Journal of Gynecology Research
Volume 7 | Issue 1
ISSN: 2576-7623

Research Article Open Access

TCF3 as a potential Tumor-related Biomarker in Breast Cancer: Comprehensive Multi-omics Analysis

Li Wentao*

School of Pharmaceutical Sciences, Zhejiang Chinese Medical University, Hangzhou 310053, China

***Corresponding Author:** Li Wentao, School of Pharmaceutical Sciences, Zhejiang Chinese Medical University, Hangzhou 310053, China, E-mail: lwt9223@163.com

Citation: Li Wentao (2023) TCF3 as a potential Tumor-related Biomarker in Breast Cancer: Comprehensive Multi-omics Analysis. J Gynecol Res 7(1): 104

Received Date: August 29, 2023 **Accepted Date:** September 29, 2023 **Published Date:** October 01, 2023

Abstract

There is growing evidence suggesting a correlation between multiple copies of transcription factor-3 (TCF3) and various malignancies, yet its impact on patients with breast cancer remains uncertain. The present study utilized the Cancer Genome Atlas (TCGA) database, Gene Expression Omnibus (GEO) database, and the UALCAN database to obtain expression profiles and clinical data from patients with breast cancer. A comparative analysis was performed to assess the differential expression levels of TCF3 in breast cancer tissues as compared to normal breast tissues. TCF3 was found to be overexpressed in breast cancer and exhibited significant associations with T stage, age, race, PAM50 subtype, histological type, ER, PR and menopause status. Subsequently, analyses were performed to determine functional enrichment. The enriched pathways encompassed xenobiotic metabolism by cytochrome P450, chemical carcinogenesis-DNA adducts, extracellular matrix organization, cornified envelope formation, collagen biosynthesis and modifying enzymes, as well as immunoregulatory interactions between lymphoid and non-lymphoid cells. Single-sample gene set enrichment analysis (GSEA) was utilized to evaluate the infiltration of immune cells. The methylation status of TCF3 was analyzed using the UALCAN and MethSurv databases. The overexpression of TCF3 is a significant contributor to decreased disease-specific survival (DSS) and progression-free interval (PFI). Likewise, the low methylation status of TCF3 is also associated with an adverse prognosis. Finally, the differential expression of TCF3 between tumor and normal tissues was validated using data from multiple sources, including GSE22820 dataset, 49 pairs of tumor and paracancerous samples sequencing data, and CPTAC dataset. Moreover, proteomic analysis of the CPTAC dataset showed that TCF3 expression was associated with alterations of multiple tumor-related pathways. Therefore, TCF3 may serve as a new tumor-related biomarker.

Keywords: TCF3 (E2A); Breast Cancer; Biomarker; RNA-sequencing; Proteomics

Introduction

Throughout the world, breast cancer is the most prevalent malignancy that poses a significant threat to women's health [1,2]. Although breast cancer treatment has improved significantly, a considerable number of patients still experience therapy resistance or relapse after treatment [3,4]. Hence, current diagnostic and therapeutic methods need to be improved urgently. In breast cancer patients, the treatment decisions and prognosis mainly depend on molecular subtyping and the TNM (Tumor-Node-Metastasis) staging [5]. However, despite administering similar treatment regimens to patients with identical tumor stages and molecular subtypes, clinical outcomes may exhibit variability. This implies that the existing staging system is inadequate in sufficiently reflecting the diverse biological characteristics of breast cancer patients and accurately predicting prognosis [6]. Both patients and tumor cells have other intrinsic characteristics that may affect clinical results. Studying and using the features as biomarkers has the potential to augment the present TNM staging system, improve the accuracy of predicting clinical outcomes, facilitate personalized treatments for patients with breast cancer, and develop novel therapies.

TCF3 (Ensembl: ENSG00000071564), also known as transcription factor E2-alpha (E2A, not to be mistaken with TCF7L1 - previously called TCF3), is a member of the E-protein family of bHLH TFs, which also consists of HEB (TCF12) and E2-2 (TCF4).[7,8] The canonical Wnt/β-catenin signaling impacts a wide range of biological activities, including stem cell self-renewal, organ morphogenesis, and tumor formation [9]. In colorectal cancer (CRC) cells, the FoxM1-Wnt/β-catenin pathway is targeted by E2A to suppress the tumor-initiating capacity, and the expression of E2A in CRC clinical specimens shows a negative correlation with patients' progression-free interval [10]. In hepatocellular carcinoma (HCC), E2A promoted oncogenic dedifferentiation and proliferation by repressing acetyl-CoA synthesis, while knockdown of E2A restored acetyl-CoA levels and inhibited tumor growth [11]. Together, although the role of E2A in cancers remain largely undefined, it may serve as a novel regulatory link between normal stem cells and malignant cells.

In this study, our objects were to investigate the correlation between TCF3 expression and its clinical, pathological and prognostic roles, explore the molecular mechanisms involved, and analyze the infiltration of immune cells in breast cancer, using bioinformatics tools. These findings may assist clinicians in diagnosing breast cancer early, optimizing treatment plans and improving overall patient outcomes. Finally, we validated the differential expression of TCF3 using Gene Expression Omnibus (GEO) data and 49 paired samples sequencing data, and investigated the relationship between TCF3 phosphoprotein and tumor stemness using CPTAC dataset. The flow chart is shown in Figure 1.

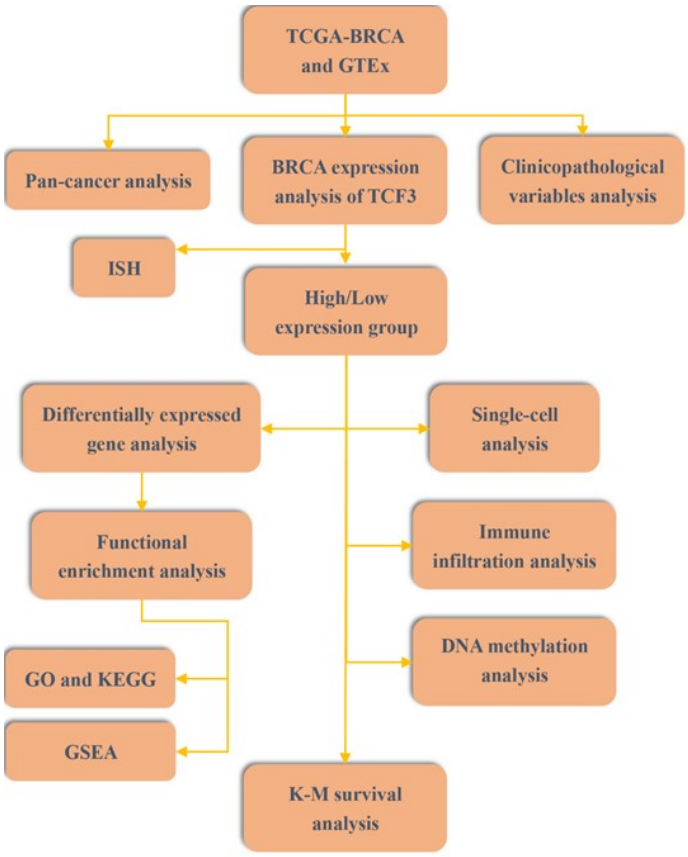

Figure 1: Analysis flow chart

Materials and Methods

Gathering and Handling of Information

The mRNA expression profiles and clinical data of breast cancer patients were collected from the Cancer Genome Atlas (TCGA) database1 and the Genotype-Tissue Expression Project (GTEx) database2. We obtained the TPM format RNA-sequencing data of TCGA and GTEx processed uniformly by the TOIL workflow [12] from the UCSC XENA database3 for pan-cancer analysis and breast cancer analysis (Figure 2A,B). We also downloaded the TPM format RNA-sequencing data of the TCGA-BRCA(invasive breast cancer) project processed by the STAR workflow from the TCGA database, and data of para-cancer and cancer samples matched with corresponding numbers were extracted for breast cancer analysis (Figure 2C). According to the characteristics of the data format, appropriate statistical methods (stats[4.2.1] package, car package) would be selected for statistics (if the statistical requirements were not met, statistical analysis would not be performed), and the ggplot 2 [3.3.6] package would be used to visualize the data.

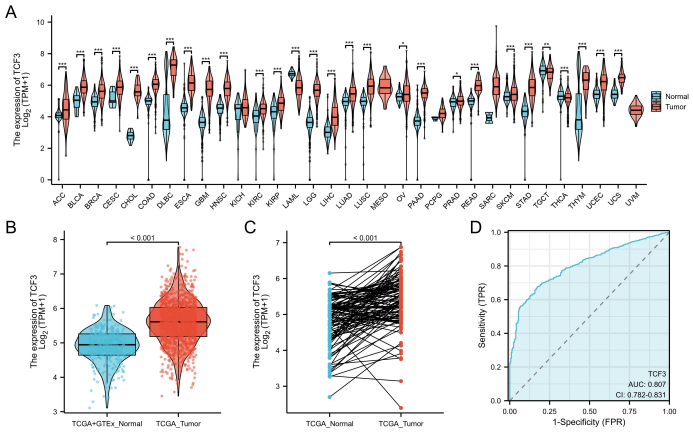

Figure 2: The study examines the expression levels of TCF3 in various tumor types, with a specific focus on breast cancer. (A) The analysis includes comparisons of TCF3 expression in tumors versus normal tissues using data from the TCGA and GTEx databases. (B) Additionally, the study investigates TCF3 expression in breast cancer compared to non-matched normal tissues in the TCGA and GTEx databases, (C) and matched normal tissues in the TCGA database. (D) Furthermore, the study evaluates the performance of ROC curves in classifying breast cancer versus normal breast tissues using data from the TCGA database. The abbreviations TCGA, GTEx, and ROC stand for the Cancer Genome Atlas, Genotype Tissue Expression Project, and receiver operating characteristic, respectively. *p < 0.05, **p < 0.05, and ***p < 0.05.

Differences in Protein Levels of TCF3 Expression

We obtained immunohistochemical images detecting TCF3 expression in normal breast tissue and breast cancer tissue from the Human Protein Atlas database[4] with an approved reliability score (Figure 3A-D). In addition, we obtained information on the subcellular localization of TCF3 expression with enhanced reliability score (Figure 3E,F).

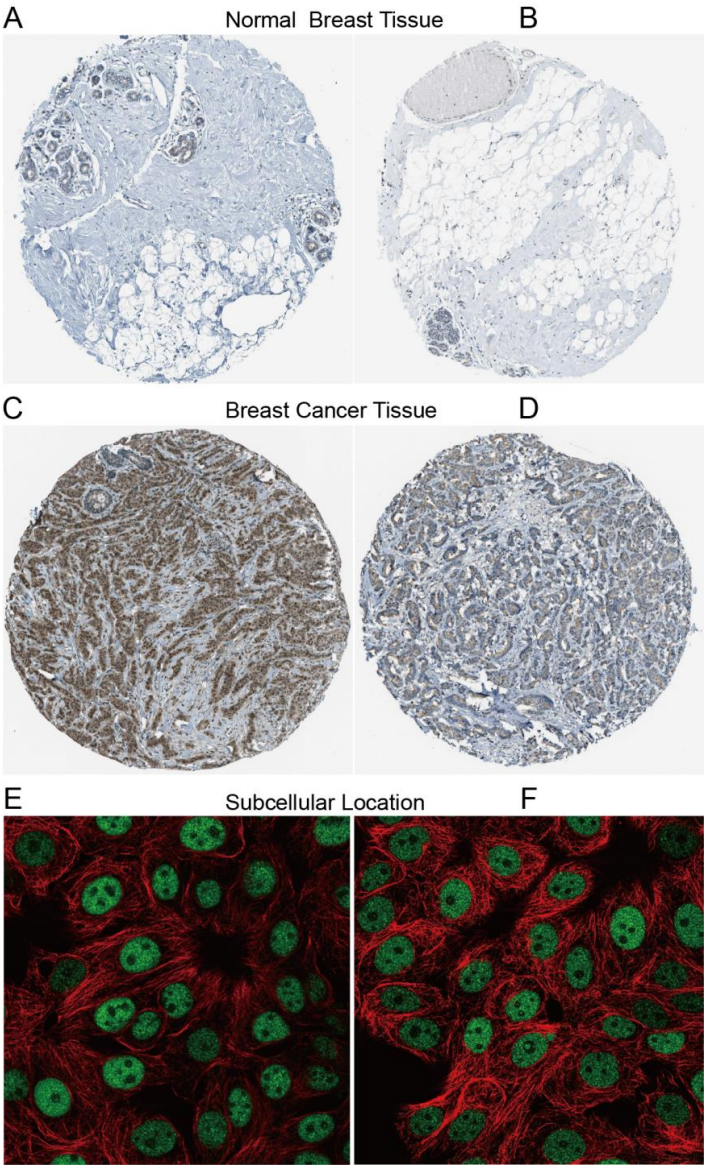

Figure 3: (A) Patient id: 2773; Female, age 23; Normal tissue, NOS (M-00100); Adipocytes Staining: Not detected; Glandular cells Staining: Low; Myoepithelial cells Staining: Not detected (B) Patient id: 2042; Female, age 75; Normal tissue, NOS (M-00100); Adipocytes Staining: Not detected; Glandular cells Staining: Low; Myoepithelial cells Staining: Not detected (C) Patient id: 2091; Female, age 40; Duct carcinoma (M-85003); Tumor cells Staining: Medium (D) Patient id: 1939; Female, age 87; Duct carcinoma (M-85003); Tumor cells Staining: Medium (E) Cell line: MCF7; Antibody: CAB018315; Green: Target protein; Red: Microtubules (F) Cell line: MCF7; Antibody: HPA062476; Green: Target protein; Red: Microtubules

TCF3 Expression and Clinicopathologic Variables

To study associations between TCF3 expression and clinicopathologic variables, we used the TPM format RNA-sequencing data of TCGA-BRCA project processed by the STAR workflow and clinical data. According to the characteristics of the data, appropriate statistical methods were selected for statistical analysis, including the Chisq test and Yates' correction Chisq test. Samples that were female and had not received neoadjuvant therapy were retained in the clinical data. PAM50 subtyping data were obtained from the study of Berger et al.[13] The visualization results were shown in Figure 4.

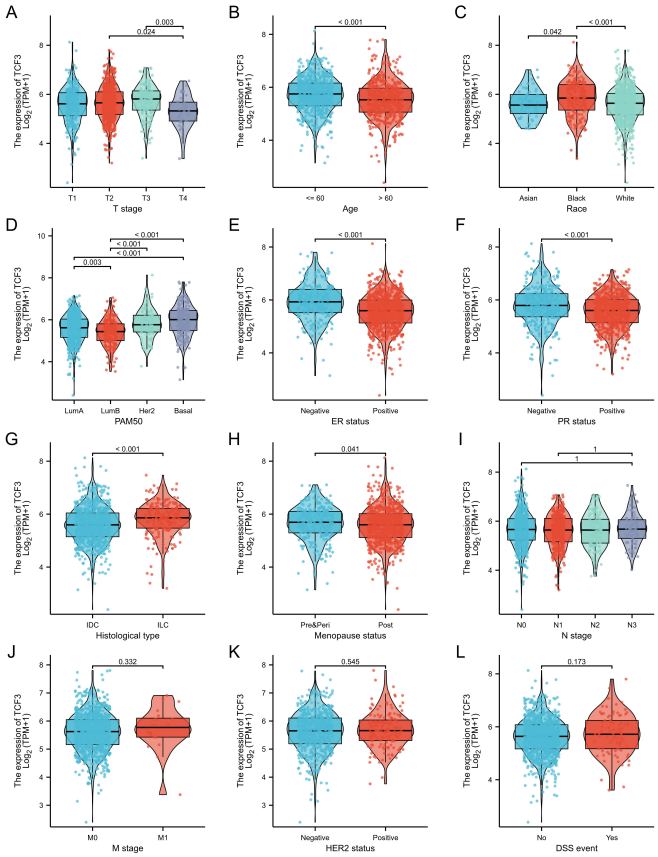

Figure 4: This study examines the associations between TCF3 expression and various clinicopathological characteristics, **(A-L)** including T stage, age, race, PAM50 subtype, ER status, PR status, histological type, menopause status, N stage, M stage, HER2 status, and DSS event. The abbreviations used in this figure are as follows: LumA for Luminal A, LumB for Luminal B, Her2 for her-2 positive, Basal for basal-like, IDC for infiltrating ductal carcinoma, ILC for infiltrating lobular carcinoma, DSS for disease-specific survival, ER for estrogen receptor, PR for progesterone receptor, and HER2 for human epidermal growth factor receptor 2.

Differential Expression Analysis of Gene

Based on the median score of TCF3 expression, the TCGA breast cancer patients were categorized into two groups, high and low TCF3 expression groups. The R package DESeq2[1.36.0][14] and edgeR[3.38.2][15] were utilized for conducting the analysis of differentially expressed genes (DEGs), with a significance threshold of adjusted p-value (p.adjust) <0.05 and |LogFC|>1. Spearman correlation analysis was utilized to assess the relevance between TCF3 and the expression of selected ten significant DEGs. The analysis results were visualized using the ggplot 2 for the volcano plot and co-expression heat map (Figure 5A,B).

Functional Enrichment Analysis

Using the R package clusterProfiler[4.4.4] and GOplot[1.0.2], enrichment analyses were carried out to study the function of DEGs, specifically through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (Figure 5C,D). After ID conversion (package org.Hs.eg.db) of the input molecule list, enrichment analysis was performed using the clusterProfiler package [16]. The GOplot package was then utilized to calculate the z-score value for each enriched entry based on the provided molecules values [17]. Gene Set Enrichment Analysis (GSEA) was conducted using the clusterProfiler R package, as depicted in Figure 5E [16,18]. Enriched function or pathway terms with a p.adjust value less than 0.05 and false discovery rate (FDR) less than 0.25 were deemed statistically significant.

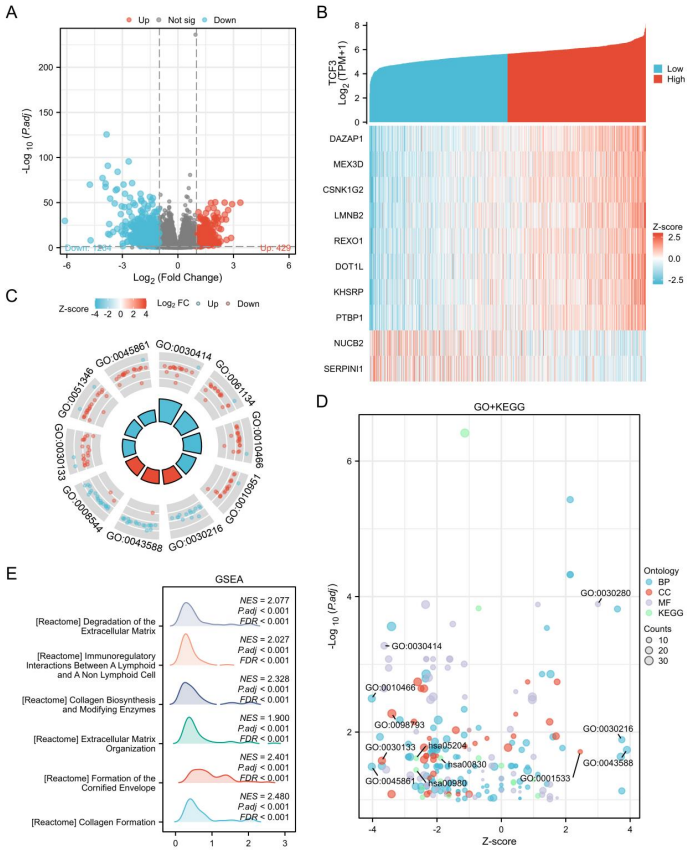

Figure 5: The figure shows TCF3-related differentially expressed genes (DEGs) and functional enrichment analysis of TCF3 in breast cancer using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA). The results are presented in two figures: (A) a volcano plot illustrating the down-regulated (blue dots) and up-regulated (red dots) DEGs, and (B) a heatmap displaying the correlation between TCF3 expression and 10 significant DEGs. (C) Circle diagram for GO analysis of DEGs. (D) Bubble diagram for GO and KEGG analysis of DEGs. (E) Gene set enrichment analysis (GSEA) of DEGs. *p < 0.05, **p < 0.05, and ***p < 0.05.

Single-cell Analysis

CancerSEA5 [19] is a specialized database that seeks to comprehensively decipher unique functional states of cancer cells at the level of individual cells, encompassing 14 distinct functional states observed in a total of 41,900 cancer single cells across 25 different cancer types. Analysis was conducted on the correlation between TCF3 expression and various tumor functions using single-cell sequencing data (Figure 6).

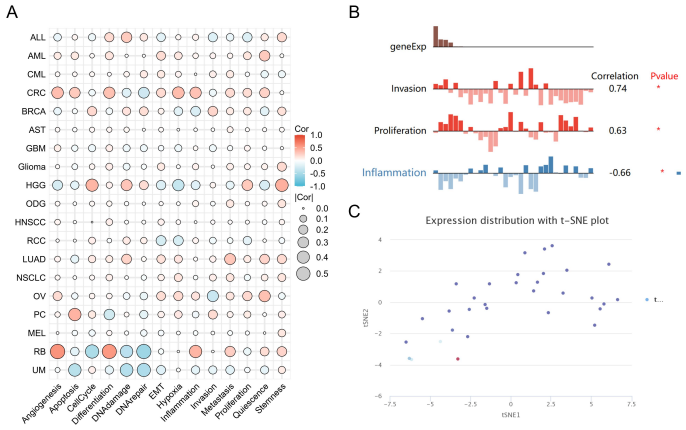

Figure 6: (A) The CancerSEA tool was utilized to investigate the correlation between TCF3 expression and various functional states in tumors at the single-cell level. *p < 0.05. (B,C) There were 3 functional states that were significantly related to TCF3 expression; TCF3 expression profiles were shown at single cells from BRCA by T-SNE diagram. ExpID: EXP0055; Dataset name: Aceto N. Mol Cancer Res. 2018 (CTC). This dataset contains 18 cell groups, which denoted different patients. Description: RNA sequencing of circulating tumor cells (CTCs) obtained from blood samples of women with metastatic estrogen receptor (ER)+ breast cancer, comparing cases with progression in bone versus visceral organs.

Immune Infiltration Analysis

The RNAseq data of the TCGA-BRCA project processed by STAR workflow were downloaded and sorted, and the TPM format data and clinical data were extracted. Based on the single-sample GSEA algorithm from the R package-GSVA [1.46.0],[20] the level of immune infiltration was calculated using 24 immune cell markers that were provided in an article published by Immunity [21] (Figure 7).

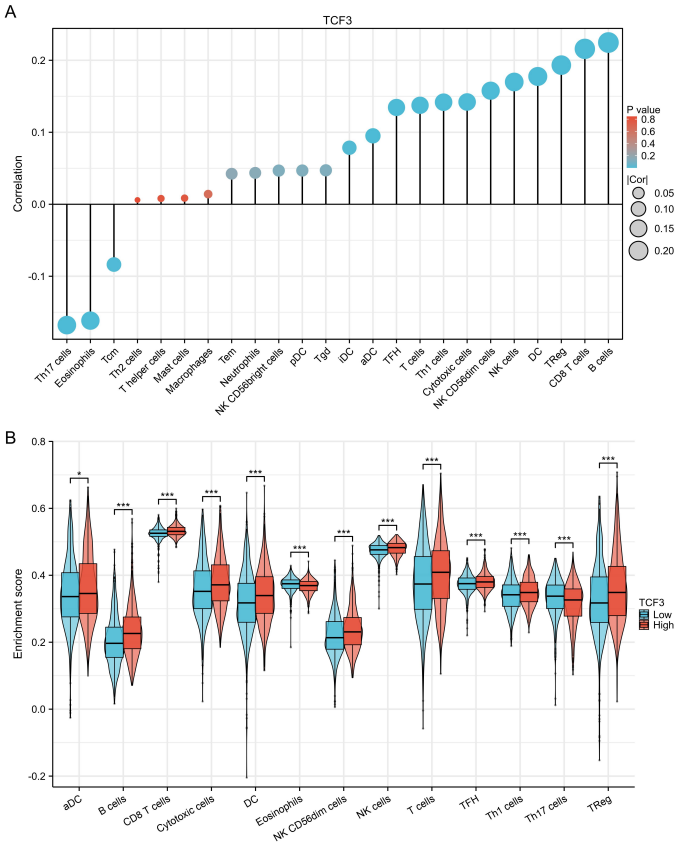

Figure 7: The present study examines the correlation between TCF3 expression and the level of immune infiltration in breast cancer. Specifically, (A) the relationship between TCF3 expression and the relative abundance of 24 distinct immune cell types is investigated, with the size of the dot representing the absolute Spearman's correlation coefficient values. Additionally, (B) a comparison for the levels of immune infiltration of 13 immune cell types is made between the high- and low-TCF3 expression groups, with significant differences.

The Spearman correlation analysis was utilized to investigate the association between TCF3 expression and the aforementioned immune cells, and we compared the extent of immune infiltration in high versus low TCF3 expression groups based on Wilcoxon rank sum test (Figure 8). Finally, the results were visualized by the ggplot 2 [3.3.6], stats[4.2.1] and car package.

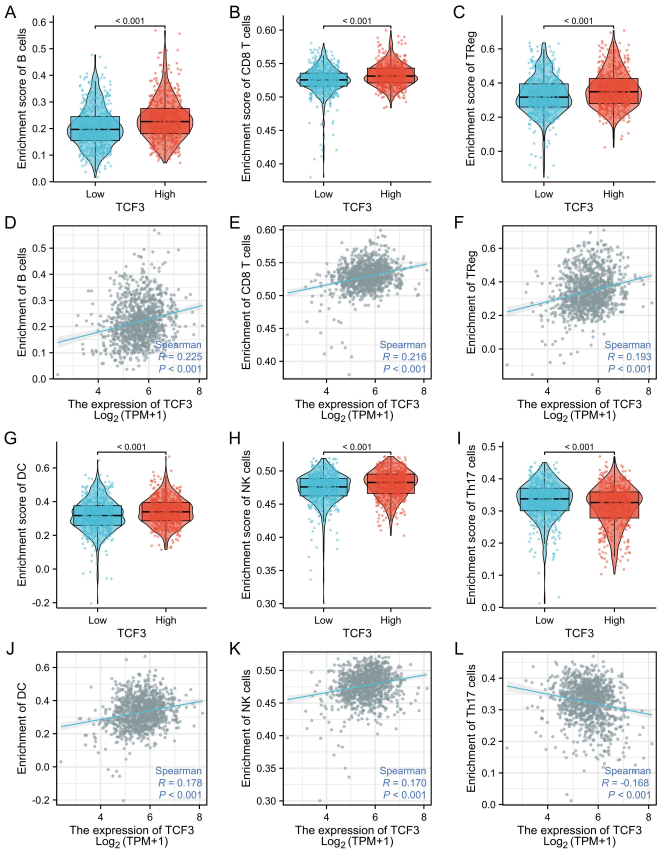

Figure 8: Comparison of immune infiltration levels of immune cells between the high- and low-TCF3 expression groups, including (A) B cells, (B) CD8+ T cells, (C) TReg cells, (G) DC, (H) NK cells, and (I) Th17 cells. Correlations between the relative enrichment scores of immune cells and the expression of TCF3, including (D) B cells, (E) CD8+ T cells, (F) TReg cells, (J) DC, (K) NK cells, (L) Th17 cells. TReg cells, regulatory T cells; DC, dendritic cells; NK cells, natural killer cells.

Analysis of Gene Methylation

To investigate the mechanism underlying TCF3 and its association with breast cancer, we employed the UALCAN database6 to examine the promoter methylation state of TCF3 (as illustrated in Figure 9A) [22]. Moreover, the MethSurv database7 was utilized to evaluate the prognostic significance of the TCF3 methylation level. This web-based tool enables the analysis of survival rates involving multiple variables utilizing of DNA methylation information (Figure 9B–I) [23].

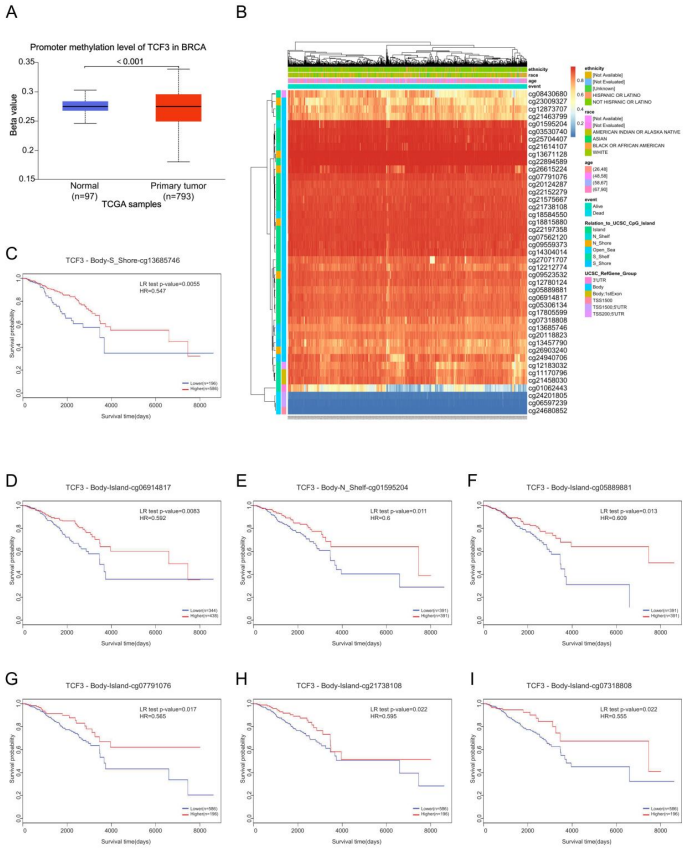

Figure 9: The impact of DNA methylation levels of TCF3 on the prognostic outcomes of individuals diagnosed with breast cancer. Data was collected on 21th February 2023. (A) The methylation level of the TCF3 promoter in breast cancer. (B) Correlation between TCF3 mRNA expression level and methylation level. (C–I) Kaplan-Meier survival curves for several methylation sites of TCF3.

Analysis of prognosis

Survival analysis was conducted using the Kaplan-Meier method along with the Logrank test (Figure 10). The survival analysis results were based on a minimum P-value approach,[24-26] and the surv_cutpoint function in the survminer package was used for the optimal grouping cut-off screening. We utilized the survival [3.3.1] R package to conduct hypothesis testing for proportional hazards and to model survival regression. We utilized the survival [3.3.1] R package to conduct hypothesis testing for proportional hazards and to model survival regression. Partial prognostic data referred to an article from Cell [27].

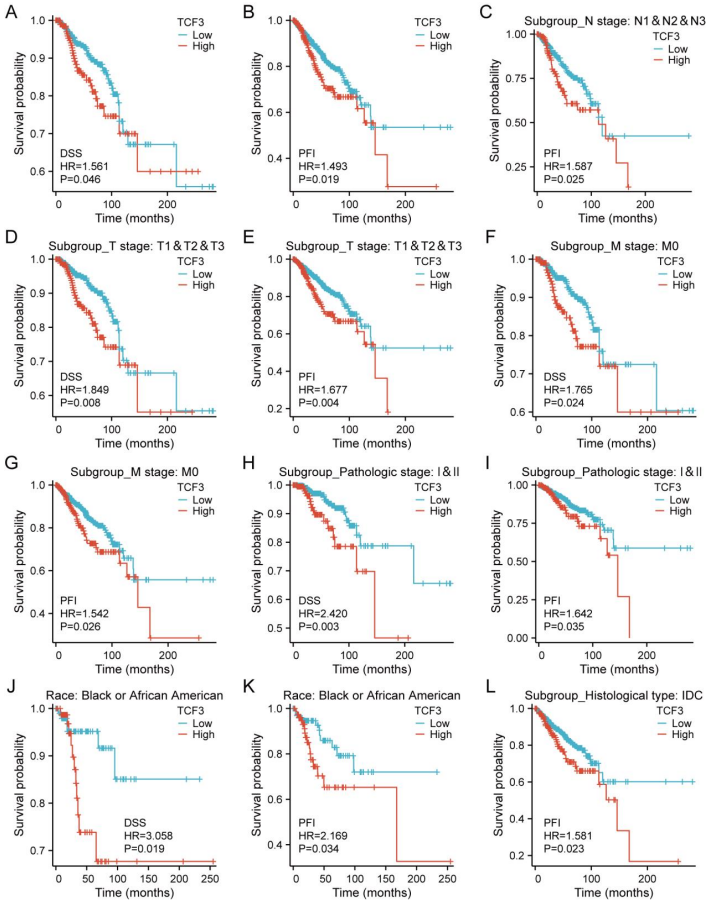

Figure 10: Prognostic values of TCF3 expression in patients with breast cancer evaluated by the Kaplan-Meier method. (A,B) Disease-specific survival (DSS) and progress free interval (PFI) for breast cancer patients with high versus low TCF3. Different subgroups: (C) PFI survival curve of N1, N2 and N3; (D,E) DSS and PFI survival curves of T1, T2 and T3; (F,G) DSS and PFI survival curves of M0; (H,I) DSS and PFI survival curves of stage I and II; (J,K) DSS and PFI survival curves of black or African American; (L) PFI survival curve of IDC. DSS, disease-specific survival; PFI, progress free interval; IDC, infiltrating ductal carcinoma.

Statistical Analysis

R (version 4.2.1) 8 was utilized for statistical computations. The Wilcoxon rank sum test was employed to evaluate the statistical significance of TCF3 expression in non-paired tissues, while the paired sample t-test was utilized for paired tissues. The assessment of correlations between clinical features classified as grouping ≥ 3 and TCF3 expression, utilized the Kruskal-Wallis test and Dunn's test. Meanwhile, the Wilcoxon rank sum test was employed to evaluate clinical features classified as grouping < 3, with a p--value < 0.05 being deemed statistically significant.

The validation of Transcriptomics and Proteomics

RNA sequencing-based analysis of TCF3 differential expression was performed by searching the GSE22820 dataset of the GEO database9 (Figure 11A) [28]. In addition, we analyzed TCF3 gene expression from sequencing data of 49 sample pairs (Figure 11B). Tissue samples were obtained from 49 breast cancer patients in our hospital, including tumor samples and paracancerous samples. Our study were also validated by proteomic data of CPTAC10 samples from the UALCAN, and we obtained TCF3 total-protein expression profile and phosphoprotein expression profile [29]. To obtain a clearer picture, we re-edited the text of the online analysis result pictures.

Results

Patient Characteristics

We obtained a cohort from TCGA, consisting of 1,087 breast cancer patients who had both clinical data and RNA-sequencing information. Among these patients, 113 individuals had matching nearby healthy tissue samples. Furthermore, to expand the pool of healthy tissues, we acquired gene expression information from 179 normal breast tissues in the GTEx database.

Elevated Levels of TCF3 Observed in Breast Cancer

TCF3 expression was found to be high in most cancer types based on the pan-cancer analysis, such as glioblastoma multiforme, brain lower grade glioma, lung adenocarcinoma, stomach adenocarcinoma, colon adenocarcinoma, HCC, pancreatic adenocarcinoma and uterine corpus endometrial carcinoma (Figure 2A), while was low in acute myeloid leukemia (LAML). In breast cancer samples, the level of TCF3 expression was considerably higher than in normal breast tissues, as indicated by Figure 2B ($p < 0.001$). Moreover, TCF3 exhibited high expression in 113 matched breast cancer tissues with a p-value of less than 0.001, as shown in Figure 2C.

In addition, TCF3 expression exhibited a high discriminatory ability [area under the ROC curve (AUC) = 0.807; 95% confidence interval (CI): 0.782-0.831] in distinguishing tumor from normal tissues, as depicted in Figure 2D, indicating its potential as a reliable predictor.

To further identify the expression of TCF3, we also performed analysis at the protein level from the Human Protein Atlas database. On 24th March 2023, information was collected and representative images are presented in Figure 3. Antibody CAB018351 was used for staining both normal breast tissue and breast cancer tissue. In addition, we can see that the TCF3 protein was localized to the nucleoplasm by antibodies CAB018351 and HPA062476.

Associations between TCF3 Expression and Clinicopathologic Variables

A significant association was observed between high expression of TCF3 and various factors, as presented in Table 1 and Figure 4, including T stage (T2 vs T4, $p = 0.024$; T3 vs T4, $p = 0.003$), PAM50 (Luminal A vs Luminal B, $p = 0.003$; Her2 vs Luminal B, $p < 0.001$; Luminal A vs Basal, $p < 0.001$; Luminal B vs Basal, $p < 0.001$), Age ($p < 0.001$), Race (Black vs. Asian, $p = 0.042$; Black vs White, $p < 0.001$), Histological type ($p < 0.001$), ER status ($p < 0.001$), PR status ($p < 0.001$) and Menopause status ($p = 0.041$).

Characteristics	Levels	Low expression of TCF3	High expression of TCF3	P value
n		543	544	
Age, n (%)	≤ 60	268 (24.7%)	335 (30.8%)	**< 0.001**
	> 60	275 (25.3%)	209 (19.2%)	
T stage, n (%)	T1	146 (13.5%)	132 (12.2%)	**0.007**
	T2	314 (29%)	317 (29.2%)	
	T3	57 (5.3%)	83 (7.7%)	
	T4	25 (2.3%)	10 (0.9%)	
N stage, n (%)	N0	255 (23.9%)	261 (24.4%)	0.826
	N1	182 (17%)	177 (16.6%)	
	N2	60 (5.6%)	56 (5.2%)	
	N3	35 (3.3%)	42 (3.9%)	
M stage, n (%)	M0	467 (50.5%)	438 (47.4%)	0.559
	M1	9 (1%)	11 (1.2%)	
Histological type, n (%)	IDC	415 (42.3%)	361 (36.8%)	**< 0.001**
	ILC	68 (6.9%)	137 (14%)	
ER status, n (%)	Negative	79 (7.6%)	161 (15.5%)	**< 0.001**
	Positive	430 (41.4%)	367 (35.3%)	
	Indeterminate	1 (0.1%)	1 (0.1%)	
PR status, n (%)	Negative	142 (13.7%)	200 (19.3%)	**0.002**
	Positive	365 (35.2%)	327 (31.5%)	
	Indeterminate	3 (0.3%)	1 (0.1%)	
HER2 status, n (%)	Negative	280 (39.1%)	280 (39.1%)	0.621
	Positive	75 (10.5%)	82 (11.4%)	
Menopause status, n (%)	Pre & Peri	127 (13%)	143 (14.7%)	0.082
	Post	376 (38.5%)	330 (33.8%)	
PAM50, n (%)	Normal	18 (1.7%)	22 (2%)	**< 0.001**
	Lum A	291 (26.8%)	273 (25.1%)	
	Lum B	139 (12.8%)	67 (6.2%)	
	Her2	31 (2.9%)	51 (4.7%)	
	Basal	64 (5.9%)	131 (12.1%)	
Race, n (%)	Asian	31 (3.1%)	29 (2.9%)	**< 0.001**
	White	387 (38.8%)	368 (36.9%)	
	Black	65 (6.5%)	117 (11.7%)	
DSS event, n (%)	No	498 (46.7%)	484 (45.4%)	0.393
	Yes	39 (3.7%)	46 (4.3%)	

Abbreviations: ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; DSS, disease-specific survival. Race (Black): Black or African American. Bold values denote two-sided P < 0.05.

Table 1: Associations between TCF3 expression and clinicopathologic characteristics

Identification of DEGs in Breast Cancer

1693 genes exhibited differential expression between high and low TCF3 expression groups, with 429 genes upregulated (25.3%) and 1264 genes downregulated (74.7%) (p.adjust <0.05, |LogFC| > 1) (Figure 5A). Next, Figure 5B illustrates the correlation between TCF3 and a set of ten differentially expressed genes, namely DAZAP1, MEX3D, CSNK1G2, LMNB2, REXO1, DOT1L, KHSRP, PTBP1, NUCB2 and SERPINI1.

Functional Enrichment Analysis

GO enrichment analysis, encompassing biological processes, cellular components, and molecular functions, demonstrated that DEGs were significantly enriched in various GO terms, including keratinocyte differentiation, epidermis development, transport vesicle, negative regulation of hydrolase activity, negative regulation of peptidase activity, and negative regulation of proteolysis (Figure 5C). Additionally, upon conducting a KEGG pathway analysis, it was observed that the pathways with a significant enrichment of DEGs were comprised of the metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis- DNA adducts, retinol metabolism, and neuroactive ligand-receptor interaction (Figure 5D). Subsequently, GSEA was applied to DEGs analysis of TCF3 expression groups, and 251 gene sets that were statistically significant (p.adjust < 0.05, FDR < 0.25). Finally, we selected 6 items with higher significance from [Reactome] database for visualization, including collagen formation, extracellular matrix organization, formation of the cornified envelope, degradation of the extracellular matrix, collagen biosynthesis and modifying enzymes, and immunoregulatory interactions between a lymphoid and a non-lymphoid cell (Figure 5E).

Single Cell Analysis

Examining the functional roles of potential molecules at the single-cell level with high precision and resolution is made possible through the essential method of single-cell transcriptome sequencing.[30,31] TCF3 expression in retinoblastoma (RB) had a negative relationship with cellcycle, DNA damage and DNA repair response, while was positively related to angiogenesis and differentiation. Additionally, the findings indicated that TCF3 expression exhibited a positive correlation with cellcycle, DNA damage and stemness in high-grade glioma (HGG) (Figure 6A). In breast cancer (BRCA), the presence of TCF3 was linked to increased invasion and proliferation, while being inversely related to inflammation (Figure 6B). Moreover, the T-SNE diagrams (Figure 6C) depicted the single-cell expression profiles of TCF3 in BRCA.

Correlation between TCF3 Expression and Immune Infiltration

The expression of TCF3 was significantly positively correlated with the levels of immune cell infiltration of B cells ($r = 0.225$, $p < 0.001$), CD8+ T cells ($r = 0.216$, $p < 0.001$), TReg cells ($r = 0.193$, $p < 0.001$), dendritic cells (DCs) ($r = 0.178$, $p < 0.001$), and natural killer (NK) cells ($r = 0.170$, $p < 0.001$), while was negatively correlated with the levels of immune cell infiltration of Th17 cells and Eosinophils (Figure 7). Furthermore, the enrichment scores of B lymphocytes, CD8-positive T lymphocytes, regulatory T cells, dendritic cells, and natural killer cells in the TCF3 high expression group were markedly higher than those in the TCF3 low expression group (all p-values < 0.001) (Figure 8).

Association between DNA Methylation and TCF3 Expression

To elucidate the underlying mechanisms responsible for the excessive expression of TCF3 in breast cancer tissues, we studied the association between TCF3 expression levels and methylation status utilizing online resources. Initially, an analysis of the UALCAN database (Figure 9A), revealed that breast tumor tissues had a significant decrease in DNA promoter methylation, compared to normal tissues ($p < 0.001$). In case of breast cancer, we observed that the majority of methylation sites in the DNA sequences of TCF3 exhibited hypomethylation, and we found a significant correlation between the extent of methylation and patient outcomes (Figure 9B). In the end, a number of methylation sites, including cg13685746, cg06914817, cg01595204, cg05889881, cg07791076, cg21738108, and cg07318808, were found to be correlated with unfavorable prognosis (Figures 9C–I). Based on this analysis, we found that individuals with TCF3 hypomethylation had a notably lower likelihood of survival compared to those with TCF3 hypermethylation.

Prognostic Significance of TCF3 in Breast Cancer

By utilizing the Kaplan-Meier methodology, we computed the correlation between patients' prognosis outcomes and TCF3 expression. The best grouping method was selected to classify patients into cohorts with high and low expression of TCF3. The group with elevated TCF3 levels exhibited a significantly poorer outcome in terms of both DSS and PFI compared to the low expression group (DSS: hazard ratio [HR] = 1.561, 95% CI = 0.966–2.521, $p = 0.046$; PFI: HR = 1.493, 95% CI = 1.037–2.151, $p = 0.019$) (Figures 10A,B). Subsequently, an assessment was carried out to establish the connection between TCF3 expression and prognosis in various subgroups. In several subgroups, patients with high TCF3 expression had a significantly worse prognosis, in terms of both DSS and PFI, such as T1/ T2/ T3, M0, stage I/ II, and Black or African American subgroups; Additionally, for PFI, there was a worse prognosis in N1/ N2/ N3 subgroup and infiltrating ductal carcinoma (IDC) subgroup. (all $p < 0.05$) (Figure 10C-L).

Transcriptomic Validation and Proteomic Analysis Of The TCF3 Expression

Consistent with the results of TCGA data analysis, both the unpaired GSE22820 dataset and our own paired samples sequencing data showed that TCF3 expression was significantly higher in tumor tissues than in normal tissues (Figure 11A,B). In the CPTAC-breast cancer dataset, the total TCF protein expression did not show significant difference based on sample types of tumor and normal tissue (Figure 11C, $p = 9.133964E-02$), but it still showed a consistent trend with the result of RNA-sequencing data.

In addition, we found that the phosphoprotein level of TCF3(NP_001129611.1:T531) was significantly different between tumor tissues and normal tissues (Figure 11D), and was associated with the alterations of multiple tumor-related pathways (Figure 11E-K), especially the P53/Rb-related pathway.

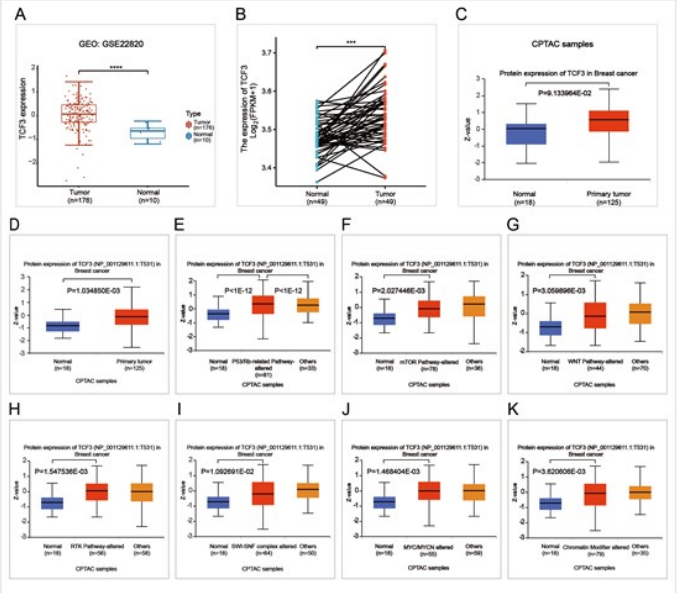

Figure 11: Transcriptomic validation and proteomic analysis of TCF3 expression differences. The expression distribution of TCF3 gene in tumor and normal tissue. ***p < 0.001, ****p < 0.0001, asterisks (*) stand for significance levels. (A) The data were derived from the GSE22820 dataset of the GEO database, and the statistical difference of two groups was compared through the Wilcox test. (B) The data were obtained from sequencing information of 49 paired tumor and paracancerous samples from breast cancer patients, and the statistical difference of two groups was compared through the paired sample t-test. Proteomic analysis and phosphoprotein analysis. The data were derived from the CPTAC samples of UALCAN database. (C) Differences in total TCF3 protein expression. (D) Differences in TCF3(NP_001129611.1:T531) phosphoprotein expression. (E-K) Association of TCF3 protein phosphorylation with several tumor-related pathways.

Discussion

There are limitations to the current biomarkers utilized for breast cancer diagnosis and prognosis prediction, owing to the disease's heterogeneity. Hence, it is crucial to explore innovative indicators for prognostic prediction and personalized therapies. In lymphopoiesis, E2A plays a crucial role, and the encoded protein is indispensable for the development of B and T lymphocytes.[32,33] The absence or reduced activity of this gene may contribute to the development of lymphoid malignancies,[34,35] while the upregulation of TCF3 has been found in various malignancies, including glioblastoma multiforme, lung adenocarcinoma, stomach adenocarcinoma, colon adenocarcinoma, HCC, pancreatic adenocarcinoma and uterine corpus endometrial carcinoma and so on. By utilizing the TCGA database, we investigated the expression of TCF3 in breast cancer and found a significantly higher level of expression compared to that in normal tissue.

This research revealed a higher level of TCF3 expression was closely associated with adverse clinicopathological features, including ER negetive breast cancer, premenopausal & perimenopausal breast cancer, and basal-like subtype. Furthermore, our findings suggested that the upregulation of TCF3 expression may potentially be used as a predictive marker for unfavorable DSS and PFI outcomes in breast cancer patients. Likewise, in a few prior studies, the expression of TCF3 has been shown to indicate poor prognosis and decreased survival rates among patients diagnosed with specific solid tumors, such as HCC, clear cell renal cell carcinoma, pancreas adenocarcinoma, lung squamous cell carcinoma, and prostate adenocarcinoma [11]. These results identify TCF3 as an attractive target for exploitation in innovative therapeutic strategies for individuals suffering from cancer.

Moreover, TCF3 has been shown to promote the development of breast cancer by modulating multiple signaling pathways associated with oncogenesis, such as E2A EMT-TFs (epithelial–mesenchymal transition-transcription factors), E2A–Snail1 axis and acetyl-CoA synthesis pathways (cancer-associated reprogramming of acetyl-CoA metabolism) [11,36]. The depletion of TCF3/E2A has been observed to result in heightened apoptosis upon exposure to doxorubicin treatment, indicating that the upregulation of TCF3/E2A expression could potentially confer protection to cancer cells against the cytotoxic effects of this widely employed chemotherapeutic agent [37]. These discoveries reveal important roles of E2A factors in tumorigenesis, breast cancer cell stemness, metastasis, and drug resistance, supporting evidence for the potential efficacy of targeting E2A proteins as a therapeutic strategy for breast cancer. However, these findings are not comprehensive enough to reveal the fundamental workings of TCF3 in breast cancer, and further research is necessary to better understand its biological function and signaling pathway. Our study employed GO/KEGG analysis and GSEA, which revealed that the group with high TCF3 expression exhibited significantly enriched pathways, such as metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis- DNA adducts, neuroactive ligand-receptor interaction, neuronal differentiation, mesenchymal to epithelial transition, extracellular matrix organization, formation of the cornified envelope, collagen biosynthesis and modifying enzymes, and immunoregulatory interactions between a lymphoid and a non-lymphoid cell. Additional experimental verification is required to validate these discoveries, which have the potential to enhance our comprehension of the biological roles associated with E2A in breast cancer.

The impact of immune cell type, density, and spatial distribution on the prognosis of solid malignancies has been demonstrated.[38,39] Furthermore, infiltrating immune cells have been shown to be a prognostic indicator for the effectiveness of both neoadjuvant chemotherapy and immune checkpoint inhibition therapy.[40,41] Given the enrichment of immunoregulatory interactions between lymphoid and non-lymphoid cells in the high TCF3 expression group, we subsequently computed the correlation between TCF3 expression and immune cell infiltration levels, revealing a positive association between TCF3 overexpression and the infiltration of B lymphocytes, CD8+ T lymphocytes, TReg lymphocytes, DCs, and NK cells. Additionally, we observed a negative correlation with the infiltration of Th17 cells and eosinophils. This may be attributed to the positive role of E proteins in lymphopoiesis, B and T lymphocyte development. Most of these cell types, such as NK cells, DCs and CD8+ T cells, have been linked to a favorable prognosis in breast cancer patients. These findings suggest that the overexpression of TCF3 may not impact breast cancer progression and prognosis through regulation of immune cell infiltrating levels.

Gene promoter methylation is a common epigenetic mechanisms that typically leads to the suppression of gene expression [42]. In the research, we conducted further investigation into the underlying mechanism of TCF3 overexpression. The findings indicate a potential correlation between TCF3 overexpression and DNA hypomethylation. Breast cancer patients with a hypomethylation status of TCF3 have a significantly unfavorable prognosis. Furthermore, breast cancer patients with elevated TCF3 levels had a considerably poorer prognosis compared to those with lower TCF3 levels, and this was also true in several subgroup analyses.

To verify the disparity in TCF3 expression between tumor and normal tissues, we obtained TCF3 gene expression information from GSE22820 dataset of GEO database and sequencing data of 49 pairs of breast cancer and adjacent tissues, and finally obtained supporting validation results. For richer validation, we also performed the proteomic analysis using the CPTAC dataset. Although there was no significant result, the trend was consistent with the previous transcriptome analysis. In the TCF3 phosphoprotein analysis, meaningful conclusions were obtained, and TCF3 phosphoprotein (NP_001129611.1:T531) was found to be associated with alterations in several tumor-related pathways.

Conclusion

Despite the novel insights provided by our current study on the connection between TCF3 expression and breast carcinogenesis, as well as its prognostic value for patients, there are still limitations that require attention. Firstly, as the majority of information for this research was gathered from online databases, we encountered difficulties in obtaining essential clinical information, including chemotherapy plans and whether patients received radiotherapy or endocrine therapy.

Secondly, additional research and thorough experimental verification are required to clarify the biological roles and underlying mechanisms of TCF3 in breast cancer for both in vitro and in vivo systems.

In summary, this research has demonstrated that the upregulation of TCF3 is a significant risk factor and an unfavorable prognostic indicator, closely linked to adverse clinical characteristics in invasive breast cancer. Based on our results, TCF3 has the potential to serve as a new biomarker for forecasting patient outcomes, and may become a new therapeutic target. Nevertheless, further elucidation is necessary to fully understand the mechanism by which TCF3 regulates breast cancer tumorigenesis and progression.

References

1. Angus L, Smid M, Wilting SM, et al. (2019) The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet. 51: 1450-8.

2. Gong C, Man EPS, Tsoi H, et al. (2018) BQ323636.1, a Novel Splice Variant to NCOR2, as a Predictor for Tamoxifen-Resistant Breast Cancer. Clin Cancer Res. 24: 3681-91.

3. Rustighi A, Zannini A, Campaner E, et al.(2017) PIN1 in breast development and cancer: a clinical perspective. Cell Death Differ.24: 200-11.

4. Cho Y, Kang HG, Kim SJ, et al. (2018) Post-translational modification of OCT4 in breast cancer tumorigenesis. Cell Death Differ.25: 1781-95.

5. Tekpli X, Lien T, Røssevold AH, et al. (2019) An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. Nat Commun. 10: 5499.

6. Plichta JK, Ren Y, Thomas SM, et al. (2020) Implications for Breast Cancer Restaging Based on the 8th Edition AJCC Staging Manual. Ann Surg. 271: 169-76.

7. Andrysik Z, Kim J, Tan AC, et al. (2013) A genetic screen identifies TCF3/E2A and TRIAP1 as pathway-specific regulators of the cellular response to p53 activation. Cell Rep. 3: 1346-54.

8. Rao C, Malaguti M, Mason JO, et al. (2020) The transcription factor E2A drives neural differentiation in pluripotent cells. Development. 147: dev184093.

9. Shy BR, Wu CI, Khramtsova GF, et al. (2013) Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/β-catenin signaling. Cell Rep. 4: 1-9.

10. Zhao H, Zhao C, Li H, et al. (2019) E2A attenuates tumor-initiating capacity of colorectal cancer cells via the Wnt/beta-catenin pathway. J Exp Clin Cancer Res. 38: 276.

11. Park S, Mossmann D, Chen Q, et al. (2022) Transcription factors TEAD2 and E2A globally repress acetyl-CoA synthesis to promote tumorigenesis. Mol Cell. 82: 4246-61.e11.

12. Vivian J, Rao AA, Nothaft FA, et al. (2017) Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 35: 314-6.

13. Berger AC, Korkut A, Kanchi RS, et al. (2018) A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell. 33: 690-705.e9.

14. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

15. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26: 139-40.

16. Yu G, Wang LG, Han Y, et al. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 16: 284-7.

17. Walter W, Sánchez-Cabo F, Ricote M (2015) GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics. 31: 2912-4.

18. Subramanian A, Tamayo P, Mootha VK, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102: 15545-50.

19. Yuan H, Yan M, Zhang G, et al. (2019) CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res.47: D900-D908.

20. Hänzelmann S, Castelo R, Guinney J (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 14: 7.

21. Bindea G, Mlecnik B, Tosolini M, et al. (2013) Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 39: 782-95.

22. Chandrashekar DS, Karthikeyan SK, Korla PK, et al. (2022) UALCAN: An update to the integrated cancer data analysis platform. Neoplasia. 25: 18-27.

23. Modhukur V, Iljasenko T, Metsalu T, et al. (2018) MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. Epigenomics. 10: 277-88.

24. Groot VP, Gemenetzis G, Blair AB, et al. (2019) Defining and Predicting Early Recurrence in 957 Patients With Resected Pancreatic Ductal Adenocarcinoma. Ann Surg. 269: 1154-62.

25. Shindoh J, Hasegawa K, Matsuyama Y, et al. (2013) Low hepatitis C viral load predicts better long-term outcomes in patients undergoing resection of hepatocellular carcinoma irrespective of serologic eradication of hepatitis C virus. J Clin Oncol. 31: 766-73.

26. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators; Devereaux PJ, Chan MT, Alonso-Coello P, et al. (2012) Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. JAMA. 307: 2295-304.

27. Liu J, Lichtenberg T, Hoadley KA, et al. (2018) An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 173: 400-16.e11.

28. Zhou T, Cai Z, Ma N, et al. (2020) A Novel Ten-Gene Signature Predicting Prognosis in Hepatocellular Carcinoma. Front Cell Dev Biol. 8: 629.

29. Zhang Y, Chen F, Chandrashekar DS, et al. (2022) Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways. Nat Commun. 13: 2669. Erratum in: Nat Commun. 13: 4688.

30. Li L, Yao W, Yan S, et al. (2021) Pan-Cancer Analysis of Prognostic and Immune Infiltrates for CXCs. Cancers (Basel) 13: 4153.

31. Liu Y, Luo G, Yan Y, et al. (2022) A pan-cancer analysis of copper homeostasis-related gene lipoyltransferase 1: Its potential biological functions and prognosis values. Front Genet 13: 1038174.

32. Geoghegan V, Guo A, Trudgian D, et al. (2015) Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. Nat Commun. 6: 6758.

33. Miyazaki M, Miyazaki K, Chen K, et al. (2017) The E-Id Protein Axis Specifies Adaptive Lymphoid Cell Identity and Suppresses Thymic Innate Lymphoid Cell Development. Immunity. 46: 818-34.e4.

34. Yu S, Zhou X, Steinke FC, et al. (2012) The TCF-1 and LEF-1 transcription factors have cooperative and opposing roles in T cell development and malignancy. Immunity. 37: 813-26.

35. Ortega M, Bhatnagar H, Lin AP, et al. (2015) A microRNA-mediated regulatory loop modulates NOTCH and MYC oncogenic signals in B- and T-cell malignancies. Leukemia. 29: 968-76.

36. López-Menéndez C, Vázquez-Naharro A, Santos V, et al. (2021) E2A Modulates Stemness, Metastasis, and Therapeutic Resistance of Breast Cancer. Cancer Res. 81: 4529-44.

37. Andrysik Z, Kim J, Tan AC, et al. (2013) A genetic screen identifies TCF3/E2A and TRIAP1 as pathway-specific regulators of the cellular response to p53 activation. Cell Rep. 3: 1346-54.

38. Kadota K, Nitadori JI, Ujiie H, et al. (2015) Prognostic Impact of Immune Microenvironment in Lung Squamous Cell Carcinoma: Tumor-Infiltrating CD10+ Neutrophil/CD20+ Lymphocyte Ratio as an Independent Prognostic Factor. J Thorac Oncol. 10: 1301-10.

39. Amicarella F, Muraro MG, Hirt C, et al. (2017) Dual role of tumour-infiltrating T helper 17 cells in human colorectal cancer. Gut. 66: 692-704.

40. Lu J, Liu X, Liao YP, et al. (2018) Breast Cancer Chemo-immunotherapy through Liposomal Delivery of an Immunogenic Cell Death Stimulus Plus Interference in the IDO-1 Pathway. ACS Nano. 12: 11041-61.

41. Inoue Y, Hazama S, Suzuki N, et al. (2017) Cetuximab strongly enhances immune cell infiltration into liver metastatic sites in colorectal cancer. Cancer Sci. 108: 455-60.

42. Yin L, Chung WO (2011) Epigenetic regulation of human β-defensin 2 and CC chemokine ligand 20 expression in gingival epithelial cells in response to oral bacteria. Mucosal Immunol. 4: 409-19.