

## ORIGINAL ARTICLE

# The subtype-specific molecular function of *SPDEF* in breast cancer and insights into prognostic significance

Ting Ye  | Jingyuan Li | Jia Feng | Jinglan Guo | Xue Wan | Dan Xie | Jinbo Liu

Department of Laboratory Medicine, The Affiliated Hospital of Southwest Medical University, Sichuan, China

## Correspondence

Ting Ye, Department of Laboratory Medicine, The Affiliated Hospital of Southwest Medical University, Luzhou, China.

Email: yeting1103@163.com

Jinbo Liu, Department of Laboratory Medicine, The Affiliated Hospital of Southwest Medical University, Luzhou, China.

Email: liulab2019@163.com

## Funding information

Doctoral Research Initiation Fund of Affiliated Hospital of Southwest Medical University, China, Grant/Award Number: 19077; University's scientific research project of Southwest Medical University, Grant/Award Number: 2020ZRQNA019; National Natural Science Youth Fund, China, Grant/Award Number: 82003138; Sichuan Science and Technology Program for key Research and Development, Grant/Award Number: 2021YFS0226

## Abstract

Breast cancer (BC) is a molecular diverse disease which becomes the most common malignancy among women worldwide. There are four BC subtypes (Luminal A, Luminal B, HER2-enriched and Basal-like) robustly established following gene expression pattern-based characterization, behave significant differences in terms of their incidence, risk factors, prognosis and therapeutic sensitivity. Thus, there is an urgent need to provide mechanism research, treatment strategies and/or prognosis evaluation based on the patient stratification of BC subtypes. The prostate-derived ETS factor *SPDEF* was first identified as an activator of prostate specific antigen, and then, the involvements in many aspects of BC have been proposed. However, the subtype-specific molecular function of *SPDEF* in BC and insights into prognostic significance have not been clearly elucidated. This study demonstrated for the first time that *SPDEF* may play a diversity role in the expression levels, clinicopathologic importance, biological function and prognostic evaluation in BC via bioinformatics and experimental evidence, which mainly depends on different BC subtyping. In summary, our findings would help to better understand the possible mechanisms of various BC subtypes and to find possible candidate genes for prognostic and therapeutic usage.

## KEYWORDS

breast cancer, molecular function, prognostic significance, *SPDEF*, subtype-specific

## 1 | INTRODUCTION

Breast cancer is the most common malignancy among women worldwide<sup>1</sup> and also a molecular diverse disease, showing different morphologic and biological characteristics and thus different clinical behaviour and treatment response. As to facilitate oncologic decision-making, the BC classification systems are developed to provide an accurate diagnosis of the disease and prediction of tumour behaviour. Hereinto, four BC subtypes have been robustly established following gene expression patterns based characterization.<sup>2</sup> These subtypes, including Luminal A, Luminal B, HER2-enriched and

Basal-like, behave significant differences in terms of their incidence, risk factors, prognosis and therapeutic sensitivity.<sup>3,4</sup> Therefore, there is an urgent need to provide mechanism research, treatment strategies and prognosis evaluation based on the patient stratification of BC subtypes.

*SPDEF* was first identified as an activator of prostate specific antigen,<sup>5</sup> which is largely restricted to epithelial tissues including the lung, stomach, colon and hormone-regulated epithelia such as the prostate, breast and ovary.<sup>6</sup> In cancer literatures, the role of *SPDEF* in BC is controversial depends on different subtypes, as several studies have demonstrated that high *SPDEF* expression was confirmed to

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd.

promote Luminal BC differentiation and correlates with poor overall survival in ER+breast cancer patients.<sup>6-9</sup> Furthermore, *SPDEF* can also promote proliferation, migration and invasion of SK-BR-3 cells through AR-PDEF pathway<sup>10</sup> or *SPDEF*-CEACAM6 oncogenic axis.<sup>11</sup> The set of above observations exhibits a possible oncogenic function of *SPDEF*. Conversely, the down-regulation of *SPDEF* in invasive basal breast cancer cell lines supports a tumour suppressive role.<sup>12,13</sup> Therefore, the discrepancies between these findings and those on *SPDEF* as an oncogene and/or a tumour suppressor have not been resolved. Further, the potential mechanisms underlying subtype-specific functions of *SPDEF* remain largely unknown.

Bioinformatics analysis has been widely applied in cancer research. In the present study, we uncovered the global expression profiles of *SPDEF*, as well as the clinicopathologic and prognostic importance in different BC subtypes through TCGA-BRCA datasets. Moreover, we verified the protein levels of *SPDEF* with immunohistochemical staining and analysed the relationships between the protein expression of *SPDEF* and clinicopathologic features in BC subtypes. These bioinformatics and clinical findings have added a new dimension to our knowledge about *SPDEF* in addition to its role only as an oncogene or a tumour suppressor in BC. Afterwards, we explored the potential functions and signal pathways of *SPDEF* in BC subtypes using GO, KEGG and hallmark effect gene set analysis, which demonstrated the potential molecular mechanisms of *SPDEF* underlying the oncogenic activity in non-TNBC (Lumina and HER2+) but tumour suppressor activity in TNBC. And lastly, we conducted the prognostic risk model of *SPDEF*-related prognosis genes, respectively, in BC subtypes, indicating a highly prognostic performance in survival surveillance. In this study, we innovatively focussed on the *SPDEF* gene in the aspects of the differential expressions, potential functions and prognostic values in multiple BC subtypes via bioinformatics and experimental evidence. The workflow of the study design is presented in Figure S1.

## 2 | MATERIALS AND METHODS

### 2.1 | *SPDEF* expression analysis in TCGA-BRCA dataset

Differential expression of *SPDEF* in non-tumourous breast tissues and different subtypes of BC tissues were obtained from The Cancer Genome Atlas Project (TCGA). The *SPDEF* mRNA levels in different subtypes of BC were evaluated using edgeR software packages.<sup>14</sup>

### 2.2 | Validation of cell lines with RT-qPCR

Cell lines were purchased from the Cell Bank of the Chinese Academy of Sciences and cultured in special medium. RNA was extracted by TRIZOL (Takara) and transcribed into cDNA using PrimeScript RT reagent Kit (Takara). The quantitative real-time PCR (qPCR) was used to detect the mRNA expression of *SPDEF* in different subtypes of BC cells. The PCR primers were

sequenced as follows: 5'- GAGCCACCTGAGGAGCCTGAG -3' (forward) and 5'- CTTGAGCACTTCGCCACCAC -3' (reverse) for *SPDEF*; 5'- CCGGAATCCCTATCTTTAGTCC -3' (forward) and 5'- GCCTTTGTTGCTCTCCAAAAT-3' (reverse) for *TBP*.

### 2.3 | Immunohistochemical staining

The paraffin-embedded tissues were obtained from the Pathology Department of the Affiliated Hospital of Southwest Medical University. And the tissue slides were deparaffinized, rehydrated and stained with the rabbit polyclonal anti-*SPDEF* antibody (AB clonal, 1:300) overnight at 4°C. Next, the slides were treated with biotinylated secondary antibody followed by incubation with streptavidin-HRP. Finally, there were stained using DAB and counterstained with haematoxylin. *SPDEF* staining was scored based on the multiplier of the positive percentage and staining intensity of the stained area as a result of the total score ranged from 0 to 6. The percentage of *SPDEF*-positive stained cells was scored as 0 (0%-25%), 1 (25%-50%) and 2 (>50%). In addition, the intensity of *SPDEF* expression was scored as 0 no staining (-), 1 weak staining (+), 2 moderate staining (++) and 3 strong staining (+++). A total score of  $\geq 4$  indicated positive *SPDEF* expression.

### 2.4 | The clinicopathologic and prognostic analysis of *SPDEF* in BC patients

The association between the *SPDEF* expression and overall survival was performed by Kaplan-Meier method.<sup>15</sup> To combine with clinical data of patients, the clinical significance of *SPDEF* expression was figured out. And the best performing threshold is used as a cut-off.

### 2.5 | GO function and KEGG pathway enrichment analysis

Aberrantly expressed genes were filtered using transcription profiles from TCGA-BRCA database. The correlation coefficients were calculated based on Pearson in order to find the *SPDEF*-related genes among differentially expressed genes ( $r > 0.4$ ,  $P < .05$ ). And then, the bioinformatic analysis of the *SPDEF*-related genes involved GO Enrichment analysis<sup>16</sup> and KEGG signal transduction pathway enrichment<sup>17</sup> were performed by R software and Bioconductor packages.<sup>18</sup>

### 2.6 | Gene set enrichment analysis

The different subtypes of BC patients were divided into high- and low-expression groups based on the median expression level of *SPDEF* from TCGA-BRCA database. Hallmark effector gene set of high *SPDEF* expression was annotated by gene set enrichment analysis (GSEA).<sup>19,20</sup> Hallmark effector gene sets were obtained from the

Molecular Signature Database (MsigDB).<sup>21</sup> The  $P$ -value  $<0.05$  and false discovery rate (FDR)  $<0.25$  were used as cut-off criterion.

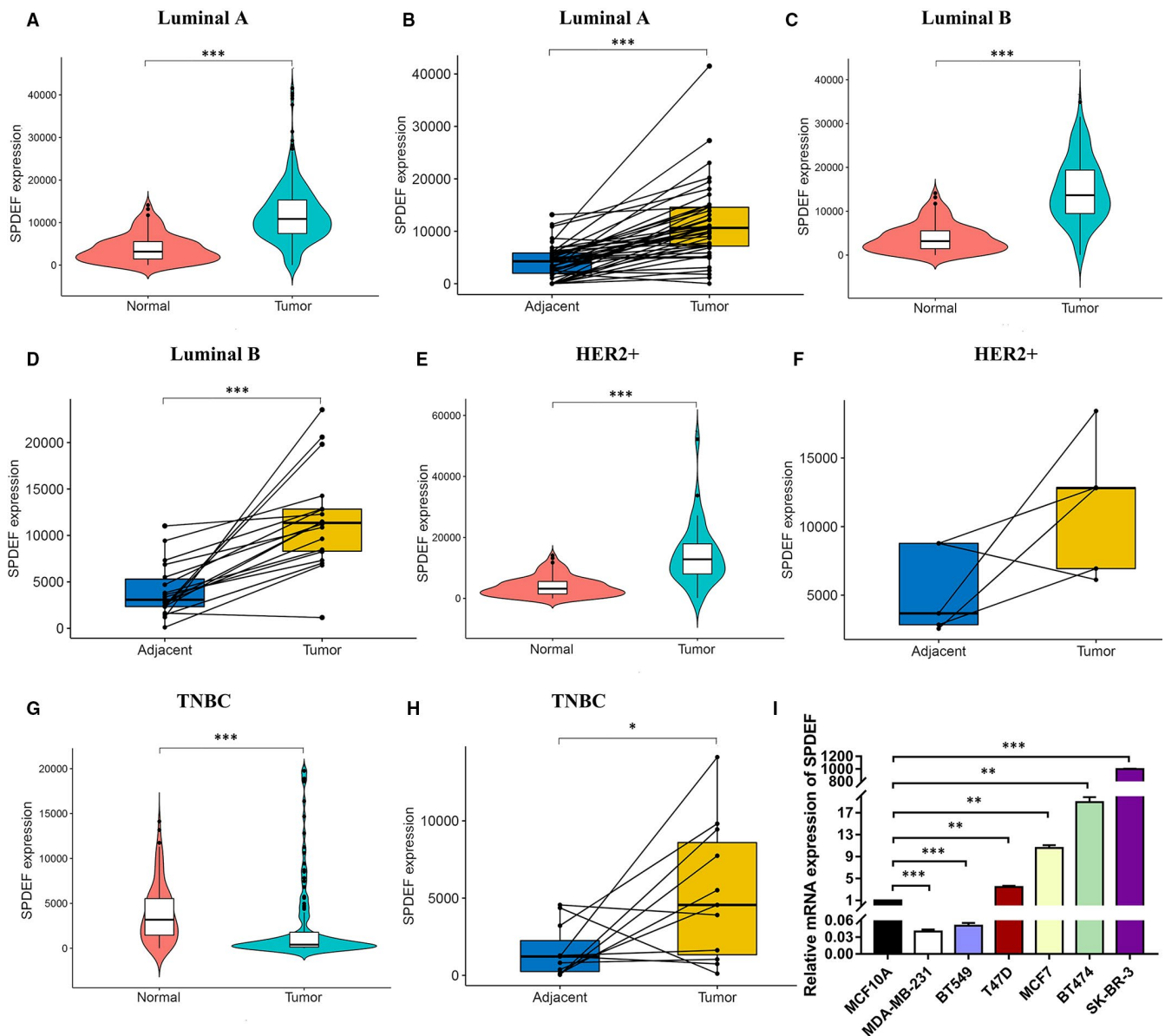
## 2.7 | Construction of prognostic risk model of BC patients based on *SPDEF*-related genes

Firstly, univariate Cox regression analysis was performed to identify significant prognostic genes in *SPDEF*-related genes from TCGA database ( $P < .05$ ). Then, the least absolute shrinkage and selection operator (LASSO) Cox<sup>22</sup> model was used to identify most critical *SPDEF*-related prognostic genes. Moreover, risk score model and predictive signature model of prognosis were built by the multivariate Cox regression.

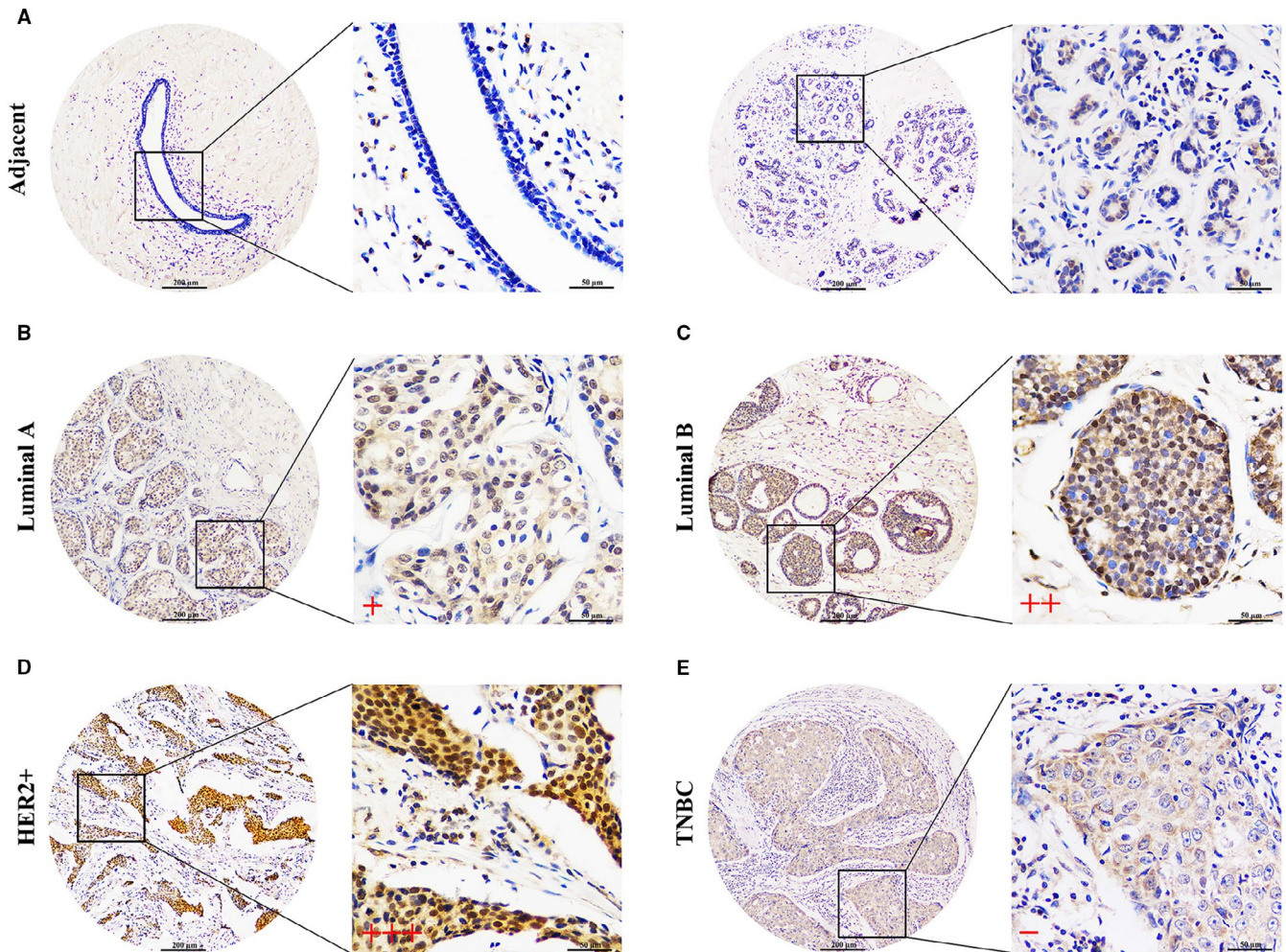
According to the median value of the risk score, all patients from TCGA database were divided into the high-risk group and low-risk group to perform the evaluation of Kaplan-Meier (K-M) survival curves.

## 2.8 | Statistical analysis

The expression levels of gene expression levels in between breast cancer and normal breast tissues were statistically compared by Student's  $t$  test and Wilcoxon signed rank sum test. Data were analysed by GraphPad Prism 7.0 software and R-4.0.2 software, which presented as mean  $\pm$  SEM. Differences were considered statistically significant when  $P < .05$ .



**FIGURE 1** The global *SPDEF* expression profiles in different BC subtypes.(A-H). *SPDEF* expression level in different subtypes of BC samples compared to normal pericarcinomatous samples. The mRNA levels of *SPDEF* in unmatched BC and matched BC were downloaded from TCGA datasets. (A-B) Luminal A, (C-D) Luminal B, (E-F) HER2+, (G-H) TNBC. I. mRNA expression of *SPDEF* in different subtypes of BC cells. \*:  $P < .05$ ; \*\*:  $P < .01$ ; \*\*\*:  $P < .001$



**FIGURE 2** Clinical validation of *SPDEF* expression in multiple BC subtypes. Representative immunohistochemical staining of *SPDEF* in multiple BC subtypes. (A) Adjacent tissue, (B) Luminal A tissue, (C) Luminal B tissue, (D) HER2+ tissue, (E) TNBC tissue

### 3 | RESULTS

#### 3.1 | The differential expressions of *SPDEF* in multiple subtypes of BC

We first analysed the mRNA expression of *SPDEF* between BC subtypes and normal (adjacent) breast tissues using TCGA database. *SPDEF* was remarkably overexpressed due to increased mRNA in non-TNBC (Luminal A, Luminal B and HER2+) compared to normal individuals or adjacent tissues (Figure 1A-F). Nevertheless, the mRNA expression levels of *SPDEF* in TNBC tissues were dramatically decreased compared with that in the normal or adjacent breast tissues (Figure 1G-H). To further validate the results of TCGA database, we conducted *SPDEF* expression analyses using the GEPIA2, TIMER website databases and GEO datasets for the expression of *SPDEF* in BC subtypes. Consistently, the *SPDEF* expression is significant up-regulated in non-TNBC compared with that in normal tissues, but opposite in TNBC (Figure S2A-C). Meanwhile, the expression of *SPDEF* in different subtype BC cells (MCF7, T47D, BT474, SK-BR-3,

MDA-MB-231, BT549) and its corresponding group (MCF 10A) was detected by RT-qPCR. The results showed that non-TNBC cells (MCF7, T47D, BT474, SK-BR-3) had elevated *SPDEF* mRNA, whereas TNBC (MDA-MB-231, BT549) cells had decreased expression in comparison with the non-malignant MCF 10A cells (Figure 1I).

Further, to verify the findings of the bioinformatic analysis, we detected the protein expression of *SPDEF* with immunohistochemical staining. The paraffin-embedded tissues were collected for *SPDEF* protein analysis, including different subtypes BC cases (Luminal A, Luminal B, HER2+ and TNBC) and their matched adjacent normal tissues. What can be clearly seen in immunohistochemical figures is that the *SPDEF* expression was significantly up-regulated in Luminal A, Luminal B and HER2+ BC tissues compared with corresponding adjacent normal tissues (Figure 2A-D). And the positive staining of *SPDEF* was mostly distributed in the nucleus. Rather, there was no significant change in *SPDEF* protein expression in TNBC tissues compared with normal tissues (Figure 2E). Taken together, our data support that *SPDEF* is up-regulated in the non-TNBC, but suppressed in TNBC. Besides, the relationships between the protein expression

**TABLE 1** Relationships between the protein expression of SPDEF and clinicopathological parameters in four molecular subtypes of BC based on IHC detection

Variables	SPDEF expression in Luminal A			SPDEF expression in Luminal B			SPDEF expression in HER2+			SPDEF expression in TNBC		
	Total (n = 42)	Negative (%)	Positive (%)	Total (n = 86)	Negative (%)	Positive (%)	Total (n = 39)	Negative (%)	Positive (%)	Total (n = 58)	Negative (%)	Positive (%)
Age at Surgery												
≤49	18	9 (50.0)	9 (50.0)	40	17 (42.5)	23 (57.5)	14	9 (64.3)	5 (35.7)	21	18 (85.7)	3 (14.3)
>49	24	14 (58.3)	10 (41.7)	46	19 (41.3)	27 (58.7)	25	12 (48.0)	13 (52.0)	37	31 (83.8)	6 (16.2)
cTNM stage												
I + II	23	14 (60.9)	9 (39.1)	49	16 (32.7)	33 (67.3)	19	14 (73.7)	5 (26.3)	41	34 (82.9)	7 (17.1)
III + IV	19	9 (47.4)	10 (52.6)	37	20 (54.1)	17 (45.9)	20	7 (35.0)	13 (65.0)	17	15 (88.2)	2 (11.8)
Lymphatic metastasis												
No	16	12 (75.0)	4 (25.0)	34	9 (26.5)	25 (73.5)	14	11 (78.6)	3 (21.4)	31	26 (83.9)	5 (16.1)
Positive	26	11 (42.3)	15 (57.7)	52	27 (51.9)	25 (48.1)	25	10 (40.0)	15 (60.0)	27	23 (85.2)	4 (14.8)
Distant Metastasis												
M0	37	20 (54.1)	17 (45.9)	82	34 (41.5)	48 (58.5)	37	20 (81.1)	17 (18.9)	56	47 (83.9)	9 (16.1)
M1	5	3 (60.0)	2 (40.0)	4	2 (50.0)	2 (50.0)	2	1 (50.0)	1 (50.0)	2	2 (100.0)	0 (0.0)
Ki-67												
<14%	42	—	—	32	15 (46.9)	17 (53.1)	6	4 (66.7)	2 (33.3)	5	5 (100.0)	0 (0.0)
≥14%	0	—	—	54	21 (38.9)	33 (61.1)	33	17 (51.5)	16 (48.5)	53	44 (83.0)	9 (17.0)

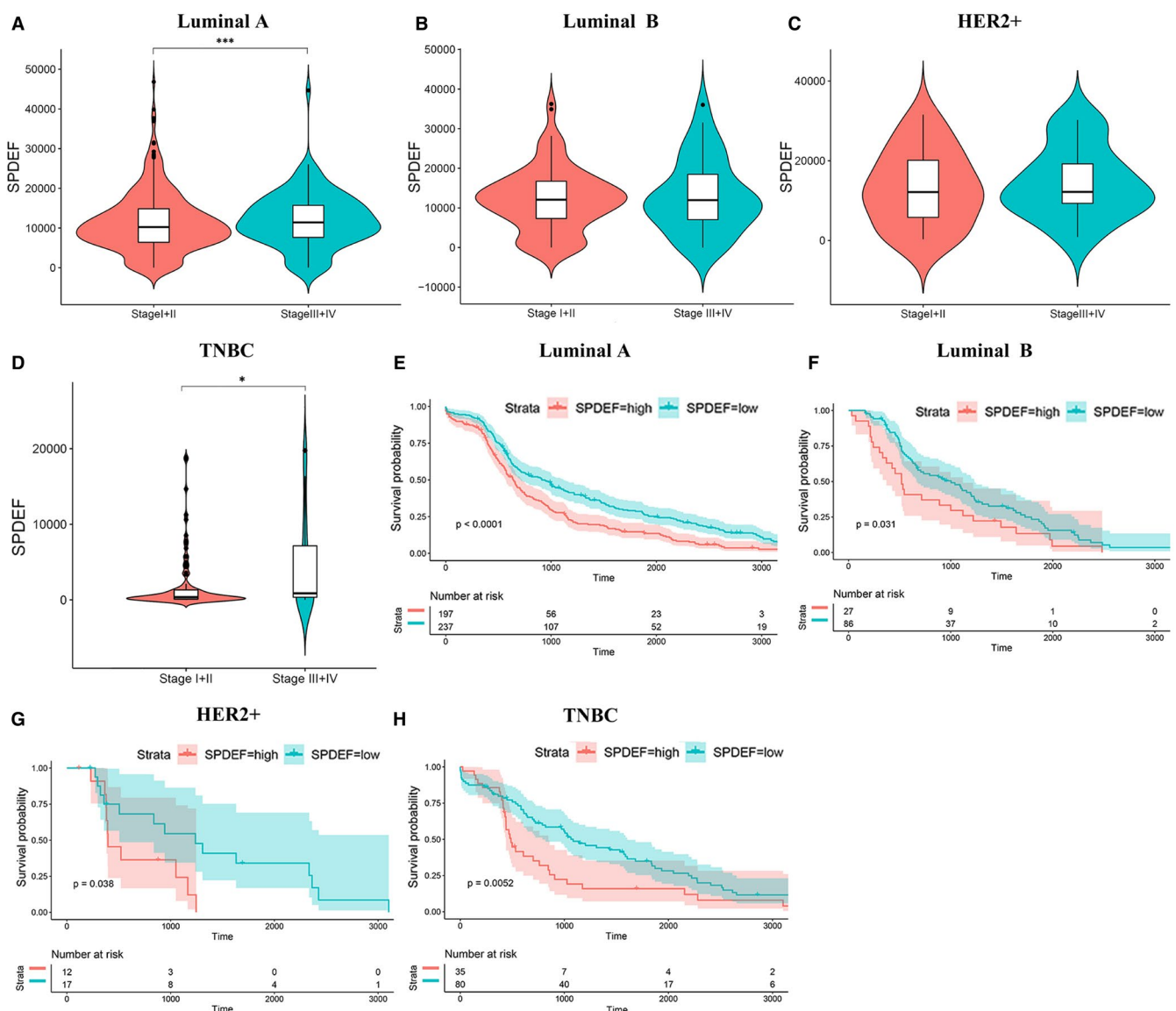
\*Bold values indicate  $P < .05$ .

of *SPDEF* and clinicopathologic features in BC subtypes are summarized in Table 1. Over-expressed protein of *SPDEF* was significantly associated with lymphatic metastasis ( $P = .039$ ) in Luminal A. As for the Luminal B and HER2+, high *SPDEF* expression was positively associated with TNM stage ( $P = .046$  in Luminal B,  $P = .023$  in HER2+) and lymphoid nodal status ( $P = .019$  in Luminal B,  $P = .043$  in HER2+). However, no significant difference was found in TNBC.

### 3.2 | The clinicopathologic and prognostic importance of *SPDEF* in different BC subtypes

In addition, we compared the transcription levels of *SPDEF* among groups of different subtype BC patients, according to different

clinicopathological characteristics (Figure 3A-D) (Table 2). It is demonstrated that no significant difference was found in age and distant metastasis status. Notably, in Luminal A, high *SPDEF* expression was positively associated with TNM stage ( $P = .004$ ), lymphoid nodal status ( $P = .023$ ), whereas in Luminal B, high *SPDEF* expression was positively associated with tumour invasion ( $P = .025$ ). As for HER2+, the overexpression of *SPDEF* was positive correlation with lymphoid nodal status ( $P = .032$ ). And it also showed a positive association between *SPDEF* increased mRNA and TNM stage ( $P = .032$ ) in TNBC. Afterwards, we analysed the prognostic value of *SPDEF* expression by examining the relationship between *SPDEF* expression and progression of BC subtyping using TCGA database by Kaplan-Meier method. Interestingly, high *SPDEF* mRNA levels are correlated with faster disease progression

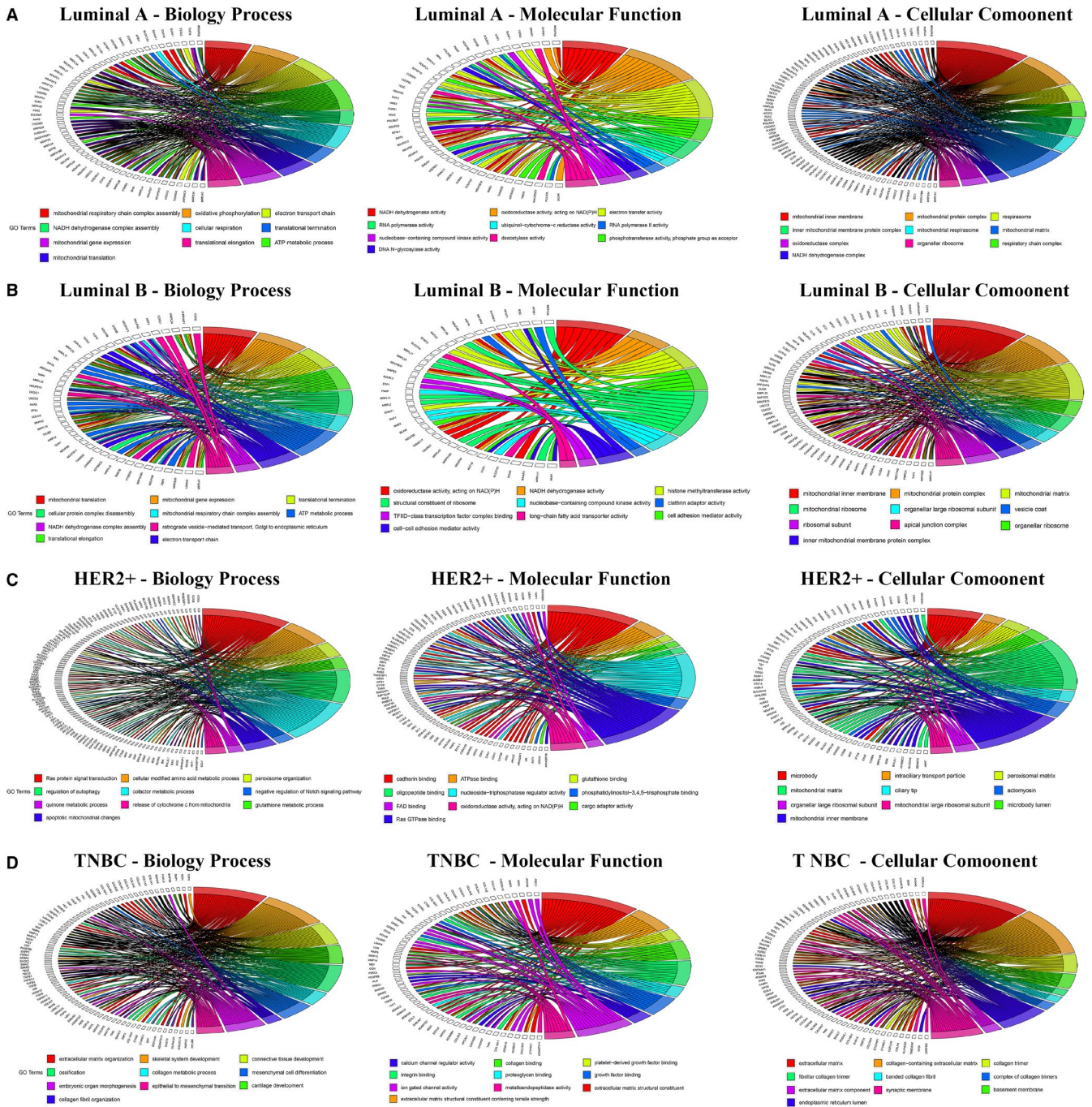


**FIGURE 3** The significance of *SPDEF* in evaluating the clinicopathologic characteristics and prognosis of BC patients across different subtypes.(A-D). Differential expressions of *SPDEF* in early and late tumour stage according to BC subtypes. (A) Luminal A, (B) Luminal B, (C) HER2+, (D) TNBC. E-H. Kaplan-Meier estimates of the overall survival of patients with different BC subtypes according to *SPDEF* levels. (E) Luminal A, (F) Luminal B, (G) HER2+, (H) TNBC. \*:  $P < .05$ , \*\*\*:  $P < .001$

**TABLE 2** The association of SPDEF mRNA expression and clinicopathological parameters in BC subtypes based on TCGA database

Variables	SPDEF mRNA expression of luminal A			SPDEF mRNA expression of luminal B			SPDEF mRNA expression of HER2+			SPDEF mRNA expression of TNBC			
	Total(n = 442)	Low (n = 221)	High (n = 221)	Total(n = 126)	Low (n = 63)	High (n = 63)	Total(n = 37)	Low (n = 19)	High (n = 18)	Total(n = 115)	Low (n = 58)	High (n = 57)	P-value
Age at surgery													
<51	121	60	61	36	16	20	10	5	5	45	25	20	0.379
≥51	321	161	160	90	47	43	27	14	13	70	33	37	
cTNM Stage													
I + II	327	177	150	90	44	46	25	15	10	92	51	41	0.032*
III + IV	110	42	68	35	19	16	12	4	8	23	7	16	
X	5	2	3	1		1							
Tumour invasion													
T1 + T2	365	188	177	107	58	49	32	16	16	99	49	50	0.616
T3 + T4	76	33	43	19	5	14	5	3	2	16	9	7	
X	1		1										
Lymphoid nodal status													
-	193	109	84	52	24	28	10	8	2	74	41	33	0.152
+	244	111	133	73	39	34	25	10	15	41	17	24	
X	5	1	4	1		1	2	1	1				
Distant metastasis status													
M0	377	195	182	101	49	52	36	19	17	98	52	46	0.137
M1	6	4	2	4	3	1	1	0	1	2	0	2	
MX	59	22	37	21	11	10	10			15	6	9	

\*Bold values indicate P < .05.



**FIGURE 4** GO annotation enrichment analysis of *SPDEF*-related genes in various BC subtyping. Main biological processes, molecular functions and cell components related to *SPDEF* were identified by *SPDEF*-related genes. (A) Luminal A, (B) Luminal B, (C) HER2+, (D) TNBC

and lower rate of overall survival (OS) in all subtypes BC (Figure 3E-H). Furthermore, BC patients with a low *SPDEF* expression exhibited a better distance metastasis-free survival (DMFS) compared with patients with a high *SPDEF* expression by the Kaplan-Meier Plotter website analysis (Figure S1D-G). Thus, high *SPDEF* expression predicts poor prognosis.

### 3.3 | The Gene Ontology functions enrichment analysis of *SPDEF*-related genes in various BC subtyping

To better understand the gene-enrichment and functional annotation analyses of *SPDEF*, we implemented GO enrichment to



discovery the functions in which the *SPDEF* participated in BC subtyping, with a threshold of  $P < .05$ . The overview schematic of analysis results is displayed in Figure 4 and Table S1. The functions of the gene *SPDEF* were enriched analysis according to the GO terms of the biological process (BP), cellular component (CC) and molecular function (MF). As the top 10 of GO enrichment illustrated in Luminal A, the GO term of 'mitochondrial respiratory chain complex assembly' (GO: 0033108) was the most significant enrichment for BP category ( $P < .001$ ). In the MF category, 'NADH dehydrogenase activity' (GO: 0003954) was the highest enrichment term ( $P < .001$ ). And the GO term of 'mitochondrial inner membrane' (GO: 0005743) was the most important of the CC category ( $P < .001$ ). In Luminal B, the GO term of 'mitochondrial translational' (GO: 0032543), 'oxidoreductase activity, acting on NAD(P)H' (GO: 0016651), 'mitochondrial inner membrane' (GO: 0005743) were the most prominent enrichment for BP ( $P < .001$ ), MF ( $P < .001$ ) and CC ( $P < .001$ ) category, respectively. And for HER2+ BC, the highest enrichment term was the 'Ras protein signal transduction' (GO: 0007265) in BP ( $P < .001$ ), 'cadherin binding' (GO: 0045296) in MF ( $P < .001$ ), 'microbody' (GO: 0042579) in CC ( $P < .001$ ).

Unlike the enrichment functions of non-triple negative BC, as the top 10 of GO enrichment illustrated in TNBC, the GO term of 'extracellular matrix organization' (GO: 0030198) was the highest enrichment term for BP category ( $P < .001$ ). For MF category, 'extracellular matrix structural constituent' (GO: 0005201) was the most significant enrichment. And the GO term of 'extracellular matrix' (GO: 0031012) was the most valuable of CC category ( $P < .001$ ).

### 3.4 | Enrichment analysis identifies the *SPDEF*-related signalling pathway in multiple BC subtypes

The deeper molecular functions of *SPDEF* were obtained via KEGG signalling pathway gene sets and evaluating hallmark effect gene sets. All the most valuable enriched pathway of each category were presented, respectively ( $P < .05$ ). Hereinto, the top five KEGG pathway enrichment analysis was shown to be significantly associated with thermogenesis, oxidative phosphorylation, retrograde endocannabinoid signalling, peroxisome and mTOR signalling pathway in Luminal A; thermogenesis, retrograde endocannabinoid signalling, oxidative phosphorylation, glucagon signalling pathway and insulin resistance in Luminal B; MAPK signalling pathway, Ras signalling pathway, endocrine resistance, prostate cancer and pancreatic cancer in HER2+ (Figure 5A-C). Notably, KEGG results in TNBC indicated enrichment mainly for PI3K-Akt signalling pathway, neuroactive ligand-receptor interaction, human papillomavirus infection, focal adhesion and calcium signalling pathway (Figure 5D).

Besides, the predefined hallmark effect gene sets of different BC subtyping were differentially enriched with the high *SPDEF* expression phenotype (Figure 5E-H). In Luminal A, *SPDEF*-related signalling pathways included DNA repair, oestrogen response

early/late, fatty acid metabolism, MYC targets V2 and oxidative phosphorylation, whereas in Luminal B, *SPDEF*-related signalling pathways included epithelial-mesenchymal transition (EMT), TGF- $\beta$  signalling, TNFA signalling via NF $\kappa$ B and UV response DN. For HER2+, the pathways enriched in adipogenesis, DNA repair, oestrogen response late, fatty acid metabolism, MYC targets V2, oxidative phosphorylation and peroxisome are similar to those of Luminal A. And TNBC-related signalling pathways include bile acid metabolism and oestrogen response early. This suggests that *SPDEF* may contribute to different biological functions in the development of various BC subtypes.

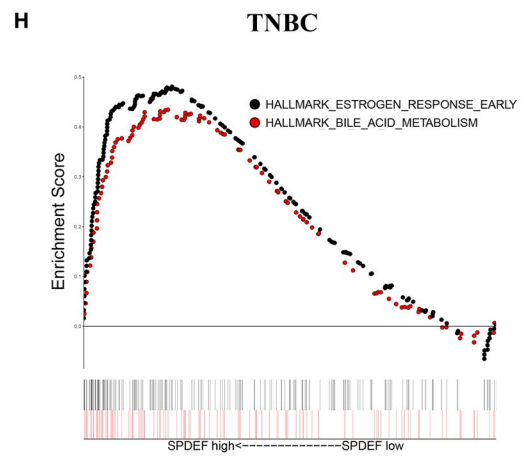
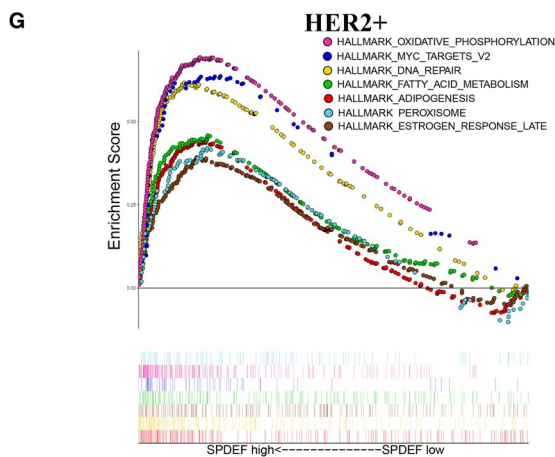
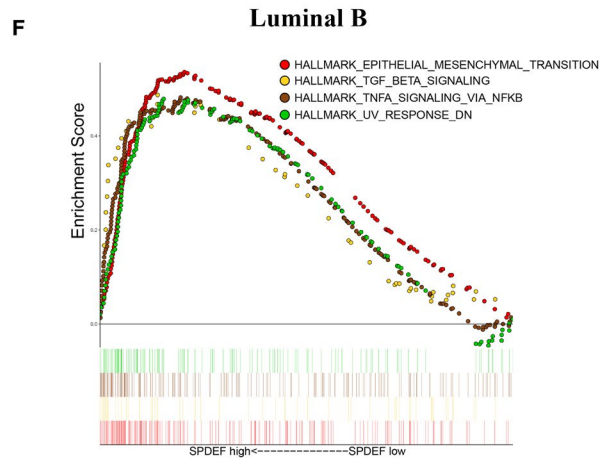
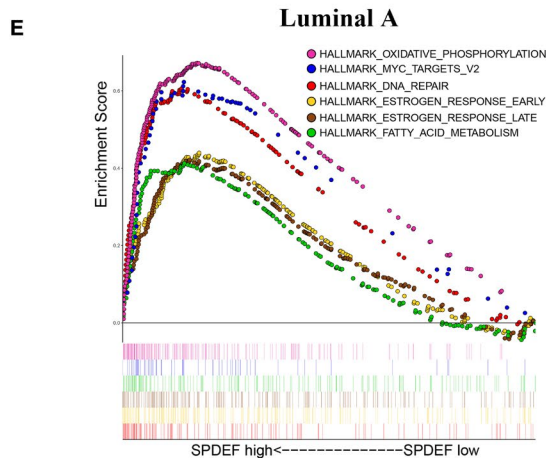
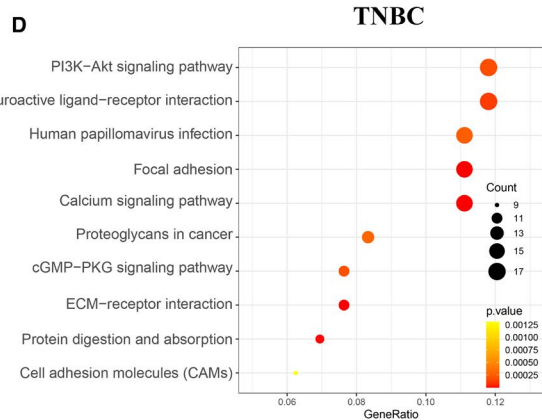
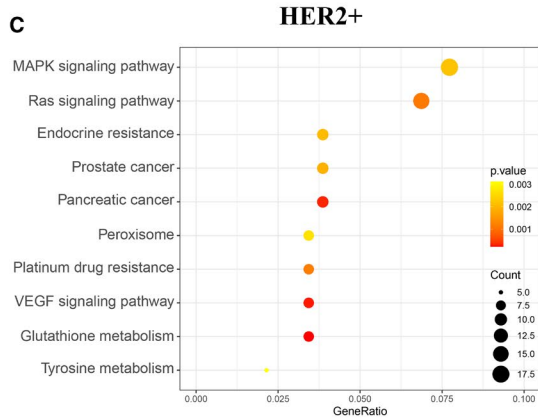
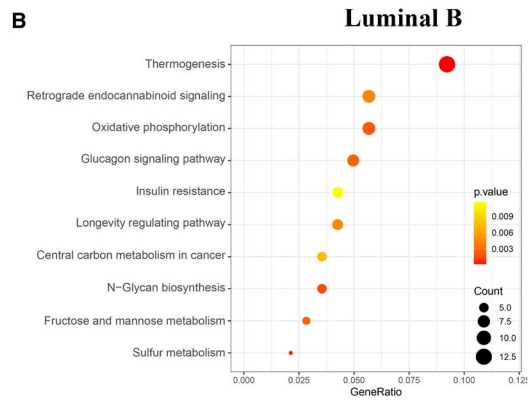
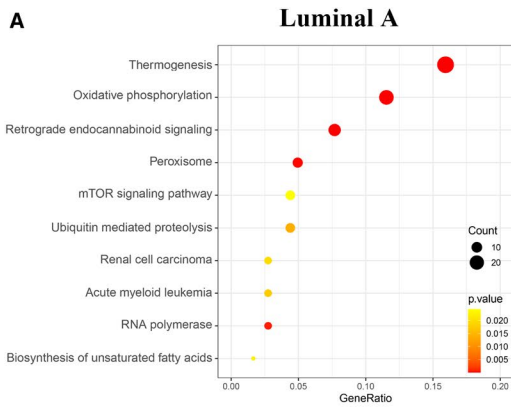
### 3.5 | Construction of the prognostic risk model of *SPDEF*-related prognosis genes in subtypes of BC

To further investigate the clinical prognostic effect of *SPDEF* in multiple BC subtyping, we firstly performed to identify prognostic genes of *SPDEF*-correlated from TCGA database by univariate Cox regression analysis. And then, we obtained 11 genes (CCDC9, UBXN1, VPS37D, SCAND1, PGLS, ZNF593, NDUFA11, RASSF7, PMF1, APEH, PRR15L) in Luminal A, 6 genes (KRT18P10, KRT18, KRT8, DCXR, CLTA, HRAS) in Luminal B, 7 genes (AP1M2, STARD3, TCAP, SNX14, CAPZB, PPIL2, KCTD15) in HER2+ and 4 genes (TFAP2B, ARFIP2, DALRD3, TRIM3) in TNBC, respectively (Figure 6A-D).

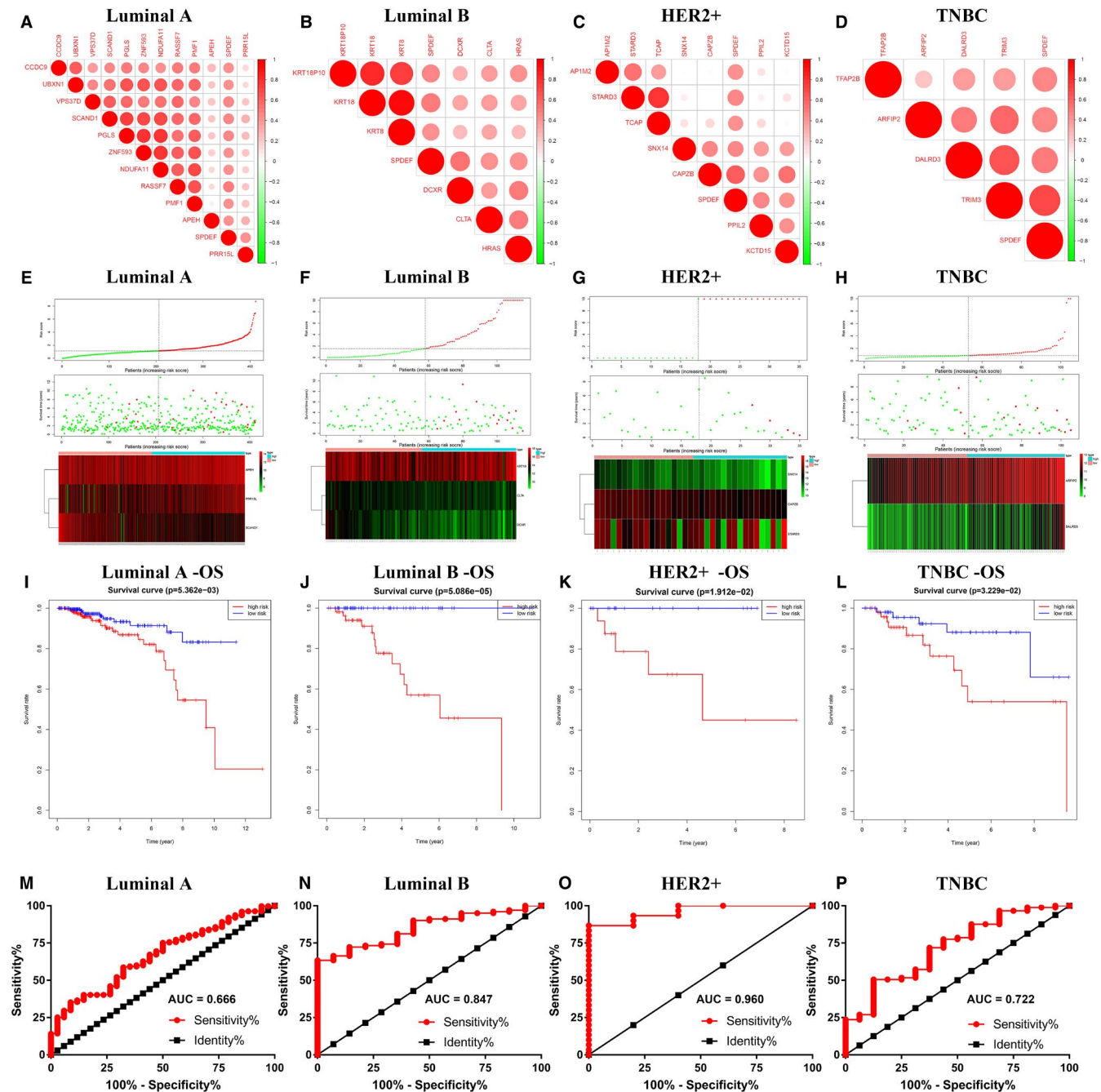
Based on the results of *SPDEF*-related prognostic genes analysis, we developed a prognostic index (PI) to stratify different subtypes of BC patients into two groups (high and low risk) and constructed a predictive model to identify the performance of PI in predicting the clinical outcome of BC subtype patients. The formula of PI is as follows:  $(0.00014 * \text{expression value of APEH}) + (0.00019 * \text{expression value of PRR15L}) + (-0.00050 * \text{expression value of SCAND1})$  in Luminal A,  $(0.00004 * \text{expression value of KRT18}) + (-0.00044 * \text{expression value of CLTA}) + (-0.00051 * \text{expression value of DCXR})$  in Luminal B,  $(0.00150 * \text{expression value of STARD3}) + (-0.01141 * \text{expression value of CAPZB}) + (-0.03763 * \text{expression value of SNX14})$  in HER2+,  $(0.00025 * \text{expression value of ARFIP2}) + (0.00071 * \text{expression value of DALRD3})$  in TNBC. The distribution of risk scores, survival status of each subject and heatmap of gene expression pattern are shown in Figure 6E-H. And the higher risk score showed a shorter survival time for patients and vice versa (Figure 6I-L). The area under the curve (AUC) of the receiver operating characteristic (ROC) was 0.666 in Luminal A, 0.847 in Luminal B, 0.960 in HER2+, 0.722 in TNBC (Figure 6M-P), indicating a high prognostic performance of the *SPDEF*-related prognostic genes in survival surveillance.

## 4 | DISCUSSION

Breast cancer is a clinically and biologically heterogeneous disease; thus, research based on BC subtypes is critical to achieve better clinical outcomes.<sup>23</sup> In cancer literature, the role of *SPDEF*, known as the prostate-derived ETS factor, that functions in BC is widely reported.



**FIGURE 5** KEGG pathways and Hallmark gene set enrichment analysis associated with *SPDEF* in subtypes of BC. (A-D). The enriched pathways among the *SPDEF* high expression group were identified in different BC subtypes by KEGG analysis. (A) Luminal A, (B) Luminal B, (C) HER2+, (D) TNBC. E-H. GSEA profiles depicted significant hallmark gene sets associated with *SPDEF* expression in different BC subtypes. (E) Luminal A, (F) Luminal B, (G) HER2+, (H) TNBC



**FIGURE 6** Prognostic risk score model analysis of *SPDEF*-related prognostic genes with different BC subtypes(A-D). Correlation analysis between *SPDEF* and co-expressed prognostic genes in subclasses of BC. (A) Luminal A, (B) Luminal B, (C) HER2+, (D) TNBC. E-H. The risk score distribution, survival status of patients and heatmap of gene expression pattern in different BC subtypes. (E) Luminal A, (F) Luminal B, (G) HER2+, (H) TNBC. I-L. The Kaplan-Meier plot for OS of patients in the different risk groups of BC subtypes. (I) Luminal A, (J) Luminal B, (K) HER2+, (L) TNBC. M-P. The ROC curves for the prognostic value of the risk score in different BC subtypes. (M) Luminal A, (N) Luminal B, (O) HER2+, (P) TNBC

Prior to the present study, we have summarized *SPDEF* as the double agent involving in expression profiles, the regulator mechanism in BC progression, as well as the role in diagnosis, treatment and

prognosis of BC with literature review.<sup>24</sup> However, the specific roles of *SPDEF* in various subtypes of BC have not been systematically evaluated and established. This study demonstrated for the first

time that *SPDEF* may play a diversity role in the expression levels, clinicopathologic importance, biological function and prognostic evaluation in BC via bioinformatics and experimental evidence, which mainly depends on different BC subtyping.

We made the following novel findings that had not been previously reported:

First, the oncogene function of *SPDEF* overexpression in non-TNBC (Luminal A, Luminal B, HER2+) and the tumour suppressor function of *SPDEF* down-regulation in TNBC have been uncovered by bioinformatics analysis (Figure 1A-H). Subsequently, the overabundance of *SPDEF* in non-TNBC (Luminal A, Luminal B, HER2+) relative to TNBC has been verified by the transcription level detection in variety BC cells (Figure 1I) and the protein analysis in paraffin-embedded tissues of BC subtypes (Figure 2A-E). Moreover, high-protein level of *SPDEF* was positively associated with lymphatic metastasis in Luminal A, with TNM stage and lymphoid nodal status in Luminal B and HER2+, but no significant difference in TNBC (Table 1). Thus, this set of observations suggests the differential expression of *SPDEF* which allowed the characteristics of the pro- and anti-oncogenic activities in various BC subtype. Future in-depth mechanism governing the regulation of *SPDEF* in BC subtypes will contribute to gain insight into the BC biology and also add a new dimension to the new treatment targets rather than treating BC as a single entity.

Second, the clinicopathologic and prognostic values of *SPDEF* in various BC subtypes have been explored and established. Here, we demonstrated that high *SPDEF* mRNA levels were positively correlated with faster disease progression in Luminal A, and TNBC (Figure 3A-D). In-depth analysis of BC from TCGA databases shows the poor overall survival of *SPDEF* high expression (Figure 3E-H), which merits further investigation to establish whether it is a new prognostic marker for the four BC subtypes. In addition, high transcript level of *SPDEF* was positively related to TNM stage, lymphoid nodal status in Luminal A, with tumour invasion in Luminal B, with lymphoid nodal status in HER2+, and even with TNM stage in TNBC (Table 2). These observations indicated that high levels of *SPDEF* expression promote the BC progression which has distinctive characteristics of subtypes, respectively, laying the foundation for future mechanism research.

Third, this study was the first attempt to predict that *SPDEF* participated in tumorigenesis and progression of BC subtypes by GO analysis, which was involved in the aspects of biological process (BP), cellular component (CC) and molecular function (MF) (Table S1). For Luminal BC, the results demonstrated that the enrichment is mainly concentrated on the mitochondrial respiratory and translational (Figure 4A-B). Consistent with our findings, recent literature sheds light on the contribution of mitochondrial respiration in BC tumorigenesis<sup>25</sup> and metastasis,<sup>26</sup> but lack subtype exploration. And the mitochondrial translational was also demonstrated to be involved in the targeted therapy for leukaemia,<sup>27,28</sup> which deserves further study in the field of BC. In addition, the bifunctional RasGAP tumour suppressor has been proved to be concomitantly suppressed in aggressive luminal B tumours and drive metastasis by activating

RAS signal transduction.<sup>29</sup> Herein, we proposed for the first time that Ras protein signal transduction was closely related to HER2+ BC by GO analysis (Figure 4C), which is worth further exploring through experimental evidences. Meanwhile, related to our analysis of TNBC (Figure 4D), extracellular matrix organization has been reported to participate in the regulation process that GREM1 promotes the invasion and metastasis of ER-negative breast.<sup>30</sup> Hence, as for TNBC, in-depth mechanistic characteristics of cancerogenesis and development referring to extracellular matrix merits further investigation.

Fourth, we have innovatively predicted the potential signalling pathways associated with *SPDEF* in BC subsets via KEGG and hallmark effect gene set analysis. Above mentioned pathways, the most enrichment pathways were referring to the thermogenesis and oxidative phosphorylation pathways in Luminal A group, the thermogenesis and EMT pathways in Luminal B, the MAPK and oxidative phosphorylation pathways in HER2+, the PI3K-Akt and oestrogen response early pathways in TNBC (Figure 5). A recent report indicated the disruption of hypoxia-inducible fatty acid-binding protein 7 induces beige fat-like differentiation and thermogenesis in breast cancer cells, in which the rise in temperature of cancer cells may impact on patients' outcomes.<sup>31</sup> EMT pathways were also proved to be responsible for metastases and therapy resistance in Luminal B type BC.<sup>32,33</sup> Additionally, seldom literature showed the MAPK pathways were involved in the metastasis of HER2+ type BC cells,<sup>34</sup> and mitochondrial oxidative phosphorylation was correlated with the promotion of chemotherapy-resistant BC stem cells.<sup>35</sup> And the evidence from a phase 1 trial verified the targeting of the PI3K/AKT/mTOR pathway for the treatment of mesenchymal TNBC.<sup>36</sup> Noteworthy, there is another study regarding the value of ER $\beta$ -targeted therapies for the treatment of TNBC patients,<sup>37</sup> which was closely correlated and consistent with the oestrogen response early pathways enriched in TNBC in our results. Taken together, *SPDEF* may carry out its regulation functions in such BC subtypes through participation in above signalling pathways. This need to be clarified by further researches.

Fifth, the prognostic risk model of *SPDEF*-related prognosis genes in subtypes of BC has been constructed for the first time, indicating a high prognostic performance in survival surveillance. The *SPDEF*-based prognostic index could be an important tool for distinguishing among various subtyping BC patients based on potential discrete outcomes (Figure 6A-H). Furthermore, this prognostic index can effectively and accurately stratify different subtypes of BC patients, which is vital for monitoring the survival of subtype-specific patients (Figure 6L). And the ROC curves revealed a high predictive value of the risk model (Figure 6M-P). Notably, there were two advantages of using the *SPDEF*-related prognosis genes to construct the prognostic risk model in different subtypes of breast cancer. On the one hand, the influence of confounding factors in the analysis process could be avoided to ensure the inclusion of *SPDEF*-related prognostic genes significantly associated with the survival outcome. On the other hand, the optimum point of the performance parameters was determined, which improved the discrimination ability of the prognostic risk model.

In summary, our findings provide new insights that can guide a more detailed assessment of BC patients in subsequent clinical trials.

In conclusion, the study we presented here indicated that specific expressions and molecular functions of *SPDEF* might lead to the occurrence and development of multiple BC subtypes. Further, high expression of *SPDEF* shows the poor OS and subtype-specific risk model of *SPDEF*-related prognosis genes indicated a high prognostic performance in survival surveillance in various BC. Overall, our findings would help to better understand the possible mechanisms of various BC subtypes and to find possible candidate genes for prognostic and therapeutic usage.

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Youth Fund, China (Grant No. 82003138), the Sichuan Science and Technology Program for key Research and Development, China (Grant No. 2021YFS0226), Doctoral Research Initiation Fund of Affiliated Hospital of Southwest Medical University, China (Grant No.19077) and University's scientific research project of Southwest Medical University, China (Grant No. 2020ZRQNA019).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

**Ting Ye:** Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Project administration (lead); Validation (lead); Visualization (lead); Writing-original draft (lead). **Jingyuan Li:** Data curation (lead); Formal analysis (lead); Validation (lead). **Jia Feng:** Data curation (equal); Formal analysis (equal). **Jinglan Guo:** Data curation (equal); Formal analysis (equal). **Xue Wan:** Data curation (equal); Formal analysis (equal). **Dan Xie:** Data curation (equal); Formal analysis (equal). **Jinbo Liu:** Conceptualization (lead); Funding acquisition (lead); Project administration (lead); Writing-review & editing (lead).

## DATA AVAILABILITY STATEMENT

All data utilized in this study are included in this article, and all data supporting the findings of this study are available on reasonable request from the corresponding authors.

## ORCID

Ting Ye  <https://orcid.org/0000-0002-5704-400X>

## REFERENCES

- Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;70(4):313.
- Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406:747-752.
- Holm J, Eriksson L, Ploner A, et al. Assessment of breast cancer risk factors reveals subtype heterogeneity. *Can Res.* 2017;77:3708-3717.
- Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast (Edinburgh, Scotland).* 2015;24(Suppl 2):S26-35.
- Oettgen P, Finger E, Sun Z, et al. PDEF, a novel prostate epithelium-specific ets transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression. *The Journal of biological chemistry.* 2000;275:1216-1225.
- Buchwalter G, Hickey MM, Cromer A, et al. PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell.* 2013;23:753-767.
- Sood AK, Saxena R, Groth J, et al. Expression characteristics of prostate-derived Ets factor support a role in breast and prostate cancer progression. *Hum Pathol.* 2007;38:1628-1638.
- Sood AK, Wang J, Mhawech-Fauceglia P, Jana B, Liang P, Geradts J. Sam-pointed domain containing Ets transcription factor in luminal breast cancer pathogenesis. *Cancer Epidemiology Biomarkers & Prevention.* 2009;18:1899-1903.
- Gunawardane RN, Sgroi DC, Wrobel CN, Koh E, Daley GQ, Brugge JS. Novel role for PDEF in epithelial cell migration and invasion. *Can Res.* 2005;65:11572-11580.
- Cao L, Xu C, Xiang G, et al. AR-PDEF pathway promotes tumour proliferation and upregulates MYC-mediated gene transcription by promoting MAD1 degradation in ER-negative breast cancer. *Mol Cancer.* 2018;17:136.
- Mukhopadhyay A, Houry T, Stein L, Shrikant P, Sood AK. Prostate derived Ets transcription factor and Carcinoembryonic antigen related cell adhesion molecule 6 constitute a highly active oncogenic axis in breast cancer. *Oncotarget.* 2013;4:610-621.
- Feldman RJ, Sementchenko VI, Gayed M, Fraig MM, Watson DK. Pdef expression in human breast cancer is correlated with invasive potential and altered gene expression. *Can Res.* 2003;63:4626-4631.
- Turner DP, Moussa O, Sauane M, Fisher PB, Watson DK. Prostate-derived ETS factor is a mediator of metastatic potential through the inhibition of migration and invasion in breast cancer. *Can Res.* 2007;67:1618-1625.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-140.
- Kishore J, Goel M, Khanna P. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res.* 2010;1:274-278.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet.* 2000;25:25-29.
- Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research.* 2000;28(1):27-30.
- Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z. CARMAweb: Comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res.* 2006;34:W498-503.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102:15545-15550.
- Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34:267-273.
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417-425.
- Li L, Huang J, Sun S, et al. Selecting pre-screening items for early intervention trials of dementia—a case study. *Stat Med.* 2004;23:271-283.
- Yeo SK, Guan JL. Breast cancer: Multiple subtypes within a tumor? *Trends in cancer.* 2017;3:753-760.
- Ye T, Feng J, Wan X, Xie D, Liu J. Double agent: *SPDEF* gene with both oncogenic and tumor-suppressor functions in breast cancer. *Cancer management and research.* 2020;12:3891-3902.
- Ikeda K, Horie-Inoue K, Suzuki T, et al. Mitochondrial supercomplex assembly promotes breast and endometrial tumorigenesis by metabolic alterations and enhanced hypoxia tolerance. *Nat Commun.* 2019;10:4108.

26. LeBleu VS, O'Connell JT, Gonzalez Herrera KN, et al. PGC-1 $\alpha$  mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nat Cell Biol*. 2014;16(992–1003):1-15.
27. Järås M, Ebert BL. Power cut: Inhibiting mitochondrial translation to target leukemia. *Cancer Cell*. 2011;20:555-556.
28. Vahedi S, Chueh FY, Chandran B, Yu CL. Lymphocyte-specific protein tyrosine kinase (Lck) interacts with CR6-interacting factor 1 (CRIF1) in mitochondria to repress oxidative phosphorylation. *BMC Cancer*. 2015;15:551.
29. Olsen SN, Wronski A, Castaño Z, et al. Loss of RasGAP tumor suppressors underlies the aggressive nature of luminal B breast cancers. *Cancer Discov*. 2017;7:202-217.
30. Neckmann U, Wolowczyk C, Hall M, et al. GREM1 is associated with metastasis and predicts poor prognosis in ER-negative breast cancer patients. *Cell communication and signaling : CCS*. 2019;17:140.
31. Kawashima M, Bensaad K, Zois CE, et al. Disruption of hypoxia-inducible fatty acid binding protein 7 induces beige fat-like differentiation and thermogenesis in breast cancer cells. *Cancer & Metabolism*. 2020;8:13.
32. Li P, Cao G, Huang Y, et al. siMTA1-loaded exosomes enhanced chemotherapeutic effect of Gemcitabine in Luminal-b type breast cancer by inhibition of EMT/HIF- $\alpha$  and autophagy pathways. *Frontiers in oncology*. 2020;10:541262.
33. Punzi S, Balestrieri C, D'Alesio C, et al. WDR5 inhibition halts metastasis dissemination by repressing the mesenchymal phenotype of breast cancer cells. *Breast Cancer Research : BCR*. 2019;21:123.
34. Chen XY, Zhou J, Luo LP, et al. Black rice Anthocyanins suppress metastasis of breast cancer cells by targeting RAS/RAF/MAPK pathway. *Biomed Res Int*. 2015;2015:1-11.
35. Lee KM, Giltane JM, Balko JM, et al. MYC and MCL1 cooperatively promote chemotherapy-resistant breast cancer stem cells via regulation of mitochondrial oxidative phosphorylation. *Cell Metab*. 2017;26:633-47.e7.
36. Basho RK, Gilcrease M, Murthy RK, et al. Targeting the PI3K/AKT/mTOR pathway for the treatment of mesenchymal triple-negative breast cancer: Evidence from a phase 1 trial of mTOR inhibition in combination with liposomal doxorubicin and bevacizumab. *JAMA oncology*. 2017;3:509-515.
37. Reese JM, Bruinsma ES, Nelson AW, et al. ER $\beta$ -mediated induction of cystatins results in suppression of TGF $\beta$  signaling and inhibition of triple-negative breast cancer metastasis. *Proc Natl Acad Sci*. 2018;115:E9580-E9589.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Ye T, Li J, Feng J, et al. The subtype-specific molecular function of *SPDEF* in breast cancer and insights into prognostic significance. *J Cell Mol Med*. 2021;25:7307–7320. <https://doi.org/10.1111/jcmm.16760>