu X, Ke Z. Ehrich Lister IN and Sihombing B: Prognostic significance of CKAP2L expression in patients with clear cell renal cell carcinoma. Front Genet 13, 2023.
40. Chiu I-J, Hsu Y-H, Chang J-S, Yang J-C, Chiu H-W, Lin Y-F. Lactotransferrin Downregulation Drives the Metastatic Progression in Clear Cell Renal Cell Carcinoma. Cancers. 2020;12:847.
41. Deng M, Zhang W, Tang H, et al. Lactotransferrin acts as a tumor suppressor in nasopharyngeal carcinoma by repressing AKT through multiple mechanisms. Oncogene. 2013;32:4273–83.
42. Kozole SL, Beningo KA. Myosin Light Chains in the Progression of Cancer. Cells 13: 2081, 2024.
43. Liu Y, Zhao J, Wang K, et al. Third-generation whole-genome sequencing reveals the role of CNTNAP2 as a tumor suppressor gene in high-risk neuroblastomas. J Transl Med. 2026;24:135.
44. Zeinali Sehrig F, Zaefizadeh M, Ahmadizadeh C, Alivand MR, Ghorbian S. In Silico and Experimental Outcomes of the Expression of miR508-5p and miR-635 in Tumor Tissues of Patients with Breast Cancer. Iran Red Crescent Med J IRCMJ. 2023;25:e189111.
45. Zhang X, Zhang M, Wang G, Tian Y, He X. Tumor promoter role of miR−647 in gastric cancer via repression of TP73. Mol Med Rep. 2018;18:3744–50.
46. Liu S, Qu D, Li W, et al. miR-647 and miR-1914 promote cancer progression equivalently by downregulating nuclear factor IX in colorectal cancer. Mol Med Rep. 2017;16:8189–99.

Figures

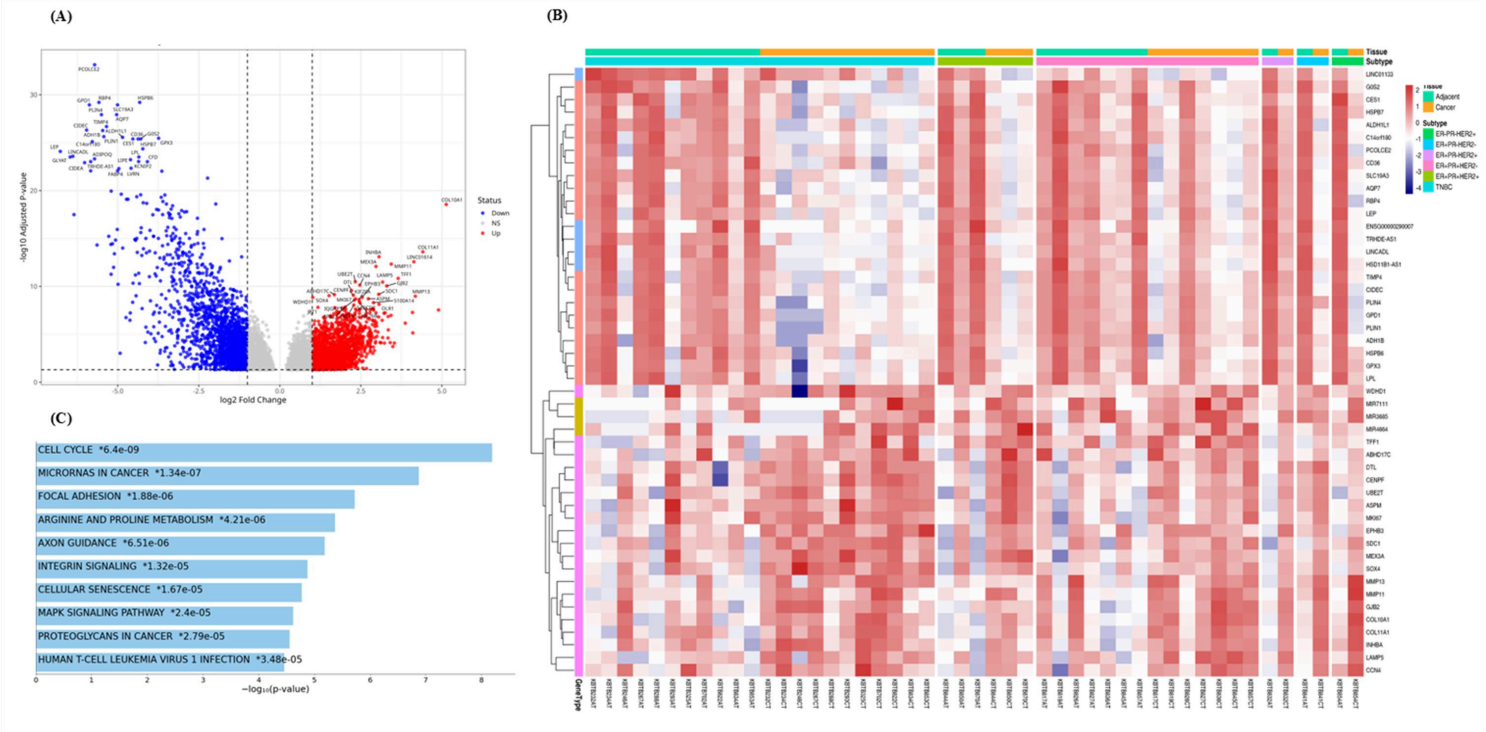


Figure 1

Transcriptome of BC patients and KEGG pathway analysis of DEGs (A) Volcano plot of DEGs. (B) Heatmap showing the expression patterns of the DEGs. (C) KEGG pathway analysis of significant DEGs.

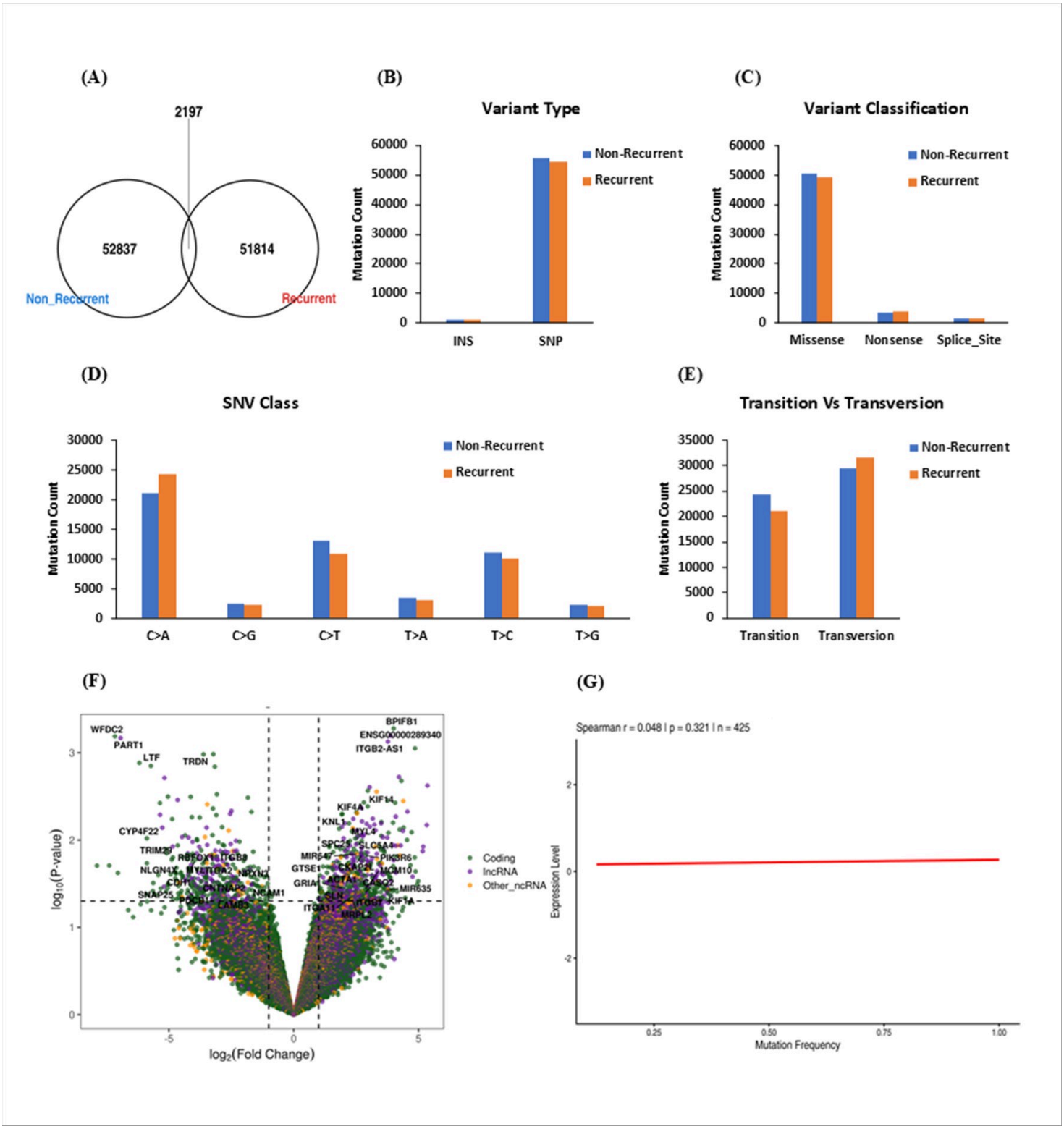


Figure 2

Genomic and transcriptomic analyses of recurrent and non-recurrent TNBC patients. **(**A) Venn diagram illustrating the overlap and distribution of somatic variants in recurrent and non-recurrent TNBC patients. The plots summarize the distribution of somatic variants in recurrent and non-recurrent TNBC patients; (B) variant types, (C) variant classes, (D) SNV classes (C>A, C>G, C>T, T>A, T>C, and T>G), (E) transitions and transversions. Orange color corresponds to recurrent TNBC patients, whereas blue color corresponds to non-recurrent TNBC patients. (F) Volcano plot of DEGs in patients with recurrent TNBC. (G) Spearman's correlation plot of mutation frequency and gene expression.

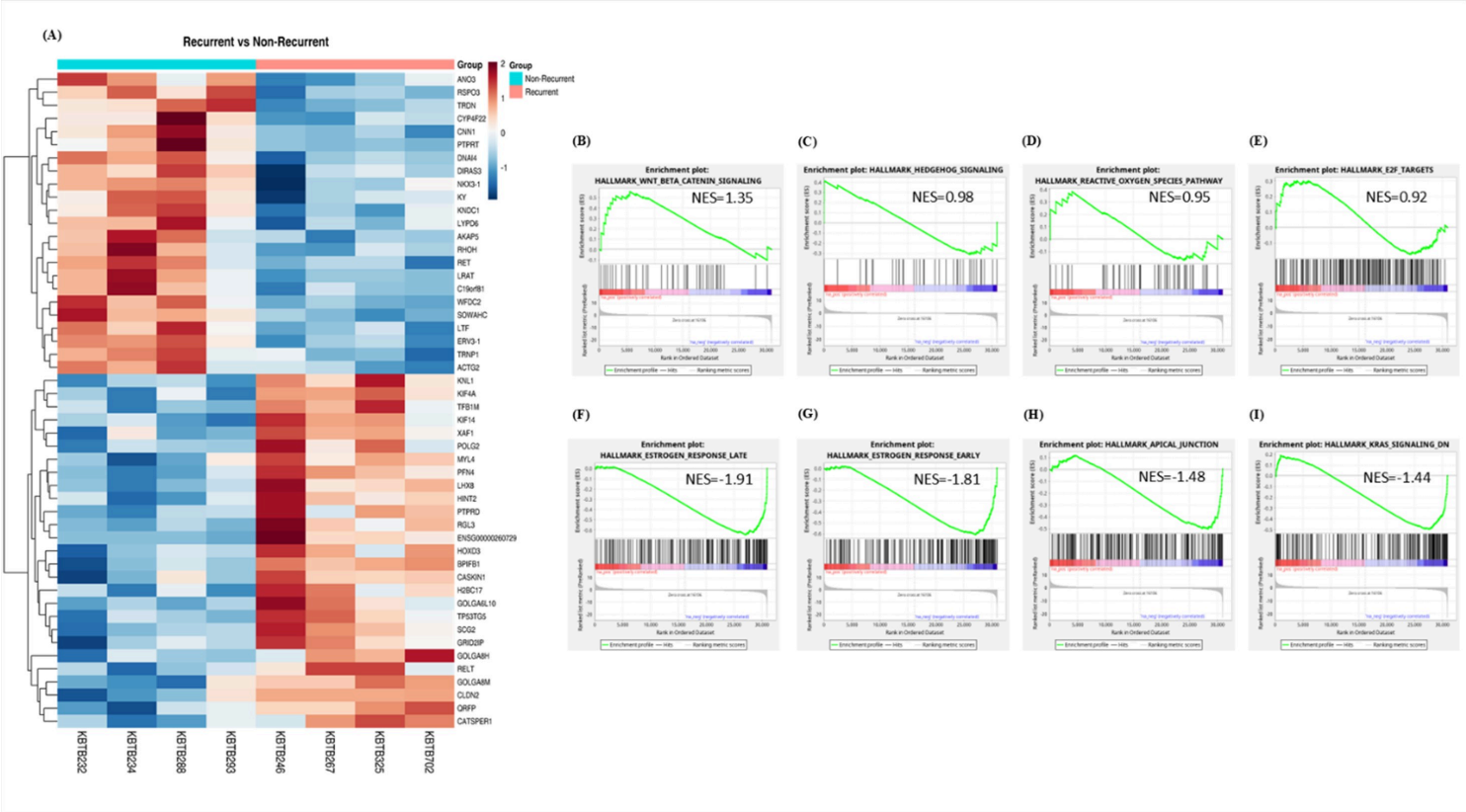


Figure 3

Gene expression profiling and GSEA of protein-coding genes in recurrent TNBC patients. (A) Heatmap showing the expression patterns of the protein-coding DEGs. GSEA analysis of protein-coding DEGs showing positively enriched pathways; (B) Wnt Beta Catenin Signalling, (C) Hedgehog Signaling, (D) Reactive Oxygen Species Pathway, (E) E2F Targets, and negatively enriched pathways; (F) Estrogen Response Late, (G) Estrogen Response Early, (H) Apical Junction, (I) KRAS Signaling DN.

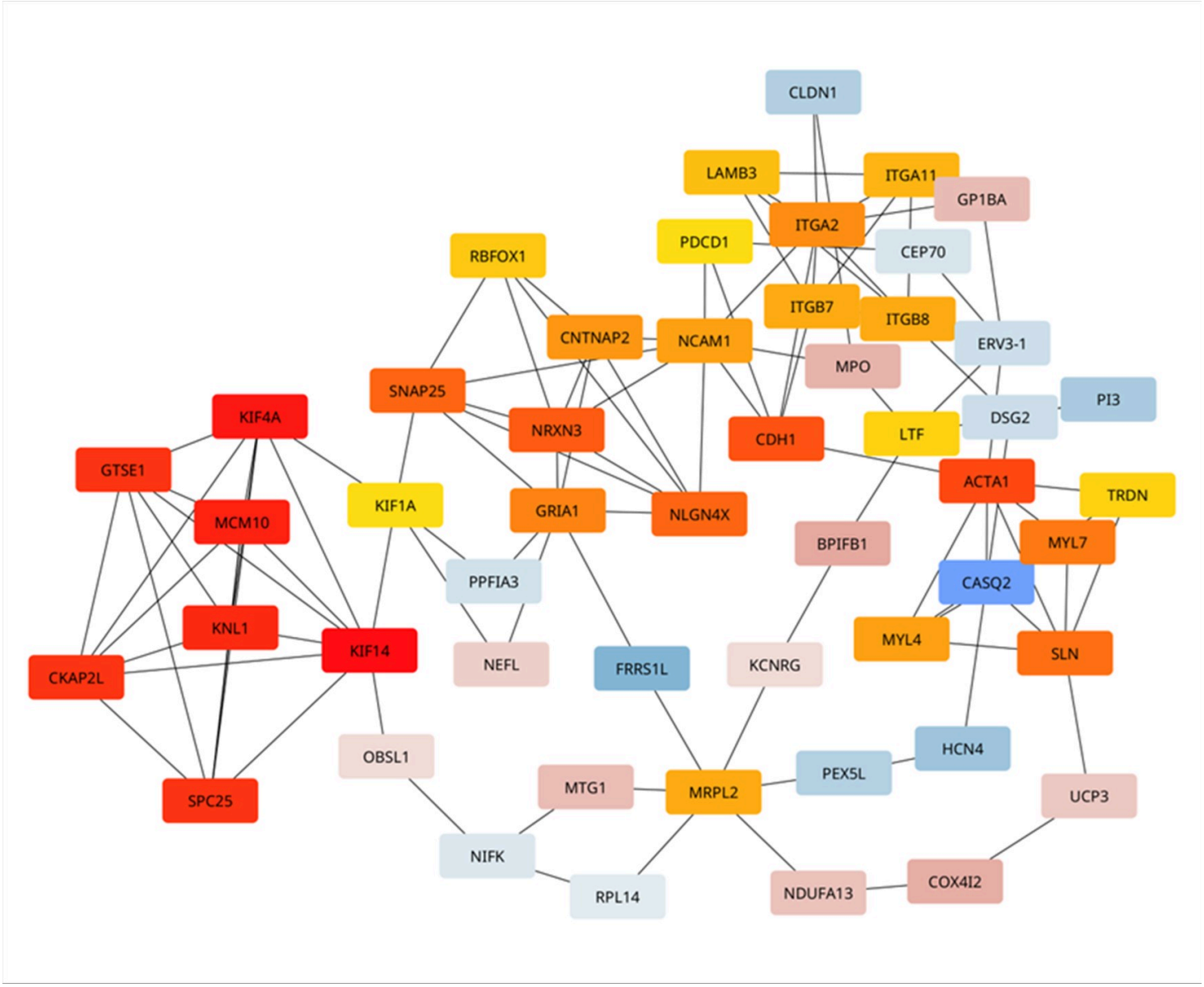


Figure 4

Regulatory network analysis of DEGs for hub gene identification in recurrent TNBC patients. Top 30 hub gene network identified from PPI network analysis of protein-coding DEGs.

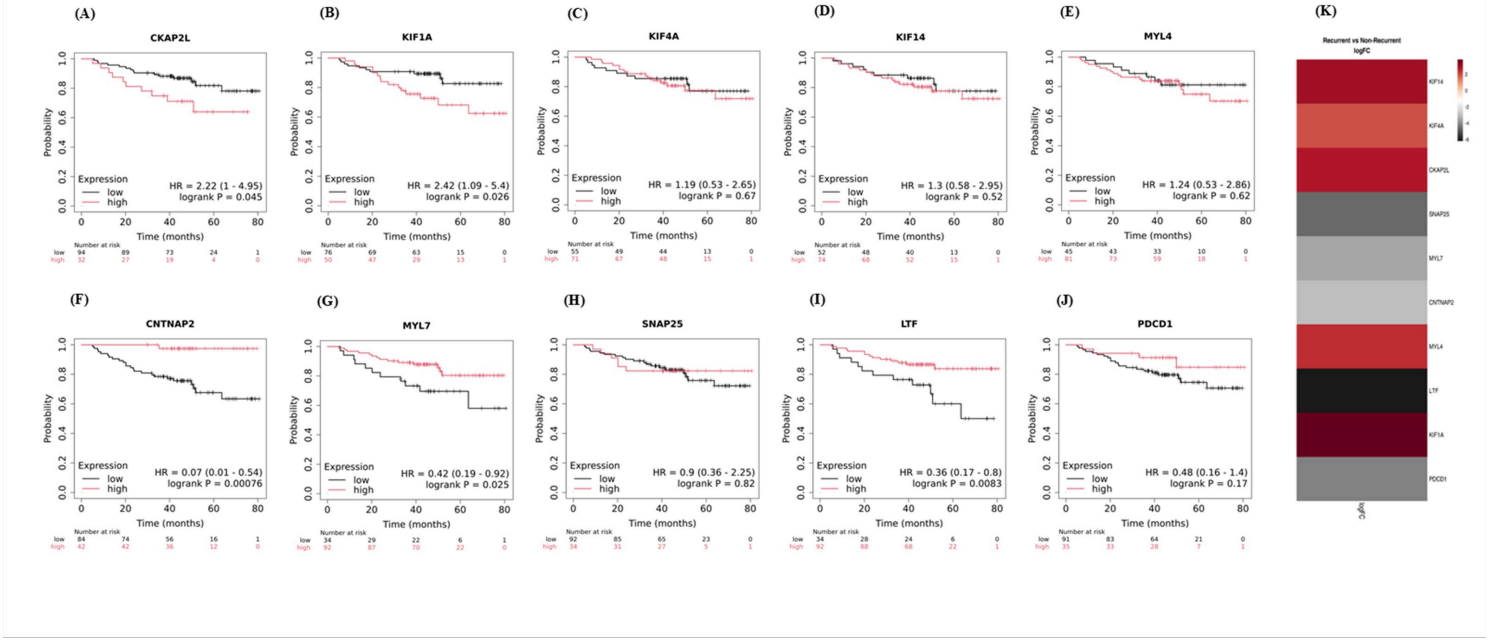


Figure 5

TCGA-based survival analysis of hub genes from recurrent TNBC patients. Kaplan-Meier survival plot of identified hub genes from the TCGA database; (A) CKAP2L, (B) KIF1A, (C) KIF4A, (D) KIF14, (E) MYL4, (F) CNTNAP2, (G) MYL7, (H) SNAP25, (I) LTF, (J) PDCD1, (K) Heatmap illustrating the expression level of hub genes.

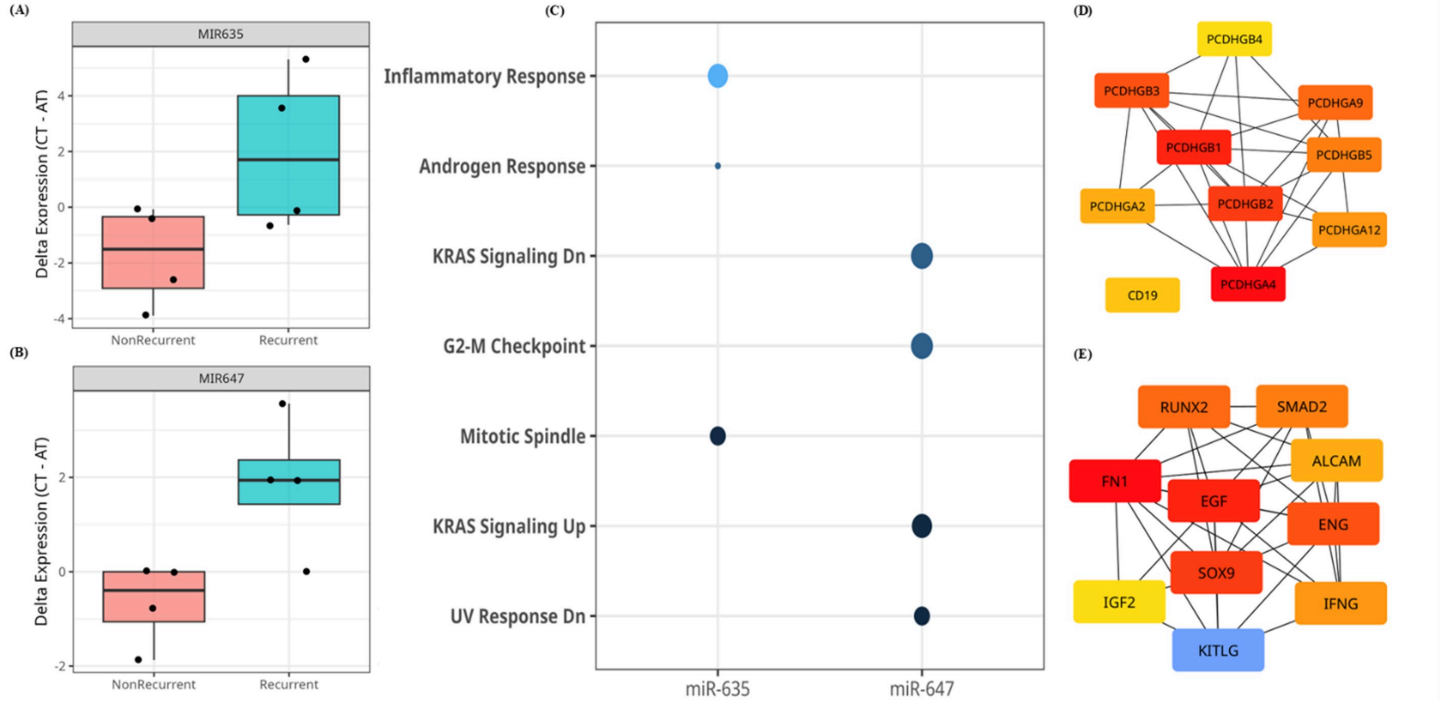


Figure 6

Pathway enrichment and regulatory network analysis of recurrent specific miRNAs and their predicted target genes. Heatmap of the expression profiles of (A) miR-635 and (B) miR-647 in recurrent and non-recurrent TNBC patients. (C) MSigDB Hallmark pathway enrichment analysis of the target genes of miR-635 and miR-647. Identified hub genes from network analysis of (D) miR-635 and (E) miR-647 target genes.