scientific reports

Check for updates

OPEN

Decoding molecular signatures and identifying therapeutic targets in triple-negative breast cancer subtypes using omics and network analysis

Debapriya Sanyal[1], Abdullah M. AlShahrani[2], Soumya Kar[1], S. Rehan Ahmad[3]✉, Chittabrata Mal[4]✉ & Sunil Kanti Mondal[1]✉

Triple-negative breast cancer (TNBC) is the most aggressive form of breast cancer. It is resistant to therapeutic applications, making it very difficult to manage. Most patients continue to receive chemotherapy. The main challenges in treating TNBC arise from its histopathological, transcriptomic, and genomic heterogeneity. While subtype classification was once essential for understanding TNBC's diverse behavior, it is now crucial to thoroughly evaluate the molecular characteristics of each subtype to improve TNBC management. Our study employed a multi-omics approach utilizing transcriptomic data to compare TNBC subtype-specific cell lines with normal breast epithelial cell lines. Our goal was to find out essential transcription factors (TFs) regulating cancer-related pathways, subtype-specific marker genes, and their key TFs. We identified a common set of genes involved in cancer related pathways and examined their alteration level with driver mutations in each subtype. Our findings indicate that, based on alteration patterns, survivability, differential gene and protein level expressions, the genes ERBB3, ERBB4, HLA-B, AKT1, GNAS, FGFR4, PIK3R3 and AXIN2 emerge as primary risk factors in TNBC. These genes, associated with specific subtypes, are potential markers and should be the focus of further studies, paving the way for personalized anticancer treatment strategies for TNBC patients.

Keywords Breast cancer, Multi-omics, Transcriptomics, Pathways, Transcription factors, Risk candidate genes, Drugs, MiRNAs

Breast cancer (BC) is the most frequent malignancy in women, and is a heterogeneous disease on the molecular level[1]. In females, BC has surpassed lung cancer and colorectal cancer and ranked at first position in the world according to the Global Cancer Observatory (GLOBOCAN) estimates. Over the past few years, researchers made enormous efforts to manage BC, and special emphasis has been placed on more biologically-directed therapies (that use the body's immune system to attack cancer cells) and treatment de-escalation to reduce the adverse effects of treatment. The two major pillars of BC management are locoregional treatment and systemic therapy. The histological and molecular characteristics of BC largely influence treatment decisions.

Several molecular alterations can drive breast carcinogenesis and accordingly, several classifications have been developed to group different tumor types. Since breast tumors differ in their histological and molecular features, numerous classifications have been proposed to classify them. Histological, intrinsic, and surrogate intrinsic subtypes are the basis for these classifications[2]. Two histological subtypes of cancer can be distinguished: ductal (now referred to as 'no special type' (NST)) and lobular invasive carcinoma, along with ductal and lobular in situ

[1]Department of Biotechnology, University of Burdwan, Golapbag, Bardhaman, West Bengal 713104, India. [2]Department of Basic Medical Science, College of Applied Medical Sciences, Khamis Mushait Campus, King Khalid University (KKU), 62561 Abha, Saudi Arabia. [3]Hiralal Mazumdar Memorial College for Women, West Bengal State University, Government of West Bengal, Kolkata, West Bengal, India. [4]Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, West Bengal, Nadia, Haringhata, West Bengal 741249, India. ✉email: professor.rehaan@gmail.com; chittabrata.mal@makautwb.ac.in; skmondal@biotech.buruniv.ac.in

preinvasive carcinoma[1]. Based on a 50-gene expression signature (PAM50), there are four intrinsic subtypes of BC: luminal A and luminal B (ER-positive), basal-like, and human epidermal growth factor receptor 2 (HER2)-enriched (without ER expression)[3,4]. Later, five surrogate intrinsic subtypes were proposed based on histology and immunohistochemistry expression of key proteins, including progesterone receptor (PR), ER, HER2, and the proliferation marker Ki67. This subtyping is majorly used in clinical practice: Luminal A-like, Luminal B-like HER2–, Luminal B-like HER2+, HER2-enriched and TNBC[2,5–7].

Luminal A tumors consist of 60–70% of all cases and show low proliferation; Luminal B-like HER2- consists of 10–20% of all cases with high proliferation and poorer prognosis than luminal A-like; Luminal B-like HER2 + and HER2-enriched consist of 13–15% of all cases with high proliferation and aggressiveness; whereas TNBC consists of only 10–15% but shows more adverse outcomes such as high invasiveness, high metastatic potential (~ 46% will develop distant metastasis involving brain and visceral organs), proneness to relapse (35–67 months average in non-TNBC patients while 19–40 months average in TNBC patients), recurrence (~ 25%), mortality rate (~ 40% within the first 5 years after diagnosis) and poor prognosis[1,5,8,9]. TNBC tumors don't express ER, PR, and HER2. Due to their unique biological behavior, they are not sensitive to endocrine therapy or HER2 treatment, and standardized TNBC treatment regimens are still lacking[8,10].

Earlier gene expression profiling of TNBC classified it as a basal-like breast cancer (BLBC) subtype based on its similarity to the BLBC gene expression profile, which can be as high as 60–90% compared to only ~ 11.5% in non-TNBCs. Depending on the epidemiological data it has been estimated that TNBC is more prevalent in premenopausal young women under 40 years (which is ~ 15–20% of all BC patients), African and Hispanic descendants, and carriers of deleterious germline mutations in BC susceptibility genes[8,11]. Based on the immunohistochemistry (IHC) profiling, TNBC is considered as a single entity but molecular profiling has revealed an unexpected level of heterogeneity in it. In 2011, after performing gene expression profiling on tumor samples, Lehmann et al. classified TNBC into six subtypes: basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal-like (ML), mesenchymal-stem like (MSL), immunomodulatory (IM) and luminal androgen receptor (LAR)[12]. Researchers found that while BL1 and BL2 are downregulated in immune response genes but enriched in DNA damage response and cell cycle genes, LAR subtype exhibits activation of hormone receptor pathways. Stem cell pathways characterize ML, whereas immune gene expression and cytokine signaling are linked to IM. Patients with TNBC have limited therapy options since they do not react to standard BC-targeted medicines. Furthermore, TNBCs exhibit decreased sensitivity for breast lineage-specific markers, making it more challenging to determine their breast origin, thereby further complicating diagnosis in metastatic scenarios[13].

Triple-negative breast cancers (TNBCs) are highly heterogeneous, and their molecular subtypes remain incompletely characterized. Presently, the treatment strategies of TNBC includes Chemotherapy (backbone therapy) using of anthracyclines and taxanes in neoadjuvant or adjuvant settings; Immunotherapy/Immune checkpoint inhibitors by monoclonal antibodies targeting PD-1/PD-L1 and CTLA-4; Targeted therapies like PARP inhibitors and inhibitors of signaling pathways (EGFR, mTOR, PI3K/AKT); Combination & multimodal approaches, and Local therapies include surgery and radiotherapy[14]. While FDA-approved drugs are available for TNBC treatment, their subtype-specific efficacy and limitations are not fully understood. At the same time, the regulatory role of miRNAs in driving subtype-specific biology of TNBC is still underexplored. Addressing these gaps is essential for improving stratified therapeutic strategies.

The primary factors contributing to the difficulty in managing TNBC are genomic, transcriptomic, and histopathological variability. As a result, despite numerous attempts, transcriptomic and genomic classifications remain largely theoretical, and the majority of patients continue to receive chemotherapy. Immunotherapy techniques, including those based on immune checkpoint inhibitors and biomarker-driven therapies, have also shown inconsistent and limited results. Therefore, comprehensive analysis of TNBC subtype-specific cell lines and patient samples holds great promise for uncovering key regulatory mechanisms underlying disease heterogeneity. Such insights can guide the development of more precise therapeutic strategies, thereby advancing the concept of personalized medicine in TNBC management[15,16].

To address these challenges, the development of multi-omics technologies, advanced biocomputational algorithms, innovative clinical trial designs, and most importantly, a deeper understanding of the immune and genomic landscape of TNBC will pave the way for individualized anticancer treatment.

Materials and methods

Dataset retrieval

The raw transcriptomic data of TNBC subtype-specific cell lines were retrieved from a public database of the Sequence Read Archive (SRA) database of NCBI (https://www.ncbi.nlm.nih.gov/sra). Due to a shortage of publicly available transcriptomic data of TNBC subtype-specific cell line replicates, highly correlated cell lines have been chosen as replicates for representing each subtype. Depending on the correlation and availability of transcriptomic datasets sequenced on the same sequencing platform (Illumina HiSeq 2500) HCC1937, HCC1143, and HCC3153 have been chosen for BL1 subtype; HDQP1, HCC70, and HCC1806 have been selected for BL2 subtype; HCC2185, CAL-148, and MDA-MB-453 have been chosen for LAR subtype; BT-549, CAL-51, and CAL-120 have been chosen for ML subtype; MDA-MB-157, MDA-MB-436, and MDA-MB-231 have been selected for MSL subtype[8] Due to the unavailability of correlated TNBC subtype-specific cell line transcriptomic data sequenced on the same platform (i.e., Illumina Hiseq 2500), the Immunomodulatory (IM) subtype has been discarded to avoid any biases in the selected dataset. To find out the differentially expressed genes (DEGs) in TNBC subtypes in comparison to the normal state, a widely popular non-treated breast epithelial cell line MCF-10 A (Rep1 Acc. Id SRR22809594, Rep2 Acc. Id SRR22809593, Rep3 Acc. Id SRR22809592) has been chosen with three replicates[17]. Only paired-end sequences were retrieved. All data has been listed in Table 1.

TNBC subtype	Cell line	Subtype correlation[A] (*p* value)	Accession ID
BL1	HCC1937	0.28 (0.00)	SRR8616179
	HCC1143	0.26 (0.00)	SRR8615819
	HCC3153	0.24 (0.00)	SRR2532349
BL2	HCC70	0.24 (0.04)	SRR8615452
	HCC1806	0.22 (0.00)	SRR8616177
	HDQ-P1	0.18 (0.17)	SRR8615278
LAR	MDA-MB-453	0.53 (0.00)	SRR8615577
	HCC2185	0.34 (0.00)	SRR2532346
	CAL-148	0.32(0.00)	SRR8615268
ML	BT-549	0.21 (0.00)	SRR8616197
	CAL-51	0.17 (0.00)	SRR8615264
	CAL-120	0.09 (0.00)	SRR8615270
MSL	MDA-MB-157	0.25 (0.00)	SRR8615756
	MDA-MB-436	0.13 (0.00)	SRR8615584
	MDA-MB-231	0.12 (0.00)	SRR8615767

Table 1. Selected TNBC subtype specific correlated cell lines with their correlation and accession IDs.

Identification of differentially expressed genes in TNBC subtypes

For the identification of DEGs, the galaxy webtool (https://galaxyproject.org/use/) and iDEP (an integrated web application for differential expression and pathway analysis of RNA-Seq data; version 2.0) have been used[18,19]. To check the quality of raw RNA-seq reads, the sequences were trimmed by a trimmomatic tool, keeping the PHRED score threshold at 30. The trimmed sequences were aligned/mapped by STAR (Spliced Transcripts Alignment to a Reference) tool against the latest human reference genome (T2T-CHM13 v2.0/hs1)[20,21]. For splice alignment, the same reference genome has been downloaded in "gtf" format from the UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/goldenPath/hs1). For counting the reads featureCounts tool has been used to reveal both transcript- and gene-level expression for individual cell lines[22]. For the DEG study, the counted read files of all the cell lines have been merged and uploaded in iDEP 2.0 (http://bioinformatics.sdstate.edu/idep96/). Differential expression analysis employed Limma's "voom with quality weights" method, with p-values adjusted for multiple testing via the Benjamini–Hochberg procedure. The differential expression between the normal and tumor groups of the mRNAs was considered statistically significant ($|\log2 \text{ fold change (FC)}| \geq 2$; $P < 0.05$)[23]. The whole workflow has been presented in Fig. 1.

GO and pathway enrichment analysis

The pathway enrichment analysis includes Gene ontology (GO) analysis. In this case Database for Annotation, Visualization, and Integrated Discovery (DAVID) was used to interpret the up-regulated and down-regulated DEGs (https://david.ncifcrf.gov/) involved in GO term and Pathway enrichment analysis[24]. The GO term and KEGG biological pathway data[25–27] were sorted by p-value, with a cut-off of 0.05 ($P < 0.05$), and only the top 10 results were selected. Then, specific GO terms and biological pathways were considered significant based on fold enrichment. All figures and graphic displays were created by the website www.bioinformatics.com.cn.

Identification of candidate genes

Metascape was used to build network clusters of upregulated genes involved in top biological pathways[28]. KEGG biological pathway data was sorted by gene count value (total genes involved in a pathway), and the top 10% of pathways were chosen. Finally, the top 5% of genes have been selected which appeared as a common set of genes (candidate genes) involved in the top 10% pathways. Further, we have performed circos plot and network cluster analysis to identify pathways connected with their common set of genes.

Identification of the mutated genes and survivability analysis

Levels of genetic alterations of the candidate genes were examined in the cBioPortal database among 9490 samples/patients reported by the Breast invasive carcinoma studies[29]. Genes were selected for survivability analysis based on putative driver mutations.

The Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/) database has been used to estimate the survivability of mutated candidate genes in certain subtypes. This analysis included previously published patient sample data. The Kaplan-Meier approach was used to calculate overall survival (OS) rates. The log-rank test was performed to compare the OS rates of each hub gene between the high- and low-expression groups, and a difference of $P < 0.05$ was considered significant. Concurrently, the hazard ratio (HR) value for each hub gene was calculated at a 95% confidence interval (95% CI). A gene's contribution to the risk of BC was indicated by an HR more than 1, while a gene's protective influence on the disease was shown by an HR lower than 1. With this technique, we identified significant and prospective risk genes in each subtype category.

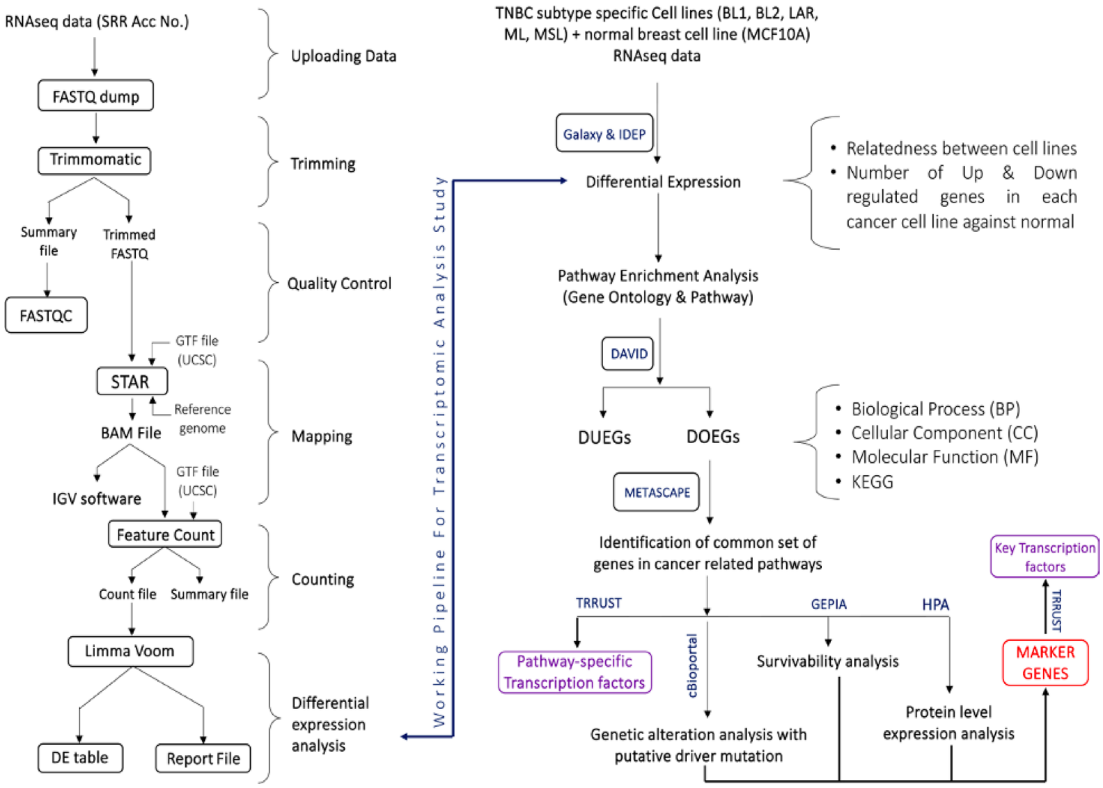


Fig. 1. Workflow of this study. First, the raw transcriptomic data of TNBC subtype-specific cell lines were retrieved. Second, differential expression analysis has been done with limma voom after trimming, mapping and counting process. Third, pathway enrichment analysis was performed with DAVID pathway enrichment analysis tool. Fourth, a common set of upregulated genes (candidate genes) involved in cancer related pathways has been identified with Metascape webtool for each subtype. Fifth, the pathway-specific transcription regulators were identified. Sixth, after checking the mutation pattern (putative driver) of the candidate genes, they were subjected to survivability analysis and validation by GEPIA and HPA database. Finally, marker genes were predicted and the key transcription factors regulating those marker genes were identified.

Validation of differential expression of risk genes

Databases like GEPIA and The Human Protein Atlas (HPA) (www.proteinatlas.org) were used to validate the differential expression of each candidate gene between BC tissue and normal breast tissue. This study was carried out by analysing gene and protein expression levels separately. Finally, the risk and potential marker genes have been identified by analyzing alteration patterns, survivability, and differential expression in breast tumors.

Identification of TFs

The common TFs regulating the differentially over expressed genes (DOEGs) involved in the top 10% of pathways was identified by a reference database of human transcriptional regulatory interactions, TRRUST[30]. Further, we identified key TFs regulating the risk genes selected earlier. We performed this task for each TNBC subtype and identified the expression levels of those TFs.

Virtual screening and molecular Docking for TNBC potential marker specific drugs

The potential marker genes are evaluated in the Drug-Gene Interaction Database (DGIdb) tool to determine their drugability and to identify molecules that target these genes. DGIdb (https://dgidb.org/) contains exper t-curated text mining and drug-gene interaction data[31]. Cytoscape tool was employed to develop the drug gene interaction network[32]. The approved and antineoplastic agents were considered as ligands for molecular docking study. The necessary PDB files of the proteins (ERBB3 PDB ID: **1M6B**; ERBB4 PDB ID: **3BCE**; FGFR4 PDB ID: **4TYI**; AKT1 PDB ID: **8UW2**; HLA-B PDB ID: **7YG3**) were collected from the RCSB Protein Data Bank. The protein files were prepared by removing solvents and other ligand molecules, adding missing hydrogens in PyMOL (v 3.1) for docking. HDOCK server (http://hdock.phys.hust.edu.cn/) was used for blind docking with the ligands. Protein-Ligand Interaction Profiler (https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index) was used to identify the interactions between protein and drug.

Identification of MiRNA targeting markers

miRTargetLink 2.0 database (https://ccb-compute.cs.uni-saarland.de/mirtargetlink2) has been used to identify interactive miRNA and marker gene networks[33]. These miRNAs target the mRNA and play vital roles in post

transcriptional modifications which may be directly linked to its dysregulation, finally resulting in cancer progression. Cytoscape tool has been used to find out common interactive miRNAs targeting multiple markers.

Results

Exploring differential gene expression patterns in TNBC subtypes

An inter-subtype assessment of differential expression patterns amongst TNBC subtypes revealed that the correlated subtype-specific cell lines used in this study clustered separately. One of the ML replicates in the PCA plot exhibited reduced variance compared with its other two replicates. The similar outcome was observed for the LAR subtype (Fig. 2A and B). For each subtype, we compared the TNBC subtype-specific cell lines to the normal, untreated breast epithelial cell line MCF10A, in order to identify the differentially over- and under-expressed genes. Finally, we identified 4337, 5013, 3459, 3634, and 4593 differentially overexpressed genes (DOEGs) in BL1, BL2, LAR, ML, and MSL respectively, and 5259, 5524, 6646, 5963, and 5600 differentially under-expressed genes (DUEGs) in the same order. Differential expression was visualized using Volcano plots for BL1 vs. Normal and LAR vs. Normal (Fig. 2C and E). The remaining volcano plots for other subtypes were provided in the Supplementary Figure S1A, S1C, and S1D. The total percentage of variance (PC1 + PC2) between the cell lines in the PCA plots of BL1, BL2, LAR, ML, and MSL was 86.06% (71.56%+14.5%) (Fig. 2D), 86.82% (68.37%+18.45%), 86.63% (67.22%+19.41%) (Fig. 2F), 86.07% (67.18%+18.89%), and 85.19% (67.51%+17.68%), respectively (Supplementary Fig. S1B, S1D and S1F). Although there was a significant variation between normal and diseased cell lines, there was comparatively little variation within every subgroup of related cancerous cell lines.

Functional classification and characteristics of the enriched gene network

Enrichment analysis of individual cell line DOEGs and DUEGs

To gain a comprehensive understanding of the functions and biological significance of the differentially overexpressed genes (DOEGs) (Table 2) and the differentially underexpressed genes (DUEGs) (Table 3), we conducted a detailed enrichment analysis. This analysis focused on categorizing the DOEGs and DUEGs according to their roles in the Gene Ontology (GO) study. By examining these aspects, we aimed to elucidate the underlying biological pathways and mechanisms influenced by the disease progression. We chose the top ten enriched terms for each subtype. The basal subtypes (BL1 and BL2) displayed unique GO terms highly enriched in immune and inflammatory responses, cellular processes, and signaling (Fig. 3). In KEGG pathways, the cytokine-cytokine receptor pathway was suppressed with several genes involved, while GABAergic synapse and cGMP-PKG signaling pathways were enriched. Simultaneously, transcriptional regulation and cellular processes

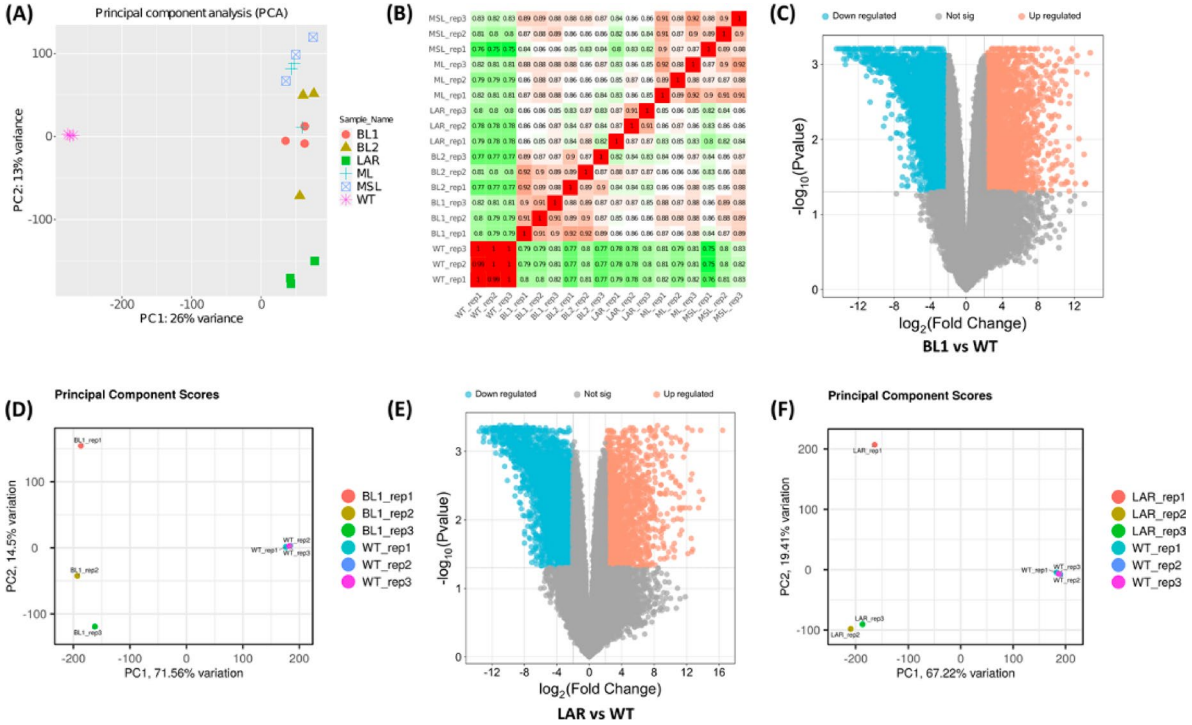


Fig. 2. Differential expression report; (**A**) PCA plot showing variation between TNBC subtype specific correlated cell lines, (**B**) Correlation analysis between TNBC subtype specific cell lines, (**C**) Volcano plot of differential expression between BL1 subtype and normal breast epithelial cell line, (**D**) PCA plot showing variation between BL1 subtype and normal breast epithelial cell line, (**E**) Volcano plot of differential expression between LAR subtype and normal breast epithelial cell line, (**F**) PCA plot showing variation between LAR subtype and normal breast epithelial cell line.

Subtype	Biological process	Cellular component	Molecular function	KEGG pathway
BL1	Positive regulation of T cell mediated cytotoxicity	Myosin complex	Transmembrane-ephrin receptor activity	Taurine and hypotaurine metabolism
	Calcium ion import across plasma membrane	High-density lipoprotein particle	Calcium-dependent phospholipid binding	cAMP signaling pathway
	Arachidonic acid metabolic process	Presynaptic active zone	Beta-amyloid binding	PI3K-Akt signaling pathway
BL2	Calcium ion-regulated exocytosis of neurotransmitter	Chloride channel complex	HMG box domain binding	Taurine and hypotaurine metabolism
	Epoxygenase P450 pathway	Hippocampal mossy fiber to CA3 synapse	Alditol: NADP + 1-oxidoreductase activity	GABAergic synapse
	Positive regulation of cell adhesion	Stereocilium	Aldehyde dehydrogenase [NAD(P)+] activity	Calcium signaling pathway
LAR	Sodium ion import across plasma membrane	Specific granule	AMP binding	Glutamatergic synapse
	Regulation of AMPA receptor activity	Semaphorin receptor complex	Water transmembrane transporter activity	Cell adhesion molecules
	Regulation of neuron apoptotic process	Endoplasmic reticulum	Oxidoreductase activity	cAMP signaling pathway
ML	Cellular response to calcium ion	Synaptic vesicle	Proteoglycan binding	Cell adhesion molecules
	Regulation of synaptic plasticity	Melanosome	Phosphatidate phosphatase activity	Oxytocin signaling pathway
	Cell migration	Presynaptic active zone	Protein disulfide isomerase activity	MAPK signaling pathway
MSL	Positive regulation of epithelial to mesenchymal transition	Cortical cytoskeleton	HMG box domain binding	Antigen processing and presentation
	Adenylate cyclase-activating G-protein coupled receptor signaling pathway	Voltage-gated calcium channel complex	MHC class II receptor activity	cAMP signaling pathway
	Glutamate secretion	Dendritic shaft	Ion channel activity	Wnt signaling pathway

Table 2. Combined pathway enrichment report on differentially over expressed genes in TNBC subtypes.

Subtype	Biological process	Cellular component	Molecular function	KEGG pathway
BL1	Positive regulation of BMP signaling pathway	Box H/ACA RNP complex	Protein tyrosine/threonine phosphatase activity	ATP-dependent chromatin remodeling
	Peptidyl-tyrosine dephosphorylation	SMAD protein complex	MAP kinase tyrosine phosphatase activity	TGF-beta signaling pathway
	Negative regulation of lens fiber cell differentiation	Transcription factor complex	Leucine binding	Cytokine-cytokine receptor interaction
BL2	BMP signaling pathway	Box H/ACA RNP complex	Protein tyrosine/threonine phosphatase activity	ATP-dependent chromatin remodeling
	Positive regulation of transcription	Cajal body	MAP kinase tyrosine phosphatase activity	p53 signaling pathway
	Negative regulation of lens fiber cell differentiation	Nucleus	Transmembrane receptor protein tyrosine kinase activator activity	Cytokine-cytokine receptor interaction
LAR	Positive regulation of ROS metabolic process	Connexin complex	Alditol: NADP+ 1-oxidoreductase activity	p53 signaling pathway
	Response to magnesium ion	Hemidesmosome	Phosphatidylinositol-3,4-bisphosphate binding	EGFR tyrosine kinase inhibitor resistance
	Cellular response to chemokine	Box H/ACA RNP complex	Glycosaminoglycan binding	ErbB signaling pathway
ML	Regulation of gene expression, epigenetic	Hemidesmosome	MAP kinase tyrosine/serine/threonine phosphatase activity	Mucin type O-glycan biosynthesis
	O-glycan processing	Box H/ACA RNP complex	Cell adhesive protein binding	Cytokine-cytokine receptor interaction
	Positive regulation of keratinocyte differentiation	AIM2 inflammasome complex	TRAIL binding	Wnt signaling pathway
MSL	Gene expression	Box H/ACA RNP complex	Protein tyrosine/threonine phosphatase activity	Aldosterone-regulated sodium reabsorption
	Negative regulation of stem cell proliferation	Basement membrane	MAP kinase tyrosine phosphatase activity	ATP-dependent chromatin remodeling
	Positive regulation of cell-matrix adhesion	Connexin complex	Aryl sulfotransferase activity	Spliceosome

Table 3. Combined pathway enrichment report on differentially under expressed genes in TNBC subtypes.

were suppressed (Fig. 4). In the LAR subtype signal transduction and synaptic activity and ion transport and membrane activity were the highly enriched GO terms. Similarly, genes involved in cell signaling and interaction were significantly downregulated (Supplementary Fig. S2A and S2B). In KEGG pathways, cell adhesion molecules and choline metabolism in cancer were enriched while the p53 signaling pathway was significantly suppressed with a low gene count (Supplementary Fig. S3A and S3B). The mesenchymal-like cluster displayed a variety of involved GO terms and pathways (Supplementary Fig. S2C and S2D). Enriched pathways in this subtype include cellular and tissue development, signal transduction, and immune response (Supplementary Fig. S3C and S3D). Moreover, the Mesenchymal stem-like subtypes showed reduced expression of genes in protein binding, enzymatic activity, and gene expression (Supplementary Fig. S2E and S2F). Genes involved in pathways like spliceosome and cytokine-cytokine receptor interaction were down-regulated while proteoglycans

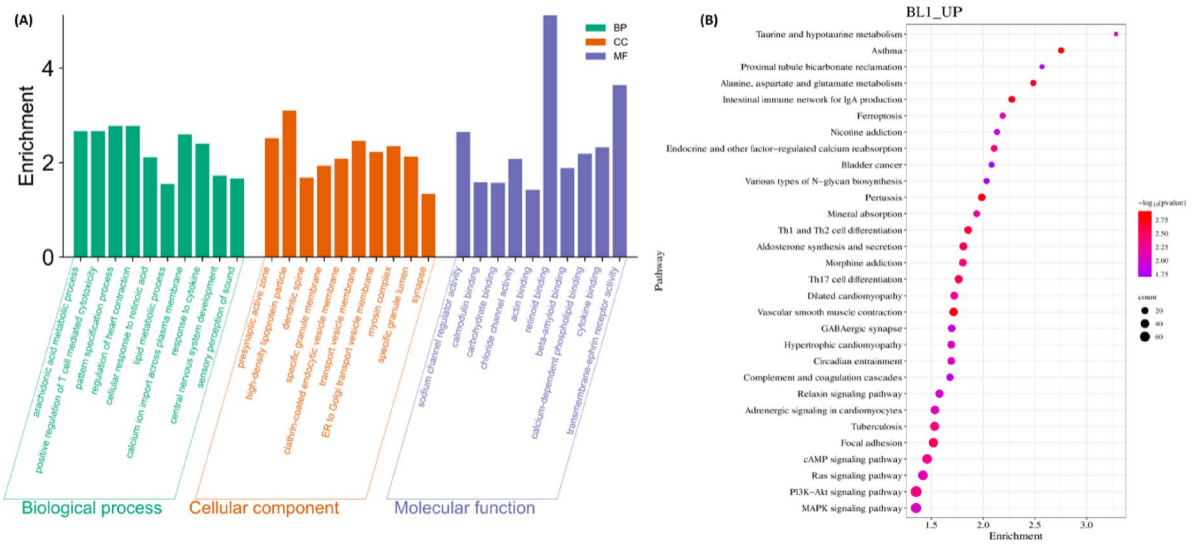


Fig. 3. Pathway enrichment report on differentially over expressed genes; (**A**) Combined GO bar plot (BP, CC and MF) and (**B**) KEGG biological pathway bubble plot on differentially over expressed genes in the BL1 TNBC subtype.

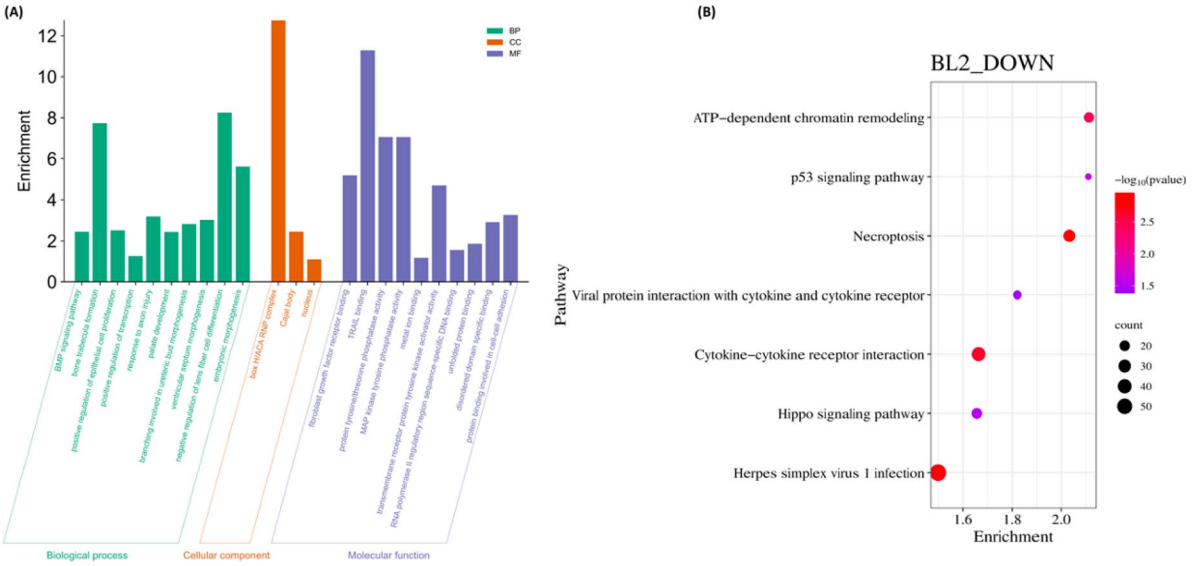


Fig. 4. Pathway enrichment report; (**A**) Combined GO bar plot (BP, CC and MF) and (**B**) KEGG biological pathway bubble plot on differentially under expressed genes in the BL2 TNBC subtype.

in cancer, cell adhesion molecules, antigen processing, and presentation were significantly enriched in KEGG (Supplementary Fig. S3E and S3F).

Visualization of candidate gene enriched pathway network in TNBC subtypes

In the Circos plot, we found pathways interconnected through a common set of genes. For the BL1 subtype, the PI3K-Akt, Ras, MAPK, human papillomavirus infection, and cAMP signaling pathways were interconnected through a common set of genes (Fig. 5A). A similar result was observed for the BL2 subtype (Supplementary Fig. S4A). The LAR subtype revealed that the PI3K-Akt, cAMP, and calcium signaling pathways were also interconnected (Supplementary Fig. S4D). In the ML subtype, several signaling pathways, including PI3K-Akt, MAPK, Ras, calcium, and those related to Alzheimer's disease, were interconnected via a common set of genes. The MSL subtype exhibited comparable results to the ML subtype.

In the BL1 subtype enrichment network, we found the PI3K-Akt signaling pathway, calcium signaling pathway, and cell adhesion molecules formed a significant network cluster. Similarly, cytokine-cytokine receptor interaction, human papillomavirus, MAPK signaling pathway, neuroactive ligand-receptor interaction, PI3K-Akt signaling pathway, cAMP signaling pathway, and Ras signaling pathway constituted a common significant

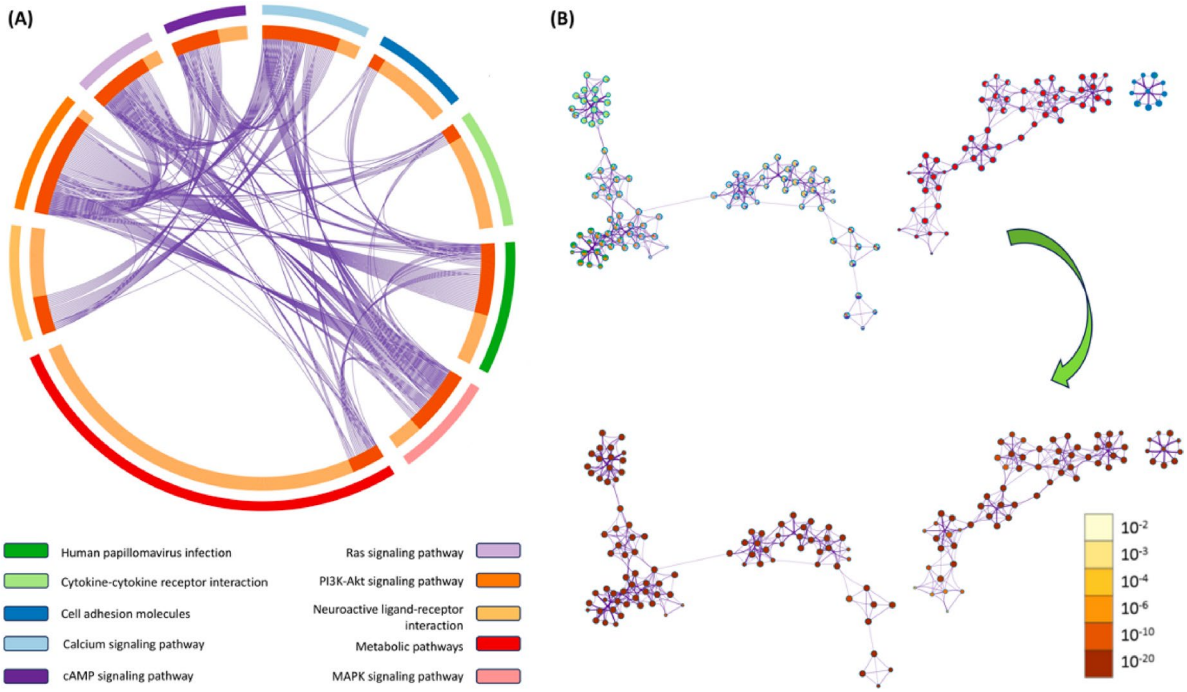


Fig. 5. Visualization of pathway specific interconnected network cluster report: (**A**) Circos plot represents upregulated pathways of BL1 subtype connected with common set of genes, (**B**) Enriched pathway network represented by small pie charts indicating their connection with each other where the color code represents the identities of pathways in BL1. The color grade bar in the subsequent network indicates the significance of the involved pathways. The darker the color shade the more significant the pathway.

interaction network (Fig. 5B). In the BL-2 subtype, the network showed that the cAMP signaling, cell adhesion, and Ras signaling pathway were connected within a network cluster (Supplementary Fig. S4B). In the LAR subtype, adrenergic signaling in cardiomyocytes, prion disease, PI3K-Akt, MAPK, and cAMP signaling pathways shared a common network cluster (Supplementary Fig. S4C). Similarly, in the ML subtype, the Ras signaling pathway, MAPK signaling pathway, Alzheimer's disease, pathways in cancer, and calcium signaling pathways were interconnected (Supplementary Fig. S5A and S5B). In the MSL subtype, the interconnected pathways were Alzheimer's disease, pathways in cancer, PI3K-Akt signaling pathway, pathways in neurodegeneration, and neuroactive ligand-receptor interaction (Supplementary Fig. S5C and S5D).

After visualizing the circos plots and network clusters, the complete list of common sets of genes has been analyzed for each subtype. The top 5% common set of genes (candidate genes) in the top 10% pathways based on total gene count have been selected for BL1 (Supplementary Table S1). A combined list of candidate genes has been presented in Table 4 for each TNBC subtype. Furthermore, we used heatmaps to visualize the primary connections among the top 10% of pathways and identified the linked GO and pathways, along with their network clusters, for each subtype (Supplementary Fig. S5).

Analysis of genetic alterations in TNBC subtypes reveals driver mutations and their frequencies

The cBioPortal database was used to assess genetic alterations in all candidate genes across 9490 samples/patients from the Breast Invasive Carcinoma study for each TNBC subtype. The searched genes were significantly more prevalent in patients/samples, as they were present in the top 10% of pathways for all subtypes. However, in terms of mutation, genes with potential driver mutations were chosen. In the BL1 subtype, searched candidate genes were mutated in 9% of queried patients/samples. Most candidate genes had missense mutations with unknown significance. ERBB3 and FGFR2 had the most missense driver mutations. ERBB4, PDGFRA, and FGFR4 had fewer missense mutations among patients/samples. PDGFRA had an inframe driver mutation. FGFR2 and FGFR4 had fusion driver mutations (Fig. 6). In the BL2 subtype, candidate genes were mutated in 8% of queried patients/samples and most candidate genes had missense mutations with unknown significance. In this case, GRIN2A, HLA-B, HLA-C, and HLA-A had truncating driver mutations. PDGFRA had inframe and missense driver mutation. GNAS had a missense driver mutation. HLA-C and HLA-E had splice site driver mutations. In the LAR subtype, searched candidate genes were mutated in 9% of queried patients/samples. Here AKT1 had the cases of most missense driver mutations with less inframe and splice site driver mutations. GNAS, ERBB4, FGFR1, and RET had missense and fusion driver mutations. PIK3R3 had truncated and splice site driver mutations among patients/samples (Supplementary Fig. S6). In the ML subtype, candidate genes were mutated in 5% of queried patients/samples. Most of the genes had missense and splice site mutations with unknown significance. FGFR2 had the most missense and fusion driver mutations. PDGFRA had both inframe

BL1	BL2	LAR	ML	MSL
PIK3CD	PIK3CD	PRKACA	MAP2K2	MAP2K2
MAPK3	PRKACA	MAPK3	MAPK3	MAPK3
PRKACA	IKBKG	CREB3L4	IKBKG	PDGFRB
MAP2K2	HLA-C	CACNA1C	IKBKB	NOS2
EGF	HLA-F	CREB3	PIK3CD	PIK3CD
NGF	MAPK3	AKT1	PRKCA	AKT1
FLT4	TNF	CACNA1D	PDGFRB	IKBKG
NGFR	GNAS	CREB3L1	PLCG1	PRKCA
FGFR4	CALML6	EGF	FGFR2	EGF
VEGFB	CALML5	GNAS	VEGFB	GRIN2A
FGF17	HLA-G	PIK3R3	FGFR1	WNT2
FGF19	ADCY7	ADCY2	FGFR4	FZD9
PDGFB	MAP2K2	MAP2K2	FGF8	HGF
PDGFRA	PLCE1	CALML6	CALML6	WNT3A
IKBKG	ADCY4	ATF6B	PRKACA	AXIN2
FGF21	HLA-E	RET	CALM3	WNT9A
NTRK2	HLA-B	FGFR1	PDGFB	DVL2
FGFR2	HLA-A	ERBB4	FLT4	CALML6
GRIN1	CALM3	CHRM1	FGF2	PRKACA
CALML6	CREB3L1	PPP2R3B	PRKCG	WNT6
EFNA2	PLD2	PDGFB	PDGFRA	FLT4
TNF	TRADD		WNT11	NTRK1
NFATC1	EGF		AXIN2	WNT11
CREB3L1	CREB3L4		WNT7B	WNT7B
GRIN2D	PLD1		FZD7	PRKCB
RET	CD40		FZD4	IL6
ERBB3	MAPK10			WNT1
ERBB4	PDGFRA			PRKCG
PLA2G4F	GRIN1			FZD7
PRKCG	GRIN2A			ARAF
PLA2G4B	NGF			
	PRKCG			
	ADCY6			

Table 4. Identified candidate genes in each TNBC subtype.

and missense driver mutation cases. AXIN2 had a truncating mutation. FGFR4 and FGFR1 had fewer missense driver mutations. In the MSL subtype, candidate genes were mutated in 9% of queried patients/samples. AKT1 had the cases of most missense driver mutations with fewer inframe and splice site driver mutations. GRIN2A and AXIN2 had truncating driver mutation, and NTRK1 had some cases of missense driver mutation (Supplementary Fig. S6).

Outcome of survivability analysis and validation of expression

To estimate the prognostic value of significantly mutated TNBC subtype-specific candidate genes in BC patients, we performed overall survival (OS) analysis using the GEPIA database and the Kaplan–Meier survival method. The hazard ratio (HR) and confidence interval (CI) yielded statistically significant results. Based on HR values, ERBB3 (HR = 1.2), FGFR4 (HR = 1), CALNAC1C (HR = 1), AKT1 (1.3), and GNAS (1.2) were identified as risk factors of BC, while ERBB4 (HR = 0.9), HLA-B (HR = 0.94), HLA-A (HR = 0.8), PIK3R3 (HR = 0.93), AXIN2 (HR = 0.84), HLA-E (HR = 0.85), GRIN2A (HR = 0.93), FGFR2 (HR = 0.89), PDGFRA (0.97) were recognized as a protective factors (Fig. 7; Supplementary Fig. S7).

The gene-level expression pattern of mutated driver genes in GEPIA showed that ERBB3, ERBB4, GNAS, FGFR4, HLA-A, HLA-B, AKT1, PDGFRA, PIK3R3, and AXIN2 at high to medium levels compared to the other genes selected for survivability analysis (Fig. 8; Supplementary Fig. S7). We also analyzed the protein level expression of these risk genes using the HPA database, filtering their expression in BC tissue from HPA pathology data. The expression of ERBB3, AKT1, and HLA-B appeared at high to medium levels (Fig. 9). GNAS, ERBB4, FGFR4, and PIK3R3 showed medium to low expression while AXIN2 was expressed at medium levels. No protein expression for PDGFRA was detected in that pathology data and expression data for HLA-A was not available. Therefore, we hypothesized that risk genes ERBB3 (BL1), ERBB4 (BL1 and LAR), GNAS (BL2 and LAR), FGFR4 (BL1 and ML), HLA-B (BL2), AKT1 (LAR and MSL), PIK3R3 (LAR), and AXIN2 (ML, MSL) could serve as markers in each TNBC subtype.

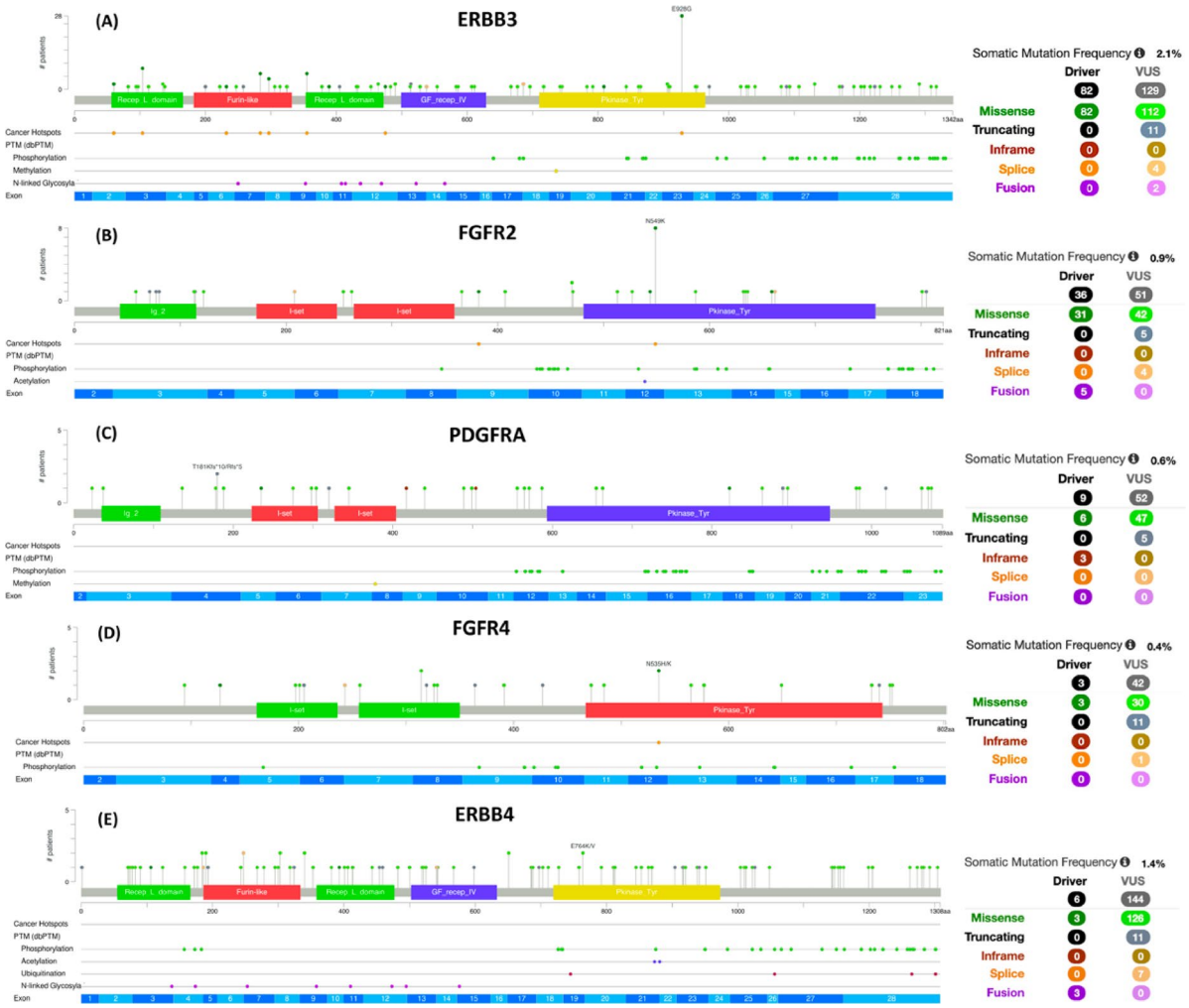


Fig. 6. Genes with putative driver mutations in BL1 TNBC subtype; (**A**) ERBB3 gene showing 2.1% somatic mutation frequency with 82 missense driver mutation, (**B**) FGFR2 gene showing 0.9% somatic mutation frequency across selected patients' samples with 31 missense and 5 fusion driver mutations, (**C**) PDGFRA gene showing 0.6% somatic mutation frequency across samples with 6 missense and 3 inframe driver mutation, (**D**) FGFR4 gene showing 0.4% mutation frequency along with 3 missense driver mutation, and (**E**) ERBB4 gene showing 1.4% mutation frequency with 3 missense and 3 fusion driver mutations.

Regulatory role of TFs in different TNBC subtypes

We identified the pathway-specific common TFs and their expression levels regulating the DOEGs for each TNBC subtype (Fig. 10A and C) and categorized the most dysregulated TFs based on our dysregulation criteria, defined as a $\log_2$ fold change ($\log_2$FC) ≥ 2 for upregulation and ≤ -2 for downregulation. In the BL1 subtype, MYCN (6.39), RFXANK (2.02), and IRF3 (2.73) genes were significantly upregulated while MYC (−3.97) and HIVEP2 (−2.48) were significantly downregulated. Here, log2FC values are provided in parentheses beside each TF. In the BL2 subtype, RFXANK (2.71) and STAT1 (2.19) were upregulated and NFKBIA (−2.77), FLI1 (−5.65), MYC (−4.42), and HIVEP2 (−2.65) were downregulated (Table 5). In the LAR subtype, MYB (2.61) and AR (11.10) were significantly upregulated, while CEBPB (−4.24), ETS2 (−3.05), TWIST1 (−8.50), STAT6 (−2.14), and ETS1 (−5.62) were significantly downregulated. In the ML subtype, HOXB7 (8.72) and SATB1 (4.69) were significantly upregulated and FLI1 (−6.91) and CEBPA (−2.48) were significantly downregulated. In the MSL subtype, MYCN (5.18) was upregulated and NFKBIA (−2.86), FLI1 (−3.13), TWIST1 (−2.75), MYC (−4.25), and TFAP2A (−3.46) were downregulated (Table 5). Figure 10 and Supplementary Table S2 summarize the common and unique pathway-specific dysregulated TFs across TNBC subtypes.

Further, we identified the key TFs regulating the risk genes and examined their differential log2FC expression levels (Supplementary Table S2). In the BL1 subtype, TFs such as TWIST1 (−5.77), TWIST2 (−8.28), and AR (2.00) were significantly dysregulated. These TFs are known key regulators of the ERBB3 risk gene according to TRRUST. The PDGFRA risk gene is regulated by CEBPD (−4.22), GLI1 (2.70), and GLI2 (3.16), which were also significantly dysregulated BL1. A similar pattern of dysregulation for TFs regulating PDGFRA in BL2 and ML subtypes. IKZF1, a transcription regulator of the FGFR4 risk gene, was significantly upregulated in the BL1 and ML subtypes. In the BL2 subtype, HIVEP2 (−2.65) and MYC (−4.42), key regulators of the HLA-B risk gene,

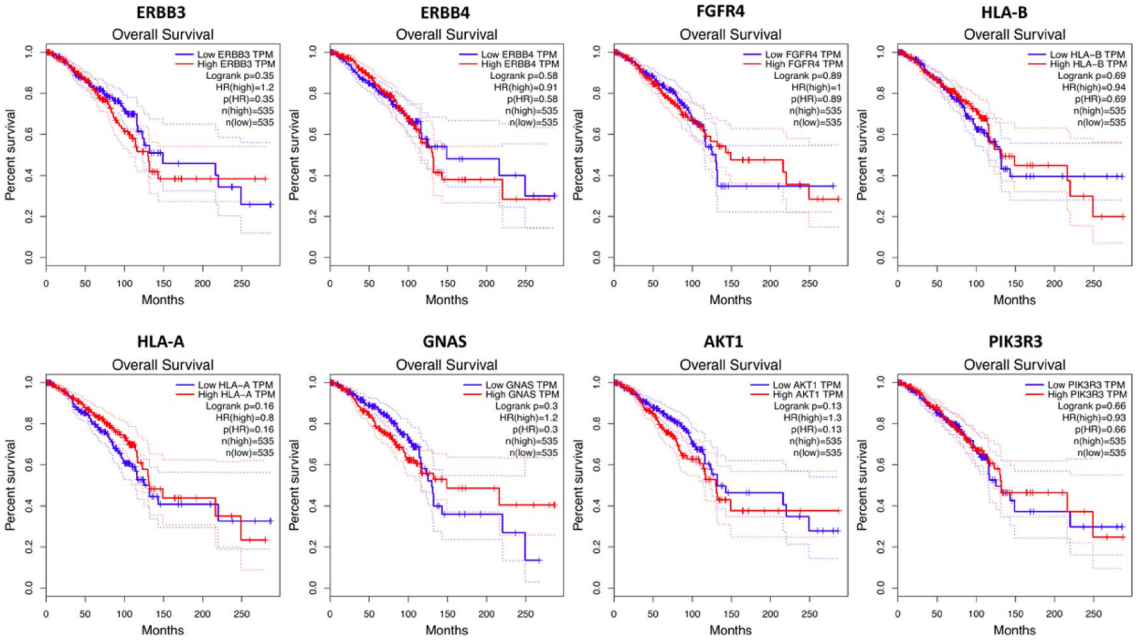


Fig. 7. Survivability analysis of driver mutated candidate genes across patients reported from GEPIA.

were significantly dysregulated (Table 6). In the LAR subtype, AR (11.10), the key TF of AKT1, was significantly upregulated. In the ML subtype, MYC (−4.14), a key regulator of AXIN2, was significantly downregulated. A similar dysregulation pattern for MYC was observed in the MSL subtype. Additionally, a TF such as GLI1 (2.70) was upregulated in the MSL subtype and serves as a key regulator of the AKT1 risk gene (Table 6).

However, the TRRUST database indicated an unknown mode of regulation for some TFs. Specifically, the regulatory mechanism of the TF CEBPD on the PDGFRA risk gene was unclear. A similar lack of clarity was observed for the TFs HIVEP2, GLI1, and MYC, which regulate the risk genes HLA-A, AKT1, and AXIN2, respectively. Furthermore, we observed substantial differences in the expression levels of these TFs (Table 6). Although TRRUST characterizes MYC regulation of the HLA-B risk gene as activation, we found that this gene was markedly downregulated.

Identified marker genes and associated drugs

The potential marker genes were assessed for their drugability in the DGIdb tool, and subsequently small-molecule drugs targeting these genes were identified. We found a total of 336 drugs including 71, 47, 52, 58, 35, 35, and 38 drugs targeting AKT1, FGFR4, ERBB3, ERBB4, GNAS, PIK3R3, and HLA-B gene respectively but no drugs for AXIN2. Among them 102 drugs are approved by United States Federal Drug Administration (US FDA) including a subset of 37 antineoplastic agents, 27 various forms of agents which may be repurposed and 37 uncategorized drugs (Supplementary Table S3). Within this FDA approved category, we found drugs targeting multiple genes (Fig. 11). Drug-gene interaction network shows that Vandetanib, Gefitinib, Cucurmin, Afatinib, Dacomatinib anhydrous interact with both ERBB3 and ERBB4 genes. Everolimus, Nintedanib esylate interact with both FGFR4 and AKT1 genes. GNAS and FGFR4 are targeted by Fluorouracil drug, while GNAS and AKT1 are targeted by Amitriptyline hydrochloride. Infigratinib drug has dual interaction with PIK3R3 and FGFR4 gene. Drugs like Lapatinib and Erlotinib interact with more than two genes. While Lapatinib targets HLA-B, ERBB3 and ERBB4 genes, Erlotinib has triple interactions with ERBB3, ERBB4 and FGFR4 genes. Most of the drugs in this group are categorized as antineoplastic drugs with a few exceptions such as Everolimus and Amitriptyline hydrochloride that are categorized as immunosuppressive and analgesic respectively in DGI database. These drugs may be repurposed as therapeutic drugs against TNBC subtype specific markers. Cucurmin, Fluorouracil and Infigratinib drugs are not properly categorized.

The protein-drug molecular docking results were summarized in Supplementary Table S4. Among the small molecule drugs, DOCETAXEL ANHYDROUS had a good docking score with 4 hydrogen bonds and 5 hydrophobic interactions and a low ligand RMSD, suggesting a reliable, stable binding pose with strong binding affinity to ERBB3. Among the monoclonal antibodies such as CETUXIMAB, PERTUZUMAB, and TRASTUZUMAB targeted to ERBB3, CETUXIMAB got the better docking score and lower ligand RMSD value indicating higher binding affinity and specificity. For AKT1, we identified several FDA approved antineoplastic drugs but TOPOTECAN HYDROCHLORIDE (H-bonds: 1 and hydrophobic bond: 2), FULVESTRANT (H-bonds: 1 and hydrophobic bond: 14), and NINTEDANIB ESYLATE (H-bonds: 3 and hydrophobic bond: 5) had better docking score and low ligand RMSD in comparison to other drug candidates. All the drugs against ERBB4 protein had similar docking scores. However, NERATINIB (H-bonds: 4 and hydrophobic bond: 6) and LAPATINIB (H-bonds: 2 and hydrophobic bond: 11) can be better fit based on docking score and ligand RMSD ratio. Among the FGFR4 targeted antineoplastic drugs, NINTEDANIB ESYLATE and PONATINIB have better

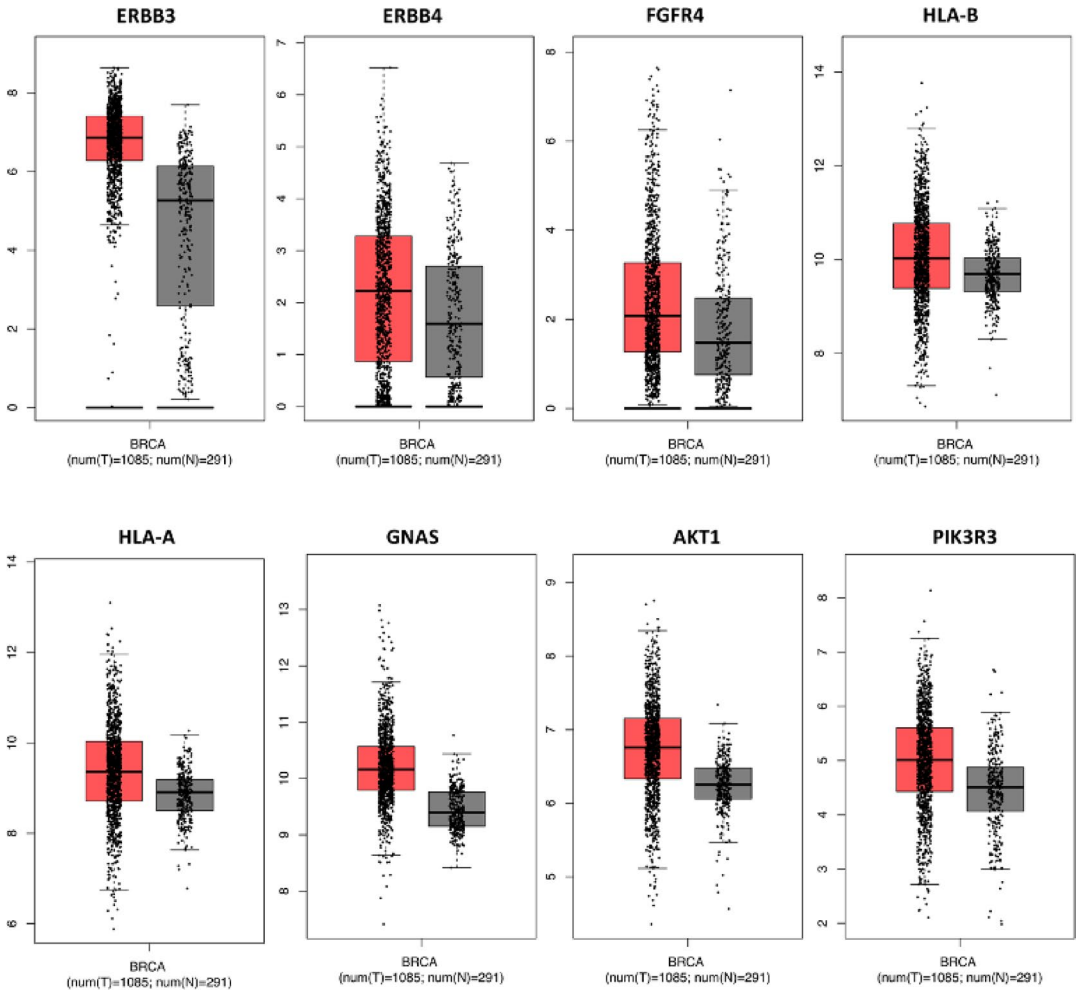


Fig. 8. Gene level expression of driver mutated candidate genes across tumor and normal samples, validation report from GEPIA.

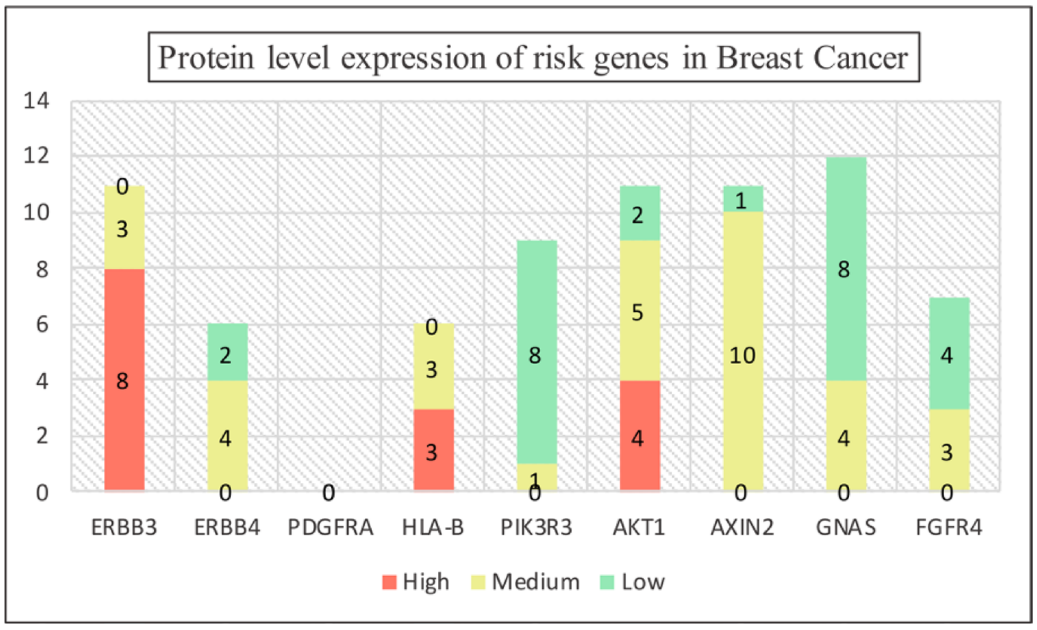


Fig. 9. Protein level expression of risk genes in breast cancer pathology data registered in HPA database.

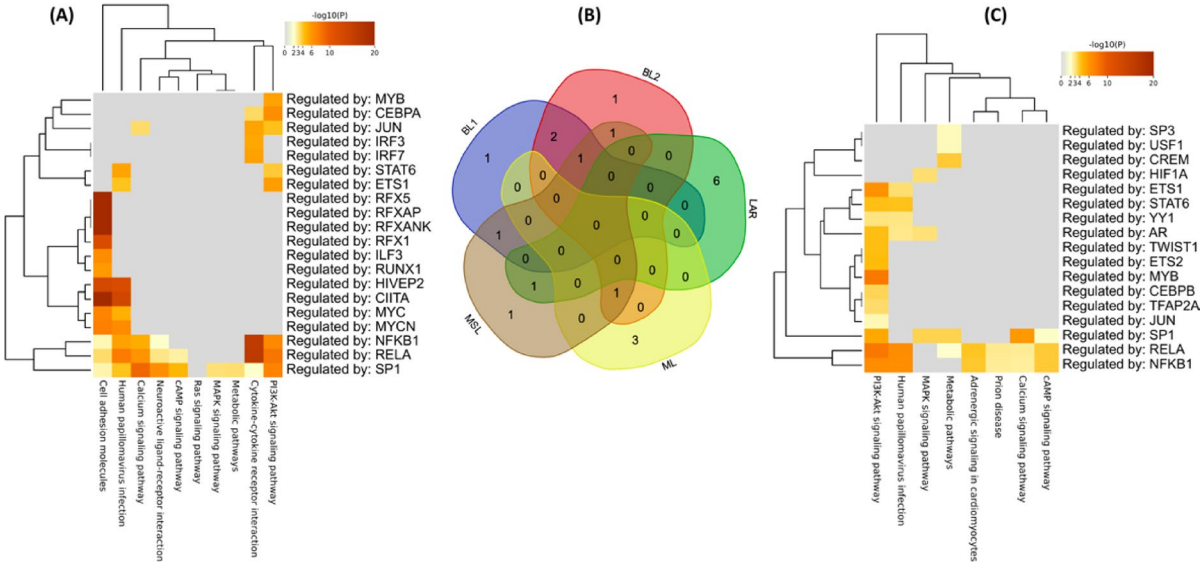


Fig. 10. Pathway specific Transcription regulator identification report; (**A**,**C**) Transcription factors regulating top 5% common set of genes (candidate genes) among top 10% upregulated pathways in BL1 & LAR TNBC subtype respectively, (**B**) Venn diagram of the essential transcription factors which were significantly dysregulated among all TNBC subtypes.

BL1	Log2FC	BL2	Log2FC	LAR	Log2FC	ML	Log2FC	MSL	Log2FC
MYCN	6.39	*RFXANK*	2.71	*AR*	11.10	*HOXB7*	8.72	*MYCN*	5.18
IRF3	2.73	*STAT1*	2.19	*MYB*	2.61	*SATB1*	4.69	**MYC**	−4.25
RFXANK	2.02	**FLI1**	−5.65	**TWIST1**	−8.50	**FLI1**	−6.91	**TFAP2A**	−3.46
MYC	−3.97	**MYC**	−4.42	**ETS1**	−5.62	**CEBPA**	−2.48	**FLI1**	−3.13
HIVEP2	−2.48	**NFKBIA**	−2.77	**CEBPB**	−4.24	EZH2	1.50	**NFKBIA**	−2.86
ETS1	1.96	**HIVEP2**	−2.65	**ETS2**	−3.05	RELA	1.43	**TWIST1**	−2.75
STAT6	1.44	STAT6	1.65	**STAT6**	−2.14	EP300	1.41	ETS1	1.89
RELA	1.39	RELA	1.39	TFAP2A	1.47	ETS1	1.38	RELA	1.48
ILF3	0.73	RFX1	1.23	RELA	1.33	STAT3	1.10	ETV4	0.63
IRF7	0.66	ILF3	1.15	JUN	1.27	TFAP2A	1.10	NFIC	0.60
JUN	0.43	IRF1	0.76	USF1	0.96	STAT1	0.66	STAT3	0.57
RFX1	0.42	RFX5	0.57	CREM	0.29	DDIT3	0.63	TP53	0.31
RFX5	0.05	RFXAP	0.52	YY1	−0.13	ETV4	0.49	USF1	0.23
CEBPA	0.03	USF1	−0.18	SP1	−0.28	CREB1	0.45	SMAD7	−0.10
RFXAP	−0.11	SP1	−0.31	NFKB1	−0.44	SP1	−0.32	HIF1A	−0.15
NFKB1	−0.15	NFKB1	−0.76	SP3	−0.73	TWIST1	−0.34	NFKB1	−0.72
RUNX1	−0.39	CIITA	−0.96	HIF1A	−1.06	NFKB1	−0.51	SP3	−0.80
SP1	−0.71	SP3	−1.05			HIF1A	−0.71	SP1	−0.88
CIITA	−0.94					SP3	−1.14	CIITA	−1.04
MYB	−1.27					EGR1	−1.62	HIVEP2	−1.87

Table 5. Data validation report on pathway-specific essential transcription regulators and their expression value. Italics represents over-expression and bold represents under-expression.

docking scores and ligand RMSD ratio with multiple non-covalent interactions. For HLA-B, small inhibitory protein molecules such as PEGINTERFERON ALFA-2 A and INTERFERON BETA-1 A had better binding affinity in comparison to the other antineoplastic agents.

MiRNAs regulating identified markers

The potential marker genes were subjected to analysis in the miRTargetLink database to identify miRNAs that play vital roles in the post-transcriptional modification of these genes. Strongly validated miRNAs were selected, while weakly validated and predicted ones were excluded. The marker genes ERBB3, ERBB4, PIK3R3, AKT1, and AXIN2 were targeted by 14, 8, 5, 36, and 10 miRNAs, respectively (Supplementary Table S5). We did not

TNBC subtype	Risk genes	Key TFs	log2FC	Mode of regulations	References (PMID)
BL1	ERBB3	*TWIST1*	− 5.77	Repression	19051271
		TWIST2	− 8.28	Repression	19051271
		YBX1	0.16	Activation	19648825
		AR	2.00	Activation	21741601
	ERBB4	WWP1	0.34	Repression	19561640
	PDGFRA	*CEBPD*	− 4.22	**Unknown**	11278956
		GLI1	2.70	Activation	21135115
		GLI2	3.16	Activation	21135115
	FGFR4	*IKZF1*	5.07	Activation	11981041
BL2	GNAS	YY1	− 0.45	Unknown	18458536
	PDGFRA	*CEBPD*	− 4.77	**Unknown**	11278956
		GLI1	2.59	Activation	21135115
		GLI2	1.11	Activation	21135115
	HLA-B	CIITA	− 0.96	Activation	11053628
		HIVEP2	− 2.65	**Unknown**	1409593
		MYC	− 4.42	***Activation***	8206526
LAR	ERBB4	WWP1	0.40	Repression	19561640
	AKT1	APC	− 0.94	Repression	15888491
		GLI1	1.32	Unknown	23580656
		NFKB1	− 0.44	Repression	18031796
		RELA	1.33	Repression	18031796
		STAT3	0.38	Unknown	16007214
		AR	11.10	Activation	11278645
		TP53	1.39	Repression	12485897
	GNAS	YY1	− 0.12	Unknown	18458536
	PIK3R3	MZF1	1.39	Unknown	23509792
ML	PDGFRA	*GLI1*	3.72	Activation	21135115
		GLI2	1.61	Activation	21135115
		CEBPD	− 3.37	**Unknown**	11278956
	AXIN2	*MYC*	− 4.14	**Unknown**	24299953
		SOX4	0.47	Unknown	22098624
		CDX2	1.59	Activation	23393221
	FGFR4	*IKZF1*	3.93	Activation	11981041
MSL	AKT1	APC	− 1.02	Repression	15888491
		GLI1	2.70	**Unknown**	23580656
		NFKB1	− 0.72	Repression	18031796
		RELA	1.48	Repression	18031796
		AR	1.68	Activation	11278645
		STAT3	0.57	Unknown	16007214
		TP53	0.31	Repression	12485897
	AXIN2	*MYC*	− 4.25	**Unknown**	24299953
		SOX4	− 1.38	Unknown	22098624

Table 6. List of key transcription factors regulating the risk genes and their mode of regulation and expression value in each TNBC subtype. Underline and italics represent downregulation and upregulation respectively, bold italics represents contradictory findings, and bold represents new findings.

find any miRNA targeting HLA-B, GNAS and FGFR4 genes. We identified 7 miRNAs targeting multiple genes (Fig. 12). For example, ERBB3 and AKT1 are targeted by hsa-miR-125b-5p, hsa-miR-22-3p, and hsa-miR-125a-5p. miRNAs like hsa-miR-302b-3p, hsa-miR-302d-3p, and hsa-miR-19a-3p have dual interaction capability with ERBB4 and AKT1 gene, while AXIN2 and AKT1 are co-targeted by hsa-miR-126-3p miRNA.

Discussions

TNBC is a complex and heterogeneous group of diseases. Classification and identification of its molecular landscape are essential for precision management. We compared individual TNBC subtypes with normal breast epithelial cells. The correlated subtype-specific cell lines clustered separately in the PCA plot. Analysis of all PCA plots indicated that the variance between normal and diseased cell lines was high, whereas the variance within the correlated cancerous cell lines of each subtype was very low. Therefore, we consider them as replicates.

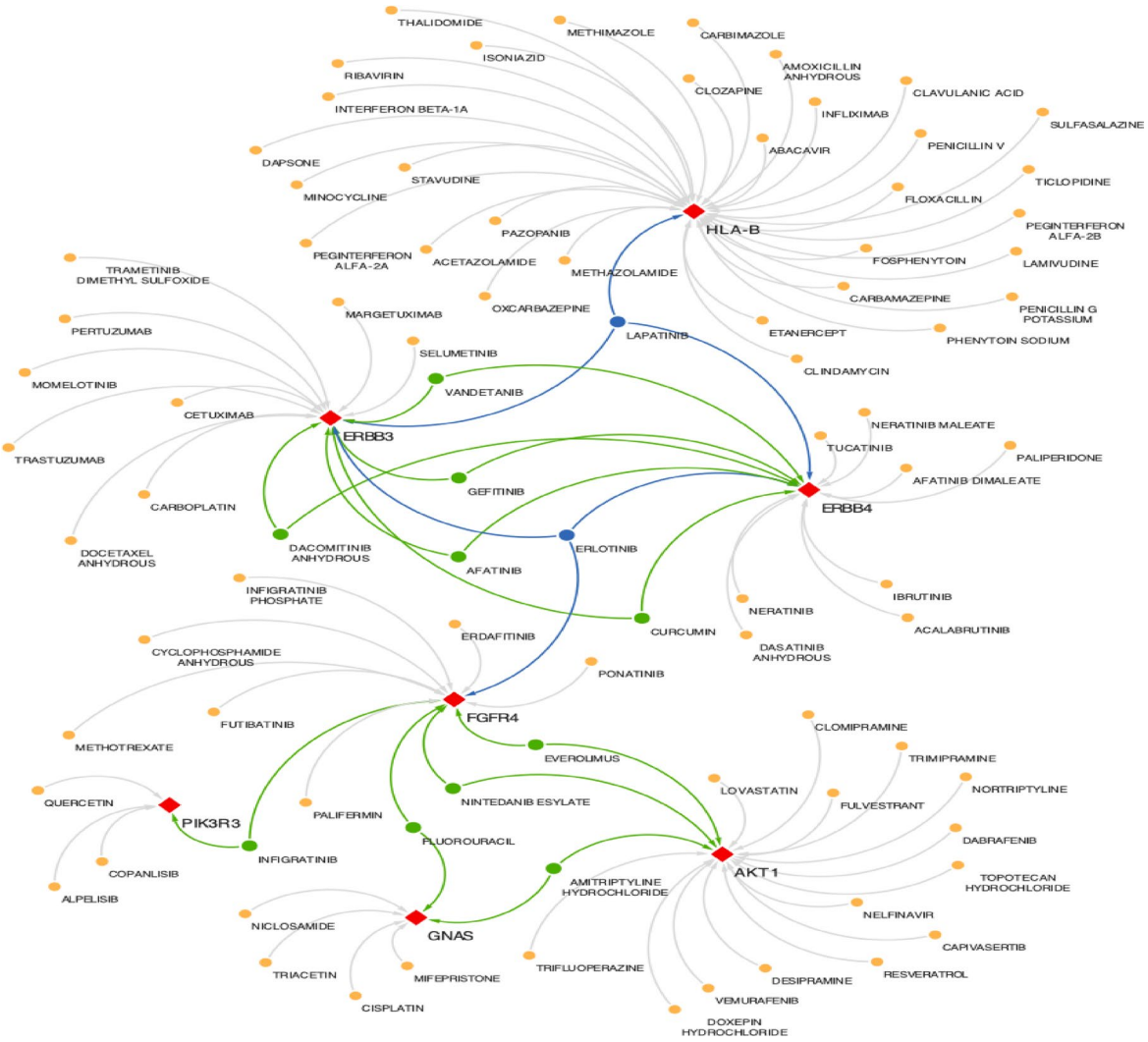


Fig. 11. Drug and targeted marker gene interacting network: Red shapes represent marker genes, orange dots represent drug interacting with single gene and green dots are for drugs targeting multiple genes.

The results of gene ontology enrichment analysis showed that the DOEGs in the most subtypes were significantly enriched in the PI3K-AKT signaling pathway, MAPK signaling pathway, and Wnt signaling pathway. The PI3K signaling pathway, which has been well known for decades, is crucial in cellular processes and is often altered in BC, thereby promoting tumor growth and reducing survival rates. Although small molecule inhibitors targeting PI3K, AKT, and mTOR have shown promising preclinical results, their effectiveness is limited by resistance and significant adverse reactions[34]. The MAPK signaling pathway is crucial in TNBC development and is associated with invasion, metastasis, and prognosis. Higher MAPK activity correlates with shorter survival in TNBC patients, indicating a poor prognosis[35]. The LAR subtype displayed unique GOs that were highly enriched in synaptic functioning and signaling, ion transport and apoptosis regulation, and natural development and differentiation. Earlier studies identified consistently upregulated Wnt-induced targets in TNBC subtypes, further indicating an association of Wnt signaling with the basal-like subtype[36]. Enriched pathways in the mesenchymal-like subtype involved focal adhesion[37]. Moreover, the mesenchymal-like subtype showed low expression of genes involved in cell signaling and gene expression. The BMP signaling pathway, ATP-dependent chromatin remodeling, and p53 signaling pathway were the most downregulated pathways identified when analyzing differentially under-expressed genes (DUEGs). Recent studies suggest that BMPs influence non-classical pathways, including PI3K/Akt and MAPKs, which are critical for tumor growth, invasion, and metastasis[38]. In the BL1 subtype, down-regulation of the notch signaling pathway was observed, which has tumor suppressive effects[29]. Moreover, in the LAR subtypes, genes involved in immune response were mostly downregulated, consistent with the tumors' rapid cell proliferation and capacity to escape the immune system.

Transcription factors are the main regulators of cancer-related gene expression. Previous studies have identified TFs such as MYCN, RFXANK, IRF3, STAT1, TWIST1, and FLI1 as frequently expressed in both primary and recurrent TNBCs. Tumor aggressiveness increases with higher expression levels of these genes[39–41]. Studies have shown that higher tumor grade and the basal-like phenotype are associated with elevated MYC

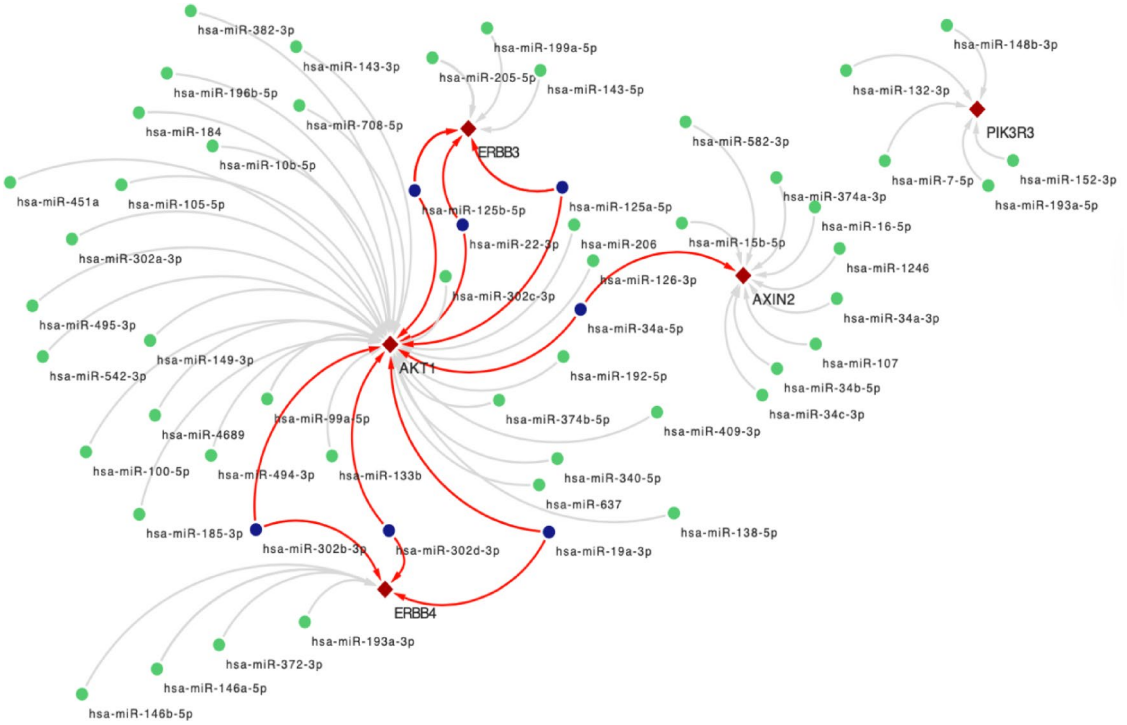


Fig. 12. miRNA and associated marker gene interacting network: Red shapes represent marker genes, green represents miRNA targeting single gene and blue dots are miRNA targeting multiple genes.

expression[42]. However, our analysis indicates that MYC is downregulated in all TNBC subtypes. This finding is intriguing and warrants further investigation.

In our study, we found that the expression level of HIVEP2 was significantly lower in BC tissue than in normal breast tissue, which is a typical event in BC cases[43]. Another TF, STAT1 suppresses mammary tumor formation and is known to be a tumor suppressor; however, its expression is frequently lost as BC progresses[44]. In contrast, we observed significant upregulation of STAT1 in the BL2 TNBC subtype. NFKBIA, a known tumor suppressor gene, is frequently deleted in BC, and this deletion is closely associated with TNBC development[45]. We also found that NFKBIA gene expression was downregulated in both BL2 and MSL subtypes. The TF MYB was upregulated in the LAR subtype. Previous studies have shown that MYB is required for tumor cell proliferation and is highly expressed in almost all androgen-receptor-positive breast tumors[46]. AR is expressed in 60–90% of breast cancers and is more prevalent in estrogen receptor-positive (ER+) tumors and luminal androgen receptor-positive TNBCs[47,48]. Our study yielded similar results for AR, particularly in the LAR subtype. CEBPB plays well-established roles in mammary gland development. It is overexpressed in BC, and has tumor-promoting function[49]. However, in the LAR subtype, we found CEBPB to be significantly downregulated. TWIST1 is a master EMT-inducing gene and plays a critical role in cancer metastasis in various malignant tumors. Overexpression of TWIST1 promotes TNBC metastasis via and is associated with poor prognosis[50]. In both LAR and MSL subtypes, TWIST1 was significantly downregulated.

Earlier studies discovered that ETS1 and other ETS family members, including ETS2, ETV1, ETV5, ELK3, and FLI1, were upregulated in TNBC cell lines compared to other BC cell lines. Elevated expression of ETS family members (e.g., ETS1) might contribute to chemotherapy resistance as well as tumor invasion and aggressiveness in TNBC cells[51]. In contrast, we detected that in the LAR TNBC subtype, ETS1 and ETS2 were significantly downregulated. These subtype-specific results present a promising area for further investigation. HOXB7 upregulates TGFβ2, driving both tumor cell-autonomous mechanisms and tumor-stromal interactions that contribute to an aggressive cancer phenotype and progression to metastasis. It has been demonstrated to enhance cell migration, invasion, epithelial-to-mesenchymal transition (EMT), and angiogenesis, facilitating tumor progression in breast carcinomas through TGFβ2 upregulation[52].

After validating the expression pattern of driver-mutated candidate genes in GEPIA, we considered ERBB3, ERBB4, PDGFRA, AKT1, PIK3R3, GNAS, HLA-A, HLA-B, FGFR4, and AXIN2 as risk genes in TNBC. Next, we hypothesized these risk genes as markers following the verification of their protein expression level using the HPA database.

ERBB3 and ERBB4 genes belong to the human epidermal growth factor receptor (HER) family. This family includes the epidermal growth factor receptor (EGFR, also known as HER1 or erbB1), HER2 (erbB2), HER3 (erbB3), and HER4 (erbB4). HER receptors are often overexpressed in human breast cancers and play critical roles in tumor initiation and progression[52,53]. HER3 is known to form heterodimers with other receptors to promote cell growth and survival-related signaling. Although there is limited understanding of the effect of HER3 in TNBC, recent studies have suggested that HER3 is a poor prognostic marker in hormone receptor-

negative BC[52–54]. After checking the mutation levels and performing survivability analysis, we conclude that ERBB3 (HER3) overexpression may be attributed to its driver mutations, which ultimately contribute to poor overall survival. TWIST1 and TWIST2 are key transcription regulators of the ERBB3 gene, and both negatively regulate its expression. In our study, we found these two TFs to be significantly downregulated in the BL1 subtype. Therefore, ERBB3 can be considered a potential marker gene and therapeutic target for BL1 TNBCs.

Like ERBB3 and other HER family genes, ERBB4 is also highly expressed in invasive breast carcinoma (IBC) and regulates cell proliferation, migration, and invasion. Recently, ERBB4 overexpression has been identified as a marker for poor prognosis in ER-negative BC[55]. Another study reported that ERBB4 exerts its effect on the proliferation, migration, and invasion of IBC cells by regulating the expression of PDGFRA[56]. In our study, we found similar results regarding the expression of ERBB4 and PDGFRA. GLI1 is one of the key transcription regulators of PDGFRA, positively regulating its expression, and showed significant upregulation in BL1, BL2, and ML subtypes. In the BL2 subtype, both ERBB4 and PDGFRA were upregulated and appeared among the top 5% of common genes within the top 10% of biological pathways. During survivability analysis, PDGFRA has been recognized as a cancer-benefit gene, i.e., these genes help cancer progression but do not cross the hazard index 1 in survival analysis. Therefore, they facilitate tumor progression but are not directly associated with mortality. However, according to the pathology data (retrieved from the HPA database) no detectable expression of PDGFRA was observed. Thus, the ERBB4/PDGFRA signaling axis could represent an important focus for future research. ERBB4 may serve as a marker gene and a crucial component of therapeutic applications in future.

Several previous studies addressed that AKT1 increases cell proliferation by regulating cell cycle proteins, including P21, P27, and cyclin D1, and impairs apoptosis[57]. While most studies recognize AKT2 as the main driver of cancer progression and metastasis, AKT1 has been reported as an important tumor initiator. AKT appears to be a reasonable therapeutic target because, in most BC cases, the PI3K/AKT signaling pathway is frequently dysregulated[57,58]. We also observed that nearly all TNBC subtypes showed a significant enrichment in the PI3K/AKT signaling pathway. This pathway was interconnected with other biological pathways through a common set of genes, supporting cancer growth, progression, survival, and metastasis. In the LAR subtype, the AKT1 gene was upregulated and appeared as one of the candidate genes. AKT1 is regulated by TF AR which was highly upregulated in LAR. Our survivability and validation analysis suggests that AKT1 may serve as a potential marker for TNBC subtype-specific studies[58,59]. Although AKT1 inhibitory drugs are approved by the FDA (Food and Drug Administration) for cancer treatment, further studies are required to apply them with high precision in TNBC subtype-specific therapies.

In both BL2 and LAR subtypes, GNAS has been identified as a risk gene based on expression, driver mutation, and survivability analyses. Conditions such as poor patient survival, tumor growth, and metastasis are strongly correlated with high GNAS expression in BC. GNAS promotes BC cell proliferation and metastasis through the PKA/PI3K/AKT/Snail1/E-cadherin axis[60]. The GNAS/PKA axis triggered the PI3K/AKT signaling pathway[61,62]. In our study, PIK3R3 and AKT1 were identified as marker genes in the LAR subtype. Thus, we predict that GNAS has a strong connection with these two markers. GNAS may be regarded as a marker gene, a potential prognostic marker, and a therapeutic target for both BL2 and LAR subtypes. These molecular leads in our work suggest combination strategies (for example AR + PIK3R3/AKT pathway targeting) that go beyond single-agent antiandrogen therapy.

While exploring subtype-specific molecular patterns, we observed that the BL2 subtype was associated with increased HLA-B expression. HLA-B expression strongly correlates with T cell activation in BC. Among the subtypes, the basal-like subtype showed the greatest association with T-cell activation, the strongest link to prognosis, and the highest expression of HLA-A and HLA-B[63]. During survivability analysis, both HLA-A and HLA-B were recognized as cancer-benefit genes. Although pathology data retrieved from the HPA database indicated a tendency toward high to medium protein expression levels, no information related to HLA-A expression was found. HIVEP2 and MYC are two of the key transcription regulators of HLA-B, both of which were significantly downregulated in the BL2 subtype. However, according to the TRRUST database, regulatory role of HIVEP2 is unknown, whereas MYC is annotated as an activator. These contradictory findings suggest the need for further investigation.

Fibroblast growth factor receptor 4 (FGFR4) has been implicated in the progression of BC and represents a promising therapeutic target for BC and other solid tumors[64]. FGFR4 activation stimulates various signaling pathways, such as the MAPK/ERK, which are critical for cell proliferation and differentiation. Once activated, it promotes cell cycle progression and survival[65]. FGFR4 also stimulates the PI3K/Akt pathway, enhancing cell survival and growth. Activation of PI3K leads to Akt phosphorylation, which promotes anti-apoptotic signaling. This contributes to the aggressive nature of BC[65,66]). Moreover, FGFR4 may influence the tumor microenvironment by modulating the behavior of stromal cells and immune cells, potentially enhancing tumor growth and metastasis. It has also been linked to EMT, a process where epithelial cells gain migratory and invasive properties[67]. This transition is crucial for cancer metastasis, making FGFR4 a significant factor in TNBC progression. In our study, we found that in both BL1 and ML subtypes, FGFR4 was upregulated and appeared as one of the candidate genes. FGFR4 emerged as a cancer risk gene, which eventually lowered the overall patient's survivability. Gene level and protein level expression analyses suggested that FGFR4 could serve as a potential marker for BL1 and ML subtypes. While identifying the key transcription regulators using TRRUST database network, we found IKZF1 is a transcription regulator of the FGFR4, and was significantly upregulated in both BL1 and ML subtypes. To date, there is no published research directly reporting that IKZF1 transcriptionally regulates FGFR4-mediated mesenchymal-like progression in triple-negative breast cancer (TNBC). While in the pituitary, IKZF1 acts as a key regulator of FGFR4[68]; its potential interaction in TNBC progression remains unexplored. Therefore, we predict there might be some relation between these two genes in TNBC progression, especially in BL1 and ML subtypes.

The role of AXIN2 in cancer progression is still a topic of debate. While AXIN2 has traditionally been identified as a tumor suppressor, recent studies suggest it may exhibit oncogenic properties in certain cancers, including BC. Emerging evidence indicates that elevated AXIN2 expression could significantly contribute to BC progression, highlighting its dual role in cancer biology[69]. AXIN2 overexpression has been shown to upregulate Snail1, a key zinc-finger TF, by modulating the nuclear export activity of GSK3 in BC cell lines. This regulatory interaction suggests a mechanism by which AXIN2 influences cancer cell behavior, contributing to tumor progression. Growing evidence connects hyperactive Wnt signaling to BC progression, primarily through the β-catenin–TCF complex, which drives epithelial–mesenchymal transition (EMT). The canonical Wnt pathway promotes tumor dedifferentiation and invasiveness via an AXIN2-dependent mechanism that stabilizes Snail1, a key EMT regulator. AXIN2 acts as a nucleocytoplasmic chaperone for GSK3β, controlling Snail1 turnover. This β-catenin–TCF-regulated AXIN2–GSK3β–Snail1 axis provides new insights into the mechanisms underlying cancer-associated EMT[17]. MYC is one of the key TFs regulating AXIN2. According to the TRRUST database, its mode of regulation is unknown. However, we observed contrasting results, with MYC showing significant down-regulation in our study. This discrepancy highlights the need for further investigation.

Our study provides new insights into the roles of transcription factors (TFs) and their target genes across triple-negative breast cancer (TNBC) subtypes. Several unexpected patterns emerged that challenge existing knowledge (Supplementary Table S6).

1. Contrary to earlier reports linking MYC to higher tumour grade and basal-like phenotypes, we observed its downregulation in BL1, BL2, and MSL subtypes. In BL2, both HLA-A and HLA-B were significantly upregulated, while their regulators HIVEP2 and MYC were downregulated. Since TRRUST annotates MYC as an activator of HLA-B and the role of HIVEP2 is unknown, these findings suggest alternative or compensatory regulatory mechanisms.
2. In the LAR subtype, we identified GNAS as a risk gene influencing the PKA/PI3K/AKT/Snail1/E-cadherin axis, and proposed ERBB4 and PIK3R3 as LAR-specific markers. These results point toward potential combination therapies targeting AR alongside PIK3R3/AKT signalling. We also found significant downregulation of ETS1 and ETS2 in LAR, which contrasts with their previously reported oncogenic roles.
3. In ML and MSL subtypes, we identified AXIN2 as a marker gene, with MYC implicated as a regulator, though its observed downregulation suggests involvement of other TFs. Additionally, IKZF1 was significantly upregulated in BL1 and ML, and TRRUST links it to FGFR4. While this regulation is known in the hematopoietic system and pituitary, its role in TNBC remains unexplored, suggesting a novel regulatory axis.
4. Surprisingly, the EMT master regulator TWIST1 was downregulated in MSL, indicating that EMT in these subtypes may be driven by TWIST1-independent pathways. Finally, while STAT1 is usually lost during breast cancer progression, it was upregulated in BL2, highlighting a unique, subtype-specific function.

Our findings highlight that TNBC subtype-specific cell lines exhibit distinct activation or suppression of pathway-related genes, along with transcriptional regulators, and subtype-specific molecular signatures (Fig. 13). These insights provide valuable leads toward unraveling the molecular heterogeneity of TNBC and hold strong potential for guiding more precise and personalized therapeutic strategies for patients.

Biomarker-based patient stratification to tailor therapy is a future emerging strategy to handle TNBC. During our study we have found many FDA approved drugs against our subtype specific predictive marker genes with categorized and non-categorized ones by the drug database. These groups of drugs may be further evaluated to choose the most effective one against a specific marker on the basis of inhibitory efficiency. Therefore, further molecular docking and dynamic simulation study is required. Drugs targeting more multiple genes can be useful for combinational therapy. Drugs that have dual interactions include Vandetanib, Gefitinib, Cucurmin, Afatinib, Dacomatinib anhydrous (for ERBB3 and ERBB4); Everolimus, Nintedanib esylate (for FGFR4 and AKT1); Fluorouracil drug (for GNAS and FGFR4); Amitriptyline hydrochloride (for GNAS and AKT1); Infigratinib (PIK3R3 and FGFR4). Drugs having triple interactions include Lapatinib (for HLA-B, ERBB3 and ERBB4) and Erlotinib (ERBB3, ERBB4 and FGFR4). We found most of the common targeted drugs among the potential markers are antineoplastic agents with a few exceptions such as Everolimus and Amitriptyline hydrochloride (categorized as immunosuppressive and analgesic respectively in DGI database). In this context, repurposing of the existing pharmacological agents that may be effective against the identified molecular targets of TNBC will pave the way for a faster and more cost-effective approach to drug development. We did not find any drugs against the AXIN2 gene which also presents an essential future scope of study. From molecular docking study we found DOCETAXEL ANHYDROUS as a small molecule agent and CETUXIMAB as a monoclonal antibody appeared to be a better therapeutic choice for targeting ERBB3 marker. Drugs like TOPOTECAN HYDROCHLORIDE, FULVESTRANT, and NINTEDANIB ESYLATE had strong binding affinity to AKT1. NINTEDANIB ESYLATE and PONATINIB appeared to be comparatively good therapeutic choices for targeting FGFR4. For HLA-B, PEGINTERFERON ALFA-2 A and INTERFERON BETA-1 A had strong binding affinity.

miRNA based therapeutic approaches involve using or blocking microRNAs (short, non-coding RNA molecules that regulate gene expression). They include miRNA mimics to replace lost or down-regulated miRNAs and antagomirs (anti-miRs) to inhibit the function of overexpressed miRNAs. For miRNAs that promote cancer disease (oncomiRs), synthetic anti-miRs (also called antagomirs) are used to block their activity and reduce gene expression they regulate[70,71]. We found a number of miRNAs targeting potential marker genes by regulating post-transcriptional modifications which might be directly linked to the aggressive cancer progression. There are miRNAs that target multiple genes such as hsa-miR-125b-5p, hsa-miR-22-3p, and hsa-miR-125a-5p miRNA (target ERBB3 and AKT1); hsa-miR-302b-3p, hsa-miR-302d-3p, and hsa-miR-19a-3p (target ERBB4 and AKT1); hsa-miR-126-3p (targets ERBB3 and AKT1). These common miRNAs may be subjected to emerging

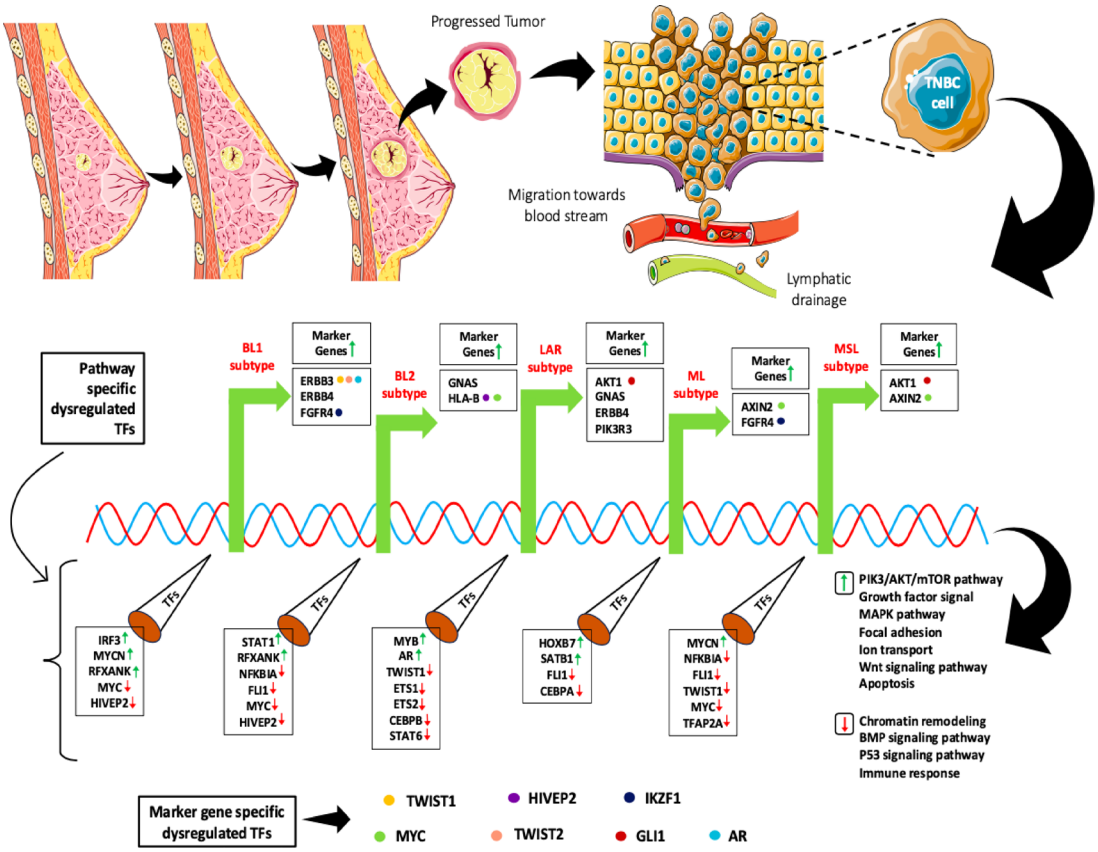


Fig. 13. Work summary: TNBC subtype-specific cell lines with distinct activation or suppression of pathway-related genes, and their essential TFs, and subtype-specific marker genes, and TFs regulating those markers.

miRNA-targetable therapeutic strategies based on post-transcriptional regulation of the specific markers to mitigate TNBC.

One of the main limitations of our study is the inability to fully validate the roles of the identified subtype-specific genes. This was primarily due to the limited availability of patient samples and the lack of complementary proteomics experiments, which would have provided stronger functional confirmation. Future studies incorporating larger patient cohorts and proteomic validation will be essential to strengthen and build upon our preliminary research.

Conclusions

This study used highly correlated cell lines as replicates to compare each TNBC subtype to a normal breast epithelial cell line. Instead of using BC patients sample repositories, we chose these cell lines for our current investigation to reduce the biological and technological variability associated with heterogeneous samples. Our findings are classified into three categories: essential TFs regulating TNBC subtype-specific pathways, subtype-specific marker genes, and TFs regulating those marker genes. IRF3, MYCN, RFXANK, MYC, and HIVEP2 are the essential TFs that regulate BL1 subtype-specific pathways. TFs such as STAT1, RFXANK, NFKBIA, FLI1, MYC, and HIVEP2 are responsible for regulating the BL2 TNBC subtype-specific pathways. MYB, AR, TWIST1, ETS1, ETS2, CEBPB, and STAT6 TFs regulate essential pathways for the LAR subtype progression. The major transcriptional regulators of pathway-specific genes for the ML subtype are HOXB7, SATB1, FLI1, and CEBPA. The most dysregulated TFs that regulate essential pathways specific to the MSL subtype are MYCN, NFKBIA, FLI1, TWIST1, MYC, and TFAP2A. The BL1 subtype-specific marker genes are ERBB3, ERBB4, and FGFR4. The BL2 subtype-specific marker genes include GNAS, and HLA-B. The LAR subtype-specific marker genes are ERBB4, AKT1, GNAS and PIK3R3. The ML subtype-specific marker genes include, FGFR4 and AXIN2. The MSL subtype-specific marker genes are AKT1 and AXIN2. The TFs regulating the marker gene ERBB3 are TWIST1, TWIST2, and AR, all showing significantly varied expression levels. Significantly varied expression has been found for HIVEP2 and MYC, which regulate the marker gene HLA-B. TFs such as GLI1 and AR showed significantly varied expression, which is associated with the regulation of AKT1. MYC is one of the major TF for AXIN2 and has shown a varied expression level. In future, the best fit drug among already approved antineoplastic agents and repurposing of the existing pharmacological agents may be effective against the identified molecular targets of TNBC. Common miRNAs such as such as hsa-miR-125b-5p, hsa-miR-22-3p, hsa-miR-125a-5p, hsa-miR-302b-3p, hsa-miR-302d-3p, hsa-miR-19a-3p, and hsa-miR-126-3p may be subjected to emerging miRNA-targetable therapeutic strategies based on their role in post-transcriptional regulation of

the specific markers to mitigate TNBC. Finally, we emphasize that these findings highlight the need for further studies on cancer marker genes and the development of personalized therapeutics to advance precision cancer treatment. Future analyses should consider precise driver mutation positions, incorporate additional proteomic research, conduct wet lab expression and validation studies, and include data from real patients with diverse racial backgrounds. Additionally, factors such as geographical location and nutrition should be taken into account for understanding the epigenetic regulations.

Data availability

This study was carried out using publicly available data from SRA [https://www.ncbi.nlm.nih.gov/sra] with acce ssion number SRR22809594, SRR22809593, SRR22809592.

Received: 1 February 2025; Accepted: 7 November 2025
Published online: 24 December 2025

References

1. Harbeck, N. F., Penaul tL lorca, F. & Cortes, J. Breast cancer. *Nat. Rev. Dis. Primers*. **5**(1), 66 (2019).
2. Dai, X. et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am. J. Cancer Res.* **5**(10), 2929 (2015).
3. Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature*. **406**(6797), 747–752 (2000).
4. Sørlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* **98**(19), 10869–10874 (2001).
5. Fragomeni, S. M., Sciallis, A. & Jeruss, J. S. Molecular subtypes and local-regional control of breast cancer. *Surg. Oncol. Clin.* **27**(1), 95–120 (2018).
6. Lehmann, B. D. et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PloS One*. **11**(6), e0157368 (2016).
7. Yam, C., Mani, S. A. & Moulder, S. L. Targeting the molecular subtypes of triple negative breast cancer: Understanding the diversity to progress the field. *Oncologist* **22**(9), 1086–1093 (2017).
8. Yin, L., Duan, J. J., Bian, X. W. & Yu, S. C. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* **22**, 1–13 (2020).
9. Mandapati, A. & Lukong, K. E. Triple negative breast cancer: approved treatment options and their mechanisms of action. *J. Cancer Res. Clin. Oncol.* 1–19 (2022).
10. Tong, L., Yu, X., Wang, S., Chen, L. & Wu, Y. Research progress on molecular subtyping and modern treatment of triple-negative breast cancer. *Breast Cancer: Targets Ther.* 647–658 (2023).
11. Marra, A., Trapani, D., Viale, G., Criscitiello, C. & Curigliano, G. Practical classification of triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug resistance, and novel therapies. *NPJ Breast Cancer*. **6**(1), 54 (2020).
12. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Investig.* **121**(7), 2750–2767 (2011).
13. Hu, H. et al. Subtyping of triple-negative breast cancers: its prognostication and implications in diagnosis of breast origin. *ESMO open*. **9**(4), 102993 (2024).
14. Obidiro, O., Battogtokh, G. & Akala, E. O. Triple negative breast cancer treatment options and limitations: future outlook. *Pharmaceutics* **15**(7), 1796 (2023).
15. Shan, R., Dai, L. J., Shao, Z. M. & Jiang, Y. Z. Evolving molecular subtyping of breast cancer advances precision treatment. *Cancer Biology Med.* **21**(9), 731 (2024).
16. Rodríguez-Bautista, R. et al. Immune milieu and genomic alterations set the triple-negative breast cancer immunomodulatory subtype tumor behavior. *Cancers* **13**(24), 6256 (2021).
17. Vale, N. et al. Normal breast epithelial MCF-10A cells to evaluate the safety of carbon dots. *RSC Med. Chem.* **12**(2), 245–253 (2021).
18. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinform.* **19**(1), 1–24 (2018).
19. Jalili, V. et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* **48**(W1), W395–W402 (2020).
20. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1), 15–21 (2013).
21. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**(6588), 44–53 (2022).
22. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**(7), 923–930 (2014).
23. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47–e47 (2015).
24. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**(W1), W216–W221 (2022).
25. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**(D1), D672–D677 (2025).
26. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**(D1), D457–D462 (2016).
27. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
28. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**(1), 1523 (2019).
29. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**(5), 401–404 (2012).
30. Han, H. et al. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* **5**(1), 11432 (2015).
31. Cannon, M. et al. DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res.* **52**(D1), D1227–D1235 (2024).
32. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape stringapp: network analysis and visualization of proteomics data. *J. Proteome Res.* **18**(2), 623–632 (2018).
33. Kern, F. et al. MiRTargetLink 2.0—interactive MiRNA target gene and target pathway networks. *Nucleic Acids Res.* **49**(W1), W409–W416 (2021).
34. Massihnia, D. et al. Triple negative breast cancer: shedding light onto the role of pi3k/akt/mtor pathway. *Oncotarget* **7**(37), 60712 (2016).
35. Jiang, W., Wang, X., Zhang, C., Xue, L. & Yang, L. Expression and clinical significance of MAPK and EGFR in triple–negative breast cancer. *Oncol. Lett.* **19**(3), 1842–1848 (2020).
36. Dey, N. et al. Wnt signaling in triple negative breast cancer is associated with metastasis. *BMC Cancer*. **13**, 1–15 (2013).

37. Liu, Y. R. et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res.* **18**, 1–10 (2016).
38. Li, S. et al. BMP9 inhibits the growth of breast cancer cells by downregulation of the PI3K/Akt signaling pathway. *Oncol. Rep.* **40**(3), 1743–1751 (2018).
39. Eskiler, G. G., Özkan, A. D., Kaleli, S. & Bilir, C. Inhibition of TLR4/TRIF/IRF3 signaling pathway by curcumin in breast cancer cells. *J. Pharm. Pharm. Sci.* **22**, 281–291 (2019).
40. Schafer, J. M. et al. Targeting MYCN-expressing triple-negative breast cancer with BET and MEK inhibitors. *Sci. Transl. Med.* **12**(534), eaaw8275 (2020).
41. Cui, Z. et al. Pan-cancer investigation of RFX family and associated genes identifies RFX8 as a therapeutic target in leukemia. *Heliyon*. **10**(15) (2024).
42. Green, A. R. et al. MYC functions are specific in biological subtypes of breast cancer and confers resistance to endocrine therapy in luminal tumours. *Br. J. Cancer*. **114**(8), 917–928 (2016).
43. Fujii, H. et al. Frequent down-regulation of HIVEP2 in human breast cancer. *Breast Cancer Res. Treat.* **91**, 103–112 (2005).
44. Koromilas, A. E. & Sexl, V. The tumor suppressor function of STAT1 in breast cancer. *Jak-Stat* **2**(2), e23353 (2013).
45. Bredel, M. et al. Deletion of the tumor suppressor NFKBIA in triple-negative breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **87**(2), S98 (2013).
46. Player, A. et al. Identification of candidate genes associated with triple negative breast cancer. *Genes Cancer*. **8**(7–8), 659 (2017).
47. Pistelli, M. et al. Androgen receptor expression in early triple-negative breast cancer: clinical significance and prognostic associations. *Cancers* **6**(3), 1351–1362 (2014).
48. Lee, M. et al. Luminal androgen receptor subtype and tumor-infiltrating lymphocytes groups based on triple-negative breast cancer molecular subclassification. *Sci. Rep.* **14**(1), 11278 (2024).
49. Matherne, M. G., Phillips, E. S., Embrey, S. J., Burke, C. M. & Machado, H. L. Emerging functions of C/EBPβ in breast cancer. *Front. Oncol.* **13**, 1111522 (2016).
50. Zhang, M. et al. Triptonide effectively inhibits triple-negative breast cancer metastasis through concurrent degradation of Twist1 and Notch1 oncoproteins. *Breast Cancer Res.* **23**, 1–16 (2021).
51. Jung, H. H. et al. Statins affect ETS1-overexpressing triple-negative breast cancer cells by restoring DUSP4 deficiency. *Sci. Rep.* **6**(1), 33035 (2016).
52. Lyu, H., Han, A., Polsdofer, E., Liu, S. & Liu, B. Understanding the biology of HER3 receptor as a therapeutic target in human cancer. *Acta Pharm. Sinica B*. **8**(4), 503–510 (2018).
53. Lyu, H. et al. HER3 functions as an effective therapeutic target in triple negative breast cancer to potentiate the antitumor activity of gefitinib and paclitaxel. *Cancer Cell Int.* **23**(1), 204 (2023).
54. Gonzalez-Conchas, G. A. et al. Epidermal growth factor receptor overexpression and outcomes in early breast cancer: A systematic review and a meta-analysis. *Cancer Treat. Rev.* **62**, 1–8 (2018).
55. Kim, J. Y. et al. Prognostic value of ERBB4 expression in patients with triple negative breast cancer. *BMC Cancer*. **16**, 1–12 (2016).
56. Gong, N., Wu, R., Ding, B. & Wu, W. ERBB4 promotes the progression of inflammatory breast cancer through regulating PDGFRA. *Translational Cancer Res.* **9**(5), 3266 (2020).
57. Xu, N., Lao, Y., Zhang, Y. & Gillespie, D. A. Akt: a double-edged sword in cell proliferation and genome stability. *J. Oncol.* **2012**(1), 951724 (2012).
58. Hinz, N. & Jücker, M. Distinct functions of AKT isoforms in breast cancer: a comprehensive review. *Cell. Communication Signal.* **17**, 1–29 (2019).
59. Ascione, L. et al. PIK3CAMutations in breast cancer subtypes other than HR-positive/HER2-negative. *J. Personalized Med.* **12**(11), 1793 (2022).
60. Jin, X. et al. Elevated expression of GNAS promotes breast cancer cell proliferation and migration via the PI3K/AKT/Snail1/E-cadherin axis. *Clin. Transl. Oncol.* **21**, 1207–1219 (2019).
61. Hunzicker-Dunn, M. E. et al. PKA and GAB2 play central roles in the FSH signaling pathway to PI3K and AKT in ovarian granulosa cells. *Proc. Nal. Acad. Sci.* **109**(44), E2979-E2988 (2012).
62. Law, N. C., White, M. F. & Hunzicker-Dunn, M. E. G protein-coupled receptors (GPCRs) that signal via protein kinase A (PKA) cross-talk at insulin receptor substrate 1 (IRS1) to activate the phosphatidylinositol 3-kinase (PI3K)/AKT pathway. *J. Biol. Chem.* **291**(53), 27160–27169 (2016).
63. Noblejas-López, M. D. M. et al. Expression of MHC class I, HLA-A and HLA-B identifies immune-activated breast tumors with favorable outcome. *Oncoimmunology*. **8**(10), e1629780 (2019).
64. Levine, K. M., Ding, K., Chen, L. & Oesterreich, S. FGFR4: A promising therapeutic target for breast cancer and other solid tumors. *Pharmacol. Ther.* **214**, 107590 (2020).
65. Liu, Y. et al. Dissecting the role of the FGF19-FGFR4 signaling pathway in cancer development and progression. *Front. cell. Dev. Biology*. **8**, 95 (2020).
66. Lang, L. & Teng, Y. Fibroblast growth factor receptor 4 targeting in cancer: new insights into mechanisms and therapeutic strategies. *Cells* **8**(1), 31 (2019).
67. Peláez-García, A. et al. FGFR4 role in epithelial-mesenchymal transition and its therapeutic value in colorectal cancer. *PloS One*. **8**(5), e63695 (2013).
68. Yu, S., Asa, S. L. & Ezzat, S. Fibroblast growth factor receptor 4 is a target for the zinc-finger transcription factor Ikaros in the pituitary. *Mol. Endocrinol.* **16**(5), 1069–1078 (2002).
69. Ahn, S. Y. & Jo, C. H. Implications of axis inhibition protein 2 in breast cancer progression. *In vivo.* **37**(2), 634–643 (2023).
70. Diener, C., Keller, A. & Meese, E. Emerging concepts of MiRNA therapeutics: from cells to clinic. *Trends Genet.* **38**(6), 613–626 (2022).
71. Menon, A., Abd-Aziz, N., Khalid, K., Poh, C. L. & Naidu, R. MiRNA: a promising therapeutic target in cancer. *Int. J. Mol. Sci.* **23**(19), 11502 (2022).

Acknowledgements

The authors would like to thank their parent institutes for providing the computational facilities.

Author contributions

C.M. conceived the study. C.M. and D.S. designed experiments. S.K.M. and C.M. supervised the project. D.S., A.M.A., S.R.A. collected data. D.S. and S.K. processed data and carried out computational analysis. D.S., S.K. and C.M. wrote the original draft. A.M.A., S.R.A., C.M. and S.K.M. revised the manuscript with comments from all authors. All authors approved the final version.

Declarations

Consent to publish
All authors have given their consent for the possible publication.

Ethics approval
This is an observational study. No ethical approval is required.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-28191-3.

Correspondence and requests for materials should be addressed to S.R.A., C.M. or S.K.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025